Working Paper 92-44
Statistics and Econometrics Series 30
December 1992

Departamento de Estadística y Econometría
Universidad Carlos III de Madrid
Calle Madrid, 126
28903 Getafe (Spain)
Fax (341) 624-9849

OPTIMALLY BOUNDING A GENERALIZED GROSS ERROR SENSITIVITY OF UNBOUNDED INFLUENCE M-ESTIMATES OF REGRESSION

Victor J. Yohai and Rubén H. Zamar*

Abstract

First we show that many robust estimates of regression which depend only on the regression residuals (including M-, S-, Tau-, least median of squares- least trimmed of squares- and some R-estimates) have infinite gross-error-sensitivity. More precisely, we show that the maximum-bias function of these estimates, called residual admissible in Yohai and Zamar (1992), is of order $\sqrt{\epsilon}$ near zero. Based on this finding we define a new robustness measure, the generalized contamination sensitivity, which generalizes Hampel's gross-error-sensitivity, and compute this measure for regression M- estimates with a general scale. Then we solve Hampel's problem of minimizing the asymptotic variance subject to a bound on the generalized contamination sensitivity, for estimates in this class. Finally, we find a sharp lower bound for the generalized contamination sensitivity of residual admissible estimates and show that it is achieved by a member of the family of the least α-quantile estimates. In the Gaussian case $\alpha = .683$.

Key Words

Bias-robustness; Maximum Bias Function; Gross Error Sensitivity; Hampel's Problem; Residual Admissibility.

*Yohai, Universidad de San Andrés, Buenos Aires; Zamar, Departamento de Estadística y Econometría, Universidad Carlos III de Madrid.

OPTIMALLY BOUNDING A GENERALIZED GROSS ERROR SENSITIVITY OF UNBOUNDED INFLUENCE M-ESTIMATES OF REGRESSION

Victor J. Yohai[1]

and

Ruben H. Zamar[2]

ABSTRACT

First we show that many robust estimates of regression which depend only on the regression residuals (including M–, S–, Tau–, least median of squares– least trimmed of squares– and some R–estimates) have infinite gross–error–sensitivity. More precisely, we show that the maximum–bias function of these estimates, called residual admissible in Yohai and Zamar (1992), is of order $\sqrt{\epsilon}$ near zero. Based on this finding we define a new robustness measure, the generalized contamination sensitivity, which generalizes Hampel's gross–error–sensitivity, and compute this measure for regression M– estimates with a general scale. Then we solve Hampel's problem of minimizing the asymptotic variance subject to a bound on the generalized contamination sensitivity, for estimates in this class. Finally, we find a sharp lower bound for the generalized contamination sensitivity of residual admissible estimates and show that it is achieved by a member of the family of the least α–quantile estimates. In the Gaussian case $\alpha = .683$.

AMS 1980 subject classifications: Primary 62F35, Secondary 62J05.

Key Words: Bias-robustness, maximum bias function, gross error sensitivity, Hampel's problem, residual admissibility

---

[1]Research supported by ONR Contracts N00014-88-K-0265 N00014-91-J-1074

[2]Research supported by Natural Sciences and Engineering Research Council of Canada Grant A9276

1. Introduction. In this paper we consider the regression model

$$y_i = \theta_0' \mathbf{x}_i + u_i \quad , \quad 1 \le i \le n, \tag{1.1}$$

where $(\mathbf{x}_1, y_1), \dots (\mathbf{x}_n, y_n)$, $\mathbf{x}_i \in R^p$, $y_i \in R$ are independent observations and the u_i have a common distribution F_0 and are independent of the $\mathbf{x}_i$. We assume, for simplicity, that the carriers $\mathbf{x}_i$ are independent random vectors with common distribution G_0. Let H_0 denote the distribution function of the pair $(\mathbf{x}_i, y_i)$ under the model (1.1), that is,

$$H_0(\mathbf{x}, y) = \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_p} F_0(y - \theta_0' \mathbf{s}) dG_0(\mathbf{s}).$$

To allow for a certain fraction ϵ of contamination, i.e., a fraction ϵ of data points which do not follow the "target" model (1.1), we consider the contamination neighborhood

$$\mathcal{H}_\epsilon = \{H : (1-\epsilon)H_0 + \epsilon H^*\},$$

where $0 < \epsilon < .5$ and H^* is an arbitrary distribution on R^{p+1}.

Let $\mathbf{T}$ be an R^p valued regression and affine equivariant functional, defined on a subset of the space of distribution functions H on R^{p+1}. This subset includes the family $\mathcal{H}_\epsilon$ and all the empirical distributions functions H_n.

We define the asymptotic bias of $\mathbf{T}$ at $H \in \mathcal{H}_\epsilon$ as

$$b(\mathbf{T}, H) = \sqrt{(\mathbf{T}(H) - \theta_0)' A(G_0)(\mathbf{T}(H) - \theta_0)}, \tag{1.2}$$

where $A(G_0)$ is and affine equivariant functional, i.e., if $\mathbf{x} \sim G_0$ and $\tilde{\mathbf{x}} = B\mathbf{x} \sim \tilde{G}$ for some non-singular pxp matrix B, then $A(\tilde{G}_0) = BA(G_0)B'$. Notice that $b(\mathbf{T}, H)$ is invariant under regression equivariant transformations.

A natural measure of the degree of robustness of an estimate $\mathbf{T}$ is given by the maximum bias $B_{\mathbf{T}}(\epsilon)$ caused by a fraction ϵ of contamination,

$$B_{\mathbf{T}}(\epsilon) = \sup_{H \in \mathcal{H}_\epsilon} b(\mathbf{T}, H).$$

The function $B_{\mathbf{T}}(\epsilon)$ was introduced in the location model by Huber (1964). It was later computed for M–estimates of scale by Martin and Zamar (1989) and for S– and GM–estimates of regression by Martin, Yohai and Zamar (1989).

In order to measure the bias for "infinitesimal" values of ϵ, Hampel (1974) introduced the influence function IF_T and the gross error sensitivity GES_T which are defined by

$$IF_{\mathbf{T}}(\mathbf{x}, y) = \left[\frac{\partial}{\partial \epsilon} b(\mathbf{T}, H_{(\mathbf{x},y),\epsilon}) \right]_{\epsilon=0}, \qquad \text{GES}_{\mathbf{T}} = \sup_{(\mathbf{x},y)} IF_{\mathbf{T}}(\mathbf{x}, y), \tag{1.3}$$

where $H_{(\mathbf{x},y),\epsilon} = (1-\epsilon)H_0 + \epsilon\delta_{(\mathbf{x},y)}$ and $\delta_{(\mathbf{x},y)}$ is a point mass distribution at $(\mathbf{x}, y)$.

He and Simpson (1991) introduced the contamination sensitivity γ^* defined by

$$\gamma^*_{\mathbf{T}} = B'_{\mathbf{T}}(0). \tag{1.4}$$

which provides a linear approximation for $B_{\mathbf{T}}(\epsilon)$ for ϵ near zero

A small difference between γ^*_T and GES is the order in which the supremum and the limit for $\epsilon \to 0$ are taken. Another small difference is on the sets where the supremum is applied: to obtain γ^* one takes the supremum over $\mathcal{H}_\epsilon$ while for the GES only point–mass contaminations are considered. However, it may be shown that in sufficiently regular cases $\text{GES}_{\mathbf{T}} = \gamma^*_{\mathbf{T}}$.

Yohai and Zamar (1992) define the class of *residual admissible* regression estimates. Roughly speaking, residual admissible estimates are those for which the empirical distribution of the absolute value of their regression residuals cannot be uniformly improved by using any other set of regression coefficients. Yohai and Zamar (1992) show that many robust estimates defined as a function of the regression residuals (including M–, S–, τ–, least median of squares– (LMS), least trimmed of squares– (LTS) and some R–estimates) are residual admissible. The formal definition of residual admissible estimates is given in Section 3.

In Section 2 we show that residual admissible estimates have infinite contamination sensitivity. Based on this finding we define a new robustness measure, the generalized contamination sensitivity, and compute this measure for regression M– estimates. In Section 3 we solve the Hampel problem of minimizing the asymptotic variance subject to a bound on the generalized contamination sensitivity, for the class of M–estimates. We also find the estimate with minimum generalized contamination sensitivity in the class of residual admissible estimators. In Section 4 we numerically compute the efficiency and the generalized contamination sensitivity of the optimal Hampel estimators. We also compare numerically the maximum bias of two estimates with efficiency 0.95: the Hampel optimal estimate and the one based on the bisquare ψ–function.

2. The contamination sensitivity of M–estimates. Given $\boldsymbol{\theta} \in R^p$ and $(\mathbf{x}, y) \in R^{p+1}$ with joint distribution H, let $F_{H,\boldsymbol{\theta}}(v)$ be the distribution function of $|y - \boldsymbol{\theta}'\mathbf{x}|$.

DEFINITION 2.1. The estimating regression functional $\mathbf{T}$ is residual admissible on $\mathcal{H}$ if given two possibly substochastic distributions F_1 and F_2 which are continuous on $(0, \infty)$ and satisfy

$$F_1(v) < F_2(v), \quad \forall \; v > 0,$$

there are not a sequence $H_n \in \mathcal{H}$ and a vector $\boldsymbol{\theta}^* \in R^p$ such that $F_{H_n,\mathbf{T}(H_n)}(v)$ and $F_{H_n,\boldsymbol{\theta}^*}(v)$ are continuous on $(0, \infty)$ and

$$\lim_{n\to\infty} F_{H_n,\mathbf{T}(H_n)}(v) = F_1(v) \quad \text{and} \quad \lim_{n\to\infty} F_{H_n,\boldsymbol{\theta}^*}(v) = F_2(v), \qquad \forall \; v > 0.$$

The following theorem shows that γ_T^* of residual admissible estimates is equal to infinity. This generalizes a similar result for spherical G_0 in Yohai and Zamar (1992).

We need the following assumptions.

A.1. (i) F_0 is twice differentiable (ii) $f_0(y) = F_0'(y)$ is even, (iii) $f_0'(y) < 0$ for $y > 0$ and (iv) $\sup |f_0'(y)| < \infty$.

A.2. (i) $\mathbf{x}$ and u are independent under H_0, (ii) G_0 has second order moments and (iii) $\Sigma_0 = \mathrm{E}_{G_0}(\mathbf{x}\mathbf{x}')$ is positive definite.

Since we only work with regression and affine equivariant estimates, we can assume without loss of generality that $A(G_0) = I$ and $\boldsymbol{\theta}_0 = 0$. Accordingly, the asymptotic bias $b(\mathbf{T}, H)$ (see (1.2)) is given by the Euclidean norm of $\mathbf{T}$,

$$b(\mathbf{T}, H) = \|\mathbf{T}(H)\|.$$

In view of **A.2.**, we can work with the natural choice

$$A(G_0) = \mathrm{E}_{G_0}(\mathbf{x}\mathbf{x}').$$

This choice not only simplifies the notations but also simplifies the statement and the proof of Theorem 2.1, 2.2 and 3.1. If $A(G_0) \neq \mathrm{E}_{G_0}(\mathbf{x}\mathbf{x}')$ then we can no longer assume without loss of generality that both, $A(G_0)$ and $\mathrm{E}_{G_0}(\mathbf{x}\mathbf{x}')$, are equal to I. Therefore, the smallest and largest eigenvalues of $\mathrm{E}_{G_0}(\mathbf{x}\mathbf{x}')$ would have to be taken into account (as appropriate) in the statement and proof of Theorem 2.1, 2.2 and 3.1. On the other hand, to prove Theorem 3.2 for general $A(G_0)$, we need to assume that F_0 is Gaussian.

The following Theorem shows that residual admissible estimates have $\gamma^* = \infty$.

THEOREM 2.1 *Suppose that* **A.1** *and* **A.2** *hold. If* $\mathbf{T}(H)$ *is residual admissible, then*

$$\lim_{\epsilon\to 0} \frac{B_{\mathbf{T}}(\epsilon)}{\epsilon} = \infty.$$

PROOF. By Theorem 4.1 and Lemma A.3 of Yohai and Zamar (1992) there exists $v^*(\epsilon)$ and $\boldsymbol{\theta}^*(\epsilon)$ such that $\|\boldsymbol{\theta}^*(\epsilon)\| \leq B_{\mathbf{T}}(\epsilon)$ and

$$F_{H_0,\boldsymbol{\theta}^*(\epsilon)}(v^*(\epsilon)) - F_{H_0,0}(v^*(\epsilon)) = -\frac{\epsilon}{1-\epsilon}. \tag{2.1}$$

By the Mean Value Theorem (MVT),

$$F_{H_0,\boldsymbol{\theta}^*(\epsilon)}(v^*(\epsilon)) - F_{H_0,0}(v^*(\epsilon)) = \boldsymbol{\theta}^*(\epsilon)' \left[\frac{\partial}{\partial\boldsymbol{\theta}} F_{H_0,\boldsymbol{\theta}}(v^*(\epsilon))\right]_{\boldsymbol{\theta}=\tilde{\boldsymbol{\theta}}(\epsilon)}, \tag{2.2}$$

where $\|\tilde{\boldsymbol{\theta}}(\epsilon))\| \leq \|\boldsymbol{\theta}^*(\epsilon)\|$.

Notice that

$$\begin{aligned} F_{H_0,\boldsymbol{\theta}}(v^*) &= \mathrm{P}_{H_0}(\boldsymbol{\theta}'\mathbf{x} - v^* \leq y \leq \boldsymbol{\theta}'\mathbf{x} + v^*) \\ &= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} [F_0(\boldsymbol{\theta}'\mathbf{x} + v^*) - F_0(\boldsymbol{\theta}'\mathbf{x} - v^*)] dG_0(\mathbf{x}). \end{aligned}$$

Using the symmetry of f_0, the MVT and the Dominated Convergence Theorem we get

$$\begin{aligned} \left\|\tfrac{\partial}{\partial\boldsymbol{\theta}} F_{H_0,\boldsymbol{\theta}}(v^*(\epsilon))\right\| &= \|\textstyle\int \cdots \int [f_0(\boldsymbol{\theta}'\mathbf{x} + v^*(\epsilon)) - f_0(\boldsymbol{\theta}'\mathbf{x} - v^*(\epsilon))]\mathbf{x} dG_0(\mathbf{x})\| \\ &= 2\,\|\textstyle\int \cdots \int f_0'(\beta(v^*(\epsilon), \boldsymbol{\theta}'\mathbf{x}))\mathbf{x}\mathbf{x}'\boldsymbol{\theta} dG_0(\mathbf{x})\| \\ &\leq 2 \sup |f_0'(y)|\, \|\mathrm{E}_{G_0}\{\mathbf{x}\mathbf{x}'\}\boldsymbol{\theta}\| = 2 \sup_y |f_0'(y)| \|\boldsymbol{\theta}\|. \end{aligned} \tag{2.3}$$

By (2.1), (2.2) and (2.3)

$$\begin{aligned} \tfrac{\epsilon}{1-\epsilon} = \left|\boldsymbol{\theta}^*(\epsilon)' \left[\tfrac{\partial}{\partial\boldsymbol{\theta}} F_{H_0,\boldsymbol{\theta}}(v^*(\epsilon))\right]_{\boldsymbol{\theta}=\tilde{\boldsymbol{\theta}}(\epsilon))}\right| &\leq \|\boldsymbol{\theta}^*(\epsilon)\| \|\tfrac{\partial}{\partial\boldsymbol{\theta}} F_{H_0,\boldsymbol{\theta}}(v^*(\epsilon))\big|_{\boldsymbol{\theta}=\tilde{\boldsymbol{\theta}}(\epsilon))}\| \\ &\leq 2 \sup_y |f_0'(y)| \|\boldsymbol{\theta}^*(\epsilon)\|^2 \leq 2 \sup_y |f_0'(y)| B_{\mathbf{T}}^2(\epsilon). \end{aligned}$$

Therefore,

$$\underline{\lim}_{\epsilon\to 0} \frac{B_{\mathbf{T}}^2(\epsilon)}{\epsilon^2} \geq \lim_{\epsilon\to 0} \frac{1}{2|\sup_y f_0'(y)|\epsilon} = \infty, \tag{2.4}$$

and the theorem follows.

An important subfamily of regression admissible estimates is the class of M–estimates of regression with general scale. These estimates are defined as (see Martin et al., 1989)

$$\mathbf{T}(H) = \arg\min_{\mathbf{t}} \mathrm{E}_H \left(\rho\left(\frac{y - \mathbf{t}'x}{s(H)}\right)\right), \tag{2.5}$$

where $s(H)$ is an estimate of the scale of the regression residuals and

A.3. ρ has the following properties: (i) even, (ii) continuous at 0, (iii) monotone on $[0,\infty)$, (iv) $\rho(0)=0$, (v) $0<\lim_{u\to\infty}\rho(u)<\infty$.

The class of M–estimates of regression with general scale includes Huber (1973) M–estimates, Yohai (1987) MM–estimates, Rousseeuw (1984) LMS–estimates and Rousseeuw and Yohai (1984) S–estimates. Yohai and Zamar (1992) show that when (i) ρ is bounded and (ii) $s(H)$ is bias–robust, the corresponding M–estimate of regression with general scale is residual admissible. On the other hand, if either one of these conditions does not hold, then the M–estimate has breakdown point equal to zero, that is, $B_{\mathbf{T}}(\epsilon)=\infty$ for all $\epsilon>0$. M–estimates which satisfy (i) and (ii) will be called "robust M–estimates".

The maximum bias function $B_{\mathbf{T}}(\epsilon)$ of robust M–estimates is continuous but not differentiable at zero. Therefore, we don't have a linear approximation for $B_{\mathbf{T}}(\epsilon)$ for small ϵ. It shall be shown in Theorem 2.2, however, that $B^2_{\mathbf{T}}(\epsilon)$ is differentiable at $\epsilon=0$ and therefore $B_{\mathbf{T}}(\epsilon)$ is proportional to $\sqrt{\epsilon}$ for ϵ near zero. More precisely, there exists a constant $\gamma^{**}_{\mathbf{T}}$ such that

$$B_{\mathbf{T}}(\epsilon)=\gamma^{**}_{\mathbf{T}}\sqrt{\epsilon}+o(\sqrt{\epsilon}).$$

The proportionality constant $\gamma^{**}_{\mathbf{T}}$ plays a role similar to the contamination sensitivity $\gamma^{*}_{\mathbf{T}}$. That is, $\gamma^{**}_{\mathbf{T}}$ is a "one–figure summary" of the behavior of $B_{\mathbf{T}}(\epsilon)$ when ϵ is small. Therefore, the constant $\gamma^{**}_{\mathbf{T}}$ will be called *the generalized contamination sensitivity* of $\mathbf{T}$. The notations

$$g(\mathbf{t},s)=\mathrm{E}_{H_0}\left[\rho\left(\frac{y-\mathbf{t}'\mathbf{x}}{s}\right)\right]$$

and

$$g^{(i)}(\mathbf{t},s)=\frac{\partial^i g(\mathbf{t},s)}{\partial \mathbf{t}^i}$$

are needed for the statement and the proof of Theorem 2.2. Since we are assuming **A.2** and $\Sigma_0=I$,

$$g^{(2)}(0,s)=\lambda(F_0,s)I, \tag{2.6}$$

where

$$\lambda(F_0,s)=\left[\frac{\partial^2}{\partial v\partial v}\mathrm{E}_{F_0}\left(\rho\left(\frac{y-v}{s}\right)\right)\right]_{v=0}.$$

If in addition to **A.3** ρ is absolutely continuous, then by **A.1**

$$\lambda(F_0,s)=\left[-\frac{1}{s}\frac{\partial}{\partial v}\mathrm{E}_{F_0}\left(\psi\left(\frac{y-v}{s}\right)\right)\right]_{v=0}=-\int_{-\infty}^{\infty}\psi\left(\frac{y}{s}\right)f_0'(y)dy, \tag{2.7}$$

where $\psi = \rho'$.

On the other hand, if ρ_a^J is the jump function with jumps at $-a$ and a, i.e.,

$$\rho_a^J(u) = \begin{cases} 0 & \text{if } |u| \leq a \\ 1 & \text{if } |u| > a, \end{cases} \tag{2.8}$$

then

$$\lambda(F_0, s) = -2f_0'(sa). \tag{2.9}$$

Before proving Theorem 2.2, we need the following lemma.

LEMMA 2.1 *Suppose that* **A.1**–**A.3** *hold, then* $g(\mathbf{t}, s) > g(0\ , s)$ *for all* $\mathbf{t} \neq 0$ *and all* $s > 0$.

PROOF. For $a \in R$ and $s > 0$, define

$$h(a, s) = E_{F_0}\left(\rho\left(\frac{y-a}{s}\right)\right).$$

We will show first that

$$h(a, s) > h(0, s) \quad \forall a \neq 0 \tag{2.10}$$

Since by **A.1** and **A.3**, $h(a, s) = h(-a, s)$, we will prove (2.10) for $a > 0$. Since

$$h(a, s) = \int_{-\infty}^{\infty} \rho\left(\frac{u}{s}\right) f_0(u + a)du,$$

using **A.1** and **A.3**, we get

$$\frac{\partial}{\partial a}h(a, s) = \int_{-\infty}^{\infty} \rho\left(\frac{u}{s}\right) f_0'(u + a)du = \int_{-\infty}^{\infty} \rho\left(\frac{u-a}{s}\right) f_0'(u)du$$

$$= \int_0^{\infty} \left(\rho\left(\frac{u-a}{s}\right) - \rho\left(\frac{u+a}{s}\right)\right) f_0'(u)du. \tag{2.11}$$

Since $(\rho(\frac{u-a}{s}) - \rho(\frac{u+a}{s}))f_0'(u) \geq 0$ for $u \geq 0$, we get

$$\frac{\partial}{\partial a}h(a, s) \geq 0 \quad \forall a.$$

Then, to prove (2.10) it is enough to show that it holds for $a \leq a_0$ for some a_0. By **A.3**, there exist $a_0 > 0$ such that $\rho(\frac{u}{s}) > \rho(\frac{a_0}{s})$ for $u > a_0$. Therefore if $0 < a < a_0$ and $u > a_0 - a$, by **A.3** we get $(\rho(\frac{u-a}{s}) - \rho(\frac{u+a}{s})) < 0$, and then by **A.1** and (2.11)

$$\frac{\partial}{\partial a}h(a, s) \geq \int_{a_0 - a}^{\infty} \left(\rho\left(\frac{u-a}{s}\right) - \rho\left(\frac{u+a}{s}\right)\right) f_0'(u)du > 0.$$

Using **A.2** we get that $g(\mathbf{t}, s) = E_{G_0}(h(\mathbf{t}'\mathbf{x}, s))$. Since by **A.2**, $P(\mathbf{t}'\mathbf{x} \neq 0) > 0$ for all $\mathbf{t} \neq 0$, the Lemma follows from (2.10).

THEOREM 2.2 *Suppose that* **A.1–A.3** *hold and*
(i) All the terms of $g^{(2)}(\mathbf{t}, s)$ *are continuous and bounded.*
(ii) $\sup_{H \in \mathcal{H}_\epsilon} |s(H) - s(H_0)| \to 0$ *as* $\epsilon \to 0$. *Then*
(a)

$$\lim_{\epsilon \to 0} B_{\mathbf{T}}(\epsilon) = 0,$$

and
(b)

$$\lim_{\epsilon \to 0} \frac{B_{\mathbf{T}}^2(\epsilon)}{\epsilon} = \frac{2\rho(\infty)}{\lambda(F_0, s(H_0))}.$$

PROOF. Let

$$J_H(\mathbf{t}, s) = \mathrm{E}_H \left(\rho \left(\frac{y - \mathbf{t}'x}{s} \right) \right).$$

By definition of $\mathbf{T}(H)$,

$$J_H(\mathbf{T}(H), s(H)) \le J_H(0, s(H)), \quad \forall H \in \mathcal{H}_\epsilon.$$

Using this, together with the monotonicity and boundness of ρ (see **A.3**) we follow that

$$g(\mathbf{T}(H), s(H)) - g(0, s(H)) \le \frac{\epsilon}{1-\epsilon}\rho(\infty), \quad \forall H \in \mathcal{H}_\epsilon, \tag{2.12}$$

and part (a) of the theorem follows from Lemma 2.1 and (i).

Since $g^{(1)}(0\ , s(H)) = 0$, a Taylor expansion of the left hand side of (2.12) gives

$$\begin{aligned} \frac{2\epsilon}{1-\epsilon}\rho(\infty) &\ge 2[g(\mathbf{T}(H), s(H)) - g(0, s(H))] \\ &= \mathbf{T}(H)' g^{(2)}(\mathbf{t}^*(H), s(H)) \mathbf{T}(H) \\ &\ge \|\mathbf{T}(H)\|^2 |\mu_1(g^{(2)}(\mathbf{t}^*, s(H)))|, \end{aligned}$$

where $\|\mathbf{t}^*(H)\| \le B_{\mathbf{T}}(\epsilon)$ and $\mu_1(A)$ denotes the eigenvalue of A with minimum module.

Therefore

$$\frac{B_T^2(\epsilon)}{\epsilon} = \frac{\sup_{H \in \mathcal{H}_\epsilon} \|\mathbf{T}(H)\|^2}{\epsilon} \le \frac{2\rho(\infty)}{\liminf_{\epsilon \to 0} \inf_{H \in \mathcal{H}_\epsilon} |\mu_1(g^{(2)}(\mathbf{t}^*(H), s(H)))|}.$$

Therefore, using (2.6), conditions (i), (ii), and part (a) of the Theorem, we get

$$\overline{\lim}_{\epsilon \to 0} \frac{B_{\mathbf{T}}^2(\epsilon)}{\epsilon} \le \frac{2\rho(\infty)}{\lambda(F_0, s(H_0))}. \tag{2.13}$$

On the other hand, let $\epsilon_n > 0$ and $k_n > 0$ be such that

$$\lim_{n\to\infty} k_n\Delta_n = \infty, \quad \Delta_n = \sqrt{\frac{2\epsilon_n\rho(\infty)}{\lambda(F_0, s(H_0))}}.$$

Let $\mathbf{a}_0$ be a unit vector and consider the sequences $\mathbf{x}_n = k_n\mathbf{a}_0$, $y_n = (1-\eta)k_n\Delta_n$, $\boldsymbol{\theta}_n = (1-\eta)\Delta_n\mathbf{a}_0$ (for some $0 < \eta < 1$) and

$$H_n = (1-\epsilon_n)H_0 + \epsilon_n\delta_{(\mathbf{x}_n, y_n)}, \quad \mathbf{x}_n = k_n\mathbf{a}_0, \quad y_n = (1-\eta)k_n\Delta_n.$$

A Taylor expansion yields

$$\begin{aligned} J_{H_n}(\boldsymbol{\theta}_n, s(H_n)) &= (1-\epsilon_n)g(\boldsymbol{\theta}_n, s(H_n)) \\ &= (1-\epsilon_n)\left[g(0, s(H_n)) + \frac{1}{2}\boldsymbol{\theta}'_n g^{(2)}(\mathbf{t}^*_n, s(H_n))\boldsymbol{\theta}_n\right] \\ &= (1-\epsilon_n)\left[g(0, s(H_n)) + \frac{1}{2}(1-\eta)^2\Delta_n^2\mathbf{a}'_0 g^{(2)}(\mathbf{t}^*_n, s(H_n))\mathbf{a}_0\right], \end{aligned}$$

where $\|\mathbf{t}^*(H_n)\| \le \Delta_n$. Therefore, using conditions (i) and (ii) and (2.6) we get

$$\lim_{n\to\infty}\frac{1}{\epsilon_n}\left[J_{H_n}(\boldsymbol{\theta}_n, s(H_n)) - (1-\epsilon_n)g(0, s(H_n))\right] = (1-\eta)^2\rho(\infty) < \rho(\infty). \tag{2.14}$$

On the other hand, if $\|\tilde{\boldsymbol{\theta}}_n\| \le (1-\delta)\|\boldsymbol{\theta}_n\|$ (for some $0 < \delta < 1$), that is, if $\tilde{\boldsymbol{\theta}}_n$ is of the form

$$\tilde{\boldsymbol{\theta}}_n = \delta_n(1-\eta)(1-\delta_n)\Delta_n\mathbf{a}_n, \quad \delta_n \le \delta, \quad \lim_{n\to\infty}\delta_n = \delta, \quad \|\mathbf{a}_n\| = 1,$$

then

$$\frac{y_n - \tilde{\boldsymbol{\theta}}'_n\mathbf{x}_n}{s(H_n)} = \frac{k_n\Delta_n(1-\eta)(1-(1-\delta_n)\mathbf{a}'_0\mathbf{a}_n)}{s(H_n)} \to \infty, \quad \text{as } n\to\infty,$$

and so,

$$\lim_{n\to\infty}\frac{1}{\epsilon_n}\left[J_{H_n}(\tilde{\boldsymbol{\theta}}_n, s(H_n)) - (1-\epsilon_n)g(0, s(H_n))\right] \ge \lim_{n\to\infty}\rho\left(\frac{y_n - \tilde{\boldsymbol{\theta}}'_n\mathbf{x}_n}{s(H_n)}\right) = \rho(\infty). \tag{2.15}$$

If n large enough, by (2.14) and (2.15),

$$J_{H_n}(\boldsymbol{\theta}_n, s(H_n)) < J_{H_n}(\tilde{\boldsymbol{\theta}}_n, s(H_n)) \;\; \forall\, \|\tilde{\boldsymbol{\theta}}_n\| \le (1-\delta)\|\boldsymbol{\theta}_n\|.$$

Therefore, for n large enough, $\|\mathbf{T}(H_n)\| \ge (1-\delta)\|\boldsymbol{\theta}_n\|$, and

$$B^2_{\mathbf{T}}(\epsilon_n) \ge \|\mathbf{T}(H_n)\|^2 \ge (1-\delta)^2\|\boldsymbol{\theta}_n\|^2 = \frac{2(1-\eta)(1-\delta)^2\epsilon_n\rho(\infty)}{\lambda(F_0, s(H_n))}. \tag{2.16}$$

Finally,

$$\underline{\lim}_{n\to\infty} \frac{B_{\mathbf{T}}^2(\epsilon_n)}{\epsilon_n} \geq \frac{2(1-\eta)(1-\delta)^2\rho(\infty)}{\lambda(F_0, s(H_0))}. \tag{2.17}$$

Since (2.17) holds for all $\delta > 0$ and all $\eta > 0$, the theorem follows from (2.13).

REMARK 2.1. Notice that, so far, $B_{\mathbf{T}}(\epsilon)$ has only been derived for S–estimates of regression, when the carriers have an elliptical distribution (see Martin et al. 1989). Therefore, Theorem 2.2 provides the useful approximation $\gamma^{**}\sqrt{\epsilon}$ for the unknown maximum bias function of other M–estimates of regression (e.g. MM–estimates) and, in general, for all the M–estimates in the case of non–elliptical carriers.

REMARK 2.2. A sufficient condition for assumption (i) in Theorem 2.2 is that f_0'' be continuous and bounded. An alternative sufficient condition is that both, $\psi = \rho'$ and f_0', be continuous and bounded.

3. Optimally bounding the generalized contamination sensitivity of M–estimates. Krasker and Welch (1982) solved Hampel's problem of optimally bounding the contamination sensitivity for regression GM–estimates, that is, they found an efficient, bounded influence regression estimate, with near optimal bias–robustness properties when ϵ is small. Unfortunately, the maximum bias behavior of the optimal GM–estimate is disappointingly bad when the dimension of $\mathbf{x}$ is large.

Martin et al. (1989) show that a certain least α–quantile estimate, $\mathbf{T}_\alpha$, defined by the property of minimizing the α-quantile of the absolute value of the regression residuals, is minimax–bias in the class of M–estimates of regression with general scale (see (2.5)) when G_0 is elliptical. The quantile α depends on ϵ and H_0, but in general $\mathbf{T}_\alpha$ is well approximated by Rousseeuw's least median of squared residuals. Unfortunately, $\mathbf{T}_\alpha$ is very inefficient at H_0 and, unless the sample size is exceedingly large, efficiency considerations must also be taken into account.

Then one wishes to find the M–estimate of regression with general scale that minimizes the generalized asymptotic variance at H_0 (i.e. the trace of the asymptotic variance) subject to a bound on the maximum bias. Unfortunately, the ensuing optimality problem is untractable except in the case that G_0 is elliptical and the error's scale (under the central model) is known; and even in this case the technical difficulties are considerable.

On the other hand, it has been shown by Martin and Zamar (1992) that in the location case the solution of the exact problem is very close to that of the Hampel's problem where the maximum bias is approximated in terms of the generalized contamination sensitivity. One then expects that the same is true for the regression case.

We now consider a generalization of the Hampel's optimality problem for the class of robust M–estimates of regression with general scale. These estimates have good maximum bias behavior independently of the dimension of $\mathbf{x}$.

To simplify the notations we also assume that $s(H_0) = 1$. It is well known that (see Yohai, 1987), under mild regularity conditions, M–estimates of regression with general scale are asymptotically normal with covariance matrix

$$AV(\psi, H_0) = V(\psi, H_0)\,(\mathrm{E}_{G_0}(\mathbf{x}\mathbf{x}'))^{-1},$$

where

$$V(\psi, H_0) = \frac{\mathrm{E}_{F_0}(\psi^2(y))}{\lambda^2(F_0, 1)}, \tag{3.1}$$

and where $\psi(y) = \rho'(y)$ and $\lambda(F_0, s)$ is given by (2.7).

On the other hand, if we denote by $\gamma^{**}(\rho, H_0) = \gamma^{**}_{\mathbf{T}}$, where $\mathbf{T}$ is an M–estimate corresponding to the function ρ, by Theorem 2.2,

$$\gamma^{**}(\rho, H_0) = \sqrt{\frac{2\rho(\infty)}{\lambda(F_0, 1)}}. \tag{3.2}$$

We find the regression M–estimate with general scale which minimizes $V(\psi, F_0)$ subject to a bound on γ^{**}.

Since $\psi^* = c\psi$ gives the same M–estimate than ψ, we can assume without loss of generality that $\lambda(F_0, 1) = 2$. Therefore using (3.1) and (3.2) the Hampel problem is equivalent to search the function ψ such that

$$\min_{\psi} \mathrm{E}_{F_0}(\psi^2(y)),$$

subject to

(i) $\rho(\infty) = \int_0^\infty \psi(u)\, du \leq K^2$

(ii) $\lambda(F_0, 1) = 2$

(iii) $\psi(0) = 0$ and $\psi(y) \geq 0$, $\forall y \geq 0$.

Using the notations

$$\Delta_0(y) = -\frac{f_0'(y)}{f_0(y)}, \quad \Delta_1(y) = \frac{1}{f_0(y)},$$

$$\langle \psi_1, \psi_2 \rangle = \int_0^\infty \psi_1(y)\psi_2(y) f_0(y) dy,$$

the Hampel problem can be rewritten as

$$\min_{\psi} \langle \psi, \psi \rangle,$$

subject to

(i) $\langle \psi, \Delta_1 \rangle \leq K^2$,

(ii) $\langle \psi, \Delta_0 \rangle = 1$,

(iii) $\psi(0) = 0$ and $\psi(y) \geq 0$, $\forall\, y \geq 0$.

For each $\beta > 0$ let

$$q_\beta(y) = \Delta_0(y) - \beta\Delta_1(y). \tag{3.3}$$

We will require that F_0 satisfy the following assumption:

A.4. $f_0(u) > 0\ \forall u$, $f_0'(u)$ continuous, and $\lim_{u\to\infty} f_0'(u) = 0$.

Then we have the following Lemma

LEMMA 3.1 *Assume* **A.1** *and* **A.4**. *Denote by* $\beta_0 = \max_y |f'(y)|$, *and by* q^+ *the positive part of the function* q. *Then for all* $\beta < \beta_0$:
(a) there exists $c_0(\beta) > 0$ *and* $c_1(\beta) > 0$ *such that* $q_\beta^+(y) = 0$ *for all* $0 \le y < c_0(\beta)$ *and all* $y > c_1(\beta)$.
(b) $\langle q_\beta^+, \Delta_0 \rangle > 0$

PROOF. Follows immediately from the assumptions and the fact that

$$q_\beta(y) = \frac{-f_0'(y) - \beta}{f_0(y)}.$$

According to Lemma 3.1 we can define the family of functions

$$\psi_\beta^*(y) = \frac{q_\beta^+(y)}{\langle q_\beta^+, \Delta_0 \rangle} \quad 0 \le \beta < \beta_0. \tag{3.4}$$

The functions in this family satisfy conditions (ii) and (iii) above. The following Theorem show that the functions ψ_β^* defined in (3.4) are solutions to the Hampel problem.

THEOREM 3.1. *Suppose that* **A.1–A.4** *hold. Then for all* $\beta < \beta_0$, *the function* ψ_β^* *solves de Hampel problem with*

$$K^2 = K^2(\beta) = \langle \psi_\beta^*, \Delta_1 \rangle = \rho_\beta^*(\infty), \tag{3.5}$$

where $\rho_\beta^*(y) = \int_0^y \psi_\beta^*(t)dt$ *is the corresponding integrated loss function.*

PROOF. Suppose that ψ satisfies (ii) and (iii) above and

$$\langle \psi, \Delta_1 \rangle \le K^2(\beta),$$

where $K^2(\beta)$ is given by (3.5).

For some $c > c_1(\beta)$ let

$$\overline{\psi}(y) = \frac{\psi(y)}{\langle \psi, \Delta_0 \rangle_c},$$

where

$$\langle \psi_1, \psi_2 \rangle_c = \int_0^c \psi_1(y)\psi_2(y) f_0(y)dy, \quad c > 0.$$

Since by **A.1** the function

$$h(c) = \frac{\langle \psi, \Delta_1 \rangle_c}{\langle \psi, \Delta_0 \rangle_c} = \frac{\rho(c)}{\int_0^c \psi(y)\Delta_0(y) f_0(y)dy} = -\frac{\rho(c)}{\int_0^c \psi(y) f_0'(y)dy},$$

is non-decreasing $\overline{\psi}$ satisfies

(a) $\langle \overline{\psi}, \Delta_1 \rangle_c \leq K^2(\beta)$,
(b) $\langle \overline{\psi}, \Delta_0 \rangle_c = 1$,
(c) $\overline{\psi}(0) = 0$ and $\overline{\psi}(y) \geq 0, \quad \forall\, y \geq 0$.

Let

$$d(\beta) = \langle q_\beta^+, \Delta_0 \rangle_c = \langle q_\beta^+, \Delta_0 \rangle,$$

where the last equality holds because $c > c(\beta)$. Since $\overline{\psi}(y) \geq 0$ for all $y \geq 0$, the definition of ψ_β^* implies that

$$\int_0^c \left[\overline{\psi}(y) - \frac{q_\beta(y)}{d(\beta)}\right]^2 f_0(y)dy \geq \int_0^c \left[\psi_\beta^*(y) - \frac{q_\beta(y)}{d(\beta)}\right]^2 f_0(y)dy,$$

and so

$$\langle \overline{\psi}, \overline{\psi} \rangle_c \geq \langle \psi_\beta^*, \psi_\beta^* \rangle_c + \frac{2}{d(\beta)}[\langle \overline{\psi}, q_\beta \rangle_c - \langle \psi_\beta^*, q_\beta \rangle_c]. \tag{3.6}$$

Since $c > c(\beta)$, $\langle \psi_\beta^*, \Delta_0 \rangle_c = \langle \psi_\beta^*, \Delta_0 \rangle$ and so

$$\begin{aligned}\langle \overline{\psi}, q_\beta \rangle_c - \langle \psi_\beta^*, q_\beta \rangle_c &= \langle \overline{\psi}, \Delta_0 \rangle_c - \beta \langle \overline{\psi}, \Delta_1 \rangle_c - \langle \psi_\beta^*, \Delta_0 \rangle + \beta \langle \psi_\beta^*, \Delta_1 \rangle \\ &= \beta[K^2(\beta) - \langle \overline{\psi}, \Delta_1 \rangle_c] \geq 0.\end{aligned}$$

Therefore, using (3.6) we get

$$\langle \overline{\psi}, \overline{\psi} \rangle_c \geq \langle \psi_\beta^*, \psi_\beta^* \rangle, \quad \forall\, c > c(\beta).$$

Finally, by the Dominated Convergence Theorem,

$$\langle \psi, \psi \rangle = \lim_{c \to \infty} \langle \overline{\psi}, \overline{\psi} \rangle_c \geq \langle \psi_\beta^*, \psi_\beta^* \rangle,$$

proving the theorem.

REMARK 3.1. In the important Gaussian case the assumptions **A.1** and **A.4** are satisfied. In this case

$$q_\beta(y) = y - \sqrt{2\pi}\beta e^{\frac{y^2}{2}}$$

and according to Lemma 3.1 ψ_β^* vanished outside the interval $(c_0(\beta), c_1(\beta))$. Inside this interval $\psi_\beta^*(y) > 0$. The corresponding loss function ρ_β^* is of the form $\rho_\beta^*(y) = 0$ for $|y| \leq c_0(\beta)$, $\rho_\beta^*(y) = \rho_\beta^*(c_1(\beta))$ for $|y| \geq c_1(\beta)$ and $\rho_\beta^*(y)$ is strictly monotone for $c_0(\beta) \leq |y| \leq c_1(\beta)$.

We finish this section by deriving the unconstrained residual admissible estimate with minimum γ^{**}.

THEOREM 3.2. *Let T^* be the M-estimate corresponding to the jump function $\rho^J_{\beta_0}$ given in (2.8). Then*

$$\gamma^{**}_{\mathbf{T}} \geq \gamma^{**}_{\mathbf{T}^*},$$

for all residual admissible estimate $\mathbf{T}$.

PROOF. From (2.4) (in the proof of Theorem 2.1) and (2.9) we have

$$\gamma^{**}_{\mathbf{T}} = \lim_{\epsilon \to 0} \frac{B_{\mathbf{T}}(\epsilon)}{\sqrt{\epsilon}} \geq [2 \sup_{y \in R} |f_0'(y)|]^{-1} = [2|f_0'(\beta_0)|]^{-1} = \gamma^{**}_{\mathbf{T}^*}.$$

REMARK 3.2. It is easy to see that $\rho^J_{\beta_0}(y)$ may be obtained as a limit of $\rho^*_\beta/\rho^*_\beta(\infty)$ when $\beta \to \beta_0$.

REMARK 3.3. One way to define an M–estimate with loss function ρ^J_α is using the least quantile estimate Q_α defined by the estimating functional

$$Q_\alpha(H) = \arg\min_{\mathbf{t} \in R^p} F^{-1}_{H,\mathbf{t}}(\alpha)$$

and taking as $\alpha = 1 - F_0(\beta_0)$. When F_0 is a normal distribution $\alpha = .683$.

4. Some numerical results. When F_0 is N(0,1) the family of optimal ψ–function obtained in Section 3 is given by

$$\psi^*_\beta(y) = \mathrm{sign}(y)[|y| - \sqrt{2\pi}\beta e^{\frac{y^2}{2}}]^+. \tag{3.7}$$

In Table 1 we show the asymptotic efficiency (AEFF) given by $V(\psi, F_0)^{-1}$ and γ^{**} for different values of β. We also show the interval $[c_0,\ c_1]$ where the function is different from 0. The limit case of $\beta = .242$ corresponds to a jump rho– function ρ^J_1.

TABLE 1 ABOUT HERE

In Table 2 we compare the asymptotic maximum bias and γ^{**} corresponding to two ψ–functions, one in the optimal family given in (3.7) and the other in the bisquare family

$$\psi^B_c(y) = y\left(1 - \frac{y^2}{c^2}\right)^2 I_{[0,c]}(|y|).$$

Both estimates have AEFF=0.95 for normal errors and the the maximum biases are computed assuming that $G_0 = \mathrm{N}(0\ , I)$ and the error scale is known.

TABLE 2 ABOUT HERE

Notice that the improvements in bias for the optimal estimate are aproximately proportional to the reduction in the value of γ^{**}. In Figure 1 we plot the corresponding psi-functions.

REFERENCES

Hampel, F. R. (1974). The influence curve and its role in robust estimation. *Ann. Statist.* **69**, 383–393.

He, X. and Simpson, D. G. (1991). Bias–robust estimation for parametric models. Unpublished manuscript.

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

Krasker, W. S. and Welsch, R. E. (1982). Efficient bounded influence regression estimation. *J. Am. Statist. Assoc.* **77**, 595–604.

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

Martin R. D. and Zamar, R. H. (1992). Efficient minimax estimates of location. To appear in the *Ann. of Statist.*

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

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

Rousseeuw P. J. and Yohai, V. J. (1984). Robust regression by means of S–estimators. In *Robust and Nonlinear Time Series Analysis*, J. Franke, W. Hardle, and R. D. Martin (eds.), Lecture Notes in Statistics **26**, Springer, New York, 256–272.

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

Yohai, V. J. and Zamar, R. H. (1992). Minimax residual admissible estimates of regression. To appear in *Ann. Statist.*.

Víctor J. Yohai
Universidad de San Andrés
Vito Dumas 284, esq. Arias
Victoria, Pcia. de Buenos Aires
Argentina

Rubén Zamar
University of British Columbia
West Mall 2021
Vancouver, British Columbia
Canada V6T 1Z2

β	c_0	c_1	AEFF	γ^{**}
0.014	0.035	2.98	0.95	2.78
0.028	0.08	2.70	0.90	2.57
0.060	0.15	2.35	0.80	2.36
0.094	0.25	2.10	0.70	2.25
0.128	0.34	1.89	0.60	2.17
0.160	0.45	1.71	0.50	2.12
0.185	0.54	1.56	0.40	2.09
0.209	0.65	1.41	0.30	2.06
0.242	1.00	1.00	0.00	2.02

Table 1. Asymptotic efficieny and generalized contamination sensitivity for the optimal ψ's.

ϵ	bisquare	optimal
0.05	0.74	0.66
0.10	1.13	1.00
0.15	1.51	1.33
0.20	1.94	1.71
0.25	2.49	2.19
0.30	3.29	2.91
γ^{**}	3.10	2.78

Table 2. Maximum biases of bisquare and Hampel–optimal M–estimates