The Annals of Statistics
2002, Vol. 30, No. 6, 1760–1781

PROJECTION ESTIMATES OF MULTIVARIATE LOCATION

BY JORGE ADROVER AND VÍCTOR YOHAI[1]

Universidad Nacional de Córdoba and CIEM, and
Universidad de Buenos Aires and CONICET

In this paper we study the maximum asymptotic bias of the projection estimate for multivariate location based on univariate estimates of location and dispersion. In particular we study the projection estimate that uses the median and median absolute deviation about the median (MAD) as univariate location and dispersion estimates respectively. This estimator may be considered a natural affine equivariant multivariate median. For spherical distributions the maximum bias of this estimate depends only on the marginal distributions, and not on the dimension, and is approximately twice the maximum bias of the univariate median. We also show that for multivariate normal distributions, its maximum bias compares favorably with those of the Donoho–Stahel, minimum volume ellipsoid and minimum covariance determinant estimates. In all these cases the maximum bias increases with the dimension p.

1. Introduction. There have been several proposals for robust estimation of multivariate location and scatter based on projections. We can cite the Stahel–Donoho (SD) estimate [Stahel (1981), Donoho (1982)], the estimates based on Tukey's concept of depth, studied by Donoho and Gasko (1992), the P-estimates of covariance matrices [Maronna, Stahel and Yohai (1992)] and their extension to multivariate location [Tyler (1994)]. These two last proposals are closely related to regression P-estimates [Maronna and Yohai (1993)]. Regression P-estimates possess outstanding robustness properties when the robustness is measured by the maximum asymptotic bias. In this paper we show that multivariate location estimates based on the P-estimate approach also have a remarkable bias performance.

The basic idea common to the different P-estimates is to transform, by means of projections, a multivariate problem into the corresponding univariate problem, which is dealt with using univariate estimates. P-estimates of multiple regression are based on univariate regression and scale estimates, while P-estimates of scatter (resp. location) require univariate estimates of dispersion (resp. location and dispersion).

Received March 2000; revised March 2002.

[1]Supported in part by Grants X083 from University of Buenos Aires, PIP 4186/96 from CONICET and PICT-99 03-06277 from ANPCyT.

AMS 2000 subject classifications. Primary 62F35; secondary 62H12.

Key words and phrases. Projection estimates, maximum bias, robust estimates, multivariate location.

Let T_n and S_n be location and dispersion univariate estimates; that is, if $y_i^* = ay_i + b$ and $a, b \in \mathbb{R}$, then

$$T_n(y_1^*, \dots, y_n^*) = aT_n(y_1, \dots, y_n) + b, \tag{1}$$
$$S_n(y_1^*, \dots, y_n^*) = |a| S_n(y_1, \dots, y_n).$$

Consider a sample $\mathbf{x}_1, \dots, \mathbf{x}_n$ in $\mathbb{R}^p$. The P-estimate approach relies on the idea that $\boldsymbol{\zeta} \in \mathbb{R}^p$ is a good center of the data if, for any direction $\mathbf{a} \in \mathbb{R}^p$, the univariate projected set $\mathbf{a}'(\mathbf{x}_1 - \boldsymbol{\zeta}), \dots, \mathbf{a}'(\mathbf{x}_n - \boldsymbol{\zeta})$ is "well centered" around 0. Then a measure of how uncentered $\mathbf{x}_1 - \boldsymbol{\zeta}, \dots, \mathbf{x}_n - \boldsymbol{\zeta}$ are is given by

$$v(\boldsymbol{\zeta}, F_n) = \sup_{\mathbf{a} \neq 0} \frac{|T_n(\mathbf{a}'(\mathbf{x}_1 - \boldsymbol{\zeta}), \dots, \mathbf{a}'(\mathbf{x}_n - \boldsymbol{\zeta}))|}{S_n(\mathbf{a}'\mathbf{x}_1, \dots, \mathbf{a}'\mathbf{x}_n)},$$

where F_n is the empirical distribution. An ideal center would be a value $\boldsymbol{\zeta}$ such that $v(\boldsymbol{\zeta}, F_n) = \mathbf{0}$, that is, such that all the projected samples are perfectly centered. But in general, for an arbitrary data cloud, such a vector $\boldsymbol{\zeta}$ does not exist. Tyler (1994) defines a P-estimate of multivariate location by

$$\mathbf{T}_P(F_n) = \arg \min_{\boldsymbol{\zeta} \in \mathbb{R}^p} v(\boldsymbol{\zeta}, F_n). \tag{2}$$

Note that

$$h(\boldsymbol{\zeta}, \mathbf{a}, F_n) = \frac{|\mathbf{a}'\boldsymbol{\zeta} - T_n(\mathbf{a}'\mathbf{x}_1, \dots, \mathbf{a}'\mathbf{x}_n)|}{S_n(\mathbf{a}'\mathbf{x}_1, \dots, \mathbf{a}'\mathbf{x}_n)}$$

is a measure of the outlyingness of $\mathbf{a}'\boldsymbol{\zeta}$ with respect to the projected sample $\mathbf{a}'\mathbf{x}_1, \dots, \mathbf{a}'\mathbf{x}_n$. Then, $\sup_{\mathbf{a} \neq 0} h(\boldsymbol{\zeta}, \mathbf{a}, F_n)$ stands for a measure of the outlyingness of $\boldsymbol{\zeta}$ with respect to the data cloud $\mathbf{x}_1, \dots, \mathbf{x}_n$. Using (1), it is immediate that

$$h(\boldsymbol{\zeta}, \mathbf{a}, F_n) = \frac{|T_n(\mathbf{a}'(\mathbf{x}_1 - \boldsymbol{\zeta}), \dots, \mathbf{a}'(\mathbf{x}_n - \boldsymbol{\zeta}))|}{S_n(\mathbf{a}'\mathbf{x}_1, \dots, \mathbf{a}'\mathbf{x}_n)}$$

and

$$v(\boldsymbol{\zeta}, F_n) = \sup_{\|\mathbf{a}\|=1} h(\boldsymbol{\zeta}, \mathbf{a}, F_n),$$

which allows for another interpretation of the P-estimate as the center point with the smallest outlyingness.

Tyler (1994) shows that the P-estimates of multivariate location have a breakdown point close to 0.5, as long as the corresponding univariate estimates of location and dispersion also have this property. It is also shown that they are equivariant. Finally, their rate of consistency is conjectured to be $n^{-1/2}$.

A good measure of the robustness of an estimate is the maximum bias, which is the maximum asymptotic bias of the estimate caused by a given fraction of contamination. Other measures used to summarize the robustness performance of an estimate, such as the breakdown point [Hampel (1971)] and the gross error sensitivity [Hampel (1974)], can be derived from the maximum bias. Riedel (1991) and He and Simpson (1993) find lower bounds for the maximum bias of equivariant estimates.

Maronna and Yohai (1993) proved that the maximum bias of the P-estimate of regression which uses the median of slopes as univariate regression estimate and the median absolute deviation about the median (MAD) as dispersion is approximately twice the lower bound for equivariant estimates. The purpose of this paper is to derive an upper bound for the maximum bias of any P-estimate, in the case of symmetric distributions, and an exact maxbias curve in the case that the central model is elliptical and the univariate location and dispersion estimates are median and MAD respectively. Throughout, this median-based P-estimate will be called the MP estimate. A lower bound for the maximum bias of any equivariant estimate of multivariate location is also derived, and it is shown that the maximum bias of the MP estimate is approximately twice that bound for small contamination. There are few results on the maximum bias behavior of multivariate location estimates. We can cite Chen and Tyler (2002), who find the maximum bias and gross error sensitivity for the location estimate based on Tukey's depth, and Croux, Haesbroeck and Rousseeuw (1997), who obtain the maximum bias over point-mass contaminations of the minimum volume ball, a nonequivariant version of the minimum volume ellipsoid (MVE) introduced by Rousseeuw (1985).

We compare the maximum bias of the MP estimate for multivariate normal distributions to other high breakdown point estimates: the MVE estimate, the minimum covariance determinant (MCD) estimate [Rousseeuw (1985)] and the SD estimate. Our results show that the maximum bias of the MP estimate is much better than those of the MVE and MCD estimates for all dimensions, and than that of the SD estimates for dimensions greater than 5. Moreover, while the maximum bias of the MP estimate is independent of the dimension, those of the MVE, MCD and SD estimates increase with the dimension p.

In Section 2 we give the basic definitions and notation. We also give a lower bound for the maximum bias of an affine equivariant estimate of multivariate location. In Section 3 we show that the P-estimates are Fisher consistent and we obtain the maximum bias for the MP estimate. In Section 4.1 we compute numerically the maximum biases of the MVE, MCD and SD estimates. In Section 4.2 we report the results of a Monte Carlo study which compares the efficiencies of the different estimates under the multivariate normal model. The Appendix contains the proofs.

2. A lower bound for the maximum bias.

In the multivariate location model we observe a p-dimensional random vector $\mathbf{X} = (X_1, \dots, X_p)$ with distribution $F_{\boldsymbol{\mu}} = F_{\mathbf{0}}(\mathbf{x} - \boldsymbol{\mu})$, where $F_{\mathbf{0}}$ is symmetric around $\mathbf{0}$; that is, if $\mathbf{X}$ has distribution $F_{\mathbf{0}}$, then $-\mathbf{X}$ also has distribution $F_{\mathbf{0}}$. An important case of symmetry are the elliptical distributions. We say that $\mathbf{X}$ has an elliptical distribution if it has a density of the form

$$f(\mathbf{x},\boldsymbol{\mu},\Sigma)=\frac{1}{(\det\Sigma)^{1/2}}f_0\big((\mathbf{x}-\boldsymbol{\mu})'\Sigma^{-1}(\mathbf{x}-\boldsymbol{\mu})\big), \tag{3}$$

where $f_0:\mathbb{R}^+\to\mathbb{R}^+$, and Σ is a $p\times p$ positive definite matrix. If $\mathbf{X}$ has a density $f(\mathbf{x},\mathbf{0},I)$, then $\mathbf{a}'\mathbf{X}$ has the same distribution for all $\mathbf{a}\in\mathbb{S}^{p-1}=\{\mathbf{a}\in\mathbb{R}^p:\|\mathbf{a}\|=1\}$. This common distribution will be denoted by H_0 and its density by h_0.

All multivariate location estimates $\mathbf{T}_n$ considered in this paper are affine equivariant; that is, given a sample $\mathbf{x}_1,\dots,\mathbf{x}_n$, a $p\times p$ matrix A and $\mathbf{b}\in\mathbb{R}^p$,

$$\mathbf{T}_n(A\mathbf{x}_1+\mathbf{b},\dots,A\mathbf{x}_n+\mathbf{b})=A\mathbf{T}_n(\mathbf{x}_1,\dots,\mathbf{x}_n)+\mathbf{b}. \tag{4}$$

To study the robustness property of the multivariate location estimate we will consider contamination neighborhoods of the target distribution. Given a fraction of contamination $\varepsilon>0$, the corresponding contamination neighborhood of $F_{\boldsymbol{\mu}}$ is defined by

$$\mathcal{V}_\varepsilon(F_{\boldsymbol{\mu}})=\big\{F=(1-\varepsilon)F_{\boldsymbol{\mu}}+\varepsilon F^*:F^*\text{ any distribution on }\mathbb{R}^p\big\}.$$

All estimates studied here are defined by means of a functional on a subset $\mathcal{F}$ of the space of all the distributions on $\mathbb{R}^p$. We will assume that $\mathcal{F}$ contains the empirical distributions and all distributions belonging to $\mathcal{V}_\varepsilon(F_{\boldsymbol{\mu}})$ and that it is closed under affine transformations. If $\mathbf{x}_1,\dots,\mathbf{x}_n$ is a random sample from some distribution F and $\mathbf{T}$ is a continuous functional, then $\mathbf{T}(F)$ is the a.s. limit value of $\mathbf{T}_n(\mathbf{x}_1,\dots,\mathbf{x}_n)$. Then it is natural to require that an estimating functional $\mathbf{T}$ have the Fisher consistency property: $\mathbf{T}(F_{\boldsymbol{\mu}})=\boldsymbol{\mu}$. In general, given $F\in\mathcal{V}_\varepsilon(F_{\boldsymbol{\mu}})$ we will have $\mathbf{T}(F)\neq\boldsymbol{\mu}$. Then we define the asymptotic bias of $\mathbf{T}$ in F by

$$b(\mathbf{T},F,\boldsymbol{\mu})=\big((\mathbf{T}(F)-\boldsymbol{\mu})'V(F_{\mathbf{0}})^{-1}(\mathbf{T}(F)-\boldsymbol{\mu})\big)^{1/2}, \tag{5}$$

where V is an affine equivariant scatter functional. The maximum asymptotic bias of an estimating functional $\mathbf{T}$ for a fraction of contamination ε is defined by

$$B(\mathbf{T},\varepsilon,F_{\boldsymbol{\mu}})=\sup_{F\in\mathcal{V}_\varepsilon(F_{\boldsymbol{\mu}})}b(\mathbf{T},F,\boldsymbol{\mu}). \tag{6}$$

The inclusion of the scatter matrix $V(F_{\mathbf{0}})$ in (5) yields a self-standardizing definition of bias which parallels the notion of self-standardizing gross error sensitivity [see Hampel, Ronchetti, Rousseeuw and Stahel (1986)]. This standardization yields a concept of maximum bias which is invariant by affine transformations when applied to an equivariant functional. Therefore, if the functional $\mathbf{T}$ is affine equivariant, the maximum bias does not depend on $\boldsymbol{\mu}$, that is, $B(\mathbf{T},\varepsilon,F_{\boldsymbol{\mu}})=B(\mathbf{T},\varepsilon,F_{\mathbf{0}})$. In the elliptical case, we will assume that the scatter matrix V used in (5) is Fisher consistent for Σ, that is, $V(F_{\boldsymbol{\mu}})=\Sigma$. In this case, if $\mathbf{T}$ is affine equivariant, then the maximum bias is also independent of Σ. For the univariate case ($p = 1$), the maximum bias of a location estimate T at an arbitrary distribution G_0 reduces to

$$B(T, \varepsilon, G_0) = \sup_{G \in \mathcal{V}_\varepsilon(G_0)} \left| \frac{T(G) - T(G_0)}{\sigma(G_0)} \right|,$$

where $\sigma(\cdot)$ is a dispersion functional.

He and Simpson (1992) introduced the contamination sensitivity of an estimate $\mathbf{T}$ as

$$\gamma(\mathbf{T}, F_{\boldsymbol{\mu}}) = \left. \frac{\partial B(\mathbf{T}, \varepsilon, F_{\boldsymbol{\mu}})}{\partial \varepsilon} \right|_{\varepsilon = 0}.$$

Observe that $\gamma(\mathbf{T}, F_{\boldsymbol{\mu}}) = \gamma(\mathbf{T}, F_{\mathbf{0}})$ because of the invariance of the bias. For small ε, the maximum bias can be approximated by

$$B(\mathbf{T}, \varepsilon, F_{\boldsymbol{\mu}}) \approx \varepsilon \gamma(\mathbf{T}, F_{\boldsymbol{\mu}}). \tag{7}$$

The contamination sensitivity $\gamma(\mathbf{T}, F_{\boldsymbol{\mu}})$ is closely related to Hampel's (1971) gross error sensitivity γ^* $(\mathbf{T}, F_{\boldsymbol{\mu}})$. In fact, it is easy to show that always

$$\gamma(\mathbf{T}, F_{\boldsymbol{\mu}}) \geq \gamma^*(\mathbf{T}, F_{\boldsymbol{\mu}}),$$

where

$$\gamma^*(\mathbf{T}, F_{\boldsymbol{\mu}}) = \sup_{\mathbf{c} \in \mathbb{R}^p} \left\| \lim_{\varepsilon \to 0} \frac{\mathbf{T}((1-\varepsilon) F_{\boldsymbol{\mu}} + \varepsilon \delta_{\mathbf{c}}) - \mathbf{T}(F_{\boldsymbol{\mu}})}{\varepsilon} \right\|,$$

and $\delta_{\mathbf{c}}$ stands for a point-mass contamination. Under very general regularity conditions $\gamma^*(\mathbf{T}, F_{\boldsymbol{\mu}}) = \gamma(\mathbf{T}, F_{\boldsymbol{\mu}})$.

Huber (1964) proved that if L_0 is a univariate symmetric distribution with unimodal density l_0 and $L_\mu = L_0(x - \mu)$, then the maximum bias of the median estimating functional T_M is minimax among the translation equivariant estimates; that is, if T is another translation equivariant estimating functional, then

$$B(T, \varepsilon, L_\mu) \geq B(T_M, \varepsilon, L_\mu) = L_0^{-1}\left(\frac{1}{2(1-\varepsilon)} \right) = d_1(\varepsilon, L_0). \tag{8}$$

The maximum bias for the median is achieved when the contaminating distribution puts all its mass at infinity.

He and Simpson (1993) obtained a lower bound for the maximum bias of equivariant estimates. Using this result we prove in the next theorem that $d_1(\varepsilon, H_0)$ is a lower bound for any equivariant multivariate location estimator when the central model is elliptical, with H_0 the univariate marginal distribution for $\boldsymbol{\mu} = \mathbf{0}$ and $\Sigma = I$. Croux, Haesbroeck and Rousseeuw (1997) derived a similar result when the covariance is known and the central model is multivariate normal.

THEOREM 1. *Assume that* $\mathbf{X}$ *has a distribution with density given by* (3), *where* f_0 *is nonincreasing. Then, for any affine equivariant estimate* $\mathbf{T}$ *of multivariate location, we have*

$$B(\mathbf{T}, \varepsilon, F_{\boldsymbol{\mu}}) \geq d_1(\varepsilon, H_0) \tag{9}$$

and

$$\gamma^*(\mathbf{T}, F_{\boldsymbol{\mu}}) \geq \frac{1}{2h_0(0)}. \tag{10}$$

The proof is given in the Appendix.

3. Fisher consistency and maximum bias of P-estimates. Given a distribution F over $\mathbb{R}^p$, we denote by $\mathcal{L}(m(\mathbf{X}), F)$ the law of $m(\mathbf{X})$ when $\mathbf{X}$ has distribution F.

Let T and S be univariate location and dispersion functionals with the following properties:

P1. They are equivariant; that is, for all a and b,

$$T(\mathcal{L}(X + a, L)) = T(\mathcal{L}(X, L)) + a \tag{11}$$

and

$$S(\mathcal{L}(bX + a, L)) = |b| S(\mathcal{L}(X, L)). \tag{12}$$

P2. The functional T is Fisher consistent for symmetric distributions; that is, if L is symmetric around 0, then $T(L) = 0$.

Given $\mathbf{a} \neq \mathbf{0}$ and $\boldsymbol{\zeta}$ in $\mathbb{R}^p$, define

$$h(\boldsymbol{\zeta}, \mathbf{a}, F) = \frac{T(\mathcal{L}(\mathbf{a}'(\mathbf{x} - \boldsymbol{\zeta}), F))}{S(\mathcal{L}(\mathbf{a}'\mathbf{x}, F))} \tag{13}$$

and

$$v(\boldsymbol{\zeta}, F) = \sup_{\mathbf{a} \neq 0} |h(\boldsymbol{\zeta}, \mathbf{a}, F)| = \sup_{\|\mathbf{a}\|=1} |h(\boldsymbol{\zeta}, \mathbf{a}, F)|.$$

Then, the estimating functional associated with the P-estimate of multivariate location introduced in Section 1 is defined by

$$\mathbf{T}_{\mathrm{P}}(F) = \arg\min_{\boldsymbol{\zeta} \in \mathbb{R}^p} v(\boldsymbol{\zeta}, F). \tag{14}$$

Theorem 2 states the Fisher consistency property of the P-estimates.

THEOREM 2. *Assume* P1 *and* P2 *and let* $F_{\boldsymbol{\mu}} = F_{\mathbf{0}}(x - \boldsymbol{\mu})$, *where* $F_{\mathbf{0}}$ *is symmetric. Then* $\mathbf{T}_{\mathrm{P}}(F_{\boldsymbol{\mu}}) = \boldsymbol{\mu}$.

The proof is given in the Appendix.

The next theorem gives an upper bound for the maximum bias of P-estimates when $\mathbf{X}$ is symmetric. Since the bias is invariant we will suppose that $\boldsymbol{\mu} = \mathbf{0}$ and $V(F_{\mathbf{0}}) = I$. Given a multivariate distribution F on $\mathbb{R}^p$, and $\mathbf{a} \in \mathbb{R}^p$, we denote by $F^{(\mathbf{a})}$ the distribution of $\mathbf{a}'\mathbf{X}$ under F. Denote by

$$B^*(T, \varepsilon, F_{\mathbf{0}}) = \sup_{\|\mathbf{a}\|=1} B(T, \varepsilon, F_{\mathbf{0}}^{(\mathbf{a})}), \tag{15}$$

where B is defined as in (6) with $V(F_{\mathbf{0}}^{(\mathbf{a})}) = 1$. Therefore $B^*(T, \varepsilon, F_{\mathbf{0}})$ is the maximum bias of the univariate location estimate T over the distributions $F_{\mathbf{0}}^{(\mathbf{a})}$ with $\|\mathbf{a}\| = 1$. Given a univariate distribution L, define

$$C^+(S, \varepsilon, L) = \sup_{L^* \in V_\varepsilon(L)} S(L^*), \qquad C^-(S, \varepsilon, L) = \inf_{L^* \in V_\varepsilon(L)} S(L^*)$$

and

$$C^*(S, \varepsilon, F_{\mathbf{0}}) = \frac{\sup_{\|\mathbf{a}\|=1} C^+(S, \varepsilon, F_{\mathbf{0}}^{(\mathbf{a})})}{\inf_{\|\mathbf{a}\|=1} C^-(S, \varepsilon, F_{\mathbf{0}}^{(\mathbf{a})})}.$$

THEOREM 3. *Assume* P1 *and* P2 *and let* $F_{\boldsymbol{\mu}} = F_{\mathbf{0}}(x - \boldsymbol{\mu})$, *where* $F_{\mathbf{0}}$ *is symmetric and standardized so that* $V(F_{\mathbf{0}}) = I$. *Then*

$$B(\mathbf{T}_{\mathrm{P}}, \varepsilon, F_{\boldsymbol{\mu}}) \le B^*(T, \varepsilon, F_{\mathbf{0}})\bigl(1 + C^*(S, \varepsilon, F_{\mathbf{0}})\bigr).$$

In particular, *if* $F_{\boldsymbol{\mu}}$ *is elliptical we have*

$$B(\mathbf{T}_{\mathrm{P}}, \varepsilon, F_{\boldsymbol{\mu}}) \le B(T, \varepsilon, H_0)\left(1 + \frac{C^+(S, \varepsilon, H_0)}{C^-(S, \varepsilon, H_0)}\right).$$

The proof is given in the Appendix.

REMARK 1. If S is a continuous functional, $C^*(S, \varepsilon, F_{\mathbf{0}})$ is approximately 1 for small ε, and therefore

$$\lim_{\varepsilon \to 0} \frac{B(\mathbf{T}_{\mathrm{P}}, \varepsilon, F_{\boldsymbol{\mu}})}{B^*(T, \varepsilon, F_{\boldsymbol{\mu}})} \le 2.$$

REMARK 2. To derive the bound of Theorem 3 we need the assumption of symmetry of $\mathbf{X}$. Under this assumption the projection estimates coincide with the center of symmetry, facilitating the derivation of the bound. Without the symmetry assumption the derivation of an upper bound for the maximum bias of $\mathbf{T}_{\mathrm{P}}$ seems intractable.

In the sequel, we focus on the exact computation of the maximum bias of the MP estimate under an elliptical central model. Without loss of generality we may assume $\boldsymbol{\mu} = \mathbf{0}$ and $\Sigma = I$.

The following notation is needed for Theorem 4. Let U be a random variable with distribution H_0 and for $c \in \mathbb{R}$, let $m_1(c, \varepsilon)$ be defined by

$$P_{H_0}\big(|U - c| \le m_1(c, \varepsilon)\big) = \frac{1 - 2\varepsilon}{2(1 - \varepsilon)},$$

and $m_2(c, \varepsilon)$ by

$$P_{H_0}\big(|U - c| \le m_2(c, \varepsilon)\big) = \frac{1}{2(1 - \varepsilon)}.$$

Let d_1 be defined as in (8). We also define $d_2 = d_2(\varepsilon, H_0) = m_2(d_1, \varepsilon)$, $d_3 = d_3(\varepsilon, H_0) = m_1(d_1, \varepsilon)$, $k(c, \varepsilon) = c/m_1(c, \varepsilon)$ and

$$d_0 = d_0(\varepsilon) = \sup_{\{c \in [0, d_1]\}} k(c, \varepsilon).$$

Since

$$\frac{1}{2} < \frac{1}{2(1 - \varepsilon)} = P_{H_0}(-d_2 - d_1 \le U \le d_2 - d_1) \tag{16}$$

and $d_1 > 0$ we obtain that $d_2 > d_1$.

REMARK 3. The quantities d_2 and d_3 have a clear meaning in this context—d_2 and d_3 represent the largest and smallest values for the median of $|U - d_1|$ under contamination respectively: d_2 arises when the contamination is placed at infinity and d_3 when the contamination is placed at d_1. The symbol d_0 stands for the largest value for the standardized median of $|U - c|$ under contamination.

The following theorem gives the maximum bias and sensitivity of the estimating functional corresponding to the MP estimate $\mathbf{T}_{\mathrm{MP}}$.

THEOREM 4. *Assume* P1 *and* P2 *and suppose that* $F_{\boldsymbol{\mu}}$ *has a density given by* (3), *where* f_0 *is strictly decreasing. Then we have the following*:

(i)

$$B(\mathbf{T}_{\mathrm{MP}}, \varepsilon, F_{\boldsymbol{\mu}}) = d_1 + d_2 d_0. \tag{17}$$

(ii) *If* $k(c, \varepsilon)$ *is nondecreasing for* $c \le d_1$, *then*

$$B(\mathbf{T}_{\mathrm{MP}}, \varepsilon, F_{\boldsymbol{\mu}}) = d_1\left(1 + \frac{d_2}{d_3}\right). \tag{18}$$

(iii) *If* $k(c, \varepsilon)$ *is nondecreasing in a neighborhood of* 0, *for all* ε *in a neighborhood of* 0, *then*

$$\gamma(\mathbf{T}_{\mathrm{MP}}, F_{\boldsymbol{\mu}}) = \gamma^*(\mathbf{T}_{\mathrm{MP}}, F_{\boldsymbol{\mu}}) = \frac{1}{h_0(0)}. \tag{19}$$

The proof is given in the Appendix.

REMARK 4. Observe that, in the elliptical case, the maximum bias of the MP estimate depends only on the marginal distribution. Therefore the maximum bias is not affected by the dimension as long as the marginal distribution remains unchanged. As we will see in Section 4, for other estimates of multivariate location such as the MVE and the SD estimates, the maximum bias increases with p even in the normal case.

The maxbias of Tukey's median [Chen and Tyler (2002)] is also dimension-free when considering elliptical distributions. However, this property no longer holds in the nonelliptical case. A counterexample can be constructed by taking a vector of p independent Cauchy random variables (as opposed to spherical Cauchy). The dependency is a factor of the square root of the dimension [Tyler (2000)]. We conjecture that the maxbias of the MP estimate is also dependent on the dimension when the marginals of the multivariate distribution consist of independent Cauchy variables. Unfortunately, we have not been able to prove this fact.

REMARK 5. If $\mathbf{X}$ is multivariate normal, then H_0 is $N(0,1)$. Numerical computations show that in this case $k(c,\varepsilon)$ is increasing for $c \le d_1$ provided $\varepsilon < 0.4088$, and therefore (18) holds. If $\varepsilon > 0.4088$, then $\partial k(c,\varepsilon)/\partial c|_{c=d_1} < 0$ and the maximum of $k(c,\varepsilon)$ is attained for $c < d_1$.

REMARK 6. From (7), (10) and (19) it follows that, under the conditions of Theorem 1, the maximum bias of $\mathbf{T}_{\mathrm{MP}}$ for small ε is approximately twice the minimum possible bias for equivariant estimates.

REMARK 7. Chen and Tyler (2002) derive the gross error sensitivity and the contamination sensitivity for Tukey's median. They conclude that in the elliptical case the contamination sensitivity is twice the gross error sensitivity. Unlike the MP estimate, both concepts do not coincide for Tukey's median. However, the contamination sensitivity for the MP estimate coincides with that of Tukey's median. A similar situation had been noticed by Adrover, Maronna and Yohai (2002) in the regression setup: the projection estimate [Maronna and Yohai (1993)] and the maximum depth estimate [Rousseeuw and Hubert (1999)] have the same contamination sensitivity.

4. Numerical results.

4.1. *Maximum bias computations.* In this section we compare the maximum biases of the MP, MVE, MCD and SD estimates when $F_{\mathbf{0}}$ is the multivariate normal model. Since there are no theoretical results for the maximum biases of the MVE, MCD and the SD estimates, they are computed numerically and the maximum is taken only over point-mass contaminations. This maximum bias will be denoted by $\overline{B}(T, \varepsilon, F_{\mathbf{0}})$.

The location and scatter MVE estimating functionals $\mathbf{T}(F)$, $V(F)$ are defined as follows: Let Σ be a positive definite $p \times p$ matrix, let $\boldsymbol{\mu} \in \mathbb{R}^p$, let F be a distribution function over $\mathbb{R}^p$ and let

$$M(\boldsymbol{\mu}, \Sigma, F) = \operatorname*{median}_F (\mathbf{X} - \boldsymbol{\mu})' \Sigma^{-1} (\mathbf{X} - \boldsymbol{\mu}).$$

Define

$$(\mathbf{T}(F), V^*(F)) = \arg \min_{\boldsymbol{\mu} \in \mathbb{R}^p,\, \det(\Sigma)=1} M(\boldsymbol{\mu}, \Sigma, F) \tag{20}$$

and

$$V(F) = M\big(\mathbf{T}(F), V^*(F), F\big) V^*(F).$$

Therefore, the location and scatter MVE estimates correspond to the center and matrix which define the minimum volume ellipsoid covering half of the data.

If A is an event based on $\mathbf{X}$, we denote by $E_F(\mathbf{X} \mid A)$ and $\operatorname{cov}_F(\mathbf{X} \mid A)$ the mean and covariance matrix of $\mathbf{X}$ given A under F respectively. To define the MCD estimating functional, consider the set $\mathcal{E}$ of all ellipsoids $E \subset \mathbb{R}^p$ such that $P_F(E) = 0.5$, and let

$$E^* = \operatorname*{arg}_{E \in \mathcal{E}} \min \det\big(\operatorname{cov}(\mathbf{X} \mid E)\big).$$

Then the location and scatter MCD estimating functionals are the expected value and covariance matrix corresponding to the data restricted to E^*; that is,

$$\mathbf{T}(F) = E_F(\mathbf{X} \mid E^*), \qquad V(F) = \operatorname*{cov}_F(\mathbf{X} \mid E^*).$$

The multivariate location SD estimating functional is defined as a weighted mean, where the weights depend on a measure of outlyingness of the observations. The outlyingness of an observation $\mathbf{y} \in \mathbb{R}^p$ is defined by

$$v(\mathbf{y}, F) = \sup_{\mathbf{a} \in \mathbb{R}^p} \frac{|\mathbf{a}'\mathbf{y} - \operatorname{median}_F(\mathbf{a}'\mathbf{X})|}{\operatorname{MAD}_F(\mathbf{a}'\mathbf{X})}.$$

Then the multivariate location SD estimating functional is defined as

$$\mathbf{T}(F) = \frac{E_F(w(v(\mathbf{X}, F))\mathbf{X})}{E_F(w(v(\mathbf{X}, F)))},$$

where w is a nonnegative and nonincreasing weight function. We consider two SD estimates with weight functions w in the Huber family

$$w_c^{\mathrm{H}}(u) = \min\left(\frac{1}{c}, \frac{1}{|u|}\right).$$

The values of c are 0—which corresponds to the limit case $w = 1/|u|$—and $\sqrt{\chi^2_{0.90,p}}$, where $\chi^2_{\alpha,p}$ is the α-quantile of the χ^2-distribution with p degrees of freedom. We have also tried other intermediate values of c, and the maximum biases in all cases were an increasing function of c.

The procedure to compute the bias for the location SD estimate is analogous to the procedure for computing the bias of the scatter SD estimate that is described in Section 3 of Maronna and Yohai (1995).

Since the MVE and MCD estimates are equivariant, the maximum bias for point-mass contaminations can be computed assuming that $F_{\mathbf{0}}$ is $N(\mathbf{0}, I)$ and considering only contaminating points of the form $(k, 0, \dots, 0)$. Then it can be shown that the minimum in (20) is attained at $\boldsymbol{\mu}$ of the form $(\mu, 0, \dots, 0)$ and Σ of the form $\operatorname{diag}(r_1, r_2, \dots, r_2)$, where $\operatorname{diag}(r_1, \dots, r_p)$ is the $p \times p$ diagonal matrix with elements $r_1, \dots, r_p$ in the diagonal. Since $\det(\Sigma) = 1$, we also have $r_2 = 1/r_1^{1/(p-1)}$. Write $\Sigma(r) = \operatorname{diag}(r, 1/r^{1/(p-1)}, \dots, 1/r^{1/(p-1)})$, and let $\kappa(k, \mu, r, \varepsilon)$ be the median of $(\mathbf{X} - (\mu, 0, \dots, 0))'\Sigma(r)^{-1}(\mathbf{X} - (\mu, 0, \dots, 0))$, when $\mathbf{X}$ has distribution $(1 - \varepsilon)N(\mathbf{0}, I) + \varepsilon\delta_{(k,0,\dots,0)}$ and

$$(21) \qquad \big(\overline{\mu}(k, \varepsilon), \overline{r}(k, \varepsilon)\big) = \arg\min_{\mu,r} \kappa(k, \mu, r, \varepsilon).$$

Then, if $\mathbf{T}$ is the location MVE estimate and $F_{\mathbf{0}}$ is the multivariate normal, we have

$$(22) \qquad \overline{B}(\mathbf{T}, \varepsilon, F_{\mathbf{0}}) = \max_k \overline{\mu}(k, \varepsilon).$$

We wrote a MATLAB function which computes $\kappa(k, \mu, r, \varepsilon)$ by numerical integration. An initial value of $\overline{\mu}(k, \varepsilon)$ is obtained by minimizing (21) over a grid of values of μ and r, with step 0.1 in both variables. Starting from this value we use the MATLAB function FMINS (Nelder–Mead simplex minimization algorithm) to obtain a final value of $\overline{\mu}(k, \varepsilon)$. We compute $\overline{\mu}(k, \varepsilon)$ for a grid of values of k (with step depending on p) obtaining an initial value for $\overline{B}(T, \varepsilon, F_{\mathbf{0}})$. Using this initial point, the maximum in (22) is obtained using a bisection maximization algorithm. A similar procedure was used to compute the maximum bias for the MCD estimate.

The maximum biases, shown in Tables 1 and 2, were computed for $2 \le p \le 10$, 15 and 20, and $\varepsilon = 0.05, 0.10, 0.15$ and 0.20. These tables show that the maximum bias of the MP estimate is always smaller than those of the MVE, MCD and the two SD estimates for $p \ge 5$. They also show that the maximum biases of the MVE and SD estimates increase with the dimension for $p \ge 5$, except for the MVE estimate when $\varepsilon = 0.05$. In this latter case, the present computing accuracy of three digits does not allow us to detect a trend. Therefore, in the case that such a trend exists, it would be negligible from a practical point of view for $p \le 20$. For fixed ε, the growth of $B(T, \varepsilon, F_{\mathbf{0}})$ for the MVE estimate is very slow, close to linear for the SD estimate and close to exponential for the MCD estimate. On the other hand, according to the content of Remark 4, the maximum bias of the MP estimate is in this case independent of p.

TABLE 1
Maximum biases for $\varepsilon = 0.05$ and $\varepsilon = 0.10$

	Estimate ($\varepsilon = 0.05$)					**Estimate ($\varepsilon = 0.10$)**				
p	**SD0**	**SD90**	**MVE**	**MCD**	**MP**	**SD0**	**SD90**	**MVE**	**MCD**	**MP**
2	0.077	0.131	0.395	0.277	0.141	0.159	0.292	0.619	0.539	0.321
3	0.095	0.149	0.395	0.307	0.141	0.200	0.333	0.69	0.655	0.321
4	0.123	0.192	0.395	0.336	0.141	0.267	0.436	0.72	0.786	0.321
5	0.160	0.226	0.395	0.369	0.141	0.361	0.523	0.73	0.935	0.321
6	0.197	0.255	0.395	0.406	0.141	0.483	0.613	0.74	1.094	0.321
7	0.246	0.298	0.395	0.440	0.141	0.617	0.717	0.75	1.277	0.321
8	0.310	0.347	0.395	0.480	0.141	0.762	0.837	0.75	1.481	0.321
9	0.368	0.394	0.395	0.520	0.141	0.898	0.951	0.75	1.703	0.321
10	0.418	0.443	0.395	0.560	0.141	1.021	1.066	0.75	1.965	0.321
15	0.715	0.719	0.395	0.781	0.141	1.746	1.751	0.765	3.846	0.321
20	1.013	1.013	0.395	1.117	0.141	2.471	2.472	0.775	7.005	0.321

4.2. *Monte Carlo efficiencies.* We performed a Monte Carlo study to compare the efficiencies under multivariate normal distribution for finite sample size of the estimates considered in Section 4.1. Since all the estimates are equivariant we consider without loss of generality only the case of zero mean and identity covariance matrix. We also include in this study the sample mean which is optimal in the normal case. We take $p = 2$–$10,15$ and 20. The sample size n was chosen as equal to 50 and 100. The number of replications was 500. For each estimate $\mathbf{T}$ we compute the mean square error (MSE) defined by

$$\frac{1}{500}\sum_{i=1}^{500}\|\mathbf{T}_i\|^2,$$

where $\mathbf{T}_i$ is the value of the estimate for the ith sample.

TABLE 2
Maximum biases for $\varepsilon = 0.15$ and $\varepsilon = 0.20$

	Estimate ($\varepsilon = 0.15$)					**Estimate ($\varepsilon = 0.20$)**				
p	**SD0**	**SD90**	**MVE**	**MCD**	**MP**	**SD0**	**SD90**	**MVE**	**MCD**	**MP**
2	0.27	0.51	0.91	0.90	0.56	0.40	0.81	1.29	1.49	0.90
3	0.36	0.56	1.03	1.20	0.56	0.57	0.98	1.48	2.22	0.90
4	0.48	0.79	1.08	1.58	0.56	0.84	1.31	1.57	3.19	0.90
5	0.67	0.95	1.12	2.04	0.56	1.18	1.59	1.65	4.46	0.90
6	0.90	1.14	1.14	2.58	0.56	1.58	1.91	1.71	6.17	0.90
7	1.13	1.33	1.16	3.23	0.56	1.98	2.25	1.78	8.53	0.90
8	1.41	1.54	1.18	4.07	0.56	2.44	2.62	1.84	11.69	0.90
9	1.67	1.77	1.20	5.05	0.56	2.91	3.02	1.90	15.92	0.90
10	1.91	1.99	1.22	6.28	0.56	3.41	3.48	1.97	21.68	0.90
15	3.27	3.28	1.31	18.00	0.56	5.76	5.77	2.30	116.8	0.90
20	4.63	4.63	1.41	49.92	0.56	8.12	8.11	2.65	412.0	0.90

TABLE 3
MSE of the mean and relative efficiencies of robust estimates for Gaussian distribution, $n = 50$

	MSE	**Efficiency**				
p	**mean**	**SD0**	**SD90**	**MVE**	**MCD**	**MP**
2	0.0386	0.81	0.97	0.21	0.24	0.77
3	0.0599	0.82	0.97	0.22	0.29	0.77
4	0.0794	0.87	0.97	0.21	0.33	0.79
5	0.1016	0.87	0.96	0.21	0.35	0.82
6	0.1149	0.87	0.95	0.20	0.35	0.82
7	0.1365	0.89	0.95	0.22	0.38	0.82
8	0.1544	0.91	0.95	0.22	0.39	0.82
9	0.1763	0.91	0.95	0.25	0.42	0.81
10	0.2031	0.92	0.95	0.26	0.43	0.82
15	0.2999	0.92	0.93	0.34	0.45	0.80
20	0.4029	0.92	0.92	0.44	0.47	0.79

We compute an approximate MP estimate as follows. Consider a sample $\mathbf{x}_1, \ldots, \mathbf{x}_n$ in $\mathbb{R}^p$. First, we compute the approximate outlyingness $\upsilon_n(\mathbf{x}_i)$ for each observation of the sample as in Section 4.1 of Maronna and Yohai (1995). According to (2), the location MP estimate is the value in $\mathbb{R}^p$ with the smallest outlyingness. Then an approximate MP estimate is computed by $\boldsymbol{\mu}_n = \mathbf{x}_{i_0}$, where

$$i_0 = \arg \min_{1 \le i \le n} \upsilon_n(\mathbf{x}_i).$$

In Table 3 for $n = 50$ and in Table 4 for $n = 100$ we report the MSE for the mean and the relative efficiencies with respect to the mean for the other estimates. We observe that, for both sample sizes, the most efficient estimates are both SD

TABLE 4
MSE of the mean and relative efficiencies of robust estimates for Gaussian distribution, $n = 100$

	MSE	**Efficiency**				
p	**mean**	**SD0**	**SD90**	**MVE**	**MCD**	**MP**
2	0.0193	0.80	0.97	0.17	0.20	0.76
3	0.0277	0.83	0.98	0.14	0.22	0.76
4	0.0390	0.89	0.98	0.13	0.28	0.76
5	0.0502	0.89	0.98	0.13	0.30	0.78
6	0.0608	0.91	0.98	0.13	0.32	0.80
7	0.0686	0.91	0.98	0.13	0.34	0.84
8	0.0809	0.92	0.99	0.13	0.35	0.86
9	0.0907	0.93	0.99	0.13	0.38	0.89
10	0.103	0.94	0.99	0.14	0.39	0.90
15	0.153	0.95	0.99	0.18	0.43	0.91
20	0.200	0.95	0.99	0.22	0.45	0.89

estimates followed by the MP estimate. The simulations show that MVE and MCD estimates are quite inefficient in agreement with the known asymptotic results of Davies (1992) and Butler, Davies and Jhun (1993). For $n = 50$, the efficiencies of the MVE, MCD and SD0 estimates increase with p. For $n = 100$, all the estimates increase their efficiency.

APPENDIX

Because of the affine equivariance of the P-estimate, without loss of generality we will assume in this appendix that the true parameters are $\boldsymbol{\mu} = \mathbf{0}$ and $\Sigma = I$.

PROOF OF THEOREM 1. We only have to verify the lower bound for contamination bias provided by Theorem 2.1 of He and Simpson (1993). Given F and G distribution functions on $\mathbb{R}^p$ and f and g the corresponding density functions, the variation norm between F and G is given by

$$d_v(F, G) = \sup_{\text{meas.}A} |F(A) - G(A)| = \tfrac{1}{2} \int |f(\mathbf{x}) - g(\mathbf{x})| \, d\mathbf{x},$$

where the last equality is Scheffé's theorem [Scheffé (1947)]. According to He and Simpson's result, an equivariant multivariate location functional $\mathbf{T}$ satisfies

$$B(\mathbf{T}, \varepsilon, F_{\mathbf{0}}) \geq \frac{1}{2} b_v\left(\frac{\varepsilon}{1-\varepsilon}; F_{\mathbf{0}}\right),$$

where $b_v(\delta, F_{\mathbf{0}}) = \sup\{\|\boldsymbol{\theta}\| : d_v(F_{\boldsymbol{\theta}}, F_{\mathbf{0}}) \leq \delta\}$ (variation gauge). Then

$$\begin{aligned}
d_v(F_{\boldsymbol{\theta}}, F_{\mathbf{0}}) = d_v(F_{\boldsymbol{\theta}/2}, F_{-\boldsymbol{\theta}/2}) &= \frac{1}{2}\left[\int \left| f_0\left(\left\|\mathbf{x} - \frac{\boldsymbol{\theta}}{2}\right\|\right) - f_0\left(\left\|\mathbf{x} + \frac{\boldsymbol{\theta}}{2}\right\|\right)\right| d\mathbf{x}\right] \\
&= \frac{1}{2}\left[\int \left(f_0\left(\left\|\mathbf{x} + \frac{\boldsymbol{\theta}}{2}\right\|\right) - f_0\left(\left\|\mathbf{x} - \frac{\boldsymbol{\theta}}{2}\right\|\right)\right) \mathbb{1}_{(-\infty,0)}(\boldsymbol{\theta}'\mathbf{x}) \, d\mathbf{x}\right. \\
&\qquad \left. + \int \left(f_0\left(\left\|\mathbf{x} - \frac{\boldsymbol{\theta}}{2}\right\|\right) - f_0\left(\left\|\mathbf{x} + \frac{\boldsymbol{\theta}}{2}\right\|\right)\right) \mathbb{1}_{(0,\infty)}(\boldsymbol{\theta}'\mathbf{x}) \, d\mathbf{x}\right] \\
&= \int \left(f_0\left(\left\|\mathbf{x} + \frac{\boldsymbol{\theta}}{2}\right\|\right) \mathbb{1}_{(-\infty,0)}(\boldsymbol{\theta}'\mathbf{x}) - f_0\left(\left\|\mathbf{x} + \frac{\boldsymbol{\theta}}{2}\right\|\right) \mathbb{1}_{(0,\infty)}(\boldsymbol{\theta}'\mathbf{x})\right) d\mathbf{x} \\
&= 2 \int f_0\left(\left\|\mathbf{x} + \frac{\boldsymbol{\theta}}{2}\right\|\right) \mathbb{1}_{(-\infty,0)}(\boldsymbol{\theta}'\mathbf{x}) \, d\mathbf{x} - 1 \\
&= 2 \int f_0(\|\mathbf{x}\|) \mathbb{1}_{(-\infty,\|\boldsymbol{\theta}\|^2/2)}(\boldsymbol{\theta}'\mathbf{x}) \, d\mathbf{x} - 1 \\
&= 2 \int f_0(\|\mathbf{x}\|) \mathbb{1}_{(-\infty,\|\boldsymbol{\theta}\|^2/2)}(x_1) \, d\mathbf{x} - 1 = 2 H_0\left(\frac{\|\boldsymbol{\theta}\|}{2}\right) - 1.
\end{aligned}$$

It follows that

$$\frac{1}{2}b_v\left(\frac{\varepsilon}{1-\varepsilon}; F_{\mathbf{0}}\right) = H_0^{-1}\left(\frac{1}{2(1-\varepsilon)}\right) = d_1(\varepsilon, H_0)$$

and (9) holds. Differentiating $d_1(\varepsilon, H_0)$ we get (10). □

PROOF OF THEOREM 2. Given any $\mathbf{a} \neq \mathbf{0}$ and $b \neq 0$, $\mathbf{a}'\mathbf{x}/b$ is a symmetric random variable around 0. Then $T(\mathcal{L}(\mathbf{a}'\mathbf{x}/b), F_{\mathbf{0}}) = 0$ and then $h(\mathbf{0}, \mathbf{a}, F_{\mathbf{0}}) = 0$. This implies that

$$v(\mathbf{0}, F_{\mathbf{0}}) = 0. \tag{23}$$

On the other hand, take $\boldsymbol{\zeta} \neq \boldsymbol{\mu}$ and put $\mathbf{a}_0 = -\boldsymbol{\zeta}$. Then

$$\frac{\mathbf{a}_0'(\mathbf{x}-\boldsymbol{\zeta})}{b} = \frac{\mathbf{a}'\mathbf{x}}{b} + \frac{\mathbf{a}_0'\mathbf{a}_0}{b}$$

and $T(\mathcal{L}(\mathbf{a}_0'(\mathbf{x}-\boldsymbol{\zeta})/b), F_{\mathbf{0}}) = \mathbf{a}_0'\mathbf{a}_0/b > 0$. Then $h(\boldsymbol{\zeta}, \mathbf{a}, F_{\mathbf{0}}) > 0$ for any $\mathbf{a} \neq \mathbf{0}$ and

$$v(\boldsymbol{\zeta}, F_{\mathbf{0}}) > 0. \tag{24}$$

From (23) and (24) the theorem follows. □

PROOF OF THEOREM 3. Let $\|\boldsymbol{\nu}_0\| > B^*(T, \varepsilon, F_{\mathbf{0}})(1 + C^*(S, \varepsilon, F_{\mathbf{0}}))$ and $\mathbf{a}_0 = \boldsymbol{\nu}_0/\|\boldsymbol{\nu}_0\|$. Then, by (15), for any $F \in \mathcal{V}_\varepsilon(F_{\mathbf{0}})$ we have

$$\begin{aligned}
\frac{|T(\mathcal{L}(\mathbf{a}_0'(\mathbf{X}-\boldsymbol{\nu}_0), F))|}{S(\mathcal{L}(\mathbf{a}_0'\mathbf{X}, F))} &= \frac{|T(\mathcal{L}(\mathbf{a}_0'\mathbf{X}, F)) - \|\boldsymbol{\nu}_0\||}{S(\mathcal{L}(\mathbf{a}_0'\mathbf{X}, F))} \\
&\geq \frac{\|\boldsymbol{\nu}_0\| - B^*(T, \varepsilon, F_{\mathbf{0}})}{S(\mathcal{L}(\mathbf{a}_0'\mathbf{X}, F))} \\
&> \frac{C^*(S, \varepsilon, F_{\mathbf{0}})B^*(T, \varepsilon, F_{\mathbf{0}})}{\sup_{\|\mathbf{a}\|=1} S(\mathcal{L}(\mathbf{a}'\mathbf{X}, F))} \\
&\geq \frac{B^*(T, \varepsilon, F_{\mathbf{0}})}{\inf_{\|\mathbf{a}\|=1} \inf_{F\in\mathcal{V}_\varepsilon(F_{\mathbf{0}})} S(\mathcal{L}(\mathbf{a}'\mathbf{X}, F))}.
\end{aligned} \tag{25}$$

Now take $\boldsymbol{\nu}_1 = \mathbf{0}$ and take any $\mathbf{a} \in \mathbb{R}^p$ such that $\|\mathbf{a}\| = 1$. Then

$$\frac{|T(\mathcal{L}(\mathbf{a}'(\mathbf{X}-\boldsymbol{\nu}_1), F))|}{S(\mathcal{L}(\mathbf{a}'\mathbf{X}, F))} \leq \frac{B^*(T, \varepsilon, F_{\mathbf{0}})|}{\inf_{\|\mathbf{a}\|=1} \inf_{F\in\mathcal{V}_\varepsilon(F_{\mathbf{0}})} S(\mathcal{L}(\mathbf{a}'\mathbf{X}, F))}. \tag{26}$$

From (25) and (26) we obtain that $\mathbf{T}_P(F) \neq \boldsymbol{\nu}_0$, proving the theorem. □

LEMMA 1. *If $h_0(t)$ is nonincreasing in $|t|$ then $m_i(x)$ is nondecreasing in $|x|$, $i = 1, 2$.*

PROOF. Let $0 < x_1 < x_2$, $a_1 = (1-2\varepsilon)/(2(1-\varepsilon))$ and $a_2 = 1 - a_1$. Then

$$\begin{aligned} a_i &= P\big(|\mathbf{a}'\mathbf{X} - x_1| \le m_i(x_1)\big) = P\big(x_1 - m_i(x_1) \le \mathbf{a}'\mathbf{X} \le x_1 + m_i(x_1)\big) \\ &\ge P\big(x_2 - m_i(x_1) \le \mathbf{a}'\mathbf{X} \le x_2 + m_i(x_1)\big), \end{aligned}$$

and $m_i(x_2) \ge m_i(x_1)$, $i = 1, 2$. □

The following lemma shows that $d_1(1 + d_2 d_0/d_1)$ is an upper bound for $B(\mathbf{T})$.

LEMMA 2. *Let $h_0(t)$ be nonincreasing in $|t|$. Then $B(\mathbf{T}_{\mathrm{MP}}, \varepsilon, F_{\mathbf{0}}) \le d_1(1 + d_2 d_0/d_1)$.*

PROOF. Let $\boldsymbol{\nu} \in \mathbb{R}^p$ be such that $\|\boldsymbol{\nu}\| > d_1$ and $\mathbf{a}_1 = \boldsymbol{\nu}/\|\boldsymbol{\nu}\|$. Let $G \in \mathcal{V}_\varepsilon(F_{\mathbf{0}})$. Then, since $|\operatorname{median}_G(\mathbf{a}_1'\mathbf{X})| \le d_1$, we obtain

$$\begin{aligned} &\sup_{\|\mathbf{a}\|=1} \big|\operatorname{median}_G\big(\mathbf{a}'(\mathbf{X} - \boldsymbol{\nu})\big)\big| \\ &\qquad \ge \big|\operatorname{median}_G(\mathbf{a}_1'\mathbf{X}) - \|\boldsymbol{\nu}\|\big| = \big|\|\boldsymbol{\nu}\| - \operatorname{median}_G(\mathbf{a}_1'\mathbf{X})\big| \ge |\|\boldsymbol{\nu}\| - d_1|. \end{aligned}$$

From Lemma 1 we obtain that $\mathrm{MAD}_G(\mathbf{a}'\mathbf{X}) \le m_2(d_1, \varepsilon) = d_2$ for all $\mathbf{a} \in S^{p-1}$ and $G \in \mathcal{V}_\varepsilon$. Therefore

$$\sup_{\|\mathbf{a}\|=1} \frac{|\operatorname{median}_G(\mathbf{a}'(\mathbf{X} - \boldsymbol{\nu}))|}{\mathrm{MAD}_G(\mathbf{a}'\mathbf{X})} \ge (\|\boldsymbol{\nu}\| - d_1)\frac{1}{d_2}. \tag{27}$$

On the other hand, we can suppose without loss of generality that

$$\sup_{\|\mathbf{a}\|=1} \frac{|\operatorname{median}_G(\mathbf{a}'(\mathbf{X} - \boldsymbol{\nu}))|}{\mathrm{MAD}_G(\mathbf{a}'\mathbf{X})} \le \sup_{\|\mathbf{a}\|=1} \frac{|\operatorname{median}_G(\mathbf{a}'\mathbf{X})|}{\mathrm{MAD}_G(\mathbf{a}'\mathbf{X})}.$$

Since $m_1(\operatorname{median}_G(\mathbf{a}'\mathbf{X}), \varepsilon) \le \mathrm{MAD}_G(\mathbf{a}'\mathbf{X})$, we get

$$\begin{aligned} \sup_{\|\mathbf{a}\|=1} \frac{|\operatorname{median}_G(\mathbf{a}'(\mathbf{X} - \boldsymbol{\nu}))|}{\mathrm{MAD}_G(\mathbf{a}'\mathbf{X})} &\le \sup_{\|\mathbf{a}\|=1} \frac{|\operatorname{median}_G(\mathbf{a}'\mathbf{X})|}{\mathrm{MAD}_G(\mathbf{a}'\mathbf{X})} \\ &\le \frac{|\operatorname{median}_G(\mathbf{a}'\mathbf{X}))|}{m_1(\operatorname{median}_G(\mathbf{a}'\mathbf{X}), \varepsilon)} \le d_0. \end{aligned} \tag{28}$$

From (27) and (28) we get that $\|\boldsymbol{\nu}\| \le d_1 + d_0 d_2$ and the result follows. □

Lemmas 3–6 show that the maximum bias produced by point-mass contaminations attains the upper bound given in Lemma 2.

The following lemma is immediate.

LEMMA 3. *Consider a distribution function L on $\mathbb{R}$ and suppose that $a = L^{-1}(1/(2(1-\varepsilon)))$ and $b = L^{-1}((1-2\varepsilon)/(2(1-\varepsilon)))$ are well defined. Let $c \in \mathbb{R}$ and $L^* = (1-\varepsilon)L + \delta_c$. Then*

$$\text{median}_{L^*}(X) = \min\bigl(b, \max(a, c)\bigr).$$

Given $\mathbf{c} \in \mathbb{R}^p$, let $G_\mathbf{c} = (1-\varepsilon)F_\mathbf{0} + \varepsilon\delta_\mathbf{c}$.

LEMMA 4. *If $\mathbf{a} \in \mathbb{S}^{p-1}$, then*

$$\text{median}_{G_\mathbf{c}}(\mathbf{a}'\mathbf{X}) = \min\bigl(-d_1, \max(d_1, \mathbf{a}'\mathbf{c})\bigr),$$

$$\text{MAD}_{G_\mathbf{c}}(\mathbf{a}'\mathbf{X}) = \begin{cases} d_2, & \mathbf{a}'\mathbf{c} \le -d_1 - d_2, \\ |\mathbf{a}'\mathbf{c} + d_1|, & -d_1 - d_2 \le \mathbf{a}'\mathbf{c} \le -d_1 - d_3, \\ d_3, & -d_1 - d_3 \le \mathbf{a}'\mathbf{c} \le -d_1, \\ m_1(\mathbf{a}'\mathbf{c}, \varepsilon), & |\mathbf{a}'\mathbf{c}| \le d_1, \\ d_3, & d_1 \le \mathbf{a}'\mathbf{c} \le d_1 + d_3, \\ |\mathbf{a}'\mathbf{c} - d_1|, & d_1 + d_3 \le \mathbf{a}'\mathbf{c} \le d_1 + d_2, \\ d_2, & \mathbf{a}'\mathbf{c} \ge d_1 + d_2, \end{cases}$$

and therefore

$$\begin{aligned} h(\boldsymbol{\nu}, \mathbf{a}, G_\mathbf{c}) &= \left| \frac{\mathbf{a}'\boldsymbol{\nu} - \text{median}_{G_\mathbf{c}}(\mathbf{a}'\mathbf{X})}{\text{MAD}_{G_\mathbf{c}}(\mathbf{a}'\mathbf{X})} \right| \\ &= \begin{cases} |\mathbf{a}'\boldsymbol{\nu} + d_1|/d_2, & \mathbf{a}'\mathbf{c} \le -d_1 - d_2, \\ |\mathbf{a}'\boldsymbol{\nu} + d_1|/|\mathbf{a}'\mathbf{c} + d_1|, & -d_1 - d_2 \le \mathbf{a}'\mathbf{c} \le -d_1 - d_3, \\ |\mathbf{a}'\boldsymbol{\nu} + d_1|/d_3, & -d_1 - d_3 \le \mathbf{a}'\mathbf{c} \le -d_1, \\ |\mathbf{a}'\boldsymbol{\nu} - \mathbf{a}'\mathbf{c}|/m_1(\mathbf{a}'\mathbf{c}, \varepsilon), & |\mathbf{a}'\mathbf{c}| \le d_1, \\ |\mathbf{a}'\boldsymbol{\nu} - d_1|/d_3, & d_1 \le \mathbf{a}'\mathbf{c} \le d_1 + d_3, \\ |\mathbf{a}'\boldsymbol{\nu} - d_1|/|\mathbf{a}'\mathbf{c} - d_1|, & d_1 + d_3 \le \mathbf{a}'\mathbf{c} \le d_1 + d_2, \\ |\mathbf{a}'\boldsymbol{\nu} - d_1|/d_2, & \mathbf{a}'\mathbf{c} \ge d_1 + d_2. \end{cases} \end{aligned}$$

PROOF. The lemma follows from Lemma 3. □

The following lemma gives the value of $\boldsymbol{\nu}$ minimizing $v(\boldsymbol{\nu}, G_\mathbf{c}) = \sup_{\mathbf{a}\in\mathbb{S}^{p-1}} h(\boldsymbol{\nu}, \mathbf{a}, G_\mathbf{c})$ when the search for the minimum is restricted to points of the form $\boldsymbol{\nu} = t\mathbf{c}$. Lemma 6 will show that this constrained estimator is actually the projection estimate; that is, the MP estimate has the same direction as the contaminating point. We will focus on the case $\|\mathbf{c}\| \ge d_1 + d_2$ since it is going to yield the maximum bias as $\|\mathbf{c}\|$ tends to infinity.

LEMMA 5. *Let $\mathbf{c} \in \mathbb{R}^p$, $l(\mathbf{c}) = \min_{t\in\mathbb{R}^+} v(t\mathbf{c}, G_\mathbf{c})$ and $T_1(\mathbf{c}) = \arg\min_{t\in\mathbb{R}^+} v(t\mathbf{c}, G_\mathbf{c})$. Then if $\|\mathbf{c}\| > d_1 + d_2$,*

$$l(\mathbf{c}) = \frac{d_0(\|\mathbf{c}\| - d_1)}{\|\mathbf{c}\| + d_0 d_2} \quad \text{and} \quad T_1(\mathbf{c}) = \frac{d_1 + d_0 d_2}{\|\mathbf{c}\| + d_0 d_2}.$$

PROOF. From Lemma 4 we get

$$v(t\mathbf{c},\mathbf{c}) = \max\left\{\left|t\frac{d_1+d_3}{d_2}-\frac{d_1}{d_2}\right|, \left|t\frac{\|\mathbf{c}\|}{d_2}-\frac{d_1}{d_2}\right|, \left|t\frac{d_1+d_3}{d_3}-\frac{d_1}{d_3}\right|, \left|t\frac{d_1}{d_3}-\frac{d_1}{d_3}\right|, |td_0-d_0|\right\}.$$

Since

$$\left|t\frac{d_1+d_3}{d_3}-\frac{d_1}{d_3}\right| > \left|t\frac{d_1+d_3}{d_2}-\frac{d_1}{d_2}\right| \quad \text{and} \quad |td_0-d_0| > \left|t\frac{d_1}{d_3}-\frac{d_1}{d_3}\right|,$$

we get

$$\begin{aligned} v(t\mathbf{c},G_{\mathbf{c}}) &= \max\left\{\left|t\frac{\|\mathbf{c}\|}{d_2}-\frac{d_1}{d_2}\right|, \left|t\frac{d_1+d_3}{d_3}-\frac{d_1}{d_3}\right|, |td_0-d_0|\right\} \\ &= \max\{q_1(t), q_2(t), q_3(t)\}. \end{aligned}$$

Note that $q_1(d_1/(d_1+d_3)) = q_2(d_1/\|\mathbf{c}\|) = q_3(1) = 0$ and $d_1/\|\mathbf{c}\| < d_1/(d_1+d_3) < 1$. Define t_0 and t_1 as the points such that $q_1(t_0) = q_3(t_0)$ and $q_2(t_1) = q_3(t_1)$. Then

$$t_0 = \frac{d_1+d_0d_3}{d_1+d_3+d_0d_3}, \qquad t_1 = \frac{d_1+d_0d_2}{\|\mathbf{c}\|+d_0d_2}.$$

Since $\|\mathbf{c}\| > d_1+d_2$, we have $t_0 > t_1$ and this implies

$$v(t\mathbf{c},G_{\mathbf{c}}) = \begin{cases} d_0 - td_0, & \text{if } t < t_1, \\ t\dfrac{\|\mathbf{c}\|}{d_2}-\dfrac{d_1}{d_2}, & \text{otherwise.} \end{cases}$$

Therefore

$$T_1(\mathbf{c}) = t_1 = \frac{d_1+d_0d_2}{\|\mathbf{c}\|+d_0d_2}.$$

This completes the proof of Lemma 5. □

The following lemma shows that the estimator follows the direction of the point-mass $\mathbf{c}$ if $\|\mathbf{c}\| \ge d_1+d_2$.

LEMMA 6. *Let $\boldsymbol{\nu} \in \mathbb{R}^p - \{\mathbf{0}\}$ such that $\boldsymbol{\nu}'\mathbf{c}/\|\mathbf{c}\|\,\|\boldsymbol{\nu}\| < 1$ and $\|\mathbf{c}\| \ge d_1+d_2$. Then $v(\boldsymbol{\nu},G_{\mathbf{c}}) > l(\mathbf{c})$.*

PROOF. We can write $\boldsymbol{\nu} = t\|\mathbf{c}\|\boldsymbol{\nu}^*$, $t \in \mathbb{R}^+$, $\boldsymbol{\nu}^* \in \mathbb{S}^{p-1}$. Let $\boldsymbol{\zeta} = c_1\mathbf{c}/\|\mathbf{c}\| + c_2\boldsymbol{\nu}^*$ be such that $\boldsymbol{\zeta}'\mathbf{c} = 0$ and $\boldsymbol{\zeta} \in S^{p-1}$. If $\boldsymbol{\nu}^{*\prime}\mathbf{c} \ge 0$, then we choose $c_1 \le 0$ and $c_2 \ge 0$. If $\boldsymbol{\nu}^{*\prime}\mathbf{c} < 0$, then we choose $c_1 > 0$ and $c_2 > 0$. Let $A = \{\mathbf{a} \in \mathbb{R}^p : \mathbf{a} = a_1\mathbf{c}/\|\mathbf{c}\| + a_2\boldsymbol{\zeta}\}$. Define the following sets: $B_1 = \{\mathbf{a} \in A \cap \mathbb{S}^{p-1} : a_1 \ge 0,\ a_2 \ge 0\}$,

$B_2 = \{\mathbf{a} \in A \cap \mathbb{S}^{p-1} : a_1 \leq 0,\ a_2 \geq 0\}$, $B_3 = \{\mathbf{a} \in A \cap \mathbb{S}^{p-1} : a_1 \leq 0,\ a_2 \leq 0\}$ and $B_4 = \{\mathbf{a} \in A \cap \mathbb{S}^{p-1} : a_1 \geq 0,\ a_2 \leq 0\}$.

For the sake of simplicity we will only consider henceforth the case $\boldsymbol{\nu}^* \in B_1$; other cases can be treated similarly. We have to consider several cases for $\boldsymbol{\nu}^*$. Let us suppose first that $|\boldsymbol{\nu}^{*\prime}\mathbf{c}| \leq d_1$. Take

$$t_1 = \frac{d_1 + d_0 d_2}{\|\mathbf{c}\| + d_0 d_2}.$$

Let $\mathbf{b}_2 \in B_2$ be such that $\cos(\mathbf{b}_2, \mathbf{c}) = -d_1/\|\mathbf{c}\|$. Then

$$\begin{aligned} v(\boldsymbol{\nu}, G_{\mathbf{c}}) &\geq \sup_{\mathbf{a} \in \mathbb{S}^{p-1}, -d_1 \geq \mathbf{a}'\mathbf{c} \geq -d_1 - d_3} \frac{1}{d_3} \big| t\|\mathbf{c}\| \cos(\mathbf{a}, \boldsymbol{\nu}^*) + d_1 \big| \\ &\geq \frac{1}{d_3} \big| t\|\mathbf{c}\| \cos(\mathbf{b}_2, \boldsymbol{\nu}^*) + d_1 \big|. \end{aligned}$$

We also have

$$\cos(\mathbf{b}_2, \boldsymbol{\nu}^*) = -\frac{d_1}{\|\mathbf{c}\|} \cos(\mathbf{c}, \boldsymbol{\nu}^*) + \sqrt{1 - \frac{d_1^2}{\|\mathbf{c}\|^2}} \sin(\mathbf{c}, \boldsymbol{\nu}^*) \geq \left(1 - \frac{2d_1^2}{\|\mathbf{c}\|^2}\right). \tag{29}$$

Since $\|\mathbf{c}\| \geq d_1 + d_2$, then $d_1/\|\mathbf{c}\| \leq 0.5$. Therefore, $\cos(\mathbf{b}_2, \boldsymbol{\nu}^*) \geq 0$. Consequently, for $t \geq t_1$, (29) and the fact that $t_1 d_1 < \|\mathbf{c}\|$ let us conclude that

$$v(\boldsymbol{\nu}, G_{\mathbf{c}}) \geq \frac{1}{d_3} \big| t_1 \|\mathbf{c}\| - d_1 \big| > \frac{1}{d_2} \big| t_1 \|\mathbf{c}\| - d_1 \big| = v(t_1 \mathbf{c}, G_{\mathbf{c}}) = l(\mathbf{c}). \tag{30}$$

Consider now the case $t < t_1$. We always have

$$\frac{|t\|\mathbf{c}\| \cos(\mathbf{a}, \boldsymbol{\nu}^*) - \|\mathbf{c}\| \cos(\mathbf{a}, \mathbf{c})|}{m_1(\mathbf{a}'\mathbf{c}, \varepsilon)} = \frac{|\mathbf{a}'\mathbf{c}|}{m_1(\mathbf{a}'\mathbf{c}, \varepsilon)} \left| t \frac{\cos(\mathbf{a}, \boldsymbol{\nu}^*)}{\cos(\mathbf{a}, \mathbf{c})} - 1 \right|. \tag{31}$$

Define the following sets for $j \in \{1, 2, 3, 4\}$,

$$M_j = \left\{ \mathbf{a} \in B_j \cap [|\mathbf{a}'\mathbf{c}| \leq d_1] : \max_{\mathbf{b} \in B_j \cap [|\mathbf{a}'\mathbf{c}| \leq d_1]} \frac{|\mathbf{b}'\mathbf{c}|}{m_1(|\mathbf{b}'\mathbf{c}|, \varepsilon)} = \frac{|\mathbf{a}'\mathbf{c}|}{m_1(|\mathbf{a}'\mathbf{c}|, \varepsilon)} \right\}. \tag{32}$$

We easily observe that $M_3 = -M_1$ and $M_4 = -M_2$. Let us take $\mathbf{a}_0 \in M_2$. If $0 \leq \cos(\mathbf{a}_0, \boldsymbol{\nu}^*)$, then we get, for $t < t_1 \leq 1$,

$$t \frac{\cos(\mathbf{a}_0, \boldsymbol{\nu}^*)}{\cos(\mathbf{a}_0, \mathbf{c})} - 1 < t - 1 \leq 0. \tag{33}$$

If $\cos(\mathbf{a}_0, \boldsymbol{\nu}^*) < 0$, we take $-\mathbf{a}_0 \in M_4$ such that

$$1 - t \frac{\cos(\mathbf{a}_0, \boldsymbol{\nu}^*)}{\cos(\mathbf{a}_0, \mathbf{c})} > 1 - t \geq 0. \tag{34}$$

Then, if $t < t_1$, we get

$$v(\boldsymbol{\nu}, G_{\mathbf{c}}) > d_0 |t - 1| = v(t\mathbf{c}, G_{\mathbf{c}}) \geq l(\mathbf{c}). \tag{35}$$

Suppose now that $d_1 \le |\boldsymbol{\nu}^{*\prime}\mathbf{c}| \le d_1 + d_3$. Since $\boldsymbol{\nu}^* \in \mathbb{S}^{p-1} \cap \{\mathbf{a} : d_1 \le \mathbf{a}'\mathbf{c} \le d_1 + d_3\}$ we get

$$\begin{aligned} v(\boldsymbol{\nu}, G_{\mathbf{c}}) &\ge \sup_{\mathbf{a}\in\mathbb{S}^{p-1},\, d_1 \le \mathbf{a}'\mathbf{c} \le d_1+d_3} \frac{|t\|\mathbf{c}\|\cos(\mathbf{a}, \boldsymbol{\nu}^*) - d_1|}{d_3} \\ &\ge \frac{1}{d_3}\big|\|\mathbf{c}\|t - d_1\big| \ge \frac{1}{d_2}\big|\|\mathbf{c}\|t - d_1\big|. \end{aligned} \tag{36}$$

If $t \ge t_1$, then

$$\frac{1}{d_2}\big|\|\mathbf{c}\|t - d_1\big| = v(t\mathbf{c}, G_{\mathbf{c}}) \ge l(\mathbf{c}). \tag{37}$$

If $t < t_1$, we take $\mathbf{a}_0 \in M_4$. Therefore, either $0 < \cos(\mathbf{a}_0, \boldsymbol{\nu}^*) < \cos(\mathbf{a}_0, \mathbf{c})$ or $\cos(\mathbf{a}_0, \boldsymbol{\nu}^*) < 0 < \cos(\mathbf{a}_0, \mathbf{c})$. From (31), (33) and (34) we conclude

$$v(\boldsymbol{\nu}, G_{\mathbf{c}}) \ge \frac{|\mathbf{a}_0'\mathbf{c}|}{m_1(\mathbf{a}_0'\mathbf{c}, \varepsilon)}\left|t\frac{\cos(\mathbf{a}_0, \boldsymbol{\nu}^*)}{\cos(\mathbf{a}_0, \mathbf{c})} - 1\right| > d_0|t-1| = v(t\mathbf{c}, G_{\mathbf{c}}) \ge l(\mathbf{c}). \tag{38}$$

Then, (36)–(38) entail that $v(\boldsymbol{\nu}, G_{\mathbf{c}}) > l(\mathbf{c})$ provided $d_1 \le |\boldsymbol{\nu}^{*\prime}\mathbf{c}| \le d_1 + d_3$.

The proof for the case $d_1 + d_3 \le |\boldsymbol{\nu}^{*\prime}\mathbf{c}| \le d_1 + d_2$ is similar. We should distinguish the cases $t \ge t_1$ and $t \le t_1$. In the case $t \ge t_1$ we obtain

$$\begin{aligned} v(\boldsymbol{\nu}, G_{\mathbf{c}}) &\ge \sup_{\mathbf{a}\in\mathbb{S}^{p-1}, d_1+d_3 \le \mathbf{a}'\mathbf{c} \le d_1+d_2} \frac{|t\|\mathbf{c}\|\cos(\mathbf{a}, \boldsymbol{\nu}^*) - d_1|}{|\|\mathbf{c}\|\cos(\mathbf{a}, \mathbf{c}) - d_1|} \\ &\ge \frac{1}{|d_1 + d_2 - d_1|}\big|\|\mathbf{c}\|t - d_1\big| = \frac{1}{d_2}\big|\|\mathbf{c}\|t - d_1\big| = v(t\mathbf{c}, G_{\mathbf{c}}) \ge l(\mathbf{c}), \end{aligned}$$

since $\boldsymbol{\nu}^* \in \mathbb{S}^{p-1} \cap \{\mathbf{a} : d_1 + d_3 \le \mathbf{a}'\mathbf{c} \le d_1 + d_2\}$. In the case $t \le t_1$, inequalities (33)–(35) are still valid and consequently $v(\boldsymbol{\nu}, G_{\mathbf{c}}) > l(\mathbf{c})$. We cope with the case $d_1 + d_2 \le |\boldsymbol{\nu}^{*\prime}\mathbf{c}|$ similarly. This concludes the proof. □

PROOF OF THEOREM 4. By Lemma 2, $B(\mathbf{T}_{\mathrm{MP}}, \varepsilon, F_{\mathbf{0}}) \le d_0 d_2 + d_1$ and, by Lemmas 5 and 6,

$$B(\mathbf{T}_{\mathrm{MP}}, \varepsilon, F_{\mathbf{0}}) \ge \lim_{u\to\infty} u\frac{d_1 + d_0 d_2}{u + d_0 d_2} = d_0 d_2 + d_1,$$

and therefore (i) follows. Part (ii) is immediate from (i). To prove (iii) observe that

$$\begin{aligned} &\frac{\partial B(\mathbf{T}_{\mathrm{MP}}, \varepsilon, F_{\mathbf{0}})}{\partial \varepsilon} \\ &\quad = \frac{\partial d_1(\varepsilon, H_0)}{\partial \varepsilon}\bigg|_{\varepsilon=0}\left(1 + \frac{d_2(0, H_0)}{d_3(0, H_0)}\right) + \frac{\partial (d_2(\varepsilon, H_0)/d_3(\varepsilon, H_0))}{\partial \varepsilon}\bigg|_{\varepsilon=0}. \end{aligned}$$

Then, since $d_1(0, H_0) = 0$ and $d_2(0, H_0) = d_3(0, H_0)$, we get

$$\gamma(\mathbf{T}_{\mathrm{MP}}, F_{\mathbf{0}}) = \frac{1}{h_0(0)}. \tag{39}$$

On the other hand, we also have that

$$\lim_{\varepsilon\to 0} d_1(\varepsilon, H_0) = 0, \qquad \lim_{\varepsilon\to 0} d_3(\varepsilon, H_0) = \lim_{\varepsilon\to 0} d_2(\varepsilon, H_0) = l_2 > 0.$$

Let $\|\mathbf{c}_1\| > l_2$. Then, by Lemmas 5 and 6, we get

$$\begin{aligned}&\frac{\mathbf{T}((1-\varepsilon)F_{\mathbf{0}} + \varepsilon\delta_{\mathbf{c}_1}) - \mathbf{T}(F_{\mathbf{0}})}{\varepsilon}\\ &\qquad = \frac{(d_1(\varepsilon, H_0) - d_1(0, H_0))}{\varepsilon}\frac{(1 + d_2(\varepsilon, H_0)\ /d_3(\varepsilon, H_0))}{\|\mathbf{c}_1\| + d_1(\varepsilon, H_0)d_2(\varepsilon, H_0)/d_3(\varepsilon, H_0)}\mathbf{c}_1.\end{aligned}$$

As $\varepsilon \to 0$ we obtain

$$\lim_{\varepsilon\to 0} \frac{\mathbf{T}((1-\varepsilon)F_{\mathbf{0}} + \varepsilon\delta_{\mathbf{c}_1}) - \mathbf{T}(F_{\mathbf{0}})}{\varepsilon} = 2\frac{\partial d_1(\varepsilon, H_0)}{\partial\varepsilon}\bigg|_{\varepsilon=0}\frac{\mathbf{c}_1}{\|\mathbf{c}_1\|}.$$

Then

$$\begin{aligned}\gamma^*(\mathbf{T}_{\mathrm{MP}}, F_{\mathbf{0}}) &= \sup_{\mathbf{c}\in\mathbb{R}^p}\left\|\lim_{\varepsilon\to 0}\frac{\mathbf{T}((1-\varepsilon)F_{\mathbf{0}} + \varepsilon\delta_{\mathbf{c}_1}) - \mathbf{T}(F_{\mathbf{0}})}{\varepsilon}\right\|\\ &\geq 2\frac{\partial d_1(\varepsilon, H_0)}{\partial\varepsilon}\bigg|_{\varepsilon=0} = \gamma(\mathbf{T}_{\mathrm{MP}}, F_{\mathbf{0}}).\end{aligned} \tag{40}$$

Since $\gamma^*(\mathbf{T}_{\mathrm{MP}}, F_{\mathbf{0}}) \leq \gamma(\mathbf{T}_{\mathrm{MP}}, F_{\mathbf{0}})$, (iii) follows from (39) and (40). □

Acknowledgments. This research was partially done while the first author was visiting the Department of Statistics, University of British Columbia, Vancouver, Canada. We thank both referees for their criticism and useful comments which contributed to the improvement of the paper.

REFERENCES

ADROVER, J. G., MARONNA, R. A. and YOHAI, V. J. (2002). Relationships between maximum depth and projection regression estimates. *J. Statist. Plann. Inference* **105** 363–375.

BUTLER, R. W., DAVIES, P. L. and JHUN, M. (1993). Asymptotics for the minimum covariance determinant estimator. *Ann. Statist.* **21** 1385–1400.

CHEN, Z. and TYLER, D. (2002). The influence function and maximum bias of Tukey's median. *Ann. Statist.* **30** 1734–1756.

CROUX, C., HAESBROECK, G. and ROUSSEEUW, P. J. (1997). Location adjustment for the minimum volume ellipsoid estimator. Technical report, Free Univ. Brussels (ULB). Available at http://homepages. ulb.ac.be/~ccroux/public.htm.

DAVIES, P. L. (1992). The asymptotics of Rousseeuw's minimum volume ellipsoid estimator. *Ann. Statist.* **20** 1828–1843.

DONOHO, D. L. (1982). Breakdown properties of multivariate location estimators. Ph.D. qualifying paper, Dept. Statistics, Harvard Univ.

DONOHO, D. L. and GASKO, M. (1992). Breakdown properties of location estimates based on half space depth and projected outlyingness. *Ann. Statist.* **20** 1803–1827.

HAMPEL, F. R. (1971). A general qualitative definition of robustness. *Ann. Math. Statist.* **42** 1887–1896.

HAMPEL, F. R. (1974). The influence curve and its role in robust estimation. *J. Amer. Statist. Assoc.* **69** 383–393.

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. (1992). Robust direction estimation. *Ann. Statist.* **20** 351–369.

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.

MARONNA, R. A., STAHEL, W. A. and YOHAI, V. J. (1992). Bias-robust estimators of multivariate scatter based on projections. *J. Multivariate Anal.* **42** 141–161.

MARONNA, R. A. and YOHAI, V. J. (1993). Bias-robust estimates of regression based on projections. *Ann. Statist.* **21** 965–990.

MARONNA, R. A. and YOHAI, V. J. (1995). The behavior of the Stahel–Donoho robust multivariate estimator. *J. Amer. Statist. Assoc.* **90** 330–341.

RIEDEL, M. (1991). Bias robustness in parametric models generated by groups. *Statistics* **22** 559–578.

ROUSSEEUW, P. J. (1985). Multivariate estimation with high breakdown point. In *Mathematical Statistics and Its Applications* (W. Grossman, G. Pflug, I. Vincze and W. Wertz, eds.) 283–297. Reidel, Dordrecht.

ROUSSEEUW, P. J. and HUBERT, M. (1999). Regression depth. *J. Amer.Statist. Assoc*. **94** 388–402.

SCHEFFÉ, H. (1947). A useful convergence theorem for probability distributions. *Ann. Math. Statist.* **18** 434–438.

STAHEL, W. A. (1981). Breakdown of covariance estimators. Research Report 31, Fachgruppe für Statistik, ETH, Zürich.

TYLER, D. E. (1994). Finite sample breakdown points of projection based multivariate location and scatter statistics. *Ann. Statist.* **22** 1024–1044.

TYLER, D. E. (2000). Personal communication.

FACULTAD DE MATEMÁTICA
ASTRONOMÍA Y FÍSICA
UNIVERSIDAD NACIONAL DE CÓRDOBA
AVDA. MEDINA ALLENDE Y HAYA DE LA TORRE
CIUDAD UNIVERSITARIA
(5000) CÓRDOBA
ARGENTINA
E-MAIL: Jorge.Adrover@famaf.unc.edu.ar

DEPARTAMENTO DE MATEMÁTICA
FACULTAD DE CIENCIAS
EXACTAS Y NATURALES
UNIVERSIDAD DE BUENOS AIRES
CIUDAD UNIVERSITARIA, PABELLÓN 1
1428 BUENOS AIRES
ARGENTINA
E-MAIL: vyohai@dm.uba.ar