*The Annals of Statistics*
1993, Vol. 21, No. 2, 965–990

# BIAS-ROBUST ESTIMATES OF REGRESSION BASED ON PROJECTIONS[1]

BY RICARDO A. MARONNA AND VICTOR J. YOHAI

*University of La Plata and C.I.C.P.B.A, and University of San Andrés, University of Buenos Aires and CONICET*

A new class of bias-robust estimates of multiple regression is introduced. If $y$ and $x$ are two real random variables, let $T(y, x)$ be a univariate robust estimate of regression of $y$ on $x$ through the origin. The regression estimate $\mathbf{T}(y, \mathbf{x})$ of a random variable $y$ on a random vector $\mathbf{x} = (x_1, \ldots, x_p)'$ is defined as the vector $\mathbf{t} \in \Re^p$ which minimizes $\sup_{\|\boldsymbol{\lambda}\|=1} |T(y - \mathbf{t}'\mathbf{x}, \boldsymbol{\lambda}'\mathbf{x})| s(\boldsymbol{\lambda}'\mathbf{x})$, where $s$ is a robust estimate of scale. These estimates, which are called projection estimates, are regression, affine and scale equivariant. When the univariate regression estimate is $T(y, x) = \text{median}(y/x)$, the resulting projection estimate is highly bias-robust. In fact, we find an upper bound for its maximum bias in a contamination neighborhood, which is approximately twice the minimum possible value of this maximum bias for any regression and affine equivariant estimate. The maximum bias of this estimate in a contamination neighborhood compares favorably with those of Rousseeuw's least median squares estimate and of the most bias-robust GM-estimate. A modification of this projection estimate, whose maximum bias for a multivariate normal with mass-point contamination is very close to the minimax bound, is also given. Projection estimates are shown to have a rate of consistency of $n^{1/2}$. A computational version of these estimates, based on subsampling, is given. A simulation study shows that its small sample properties compare very favorably to those of other robust regression estimates.

**1. Introduction.** An important measure of the robustness of an estimator is its maximum bias in a contamination neighborhood of a central model. Two different approaches have been used to deal with the maximum asymptotic bias:

1. *The local approach* which studies the bias caused by a small fraction $\varepsilon$ of contamination. This approach is based on the concept of influence curve introduced by Hampel (1974), which gives a linear approximation which is valid for infinitesimal $\varepsilon$, to the bias due to contamination.
2. *The global approach* which takes into account the bias caused by both small and large values of $\varepsilon$. One important measure of global robustness is the breakdown point introduced by Hampel (1971). In recent years several

Received July 1990; revised June 1992.
[1]This research was partially done while the authors were visiting the Department of Statistics at the University at Washington supported by ONR Contracts CN00014-88-K0265 and CN00014-91-J-1074.
*AMS* 1991 *subject classifications*. Primary 62F35; secondary 62J05.
*Key words and phrases.* Minimax bias, regression, robust estimates.

proposals of high breakdown estimates for regression have been made. We can cite among others: the least median of squares (LMS-estimate) [Rousseeuw (1984)], the least trimmed squares (LTS) [Rousseeuw (1984)], MM-estimates [Yohai (1987)] and tau-estimates [Yohai and Zamar (1988)].

Both the influence function and the breakdown point may be insufficient to adequately describe the bias. It is more reliable to directly deal with the bias under contamination, summarizing it through the maximum asymptotic bias for fractions of contamination smaller than the breakdown point. In this setting it seems natural to search for estimates which minimize the maximum asymptotic bias in an $\varepsilon$-contamination neighborhood (minimax-bias estimates). Several authors have studied the robustness problem by using a minimax-bias approach. Huber (1964) has shown that the median is the minimax-bias estimate for location in the class of shift equivariant estimators. Martin and Zamar (1989) have found minimax-bias estimators for scale in the class of M-estimates. Riedel (1991) has shown the existence of minimax-bias equivariant estimates for general models with a parameter space endowed with a group structure. Donoho and Liu (1988) have shown that minimum distance estimates are highly bias-robust for some parametric and semiparametric models.

Martin, Yohai and Zamar (1989) have found minimax-bias estimators for two different classes of regression estimates: M-estimates with general scale and GM-estimates. In particular, they show that for $p = 1$ and no intercept, the median of the slopes, which is the minimax GM-estimate, is also minimax in the class of all regression- and affine-equivariant estimates.

In this article we introduce a new class of regression estimators which are highly bias-robust. They are based on (univariate) regressions of the response variable with respect to all one-dimensional projections of the carriers; and will be called *projection estimators* (or P-estimators for short).

These estimates have simultaneously good local and global robustness properties, since they possess both bounded influence and high breakdown point. Numerical computations of the asymptotic bias and Monte Carlo estimations of the MSE under contamination for finite sample sizes, show that P-estimates compare favorably with other robust estimates.

In Section 2 we define the target model and give formal definitions of the robustness concepts used in the paper. In Section 3 we define P-estimators and prove their robustness properties. We also find a lower bound for the maximum bias of any equivariant estimate, and prove that there exists a P-estimate whose maximum bias is approximately twice this bound. In Section 4 we give a new estimate which is obtained by modifying a P-estimate. The maximum asymptotic bias of this estimate is computed numerically for a multivariate normal central model with mass-point contamination, and shown to be very close to the minimax bound obtained in Section 3. In Section 5 we give a computable version of the P-estimates and deal with its finite-sample breakdown point. In Section 6 we prove that these estimators have a rate of consistency $n^{1/2}$ and we give a further result on minimax bias. In Section 7 we compare the P-estimates with the LMS- and the minimax GM-estimates by

computing the asymptotic maximum biases and by Monte Carlo simulation. The Appendix contains some proofs.

**2. Target model and robustness definitions.** We assume the target model is the linear model

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

where $'$ denotes transpose, $\mathbf{x} = (x_1, \ldots, x_p)'$ is a random vector in $\mathfrak{R}^p$ with distribution $G_0$, $\boldsymbol{\alpha}_0 = (\alpha_{10}, \ldots, \alpha_{p0})'$ is the vector of the true regression parameters, and the error $u$ is a random variable with distribution $F_0$, independent of $\mathbf{x}$. We shall call $H_0$ the joint distribution of $(y, \mathbf{x})$ under this model, so that

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

Let $\mathbf{T}$ be an $\mathfrak{R}^p$ valued functional defined on the set of distributions $H$ on $\mathfrak{R}^{p+1}$. Then, given a sample $z_1 = (y_1, \mathbf{x}_1), \ldots, z_n = (y_n, \mathbf{x}_n)$, we define the corresponding estimate of $\boldsymbol{\alpha}_0$ as $\mathbf{T}_n = \mathbf{T}(H_n)$, where $H_n$ is the empirical distribution of the sample. This estimate is Fisher consistent if $\mathbf{T}(H_0) = \boldsymbol{\alpha}_0$.

As a general notation, if $g$ and $h$ are any functions defined respectively on $\mathfrak{R}^{p+1}$ and $\mathfrak{R}^p$, then $\mathscr{L}(H, g(y, \mathbf{x}), h(\mathbf{x}))$ will denote the joint distribution of $g(y, \mathbf{x})$ and $h(\mathbf{x})$ when $(y, \mathbf{x})$ have joint distribution $H$. The $H$ will be dropped when this causes no confusion.

An estimate $\mathbf{T}$ is regression, affine and scale equivariant, respectively, if $\mathbf{T}(\mathscr{L}(y + \mathbf{x}'\mathbf{t}, \mathbf{x})) = \mathbf{T}(\mathscr{L}(y, \mathbf{x})) + \mathbf{t}$ for all $\mathbf{t} \in \mathfrak{R}^p$, $\mathbf{T}(\mathscr{L}(y, R\mathbf{x})) = R'^{-1}\mathbf{T}(\mathscr{L}(y, \mathbf{x}))$ for all nonsingular $p \times p$ matrices $R$, and $\mathbf{T}(\mathscr{L}(by, \mathbf{x})) = b\mathbf{T}(\mathscr{L}(y, \mathbf{x}))$ for all $b \in \mathfrak{R}$.

In order to study the robustness of an estimate, we consider the $\varepsilon$-contamination neighborhood of $H_0$

$$\mathscr{V}_{H_0, \varepsilon} = \{H\colon H = (1 - \varepsilon)H_0 + \varepsilon H_1,\ H_1 \text{ any distribution}\}.$$

Let $V(G)$ be an affine-equivariant scatter matrix, that is, for each distribution $G$ on $\mathfrak{R}^p$, $V(G)$ is a symmetric, nonnegative definite $p \times p$-matrix satisfying

$$V(\mathscr{L}(R\mathbf{x})) = RV(\mathscr{L}(\mathbf{x}))R'.$$

The bias of $\mathbf{T}$, at $H \in \mathscr{V}_{H_0, \varepsilon}$ is defined by

$$b(\mathbf{T}, H) = \left[(\mathbf{T}(H) - \boldsymbol{\alpha}_0)'V(G_0)^{-1}(\mathbf{T}(H) - \boldsymbol{\alpha}_0)\right]^{1/2}. \tag{2.3}$$

Therefore if $\mathbf{T}$ is regression and affine equivariant, the bias will be regression and affine invariant.

When $T$ is an estimate of univariate regression through the origin, the bias is defined by

$$b(T, H) = |T(H) - \alpha_0|. \tag{2.4}$$

The robustness of an estimate can be measured by the maximum bias of $\mathbf{T}$ for $H \in \mathscr{V}_{H_0,\varepsilon}$, that is, by

$$\beta(\mathbf{T}, H_0, \varepsilon) = \sup\{b(\mathbf{T}, H): H \in \mathscr{V}_{H_0,\varepsilon}\}. \tag{2.5}$$

Since we are going to deal only with affine equivariant estimates, we will henceforth assume, without loss of generality, that $V(G_0) = I$.

Given a class of regression and affine equivariant estimates $\mathscr{T}$, a minimax bias estimate for the target model $H_0$ is an estimate $\mathbf{T}^* \in \mathscr{T}$ satisfying

$$\beta(\mathbf{T}^*, H_0, \varepsilon) = \inf_{\mathbf{T}\in\mathscr{T}} \beta(\mathbf{T}, H_0, \varepsilon).$$

Another robustness measure related to $\beta(\mathbf{T}, H_0, \varepsilon)$ is the asymptotic breakdown point introduced by Hampel (1971). It is defined here by

$$\varepsilon^*(\mathbf{T}, H_0) = \inf\{\varepsilon: \beta(\mathbf{T}, H_0, \varepsilon) = \infty\},$$

which gives the smallest fraction of outliers which make the estimate unbounded.

As a measure of the bias for "infinitesimal" values of $\varepsilon$, we shall use

$$\gamma^*(\mathbf{T}, H_0) = \frac{\partial}{\partial\varepsilon}\beta(\mathbf{T}, H_0, \varepsilon)\Big|_{\varepsilon=0}.$$

In sufficiently regular cases $\gamma^*$ coincides with Hampel's (1974) definition of gross error sensitivity.

## 3. Regression estimates based on projections.

3.1. *Definition of projection estimates.* Let $T$ be an estimating functional of univariate regression through the origin, which is regression, scale and affine equivariant. Then, for any pair $(y, x)$ of random variables

$$T(\mathscr{L}(b(y + tx), ax)) = \frac{b}{a}(T(\mathscr{L}(y, x)) + t). \tag{3.1}$$

Let $s$ be a scale estimating functional, so that for any random variable $z$

$$s(\mathscr{L}(az)) = |a|s(\mathscr{L}(z)). \tag{3.2}$$

The type 1 projection estimate (P1-estimate) of multiple regression associated with $T$ and $s$ is defined as follows. Let $H$ be the distribution of $(y, \mathbf{x})$, where $y \in \mathfrak{R}$ and $\mathbf{x} \in \mathfrak{R}^p$. For $\boldsymbol{\alpha}, \boldsymbol{\lambda} \in \mathfrak{R}^p$, define

$$A(\boldsymbol{\alpha}, \boldsymbol{\lambda}) = |T(\mathscr{L}(y - \boldsymbol{\alpha}'\mathbf{x}, \boldsymbol{\lambda}'\mathbf{x}))|s(\mathscr{L}(\boldsymbol{\lambda}'\mathbf{x})) \tag{3.3}$$

and

$$C(\boldsymbol{\alpha}) = \sup_{\|\boldsymbol{\lambda}\|=1} A(\boldsymbol{\alpha}, \boldsymbol{\lambda}), \tag{3.4}$$

where $\|\ \|$ denotes Euclidean norm. The P1-estimate $\mathbf{T}$ is defined by

$$C(\mathbf{T}(H)) = \min_{\boldsymbol{\alpha}\in\mathfrak{R}^p} C(\boldsymbol{\alpha}). \tag{3.5}$$

The heuristic motivation for the P1-estimates is simple. If $(y, \mathbf{x})$ satisfies the linear model (2.1), then $y - \boldsymbol{\alpha}_0'\mathbf{x} = u$ is independent of the *projection* $\boldsymbol{\lambda}'\mathbf{x}$ for all $\boldsymbol{\lambda}$. Therefore if $T$ is Fisher-consistent as an estimate of univariate regression, then $T(\mathscr{L}(y - \boldsymbol{\alpha}_0'\mathbf{x}, \boldsymbol{\lambda}'\mathbf{x})) = 0$ for all $\boldsymbol{\lambda}$ and therefore $C(\boldsymbol{\alpha}_0) = 0$. When the distribution $H$ does not correspond to a perfect linear model, as will be the case for example with the empirical distributions, the equation $C(\boldsymbol{\alpha}) = 0$ does not have in general a solution. Therefore the estimate is defined by minimizing $C$. The normalizing factor $s(\mathscr{L}(\boldsymbol{\lambda}'x))$ is necessary to guarantee the equivariance of the estimate.

The type 2 projection estimates (P2-estimates) replace the scale estimate $s$ by a scatter matrix estimate $S$ of $\mathbf{x}$. Let $S$ be a scatter matrix estimating functional on $\mathfrak{R}^p$ which is affine equivariant. The P2-estimate is defined similarly to the P1-estimate, but replacing (3.3) by

$$A(\boldsymbol{\alpha}, \boldsymbol{\lambda}) = |T(\mathscr{L}(y - \boldsymbol{\alpha}'\mathbf{x}, \boldsymbol{\lambda}'\mathbf{x}))|(\boldsymbol{\lambda}'S(\mathscr{L}(\mathbf{x}))\boldsymbol{\lambda})^{1/2}. \tag{3.6}$$

Then the P2-estimate is defined by (3.6), (3.4) and (3.5). The P1- and P2-estimates will generically be called P-estimates.

The idea of considering all the projections of the data was first proposed independently by Stahel (1981) and Donoho (1982) in the context of multivariate analysis.

REMARK 3.1. Because of the equivariance of $T$, $s$ and $S$, we can take the supremum over $\boldsymbol{\lambda} \in \mathfrak{R}^p$, $\boldsymbol{\lambda} \neq 0$, in (3.4), without the restriction $\|\boldsymbol{\lambda}\| = 1$.

REMARK 3.2. If the univariate regression estimate is the least squares (LS) estimator, then it is immediate that the corresponding P-estimates coincide with the LS-estimate.

Theorem 3.1 states the equivariance of P-estimates.

THEOREM 3.1. *The* P1- *and* P2-*estimators are regression, affine and scale equivariant.*

The proof is straightforward and may be found in Maronna and Yohai (1989).

Another property of the P-estimates is Fisher-consistency. Suppose that the univariate regression estimate $T$ is Fisher-consistent for the error distribution $F_0$, that is, if $(y, x)$ have joint distribution $H_0(y, x) = F_0(y)G_0(x)$, where $G_0$ does not assign probability 1 to 0, then $T(H_0) = 0$. Then, Theorem 3.2 states that the P-estimates are also Fisher-consistent.

THEOREM 3.2. *Suppose that the univariate regression estimate $T$ is Fisher-consistent for the error distribution $F_0$ and that $\mathscr{L}(y, \mathbf{x}) = H_0$ given by* (2.2).

(a) *If $s(\mathscr{L}(\boldsymbol{\lambda}'\mathbf{x})) > 0$ for all $\boldsymbol{\lambda} \neq \mathbf{0}$, then the P1-estimate defined by* (3.3), (3.4) *and* (3.5) *is Fisher-consistent, that is, $T(H_0) = \boldsymbol{\alpha}_0$.*

(b) *If $S(G_0)$ is positive definite, then the P2-estimate defined by* (3.6), (3.4) *and* (3.5) *is also Fisher-consistent.*

The proof is straightforward and may be found in Maronna and Yohai (1989).

3.2. *Robustness properties of projection estimates.* Given a scale estimating functional $s$ and a distribution $G_0$ on $\Re^p$, define

$$d^+(s, G_0, \varepsilon) = \sup_{\|\boldsymbol{\lambda}\|=1} \{\sup s(\mathscr{L}(G, \boldsymbol{\lambda}'\mathbf{x})) : G \in \mathscr{V}_{G_0,\varepsilon}\},$$

$$d^-(s, G_0, \varepsilon) = \inf_{\|\boldsymbol{\lambda}\|=1} \{\inf s(\mathscr{L}(G, \boldsymbol{\lambda}'\mathbf{x})) : G \in \mathscr{V}_{G_0,\varepsilon}\}.$$

If $S$ is a scatter estimating functional on $\Re^p$ and $G_0$ a distribution function on $\Re^p$, define

$$d^+(S, G_0, \varepsilon) = \sup\{\delta_p(S(G)) : G \in \mathscr{V}_{G_0,\varepsilon}\},$$

$$d^-(S, G_0, \varepsilon) = \inf\{\delta_1(S(G)) : G \in \mathscr{V}_{G_0,\varepsilon}\},$$

where $\delta_1(S(G)) \le \cdots \le \delta_p(S(G))$ are the eigenvalues of $S(G)$. The following theorem gives an upper bound for the maximum bias $\beta(\mathbf{T}, H_0, \varepsilon)$ of P-estimates.

THEOREM 3.3. *Let $(y, \mathbf{x})$ have distribution $H_0$ given by* (2.2). *Assume that the univariate regression estimate $T$ is Fisher-consistent for $H_0$. Let $\mathbf{T}$ be a* P1-(P2-)*estimate, based on $T$ with scale $s$ (scatter matrix $S$). Then*

$$\beta(\mathbf{T}, H_0, \varepsilon) \le \sup\{\beta(T, L_{\boldsymbol{\lambda}}, \varepsilon) : \|\boldsymbol{\lambda}\| = 1\}\left(1 + \frac{d^+(G_0, \varepsilon)}{d^-(G_0, \varepsilon)}\right),$$

*where $d^{\pm}(G_0, \varepsilon)$ means $d^{\pm}(s, G_0, \varepsilon)$ for* P1*-estimates and $d^{\pm}(S, G_0, \varepsilon)$ for* P2*-estimates, and $L_{\lambda} = \mathscr{L}(H_0, y, \boldsymbol{\lambda}'\mathbf{x})$ when $\boldsymbol{\alpha}_0 = 0$.*

PROOF. We will prove the theorem only for P1-estimates, the proof for P2-estimates being completely similar.

Because of the regression equivariance of the univariate regression estimate $T$, and of the P1-estimate $\mathbf{T}$, we can suppose that $\boldsymbol{\alpha}_0 = \mathbf{0}$.

Let $H \in \mathscr{V}_{H_0,\varepsilon}$, and put $\boldsymbol{\alpha} = \mathbf{T}(H)$. We have to prove that

$$\|\boldsymbol{\alpha}\| \le \sup\{\beta(T, L_{\boldsymbol{\lambda}}, \varepsilon) : \|\boldsymbol{\lambda}\| = 1\}\left(1 + \frac{d^+(G_0, \varepsilon)}{d^-(G_0, \varepsilon)}\right). \tag{3.7}$$

For any $\boldsymbol{\lambda} \in \Re^p$ with $\|\boldsymbol{\lambda}\| = 1$, we have $|T(\mathscr{L}(y, \boldsymbol{\lambda}'\mathbf{x}))| \le \beta(T, L_{\boldsymbol{\lambda}}, \varepsilon)$, and hence

$$C(\mathbf{0}) \le \sup\{\beta(T, L_{\boldsymbol{\lambda}}, \varepsilon) : \|\boldsymbol{\lambda}\| = 1\} d^+(G_0, \varepsilon). \tag{3.8}$$

Take any $\boldsymbol{\alpha} \in \Re^p$, $\boldsymbol{\alpha} \ne 0$ and put $\hat{\boldsymbol{\lambda}} = \boldsymbol{\alpha}/\|\boldsymbol{\alpha}\|$. Then, the regression equivariance

of $T$ implies

$$T\big(\mathscr{L}(y - \boldsymbol{\alpha}'\mathbf{x}, \hat{\boldsymbol{\lambda}}'\mathbf{x})\big) = T\big(\mathscr{L}(y, \hat{\boldsymbol{\lambda}}'\mathbf{x})\big) - \|\boldsymbol{\alpha}\|. \tag{3.9}$$

If

$$\|\boldsymbol{\alpha}\| \le \sup\{\beta(T, L_{\boldsymbol{\lambda}}, \varepsilon): \|\boldsymbol{\lambda}\| = 1\}, \tag{3.10}$$

then (3.7) holds. Suppose that (3.10) does not hold, then from (3.9) we get

$$\sup\{|T(\mathscr{L}(y - \boldsymbol{\alpha}'\mathbf{x}, \boldsymbol{\lambda}'\mathbf{x}))|: \|\boldsymbol{\lambda}\| = 1\} \ge \|\boldsymbol{\alpha}\| - \sup\{\beta(T, L_{\boldsymbol{\lambda}}, \varepsilon): \|\boldsymbol{\lambda}\| = 1\},$$

and then $C(\boldsymbol{\alpha}) \ge (\|\boldsymbol{\alpha}\| - \sup\{\beta(\mathrm{T}, L_{\boldsymbol{\lambda}}, \varepsilon)$: $\|\boldsymbol{\lambda}\| = 1\})d^{-}(G_0, \varepsilon)$. Using (3.8) and (3.5) we get (3.7). □

The following theorem shows that, roughly speaking, if the univariate estimate and the scale (or scatter) estimates have high breakdown point, so has the P-estimate.

THEOREM 3.4. *Let $H$ be the distribution of $(y, \mathbf{x})$ and $G$ the distribution of $\mathbf{x}$ and let $\mathbf{T}$ be a* P1-, (P2-)*estimate corresponding to an univariate regression estimate $T$. Let $\varepsilon > 0$ be such that* $\sup\{\beta(T, L_{\boldsymbol{\lambda}}, \varepsilon)$: $\|\boldsymbol{\lambda}\| = 1\} < \infty$, *and assume that the scale $s$ (scatter matrix $S$) satisfies $d^{+}(G, \varepsilon) < \infty$ and $d^{-}(G, \varepsilon) > 0$. Then $\varepsilon^{*}(T, H) \ge \varepsilon$.*

The proof may be found in Maronna and Yohai (1989). A finite sample breakdown point result is proved in Section 5.

3.3. *The median of slopes as univariate regression estimate.* We will study in particular the P-estimates obtained when the univariate regression estimate is given by

$$T_0(H) = \operatorname{median}_H\left[\left(\frac{y}{x}\right)\middle| x \ne 0\right]. \tag{3.11}$$

Martin, Yohai and Zamar (1989) proved that if $p = 1$ and $F_0$ has a symmetric and unimodal density, then this estimate is minimax in the class of all regression equivariant estimates. They also proved that in this case

$$\beta(T_0, H_0, \varepsilon) = \left|J^{-1}\left(\frac{1}{2(-\varepsilon)}\right)\right|, \tag{3.12}$$

where $J$ is the distribution function of $y/x$ under $\alpha_0 = 0$.

The P1- or P2-estimates based on the univariate regression estimate $T_0$ given in (3.11) will be called *median projection estimates* (MP-estimates). For $\boldsymbol{\alpha}$ and $\boldsymbol{\lambda} \in \mathfrak{R}^p$, and a distribution function $H$ on $\mathfrak{R}^{p+1}$ define

$$J_{H,\boldsymbol{\alpha},\boldsymbol{\lambda}}(t) = P_H\left(\frac{y - \boldsymbol{\alpha}'\mathbf{x}}{\boldsymbol{\lambda}'\mathbf{x}} \le t \middle| \boldsymbol{\lambda}'\mathbf{x} \ne 0\right), \tag{3.13}$$

and

$$J_{\lambda}(t) = P_{H_0, \alpha_0, \lambda}(t), \tag{3.14}$$

where $H_0$ is given by (2.2).

From Theorem 3.3 and (3.12) we get the following upper bound for the maximum bias of an MP-estimate $\mathbf{T}_0$ when $\mathscr{L}(y, \mathbf{x}) = H_0$ given by (2.2),

$$\beta(\mathbf{T}_0, H_0, \varepsilon) \leq \sup_{\|\lambda\|=1} \left| J_{\lambda}^{-1}\left(\frac{1}{2(1-\varepsilon)}\right)\right| \left(1 + \frac{d^+(G_0, \varepsilon)}{d^-(G_0, \varepsilon)}\right). \tag{3.15}$$

The following theorem gives a lower bound for $\beta(\mathbf{T}, H_0, \varepsilon)$ for any estimate $\mathbf{T}$ which is (i) regression and affine equivariant, and (ii) Fisher-consistent.

THEOREM 3.5. *Assume that* $(y, \mathbf{x})$ *has distribution* $H_0$ *given by* (2.2), *where* $F_0$ *has a density* $f_0(u)$ *which is symmetric and decreasing in* $|u|$. *Let* $\mathbf{T}$ *be any estimate satisfying* (i) *and* (ii) *as above, then*

$$\beta(\mathbf{T}, H_0, \varepsilon) \geq \sup_{\|\lambda\|=1} J_{\lambda}^{-1}\left(\frac{1}{2(1-\varepsilon)}\right). \tag{3.16}$$

PROOF. Since $\mathbf{T}$ is regression and affine equivariant we can assume that $\boldsymbol{\alpha}_0 = \mathbf{0}$. Since $V(G_0) = I$ is invariant for orthogonal transformations it is enough to show that

$$\beta(\mathbf{T}, H_0, \varepsilon) \geq \left| J_{e_1}^{-1}\left(\frac{1}{2(1-\varepsilon)}\right)\right|, \tag{3.17}$$

where $\mathbf{e}_1 = (1, 0, \ldots, 0)'$.

The proof of (3.17) follows very closely the one given by Huber (1981) to prove that the median is the minimax-bias estimate for location, and therefore we omit details. Let $q = J_{\mathbf{e}_1}^{-1}(1/(2(1-\varepsilon)))$. In order to make the proof easier we will assume that $G_0$ has a density $g_0$, but this assumption is not necessary. Then $H_0$ has density $h_0 = f_0(y)g_0(\mathbf{x})$. Consider the following density of $(y, \mathbf{x})$

$$h^+(y, \mathbf{x}) = \begin{cases} (1-\varepsilon) f_0(y) g_0(\mathbf{x}), & \text{if } (y/x_1) \leq q, \\ (1-\varepsilon) f_0(y - 2qx_1) g_0(\mathbf{x}), & \text{if } (y/x_1) > q. \end{cases}$$

It is easy to prove that $h^+$ is a density function, and that the corresponding distribution $H^+ \in \mathscr{V}_{H_0, \varepsilon}$. Let $H^-$ be the distribution corresponding to $(y - 2qx_1, \mathbf{x})$, when $(y, \mathbf{x})$ has distribution $H^+$. Then, it may be proved that also $H^- \in \mathscr{V}_{H_0, \varepsilon}$. Let $T_1$ be the first coordinate of $\mathbf{T}$; then the regression equivariance of $\mathbf{T}$ yields $T_1(H^+) - T_1(H^-) = 2q$ and thus $\beta(\mathbf{T}, H_0, \varepsilon) \geq q$. □

Suppose that the scale $s$ or the scatter matrix $S$ used to define the P-estimates is weakly continuous as it happens with robust estimates. Then

$$\lim_{\varepsilon \to 0} \frac{d^+(G_0, \varepsilon)}{d^-(G_0, \varepsilon)} = 1, \tag{3.18}$$

and hence for small $\varepsilon$ the upper bound given in (3.15) will be approximately twice the lower bound obtained in Theorem 3.5. A similar result was proved by Donoho and Liu (1988) for minimum distance estimates. In fact they prove that for some parametric and semiparametric models, the maximum bias of these estimates is not larger than twice the minimax value.

Suppose that (3.18) holds, and that $\mathbf{T}_0$ is an MP-estimate. Then, differentiating (3.15) we get an upper bound for the value of $\gamma^*$:

$$\gamma^*(\mathbf{T}_0, H_0) \leq 2\frac{\partial}{\partial\varepsilon} \sup_{\|\boldsymbol{\lambda}\|=1} \left| J_{\boldsymbol{\lambda}}^{-1}\left(\frac{1}{2(1-\varepsilon)}\right)\right|_{\varepsilon=0}. \tag{3.19}$$

In the case that $(y, \mathbf{x})$ is $N(\mathbf{0}, I)$, we have

$$J_{\boldsymbol{\lambda}}(t) = \frac{1}{\pi}\arctan(t) + \frac{1}{2}, \tag{3.20}$$

which implies,

$$J_{\boldsymbol{\lambda}}^{-1}\left(\frac{1}{2(1-\varepsilon)}\right) = \tan\left(\frac{\pi\varepsilon}{2(1-\varepsilon)}\right), \tag{3.21}$$

and therefore (3.19) and (3.21) yield

$$\gamma^*(\mathbf{T}_0, H_0) \leq \pi. \tag{3.22}$$

It turns out, as we will see in Theorem 4.1, that in the case that $G_0$ is elliptical, this bound is equal to $\gamma^*$.

In Table 1, see subsection 7.1, we give the values of $\gamma^*$ for the MP-estimates (it is the same for the two versions P1 and P2) and for the minimax-bias GM-estimate MGM when $(y, \mathbf{x})$ is $N(\mathbf{0}, I)$. We observe that the MP-estimates have smaller $\gamma^*$ than the MGM-estimate for $p \geq 6$. In Section 4 we will derive an estimate which has smaller $\gamma^*$ than $T_{GM}$ for all $p$.

**4. A modified P-estimate.** In this section we define new estimates by applying a correction to the MP-estimates defined in subsection 3.3. Let $\mathbf{T}_0$ be an MP-estimate and define $\hat{\boldsymbol{\lambda}} = \hat{\boldsymbol{\lambda}}(H) \in \mathfrak{R}^p$ as a vector such that $\|\hat{\boldsymbol{\lambda}}\| = 1$ and

$$C(\mathbf{T}_0(H)) = A(\mathbf{T}_0(H), \hat{\boldsymbol{\lambda}}). \tag{4.1}$$

Then $\hat{\boldsymbol{\lambda}}$ may be considered as the direction "most correlated" with the residuals $y - \mathbf{T}_0(H)'\mathbf{x}$.

The new estimate $\mathbf{T}_1$ modifies $\mathbf{T}_0$ precisely in the direction $\hat{\boldsymbol{\lambda}}$ by eliminating its correlation with the residuals. It is defined by

$$\mathbf{T}_1(H) = \mathbf{T}_0(H) + T_0\big(\mathscr{L}(y - \mathbf{T}_0(H), \hat{\boldsymbol{\lambda}}'\mathbf{x})\big)\hat{\boldsymbol{\lambda}}. \tag{4.2}$$

Then $T_0(\mathscr{L}(y - \mathbf{T}_1(H), \hat{\boldsymbol{\lambda}}'\mathbf{x})) = 0$, and we may consider that the "correlation" between the residuals $y - \mathbf{T}_1(H)$ and $\hat{\boldsymbol{\lambda}}$ is 0. We shall call $\mathbf{T}_1$ *corrected median projection estimate* (CMP-estimate).

It seems difficult to compute the exact value of the maximum bias $\beta$ for MP- and CMP-estimates. To gain insight into the comparative behavior of

these two estimates we start by computing $\beta$ under the following simplifications.

(i) Only point mass contaminating distributions are considered. Hence the maximum bias is defined by

$$\beta^*(\mathbf{T}, H_0, \varepsilon) = \sup_{y, \mathbf{x}} \{b(\mathbf{T}, H): H = (1 - \varepsilon)H_0 + \varepsilon\delta_{y,\mathbf{x}}, y \in \Re, \mathbf{x} \in \Re^p\}.$$

(ii) The distribution $G_0$ of $\mathbf{x}$ is spherical. In this case the scale normalizing factor in (3.3) and (3.6) is constant and can be omitted.

The MP-estimate $\mathbf{T}_0$ and the CMP-estimate $\mathbf{T}_1$ defined by omitting the scale factor in $A(\boldsymbol{\alpha}, \boldsymbol{\lambda})$ will be henceforth referred to as the "nonaffine equivariant versions."

When $G_0$ is spherical, $J_{\boldsymbol{\lambda}}$ given by (3.14) is the same for all $\|\boldsymbol{\lambda}\| = 1$. This common distribution will be denoted by $J^*$. Applying Theorem 3.5 to this case yields for any regression and affine equivariant estimate $\mathbf{T}$

$$\beta(\mathbf{T}, H, \varepsilon) \geq K_{0,\varepsilon}, \tag{4.3}$$

where

$$K_{0,\varepsilon} = J^{*-1}\left(\frac{1}{2(1 - \varepsilon)}\right). \tag{4.4}$$

On the other hand, for the nonaffine equivariant version of the MP-estimate $\mathbf{T}_0$, the inequality (3.15) becomes

$$\beta(\mathbf{T}_0, H_0, \varepsilon) \leq 2K_{0,\varepsilon}. \tag{4.5}$$

Theorem 4.1 shows that (4.5) is an equality, and that the lower bound (4.3) is attained by the nonaffine equivariant version of the CMP-estimate $\mathbf{T}_1$.

THEOREM 4.1. *Let $H_0$ be given by* (2.2), *where $F_0$ has a density $f_0(y)$ which is symmetric and decreasing in $y$ for $y \geq 0$, and $G_0$ spherical. Let $\mathbf{T}_0$ and $\mathbf{T}_1$ the nonaffine equivariant versions of the MP- and the CMP-estimates, respectively. Then*:

(a) $\beta^*(\mathbf{T}_0, H_0, \varepsilon) = 2K_{0,\varepsilon}$.
(b) $\beta^*(\mathbf{T}_1, H_0, \varepsilon) = K_{0,\varepsilon}$.

This theorem is proved in the Appendix.

Despite the fact that Theorem 4.1 refers to the nonaffine equivariant versions of MP- and CMP-estimates, it suggests what the behavior of the actual (affine equivariant) estimates may be. In Section 7.1 we present numerical computations of $\beta^*(\mathbf{T}_i, H_0, \varepsilon)$, $i = 0, 1$, for the actual estimates when $H_0$ is the multivariate normal distribution.

## 5. Computing algorithm.

5.1. *Approximate P-estimates.* Since the numerical calculation of P-estimators for finite samples seems extremely complicated because of the two extrema involved, we have tried an approximate method which is based on a subsampling scheme similar to the one used by Rousseeuw (1984) to compute the LMS-estimate, and which would yield an estimator which behaves similarly to the P-estimate.

Let $\{(\mathbf{x}_i, y_i),\ i = 1, \ldots, n\}$ be a sample of size $n$, $D_n \subseteq \Re^p$ be a set (depending on the sample) of possible values of the estimator, and for each $\boldsymbol{\alpha} \in D_n$ let $D_n^*(\boldsymbol{\alpha}) \subseteq \{\boldsymbol{\lambda} \in \Re^p\colon \|\boldsymbol{\lambda}\| = 1\}$ be a set of possible projection directions. Define for P1-estimators $s_n(\boldsymbol{\lambda}) = s(\mathscr{L}(G_n, \boldsymbol{\lambda}'\mathbf{x}))$ and for P2-estimates $s_n(\boldsymbol{\lambda}) = (\boldsymbol{\lambda}' S(G_n)\boldsymbol{\lambda})^{1/2}$, where $G_n$ is the empirical distribution of the $\mathbf{x}_i$'s. Let

$$t_n(\boldsymbol{\alpha}, \boldsymbol{\lambda}) = T\big(\mathscr{L}(H_n, y - \boldsymbol{\alpha}'\mathbf{x}, \boldsymbol{\lambda}'\mathbf{x})\big),$$

$A_n(\boldsymbol{\alpha}, \boldsymbol{\lambda}) = |t_n(\boldsymbol{\alpha}, \boldsymbol{\lambda})| s_n(\boldsymbol{\lambda})$, $C_n(\boldsymbol{\alpha}) = \max_{\boldsymbol{\lambda} \in D_n^*(\boldsymbol{\alpha})} A_n(\boldsymbol{\alpha}, \boldsymbol{\lambda})$.

Finally define the approximate P-estimate by

$$\mathbf{T}_n = \arg\min_{\boldsymbol{\alpha} \in D_n} C_n(\boldsymbol{\alpha}). \tag{5.1}$$

The approximate MP-estimates $\mathbf{T}_{n0}$ are defined by (5.1) using for univariate regression the estimate $T_0$ given by (3.11).

Similarly the approximate CMP-estimates are defined by finding a direction $\hat{\boldsymbol{\lambda}}_n \in D_n^*$ maximizing $A_n(\boldsymbol{\lambda}, \mathbf{T}_{n0})$ and defining

$$\mathbf{T}_{n1} = \mathbf{T}_{n0} + t_n\big(\hat{\boldsymbol{\lambda}}_n, \mathbf{T}_{n0}\big)\hat{\boldsymbol{\lambda}}_n. \tag{5.2}$$

It is clear that when $D_n = \Re^p$ and $D_n^* = S_p$ the estimates given by (5.1) are the P-estimates defined in Section 3.

We propose to take for $\boldsymbol{\alpha} \in D_n$

$$D_n^*(\boldsymbol{\alpha}) = \big\{\boldsymbol{\lambda} = (\boldsymbol{\alpha}^* - \boldsymbol{\alpha})/\|\boldsymbol{\alpha}^* - \boldsymbol{\alpha}\|\colon \boldsymbol{\alpha}^* \in D_n,\ \boldsymbol{\alpha}^* \neq \boldsymbol{\alpha}\big\}. \tag{5.3}$$

The rationale for this choice is that, if model (2.1) holds and $\boldsymbol{\alpha}_0$ is the true parameter, then for each $\boldsymbol{\alpha}$, the direction $\boldsymbol{\lambda}$ such that $y - \boldsymbol{\alpha}'\mathbf{x}$ has the highest correlation with $\boldsymbol{\alpha}'\mathbf{x}$ is $\boldsymbol{\lambda} = (\boldsymbol{\alpha}_0 - \boldsymbol{\alpha})/\|\boldsymbol{\alpha}_0 - \boldsymbol{\alpha}\|$. It is easy to show that if the set $D_n$ is regression, affine and scale equivariant and $D_n^*(\boldsymbol{\alpha})$ is given by (5.3), the estimates defined by (5.1) are regression, affine and scale equivariant too.

For our computer approximations, $D_n$ will be the finite equivariant set obtained as follows. Generate $N$ subsamples of size $p$ from the sample. For the $k$th subsample ($k = 1, \ldots, N$) fit a hyperplane ($y = \boldsymbol{\alpha}_k'\mathbf{x}$) containing the $p$ points. Then $D_n = \{\boldsymbol{\alpha}_k\colon k = 1, \ldots, N\}$. The sets $D_n^*(\boldsymbol{\alpha})$ are defined by (5.3).

Actually, not all $N(N-1)$ values of $A_n(\boldsymbol{\alpha}_i, \boldsymbol{\lambda}_j)$ need to be computed. For a given $\boldsymbol{\alpha}_k$, let $C_j = \min_{k \le j} C_n(\boldsymbol{\alpha}_k)$, which is a decreasing function of $j$. Then if for some $h$,

$$A_n(\boldsymbol{\alpha}_{j+1}, \boldsymbol{\lambda}_h) > C_j, \tag{5.4}$$

we may drop $\boldsymbol{\alpha}_{j+1}$ as a candidate for the minimum, for in this case

$$C_n(\boldsymbol{\alpha}_{j+1}) > C_j, \tag{5.5}$$

and therefore we need not compute $A_n(\boldsymbol{\alpha}_{j+1}, \boldsymbol{\lambda}_i)$ for $i > h$.

It is easy to show that the expected number of times that (5.5) does not occur is asymptotically $\cong \ln N$. By conditioning on (5.4), it follows that the expected waiting time for the first $h$ such that (5.4) holds is also $\cong \ln N$. Then the expected number of times that $A(\boldsymbol{\alpha}_i, \boldsymbol{\lambda}_k)$ must be computed is $\cong N \ln N + (N - \ln N)\ln N \cong 2N \ln N$.

Summing up, the algorithm is as follows:

For $k = 1, \ldots, N$ do
    Generate a random subsample of size $p$ from the sample, and compute the coefficient vector $\boldsymbol{\alpha}_k$ of the hyperplane $y = \boldsymbol{\alpha}_k' \mathbf{x}$ containing the $p$ points
End do
Let $C = \infty$ [this is the current minimum of $C_n(\boldsymbol{\alpha})$]
For $k = 1, \ldots, N$ do:
    let $A = 0$ [this is the current minimum of $A_n$]
    For $j = 1, \ldots, N$, $j \neq k$ do while $A < C$:
        Let $A = \max\{A, A_n(\boldsymbol{\alpha}_k, (\boldsymbol{\alpha}_k - \boldsymbol{\alpha}_j)/\|\boldsymbol{\alpha}_k - \boldsymbol{\alpha}_j\|)\}$
    End do
    If $A < C$, then $C = A$ and $T = \boldsymbol{\alpha}_k$
End do

The breakdown behavior of the approximate MP-estimates is studied in subsection 5.2. As it will be proved there, $N$ may be chosen so as to guarantee in some sense a high breakdown point with high probability. We think however that $N$ should be chosen so that the approximate estimator not only has a high breakdown point, but also keeps the favorable features of the "exact" P-estimator, that is, we want an estimator whose bias under contamination in not only bounded, but as small as possible. Theorem 6.2 shows that this can be attained at least for large $n$ and $N$. In our simulation studies we have chosen $N = 200$, and as it will be seen in subsection 7.2, the behavior of the resulting estimator compares favorably with the other estimates considered there.

The computing time required for the algorithm for $n = 50$, $p = 5$ and $N = 200$ in a PC with the INTEL 80386 (33 MHz) processor and mathematical co-processor is around 30 seconds. For $N = 500$, the computing time is around 90 seconds.

5.2. *Breakdown point.* The finite sample breakdown point was defined by Donoho and Huber (1983). Let the sample be $\mathbf{Z}_n = \{\mathbf{z}_1, \ldots, \mathbf{z}_n\} \subseteq \Re^{p+1}$, with $\mathbf{z}_i = (\mathbf{x}_i, y_i)$. Let $\mathbf{Z}_{n,m}$ be any contaminated sample of size $n$ obtained by replacing $m$ observations of the original sample $\mathbf{Z}_n$ by arbitrary outliers. Call $\mathcal{Z}$ the set all these $Z_{n,m}$'s.

The regression estimator $\mathbf{T}_n$ defined for samples of size $n$ is said to break down at $\mathbf{Z}_n$ for a given $m$, if $\sup \|\mathbf{T}_n(\mathbf{Z}_{n,m})\| = \infty$, where the supremum is over all $\mathbf{Z}_{n,m} \in \mathscr{Z}$. Let $m_0$ be the minimum $m$ such that $\mathbf{T}_n$ breaks down. Then the finite sample breakdown point of $\mathbf{T}_n$ at $\mathbf{Z}_n$ is $\varepsilon^*(\mathbf{T}, \mathbf{Z}_n) = m_0/n$.

Since the computational version of our estimates contains a randomization, the resulting value for a fixed sample will be random, and hence we have to modify our treatment accordingly.

To simplify the exposition we consider the P1 version of the approximate MP-estimate defined by (5.1) with initial estimator $T_0$ given by (3.11) and scale $s(r_1, \ldots, r_n) = \text{median}\{|r_i|\}$. The set $D_n$ is defined in subsection 5.1 by taking $N$ random subsamples of size $p$ from the contaminated sample $\mathbf{Z}_{n,m}$, and the sets $D_n^*(\boldsymbol{\alpha})$ are defined by (5.3). Since now the set $D_n$ is random, we will write $\mathbf{T}_{n0}(\mathbf{z}, D_n)$ to indicate the dependence of the estimate on this set.

Call $\pi(n, m, p, N)$ the probability that at least one subsample of size $p$ is contained in $\mathbf{Z}_n$. Then $\pi(n, m, p, N) = 1 - (1 - \beta)^N$, where

$$\beta = \binom{n-m}{p} \Big/ \binom{n}{p} \cong (m/n)^p.$$

The following theorem can be interpreted as stating that the approximate MP-estimate $\mathbf{T}_{n0}$, where the set $D_n$ is defined by using $N$ random subsamples of size $p$, does not break down for $m < [n/2] - (p - 1)$ with probability at least $\pi(n, m, p, N)$. This probability can be taken as close to 1 as desired by choosing $N$ large enough.

THEOREM 5.1. *Let $\mathbf{Z}_n$ be such that every subset $\{\mathbf{x}_{i_j},\ j = 1, \ldots, p\}$ is linearly independent. Let $m < [n/2] - (p - 1)$, where $[x]$ denotes integer part of $x$. Then there exists $K$ (depending only on $\mathbf{Z}_n$) such that for all $\mathbf{Z}_{n,m} \in \mathscr{Z}$, $\|\mathbf{T}_{n0}(\mathbf{Z}_{n,m}, D_n)\| \le K$ with probability larger than $\pi(n, m, p, N)$.*

PROOF. Let $\mathbf{Z}_n = \{\mathbf{z}_1, \ldots, \mathbf{z}_n\}$ and $\mathbf{Z}_{n,m} = \{\mathbf{z}_1^*, \ldots, \mathbf{z}_n^*\} \in \mathscr{Z}$. Put $\mathbf{z}_i = (y_i, \mathbf{x}_i)$ and $\mathbf{z}_i^* = (y_i^*, \mathbf{x}_i^*)$. Let $n_1 = n - m$; we may assume that the elements of $\mathbf{Z}_{n,m}$ are numbered so that $\mathbf{z}_i^* \in \mathbf{Z}_n$ for $1 \le i \le n_1$. Note that by hypothesis $n_1 \ge n/2 + p - 1$. To each subset of size $p$ from $\mathbf{Z}_n$, there corresponds a vector $\boldsymbol{\alpha}$ obtained by fitting a hyperplane. Call $B$ the set of all such $\boldsymbol{\alpha}$'s. To each subset of $N$ subsamples from $Z_{n,m}$, there corresponds a set $D_n = \{\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_N\}$ obtained by fitting each subsample. Call $\mathscr{D}$ the set of all $D_n$'s which have a nonvoid intersection with $B$.

In order to prove the theorem it is enough to show that

$$\sup_{\mathbf{Z}_{n,m} \in \mathscr{Z}} \sup_{D_n \in \mathscr{D}} \|\mathbf{T}_{n0}(\mathbf{Z}_{n,m}, D_n)\| = K < \infty. \tag{5.6}$$

For a given $D_n \in \mathscr{D}$ and $\boldsymbol{\lambda} \in \mathfrak{R}^p$, let $I_{\boldsymbol{\lambda}} = \{i \le n_1 \colon \boldsymbol{\lambda}'\mathbf{x}_i \ne 0\}$, $n_2(\boldsymbol{\lambda}) = \#I_{\boldsymbol{\lambda}}$ and $p_1(\boldsymbol{\lambda}) = n_1 - n_2(\boldsymbol{\lambda})$. Since every set of $p$ $\mathbf{x}_i$'s is linearly independent, we have $p_1(\boldsymbol{\lambda}) < p$. Call, respectively, $q_0(\boldsymbol{\alpha}, \boldsymbol{\lambda})$ and $q_1(\boldsymbol{\alpha}, \boldsymbol{\lambda})$ the $(p - p_1(\boldsymbol{\lambda}))$ and the $(n_1 - p + 1)$ smallest values of $(y_i - \boldsymbol{\alpha}'\mathbf{x}_i)/(\boldsymbol{\lambda}'\mathbf{x}_i)$, $i \in I_{\boldsymbol{\lambda}}$. By hypothesis $p \le n/2 \le n_1 - p + 1$, and this implies that $t_n(\boldsymbol{\alpha}, \boldsymbol{\lambda}) \in [q_0(\boldsymbol{\lambda}, \boldsymbol{\alpha}), q_1(\boldsymbol{\lambda}, \boldsymbol{\alpha})]$.

Since $q_0$ and $q_1$ are continuous functions of $\boldsymbol{\lambda}$ we have that for any $\boldsymbol{\alpha}$, $\sup_{\|\boldsymbol{\lambda}\|=1}|q_i(\boldsymbol{\alpha}, \boldsymbol{\lambda})| < \infty$, $i = 1, 2$, and therefore since the set $B$ is finite

$$\sup_{Z_{n,m}\in\mathscr{Z},\, \boldsymbol{\alpha}\in G,\, \|\boldsymbol{\lambda}\|=1} \left|t_n(\boldsymbol{\alpha}, \boldsymbol{\lambda})\right| = t_0 < \infty. \tag{5.7}$$

At the same time, it follows from the fact that $n_1 > n/2$ that $s_n(\boldsymbol{\lambda}) = \text{median}\{|\boldsymbol{\lambda}'\mathbf{x}_1^*|, \ldots, |\boldsymbol{\lambda}'\mathbf{x}_n^*|\} \le \max_{1\le i\le n}\|\mathbf{x}_i\| = s_0 < \infty$. This and (5.7) imply that

$$\min_{\boldsymbol{\alpha}\in D_n} C_n(\boldsymbol{\alpha}) \le C_0, \quad \forall\, \mathbf{Z}_{n,m} \in \mathscr{Z}, \forall\, D_n \in \mathscr{D}, \tag{5.8}$$

where $C_0 = t_0 s_0$.

We shall now prove that there exists $K$ such that

$$\inf_{\boldsymbol{\alpha}\in D_n,\, \|\boldsymbol{\alpha}\|>K} C_n(\boldsymbol{\alpha}) > C_0, \quad \forall\, \mathbf{Z}_{n,m} \in \mathscr{Z}, \forall\, D_n \in \mathscr{D}, \tag{5.9}$$

where $C_0$ is defined below (5.8).

Given $\boldsymbol{\lambda} \in \Re^p$ with $\|\boldsymbol{\lambda}\| = 1$, call $|\boldsymbol{\lambda}'\mathbf{x}|_{(p)}$ the $p$th smallest value of $|\boldsymbol{\lambda}'\mathbf{x}_i|$, $1 \le i \le n$. Since this is a continuous function of $\boldsymbol{\lambda}$ which is positive for $\boldsymbol{\lambda} \ne 0$, we have $\inf_{\|\boldsymbol{\lambda}\|=1}|\boldsymbol{\lambda}'\mathbf{x}|_{(p)} = \delta > 0$, and hence $s_n(\boldsymbol{\lambda}) \ge \delta$ for all $\|\boldsymbol{\lambda}\| = 1$ and $\mathbf{Z}_{n,m} \in \mathscr{Z}$.

Given any $D_n \in \mathscr{D}$, let $\boldsymbol{\alpha}$ and $\boldsymbol{\alpha}^*$ be two different elements of $D_n$, with $\boldsymbol{\alpha}^* \in B$; then $\boldsymbol{\lambda} = (\boldsymbol{\alpha}^* - \boldsymbol{\alpha})/\|\boldsymbol{\alpha}^* - \boldsymbol{\alpha}\| \in D_n^*(\boldsymbol{\alpha})$. Put $r = \|\boldsymbol{\alpha} - \boldsymbol{\alpha}^*\|$. Then $\boldsymbol{\alpha} = \boldsymbol{\alpha}^* - r\boldsymbol{\lambda}$, and hence $t_n(\boldsymbol{\alpha}, \boldsymbol{\lambda}) = \text{median}\{(y_i^* - \boldsymbol{\alpha}'\mathbf{x}_i^*)/\boldsymbol{\lambda}'\mathbf{x}_i^*:\ \boldsymbol{\lambda}'\mathbf{x}_i^* \ne 0\} = t_n(\boldsymbol{\alpha}^*, \boldsymbol{\lambda})) + r \ge \|\boldsymbol{\alpha}\| - v$, where $v = \sup_{\boldsymbol{\alpha}^*\in G,\ \|\boldsymbol{\lambda}\|=1}(|q_0(\boldsymbol{\lambda}, \boldsymbol{\alpha}^*)| + \|\boldsymbol{\alpha}^*\|)$. Then (5.9) holds if $K = v + C_0/\delta$. Finally (5.8), (5.9) and (5.1) imply (5.6). □

## 6. Asymptotic theory.

6.1. *Rate of convergence.* To show that $P$ estimators have a rate of convergence of $n^{1/2}$, three previous technical results are needed.

Given a closed set $A$ in a Euclidean space, define $L^\infty(A)$ as the set of all bounded and measurable real functions on $A$ metrized with the supremum norm. Let $\mathscr{S}_q = \{\boldsymbol{\lambda} \in \Re^q: \|\boldsymbol{\lambda}\| = 1\}$. Let $H$ be a distribution on $\Re^q$, let $\mathbf{z}_1, \ldots, \mathbf{z}_n$ be i.i.d. random vectors in $\Re^q$ with distribution $H$, and call $H_n$ the respective empirical distribution. Define for $\boldsymbol{\gamma} \in \mathscr{S}_q$ the process $U_n(\boldsymbol{\gamma}) = n^{1/2}(P_{H_n}(\boldsymbol{\gamma}'\mathbf{z} \le 0) - P_H(\boldsymbol{\gamma}'\mathbf{z} \le 0))$. Then it is proved in Proposition 1 of Section 4 of Beran and Millar (1986), that $U_n$ converges weakly as a random element of $L^\infty(\mathscr{S}_q)$ to a Gaussian process. Define now, for $\boldsymbol{\gamma}_1$ and $\boldsymbol{\gamma}_2$ in $\Re^q$, with $(\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2)$ in $\mathscr{S}_{2q}$, the process

$$V_n(\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2) = n^{1/2}\big(P_{H_n}(\boldsymbol{\gamma}_1'\mathbf{z} \le 0, \boldsymbol{\gamma}_2'\mathbf{z} \le 0) - P_H(\boldsymbol{\gamma}_1'\mathbf{z} \le 0, \boldsymbol{\gamma}_2'\mathbf{z} \le 0)\big).$$

Then, using exactly the same method of proof of Beran and Millar, the following lemma may be proved.

LEMMA 6.1. *$V_n$ converges weakly as random elements of $L^\infty(\mathscr{S}_{2q})$ to a Gaussian process.*

Let now $H_0$ be the distribution on $\Re^{p+1}$ given by (2.2). To simplify the proofs, it will be assumed in this section that $P_{G_0}(\lambda'\mathbf{x} = 0) = 0$ for all $\lambda \in \mathscr{S}_p$. Let $(y_1, \mathbf{x}_1), \ldots, (y_n, \mathbf{x}_n)$ be i.i.d. random vectors in $\Re^{p+1}$ with distribution $H_0$, and let $H_n$ be the corresponding empirical distribution. Define the processes

$$W_n(\alpha, \lambda, t) = \sqrt{n}\left(J_{H_n, \alpha, \lambda}(t) - J_{H_0, \alpha, \lambda}(t)\right),$$

where $J_{H, \alpha, \lambda}$ is defined in (3.13). Then we have the following lemma.

LEMMA 6.2. *The sequence of empirical processes $W_n$ converges weakly as random elements of $L^\infty(\Re^p \times \mathscr{S}_p \times \Re)$ to a Gaussian process.*

PROOF. Note that

$$P((y - \alpha'\mathbf{x})/\lambda'\mathbf{x} \le t) = P(y - \alpha'\mathbf{x} - t\lambda'\mathbf{x} \le 0, \lambda'\mathbf{x} \ge 0) + P(y - \alpha'\mathbf{x} - t\lambda'\mathbf{x} \ge 0, \lambda'\mathbf{x} \le 0).$$

Put $q = p + 1$, $H = H_0$ and $z = (y, \mathbf{x})$. Then the first term above is dealt with by applying Lemma 6.1 with $\gamma_1 = (1, -\alpha - t\lambda)/a$ and $\gamma_2 = -\alpha/a$, where $a = (1 + \|\alpha + t\lambda\|^2 + \|\alpha\|^2)^{1/2}$. The second term is dealt with likewise. □

LEMMA 6.3. *Suppose $H_0 = F_0 G_0$ as in* (2.2). *Assume also that*:

(i) *$F_0$ has a continuous and bounded density $f_0$.*
(ii) *$E_{G_0}(\|x\|) < \infty$.*

*Recall the definitions* (3.13) *and* (3.14). *Then*

(a) *$J_\lambda^{-1}(u)$ exists for all $\lambda \in \mathscr{S}_p$ and $u \in (0, 1)$, and is differentiable in $u$ with a continuous derivative.*

(b) *For each $c$ there exist constants $\gamma_1$, $\gamma_2$ and $\gamma_3$ (depending on $c$) such that if $\|\alpha\| \le \gamma_1$ and $|\delta| \le \gamma_2$, then for all $\lambda \in \mathscr{S}_p$, $J_{H_0, \alpha, \lambda}(t) \le c + \delta$ implies $t \le J_\lambda^{-1}(c) + \gamma_3(\|\alpha\| + |\delta|)$, and $J_{H_0, \alpha, \lambda}(t) \ge c - \delta$ implies $t \ge J_\lambda^{-1}(c) - \gamma_3(\|\alpha\| + |\delta|)$.*

PROOF. We begin by calculating the partial derivatives $\partial J_{H_0, \alpha, \lambda}(t)/\partial t$ and $\partial J_{H_0, \alpha, \lambda}(t)/\partial \alpha$. Conditioning on $\mathbf{x}$ yields

$$P_{H_0}\left(\frac{y - \alpha'\mathbf{x}}{\lambda'\mathbf{x}} \le t \middle| \mathbf{x}\right) = F_0(\alpha'\mathbf{x} + t\lambda'\mathbf{x}) I(\lambda'\mathbf{x} > 0) + (1 - F_0(\alpha'\mathbf{x} + t\lambda'\mathbf{x})) I(\lambda'\mathbf{x} < 0). \tag{6.1}$$

Hence given $t' \in \Re$, the mean value theorem implies

$$J_{H_0, \alpha, \lambda}(t') - J_{H_0, \alpha, \lambda}(t) = E\left[\left(I(\lambda'\mathbf{x} > 0) - E(\lambda' x < 0)\right) f_0(\eta) \lambda'\mathbf{x}(t' - t)\right],$$

where $\eta$ is such that $|\eta - t\lambda'\mathbf{x} - \alpha'\mathbf{x}| \le |(t' - t)\lambda'\mathbf{x}|$. Hence the dominated

convergence theorem implies

$$\frac{\partial J_{H_0,\alpha,\lambda}(t)}{\partial t} = E\left[ f_0(t\lambda'\mathbf{x})|\lambda'\mathbf{x}|\right], \tag{6.2}$$

which, by the assumption that $P_{G_0}(\lambda'x = 0) = 0$, is everywhere positive and continuous.

In particular for $\alpha = \mathbf{0}$, all assertions in (a) follow immediately.

Proceeding as in the proof of (6.2), we can show that

$$\frac{\partial J_{H_0,\alpha,\lambda}(t)}{\partial \alpha} = E\left[ f_0(t\lambda'\mathbf{x})(I(\lambda'\mathbf{x} > 0) - I(\lambda'\mathbf{x} < 0))\mathbf{x}\right].$$

Then (ii) implies that $\partial J_{H_0,\alpha,\lambda}(t)/\partial\alpha$ is bounded and continuous. Hence there exists $A$ such that

$$\left|J_{H_0,\alpha,\lambda}(t) - J_{H_0,\mathbf{0},\lambda}(t)\right| \le A\|\alpha\| \quad \text{for all } (\lambda, \alpha, t) \in \mathscr{S}_p \times \Re^p \times \Re. \tag{6.3}$$

To prove (b), (6.3) implies that $J_{H_0,\alpha,\lambda}(t) \ge J_{H_0,\mathbf{0},\lambda}(t) - A\|\alpha\|$. Hence

$$J_{H_0,\alpha,\lambda}(t) \le c + \delta \quad \text{implies } J_{H_0,\mathbf{0},\lambda}(t) \le c + \delta + A\|\alpha\|. \tag{6.4}$$

This implies $t \le J_\lambda^{-1}(c + \delta + A\|\alpha\|) = J_\lambda^{-1}(c) + (A\|\alpha\| + \delta)(J_\lambda^{-1})'(\eta)$, where $|\eta - c| \le A\|\alpha\| + \delta$. This proves the first implication in (b). The second one is proved likewise. □

In the next theorem we prove the $n^{1/2}$ consistency of the MP-estimates $\mathbf{T}_0$ and $\mathbf{T}_1$ as well as of their computational versions defined in Section 5. Let $\mathbf{T}_{n0}$ and $\mathbf{T}_{n1}$ be as defined in (5.1) and (5.2), respectively, with $D_n$ an equivariant set, $D_n^*$ given by (5.3) and the univariate regression estimate $T$ given by (3.11). To simplify notation define $s_0(\lambda)$ as $s(\mathscr{L}(G_0, \lambda'\mathbf{x}))$ for P1-estimators and as $(\lambda' S(G_0)\lambda)^{1/2}$ for P2-estimators.

THEOREM 6.1. *Let $(y_1, \mathbf{x}_1), \ldots, (y_n, \mathbf{x}_n)$ be i.i.d. $(p+1)$-dimensional random vectors with distribution $H_0$ given by* (2.2), *where $F_0$ is symmetric. Assume also that*:

(i) *The assumptions of the Lemma* 6.3 *are satisfied.*
(ii) $0 < \inf_{\lambda \in \mathscr{S}_p} s_0(\lambda) \le \sup_{\lambda \in \mathscr{S}_p} s_0(\lambda) < \infty$.
(iii) $p\lim_{n\to\infty} s_n(\lambda) = s_0(\lambda)$ *uniformly in* $\lambda \in \mathscr{S}_p$.
(iv) $n^{1/2} \inf_{\alpha \in D_n} \|\alpha - \alpha_0\| = O_P(1)$.

*Then*:

(a) $n^{1/2}(\mathbf{T}_{n0} - \alpha_0) = O_P(1)$.
(b) $n^{1/2}(\mathbf{T}_{n1} - \alpha_0) = O_P(1)$.

PROOF. Because of the equivariance of the estimate we can assume $\alpha_0 = \mathbf{0}$.

Put $U_n = n^{1/2} \sup_{\lambda \in \mathscr{S}_p, \alpha \in \Re^p, t \in \Re} |J_{H_n,\alpha,\lambda}(t) - J_{H_0,\alpha,\lambda}(t)|$. Then, Lemma 6.2, implies that $U_n = O_P(1)$. By assumption (iv), we can choose a sequence $\alpha_n \in D_n$

such that $n^{1/2}\boldsymbol{\alpha}_n = O_P(1)$. By the definition of $t_n$ in subsection 5.1, $J_{H_n,\boldsymbol{\alpha}_n,\boldsymbol{\lambda}}(t_n(\boldsymbol{\lambda},\boldsymbol{\alpha}_n)) = 0.5$, and this together with the definition of $U_n$, imply that $0.5 - U_n n^{-1/2} \le J_{H_0,\boldsymbol{\alpha},\boldsymbol{\lambda}_n}(t_n(\boldsymbol{\lambda},\boldsymbol{\alpha}_n)) \le 0.5 + U_n n^{-1/2}$. Hence Lemma 6.3(b) implies that

(6.5) $|t_n(\boldsymbol{\lambda},\boldsymbol{\alpha}_n)|$ can be bounded independently of $\boldsymbol{\lambda}$ by $O_P\big(n^{-1/2}U_n + \|\boldsymbol{\alpha}_n\|\big)$,

and therefore (ii) and (iii) imply that $n^{1/2}C_n(\boldsymbol{\alpha}_n) = O_P(1)$.

Put now $\boldsymbol{\lambda}_n = (\boldsymbol{\alpha}_n - \mathbf{T}_{n0})/\|\boldsymbol{\alpha}_n - \mathbf{T}_{n0}\|$. Then

$$J_{H_n,\mathbf{T}_{n0},\boldsymbol{\lambda}_n}(t_n(\boldsymbol{\lambda}_n,\mathbf{T}_{n0})) = J_{H_n,\boldsymbol{\alpha}_n,\boldsymbol{\lambda}_n}(t_n(\boldsymbol{\lambda}_n,\mathbf{T}_{n0})) - \|\boldsymbol{\alpha}_n - \mathbf{T}_{n0}\| = 0.5,$$

which implies

$$\|\boldsymbol{\alpha}_n - \mathbf{T}_{n0}\| = t_n(\boldsymbol{\lambda}_n,\mathbf{T}_{n0}) - t_n(\boldsymbol{\lambda}_n,\boldsymbol{\alpha}_n). \tag{6.6}$$

Since $\boldsymbol{\alpha}_n = O_P(n^{-1/2})$, to prove (a) it suffices to show that both terms in the right-hand side of (6.6) are $O_P(n^{-1/2})$. According to the definition of $\mathbf{T}_{n0}$, $s_n(\boldsymbol{\lambda}_n)|t_n(\boldsymbol{\lambda}_n,\mathbf{T}_{n0})| \le C_n(\boldsymbol{\alpha}_n)$, which was proved to be $O_P(n^{-1/2})$. Since (ii) and (iii) imply that $s_n$ is bounded away from 0, we have $t_n(\boldsymbol{\lambda}_n,\mathbf{T}_{n0}) = O_P(n^{-1/2})$. The bound on the other terms follows from (6.5).

Assertion (b) follows immediately from (a) and the fact that

$$n^{1/2}t_n\big(\hat{\boldsymbol{\lambda}}_n,\mathbf{T}_{n0}\big) = O_P(1). \qquad \square$$

REMARK 6.1. For P2-estimators (ii) holds if $S(G_0)$ is positive definite, and (iii) holds if $S(G_n)$ converges in probability to $S(G_0)$. For P1-estimators, (ii) and (iii) hold when $s(u_1,\ldots,u_n) = \text{median}\,(|u_1|,\ldots,|u_n|)$ and $P_{G_0}(\boldsymbol{\lambda}'\mathbf{x}) > 0.5$ for all $\boldsymbol{\lambda} \neq \mathbf{0}$. As to the validity of (iv): If $D_n$ is chosen as in the algorithm in Section 5, namely, the set of hyperplane coefficients which fit subsamples of size $p$, then (iv) requires $N$ to be at least of order $n^{p/2}$, which would yield impractically large numbers. This difficulty can be avoided by including in $D_n$ an estimator known to possess $n^{1/2}$-consistency at the model $H_0$, as the LS-estimator, or an M- or GM-estimator.

REMARK 6.2. Maronna and Yohai (1989) give heuristic arguments to show that $n^{1/2}(\mathbf{T}_{n0} - \boldsymbol{\alpha}_0)$ has a nonnormal limit distribution.

6.2. *Asymptotic bias of approximate MP-estimates.* The following theorem shows that the maximum asymptotic bias of the approximate MP-estimates $T_{n0}$ is bounded by the right-hand side of (3.15) in a stronger sense that the one defined in Section 1.

THEOREM 6.2. *Let* $\mathbf{z}_1 = (y_1,\mathbf{x}_1),\ldots,\mathbf{z}_n = (y_n,\mathbf{x}_n),\ldots$ *be i.i.d. random vectors with distribution* $H = (1-\varepsilon)H_0 + \varepsilon H_*$, *where* $H_0$ *is given by* (2.2),

*with $F_0$ symmetric. Assume also the following*:

(i) *All the assumptions of Lemma* 6.3 *hold.*

(ii) $$p \limsup_{n\to\infty} \sup_{\boldsymbol{\lambda}\in\mathscr{S}_r} s_n(\boldsymbol{\lambda}) \le d^+(G_0,\varepsilon) < \infty.$$

(iii) $$p \liminf_{n\to\infty} \inf_{\boldsymbol{\lambda}\in\mathscr{S}_p} s_n(\boldsymbol{\lambda}) \ge d^-(G_0,\varepsilon) > 0.$$

(iv) $$p \lim_{n\to\infty} \inf\{\|\boldsymbol{\alpha}-\boldsymbol{\alpha}_0\|: \boldsymbol{\alpha}\in D_n\} = 0.$$

*Then*

$$p \limsup_{n\to\infty} \|\mathbf{T}_{n0}-\boldsymbol{\alpha}_0\| \le \sup_{\boldsymbol{\lambda}\in\mathscr{S}_p} J_{\boldsymbol{\lambda}}^{-1}\left(\frac{1}{2(1-\varepsilon)}\right)\left(1+\frac{d^+(G_0,\varepsilon)}{d^-(G_0,\varepsilon)}\right).$$

PROOF. By equivariance we assume again $\boldsymbol{\alpha}_0 = \mathbf{0}$. Let $\mathbf{z}_i^* = (y_i^*, \mathbf{x}_i^*)$ be the $i$th element of the sample corresponding to $H_0$, $m_n$ the number of such elements found in the first $n$ observations and $\varepsilon_n = 1-(m_n/n)$. Let $H_n^*$ and $H_n$ be the empirical distributions of $\{\mathbf{z}_i^*,\ 1\le i\le m_n\}$ and $\{\mathbf{z}_i,\ 1\le i\le n\}$, respectively.

By assumption (iv) we can choose $\boldsymbol{\alpha}_n\in D_n$ such that $\lim_{n\to\infty}\boldsymbol{\alpha}_n = \mathbf{0}$ in probability. By definition $J_{H_n,\boldsymbol{\alpha}_n,\boldsymbol{\lambda}}(t_n(\boldsymbol{\lambda},\boldsymbol{\alpha}_n)) = 0.5$, and this implies

$$1-\frac{1}{2(1-\varepsilon_n)} \le J_{H_n^*,\boldsymbol{\alpha}_n,\boldsymbol{\lambda}}(t_n(\boldsymbol{\lambda},\boldsymbol{\alpha}_n)) \le \frac{1}{2(1-\varepsilon_n)}. \tag{6.7}$$

Since $\mathbf{z}_1^*,\dots,\mathbf{z}_{m_n}^*$ are i.i.d. with distribution $H_0$, by Lemma 6.2 we have

$$U_n = (m_n)^{1/2} \sup_{\boldsymbol{\lambda}\in\mathscr{S}_p,\ \boldsymbol{\alpha}\in\Re^p,\ t\in\Re} \left|J_{H_n^*,\boldsymbol{\alpha},\boldsymbol{\lambda}}(t) - J_{H_0,\boldsymbol{\alpha},\boldsymbol{\lambda}}(t)\right| = O_P(1). \tag{6.8}$$

Using (6.7), (6.8) and the fact that $p\lim_{n\to\infty}\varepsilon_n = \varepsilon$, for all $\delta > 0$

$$\begin{aligned} 1-\frac{1}{2(1-\varepsilon)} - \delta &\le p \liminf_{n\to\infty} J_{H_0,\boldsymbol{\alpha}_n,\boldsymbol{\lambda}}(t_n(\boldsymbol{\lambda},\boldsymbol{\alpha}_n)) \\ &\le p \limsup_{n\to\infty} J_{H_0,\boldsymbol{\alpha}_n,\boldsymbol{\lambda}}(t_n(\boldsymbol{\lambda},\boldsymbol{\alpha}_n)) \\ &\le \frac{1}{2(1-\varepsilon)} + \delta \quad \forall\,\boldsymbol{\lambda}\in\mathscr{S}_p. \end{aligned}$$

Application of both implications in Lemma 6.3(b) and the symmetry of $F_0$ yield

$$p \limsup_{n\to\infty} |t_n(\boldsymbol{\lambda},\mathbf{a}_n)| \le \sup_{\boldsymbol{\lambda}\in\mathscr{S}_p} J_{\boldsymbol{\lambda}}^{-1}\left(\frac{1}{2(1-\varepsilon)}\right). \tag{6.9}$$

Put now $\boldsymbol{\lambda}_n = (\boldsymbol{\alpha}_n - \mathbf{T}_{n0})/\|\boldsymbol{\alpha}_n - \mathbf{T}_{n0}\|$. Using (6.6), (6.9) and the fact that, by the definition of $\mathbf{T}_{n0}$, $s_n(\boldsymbol{\lambda}_n)|t_n(\boldsymbol{\lambda}_n, T)| = A_n(\boldsymbol{\lambda}_n, \mathbf{T}_{n0}) \leq C_n(\boldsymbol{\alpha}_n)$, we get

$$p \limsup_{n\to\infty} \|\boldsymbol{\alpha}_n - \mathbf{T}_{n0}\| \tag{6.10}$$
$$\leq \left( \sup_{\boldsymbol{\lambda}\in\mathscr{S}_p} J_{\boldsymbol{\lambda}}^{-1}\left(\frac{1}{2(1-\varepsilon)}\right)\right)\left(1 + \frac{\sup_{\boldsymbol{\lambda}\in\mathscr{S}_p} s_n(\boldsymbol{\lambda})}{\inf_{\boldsymbol{\lambda}\in\mathscr{S}_p} s_n(\boldsymbol{\lambda})}\right),$$

and since $p \lim_{n\to\infty} \boldsymbol{\alpha}_n = \mathbf{0}$, the theorem follows. □

REMARK 6.3. For P2-estimators, (ii) means that the matrix $S$ has an asymptotic breakdown point smaller than $\varepsilon$ at $G_0$, and (iii) holds if $S(G_n) \to S(G)$ in probability for all $G \in \mathscr{V}_{H_0,\varepsilon}$. For P1-estimators, (ii) and (iii) hold when $s(u_1, \ldots, u_n) = \text{median}(|u_1|, \ldots, |u_n|)$ and $P_{G_0}(\boldsymbol{\lambda}'\mathbf{x} \neq 0) > 0.5/(1-\varepsilon)$ for all $\boldsymbol{\lambda} \neq \mathbf{0}$. If $D_n$ is chosen as in the algorithm in Section 5 based on resampling, the validity of (iv) requires that $N$ tends to $\infty$ with $n$.

## 7. Numerical evaluation of P-estimates.

7.1. *Asymptotic bias.* We computed numerically $\beta^*(\mathbf{T}, H_0, \varepsilon)$, when $\mathbf{T}$ is an MP-estimator $\mathbf{T}_0$ or a CMP-estimator $\mathbf{T}_1$ and $H_0$ is the multivariate normal. Due to the equivariance properties, these computations were done without loss of generality assuming that $H_0$ is $(p+1)$-dimensional $N(0, I)$. We considered P2-versions of the MP- and CMP-estimators, and the scatter matrix $S$ used was the one given in Maronna, Stahel and Yohai (1992). This scatter matrix is defined using ideas similar to those used here to introduce the P-estimates, and is highly bias-robust. In Table 1 we show the values of $\beta^*$ for $\varepsilon = 0.05, 0.10, 0.15$ and $0.20$, and of $\gamma^*$. In this table we also reproduce part of Table 1 of Martin, Yohai and Zamar (1989), showing the values of $\beta$ for the bias-minimax M-estimator, the LMS-estimator and the bias-minimax GM-estimate. The minimax-bias M-estimate corresponds to minimizing the $\delta$-quantile of the residual squares. When $\varepsilon \to 0.5$ then $\delta \to 0.5$ and the minimax-bias M-estimate approaches the LMS. The minimax GM-estimator is the weighted $L_1$-estimate given by

$$T(F) = \arg\min_{\boldsymbol{\alpha}} E\left(\frac{|y - \boldsymbol{\alpha}'\mathbf{x}|}{(\mathbf{x}'\Sigma^{-1}\mathbf{x})^{1/2}}\right), \tag{7.1}$$

where $\Sigma$ is the covariance matrix of $\mathbf{x}$. Since $\Sigma$ is not known, it was replaced by the robust scatter estimate proposed by Tyler (1987). The values of the first row of this table (the value of $\beta^*$ for the minimax GM-estimates when $p = 1$) coincides with $K_{0,\varepsilon}$ defined by (4.4). Therefore we observe that the values of $\beta^*(\mathbf{T}_0, H_0, \varepsilon)$ are the same as those given by Theorem 4.1 for the nonaffine

TABLE 1
*Maximum biases of* M-, GM- *and* P-*estimates*

| $p$ | $\gamma^*$ | $\varepsilon = 0.05$ | $\varepsilon = 0.10$ | $\varepsilon = 0.15$ | $\varepsilon = 0.20$ |
|---|---|---|---|---|---|
| | | GM-estimates | | | |
| 1 | 1.57 | 0.083 | 0.18 | 0.28 | 0.41 |
| 2 | 2.00 | 0.11 | 0.25 | 0.42 | 0.68 |
| 3 | 2.35 | 0.12 | 0.29 | 0.60 | 1.39 |
| 4 | 2.67 | 0.15 | 0.39 | 0.95 | $\infty$ |
| 5 | 2.94 | 0.19 | 0.49 | 2.85 | $\infty$ |
| 6 | 3.20 | 0.21 | 0.62 | $\infty$ | $\infty$ |
| 10 | 4.06 | 0.31 | $\infty$ | $\infty$ | $\infty$ |
| 15 | 4.94 | 0.62 | $\infty$ | $\infty$ | $\infty$ |
| | | S-estimates (all $p$) | | | |
| Minimax | $\infty$ | 0.49 | 0.77 | 1.05 | 1.37 |
| LMS | $\infty$ | 0.53 | 0.83 | 1.07 | 1.52 |
| | | P-estimates (all $p$) | | | |
| MP | 3.14 | 0.163 | 0.36 | 0.56 | 0.82 |
| CMP | 1.57 | 0.085 | 0.19 | 0.31 | 0.50 |

equivariant version. We also observe that the values of $\beta^*(\mathbf{T}_1, H_0, \varepsilon)$ are very close to those given by Theorem 4.1.

We verified numerically that the P1-version of the estimate $\mathbf{T}_0$, using as scale $s(\mathbf{u}) = \text{median}|u_i|$ behaves similarly to the P2-version. The P1-version of $\mathbf{T}_1$ behaves similarly to the P2-version too, except when the contamination point has a very high leverage. In the latter case the bias of the P1-version is larger than that of the P2-version.

The reason for this difference in behavior of the two versions of the CMP-estimate is the following. Suppose that we contaminate the distribution $G_0$ of $\mathbf{x}$ with the point mass $\delta_{\mathbf{x}_0}$, where $\mathbf{x}_0 = x_0\boldsymbol{\lambda}_0$, with $\|\boldsymbol{\lambda}_0\| = 1$. Then if $|x_0|$ is very large, this contamination will have a significant increasing effect on $s(\mathscr{L}(\boldsymbol{\lambda}'\mathbf{x}))$ even when $\boldsymbol{\lambda}$ is almost orthogonal to $\boldsymbol{\lambda}_0$, that is, when $\boldsymbol{\lambda}_0'\boldsymbol{\lambda}$ is very close to 0. This will significantly influence the direction $\boldsymbol{\lambda}$ used to define $\mathbf{T}_1$. On the other hand, when we use a robust scatter matrix $S$, the contamination will have only an increasing effect along the direction $\boldsymbol{\lambda}_0$ of $S(G)$. Since this effect is bounded, the effect of the contamination on $\boldsymbol{\lambda}'S(H)\boldsymbol{\lambda}$ will be very small when $\boldsymbol{\lambda}_0'\boldsymbol{\lambda}$ is small. In view of these results we omit the values for the P1-versions. In Table 1 we also observe that the minimax GM-estimate behaves well for $\varepsilon$ and $p$ small. The CMP-estimate $\mathbf{T}_1$ behaves uniformly in $\varepsilon$ better than the minimax GM-estimate for $p \geq 2$ and than $\mathbf{T}_0$ for $p \geq 6$. Both $\mathbf{T}_0$ and $\mathbf{T}_1$ are always much better than the LMS and minimax M-estimates in terms of maximum bias.

7.2. *Monte Carlo finite sample size results.* In order to understand and compare the behavior of the P-estimators for finite sample size, we have run a

small simulation experiment. The estimates included were:

1. LS: the least-squares estimator.
2. LMS: the least median of squares estimator, computed by subsampling, with 200 subsamples.
3. MP: The P1-version of the approximate MP-estimate computed by the algorithm described in subsection 5.1, with scale $s(\mathbf{u}) = \text{median}|u_i|$. The number of subsamples $N$ was taken as 200.
4. CMP: The P1-version of the approximate CMP-estimate computed as described in subsection 5.1.
5. GM: The minimax-bias GM-estimator which is the weighted $L_1$ estimate given in (7.1)

Besides, we have included the "one-step reweighted estimators" starting from each of the estimates 2–5. These one-step estimates are obtained as a weighted LS-estimator, where each observation $(y_i, \mathbf{x}_i)$ is weighted by $w((y_i - \mathbf{T}'\mathbf{x}_i)/\hat{\sigma})$, where $\mathbf{T}$ is the starting estimate, $\hat{\sigma}$ is the scale estimate of residual error given by

$$\hat{\sigma} = 1.481 \,\text{median}|y_i - \mathbf{T}'\mathbf{x}_i| \tag{7.2}$$

and $w$ is a weight function.

Here, as proposed by Rousseeuw and Leroy (1987), we use the "hard rejection" weight function, which is defined by $w(t) = I(|t| \leq a)$, with $a = 2.5$. The one-step estimator based on LMS will be called 1-LMS, and so on. The purpose of using these one-step reweighted estimates is to gain efficiency under a perfectly observed regression model with normal errors while keeping most of the robustness of the starting estimate.

One-step Newton–Raphson M-estimates were studied by Bickel (1975). He showed that they behave asymptotically with the same efficiency as the fully iterated M-estimates. However a Monte Carlo study by Rousseeuw and Leroy (1987) showed that, at least for sample size 40, if the starting value is the LMS-estimate, then one-step Newton–Raphson M-estimates have larger mean square error than one step reweighted least square estimates; both under normal errors and under outlier contamination. Some preliminary Monte Carlo experiments confirmed that the same happens when the starting values are the MP- and CMP-estimates. For this reason we did not include one-step Newton–Raphson estimates in our Monte Carlo study.

In spite of their better asymptotic bias (see subsection 7.1) P2-estimates were not included because of the computational complexity of the scatter matrix involved, which would have required too much computer time.

In order to control the computer time, we had to choose a restricted set of sampling situations. We took two values of $p$, the number of independent variables: 2 and 5. In both cases we considered regression models without intercept. The sample size $n$ was chosen equal to 50. We consider $50(1 - \varepsilon)$ observations $(y, \mathbf{x})$ sampled from a $(p + 1)$-multivariate normal distribution

TABLE 2
*Maximum Total Square Errors*

| | | Estimates | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| $p$ | $\varepsilon$ | LS | LMS | MP | CMP | GM | 1-LMS | 1-MP | 1-CMP | 1-GM |
| 2 | 0.00 | 0.04 | 0.23 | 0.10 | 0.12 | 0.09 | 0.07 | 0.05 | 0.05 | 0.05 |
| | 0.10 | 1.98 | 0.52 | 0.22 | 0.24 | 0.17 | 0.37 | 0.17 | 0.16 | 0.18 |
| | 0.20 | 8.70 | 1.81 | 0.60 | 0.61 | 0.85 | 1.55 | 0.51 | 0.49 | 0.63 |
| 5 | 0.00 | 0.11 | 0.60 | 0.39 | 0.42 | 0.18 | 0.17 | 0.14 | 0.13 | 0.13 |
| | 0.10 | 2.26 | 1.66 | 0.82 | 0.94 | 0.67 | 0.94 | 0.50 | 0.36 | 0.51 |
| | 0.20 | 9.38 | 6.25 | 2.91 | 2.21 | 10.80 | 5.02 | 2.32 | 0.64 | 10.31 |

$N(\mathbf{0}, I)$ and the remaining $50\varepsilon$ observations are fixed at a value of the form $\tilde{\mathbf{z}} = (\tilde{y}, \tilde{x}(1, 0 \ldots, 0)')$. Therefore the true parameter $\boldsymbol{\alpha}_0 = \mathbf{0}$. Because of the equivariance properties of the estimates considered, the results that we obtain apply when the central observations come from a $N(\mathbf{0}, \Sigma)$, with arbitrary $\Sigma$ and the contamination is taken in an arbitrary direction. The contamination model used here takes a fixed number $(n(1 - \varepsilon))$ of contaminated observations, and therefore it is a little different from taking i.i.d. observations with underlying distribution $F = (1 - \varepsilon)F_0 + \varepsilon\delta_{\tilde{\mathbf{z}}}$. However the two contamination models give the same asymptotic values when $n \to \infty$.

We chose $\varepsilon = 0$, 0.10 and 0.20. The contaminating value $\tilde{x}$ was chosen as 10 and the contaminating slopes $sl = \tilde{y}/\tilde{x}$ were taken between 0.25 and 1.50 for $\varepsilon = 0.10$, and between 0.25 and 3 for $\varepsilon = 0.20$; in both cases with increments of 0.25.

For each estimator and sampling situation, we computed the total mean square error MSE = average$(\sum_{j=1}^{p} T_j^2)$, where the averages are taken over all Monte Carlo trials. For reason of space we only report, for each $p$ and $\varepsilon > 0$, the maximum of the MSE's corresponding to all values of $sl$. The results together with the MSE's for $\varepsilon = 0$ are given in Table 2. The complete tables can be requested from the authors.

The number of replications of the Monte Carlo study was 500 for $p = 2$ and 200 for $p = 5$.

We observe that the P-estimates, especially the CMP-estimates, behave clearly more robustly than the LMS- and the GM-estimates. The one-step versions of these estimates keep most of their robustness properties, while showing important gains in efficiency under the noncontaminated Gaussian model. For $p = 2$ and $\varepsilon = 0.10$ the GM-estimate and its one-step version have a behavior similar to the MP- and CMP-estimates. However for $p = 5$ the MP- and CMP-estimates are clearly more robust. For $p = 5$ the results are clearly favorable to the P-estimates, especially to $T_1$, even for $\varepsilon = 0.1$. We remark that these results were obtained using a rather low value of $N$. It seems quite plausible to expect improvements in the robustness properties of these estimates if larger values of $N$ are used.

## APPENDIX

In this section we will give a proof of Theorem 4.1. Since the MP-estimate $\mathbf{T}_0$ and the CMP-estimate $\mathbf{T}_1$ are regression equivariant we will assume from now on that $\boldsymbol{\alpha}_0 = \mathbf{0}$.

Note that since $G_0$ is spherical, the normalizing factor $s(\mathscr{L}(\boldsymbol{\lambda}'\mathbf{x}))$ in (3.3) is constant. Hence it will be omitted throughout the proof.

LEMMA A.1. *Let $M_0(z)$ be a distribution function on $\Re$ such that $q_1 = M_0^{-1}((1 - 2\varepsilon)/(2(1 - \varepsilon)))$ and $q_2 = M_0^{-1}(1/(2(1 - \varepsilon)))$ are uniquely determined. Define $M = (1 - \varepsilon)M_0 + \varepsilon\delta_{z_0}$. Then* $\operatorname{median}_M(z) = q_1$ *if* $z_0 < q_1$, $= z_0$ *if* $q_1 \le z_0 \le q_2$, *and* $= q_2$ *if* $z_0 > q_2$.

The proof of this lemma is immediate, and therefore we omit it.

Assume $\mathscr{L}(y, \mathbf{x}) = H_0$ with $\boldsymbol{\alpha}_0 = 0$ and $G_0 = \mathscr{L}(\mathbf{x})$ spherical. Let $J_{H,\boldsymbol{\alpha},\boldsymbol{\lambda}}(t)$ as defined in (3.13). For any $r \le 0$ call $J_r^*(t) = J_{H_0,\boldsymbol{\alpha},\boldsymbol{\lambda}}(t)$ with $\boldsymbol{\alpha}'\boldsymbol{\lambda} = 0$, $\|\boldsymbol{\lambda}\| = 1$ and $\|\boldsymbol{\alpha}\| = r$. Note that $J_0^*$ coincides with the function $J^*$ used in (4.4). We prove the following lemma.

LEMMA A.2. *Suppose $F_0$ has a density $f_0(u)$ symmetric and strictly decreasing for $u \ge 0$. Then*:

(a) *For any $u > 0.5$, $J_r^{*-1}(u)$ is increasing in $r$; and $J_r^{*-1}(1 - u) = -J_r^{*-1}(u)$ for any $0 \le u \le 1$.*

(b) *If $\boldsymbol{\alpha}$ and $\boldsymbol{\lambda} \in \Re^p$ with $\|\boldsymbol{\lambda}\| = 1$, then*

$$J_{H_0,\boldsymbol{\alpha},\boldsymbol{\lambda}}^{-1}(u) = J_{v(\boldsymbol{\alpha},\boldsymbol{\lambda})}^{*-1}(u) - \boldsymbol{\lambda}'\boldsymbol{\alpha},$$

*where $v(\boldsymbol{\alpha}, \boldsymbol{\lambda}) = (\|\boldsymbol{\alpha}\|^2 - (\boldsymbol{\lambda}'\boldsymbol{\alpha})^2)^{1/2}$.*

The proof of this lemma is also straightforward and may be found in Maronna and Yohai (1989).

LEMMA A.3. *Let $H_0(y, \mathbf{x}) = F_0(y)G_0(\mathbf{x})$ be the distribution of $(y, \mathbf{x})$. Suppose that $F_0$ has a density $f_0(y)$ symmetric and strictly decreasing for $y \ge 0$ and that $G_0$ is spherical. Let $H = (1 - \varepsilon)H_0 + \varepsilon\delta_{(y_0, x_0, \mathbf{e}_1)}$, where $\mathbf{e}_1 = (1, 0, \ldots, 0)'$, $x_0 > 0$ and $y_0 > 0$. Let $\mathbf{T}_0$ and $\mathbf{T}_1$ be the nonaffine equivariant versions of the* MP- *and* CMP-*estimates, respectively, and $z_0 = y_0/x_0$, then*:

(a) *If $z_0 < K_{0,\varepsilon}$, then* (3.5) *defining $\mathbf{T}_0(H)$ has infinite solutions. These solutions are all the vectors $\hat{\boldsymbol{\alpha}} = (\hat{\alpha}_1, \ldots, \hat{\alpha}_p)'$ such that $\hat{\alpha}_1 = z_0$ and $\|\hat{\boldsymbol{\alpha}}\| \le K_{0,\varepsilon}$. Besides $\mathbf{T}_1(H) = \mathbf{T}_0(H)$.*

(b) *If $K_{0,\varepsilon} \le z_0 < 2K_{0,\varepsilon}$, then $\mathbf{T}_0(H) = (z_0, 0, \ldots, 0)'$ and $\mathbf{T}_1 = (K_{0,\varepsilon}, 0, \ldots, 0)'$.*

(c) *If $z_0 \ge 2K_{0,\varepsilon}$, then $\mathbf{T}_0(H) = \mathbf{0}$ and $\mathbf{T}_1(H) = K_{0,\varepsilon}\hat{\boldsymbol{\lambda}}$, where $\|\hat{\boldsymbol{\lambda}}\| = 1$.*

PROOF. Put $K_{r,\varepsilon} = J_r^{*-1}(1/(2(1-\varepsilon)))$ [observe that $K_{0,\varepsilon}$ coincides with the definition given in (4.4)]. Take $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_p)'$ and $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_p)'$ with $\|\boldsymbol{\lambda}\| = 1$; then Lemmas A.1 and A.2 and (3.11) imply that for $\lambda_1 \neq 0$,

$$T_0(\mathscr{L}(H, y - \boldsymbol{\alpha}'\mathbf{x}, \boldsymbol{\lambda}'\mathbf{x}))$$

$$= \begin{cases} -K_{v(\boldsymbol{\alpha},\boldsymbol{\lambda}),\varepsilon} - \boldsymbol{\lambda}'\boldsymbol{\alpha}, & \text{if } (z_0 - \alpha_1)/\lambda_1 < -K_{v(\boldsymbol{\alpha},\boldsymbol{\lambda}),\varepsilon} - \boldsymbol{\lambda}'\boldsymbol{\alpha}, \\ (z_0 - \alpha_1)/\lambda_1, & \text{if } -K_{v(\boldsymbol{\alpha},\boldsymbol{\lambda}),\varepsilon} - \boldsymbol{\lambda}'\boldsymbol{\alpha} \le (z_0 - \alpha_1)/\lambda_1 \\ & \qquad \le K_{v(\boldsymbol{\alpha},\boldsymbol{\lambda}),\varepsilon} - \boldsymbol{\lambda}'\boldsymbol{\alpha}, \\ K_{v(\boldsymbol{\alpha},\boldsymbol{\lambda}),\varepsilon} - \boldsymbol{\lambda}'\boldsymbol{\alpha}, & \text{if } (z_0 - \alpha_1)/\lambda_1 > K_{v(\boldsymbol{\alpha},\boldsymbol{\lambda}),\varepsilon} - \boldsymbol{\lambda}'\boldsymbol{\alpha}. \end{cases} \tag{A.1}$$

For $\lambda_1 = 0$, the conditioning in (3.11) implies

$$T_0(\mathscr{L}(H, y - \boldsymbol{\alpha}'\mathbf{x}, \boldsymbol{\lambda}'\mathbf{x})) = \operatorname{median}_{H_0}\left(\frac{y' - \boldsymbol{\alpha}'\mathbf{x}}{\boldsymbol{\lambda}'\mathbf{x}}\right) = J^{-1}_{H_0,\boldsymbol{\alpha},\boldsymbol{\lambda}}(0.5) = -\boldsymbol{\lambda}'\boldsymbol{\alpha}$$

by Lemma A.2. Hence $\sup_{\|\boldsymbol{\lambda}\|=1,\ \lambda_1=0} |T_0(\mathscr{L}(H, y - \boldsymbol{\alpha}'\mathbf{x}, \boldsymbol{\lambda}'\mathbf{x}))| = \|\boldsymbol{\alpha}\|^2 - \alpha_1^2$.

We will first show that

$$\|\boldsymbol{\alpha}\| > 0 \quad \text{and} \quad \alpha_1 \neq z_0 \quad \text{imply} \quad C(\boldsymbol{\alpha}) > C(\mathbf{0}). \tag{A.2}$$

Note that if we put $C_0(\boldsymbol{\alpha}) = \sup_{\|\boldsymbol{\lambda}\|=1,\ \lambda_1=0} A(\boldsymbol{\alpha}, \boldsymbol{\lambda})$, then $C(\boldsymbol{\alpha}) = \max(C_0(\boldsymbol{\alpha}), \|\boldsymbol{\alpha}\|^2 - \alpha_1^2)$, and hence $C(0) = C_0(0)$. Thus we only need to consider $\lambda_1 \neq 0$ to prove (A.2).

Suppose first that $\alpha_1 < z_0$. Take a sequence $\boldsymbol{\lambda}_n = (\lambda_{n,1}, \ldots, \lambda_{n,p})' \in \mathfrak{R}^p$, with $\|\boldsymbol{\lambda}_n\| = 1$, $\lambda_{n,1} > 0$, $\lim_{n\to\infty} \lambda_{n,1} = 0$ and $\lim_{n\to\infty} \boldsymbol{\alpha}'\boldsymbol{\lambda}_n \le 0$. Put $k = -\lim_{n\to\infty} \boldsymbol{\alpha}'\boldsymbol{\lambda}_n \ge 0$ and $\mu = \lim_{n\to\infty} \nu(\boldsymbol{\alpha}, \boldsymbol{\lambda}_n) = \|\boldsymbol{\alpha}\|^2 - k^2$. Then since $(z_0 - \alpha_1)/\lambda_{n,1} \to \infty$, using (A.1) we get $\lim_{n\to\infty} T_0(\mathscr{L}(H, y - \boldsymbol{\alpha}'\mathbf{x}, \boldsymbol{\lambda}_n'\mathbf{x})) = K_{\mu,\varepsilon} + k$. If $\mu = 0$, then $k = \|\boldsymbol{\alpha}\|$ and hence $K_{\mu,\varepsilon} + k > K_{0,\varepsilon}$. If $\mu > 0$, then by part (b) of Lemma A.2, $K_{\mu,\varepsilon} > K_{0,\varepsilon}$. Thus we conclude that

$$C(\boldsymbol{\alpha}) > K_{0,\varepsilon}. \tag{A.3}$$

In the case that $\alpha_1 > z_0$, (A.3) is proved similarly by taking a sequence $\boldsymbol{\lambda}_n$ with $\lambda_{n,1} < 0$.

On the other hand, for any $\|\boldsymbol{\lambda}\| = 1$, (A.1) implies $|T_0(\mathscr{L}(H, y, \boldsymbol{\lambda}'\mathbf{x}))| \le K_{0,\varepsilon}$ and hence

$$C(\mathbf{0}) \le K_{0,\varepsilon}. \tag{A.4}$$

Since by part (b) of Lemma A.2 $K_{\|\boldsymbol{\alpha}\|,\varepsilon} > K_{\mathbf{0},\varepsilon}$, (A.3) and (A.4) imply (A.2).

Then according to (A.2), in order to find the minimum of $C(\boldsymbol{\alpha})$, we have to compare $\boldsymbol{\alpha} = \mathbf{0}$ with the $\boldsymbol{\alpha}$'s having $\alpha_1 = z_0$. Take $\boldsymbol{\alpha}$ with $\alpha_1 = z_0$; then (A.1) yields

$$A(\boldsymbol{\alpha}, \boldsymbol{\lambda}) = \begin{cases} 0, & \text{if } K_{v(\boldsymbol{\alpha},\boldsymbol{\lambda}),\varepsilon} - |\boldsymbol{\lambda}'\boldsymbol{\alpha}| > 0, \\ \boldsymbol{\lambda}'\boldsymbol{\alpha} - K_{v(\boldsymbol{\alpha},\boldsymbol{\lambda}),\varepsilon}, & \text{if } K_{v(\boldsymbol{\alpha},\boldsymbol{\lambda}),\varepsilon} - |\boldsymbol{\lambda}'\boldsymbol{\alpha}| \le 0. \end{cases} \tag{A.5}$$

Since $K_r$ is an increasing function of $r$, it is easy to see that $K_{v(\boldsymbol{\alpha},\boldsymbol{\lambda}),\varepsilon} - |\boldsymbol{\lambda}'\boldsymbol{\alpha}|$ takes its minimum value as a function of $\boldsymbol{\lambda}$ when $\boldsymbol{\lambda} = \hat{\boldsymbol{\lambda}} = \boldsymbol{\alpha}/\|\boldsymbol{\alpha}\|$. Therefore it

follows from (A.5) and the fact that $K_{v(\boldsymbol{\alpha},\hat{\boldsymbol{\lambda}}),\varepsilon} - \hat{\boldsymbol{\lambda}}'\boldsymbol{\alpha} = K_{0,\varepsilon} - \|\boldsymbol{\alpha}\|$ that

$$C(\boldsymbol{\alpha}) = \begin{cases} 0, & \text{if } \|\boldsymbol{\alpha}\| < K_{0,\varepsilon}, \\ \|\boldsymbol{\alpha}\| - K_{0,\varepsilon}, & \text{if } \|\boldsymbol{\alpha}\| \geq K_{0,\varepsilon}. \end{cases} \tag{A.6}$$

On the other hand, (A.1) implies $A(\mathbf{0}, \boldsymbol{\lambda}) \leq K_{0,\varepsilon}$ for any $\boldsymbol{\lambda} \in \Re^p$ with $\|\boldsymbol{\lambda}\| = 1$. Besides, by taking a sequence $\boldsymbol{\lambda}_n$, with $\lambda_{n,1} > 0$ and $\lim_{n\to\infty} \lambda_{n,1} = 0$ we get, again from (A.1), $\lim_{n\to\infty} A(\mathbf{0}, \boldsymbol{\lambda}_n) = K_{0,\varepsilon}$. Hence

$$C(\mathbf{0}) = K_{0,\varepsilon}. \tag{A.7}$$

If $z_0 < K_{0,\varepsilon}$, then comparing (A.6) and (A.7) it results that $C(\boldsymbol{\alpha})$ is minimized by taking any $\hat{\boldsymbol{\alpha}} = (\hat{\alpha}_1, \ldots, \hat{\alpha}_p)'$ with $\hat{\alpha}_1 = z_0$ and $\|\boldsymbol{\alpha}\| \leq K_{0,\varepsilon}$. This implies part (a) of the lemma.

Parts (b) and (c) are proved similarly. Details can be found in Maronna and Yohai (1989). □

PROOF OF THEOREM 4.1. Let $y_0 \in \Re$ and $\mathbf{x}_0 \in \Re^p$. Define $H = (1 - \varepsilon)H_0 + \varepsilon\delta_{(y_0,\mathbf{x}_0)}$. Then the symmetry of $F_0$ and the sphericity of $G_0$ imply that $\|\mathbf{T}_0(H)\|$ and $\|\mathbf{T}_1(H)\|$ depend only on $|y_0|$ and $\|\mathbf{x}_0\|$. Then Theorem 4.1 follows from Lemma A.3. □

## REFERENCES


BERAN, R. and MILLAR, P. W. (1986). Confidence sets for multivariate distributions. *Ann. Statist.* **14** 431–443.

BICKEL, P. J. (1975). One-step Huber estimates in the linear model. *J. Amer. Statist. Assoc.* **70** 597–616.

DONOHO, D. L. (1982). Breakdown properties of multivariate location estimators. Qualifying paper, Harvard Univ.

DONOHO, D. L. and HUBER, P. J. (1983). The notion of breakdown point. In *Festschrift for Erich L. Lehmann* (P. J. Bickel, K. A. Doksum and J. L. Hodges, Jr., eds.) 157–184. Wadsworth, Belmont, CA.

DONOHO, D. L. and LIU, R. C. (1988). The "automatic" robustness of minimum distance functionals. *Ann. Statist.* **16** 552–586.

HAMPEL, F. R. (1971). A general definition of qualitative 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.

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

HUBER, P. J. (1973). Robust regression: Asymptotics, conjectures and Monte Carlo. *Ann. Statist.* **1** 799–821.

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

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. (1989). A new class of bias-robust estimates for regression. Technical Report 177, Dept. Statistics, Univ. Washington, Seattle.

MARONNA, R. A. and YOHAI, V. J. (1990). The maximum bias of robust covariances. *Comm. Statist. Theory Methods* **19** 3925–3933.

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.

RIEDEL, M. (1991). Bias robustness in parametric models generated by groups. Technical Report, Sektion Mathematik, Karl-Marx-Universität, Leipzig.

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

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

ROUSSEEUW, P. J. and YOHAI, V. J. (1984). Robust regression by means of S-estimators. *Robust and Nonlinear Time Series. Lecture Notes in Statist.* **26** 256–272. Springer, New York.

STAHEL, W. A. (1981). Robuste Schätzungen: Infinitesimale Optimalität und Schätzungen von Kovarianzmatrizen. Ph.D. dissertation, ETH, Zürich.

TYLER, D. (1987). A distribution free M-estimator of multivariate scatter. *Ann. Statist.* **15** 234–251.

YOHAI, V. J. (1987). High breakdown-point and high efficiency robust estimates for regression *Ann. Statist.* **15** 642–665.

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

DEPARTAMENTO DE MATEMÁTICA
FACULTAD DE CIENCIAS EXACTAS
UNIVERSIDAD NACIONAL DE LA PLATA
CASILLA DE CORREO N° 172
1900 LA PLATA
ARGENTINA

DEPARTAMENTO DE ECONOMÍA
UNIVERSIDAD DE SAN ANDRÉS
VITO DUMAS ESQ. ARIAS
1644 VICTORIA
PROVINCIA DE BUENOS AIRES
ARGENTINA