NANZAN-TR-2014-01

The Maximum Asymptotic Bias of Robust Regression Estimates over (c, γ)- Contamination Neighborhoods

Masakazu Ando and Miyoshi Kimura

March 2015

Technical Report of the Nanzan Academic Society
Information Sciences and Engineering

The Maximum Asymptotic Bias of Robust Regression Estimates over (c, γ)- Contamination Neighborhoods

Masakazu Ando [1] and Miyoshi Kimura [2]

Chiba Institute of Tecnology and Nanzan University

Abstract. When the observations may be contaminated in the linear model with the intercept, a certain large class of robust regression estimates including S-estimates, τ-estimates and CM-estimates is considered. The (c, γ)-contamination neighborhood, which is a generalization of the neighborhoods defined in terms of ε-contamination and total variation, is used for describing the contamination of the observations. Lower and upper bounds for the maximum asymptotic bias of the regression estimates over (c, γ)-contamination neighborhoods are derived without imposing elliptical regressors. As important special cases, the lower and upper bounds for S-estimates, τ-estimates and CM-estimates under the Gaussian model are obtained.

AMS 2000 Subject classifications: Primary 62F35, Secondary 62H12, 62J05

Key words: maximum asymptotic bias, robust regression estimates, S-estimates, τ-estimates, CM-estimates, (c, γ)-contamination neighborhoods

1. Introduction

In the case where the observations may be contaminated in the location model, Huber (1964) introduced the maximum asymptotic bias $B_T(\varepsilon)$ of a location estimate T over the ε-contamination neighborhood. The $B_T(\varepsilon)$ is one of the most informative global quantitative measures to assess robustness of T, because $B_T(\varepsilon)$ shows the whole performance of T from $\varepsilon = 0$ (the central model distribution) to the breakdown point and its derivative $B'_T(0)$ equals the gross error sensitivity under some regularity conditions. Huber (1964) established that the median minimizes $B_T(\varepsilon)$ among translation equivariant location estimates. Martin and Zamer (1989, 1993) obtained minimax bias robust scale estimates. Adrover (1998) derived minimax bias robust dispersion matrix estimates.

As for the linear regression model, in the case of the zero-intercept and elliptical regressors, Martin, Yohai and Zamer (1989) obtained the minimax bias estimates in the respective classes of M-estimates with general scale and GM- estimates of regression. In particular, they showed that the least median of square estimate (LMS) introduced by

[1] Department of Risk Science in Finance and Management, Chiba Institute of Technology, 2-17-1 Tsudanuma, Narashino, Chiba, 275-0016, JAPAN. E-mail address: andomasa@sun.it-chiba.ac.jp

[2] Department of Systems and Mathematical Science, Nanzan University, 27 Seirei-cho, Seto, Aichi, 489-0863, JAPAN. E-mail address: kimura@ms.nanzan-u.ac.jp

Rousseeuw (1984) is nearly minimax. Yohai and Zamer (1993) extended this result to the larger class of residual admissible estimates. Berrendero and Zamer (2001) obtained maximum asymptotic bias of robust regression estimates in a broad class, which includes S-estimates, τ-estimates and R-estimates, without requiring zero-intercept and/or elliptical regressors. Berrendero, Mendes and Tyler (2007) derived the maximum asymptotic bias of MM-estimates and the constrained M-estimates (CM-estimates) of regression, and compared them to those of S-estimates and τ -estimates in detail. All the authors mentioned above adopt the ε-contamination neighborhood to describe deviation from the central model.

On the other hand, in order to describe deviation from the central model Ando and Kimura (2003) introduced the (c, γ) - contamination neighborhood (the (c, γ) - neighborhood, for short) , which is a generalization of Rieder's (ε, δ) -neighborhood and includes the neighborhoods defined in terms of ε - contamination and total variation. They gave a characterization of the (c, γ) - neighborhood and applied it to bias-robustness study of estimates. Among their achievements, there are the extensions of Huber's (1964) and He and Simpson's (1993) results. The former states that the median minimizes the maximum asymptotic bias $B_T(c, \gamma)$ over (c, γ) - neighborhoods among translation equivariant location estimates. Ando and Kimura (2004) derived the lower and upper bounds for $B_S(c, \gamma)$ of regression S-estimates over (c, γ) - neighborhoods in the zero-intercept linear model with elliptical regressors, and showed that in the case of Rieder's (ε, δ) - neighborhood the lower and upper bounds coincide and become $B_S(c, \gamma)$. Ando, Kakiuchi and Kimura (2009) gave the applications of the (c, γ) - neighborhoods to nonparametric confidence intervals and tests for the median.

In this paper, following Berrendero and Zamer (2001), without imposing the zero-intercept and/or elliptical regressors , we derive the lower and upper bounds for $B_T(c, \gamma)$ of estimates T in the large class. In the case of ε-contamination neighborhoods, the lower and upper bounds coincide and the results are reduced to Theorems 1 and 2 of Berrendero and Zamar (2001). As important special cases, we obtain the lower and upper bounds for the maximum asymptotic bias $B_S(c, \gamma)$, $B_\tau(c, \gamma)$ and $B_{CM}(c, \gamma)$ of S-estimates, τ-estimates and CM-estimates under the Gaussian model. We give some tables of the lower and upper bounds for $B_\tau(c, \gamma)$ of τ-estimates based on Huber score function and show a selective graph to visualize the difference between the two bounds. We should emphasize that the characterization (Proposition 2.1) of the (c, γ)-neighborhoods is indispensable to the derivation of our results in the paper.

The paper is organized as follows. Section 2 presents basic definitions and preliminary results. Section 3 gives the lower and upper bounds for $B_T(c, \gamma)$ which are our main results. Section 4 shows the lower and upper bounds for $B_S(c, \gamma)$, $B_\tau(c, \gamma)$ and $B_{CM}(c, \gamma)$ of under the Gaussian model. The tables and a graph of the lower and upper bounds for $B_\tau(c, \gamma)$ of the τ-estimate based on Huber score functions are also exhibited. All the proofs of lemmas and theorems are collected in section 5.

2. Preliminaries

We consider the linear regression model

$$y \quad = \quad \alpha_0 + \boldsymbol{\theta}_0' \boldsymbol{x} + u,$$

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, α_0 is the true intercept parameter in R and the error u is a random variable independent of $\boldsymbol{x}$. Let F_0 be the nominal distribution function of u and G_0 the nominal distribution function of $\boldsymbol{x}$. Then the nominal distribution function H_0 of $(y, \boldsymbol{x})$ is

$$H_0(y, \boldsymbol{x}) = \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_p} F_0(y - \alpha_0 - \boldsymbol{\theta}_0' \boldsymbol{s}) dG_0(\boldsymbol{s}). \tag{2.1}$$

Let $\mathcal{M}$ be the set of all distribution functions 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 a R^p-valued functional defined on $\mathcal{M}$. Given a sample of independent observations $(y_1, \boldsymbol{x}_1), \cdots, (y_n, \boldsymbol{x}_n)$ 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.

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

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

where $\boldsymbol{A}$ is an affine equivariant covariance functional of $\boldsymbol{x}$ under G_0. Since we only work with regression and affine equivariant estimates and $b_{\boldsymbol{A}}(\boldsymbol{T}, H)$ is invariant under regression and affine transformations, we can assume without loss of generality that $\boldsymbol{\theta}_0 = \boldsymbol{0}$ and $\boldsymbol{A} = \boldsymbol{I}_p$ (the identity matrix). Therefore the asymptotic bias $b_{\boldsymbol{A}}(\boldsymbol{T}, H)$ is given by

$$b(\boldsymbol{T}, H) = \|\boldsymbol{T}(H)\|, \tag{2.2}$$

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

In order to describe deviation from the nominal distribution H_0 we adopt the following neighborhood of H_0, which was introduced by Ando and Kimura (2003):

$$\mathcal{P}_{H_0}(c, \gamma) = \{H \in \mathcal{M} \ : \ H(B) \leq c\, H_0(B) + \gamma, {}^{\forall} B \in \mathcal{B}^{p+1}\}, \tag{2.3}$$

where $0 \leq \gamma < 1$ and $1 - \gamma \leq c < \infty$. Note that $H_0(H)$ is used as both a distribution function and a probability measure for convenience. The neighborhood $\mathcal{P}_{H_0}(c, \gamma)$, which is called a (c, γ) -contamination neighborhood ((c, γ)-neighborhood, for short), is a generalization of Rieder's (1977) (ε, δ)-neighborhood and includes ε-contamination and total variation neighborhoods: Let ε and δ be some given constants such that $\varepsilon \geq 0, \delta \geq 0$ and $\varepsilon + \delta < 1$. Then we have the ε-contamination neighborhood $\mathcal{P}_{H_0}(1 - \varepsilon, \varepsilon)$ for $c = 1 - \varepsilon$ and $\gamma = \varepsilon$, the total variation neighborhood $\mathcal{P}_{H_0}(1, \delta)$ for $c = 1$ and $\gamma = \delta$, and Rieder's (1977) (ε, δ)-neighborhood $\mathcal{P}_{H_0}(1 - \varepsilon,\ \varepsilon + \delta)$ for $c = 1 - \varepsilon$ and $\gamma = \varepsilon + \delta$. We can see that $\mathcal{P}_{H_0}(c, \gamma)$ is generated by a special capacity v defined as

$$v(B) = \begin{cases} \min\{cH_0(B) + \gamma,\ 1\} & \text{if } \ B \neq \phi, \quad B \in \mathcal{B}^{p+1}, \\ 0, & \text{if } \ B = \phi. \end{cases}$$

This means that the (c,γ)-neighborhood $\mathcal{P}_{H_0}(c,\gamma)$ has nice properties for developing minimax theory in robust inference (see Bednarski, 1981). Ando and Kimura (2003) gave the following useful characterization of $\mathcal{P}_{H_0}(c,\gamma)$.

Proposition 2.1. *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 $\boldsymbol{T}$ over $\mathcal{P}_{H_0}(c,\gamma)$ is defined as

$$B_{\boldsymbol{T}}(c,\gamma) = \sup\{\|\boldsymbol{T}(H)\|\ :\ H \in \mathcal{P}_{H_0}(c,\gamma)\}. \tag{2.4}$$

We consider the following class of robust estimates defined as

$$(T_0(H), \boldsymbol{T}(H)) = \arg\min_{\alpha,\boldsymbol{\theta}} J(F_{H,\alpha,\boldsymbol{\theta}}), \tag{2.5}$$

where $J(\cdot)$ is a robust loss functional defined on the set of all distributions on the real line and $F_{H,\alpha,\theta}$ is the distribution of the absolute residual $|y-\alpha-\boldsymbol{\theta}'\boldsymbol{x}|$ under H. This class of estimates includes the well-known robust estimates such as S-estimates, τ-estimates, CM-estimates and R-estimates. We assume that J, F_0 and G_0 satisfy the following conditions A1 and A2 corresponding to Berrendero and Zamar (2001).

Let $\mathcal{L}^+$ be the set of all distributions on $[0,\infty)$ and let $\mathcal{L}_c^+$ be the subset of $\mathcal{L}^+$ of all continuous distributions on $(0,\infty)$.

A1. (a) Let $F \in \mathcal{L}^+$ and $G \in \mathcal{L}^+$. If $F(v) \le G(v)$ $(F(v) < G(v))$ for every $v \ge 0$, then $J(F) \ge J(G)$ $(J(F) > J(G))$.

(b) Let $\{F_n\}$ and $\{G_n\}$ be sequences of $F_n \in \mathcal{L}_c^+$ and $G_n \in \mathcal{L}_c^+$ $(n = 1, 2, \cdots)$ such that $F_n(v) \to F(v)$ and $G_n(v) \to G(v)$, where F and G are possibly sub-stochastic and continuous on $(0,\infty)$ with $G(\infty) \ge 1-\gamma$. If $G(v) \ge F(v)$ $(G(v) > F(v))$ for every $v > 0$, then $\lim_{n\to\infty} J(F_n) \ge \lim_{n\to\infty} J(G_n)$ $(\lim_{n\to\infty} J(F_n) > \lim_{n\to\infty} J(G_n))$.

(c) If $F \in \mathcal{L}_c^+$ and $G \in \mathcal{L}^+$, then

$$J((1-\gamma)\,F + \gamma\delta_\infty) = \lim_{n\to\infty} J((1-\gamma)\,F + \gamma U_n) \ge J((1-\gamma)\,F + \gamma G),$$

where U_n stands for the uniform distribution function on $[n-\frac{1}{n}, n+\frac{1}{n}]$.

A2. F_0 has an even and strictly unimodal density f_0 with $f_0(v) > 0$ for every $v \in R$, and $P_{G_0}(\boldsymbol{\theta}'\boldsymbol{x} = a) < 1$, for every $\boldsymbol{\theta} \in R^p$ $(\boldsymbol{\theta} \ne 0)$ and $a \in R$.

Remark 2.1 The ε-monotonicity condition A1(b) guarantees that the corresponding estimate $\boldsymbol{T}$ is residual admissible (see Yohai and Zamar, 1993, for the definition of residual admissible estimates). We should emphasize that A2 does not require ellipticity nor continuity of regressor's distribution.

3. Main results

We give three lemmas and two main theorems , whose proofs are collected in Section 5. First we introduce the family of measures (improper distributions) . Let $\xi = \{W_{\alpha,\boldsymbol{\theta}} : \alpha \in R, \boldsymbol{\theta} \in R^p\}$ be a family of $W_{\alpha,\boldsymbol{\theta}} \in \mathcal{W}_{H_0,\lambda}$ and let

$$F^{\xi}_{\alpha,\boldsymbol{\theta}}(v) = (H_0 - W_{\alpha,\boldsymbol{\theta}})(|y - \alpha - \boldsymbol{\theta}'\boldsymbol{x}| \le v), \quad {}^{\forall}v \ge 0. \tag{3.1}$$

Also, let

$$d_{\xi} = J(c\,F^{\xi}_{0,\mathbf{0}} + \gamma\delta_{\infty}) \tag{3.2}$$

and

$$m_{\xi}(t) = \inf_{\|\boldsymbol{\theta}\|=t} \inf_{\alpha\in R} J(c\,F^{\xi}_{\alpha,\boldsymbol{\theta}} + \gamma\delta_0), \tag{3.3}$$

where δ_0 and δ_{∞} are the point mass distributions at 0 and ∞, respectively. Note that $F^{\xi}_{\alpha,\boldsymbol{\theta}}$ is used as both distribution function and measure on $(R, \mathcal{B})$.

We consider the following conditions of $F^{\xi}_{\alpha,\boldsymbol{\theta}}$:

A3. (a) $F^{\xi}_{k\alpha,k\boldsymbol{\theta}}(v)$ is strictly decreasing in $k > 0$ for $0 < F^{\xi}_{k\alpha,k\boldsymbol{\theta}}(v) < (1-\gamma)/c$.

(b) $F^{\xi}_{\alpha,\boldsymbol{\theta}}$ satisfies $0 < F^{\xi}_{\alpha,\boldsymbol{\theta}}(v) \le F^{\xi}_{0,\mathbf{0}}(v), \ {}^{\forall}v > 0$.

Let two families $\hat{\xi} = \{\hat{W}_{\alpha,\boldsymbol{\theta}}\}$ and $\xi^* = \{W^*_{\alpha,\boldsymbol{\theta}}\}$ be defined as follows:

$$\hat{W}_{\alpha,\boldsymbol{\theta}}(B) = H_0\left(B \cap \left\{|y - \alpha - \boldsymbol{\theta}'\boldsymbol{x}| \ge a_{\alpha,\boldsymbol{\theta}}\left(\tfrac{c+\gamma-1}{c}\right)\right\}\right), \quad {}^{\forall}B \in \mathcal{B}^{p+1} \tag{3.4}$$

$$W^*_{\alpha,\boldsymbol{\theta}}(B) = H_0\left(B \cap \left\{|y - \alpha - \boldsymbol{\theta}'\boldsymbol{x}| \le a_{\alpha,\boldsymbol{\theta}}\left(\tfrac{1-\gamma}{c}\right)\right\}\right), \quad {}^{\forall}B \in \mathcal{B}^{p+1}, \tag{3.5}$$

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

$$H_0\left(|y - \alpha - \boldsymbol{\theta}'\boldsymbol{x}| \ge a_{\alpha,\boldsymbol{\theta}}(\eta)\right) = \eta.$$

Let $\mathcal{F}_\lambda$ be the set of all $\xi = \{W_{\alpha,\boldsymbol{\theta}} : \alpha \in R, \boldsymbol{\theta} \in R^p\}$ satisfying A3. Here we note that $\hat{\xi}$ belongs to $\mathcal{F}_\lambda$ but ξ^* does not belong to $\mathcal{F}_\lambda$ (ξ^* does not satisfy A3(b)). Then we obtain the following three lemmas, which correspond to Lemmas 4, 5 and 6 in Berrendero and Zamar (2001).

Lemma 3.1. *Under A1(b) and A2, for any $\xi = \{W_{\alpha,\boldsymbol{\theta}}\} \in \mathcal{F}_\lambda$ there exists $\alpha(\boldsymbol{\theta}) \in R$ such that*

$$J(c\,F^{\xi}_{\alpha(\boldsymbol{\theta}),\boldsymbol{\theta}} + \gamma\delta_0) = \inf_{\alpha \in R} J(c\,F^{\xi}_{\alpha,\boldsymbol{\theta}} + \gamma\delta_0).$$

Moreover, for any $t > 0$ there exists $K_t > 0$ such that $|\alpha(\boldsymbol{\theta})| \le K_t$ for every $\boldsymbol{\theta} \in \{\boldsymbol{\theta} : \|\boldsymbol{\theta}\| = t\}$.

Lemma 3.2. *Under A2, $F^{\hat{\xi}}_{k\alpha,k\boldsymbol{\theta}}(v)$ and $F^{\xi^*}_{k\alpha,k\boldsymbol{\theta}}(v)$ are strictly decreasing in $k > 0$ for $0 < F^{\hat{\xi}}_{k\alpha,k\boldsymbol{\theta}}(v) < (1-\gamma)/c$ and $0 < F^{\xi^*}_{k\alpha,k\boldsymbol{\theta}}(v) < (1-\gamma)/c$, respectively.*

Lemma 3.3. *Let $m_\xi(t)$ be as in (3.3). Then, under A1(b) and A2, for any $\xi = \{W_{\alpha,\boldsymbol{\theta}}\} \in \mathcal{F}_\lambda$ the following results hold:*

(a) *There exist $\boldsymbol{\theta}_t \in R^p$ and $\alpha(\boldsymbol{\theta}_t) \in R$ such that $\|\boldsymbol{\theta}_t\| = t$ and $m_\xi(t) = J(c\,F^{\xi}_{\alpha(\boldsymbol{\theta}_t),\boldsymbol{\theta}_t} + \gamma\delta_0)$.*

(b) *$m_\xi(t)$ is strictly increasing.*

Using these lemmas, we can derive the following theorem which gives lower and upper bounds for the maximum asymptotic bias $B_{\boldsymbol{T}}(c,\gamma)$ of $\boldsymbol{T}$.

Theorem 3.1. *Let $\boldsymbol{T}$ be a regression estimate defined by (2.5). Assume A1 and A2. Then it holds that*

$$\underline{B}_{\boldsymbol{T}}(c,\gamma) \le B_{\boldsymbol{T}}(c,\gamma) \le \overline{B}_{\boldsymbol{T}}(c,\gamma), \tag{3.6}$$

where

$$\underline{B}_{\boldsymbol{T}}(c,\gamma) = \sup_{\xi \in \mathcal{F}_\lambda} m_\xi^{-1}(d_\xi) \quad \textit{and} \quad \overline{B}_{\boldsymbol{T}}(c,\gamma) = m_{\hat{\xi}}^{-1}(d_{\xi^*})$$

.

The function $m_\xi(t)$ is simplified under symmetry and unimodality assumptions on the regressors distribution.

Theorem 3.2. *Let $\boldsymbol{T}$ be a regression estimate defined by (2.5). Assume A1 and A2, and that under G_0 the distribution of $\boldsymbol{\theta}'\boldsymbol{x}$ is symmetric, unimodal and only depends on $\|\boldsymbol{\theta}\|$ for all $\boldsymbol{\theta} \neq \mathbf{0}$. Then, for any $\xi = \{W_{\alpha,\boldsymbol{\theta}}\} \in \mathcal{F}_\lambda$ it holds that*

$$\inf_{\alpha \in R} J(c\,F^{\xi}_{\alpha,\boldsymbol{\theta}} + \gamma\delta_0) = J(c\,F^{\xi}_{0,\boldsymbol{\theta}} + \gamma\delta_0) = m_\xi(\|\boldsymbol{\theta}\|).$$

and the inequlities (3.6) are simplified.

Remark 3.1 When $c = 1 - \varepsilon$ and $\gamma = \varepsilon$ (i.e., the ε-contamination case) , Theorems 3.1 and 3.2 reduce to Theorems 1 and 2 of Berrendero and Zamar (2001), respectively. In this case, we have $\lambda = 0$ and $\hat{\xi} = \xi^*$, and hence $\overline{B}_{\boldsymbol{T}}(c,\,\gamma) = \underline{B}_{\boldsymbol{T}}(c,\,\gamma)$.

4. S-, τ-and CM-estimates under the Gaussian model

As important special cases, we consider S-estimates, τ-estimates and CM-estimates in the case that H_0 is the multivariate normal distribution $N(\mathbf{0}, \boldsymbol{I}_{p+1})$ with mean vector $\mathbf{0}$ and covariance matrix $\boldsymbol{I}_{p+1}$. We denote by ϕ the density of the standard normal distribution $N(0,\ 1)$. For any $\xi = \{W_{\alpha,\boldsymbol{\theta}}\} \in \mathcal{F}_\lambda$ let $\varphi^{\xi}_{\alpha,\boldsymbol{\theta}}$ denote the density of $F^{\xi}_{\alpha,\boldsymbol{\theta}}$. Let $\mathcal{F}^{\circ}_\lambda$ be the set of all $\xi = \{W_{\alpha,\boldsymbol{\theta}}\} \in \mathcal{F}_\lambda$ such that $\varphi^{\xi}_{0,\boldsymbol{\theta}}$ is expressed in the form of

$$\varphi^{\xi}_{0,\boldsymbol{\theta}}(v) = \frac{1}{\sqrt{1+\|\boldsymbol{\theta}\|^2}}\phi_\xi\left(\frac{v}{\sqrt{1+\|\boldsymbol{\theta}\|^2}}\right), \quad {}^{\forall}v \geq 0, \tag{4.1}$$

where ϕ_ξ $(0 \leq \phi_\xi \leq 2\phi)$ is some measurable function defined on $[0, \infty)$ such that

$$\int_0^\infty \phi_\xi(v)dv = \frac{1-\gamma}{c}.$$

We can easily see that $\hat{\xi} = \{\hat{W}_{\alpha,\boldsymbol{\theta}}\}$ belongs to $\mathcal{F}^{\circ}_\lambda$.

Let ρ_1 and ρ_2 be score functions satisfying the following conditions:

A4. (a) The functions ρ_1 and ρ_2 are even, bounded, monotone on $[0, \infty)$, continuous at 0 with $0 = \rho_i(0) < \rho_i(\infty) = 1,\ \ i = 1, 2$ and with at most a finite number of discontinuities.

(b) The function ρ_2 is differentiable with $2\rho_2(v) - \rho_2'(v)v \geq 0$.

The S-estimate (Rousseeuw and Yohai,1984) is defined with $J(F) = S(F)$, where

$$S(F) \;=\; \inf\left\{s > 0 \;:\; E_F\left[\rho_1\left(\frac{v}{s}\right)\right] \leq b\right\}, \quad 0 < b < 1. \tag{4.2}$$

For any $\xi = \{W_{\alpha,\boldsymbol{\theta}}\} \in \mathcal{F}_\lambda$ let

$$g_{\xi,i}(s) = E_{F^{\xi}_{0,\mathbf{0}}}\left[\rho_i\left(\frac{v}{s}\right)\right] = \int_0^{\infty} \rho_i\left(\frac{v}{s}\right)\varphi^{\xi}_{0,\mathbf{0}}(v)dv, \;\; i = 1, 2.$$

The following theorem gives the lower and upper bounds for the maximum asymptotic bias $B_S(c,\gamma)$ of S-estimates based on ρ_1.

Theorem 4.1. *Assume that H_0 is the multivariate normal distribution $N(\mathbf{0}, \boldsymbol{I}_{p+1})$. Then*

$$\begin{aligned} &\underline{B}_S(c,\,\gamma) \leq B_S(c,\,\gamma) \leq \overline{B}_S(c,\,\gamma), && \text{if} \quad \gamma < \min(b, 1-b), \\ &B_S(c,\,\gamma) = \infty, && \text{if} \quad \gamma \geq \min(b, 1-b), \end{aligned}$$

where

$$\underline{B}_S(c,\,\gamma) \;=\; \sup_{\xi \in \mathcal{F}^{\circ}_{\lambda}} \left(\left\{g^{-1}_{\xi,1}\left(\frac{b-\gamma}{c}\right) \Big/ g^{-1}_{\xi,1}\left(\frac{b}{c}\right)\right\}^2 - 1\right)^{1/2} \tag{4.3}$$

and

$$\overline{B}_S(c,\,\gamma) \;=\; \left(\left\{g^{-1}_{\xi^*,1}\left(\frac{b-\gamma}{c}\right) \Big/ g^{-1}_{\hat{\xi},1}\left(\frac{b}{c}\right)\right\}^2 - 1\right)^{1/2}. \tag{4.4}$$

The τ-estimate (Yohai and Zamar, 1988) is defined with $J(F) = \tau^2(F)$, where

$$\tau^2(F) = S^2(F) E_F\left[\rho_2\left(\frac{v}{S(F)}\right)\right]. \tag{4.5}$$

As shown in Yohai and Zamar (1988), τ-estimates inherit the breakdown point of the initial S-estimate defined by ρ_1 and their efficiencies are mainly determined by ρ_2. The following theorem gives the lower and upper bounds for the maximum asymptotic bias $B_\tau(c,\gamma)$ of τ-estimates which shows how $B_\tau(c,\gamma)$ relates to the maximum asymptotic bias $B_S(c,\gamma)$ of the initial S-estimates based on ρ_1.

Theorem 4.2. *Assume that H_0 is the multivariate normal distribution $N(\mathbf{0},\boldsymbol{I}_{p+1})$. Then*

$$\underline{B}_\tau(c,\,\gamma) \le B_\tau(c,\,\gamma) \le \overline{B}_\tau(c,\,\gamma),$$

where

$$\underline{B}_\tau(c,\gamma) = \sup_{\xi\in\mathcal{F}_\lambda^\circ}\left\{\left[\frac{g_{\xi,1}^{-1}\left(\frac{b-\gamma}{c}\right)}{g_{\xi,1}^{-1}\left(\frac{b}{c}\right)}\right]^2 H_{\xi,\xi}(c,\gamma)-1\right\}^{1/2}, \tag{4.6}$$

$$\overline{B}_\tau(c,\,\gamma) = \{[1+\overline{B}_S^2(c,\gamma)]H_{\xi^*,\hat{\xi}}(c,\gamma)-1\}^{1/2}, \tag{4.7}$$

$$H_{\xi_1,\xi_2}(c,\gamma) = \left[\overline{g}_{\xi_1}\left(\frac{b-\gamma}{c}\right)+\frac{\gamma}{c}\right]/\overline{g}_{\xi_2}\left(\frac{b}{c}\right) \quad \textit{and} \quad \overline{g}_\xi(t) = g_{\xi,2}[g_{\xi,1}^{-1}(t)].$$

The CM estimate (Mendes and Tyler, 1996) is defined by $J(F)=CM(F)$, where

$$CM(F) = \inf_{s\ge S(F)}\left\{aE_F\left[\rho_1\left(\frac{v}{s}\right)\right]+\log s\right\}, \tag{4.8}$$

a is a tuning constant, and $S(F)$ is given by (4.2). We let

$$\kappa_{c,\gamma}^\xi = g_{\xi,1}^{-1}\left(\frac{b-\gamma}{c}\right), \quad \eta_{c,\gamma}^\xi = g_{\xi,1}^{-1}(\frac{b}{c}).$$

and

$$A_{a,c,\gamma}^\xi(s) = acg_{\xi,1}(s)+\log s. \tag{4.9}$$

The following theorem gives the lower and upper bounds for the maximum asymptotic bias $B_{CM}(c,\gamma)$ of CM-estimates based on ρ_1.

Theorem 4.3. *Assume that H_0 is the normal distribution $N(\mathbf{0},\boldsymbol{I}_{p+1})$. Then*

$$\underline{B}_{CM}(c,\,\gamma) \le B_{CM}(c,\,\gamma) \le \overline{B}_{CM}(c,\,\gamma),$$

where

$$\underline{B}_{CM}(c,\,\gamma) \;=\; \sup_{\xi\in\mathcal{F}_\lambda^\circ}\{\exp[2a\gamma+2D_{\xi,\xi,a}(c,\gamma)]-1\}^{1/2}\,, \tag{4.10}$$

$$\overline{B}_{CM}(c,\,\gamma) \;=\; \{\exp[2a\gamma+2D_{\xi^*,\hat{\xi},a}(c,\gamma)]-1\}^{1/2}, \tag{4.11}$$

$$D_{\xi_1,\xi_2,a}(c,\gamma) = \inf_{s\geq\kappa_{c,\gamma}} A^{\xi 1}_{a,c,\gamma}(s) - \inf_{s\geq\eta_{c,\gamma}} A^{\xi_2}_{a,c,\gamma}(s).$$

Remark 4.1. When $c = 1-\varepsilon$ and $\gamma=\varepsilon$, Theorems 4.1, 4.2 and 4.3 are reduced to (3.24) of Martin et al. (1989), Theorem 3 of Berrendero and Zamar (2001) and Theorem 4.1 of Bererendero et al. (2007), respectively.

Remark 4.2. The upper bound $\overline{B}_S(c,\gamma)$ in (4.4) is the same as (4.7) in Ando and Kimura (2004). Note that $h_\xi(\tau)$ in the equality (4.7) satisfies the relation $h_\xi(\tau) = g_{\xi,1}(\frac{1}{\tau})$. We notice that when ρ_1 is a jump function, $\overline{B}_S(c,\gamma) = B_S(c,\gamma)$ holds for $c \leq 1$ (see Theorem 4.1 of Ando and Kimura, 2004).

Remark 4.3. The arguments concerning the intercept estimates can be seen in Section 7 of Berrendero and Zamar (2001). Here, we should point out that the same arguments also hold for our (c,γ)-neighborhood case.

Table 1 exhibits the upper bounds $\overline{B}_\tau(c,\gamma)$ for the τ-estimate based on Huber score functions $\rho_1=\rho_H$ with $c_H = 1.041$ and $\rho_2=\rho_H$ with $c_H = 2.832$, where $\rho_H(v) = \min\{(v/c_H)^2, 1\}$. The constants c_H are chosen so that the τ-estimate has 95% efficiency and 0.5 breakdown point (i.e., b=0.5). We have $\overline{B}_\tau(1-\gamma,\gamma) = B_\tau(1-\gamma,\gamma)$ for γ -contamination case. When $\rho_1=\rho_2$, τ-estimates reduce to S-estimates. See Ando and Kimura (2004) for the values of $B_S(c,\gamma)$. On the other hand, it is difficult to find the exact values of the lower bounds $\underline{B}_\tau(c,\gamma)$. In order to obtain their good approximate values we need to find $\xi\in\mathcal{F}_\lambda^\circ$ which makes the inside of the supremum in (4.6) as large as possible. Such ξ depends on c and γ. Table 2 presents lower bounds of $\underline{B}_\tau(c,\gamma)$. We obtained their bounds using the set $\{\xi_1,\cdots,\xi_{12}\}$ which consists of various types of ξ given below. Acoording to (4.1), we define $\phi_{\xi_0},\cdots,\phi_{\xi_{12}}$ as follows: Let

$$\phi_{\xi_i}(v) = \begin{cases} 2\phi(v) & \text{if} \quad 0\leq v< a_i \\ 0 & \text{if} \quad a_i\leq v< a_{i+1} \\ 2\phi(v) & \text{if} \quad a_{i+1}\leq v<\infty \end{cases}$$

where

$$a_i = \Phi^{-1}\left(\frac{1}{2}+id\right), \qquad d=\frac{1-\gamma}{2c}, \qquad \text{i=0, 1, } \cdots \text{, 10,}$$

and Φ denotes the distribution function of $N(0,1)$. Let

$$\phi_{\xi_{11}}(v) = \begin{cases} 2\phi(a) & \text{if} \quad 0 \le v < a \\ 2\phi(v) & \text{if} \quad a \le v < \infty, \end{cases}$$

where a is the constant such that

$$\Phi(a) - \phi(a) = \frac{2c+\gamma-1}{2c},$$

and let $\phi_{\xi_{12}}(v) = \dfrac{2(1-\gamma)}{c}\phi(v),\ v \ge 0$. The lower bounds in Table 2 were obtained from

$$\max_{1\le i\le 12}\left\{\left[\frac{g^{-1}_{\xi_i,1}\left(\frac{0.5-\gamma}{c}\right)}{g^{-1}_{\xi_i,1}\left(\frac{0.5}{c}\right)}\right]^2 H_{\xi_i,\xi_i}(c,\gamma) - 1\right\}^{1/2}. \tag{4.12}$$

Here we note that $\xi_0 = \xi^*$, $\xi_{10} = \hat{\xi}$ and ξ^* is not a member of $\mathcal{F}^{\circ}_{\lambda}$. For the purpose of getting good approximated values, we included different types of ξ as candidates in taking the maximum values (ξ^* hardly affects the maximum values).

Figure 1 gives a graph of $\bar{B}_\tau(c,\gamma)$ and the lower bounds in Table 2 for c=1.2. This graph shows that the lower and upper bounds of $B_\tau(c,\gamma)$ are useful .

Table 1: $\overline{B}_\tau(c,\ \gamma)$ (Huber score function)

$c \setminus \gamma$	0.00	0.01	0.02	0.03	0.05	0.10	0.15	0.20	0.25	0.35	0.45
0.55	—	—	—	—	—	—	—	—	—	—	20.22
0.65	—	—	—	—	—	—	—	—	—	5.05	26.03
0.75	—	—	—	—	—	—	—	—	2.63	6.54	32.17
0.80	—	—	—	—	—	—	—	2.00	3.11	7.31	35.35
0.85	—	—	—	—	—	—	1.52	2.44	3.62	8.09	38.58
0.90	—	—	—	—	—	1.10	1.92	2.87	4.13	8.89	41.87
0.95	—	—	—	—	0.71	1.46	2.27	3.29	4.64	9.70	45.20
0.97	—	—	—	0.53	0.86	1.58	2.40	3.45	4.84	10.03	46.55
0.98	—	—	0.42	0.62	0.93	1.64	2.47	3.53	4.94	10.19	47.22
0.99	—	0.29	0.53	0.70	0.99	1.70	2.54	3.61	5.04	10.35	47.90
1.00	0.00	0.42	0.61	0.76	1.04	1.76	2.60	3.69	5.14	10.52	48.58
1.10	0.84	0.97	1.11	1.24	1.52	2.29	3.25	4.48	6.14	12.19	55.47
1.20	1.20	1.34	1.48	1.63	1.93	2.80	3.88	5.28	7.15	13.89	62.50
1.50	2.08	2.28	2.47	2.67	3.10	4.30	5.79	7.69	10.22	19.16	84.36
2.00	3.47	3.77	4.07	4.38	5.02	6.84	9.06	11.87	15.56	28.42	123.01
3.00	6.29	6.83	7.37	7.91	9.05	12.18	15.96	20.69	26.87	48.10	204.65
5.00	12.34	13.38	14.43	15.50	17.69	23.69	30.82	39.70	51.23	90.60	381.21

Table 2: Lower bouunds of $\underline{B}_\tau(c,\ \gamma)$ (Huber score function)

$c \setminus \gamma$	0.00	0.01	0.02	0.03	0.05	0.10	0.15	0.20	0.25	0.35	0.45
0.55	—	—	—	—	—	—	—	—	—	—	20.22
0.65	—	—	—	—	—	—	—	—	—	5.05	25.45
0.75	—	—	—	—	—	—	—	—	2.63	5.82	31.09
0.80	—	—	—	—	—	—	—	2.00	2.71	6.27	33.97
0.85	—	—	—	—	—	—	1.52	2.06	2.80	6.75	36.90
0.90	—	—	—	—	—	1.10	1.58	2.10	2.92	7.24	39.86
0.95	—	—	—	—	0.71	1.16	1.62	2.16	3.04	7.75	42.86
0.97	—	—	—	0.53	0.72	1.18	1.63	2.18	3.09	7.96	44.07
0.98	—	—	0.42	0.53	0.73	1.18	1.64	2.19	3.11	8.06	44.68
0.99	—	0.29	0.43	0.54	0.74	1.19	1.64	2.20	3.14	8.17	45.29
1.00	0.00	0.30	0.43	0.55	0.75	1.20	1.65	2.22	3.16	8.27	45.90
1.10	0.00	0.32	0.46	0.58	0.78	1.23	1.68	2.36	3.47	9.36	52.07
1.20	0.00	0.33	0.48	0.60	0.80	1.25	1.76	2.58	3.87	10.47	58.35
1.50	0.00	0.35	0.50	0.62	0.83	1.40	2.21	3.44	5.27	13.93	77.81
2.00	0.00	0.36	0.52	0.71	1.02	1.96	3.29	5.11	7.72	19.99	111.85
3.00	0.00	0.53	0.84	1.13	1.74	3.40	5.59	8.62	12.92	32.80	184.68
5.00	0.00	1.11	1.72	2.26	3.34	6.41	10.50	16.12	24.03	60.15	341.05

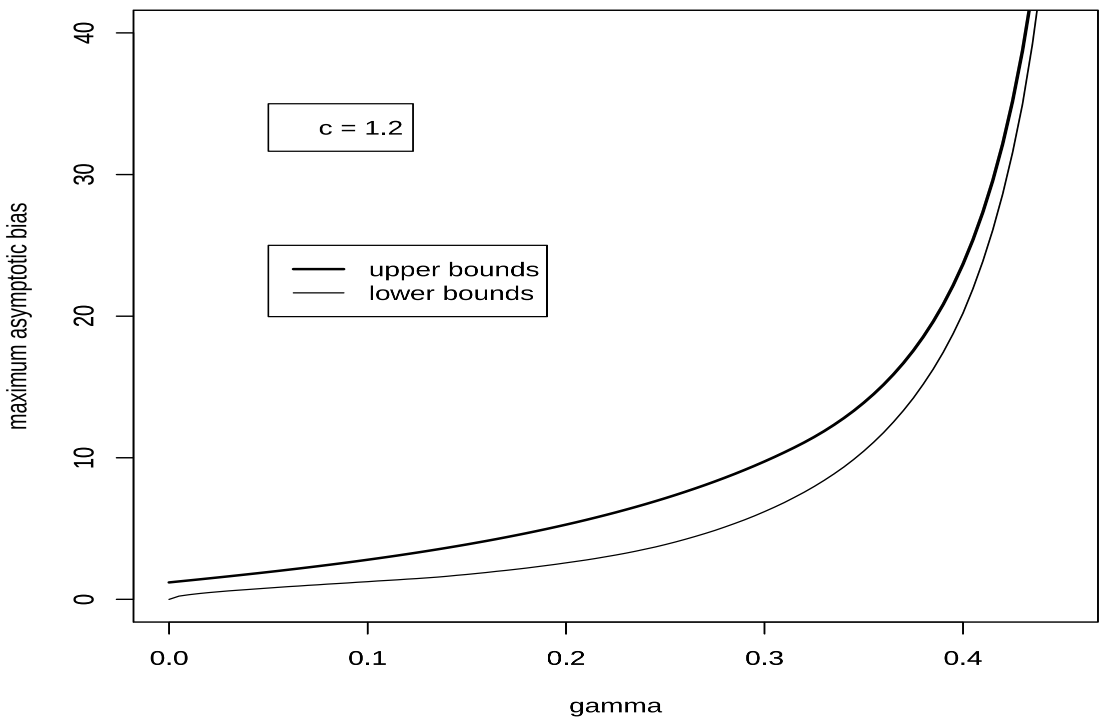

Figure 1: Lower and upper bounds of $B_\tau(c, \gamma)$ for c=1.2

5 Proofs

Proof. of Lemma 3.1 First we note that by A2 $J(c\,F^{\xi}_{\alpha,\boldsymbol{\theta}}+\gamma\delta_0)$ is a continuous function of α and $\boldsymbol{\theta}$. Since, for any $v>0$, $\lim_{|\alpha|\to\infty} F^{\xi}_{\alpha,\boldsymbol{\theta}}(v) < F^{\xi}_{0,\boldsymbol{\theta}}(v)$, it also follows from A1(b) that

$$\lim_{|\alpha|\to\infty} J(c\,F^{\xi}_{\alpha,\boldsymbol{\theta}}+\gamma\delta_0) > J(c\,F^{\xi}_{0,\boldsymbol{\theta}}+\gamma\delta_0).$$

Therefore, for any $\boldsymbol{\theta}\in R^p$ there exists $K_{\boldsymbol{\theta}}$ such that the infimum is attained in the compact set $[-K_{\boldsymbol{\theta}}, K_{\boldsymbol{\theta}}]$. Denoting by $\alpha(\boldsymbol{\theta})$ the value of α which gives the infimum (= the minimum), we obtain the first assertion of the lemma. We note that $\alpha(\boldsymbol{\theta})$ and K_t depend on ξ.

Assume that the second assertion of the lemma is not true. Then, there exist some $t>0$ and a sequence $\{\boldsymbol{\theta}_n\}\in\{\boldsymbol{\theta}:\|\boldsymbol{\theta}\|=t\}$ such that $\lim_{n\to\infty}|\alpha(\boldsymbol{\theta}_n)|=\infty$. Suppose without loss of generality that $\boldsymbol{\theta}_n\to\tilde{\boldsymbol{\theta}}$. For any $\alpha>0$ and $v>0$ we have

$$\lim_{n\to\infty}[c\,F^{\xi}_{\alpha(\boldsymbol{\theta}_n),\boldsymbol{\theta}_n}(v)+\gamma\delta_0(v)] = \gamma \le c\,F^{\xi}_{\alpha,\tilde{\boldsymbol{\theta}}}(v)+\gamma\delta_0(v).$$

Hence

$$\lim_{n\to\infty} J(c\,F^{\xi}_{\alpha(\boldsymbol{\theta}_n),\boldsymbol{\theta}_n}+\gamma\delta_0) \;\ge\; J(c\,F^{\xi}_{\alpha,\tilde{\boldsymbol{\theta}}}+\gamma\delta_0). \tag{5.1}$$

On the other hand, the definition of $\alpha(\boldsymbol{\theta})$ implies that for any $\alpha\in R$,

$$\lim_{n\to\infty} J(c\,F^{\xi}_{\alpha(\boldsymbol{\theta}_n),\boldsymbol{\theta}_n}+\gamma\delta_0) \;\le\; J(c\,F^{\xi}_{\alpha,\tilde{\boldsymbol{\theta}}}+\gamma\delta_0). \tag{5.2}$$

It follows from (5.1) and (5.2) that $J(c\,F^{\xi}_{\alpha,\tilde{\boldsymbol{\theta}}}+\gamma\delta_0)$ does not depend on α. This contradicts $\lim_{|\alpha|\to\infty} J(c\,F^{\xi}_{\alpha,\tilde{\boldsymbol{\theta}}}+\gamma\delta_0) > J(c\,F^{\xi}_{0,\tilde{\boldsymbol{\theta}}}+\gamma\delta_0)$, which implies the second assertion. □

Proof of Lemma 3.2. We note that $F^{\hat{\xi}}_{k\alpha,k\boldsymbol{\theta}}$ and $F^{\xi^*}_{k\alpha,k\boldsymbol{\theta}}$ are expressed in the form of

$$F^{\hat{\xi}}_{k\alpha,k\boldsymbol{\theta}}(v) = \min\left(F_{H_0,k\alpha,k\boldsymbol{\theta}}(v),\ \frac{1-\gamma}{c}\right),\quad \forall v\ge 0,$$

and

$$F^{\xi^*}_{k\alpha,k\boldsymbol{\theta}}(v) = \max\left(F_{H_0,k\alpha,k\boldsymbol{\theta}}(v) - \frac{c+\gamma-1}{c}, 0\right),\quad \forall v\ge 0,$$

where $F_{H_0,k\alpha,k\boldsymbol{\theta}}(v)$ is the distribution function of $|y-k\alpha-k\boldsymbol{\theta}'\boldsymbol{x}|$ under H_0. By Lemma 5 of Berrendero and Zamar (2001), $F_{H_0,k\alpha,k\boldsymbol{\theta}}(v)$ is strictly decreasing in $k>0$. Therefore, $F^{\hat{\xi}}_{\alpha,\boldsymbol{\theta}}(v)$ and $F^{\xi^*}_{\alpha,\boldsymbol{\theta}}(v)$ are strictly decreasing in $k>0$. □

Proof of Lemma 3.3. By Lemma 3.1, we have

$$m_{\xi}(t) = \inf_{\|\boldsymbol{\theta}\|=t} M_{\xi}(\boldsymbol{\theta}) = \inf_{\|\boldsymbol{\theta}\|=t}\ \inf_{[-K_t,K_t]} J(c\,F^{\xi}_{\alpha,\boldsymbol{\theta}}+\gamma\delta_0),$$

where $J(c\,F^{\xi}_{\alpha,\boldsymbol{\theta}} + \gamma\delta_0)$ is uniformly continuous on the compact set $\{\boldsymbol{\theta} : \|\boldsymbol{\theta}\| = t\} \times [-K_t, K_t]$. Therefore, $M_\xi(\boldsymbol{\theta})$ is continuous on the compact set $\{\boldsymbol{\theta} : \|\boldsymbol{\theta}\| = t\}$ and there exists $\|\boldsymbol{\theta}_t\| = t$ such that $M_\xi(\boldsymbol{\theta}_t) = \inf_{\|\boldsymbol{\theta}\|=t} M_\xi(\boldsymbol{\theta})$. This implies the assertion (a).

To show the assertion (b) let t_1 and t_2 be such that $t_1 > t_2$. Define $k = t_2/t_1 < 1$. Applying the assertion (a), there exist $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$ such that $m_\xi(t_1) = M_\xi(\boldsymbol{\theta}_1)$ and $m_\xi(t_2) = M_\xi(\boldsymbol{\theta}_2)$. Since, by $\xi \in \mathcal{F}_\lambda$

$$F^{\xi}_{\alpha(\boldsymbol{\theta}_1),\boldsymbol{\theta}_1}(v) < F^{\xi}_{k\alpha(\boldsymbol{\theta}_1),k\boldsymbol{\theta}_1}(v),$$

it follows from A1(a) and the definition of $\alpha(\boldsymbol{\theta})$ that

$$m_\xi(t_1) > J(c\,F^{\xi}_{k\alpha(\boldsymbol{\theta}_1),k\boldsymbol{\theta}_1} + \gamma\delta_0) \geq J(c\,F^{\xi}_{\alpha(k\boldsymbol{\theta}_1),k\boldsymbol{\theta}_1} + \gamma\delta_0). \tag{5.3}$$

Also, by the definition of $m_\xi(t)$ and $\|k\boldsymbol{\theta}_1\| = t_2$

$$m_\xi(t_2) \leq M_\xi(k\boldsymbol{\theta}_1) = J(c\,F^{\xi}_{\alpha(k\boldsymbol{\theta}_1),k\boldsymbol{\theta}_1} + \gamma\delta_0). \tag{5.4}$$

The inequalities (5.3) and (5.4) imply the assertion (b). □

Proof of Theorem 3.1. Let t^* be such that $d_{\xi^*} = m_{\hat{\xi}}(t^*)$. First, we show $B_{\boldsymbol{T}}(c,\gamma) \leq t^*$. Let $\tilde{\boldsymbol{\theta}} \in R^p$ be such that $\|\tilde{\boldsymbol{\theta}}\| = t > t^*$. It is enough to show that for any $H \in \mathcal{P}_{H_0}(c,\gamma)$ and any $\alpha \in R$ we have

$$J(F_{H,\alpha,\tilde{\boldsymbol{\theta}}}) > J(F_{H,0,\mathbf{0}}). \tag{5.5}$$

It is clear that for any $H = c\,(H_0 - W) + \gamma K \in \mathcal{P}_{H_0}(c,\gamma)$, $\alpha \in R$ and $v > 0$,

$$F_{H,\alpha,\tilde{\boldsymbol{\theta}}}(v) \;=\; c\,F^{\xi}_{\alpha,\tilde{\boldsymbol{\theta}}}(v) + \gamma F_{K,\alpha,\tilde{\boldsymbol{\theta}}}(v) \leq c\,F^{\hat{\xi}}_{\alpha,\tilde{\boldsymbol{\theta}}}(v) + \gamma\delta_0(v), \tag{5.6}$$

where $\xi = \{W_{\alpha,\boldsymbol{\theta}}\} \in \mathcal{F}_\lambda$ is defined as $W_{\alpha,\boldsymbol{\theta}} = W$ for any $\alpha \in R$ and $\boldsymbol{\theta} \in R^p$. From (5.6), A1(a), the definition of $m_\xi(t)$ and Lemma 3.3(b) it follows that for any $H \in \mathcal{P}_{H_0}(c,\gamma)$

$$J(F_{H,\alpha,\tilde{\boldsymbol{\theta}}}) \;\geq\; J(c\,F^{\hat{\xi}}_{\alpha,\tilde{\boldsymbol{\theta}}} + \gamma\delta_0) \geq m_{\hat{\xi}}(t) > m_{\hat{\xi}}(t^*). \tag{5.7}$$

The condition $d_{\xi^*} = m_{\hat{\xi}}(t^*)$ and A1(c) imply

$$m_{\hat{\xi}}(t^*) = \lim_{n\to\infty} J(c\,F^{\xi^*}_{0,\mathbf{0}} + \gamma U_n) \geq \lim_{n\to\infty} J(c\,F^{\xi}_{0,\mathbf{0}} + \gamma U_n) \geq J(F_{H,0,\mathbf{0}}). \tag{5.8}$$

Noting $t^* = m_{\hat{\xi}}^{-1}(d_{\xi^*})$, we obtain $B_{\boldsymbol{T}}(c,\gamma) \leq \overline{B}_{\boldsymbol{T}}(c,\gamma)$ from (5.7) and (5.8).

Next, we show $B_{\boldsymbol{T}}(c,\gamma) \geq m_\xi^{-1}(d_\xi)$, $^\forall\xi \in \mathcal{F}_\lambda$. Let $t_1 = m_\xi^{-1}(d_\xi)$ and let $t < t_1$. We find a distribution $H \in \mathcal{P}_{H_0}(c,\gamma)$ such that $\|\boldsymbol{T}(H)\| \geq t$. By Lemma 3.3(a), there exist $\boldsymbol{\theta}_t$ and α_t such that $m_\xi(t) = J(c\,F^{\xi}_{\alpha_t,\boldsymbol{\theta}_t} + \gamma\delta_0)$. Define $\tilde{H}_n = \delta_{(y_n,\boldsymbol{x}_n)}$, where $\boldsymbol{x}_n = n\boldsymbol{\theta}_t$ and

y_n is uniformly distributed on the interval $[\alpha_t + nt^2 - \frac{1}{n}, \alpha_t + nt^2 + \frac{1}{n}]$. If F_n is the uniform distribution function on $[-\frac{1}{n}, \frac{1}{n}]$, then for any $\boldsymbol{\beta} \in R^p$, $v > 0$ and $\alpha \in R$

$$\begin{aligned} F_{\tilde{H}_n,\alpha,\boldsymbol{\beta}}(v) &= F_n(v + \alpha - \alpha_t - n(t^2 - \boldsymbol{\beta}'\boldsymbol{\theta}_t)) \\ &\qquad -F_n(-v + \alpha - \alpha_t - n(t^2 - \boldsymbol{\beta}'\boldsymbol{\theta}_t)). \end{aligned} \tag{5.9}$$

For any $\xi = \{W_{\alpha,\boldsymbol{\theta}}\} \in \mathcal{F}_\lambda$ let $H_n^\xi(\alpha, \boldsymbol{\theta}) = c\,(H_0 - W_{\alpha,\boldsymbol{\theta}}) + \gamma\tilde{H}_n \in \mathcal{P}_{H_0}(c, \gamma)$. Suppose that $\sup_{n,\alpha,\boldsymbol{\theta}} \|\boldsymbol{T}(H_n^\xi(\alpha, \boldsymbol{\theta}))\| < t$ to find a contradiction. Then, for any $\alpha \in R$ and $\boldsymbol{\theta} \in R^p$ there exists a convergent subsequence, $\{\boldsymbol{T}(H_n^\xi(\alpha, \boldsymbol{\theta}))\}$, such that

$$\lim_{n\to\infty} \boldsymbol{T}(H_n^\xi(\alpha, \boldsymbol{\theta})) = \lim_{n\to\infty} \boldsymbol{\theta}_n^\xi(\alpha, \boldsymbol{\theta}) = \tilde{\boldsymbol{\theta}}^\xi(\alpha, \boldsymbol{\theta}), \quad where \quad \|\tilde{\boldsymbol{\theta}}^\xi(\alpha, \boldsymbol{\theta})\| = \tilde{t}^\xi(\alpha, \boldsymbol{\theta}) < t.$$

Since $t^2 - \boldsymbol{\theta}_t'\boldsymbol{\theta}_t = 0$, it follows from (5.9) that

$$\lim_{n\to\infty} F_{\tilde{H}_n,\alpha_t,\boldsymbol{\theta}_t}(v) = 1, \quad {}^\forall v > 0. \tag{5.10}$$

We show that for any $\alpha \in R$ and $\boldsymbol{\theta} \in R^p$ the subsequence of intercepts corresponding to $\boldsymbol{\theta}_n^\xi(\alpha, \boldsymbol{\theta})$, denoted by $\{T_0(H_n^\xi(\alpha, \boldsymbol{\theta}))\} = \{\alpha_n^\xi(\alpha, \boldsymbol{\theta})\}$ converges to a finite $\tilde{\alpha}^\xi(\alpha, \boldsymbol{\theta})$. To do this, assume $\lim_{n\to\infty} |\alpha_n^\xi(\alpha^*, \boldsymbol{\theta}^*)| = \infty$ for some $\alpha^* \in R$ and $\boldsymbol{\theta}^* \in R^p$. Then, it follows from (5.10) that

$$\begin{aligned} &\lim\nolimits_{n\to\infty} F_{H_n^\xi(\alpha^*,\boldsymbol{\theta}^*),\alpha_n^\xi(\alpha^*,\boldsymbol{\theta}^*),\boldsymbol{\theta}_n^\xi(\alpha^*,\boldsymbol{\theta}^*)}(v) \\ &\qquad = \gamma \lim_{n\to\infty} F_{\tilde{H}_n,\alpha_n^\xi(\alpha^*,\boldsymbol{\theta}^*),\boldsymbol{\theta}_n^\xi(\alpha^*,\boldsymbol{\theta}^*)}(v) \\ &\qquad < c\,F_{(H_0-W_{\alpha^*,\boldsymbol{\theta}^*}),\alpha_t,\boldsymbol{\theta}_t}(v) + \gamma\delta_0(v) \\ &\qquad = \lim_{n\to\infty} F_{H_n^\xi(\alpha^*,\boldsymbol{\theta}^*),\alpha_t,\boldsymbol{\theta}_t}(v), \qquad {}^\forall v > 0. \end{aligned} \tag{5.11}$$

Hence, by A1(b) we have

$$J(F_{H_n^\xi(\alpha^*,\boldsymbol{\theta}^*),\alpha_n^\xi(\alpha^*,\boldsymbol{\theta}^*),\boldsymbol{\theta}_n^\xi(\alpha^*,\boldsymbol{\theta}^*)}) > J(F_{H_n^\xi(\alpha^*,\boldsymbol{\theta}^*),\alpha_t,\boldsymbol{\theta}_t})$$

for large enough n. This fact contradicts the definition of $(\alpha_n^\xi(\alpha^*, \boldsymbol{\theta}^*), \boldsymbol{\theta}_n^\xi(\alpha^*, \boldsymbol{\theta}^*))$. Therefore, for any α and $\boldsymbol{\theta}$ we have $\lim_{n\to\infty} |\alpha_n^\xi(\alpha, \boldsymbol{\theta})| = \tilde{\alpha}^\xi(\alpha, \boldsymbol{\theta}) < \infty$. Since $t^2 - |\boldsymbol{\theta}_t'\tilde{\boldsymbol{\theta}}^\xi(\alpha, \boldsymbol{\theta})| = t^2 - t\tilde{t}^\xi(\alpha, \boldsymbol{\theta}) > 0$, it follows from (5.9) that

$$\lim_{n\to\infty} F_{\tilde{H}_n,\alpha_n^\xi(\alpha,\boldsymbol{\theta}),\boldsymbol{\theta}_n^\xi(\alpha,\boldsymbol{\theta})}(v) = 0, \quad {}^\forall v > 0. \tag{5.12}$$

Hence, by (5.12) and $\xi \in \mathcal{F}_\lambda$ we have

$$\begin{aligned} \lim_{n\to\infty} F_{H_n^\xi(\alpha,\boldsymbol{\theta}),\alpha_n^\xi(\alpha,\boldsymbol{\theta}),\boldsymbol{\theta}_n^\xi(\alpha,\boldsymbol{\theta})}(v) &= c\,F^\xi_{\tilde{\alpha}^\xi(\alpha,\boldsymbol{\theta}),\tilde{\boldsymbol{\theta}}^\xi(\alpha,\boldsymbol{\theta})}(v) \\ &\leq c\,F^\xi_{0,\mathbf{0}}(v) \\ &= \lim_{n\to\infty} [c\,F^\xi_{0,\mathbf{0}}(v) + \gamma U_n(v)], \quad {}^\forall v > 0. \end{aligned} \tag{5.13}$$

By A1(b) and A1(c) we have

$$\begin{aligned}\lim_{n\to\infty} J(F_{H_n^\xi(\alpha,\boldsymbol{\theta}),\alpha_n^\xi(\alpha,\boldsymbol{\theta}),\boldsymbol{\theta}_n^\xi(\alpha,\boldsymbol{\theta})}) &\geq \lim_{n\to\infty} J(c\,F_{0,\mathbf{0}}^\xi + \gamma U_n) \\ &= d_\xi = m_\xi(t_1).\end{aligned} \tag{5.14}$$

From (5.10) it follows that

$$\lim_{n\to\infty} F_{H_n^\xi(\alpha,\boldsymbol{\theta}),\alpha_t,\boldsymbol{\theta}_t}(v) = c\,F_{\alpha_t,\boldsymbol{\theta}_t}^\xi(v) + \gamma\delta_0(v) \tag{5.15}$$

The equation (5.15) and Lemma 3.3(b) imply

$$\lim_{n\to\infty} J(F_{H_n^\xi(\alpha,\boldsymbol{\theta}),\alpha_t,\boldsymbol{\theta}_t}) = J(c\,F_{\alpha_t,\boldsymbol{\theta}_t}^\xi + \gamma\delta_0) = m_\xi(t) < m_\xi(t_1). \tag{5.16}$$

By (5.14) and (5.16), we have

$$J(F_{H_n^\xi(\alpha,\boldsymbol{\theta}),\alpha_n^\xi(\alpha,\boldsymbol{\theta}),\boldsymbol{\theta}_n^\xi(\alpha,\boldsymbol{\theta})}) > J(F_{H_n^\xi(\alpha,\boldsymbol{\theta}),\alpha_t,\boldsymbol{\theta}_t})$$

for large enough n. This inequality is a contradiction because of $(\alpha_n^\xi(\alpha,\boldsymbol{\theta}),\boldsymbol{\theta}_n^\xi(\alpha,\boldsymbol{\theta})) = \arg\min_{\eta,\boldsymbol{\beta}} J(F_{H_n^\xi(\alpha,\boldsymbol{\theta}),\eta,\boldsymbol{\beta}})$. Thus, for any $t < t_1$ we obtain $\sup_{n,\alpha,\boldsymbol{\theta}} \|\boldsymbol{T}(H_n^\xi(\alpha,\boldsymbol{\theta}))\| \geq t$. This completes the proof. $\Box$

Proof of Theorem 3.2. It is easy to check that

$$F_{\alpha,\boldsymbol{\theta}}^{\hat\xi}(v) = (H_0 - \hat W_{\alpha,\boldsymbol{\theta}})(-v+\alpha \leq y - \boldsymbol{\theta}'\boldsymbol{x} \leq v+\alpha), \quad {}^\forall v > 0.$$

By the symmetry and unimodality assumptions on F_0 and G_0 and the definition of $\hat W_{\alpha,\boldsymbol{\theta}}$, we have for all $\alpha \in R$,

$$F_{\alpha,\boldsymbol{\theta}}^{\hat\xi}(v) \;\leq\; (H_0 - \hat W_{0,\boldsymbol{\theta}})(-v \leq y - \boldsymbol{\theta}'\boldsymbol{x} \leq v) = F_{0,\boldsymbol{\theta}}^{\hat\xi}(v), \quad {}^\forall v > 0,$$

and therefore, from A1(a), it follows that

$$J(c\,F_{\alpha,\boldsymbol{\theta}}^{\hat\xi} + \gamma\delta_0) \;\geq\; J(c\,F_{0,\boldsymbol{\theta}}^{\hat\xi} + \gamma\delta_0), \quad {}^\forall \alpha \in R.$$

This implies the first equality of the lemma. It is easy to see that $J(c\,F_{0,\boldsymbol{\theta}}^{\hat\xi} + \gamma\delta_0)$ only depends on $\boldsymbol{\theta}$ through the value of $\|\boldsymbol{\theta}\|$, because $F_{0,\boldsymbol{\theta}}^{\hat\xi}$ is so. $\Box$

Proof of Theorem 4.1. It follows from (4.2) and Theorem 3.2 that

$$d_{\xi^*} = S(c\,F_{0,\mathbf{0}}^{\xi^*} + \gamma\delta_\infty) = g_{\xi^*,1}^{-1}\left(\frac{b-\gamma}{c}\right)$$

and

$$m_{\hat\xi,S}(\|\boldsymbol{\theta}\|) = S(c\,F_{0,\boldsymbol{\theta}}^{\hat\xi} + \gamma\delta_0) = \sqrt{1+\|\boldsymbol{\theta}\|^2}\, g_{\hat\xi,1}^{-1}\left(\frac{b}{c}\right).$$

Hence, solving $m_{\hat{\xi},S}(\|\boldsymbol{\theta}\|) = d_{\xi^*}$ in $\|\boldsymbol{\theta}\|$, we obtain (4.4). Similarly, we can obtain (4.3). Assume $b \leq 0.5$. Then we have $\min(b, 1-b) = b$,

$$\lim_{\gamma \uparrow b} g^{-1}_{\xi^*,1}\left(\frac{b-\gamma}{c}\right) = \infty \quad \text{and} \quad \lim_{\gamma \uparrow b} g^{-1}_{\xi^\circ,1}\left(\frac{b-\gamma}{c}\right) = \infty,$$

where $\xi^\circ = \{W^\circ_{\alpha,\boldsymbol{\theta}}\}$, $W^\circ_{\alpha,\boldsymbol{\theta}} = [(c+\gamma-1)/c]\, H_0$. Therefore

$$\lim_{\gamma \uparrow b} \overline{B}_S(c,\gamma) = \lim_{\gamma \uparrow b} \underline{B}_S(c,\gamma) = \infty.$$

This completes the proof. □

Proof of Theorem 4.2. It is seen from (4.5) and Theorem 3.2 that

$$\begin{aligned} d_{\xi^*} &= \tau^2(c\,F^{\xi^*}_{0,\mathbf{0}} + \gamma\delta_\infty) \\ &= \left[g^{-1}_{\xi^*,1}\left(\frac{b-\gamma}{c}\right)\right]^2 \left[c\,\overline{g}_{\xi^*}\left(\frac{b-\gamma}{c}\right) + \gamma\right] \end{aligned}$$

and that

$$\begin{aligned} m_{\hat{\xi},\tau}(\|\boldsymbol{\theta}\|) &= \tau^2(c\,F^{\hat{\xi}}_{0,\boldsymbol{\theta}} + \gamma\delta_0) \\ &= m^2_{\hat{\xi},S}(\|\boldsymbol{\theta}\|)\cdot c\,E_{F^{\hat{\xi}}_{0,\boldsymbol{\theta}}}\left[\rho_2\left(\frac{y-\boldsymbol{\theta}'\boldsymbol{x}}{m_{\hat{\xi},S}(\|\boldsymbol{\theta}\|)}\right)\right] \qquad (5.17) \\ &= (1+\|\boldsymbol{\theta}\|^2)\left[g^{-1}_{\hat{\xi},1}\left(\frac{b}{c}\right)\right]^2 c\,\overline{g}_{\hat{\xi}}\left(\frac{b}{c}\right). \end{aligned}$$

Solving $m_{\hat{\xi},\tau}(\|\boldsymbol{\theta}\|) = d_{\xi^*}$, we obtain.

$$\|\boldsymbol{\theta}\| = m^{-1}_{\hat{\xi},\tau}(d_{\xi^*}) = \{(1+\overline{B}_S(c,\gamma)^2)H_{\xi^*,\hat{\xi}}(c,\gamma) - 1\}^{1/2}.$$

which implies (4.7). Similarly, we can obtain (4.6). □

Proof of Theorem 4.3. It is easily seen from (4.8) and Theorem 3.2 that

$$\begin{aligned} d_{\xi^*} &= CM(c\,F^{\xi^*}_{0,\mathbf{0}} + \gamma\delta_\infty) \\ &= \inf_{s \geq \sigma(cF^{\xi^*}_{0,\mathbf{0}} + \gamma\delta_\infty)} \left\{aE_{cF^{\xi^*}_{0,\mathbf{0}}+\gamma\delta_\infty}\left[\rho_1\left(\frac{v}{s}\right)\right] + \log s\right\} \\ &= \inf_{s \geq \sigma^{\xi^*}_{b,c,\gamma}} A^{\xi^*}_{a,c,\gamma}(s) + a\gamma. \end{aligned}$$

We also see that

$$\begin{aligned} m_{\hat{\xi},CM}(\|\boldsymbol{\theta}\|) &= CM(c\,F^{\hat{\xi}}_{0,\boldsymbol{\theta}} + \gamma\delta_0) \\ &= \inf_{s \geq (1+\|\boldsymbol{\theta}\|^2)^{1/2}\eta^{\hat{\xi}}_{b,c,\gamma}} \left\{acg_{\hat{\xi},1}\left(\frac{s}{(1+\|\boldsymbol{\theta}\|^2)^{1/2}}\right) + \log s\right\} \qquad (5.18) \\ &= \inf_{s \geq \eta^{\hat{\xi}}_{b,c,\gamma}} A^{\hat{\xi}}_{a,c,\gamma}(s) + \frac{1}{2}\log(1+\|\boldsymbol{\theta}\|^2). \qquad (5.19) \end{aligned}$$

Therefore, it follows from $m_{\hat{\xi},CM}(\|\boldsymbol{\theta}\|) = d_{\xi^*}$ that

$$\|\boldsymbol{\theta}\| = \{\exp[2a\gamma + 2D_{\xi^*,\hat{\xi},a}(c,\gamma)] - 1\}^{1/2}. \tag{5.20}$$

which is (4.11).

Similarly, we obtain (4.10). $\square$

Acknowledgments. The authors thank Shuhei Ando for his programming support, which was very helpuful to making Table 2.

References

Adrover,J.G.(1998). Minimax bias robust estimation of the dispersion matrix of a multivariate distribution, *Ann. Statist.* **26**, 2301-2320.

Ando,M. and Kimura,M.(2003). A characterization of the neighborhoods defined by certain special capacities and their applications to bias-robustness of estimates, *J. Statist. Plann. Inference.* **116**, 61-90.

Ando,M. and Kimura,M.(2004). The maximum asymptotic bias of S-estimates for regression over the neighborhoods defined by certain special capacities, *J. Multivariate Anal.* **90**, 407-425.

Ando,M., Kakiuchi,I. and Kimura,M.(2009). Robust nonparametric confidence intervals and tests for the median in the presence of (c,γ)-contamination, *J. Statist. Plann. inference.* **139**, 1836-1846.

Bednarski, T.(1981). On solutions of minimax test problems for special capacities.*Z. Wakreck. verw. Gebiete.* **10**, 269-278.

Berrendero,J.R. and Zamar,R.H.(2001). Maximum bias curves for robust regression with non-elliptical regressors, *Ann. Statist.* **29**, 224-251.

Berrendero,J.R., Mendes,B.V.M. and Tyler,D.E.(2007). On the maximum bias functions of MM-estimates and constrained M-estimates of regression, *Ann. Statist.* **35**, 13-40.

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

Huber,P.J.(1964). Robust estimation of a location parameter, *Ann. Math. Statist.* **35**, 73-101.

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

17, 1608-1630.

Martin,R.D. and Zamar,R.H.(1989). Asymptotically min-max robust M-estimates of scale for positive random variables, *J. Amer. Statist. Assoc.* **84**, 494-501.

Martin,R.D. and Zamar,R.H.(1993). Bias robust estimation of scale, *Ann. Statist.* **21**, 991-1017.

Mendes, B.V.M. and Tyler, D.E.(1996). Constrained M-estimation for regression, *In Robust Statistics, Data Analysis and Computer Intensive Methods. Lecture Notes in Statist*, **109**, Springer, New York, 299-320.

Rieder,H.(1977). Least favorable pairs for special capacities. *Ann. Statist.* **6**, 1080-1094.

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 Statist*, **26**, Springer, New York, 256-272.

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.