ON THE MAXIMUM BIAS FUNCTIONS OF *MM*-ESTIMATES AND CONSTRAINED *M*-ESTIMATES OF REGRESSION

By Jose R. Berrendero, Beatriz V.M. Mendes and David E. Tyler [1]

Universidad Autonoma de Madrid, Federal University of Rio de Janeiro and Rutgers University

We derive the maximum bias functions of the *MM*-estimates and the constrained *M*-estimates or *CM*-estimates of regression, and compare them to the maximum bias functions of the *S*-estimates and the τ-estimates of regression. In these comparisons, the *CM*-estimates tend to exhibit the most favorable bias-robustness properties. Also, under the gaussian model, it is shown how one can construct a *CM*-estimate which has a smaller maximum bias function than a given *S*-estimate, i.e. the resulting *CM*-estimate *dominates* the *S*-estimate in terms of maxbias, and at the same time is considerably more efficient.

1. Introduction. An important consideration for any estimate is an understanding of its robustness properties. Different measures exist which try to reflect the general concept known as robustness. One such measure is the maximum bias function, which measures the maximum possible bias of an estimate under ϵ contamination. In this paper we study

1

the maximum bias functions for the *MM*-estimates and the constrained *M*-estimates or *CM*-estimates of regression and compare them to the maximum bias functions for the *S*-estimates and the τ-estimates of regression.

The maximum bias function for Rousseeuw and Yohai's [10] *S*-estimates of regression were originally derived by Martin, Yohai and Zamar [7] under the assumption that the independent variables follow an elliptical distribution and that the intercept term is known. More recently, Berrendero and Zamar [1] derived the maximum bias functions for the *S*-estimates of regression under much broader conditions. Further general results on the maximum bias function can be found in [4]. The method used by [1] applies to a wide class of regression estimates. For example, it allows one to obtain the maximum bias functions of Yohai and Zamar's [12] τ-estimates of regression. Unfortunately, it does not apply to Yohai's [11] *MM*-estimates of regression, which are arguably the most popular high breakdown point estimates of regression. The MM-estimates, for example, are the default robust regression estimates in *S-plus*.

The original motivation for the current paper was thus to derive the maximum bias functions of the *MM*-estimates of regression and compare them to the maximum bias functions of the *S*-estimates and τ-estimates of regression. A lesser known high breakdown point estimate of regression, namely Mendes and Tyler's [8] constrained *M*-estimates of regression (or *CM*-estimates for short) has also been included in the study since their maximum bias functions can be readily obtained by applying the general method given by [1]. Expressions for the maximum bias functions of the *MM*-estimates and the *CM*-estimates are derived in sections 3 and 4. Comparisons between the *S*, τ, *MM*, and *CM*-estimates based on bi-weight score functions are given in Section 5. It turns out that in these comparisons, the *CM*-estimates tend to exhibit the most favorable robustness properties.

Consequently, a more detailed theoretical comparison between the maximum bias function of the *S*-estimates and the *CM*-estimates of regression, which helps explain the computational comparisons made in section 5, is given in section 6. In particular, under the gaussian model, it is shown how one can construct a *CM*-estimate of regression so that its maximum bias function *dominates* that of a given *S*-estimate of regression. That is, the maximum bias function of the *CM*-estimate is smaller for some level of contamination ϵ and is never larger for any value of ϵ. The *S*-estimate is thus said to be *bias-inadmissible* at the gaussian model.

Section 2 reviews the notion of the maximum bias function in the regression setting, as well as the definitions of the *S*-estimates, the *MM*-estimates and the *CM*-estimates for regression. Technical proofs are given in Section 7, an appendix.

2. The regression model and the concept of maximum bias. We follow the general setup given in [7]. Specifically, we consider the linear regression model

$$(2.1) \qquad \qquad y \; = \; \alpha_o + \mathbf{x}' \boldsymbol{\theta}_o + u,$$

where $y \in \mathbb{R}$ represents the response, $\mathbf{x} = (x_1, x_2, ..., x_p)' \in \mathbb{R}^p$ represents a random vector of explanatory variables, $\alpha_o \in \mathbb{R}$ and $\boldsymbol{\theta}_o \in \mathbb{R}^p$ are the true intercept and slope parameters respectively, and the random error term $u \in \mathbb{R}$ is assumed to be independent of $\mathbf{x}$. Let F_o and G_o represent the distribution functions of u and $\mathbf{x}$ respectively, and let H_o represent the corresponding joint distribution function of $(y, \mathbf{x})$. The following assumptions on the distribution H_o are assumed throughout the paper.

A1) F_o is absolutely continuous with density f_o which is symmetric, continuous and strictly decreasing on $\mathbb{R}^+$.

A2) $P_{G_o}(\mathbf{x}'\boldsymbol{\theta} = c) < 1$, for any $\boldsymbol{\theta} \in \mathbb{R}^p, \boldsymbol{\theta} \neq \mathbf{0}, c \in \mathbb{R}$.

As in [7] and [1], we focus on the estimation of the slope parameters $\boldsymbol{\theta}_o$. One reason for doing so, it that once given a good estimate of the slope parameters, the problem of estimating the intercept term and the residual scale reduces to the well studied univariate location and scale problem. Let $\mathbf{T}$ represent some $\mathbb{R}^p$ valued functional defined on $\mathcal{H}$, a space of distribution functions on $\mathbb{R}^{p+1}$ which includes some weak neighborhood of H_o, and such that $\mathbf{T}(H_o) = \boldsymbol{\theta}_o$. For large enough n, $\mathcal{H}$ almost surely contains the empirical distribution function H_n corresponding to a random sample $\{(\mathrm{y}_1, \mathbf{x}_1), ..., (\mathrm{y}_n, \mathbf{x}_n)\}$ from H_o. Furthermore, we assume that $\mathbf{T}$ is weakly continuous at H_o and so the statistic $\mathbf{T}_n = \mathbf{T}(H_n)$ is a consistent estimate of $\boldsymbol{\theta}_o$.

All functionals $\mathbf{T}$ considered in this paper are regression equivariant, as defined e.g. in [7]. For such functionals, a natural invariant measure of the "asymptotic" bias of $\mathbf{T}$ at H is given by

$$(2.2) \qquad b_{\Sigma_o}(\mathbf{T}, H) = \begin{cases} \left((\mathbf{T}(H) - \boldsymbol{\theta}_o)' \Sigma_o (\mathbf{T}(H) - \boldsymbol{\theta}_o)\right)^{1/2} & H \in \mathcal{H} \\ \\ \infty & H \notin \mathcal{H} . \end{cases}$$

Here, $\Sigma_o = \Sigma(G_o)$ is taken to an affine equivariate scatter matrix for the regressors $\mathbf{x}$ under G_o. We can thus presume without loss of generality that $(\alpha_o, \boldsymbol{\theta}_o) = \mathbf{0}$ and $\Sigma_\mathbf{o} = \mathbf{I}$. Hence, the asymptotic bias of $\mathbf{T}$ at H becomes the Euclidean norm of $\mathbf{T}$,

$$(2.3) \qquad b(\mathbf{T}, H) = \begin{cases} \| \mathbf{T}(H) \| & H \in \mathcal{H} \\ \\ \infty & H \notin \mathcal{H} , \end{cases}$$

where $\mathcal{H}$ is the class of distributions such that $\| \mathbf{T}(H) \| < \infty$. The maximum asymptotic bias of $\mathbf{T}$ over ϵ-contaminated neighborhoods V_ϵ of H_o, i.e., $V_\epsilon = \{H \mid H = (1-\epsilon)H_o + \epsilon H^*, \ H^* \in \mathcal{H}^*\}$ where $\mathcal{H}^*$ is the set of all distribution functions on $\mathbb{R}^{p+1}$, is defined to be

$$(2.4) \qquad B_\mathbf{T}(\epsilon) \ = \ \sup\{b(\mathbf{T}, H) \mid H \in V_\epsilon\} ,$$

and the asymptotic breakdown point is subsequently defined to be

$$(2.5) \qquad \epsilon^* \;=\; \inf\{\epsilon \mid B_{\mathbf{T}}(\epsilon) = \infty\}.$$

From an applied perspective, regardless of Σ_o, it may be of interest to derive upper bounds for the Euclidean distance between $\mathbf{T}(H)$ and $\boldsymbol{\theta}_o$, i.e. for $\|\mathbf{T}(H) - \boldsymbol{\theta}_o\|$. This measure is referred to as a *bias bound* by Berrendero and Zamar [1], wherein they use it for adjusting confidence intervals for $\boldsymbol{\theta}$ to include the possibility of bias introduced by a contaminated model. Note that the *bias bound* is regression and scale equivariant but not affine equivariant, and hence is not directly related to the maximum bias (2.4). In [1], some results are given for computing *bias bounds* taking the maximum bias function as a starting point.

2.1. *M-estimates with general scale.* The *S*, *MM* and *CM*-estimates of regression all lie within the class of *M*-estimates with general scale considered in [7]. An *M*-estimate, or more appropriately an *M*-functional, with general scale for the regression parameters α_o and $\boldsymbol{\theta}_o$, say $\mathrm{t}(H)$ and $\mathbf{T}(H)$ respectively, can be defined as the solution which minimizes

$$(2.6) \qquad \mathrm{E}_H \left[\rho \left(\frac{\mathrm{y} - \alpha - \mathbf{x}'\boldsymbol{\theta}}{\sigma(H)} \right) \right]$$

over all $\alpha \in \mathbb{R}$ and $\boldsymbol{\theta} \in \mathbb{R}^p$, where ρ is some nonnegative symmetric function and $\sigma(H)$ is some scale functional. The scale functional $\sigma(H)$ may be determined simultaneously or independently of $\{\mathrm{t}(H), \mathbf{T}(H)\}$. We assume throughout the paper that $\sigma(H)$ is regression invariant and residual scale equivariant, again as defined e.g. in [7]. Throughout, it is assumed that the function ρ satisfies the following conditions:

A3) (i) ρ is symmetric and nondecreasing on $[0, \infty)$ with $\rho(0) = 0$,

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

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

If the function ρ is also differentiable, then $(\mathrm{t}(H), \mathbf{T}(H))$ is a solution to the $p + 1$ simultaneous *M*-estimating equations

$$(2.7) \quad \mathrm{E}_H \left\{ \psi \left(\frac{\mathrm{y} - \alpha - \mathbf{x}'\boldsymbol{\theta}}{\sigma(H)} \right) \mathbf{x} \right\} = \mathbf{0}, \text{ and } \mathrm{E}_H \left\{ \psi \left(\frac{\mathrm{y} - \alpha - \mathbf{x}'\boldsymbol{\theta}}{\sigma(H)} \right) \right\} = 0,$$

where $\psi(u) \propto \rho'(u)$. By Condition A3(i), ψ is an odd function, nonnegative on $[0, \infty)$. Condition A3(ii) implies that these *M*-estimates are redescending, i.e. $\psi(u) \to 0$ as $u \to \infty$. A popular choice for *M*-estimates are Tukey's biweighted *M*-estimates, which correspond to choosing $\rho(u)$ to be

$$(2.8) \qquad\qquad \rho_T(u) = \begin{cases} 3u^2 - 3u^4 + u^6 & \text{for } |u| \leq 1 \\ 1 & \text{for } |u| > 1. \end{cases}$$

Note that this gives rise to the biweight ψ function $\psi_T(u) = u\{(1 - u^2)_+\}^2$.

The *S*-estimates for the intercept, slopes and scale are defined to be the solution $\{\mathrm{t}_s(H), \mathbf{T}_s(H), \sigma_s(H)\}$ to the problem of minimizing $\sigma \in \mathbb{R}^+$ subject to the constraint

$$(2.9) \qquad\qquad \mathrm{E}_H \left[\rho \left(\frac{\mathrm{y} - \alpha - \mathbf{x}'\boldsymbol{\theta}}{\sigma} \right) \right] \leq b$$

for some fixed value b, $0 < b < 1$. The breakdown point of the *S*-estimate of regression is $\epsilon^* = \min\{b, 1 - b\}$. A drawback to the *S*-estimates is that the tuning constant b not only determines the breakdown point but it also determines the efficiency of the estimate. To obtain a reasonable efficiency under a normal error model, one must usually substantially decrease the breakdown point.

This problem with tuning the *S*-estimates of regression motivated Yohai [11] to introduce the *MM*-estimates of regression, which can be tuned to have high efficiency under normal error while simultaneously maintaining a high breakdown point. Let ρ_1 and ρ_2 be a pair of loss functions satisfying A3, and with $\rho_1 > \rho_2$. Set $b = \mathrm{E}_{F_o}\rho_1(Y)$. *MM*-estimates are defined to be the solution $\{\mathrm{t}_{mm}(H), \mathbf{T}_{mm}(H)\}$ which minimizes

$$L_H(\alpha, \boldsymbol{\theta}) = \mathrm{E}_H \left[\rho_2 \left(\frac{\mathrm{y} - \alpha - \mathbf{x}'\boldsymbol{\theta}}{s(H)} \right) \right],$$

where $s(H) \doteq \sigma_s(H)$ is a preliminary *S*-functional of scale defined above based on $\rho = \rho_1$. The breakdown point of the *MM*-estimates only depends on ρ_1, and is given by $\epsilon^* = \min\{b, 1-b\}$. On the other hand, their asymptotic distribution is determined exclusively by ρ_2. This allows the *MM*-estimates to be tuned so that they possess both high breakdown point and high efficiency.

The *CM*-estimates are another class of regression estimates which can be tuned to have high efficiency at the normal model while maintaining a high breakdown point. The *CM*-estimates for the intercept, slopes and scale are defined to be the solution $(\mathrm{t}_{cm}(H), \mathbf{T}_{cm}(H), \sigma_{cm}(H))$ which minimizes

$$(2.10) \qquad \mathrm{L}_H(\alpha, \boldsymbol{\theta}, \sigma) \;=\; c\, \mathrm{E}_H \left[\rho \left(\frac{\mathrm{y} - \alpha - \mathbf{x}'\boldsymbol{\theta}}{\sigma} \right) \right] + \; \log \sigma$$

subject to the constraint (2.9), where $c > 0$ represents a tuning constant. As with the *S*-estimates of regression, the asymptotic breakdown point of the *CM*-estimates of regression is $\epsilon^* = \min\{b, 1-b\}$. Unlike the *S*-estimates of regression, though, the *CM*-estimates of regression can be tuned through the constant c in order to obtain a reasonably high efficiency without affecting the breakdown point.

We again emphasize that our focus here is on the slope functionals $\mathbf{T}(H)$ rather than on the intercept functionals $\mathrm{t}(H)$ or the scale functionals $\sigma(H)$. Given a good slope functional, one may wish to consider the wider range of location and scale functionals based on the distribution of $\mathrm{y} - \mathbf{x}'\mathbf{T}(H)$, such as its median and median absolute deviation, rather than those arising from an *S*, *MM* or *CM*-estimate of regression.

3. Maximum bias functions.

3.1. *Maximum bias functions for MM-estimates.* If $F_{H,\alpha,\boldsymbol{\theta}}$ is the distribution function of the absolute residuals $|\mathrm{y} - \alpha - \mathbf{x}'\boldsymbol{\theta}|$, then Berrendero and Zamar [1] give an expression for the maximum bias function for any estimate whose definition can be

expressed in the form

(3.11) $$(t(H), \mathbf{T}(H)) = \arg \min_{(\alpha, \boldsymbol{\theta})} J(F_{H, \alpha, \boldsymbol{\theta}}),$$

where $J(F)$ is a functional possessing certain monotonic properties. The *S*, τ, and *CM*-estimates are of this form. Application of their general results to the *S* and the τ-estimates are given [1]. Application of these results to the *CM*-estimates are presented in section 3.2.

The *MM*-estimates, however, cannot be expressed in the form (3.11) and so a different approach is needed in order to study its bias behavior. Let $B_{MM}(\epsilon)$ be the maximum bias function of an *MM*-estimate of regression. In this subsection, lower and upper bounds for $B_{MM}(\epsilon)$ are obtained under quite general conditions. In some important cases these two bounds are often equal and so allow for the determination of the maximum bias function exactly.

Let $\underline{s} = \inf_{H \in V_\epsilon} s(H)$, $\overline{s} = \sup_{H \in V_\epsilon} s(H)$, and

$$m(t, s) = \inf_{\|\boldsymbol{\theta}\| = t} \inf_{\alpha \in \mathbb{R}} E_{H_o} \rho_2 \left(\frac{\mathrm{y} - \alpha - \mathbf{x}'\boldsymbol{\theta}}{s} \right) - E_{H_o} \rho_2 \left(\frac{\mathrm{y}}{s} \right).$$

The following two functions play a key role in the developments below:

(3.12) $$h_1(t) = m(t, \overline{s}), \quad \text{and} \quad h_2(t) = \inf_{\underline{s} \le s \le \overline{s}} m(t, s).$$

THEOREM 3.1. *Let* $\mathbf{T}_{mm}$ *be an MM–estimate of the regression slopes with loss functions* ρ_i, $i = 1, 2$, *satisfying A3. Assume that the maximum bias function of the S-estimate with score function* ρ_1, $B_S(\epsilon)$, *satisfies* $B_S(\epsilon) < h_1^{-1}[\epsilon/(1 - \epsilon)]$. *Under A1 and A2, the maximum bias function of* $\mathbf{T}_{mm}$, $B_{MM}(\epsilon)$, *satisfies*

(3.13) $$h_1^{-1} \left(\frac{\epsilon}{1 - \epsilon} \right) \le B_{MM}(\epsilon) \le h_2^{-1} \left(\frac{\epsilon}{1 - \epsilon} \right).$$

Note that the condition $B_S(\epsilon) < h_1^{-1}[\epsilon/(1 - \epsilon)]$ of the above theorem together with (3.13) implies that $B_S(\epsilon) < B_{MM}(\epsilon)$. This condition usually holds for an appropriately

choosen ρ_1 function. Thus, an *MM*-estimate does not improve upon the maximum bias of the initial *S*-estimate. The trade-off though is that with an appropriately choosen ρ_2 function, the *MM*-estimate can greatly improve upon the efficiency of the initial *S*-estimate.

Upper and lower bounds for the maximum bias of MM-estimates have also been obtained respectively by Hennig [5], Theorem 3.1, and by Martin et. al. [7], Lemma 4.1, under the assumption of unimodal elliptically distributed regressors. For this special case, the upper bound given in (3.13) and in [5] agree. On the other hand, the lower bound given in [7], namely $B_{MM}(\epsilon) \geq h_0^{-1}[\epsilon/(1-\epsilon)]$, where $h_0(t) = \sup_{\underline{s} \leq s \leq \overline{s}} m(t,s)$ is not as tight as that given in (3.13).

In our setup, the assumption of unimodal elliptical regressors is equivalent to:

A2*) Under G_o, the distribution of $\mathbf{x}'\boldsymbol{\theta}$ is absolutely continuous, with a symmetric, unimodal density and depends on $\boldsymbol{\theta}$ only through $\|\boldsymbol{\theta}\|$ for all $\boldsymbol{\theta} \neq \mathbf{0}$.

Under this condition, we can define

$$(3.14) \qquad g(s,t) = \mathrm{E}_{H_o}\left[\rho\left(\frac{\mathbf{y} - \mathbf{x}'\boldsymbol{\theta}}{s}\right)\right],$$

where $\boldsymbol{\theta}$ is any vector such that $\|\boldsymbol{\theta}\| = t$. Under conditions A1, A2*, and A3, it is shown in Lemma 3.1 of Martin, Yohai and Zamar [7] that g is continuous, strictly increasing with respect to $\|\boldsymbol{\theta}\|$ and strictly decreasing in s for $s > 0$.

If A2* holds, then $\underline{s}$ and $\overline{s}$ are defined so that $g_1(\underline{s},0) = b/(1-\epsilon)$ and $g_1(\overline{s},0) = (b-\epsilon)/(1-\epsilon)$ respectively, and $m(t,s) = g_2(s,t) - g_2(s,0)$, where $g_i(s,t)$ is defined as in (3.14) after replacing ρ with ρ_i.

3.2. *Maximum bias curves for CM-estimates.* A *CM*-estimate of regression $\{t_{cm}(H), \mathbf{T}_{cm}(H)\}$ can be expressed in the form (3.11) with J taken to be

$$(3.15) \qquad J_{CM}(F) = \inf_{s \geq \sigma(F)} c\mathrm{E}_F[\rho(\mathrm{y}/s)] + \log s,$$

and where $\sigma(F)$ is the M-scale defined as the solution to the equation

$$(3.16) \qquad \mathrm{E}_F\left[\rho\left(\mathrm{y}/\sigma(F)\right)\right] = b.$$

Consequently, application of the general method in [1] for computing maximum bias functions leads to the following result.

THEOREM 3.2.   *Let $\mathbf{T}_{cm}$ be a CM-estimate of the regression slopes based on a function ρ satisfying A3, and suppose H_o satisfies A1 and A2. Define*

$$r_{cm}(\epsilon) = J_{CM}[(1-\epsilon)F_{H_0,0,\mathbf{0}} + \epsilon\delta_{\infty}],$$

and let

$$(3.17) \qquad m_{cm}(t) = \inf_{\|\boldsymbol{\theta}\|=t} \inf_{\alpha \in \mathbb{R}} J_{CM}[(1-\epsilon)F_{H_0,\alpha,\boldsymbol{\theta}} + \epsilon\delta_0].$$

Then, the maximum bias function of $\mathbf{T}_{cm}$, denoted by $B_{CM}(\epsilon)$, is given by

$$(3.18) \qquad B_{CM}(\epsilon) = m_{cm}^{-1}[r_{cm}(\epsilon)].$$

This general result can be given a simpler representation when condition A2* also holds. In particular, in the definition of $m_{cm}(t)$, the infimum is obtained when $\alpha = 0$ and $\boldsymbol{\theta}$ is any vector such that $\|\boldsymbol{\theta}\| = t$. This gives

$$m_{cm}(t) = \inf\{A_{c,\epsilon}(s,t) \mid s \geq m_s(t)\},$$

where $A_{c,\epsilon}(s,t) = c(1-\epsilon)g(s,t) + \log s$ and $m_s(t) = g_{(1)}^{-1}(b/(1-\epsilon),t)$, with $g(s,t)$ being defined in (3.14) and $g_{(1)}^{-1}(\cdot,t)$ being the inverse of g with respect to s. Also, it is easy to verify that

$$r_{cm}(\epsilon) = \inf\{A_{c,\epsilon}(s,0) \mid s \geq r_s(\epsilon)\} + c\epsilon,$$

where $r_s(\epsilon) = g_{(1)}^{-1}\left((b - \epsilon)/(1 - \epsilon), 0\right)$.

4. Maximum bias functions for two special cases. Maximum bias functions in general tend to have rather complicated expressions. At some model distributions though these expressions can be substantially simplified. This is possible for two special cases considered here, namely the gaussian and the cauchy models. These simplified expressions are useful for computing and comparing the maximum bias curves of various estimates for these models, which is done in section 5.

4.1. *Maximum bias functions under the gaussian model.* We assume thoughout this section not only that the error term but also that the regressor variables arise from a multivariate normal distribution. That is, we assume H_o has a joint $\mathrm{N}(\mathbf{0}, \mathbf{I}_{p+1})$ distribution, and refer to this as the gaussian model. Let $g(s) = \mathrm{E}_\Phi \rho(Z/s)$, where Z is a standard normal random variable, and define $\sigma_{b,\epsilon} \doteq g^{-1}[(b-\epsilon)/(1-\epsilon)]$ and $\gamma_{b,\epsilon} \doteq g^{-1}[b/(1-\epsilon)]$. Martin, Yohai and Zamar [7] show that the maximum bias function for an *S*-estimate of the regression slope under the gaussian model and based on a function ρ satisfying A3 is given by

$$(4.19) \qquad B_S(\epsilon) = \left[\left(\frac{\sigma_{b,\epsilon}}{\gamma_{b,\epsilon}}\right)^2 - 1\right]^{1/2}.$$

To obtain an expression for the maximum bias function of a *CM*-estimate of regression under the gaussian model, let

$$(4.20) \qquad A_{c,\epsilon}(s) = c(1 - \epsilon)g(s) + \log s,$$

Also, define $D_c(\epsilon) = \inf_{s \geq \sigma_{b,\epsilon}} A_{c,\epsilon}(s) - \inf_{s \geq \gamma_{b,\epsilon}} A_{c,\epsilon}(s)$. We then have the following result.

THEOREM 4.1.   *Let $\mathbf{T}_{cm}$ be a CM-estimate of the regression slopes based on a function ρ satisfying A3, and assume H_o is multivariate normal. It then holds that*

$$(4.21) \qquad B_{CM}(\epsilon) = \{\exp\left[2c\epsilon + 2D_c(\epsilon)\right] - 1\}^{1/2}.$$

Turning now to the *MM*-estimates, let $g_i(s) = \mathrm{E}_\Phi \rho_i(Z/s)$ for $i = 1, 2$, where Z is a standard normal random variable. Under the gaussian model,

$$m(t, s) = g_2\left(\frac{s}{(1 + t^2)^{1/2}}\right) - g_2(s).$$

Moreover, $\overline{s} = g_1^{-1}[(b - \epsilon)/(1 - \epsilon)]$, and $\underline{s} = g_1^{-1}[b/(1 - \epsilon)]$. Since ρ_1 is the same ρ-function used in defining the preliminary *S*-estimate, we have $\overline{s} = \sigma_{b,\epsilon}$ and $\underline{s} = \gamma_{b,\epsilon}$. Hence, $B_{MM}(\epsilon) \geq \ell(\epsilon)$, where

$$(4.22) \qquad \ell(\epsilon) = h_1^{-1}\left(\frac{\epsilon}{1 - \epsilon}\right) = \left[\left(\frac{\sigma_{b,\epsilon}}{g_2^{-1}\left[g_2(\sigma_{b,\epsilon}) + \epsilon/(1 - \epsilon)\right]}\right)^2 - 1\right]^{1/2}.$$

A simpler form for the upper bound which can be used for computational purposes can be obtained under some additional regularity conditions on $g_2(t)$. These conditions hold in most cases of interest.

A4) (i) $g(s)$ is continuously differentiable, and

(ii) $\phi(s) \doteq -sg'(s)$ is unimodal, with its maximum being obtained at σ_M. Set $K \doteq \phi(\sigma_M)$.

THEOREM 4.2.   *In addition to the assumptions of Theorem 3.1, suppose that $g_2(s)$ satisfies A4. Then, when H_o is multivariate normal,*

$$\ell(\epsilon) \leq B_{MM}(\epsilon) \leq \max\{\ell(\epsilon), u(\epsilon)\},$$

where $\ell(\epsilon)$ is given in (4.22), and

$$u(\epsilon) = \left[\left(\frac{\gamma_{b,\epsilon}}{g_2^{-1}\left[g_2(\gamma_{b,\epsilon}) + \epsilon/(1 - \epsilon)\right]}\right)^2 - 1\right]^{1/2}.$$

The upper bound in Theorem 4.2 coincides with that obtained by Hennig [5]. However, the tighter lower bound gives us further insight into the maximum bias and enables us to determine when the bounds are actually an equality. Obviously, when ϵ is such that $u(\epsilon) \leq \ell(\epsilon)$, then $B_{MM}(\epsilon) = \ell(\epsilon)$. This occurs in many important cases for a wide range of ϵ values.

As an example, consider the biweight loss function ρ_T defined by (2.8). If we choose $\rho_1(u) = \rho_T(u/k_1)$ and $\rho_2(u) = \rho_T(u/k_2)$ with tuning constants $k_1 = 1.56$ and $k_2 = 4.68$, and choose $b = 0.5$, then the resulting MM-estimate has a 50% breakdown point and is asymptotically 95% efficient under the gaussian model. For this case, it can be verified that the condition $B_S(\epsilon) < h_1^{-1}[\epsilon/(1 - \epsilon)]$ in Theorem 3.13 holds. From (4.22), it can be noted that this condition is equivalent to $g_2(\gamma_{b,\epsilon}) - g_2(\sigma_{b,\epsilon}) < \epsilon/(1 - \epsilon)$. It can also be verified that the corresponding ϕ_2 function is unimodal. A plot of ϕ_2 is displayed in the left hand graph of Figure 1. The bounds given in Theorem 4.2 for this MM-estimate are displayed in the right hand graph of Figure 1. Both bounds coincide, and therefore the exact maximum bias function is known for, roughly, $\epsilon \leq 0.33$.
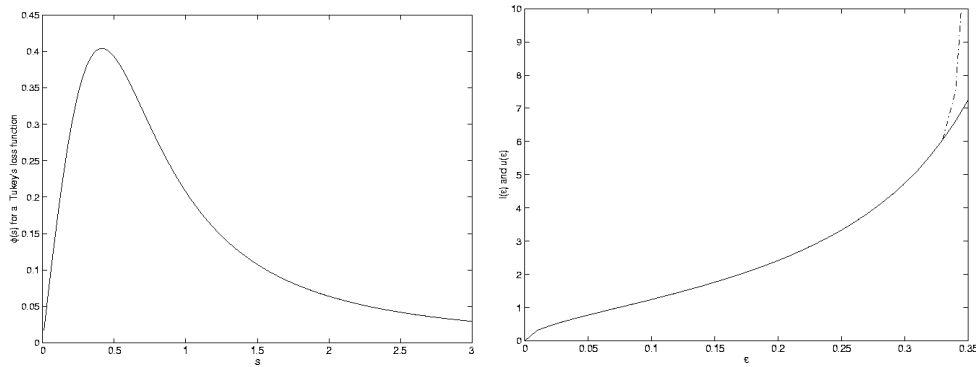

FIG. 1.   *The graph on the left represents the function $\phi(s)$ for a biweight ρ function. The graph on the right gives the maximum bias bounds ($\ell(\epsilon)$, solid line; $u(\epsilon)$ dotted-dashed line) for an MM-estimate based on biweight loss functions with 50% breakdown point and 95% efficiency under the gaussian model.*

4.2. *Maximum bias functions under the cauchy model.* We now assume that the error term and the regressors follow independent cauchy distributions rather than normal distributions. That is, we assume $\mathbf{x}_1, \ldots, \mathbf{x}_n$ and y have independent standard cauchy distributions, and refer to this as the cauchy model. Note that in this case, the distribution of the regressors is not elliptically symmetric. The derivations for the cauchy model follow closely those given for the gaussian model.

Let $g(s) = \mathrm{E}_\Phi \rho(Z/s)$, where Z is now a standard cauchy random variable, and again let $\sigma_{b,\epsilon} \doteq g^{-1}[(b-\epsilon)/(1-\epsilon)]$ and $\gamma_{b,\epsilon} \doteq g^{-1}[b/(1-\epsilon)]$. In the appendix, we show the maximum bias function for an *S*-estimate of regression to be

$$(4.23) \qquad B_S(\epsilon) = \frac{\sigma_{b,\epsilon}}{\gamma_{b,\epsilon}} - 1,$$

and for a *CM*-estimates of regression to be

$$(4.24) \qquad B_{cm}(\epsilon) = \exp\{D_c(\epsilon) + c\epsilon\} - 1,$$

with $D_c(\epsilon)$ being analogous to its definition given after equation (4.20). Upper and lower bound for the maximum bias function for the *MM*-estimates of regression are shown in the appendix to be

$$(4.25) \qquad \ell(\epsilon) \le B_{MM}(\epsilon) \le \max\{\ell(\epsilon), u(\epsilon)\}, \quad \text{where}$$

$$\ell(\epsilon) = \frac{\sigma_{b,\epsilon}}{g_2^{-1}\left[g_2(\sigma_{b,\epsilon}) + \epsilon/(1-\epsilon)\right]} - 1, \text{ and } u(\epsilon) = \frac{\gamma_{b,\epsilon}}{g_2^{-1}\left[g_2(\gamma_{b,\epsilon}) + \epsilon/(1-\epsilon)\right]} - 1.$$

The conditions given in (4.19), Theorem 4.1 and Theorem 4.2 for the gaussian model are also being assumed here for (4.23), (4.24), and (4.25) respectively for the cauchy model. For an *MM*-estimate of regression, condition A4 can again be shown to hold when using a biweight loss function.

It is somewhat surprising that the expressions for $B_S(\epsilon)$, $B_{CM}(\epsilon)$ and $B_{MM}(\epsilon)$ are of order $o(\epsilon)$ as $\epsilon \to 0$ under the cauchy model in contrast with the usual $\sqrt{\epsilon}$ order. This

is not a contradiction, however, of known results which establish general $\sqrt{\epsilon}$ order for the maximum bias functions of regression estimates based on residuals since such results require either elliptical regressors, as in Yohai and Zamar [13], or the existence of second moments for the regressors, as in He [3] or Yohai and Zamar [14].

5. Maximum bias curve comparisons.

5.1. *Comparisons at the gaussian model.*   Most estimators need to be tuned so that they perform reasonably well at some important model, as well as being robust to deviations from the model. In practice, one oftens tunes an estimate so that it has good efficiency at the gaussian model as well as a high breakdown point. For smooth ρ-functions, both the *MM* and *CM*-estimates of regression can be tuned to have a 50% breakdown point and 95% asymptotic relative efficiency at the gaussian model. This is also true for the class of τ-estimates, see Yohai and Zamar [12] for the details. Thus, these estimates cannot be ranked on the basis of their efficiency and breakdown point alone. Comparing their maximum bias behavior at the gaussian model gives further insight into how these estimates are affected by deviations from the model.

Here, we again consider the estimates associated with the family of Tukey's biweight loss function (2.8). The 95% efficient biweight *MM*-estimate with a 50% breakdown point has been discussed in the previous subsection. A 95% efficient biweight *CM*-estimate with a 50% breakdown point, is obtained by choosing $\rho(u) \doteq \rho_T(u)$, $b = 0.5$, and the tuning constant $c = 4.835$, see [8] for details. In contrast, a 95% efficent biweight *S*-estimate of regression has a 12% breakdown point, whereas a biweight *S*-estimate with a 50% breakdown point is only 28.7% efficient at the gaussian model.

Figure 2, represents the maximum bias functions at the gaussian model of the *MM*-, *CM*- and τ-estimates based on biweight functions, and tuned so that they have 95%
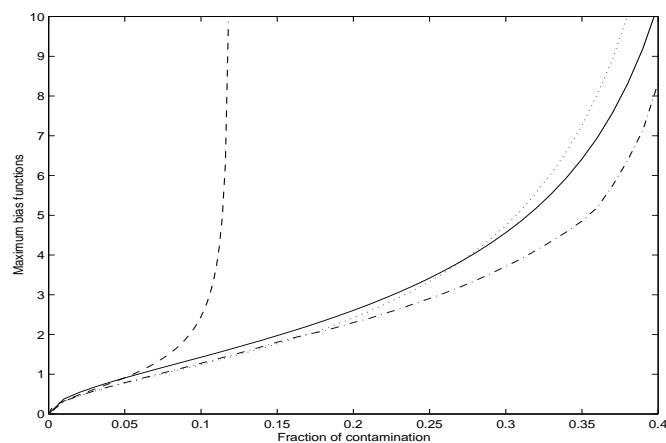
FIG. 2. *Maximum bias functions for a biweight S-estimate (dashed line) MM-estimate (dotted line, lower bound), τ-estimate (solid line) and CM-estimate (dashed-dotted line). All the estimates have 95% efficiency under the gaussian model. The S-estimate has a breakdown point of 12%, whereas the others have a 50% breakdown point.*

(asymptotic) efficiency under the gaussian model and a 50% breakdown point, as well as that of the 95% efficient biweight *S*-estimate. We observe that up to $\epsilon \approx 0.28$, the τ-estimate has a larger bias than the *MM*-estimate, and then a smaller bias afterwards. The τ-estimate, though, has a larger bias than the *CM*-estimate over essentially the entire range of ϵ. Up to $\epsilon \approx 0.20$, *MM*- and *CM*-estimates are roughly equivalent, although for larger fractions of contamination the *CM*-estimate is clearly better.

As a further comparison, Figure 3 again shows the maximum bias function at the gaussian model of the above 95% efficient biweight *MM* and *CM*-estimates, as well as the less efficient 50% breakdown point biweight *S*-estimate. Also, included in Figure 3 is the biweight *CM*-estimate having a 50% breakdown point and an asymptotic relative efficiency of 61.1% at the gaussian model, which corresponds to choosing the tuning constant $c = 2.568$. (The efficiency of the *CM*-estimate based on a biweight function with $b = 1/2$ and $c = 2.568$ under the gaussian model is incorrectly reported as 28.7%
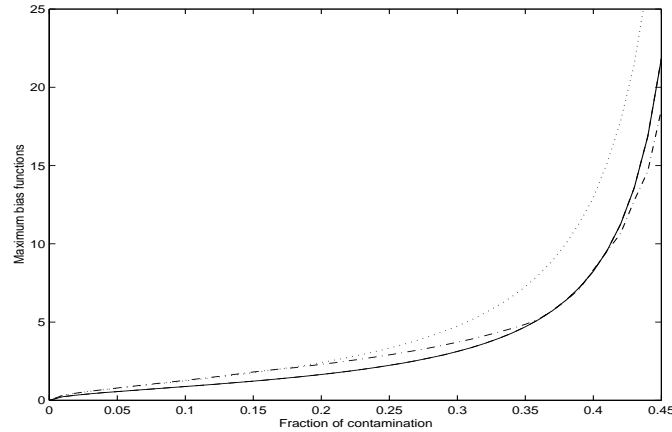
FIG. 3. *Maximum bias functions for a biweight S-estimate (solid line), MM-estimate (dotted line, lower bound), and two CM-estimates (dotted-dashed line and solid line). The plot for the S-estimate and the second CM-estimate are almost identical. All estimates have a 50% breakdown point. The MM-estimate and the first CM-estimates (dotted-dashed line) has 95% efficiency under the gaussian model. The second CM-estimate (solid line) has an efficiency of 61.1%, whereas the efficiency of the S-estimate is 28.7%.*

rather than 61.1% in Table 1 of Mendes and Tyler [8]. The rest of Table 1 of [8] is correct).

The maximum bias of the 95% efficient *MM*-estimate is uniformly larger than that of the corresponding *S*-estimate. This is consistent with the general result given in Theorem 3.1. The increase in bias for the *MM*-estimate is compensated by its increase in efficiency. A curious observation, though, is that for large fractions of contamination the maximum bias of the 95% efficient *CM*-estimate is lower than that of the 28.7% efficient *S*-estimate. Furthermore, the maximum bias of the 61.1% efficient *CM*-estimate is almost identical to, and as shown theoretically in the next section, is never larger than that of the 28.7% efficient *S*-estimate. That is, there is no trade-off between increase efficiency and maximum bias for this *CM*-estimate relative to the *S*-estimate. In practice, given that the maximum bias function of the 95% efficient *CM*-estimate does not greatly

differ from that of the 61.1% estimate, the 95% efficient estimate would be preferable.

5.2. *Comparisons at the cauchy model.*   We consider now the maximum bias behavior of *S*, *MM* and *CM*-estimates at the cauchy model. Figure 4 shows the maximum bias function at the cauchy model for the *MM*, and *CM*-estimates which are 95% efficient at the gaussian model as well as the 28.7% efficient biweight *S*-estimate and the 61.1% efficient *CM*-estimate discussed in section 5.1. The breakdown point of each of these estimates remains 50% under the cauchy model. The estimates though are not re-tuned here for the cauchy model. Rather, our goal is to make further comparisons among the same estimates. In practice, given a specific estimate, one would wish to evaluate its robustness properties under various scenarios. From Figure 4, it can be noted that the 95% efficient *CM*-estimate tends to have the better maximum bias behavior at the cauchy model, even better than that of the 61.1% efficient *CM*-estimate.

5.3. *Other considerations.*   Aside from maximum bias functions, a classical way of evaluating the robustness of an estimate as it deviates from normality is to consider its efficiency at other distributions. The asymptotic efficiencies at the gaussian model discussed in section 5.1 depend on the distribution of the error term being normal. They do not however depend on the distribution of the carrier being normal, only that they possess second moments. This is also true for the asymptotic efficiencies at other symmetric error distributions, see e.g. Maronna, Bustos and Yohai [6]. In particular, they note that the asymptotic variance-covariance matrix of $\widehat{\boldsymbol{\theta}} = \mathbf{T}_n$ has the form $\sigma_u^2 \, \boldsymbol{\Sigma}_\mathbf{x}$, where $\boldsymbol{\Sigma}_\mathbf{x}$ is the variance-covariance matrix of the carriers $\mathbf{x}$ and σ_u depends only on the distribution of the error term u.
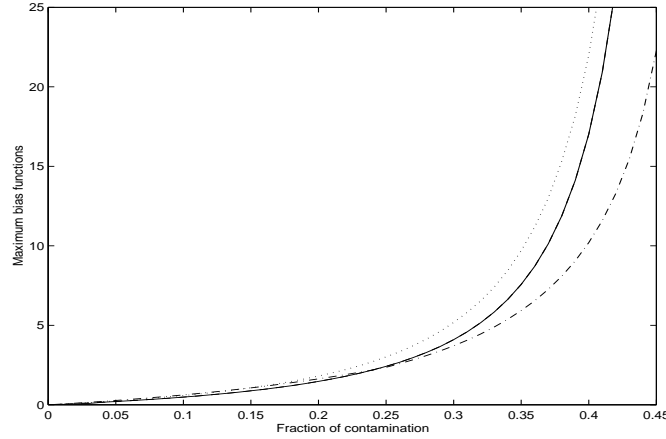
FIG. 4. *Maximum bias functions for a biweight S-estimate (solid line), an MM-estimate (dotted line, lower bound), and two CM-estimates (dotted-dashed line and solid line). The plot for the S-estimate and the second CM-estimate are almost identical. All estimates have a 50% breakdown point under the cauchy model. The MM-estimate and the first CM-estimate have 95% efficiency at the gaussian model, whereas the second CM-estimate and the S-estimate have efficiencies of 61.1% and 28.7% respectively at the gaussian model.*

In Table 1, we again consider the 95% efficient biweight *S*, *MM*, and *CM*-estimates, the 28.7% efficient biweight *S*-estimate and the 61.1% efficient *CM*-estimate discussed in section 5.1, where the efficiency is take under a normal error model. These estimates are labeled *S95, MM95, CM95, S28* and *CM61* respectively. For these estimates, we compute their asymptotic variances σ_u^2 (AVAR) under a variety of symmetric error models. Besides the standard normal (NORM), these models include the slash (SL), the cauchy (CAU), the t_3-distribution (T3), the double exponential (DE), a 90-10% mixture of a standard normal and a normal with mean zero and variance 9 (CN), and the uniform distribution on $(-1, 1)$ (UNIF). Each of these distributions are normalized so that their interquartile ranges are all equal to that of the standard normal, namely 1.3490. This corresponds to multiplying the SL, CAU, T3, DE, CN or UNIF random variable by $0.4587, 0.6745, 0.8818, 0.9731, 0.9248$ and 1.3490 respectively. Also included in Table

1 are the residual gross error sensitivities (RGES), see Hampel, et. al. [2]. Formulas for AVAR and RGES can be found in [8].

		NORM	SL	CAU	T3	DE	CN	UNIF
S95	AVAR	1.053	1.798	2.209	1.257	1.429	1.091	0.771
	RGES	1.770	3.277	3.716	2.146	2.258	1.942	1.415
MM95	AVAR	1.053	1.230	1.312	1.221	1.368	1.087	0.713
	RGES	1.770	2.146	2.243	1.953	2.038	1.844	1.548
CM95	AVAR	1.053	1.159	1.202	1.227	1.396	1.088	0.755
	RGES	1.770	1.995	2.061	1.988	2.138	1.835	1.439
CM61	AVAR	1.637	1.330	1.059	2.091	1.528	2.891	1.128
	RGES	1.838	1.900	1.765	2.285	2.045	2.619	1.405
S28	AVAR	3.484	1.330	1.059	2.091	1.528	2.891	120.336
	RGES	2.850	1.900	1.765	2.285	2.045	2.619	15.621

TABLE 1

Asymptotic variances and residual gross error sensitivities of some S, MM, and CM estimates of regression under symmetric error distributions.

From Table 1, it can be noted that the estimates *MM95* and *CM95* behave similarly with respect to asymtotic variance and residual gross error sensitivity, with *CM95* being slightly better at the longer tailed slash and cauchy distributions and the *MM95* being slightly better at the more moderate t_3 and double exponential distributions. Both *MM95* and *CM95* perform better than *S95* at longer tailed distributions. The behavior of *S28* and *CM61* are the same except at the normal and uniform distributions. At longer tailed distributions, equality tends to hold for the constrain (2.9) on *CM61* and so as an estimate it is asymptotically equivalent to *S28* at these distributions. At the normal and the uniform

distributions, there is a considerable difference in favor of *CM61*. Curiously, the behavior of *S28* and *CM61* at the cauchy distribution is better than that of *MM95* and *CM95*. However, based on the overall behavior of the asymptotic variances and residual gross error sensitivities alone, either *MM95* and *CM95* would be preferable in practice.

6. Bias-inadmissibility of S-estimates at the gaussian model. Throughout this section, we assume the gaussian model. In section 5.1, it was noted that under the gaussian model the maximum bias function of the 61.1% efficient biweight *CM*-estimate is never smaller than that of the 27.78% efficient biweight *S*-estimate. In this section, we verify this result theoretically rather than computationally. Moveover, we note this result is not specific to the use of the biweight estimates. In general, we show that for a given *S*-estimate, it is usually possible to tune the corresponding *CM*-estimates (through the value of c) so that $B_{CM}(\epsilon) \leq B_S(\epsilon)$ for all ϵ, and with strict inequality for at least one value of ϵ. In such a case, we will say that, with respect to the maximum bias criterion, the estimate T_S is *inadmissible* at the gaussian model since it can be *dominated* by T_{CM}.

To show this, we need to compare carefully the maximum bias functions of the *CM*-estimates and the *S*-estimates. An alternative representation for $B_{CM}(\epsilon)$ in terms of $B_S(\epsilon)$ at the normal model [see equations (4.21) and (4.19)] is given by

$$(6.26) \qquad \log[1 + B_{CM}^2(\epsilon)] = \log[1 + B_S^2(\epsilon)] + 2d_c(\epsilon),$$

where $d_c(\epsilon) = h_c(\epsilon, \gamma_{b,\epsilon}) - h_c(\epsilon, \sigma_{b,\epsilon})$ and

$$(6.27) \qquad h_c(\epsilon, \sigma) = A_{c,\epsilon}(\sigma) - \inf_{s \geq \sigma} A_{c,\epsilon}(s),$$

The functionals T_{CM} and T_S in (6.26) are understood to be defined by using the same ρ and the same value of b. From representation (6.26), we see that what we need to consider is the sign of $d_c(\epsilon)$ in terms of c and ϵ. The following result represents a first step in

determining appropriate values of the tuning constant c necessary for showing the bias inadmissibility of an *S*-estimate. The value of K below is defined within Condition A4.

THEOREM 6.1.   *Suppose that ρ is such that conditions A3 and A4 hold.*

(i) If $c \leq 1/K$, then $B_{CM}(\epsilon) = B_S(\epsilon)$ for all ϵ.

(ii) For any ϵ such that $c > c(\epsilon) \doteq \epsilon^{-1} \log(\sigma_{b,\epsilon}/\gamma_{b,\epsilon})$, it holds $B_{CM}(\epsilon) > B_S(\epsilon)$.

As a consequence, for the *CM*-estimate to improve upon the maximum bias function of the *S*-estimate, one needs to choose $c > 1/K$. On the other hand, if $c_o \doteq \inf\{c(\epsilon) : 0 < \epsilon < b\}$, then we also need to choose $c \leq c_o$. This range is not empty, since as shown in the appendix,

$$(6.28) \qquad\qquad 1/K < (1-b)/K + b/\phi(\gamma_{b,0}) \leq c_o,$$

where $\phi(s)$ is defined within Condition A4.

For $c \leq 1/K$, the *CM*-functional is the same as the *S*-functional at H_o as well as at any H in an ϵ-contaminated neighborhood of H_o. This is because equality is obtained in the constraint (2.9) for the *CM*-estimate, and when equality is obtained the *CM*-estimate gives the same solution as the corresponding *S*-estimate. Thus, for $c \leq 1/K$, the *CM*-estimate has the same maximum bias function as the corresponding *S*-estimate. On the other hand, for large values of c, the *CM*-estimate tends to give a solution similar to the least squares solution, and so one expects the maximum bias function to be unacceptably large even though the breakdown point may be close to 1/2. In fact, one can note from (4.21) that for any ϵ, $B_{CM}(\epsilon) \to \infty$ as $c \to \infty$.

Varying the tuning constant c may decrease the maximum bias for some values of ϵ, while increasing the maximum bias for other values of ϵ. The question we address now is whether it is possible to find a moderate value of c (necessarily between $1/K$ and c_o)

such that the maximum bias function of the *CM*-estimate improves upon the maximum bias function of the *S*-estimate.

The following result shows that, in most cases of interest, the condition $c \leq c_o$ is not only necessary but also sufficient to obtain $B_{CM}(\epsilon) \leq B_S(\epsilon)$ for all ϵ. The value of σ_M below is also defined within Condition A4.

THEOREM 6.2. *Suppose that the assumptions of Theorem 6.1 hold. If $c \leq c_o$ and $g(\sigma_M) \leq b$, then $B_{CM}(\epsilon) \leq B_S(\epsilon)$ for all $\epsilon > 0$.*

REMARK 6.1. *This result cannot be improved upon. That is, if $c > c(\epsilon)$, then $B_S(\epsilon) < B_{CM}(\epsilon)$ by Theorem 6.1. Also, if $c \leq c_o$ and $g(\sigma_M) > b$, then either $B_S(\epsilon) < B_{CM}(\epsilon)$ for some ϵ or $B_S(\epsilon) = B_{CM}(\epsilon)$ for all ϵ. This remark is verified in the appendix.*

In order to show that an *S*-estimate can be *dominated* by a *CM*-estimate with c chosen so that $1/K < c \leq c_o$, it remains to be shown that for some $0 < \epsilon < b$, $B_{CM}(\epsilon) < B_S(\epsilon)$. For specific examples, this can be checked numerically. Under additional assumptions, though, this can be shown analytically.

THEOREM 6.3. *Suppose that the assumptions of Theorem 6.2 hold. Furthermore, suppose that $g(s)$ is convex, and*

$$(6.29) \qquad \phi(\sigma_{b,0}) \geq \frac{[1 - g(\sigma_M)]^2(1 - b)}{2 - [b + g(\sigma_M)]}.$$

Then, for any value c such that

$$c_1 \doteq \frac{\log(\sigma_M/\sigma_{b,0})}{b - g(\sigma_M)} < c \leq \frac{1}{\phi(\sigma_{b,0})} = c_o,$$

*the CM-estimate of regression **dominates** the S-estimate of regression with respect to the maximum bias function. Furthermore, this range of values for c is not empty.*

REMARK 6.2. *From the proof of Theorem 6.3 it follows that a condition more general than (6.29) under which the conclusions also hold is $c_o = \lim_{\epsilon \to 0^+} c(\epsilon)$. However, (6.29) is easier to check and holds in most cases of interest.*

Consider the biweight *S*-estimate with breakdown point $b \leq 1/2$. It can be verified that the conditions of Theorem 6.3 holds whenever $b > 0.410$, and so any such biweight *S*-estimate is inadmissible with respect to maximum bias at the gaussian model. For $b = 1/2$, i.e. the 27.78% efficient biweight *S*-estimate, the value of $c = 2.568$ falls within the interval given in Theorem 6.3. Hence, the 61.1% efficient biweight *CM*-estimate *dominates* the 27.78% efficient biweight *S*-estimate with respect to maximum bias at the gaussian model. As noted in section 5.1, although the decrease in maximum bias is negligible, the increase in efficiency is not.

As another example, consider the α-quantile regression estimates. These correspond to *S*-estimates with $\rho(u) = I\{|u| \geq 1\}$ and $b = 1 - \alpha$. It is straightforward to to verify that the conditions of Theorem 6.3 hold in this case whenever $b > 0.3173$, and so the α-quantile regression estimates with $\alpha < 0.6837$ are inadmissible at the gaussian model with respect to maximum bias. Again the decrease in maxbias is not large. For example, for the special case $\alpha = b = 0.5$, for which the resulting α-quantile estimate corresponds to Rousseeuw's [9] least median of squares estimate (*LMS*), the best improvement is only 95.7% of the *LMS* bias.

The α-quantile estimates are often referred to as minimax bias regression estimates. Martin, Yohai and Zamar [7] show that within the class of *M*-estimates of regression with general scale, an α-quantile estimate minimizes the maximum bias at ϵ, with the value of α depending on ϵ. Yohai and Zamar [13] generalize this minimax result to the class of all residual admissible estimates of regression. At the gaussian model, an α-quantile estimate can be shown to have minimax bias for some ϵ whenever $0.500 < \alpha < 0.6837$,

or equivalently when $0.3173 < b < 0.500$. Despite having minimax bias at the gaussian model for a given ϵ, these α-quantile regression estimates are still inadmissible at the gaussian model with respect to maximum bias. In particular, a *CM*-estimate can be constructed which also has minimax bias at the given ϵ, never larger bias at any other ϵ, and smaller bias for some other ϵ. Although the decrease in the maximum bias may not be of practical importance, these observations expose some limitations of the notion of minimax bias.

The minmax bias results given in [13] for the α-quantile regression estimates apply more generally than to just the gaussian model. They also apply to models having a symmetric unimodal error term along with elliptically distributed carriers. At such models, though, we conjecture that the α-quantile regression estimates may again be inadmissible with respect to maximum bias, but we do not pursue this topic further here. The value of α which attains the minimum maxbias at a particular ϵ is not only dependent on the value of ϵ but also dependent on the particular model. That is, a particular α-quantile estimate is not necessarily minimax at ϵ over a range of models but is only known to be minimax at ϵ at a specific model. Any estimate which can be shown to *dominate* an α-quantile estimate would most likely need to be model specific.

7. Appendix. In this section we include the proof of the results and other technical questions.

Proof of Theorem 3.1: It can be shown following the proofs of Lemmas 4, 5 and 6 in [1] that, for all $s > 0$ and $t \in \mathbb{R}$, there exist $\alpha_t \in \mathbb{R}$ and $\boldsymbol{\theta}_t \in \mathbb{R}^p$ such that

$$m(t,s) = \mathrm{E}_{H_o} \rho_2 \left(\frac{y - \alpha_t - \mathbf{x}' \boldsymbol{\theta}_t}{s} \right) - \mathrm{E}_{H_o} \rho_2 \left(\frac{y}{s} \right).$$

Also, we can show that $m(t,s)$ is a strictly increasing function of t, for all $s > 0$. It follows that $h_1(t)$ is also strictly increasing.

We show first that $B_{MM}(\epsilon) \leq t_2$, where t_2 is such that $h_2(t_2) = \epsilon/(1 - \epsilon)$. Let $\tilde{\boldsymbol{\theta}} \in \mathbb{R}^p$ be such that $\tilde{t} = \|\tilde{\boldsymbol{\theta}}\| > t_2$. We shall prove that

$$(7.30) \quad \mathrm{E}_H \rho_2 \left(\frac{y - \alpha - \mathbf{x}'\tilde{\boldsymbol{\theta}}}{s(H)} \right) > \mathrm{E}_H \rho_2 \left(\frac{y}{s(H)} \right), \quad \text{for each } \alpha \in \mathbb{R} \text{ and } H \in V\epsilon.$$

Let $H = (1 - \epsilon)H_o + \epsilon\tilde{H}$. We have that:

$$m[\tilde{t}, s(H)] > m[t_2, s(H)] \geq \inf_{\underline{s} \leq s \leq \overline{s}} m(t_2, s) = h_2(t_2) = \frac{\epsilon}{1 - \epsilon}.$$

Therefore, for each $\alpha \in \mathbb{R}$ and $H \in V\epsilon$,

$$\mathrm{E}_{H_o}\rho_2 \left(\frac{y - \alpha - \mathbf{x}'\tilde{\boldsymbol{\theta}}}{s(H)} \right) - \mathrm{E}_{H_o}\rho_2 \left(\frac{y}{s(H)} \right) > \frac{\epsilon}{1 - \epsilon},$$

that is,

$$(1 - \epsilon)\mathrm{E}_{H_o}\rho_2 \left(\frac{y - \alpha - \mathbf{x}'\tilde{\boldsymbol{\theta}}}{s(H)} \right) > (1 - \epsilon)\mathrm{E}_{H_o}\rho_2 \left(\frac{y}{s(H)} \right) + \epsilon.$$

It follows that, for every $\alpha \in \mathbb{R}$ and $H \in V_\epsilon$,

$$\mathrm{E}_H\rho_2 \left(\frac{y - \alpha - \mathbf{x}'\tilde{\boldsymbol{\theta}}}{s(H)} \right) \geq (1 - \epsilon)\mathrm{E}_{H_o}\rho_2 \left(\frac{y - \alpha - \mathbf{x}'\tilde{\boldsymbol{\theta}}}{s(H)} \right)$$

$$> (1 - \epsilon)\mathrm{E}_{H_o}\rho_2 \left(\frac{y}{s(H)} \right) + \epsilon \geq \mathrm{E}_H\rho_2 \left(\frac{y}{s(H)} \right),$$

that is, inequality (7.30) holds. The last inequality above follows from A3(ii).

Next, we show that $B_{MM}(\epsilon) \geq t_1$, where t_1 is such that $h_1(t_1) = \epsilon/(1 - \epsilon)$. Since $B_S(\epsilon) < t_1$, we can select an arbitrary $t > 0$ such that $B_S(\epsilon) < t < t_1$. It is enough to show that $B_{MM}(\epsilon) \geq t$. We know that there exist $\alpha_t \in \mathbb{R}$ and $\boldsymbol{\theta}_t \in \mathbb{R}^p$ such that

$$h_1(t) = m(t, \overline{s}) = \mathrm{E}_{H_o}\rho_2 \left(\frac{y - \alpha_t - \mathbf{x}'\boldsymbol{\theta}_t}{\overline{s}} \right) - \mathrm{E}_{H_o}\rho_2 \left(\frac{y}{\overline{s}} \right).$$

Since h_1 is strictly increasing, $h_1(t) < h_1(t_1) = \epsilon/(1 - \epsilon)$. It follows that

$$(7.31) \qquad (1 - \epsilon)\mathrm{E}_{H_o}\rho_2 \left(\frac{y - \alpha_t - \mathbf{x}'\theta_t}{\overline{s}} \right) < (1 - \epsilon)\mathrm{E}_{H_o}\rho_2 \left(\frac{y}{\overline{s}} \right) + \epsilon.$$

Define the following sequence of contaminating distributions: $\tilde{H}_n = \delta_{(y_n, \mathbf{x}_n)}$ where $\mathbf{x}_n = n\boldsymbol{\theta}_t$ and $y_n = \alpha_t + \mathbf{x}'_n \boldsymbol{\theta}_t = \alpha_t + nt^2$. Let $H_n = (1 - \epsilon)H_o + \epsilon\tilde{H}_n$ and $\boldsymbol{\theta}_n = \mathbf{T}(H_n)$.

Suppose that $\sup_n \|\boldsymbol{\theta}_n\| < t$ in order to find a contradiction. Under this assumption, there exists a convergent subsequence, denoted also by $\{\boldsymbol{\theta}_n\}$, such that $\lim_{n\to\infty} \boldsymbol{\theta}_n = \tilde{\boldsymbol{\theta}}$, where $\|\tilde{\boldsymbol{\theta}}\| = \tilde{t} < t$. Assume for a moment that the sequence of intercept functionals evaluated at H_n, $\alpha_n = t(H_n)$, satisfies $\lim_{n\to\infty} |\alpha_n| = \infty$. Then,

$$\lim_{n\to\infty} \mathrm{E}_{H_n}\rho_2\left(\frac{y - \alpha_n - \mathbf{x}'\boldsymbol{\theta}_n}{s(H_n)}\right) = (1 - \epsilon) + \epsilon \lim_{n\to\infty} \rho_2\left(\frac{y_n - \alpha_n - \mathbf{x}'_n\boldsymbol{\theta}_n}{s(H_n)}\right)$$
$$> (1 - \epsilon) \lim_{n\to\infty} \mathrm{E}_{H_o}\rho_2\left(\frac{y - \alpha_t - \mathbf{x}'\boldsymbol{\theta}_t}{s(H_n)}\right) = \lim_{n\to\infty} \mathrm{E}_{H_n}\rho_2\left(\frac{y - \alpha_t - \mathbf{x}'\boldsymbol{\theta}_t}{s(H_n)}\right),$$

but this fact contradicts the definition of $(\alpha_n, \boldsymbol{\theta}_n)$. Notice that $0 < \underline{s} < s(H_n) < \overline{s} < \infty$ implies that $\lim_{n\to\infty} \mathrm{E}_{H_o}\rho_2[(y - \alpha_t - \mathbf{x}'\boldsymbol{\theta}_t)/s(H_n)] < 1$ which in turn implies the strict inequality above. Therefore, we can assume without loss of generality that $\lim_{n\to\infty} \alpha_n = \tilde{\alpha}$, for some finite $\tilde{\alpha} \in \mathrm{I\!R}$. As a consequence, we have that

$$(7.32) \quad \lim_{n\to\infty} \left|\frac{y_n - \alpha_n - \mathbf{x}'_n\boldsymbol{\theta}_n}{s(H_n)}\right| = \infty, \quad \text{and} \quad \left|\frac{y_n - \alpha_t - \mathbf{x}'_n\boldsymbol{\theta}_t}{s(H_n)}\right| = 0, \quad \text{for each } n.$$

We prove now that $\lim_{n\to\infty} s(H_n) = \overline{s}$, for any convergent subsequence $s(H_n)$. Let $s_\infty = \lim_{n\to\infty} s(H_n)$. Notice that $\overline{s}$ satisfies the equation

$$(7.33) \qquad\qquad (1 - \epsilon)\mathrm{E}_{H_o}\rho_1(y/\overline{s}) + \epsilon = b.$$

Let $(\gamma_n, \boldsymbol{\beta}_n) = (t_1(H_n), \mathbf{T}_1(H_n))$ be the regression S–estimate based on ρ_1. We know that $\|\boldsymbol{\beta}_n\| \leq B_S(\epsilon) < t$, for all n, so that without loss of generality $\lim_{n\to\infty} \boldsymbol{\beta}_n = \tilde{\boldsymbol{\beta}}$, where $\|\tilde{\boldsymbol{\beta}}\| < t$. Assume that $\lim_{n\to\infty} |\gamma_n| = \infty$. Since

$$(7.34) \qquad\qquad \mathrm{E}_{H_n}\rho_1\left(\frac{y - \gamma_n - \mathbf{x}'\boldsymbol{\beta}_n}{s(H_n)}\right) = b,$$

letting $n \to \infty$, it follows that

$$b = \lim_{n\to\infty} \mathrm{E}_{H_n}\rho_1\left(\frac{y - \gamma_n - \mathbf{x}'\boldsymbol{\beta}_n}{s(H_n)}\right) = (1 - \epsilon) + \epsilon \lim_{n\to\infty} \rho_1\left(\frac{y_n - \gamma_n - \mathbf{x}'_n\boldsymbol{\beta}_n}{s(H_n)}\right)$$

$$> (1 - \epsilon) \lim_{n \to \infty} \mathrm{E}_{H_o} \rho_1 \left(\frac{y - \alpha_t - \mathbf{x}' \boldsymbol{\theta}_t}{s(H_n)} \right) = \mathrm{E}_{H_n} \rho_1 \left(\frac{y - \alpha_t - \mathbf{x}' \boldsymbol{\theta}_t}{s(H_n)} \right).$$

Then, there exists $s_n < s(H_n)$ such that

$$\mathrm{E}_{H_n} \rho_1 \left(\frac{y - \alpha_t - \mathbf{x}' \boldsymbol{\theta}_t}{s_n} \right) = b$$

but this fact contradicts the definition of $(\gamma_n, \boldsymbol{\beta}_n)$. Therefore, we can also assume without loss of generality that $\lim_{n \to \infty} \gamma_n = \tilde{\gamma}$, for some finite $\tilde{\gamma} \in \mathbb{R}$. As a consequence, letting $n \to \infty$ in (7.34) we obtain

$$b = (1 - \epsilon) \mathrm{E}_{H_o} \rho_1 \left(\frac{y - \tilde{\gamma} - \mathbf{x}' \tilde{\boldsymbol{\beta}}}{s_\infty} \right) + \epsilon \ge (1 - \epsilon) \mathrm{E}_{H_o} \rho_1 (y/s_\infty) + \epsilon.$$

Comparing the last equation with (7.33), we deduce that $s_\infty \ge \overline{s}$. Since $\overline{s} = \sup_{H \in V_\epsilon} s(H)$, then $s_\infty = \overline{s}$. We use this fact to obtain equations (7.35) and (7.36) below.

Equations (7.31) and (7.32) imply,

$$(7.35) \quad \lim_{n \to \infty} E_{H_n} \rho_2 \left(\frac{y - \alpha_n - \mathbf{x}' \boldsymbol{\theta}_n}{s(H_n)} \right) = (1 - \epsilon) E_{H_o} \rho_2 \left(\frac{y - \tilde{\alpha} - \mathbf{x}' \tilde{\boldsymbol{\theta}}}{\overline{s}} \right) + \epsilon$$

$$\ge (1 - \epsilon) E_{H_o} \rho_2 \left(\frac{y}{\overline{s}} \right) + \epsilon > (1 - \epsilon) E_{H_o} \rho_2 \left(\frac{y - \alpha_t - \mathbf{x}' \boldsymbol{\theta}_t}{\overline{s}} \right)$$

On the other hand, applying (7.32),

$$(7.36) \quad \lim_{n \to \infty} E_{H_n} \rho_2 \left(\frac{y - \alpha_t - \mathbf{x}' \boldsymbol{\theta}_t}{s(H_n)} \right) = (1 - \epsilon) E_{H_o} \rho_2 \left(\frac{y - \alpha_t - \mathbf{x}' \boldsymbol{\theta}_t}{\overline{s}} \right).$$

Therefore, for large enough n,

$$E_{H_n} \rho_2 \left(\frac{y - \alpha_n - \mathbf{x}' \boldsymbol{\theta}_n}{s(H_n)} \right) > E_{H_n} \rho_2 \left(\frac{y - \alpha_t - \mathbf{x}' \boldsymbol{\theta}_t}{s(H_n)} \right).$$

This last inequality is a contradiction with the definition of $(\alpha_n, \boldsymbol{\theta}_n)$. For every $t > 0$ such that $B_S(\epsilon) < t < t_1$ we have found a sequence of distributions $\{H_n\}$ in the neighborhood V_ϵ such that $\sup_n \|\mathbf{T}(H_n)\| \ge t$. Therefore $B_{MM}(\epsilon) \ge t_1$. $\square$

Proof of Theorem 3.2. It is enough to check that the functional $J(F)$ defined in (3.15) satisfies condition A1 in [1]. For instance, the monotonicity condition A1(a) follows immediately from the monotonicity of the M-scale $\sigma(F)$. $\square$

Proof of Theorem 4.1: We will apply Theorem 3.2. Let $F_o = (1 - \epsilon)F_{H_o,0,\mathbf{0}} + \epsilon\delta_\infty$. Then, $\sigma(F_o) = \sigma_{b,\epsilon}$ and

(7.37)
$$r_{cm}(\epsilon) = J_{CM}(F_o) = \inf_{s \geq \sigma_{b,\epsilon}} A_{c,\epsilon}(s) + c\epsilon.$$

On the other hand, if $\|\boldsymbol{\theta}\| = t$ and $F_t = (1 - \epsilon)F_{H_o,0,\boldsymbol{\theta}} + \epsilon\delta_o$. Then, when H_o is multivariate normal, we have that $\sigma(F_t) = (1 + t^2)^{1/2}\gamma_{b,\epsilon}$ and

(7.38)
$$m_{cm}(t) = J_{CM}(F_t) = \frac{1}{2}\log(1 + t^2) + \inf_{s \geq \gamma_{b,\epsilon}} A_{c,\epsilon}(s).$$

From (3.18), we know that $B_{CM}(\epsilon) = t_\epsilon$, where $m_{cm}(t_\epsilon) = r_{cm}(\epsilon)$. Matching the expressions in equations (7.37) and (7.38), and solving for t yields the result. $\square$

Proof of Theorem 4.2: Let $t \in \mathbb{R}$ arbitrary. Under the assumptions, the function $m(t, s)$ is continuously differentiable with respect to s, with derivative given by

$$\frac{\partial m(t,s)}{\partial s} = \frac{1}{s}\left[\phi_2(s) - \phi_2\left(\frac{s}{(1+t^2)^{1/2}}\right)\right].$$

Since $\phi_2(s)$ is unimodal, for each ϵ we have that $m(t, s)$ is (a) strictly increasing for $s \in [\underline{s}, \overline{s}]$, (b) strictly decreasing for $s \in [\underline{s}, \overline{s}]$, or (c) it has a unique critical point $\tilde{s} \in (\underline{s}, \overline{s})$, which is a local maximum. In any of the three cases, the global minimum of $m(t, s)$ for $s \in [\underline{s}, \overline{s}]$ is attained at one of the two extremes of the interval. That is,

$$h_2(t) = \inf_{\underline{s} \leq s \leq \overline{s}} m(t, s) = \min\{m(t, \underline{s}), m(t, \overline{s})\}.$$

From Theorem 3.1, an upper bound for the maximum bias is given by the value of t_ϵ such that $h_2(t_\epsilon) = \epsilon/(1 - \epsilon)$. If $h_2(t_\epsilon) = m(t_\epsilon, \overline{s})$, then $h_1(t_\epsilon) = h_2(t_\epsilon)$, and therefore

$t_\epsilon = \ell(\epsilon)$. On the other hand, if $h_2(t_\epsilon) = m(t_\epsilon, \underline{s})$, then we have that $t_\epsilon = u(\epsilon)$. Hence, the result follows. $\square$

Proof of (4.23). We apply Theorem 1 in [1]. Following the notation in that paper, we have that $c = \sigma_{b,\epsilon}$. On the other hand,

$$m(t) \doteq \inf_{\|\boldsymbol{\theta}\|=t} \inf_{\alpha \in \mathbb{R}} J_S[(1-\epsilon)F_{H_o,\alpha,\boldsymbol{\theta}}+\epsilon\delta_0] = \inf_{\|\boldsymbol{\theta}\|=t} J_S[(1-\epsilon)F_{H_o,0,\boldsymbol{\theta}}+\epsilon\delta_0] = \inf_{\|\boldsymbol{\theta}\|=t} S(\boldsymbol{\theta}),$$

where $S(\boldsymbol{\theta})$ is such that

$$(7.39) \qquad\qquad (1-\epsilon)\mathrm{E}_{H_0}\rho\left(\frac{y-\mathbf{x}'\boldsymbol{\theta}}{S(\boldsymbol{\theta})}\right) = b.$$

Since $y - \mathbf{x}'\boldsymbol{\theta}$ is distributed as $(1 + \sum_i |\theta_i|)Z$, where Z is standard Cauchy, we have that (7.39) amounts to

$$(1-\epsilon)g\left(\frac{S(\boldsymbol{\theta})}{1+\sum_i |\theta_i|}\right) = b.$$

Therefore,

$$S(\boldsymbol{\theta}) = \left[1 + \sum_i |\theta_i|\right]\gamma_{b,\epsilon}$$

and

$$(7.40) \qquad\qquad m(t) = \inf_{\|\boldsymbol{\theta}\|=t}\left[1+\sum_i |\theta_i|\right]\gamma_{b,\epsilon} = (1+t)\gamma_{b,\epsilon}.$$

Finally, since $B_s(\epsilon) = t$, where $m(t) = \sigma_{b,\epsilon}$, the result follows from (7.40). $\square$

Proof of (4.24). Clearly, Under the Cauchy model, the expression for $r_{cm}(\epsilon)$ is formally the same as that corresponding to the Gaussian model. We just have to compute $g(s)$ with respect to the Cauchy distribution instead of the normal. On the other hand, under the Cauchy model it is not difficult to check that

$$m_{cm}(t) = \log(1+t) + \inf_{s \geq \gamma_{b,\epsilon}} A_{c,\epsilon}(s),$$

where $A_{c,\epsilon}(s)$ is defined by (4.20). Since the bias satisfies $m[B_{cm}(\epsilon)] = r_{cm}(\epsilon)$, the result follows. $\square$

Proof of (4.25). The same arguments as above yield the following expression for the function $m(t, s)$ under the cauchy model:

$$m(t, s) = g_2 \left(\frac{s}{1 + t} \right) - g_2(s).$$

From this expression the computation of $\ell(\epsilon)$ and $u(\epsilon)$ under the cauchy model is straightforward:

$$\ell(\epsilon) = h_1^{-1} \left(\frac{\epsilon}{1 - \epsilon} \right) = \frac{\sigma_{b,\epsilon}}{g_2^{-1} \left[g_2(\sigma_{b,\epsilon}) + \epsilon/(1 - \epsilon) \right]} - 1,$$

and

$$u(\epsilon) = \frac{\gamma_{b,\epsilon}}{g_2^{-1} \left[g_2(\gamma_{b,\epsilon}) + \epsilon/(1 - \epsilon) \right]} - 1.$$

Since we are assuming that $\phi(s)$ is unimodal, the same proof as in the case of the gaussian model yields (4.25).

Proof of Theorem 6.1. (i) Computing the derivative of $A_{c,\epsilon}(s)$ with respect to s, we see that $A_{c,\epsilon}(s)$ is non decreasing when $c < [(1 - \epsilon)\phi(s)]^{-1}$. Since $K^{-1} < [(1 - \epsilon)\phi(s)]^{-1}$ for all ϵ and $s > 0$, the condition $c \leq K^{-1}$ implies that $A_{c,\epsilon}(s)$ is non decreasing for all ϵ and $s > 0$. As a consequence, $h_c(\epsilon, \sigma) = 0$ for all ϵ and $\sigma > 0$. Then, $d_c(\epsilon) = 0$ for all ϵ what implies that $B_{CM}(\epsilon) = B_S(\epsilon)$ for all ϵ.

(ii) Since $g(\sigma_{b,\epsilon}) = (b - \epsilon)/(1 - \epsilon)$ and $g(\gamma_{b,\epsilon}) = b/(1 - \epsilon)$, it follows that

(7.41) $$A_{c,\epsilon}(\sigma_{b,\epsilon}) < A_{c,\epsilon}(\gamma_{b,\epsilon}) \Leftrightarrow c > c(\epsilon).$$

However, if $A_{c,\epsilon}(\sigma_{b,\epsilon}) < A_{c,\epsilon}(\gamma_{b,\epsilon})$, then $d_c(\epsilon) > 0$ and hence $B_S(\epsilon) < B_{CM}(\epsilon)$. $\square$

Proof of (6.28). By using implicit differentiation, one obtains

$$\frac{\partial \sigma_{b,\epsilon}}{\partial \epsilon} = \frac{1}{(1 - \epsilon)^2} \frac{(1 - b)\sigma_{b,\epsilon}}{\phi(\sigma_{b,\epsilon})}, \quad \text{and} \quad \frac{\partial \gamma_{b,\epsilon}}{\partial \epsilon} = \frac{1}{(1 - \epsilon)^2} \frac{-b \, \gamma_{b,\epsilon}}{\phi(\gamma_{b,\epsilon})}.$$

This then gives

$$(7.42) \qquad \frac{\partial\, \epsilon c(\epsilon)}{\partial\, \epsilon} = \frac{1}{(1-\epsilon)^2} \left(\frac{1-b}{\phi(\sigma_{b,\epsilon})} + \frac{b}{\phi(\gamma_{b,\epsilon})} \right) \geq \frac{1-b}{K} + \frac{b}{\phi(\sigma_{b,0})}.$$

The last inequality follows since as noted previously, $\gamma_{b,\epsilon} < \sigma_{b,0} < \sigma_M$. This then implies (6.28). $\square$

Proof of Theorem 6.2. For $c \leq 1/K$, it has already been noted that the maximum bias functions are the same, and so we only need to consider $1/K < c \leq c_o$. In general, for $c > 1/K$ and under assumption A4, the function $A_{c,\epsilon}(s)$ has the following properties:

i) $A_{c,\epsilon}(0) = -\infty$ and $A_{c,\epsilon}(\infty) = \infty$.

ii) $A_{c,\epsilon}(s)$ has two critical points, say $\sigma_L(c,\epsilon) \leq \sigma_U(c,\epsilon)$, with

$\qquad A_{c,\epsilon}(s) \Uparrow$ over 0 to $\sigma_L(c,\epsilon)$,

$\qquad A_{c,\epsilon}(s) \Downarrow$ over

$\qquad \sigma_L(c,\epsilon)$ to $\sigma_U(c,\epsilon)$, and

$\qquad A_{c,\epsilon}(s) \Uparrow$ over $\sigma_U(c,\epsilon)$ to ∞.

iii) $A_{c,\epsilon}(s)$ is concave for $s < \sigma_M$ and convex for $s > \sigma_M$.

Note that the critical points of $A_{c,\epsilon}(s)$ correspond to the two solutions to $\phi(s) = 1/[(1-\epsilon)c]$. The value of σ_M, though, does not depend on c or ϵ. Graphs of a typical function $A_{c,\epsilon}(\sigma)$ for two different values of ϵ are given in Figure 5.

Some further properties which are easy to verify are the following.

a) $\gamma_{b,\epsilon}$, $\sigma_{b,\epsilon}$, $\sigma_L(c,\epsilon)$, $\sigma_U(c,\epsilon)$, and $A_{c,\epsilon}(s)$ are continuous in ϵ.

b) As $\epsilon \Uparrow$: $\gamma_{b,\epsilon} \Downarrow$, $\sigma_{b,\epsilon} \Uparrow$, $\sigma_L(c,\epsilon) \Uparrow$, $\sigma_U(c,\epsilon) \Downarrow$, and $A_{c,\epsilon}(s) \Downarrow$.

c) $\gamma_{b,\epsilon} \leq \sigma_{b,\epsilon}$ with $\gamma_{b,0} = \sigma_{b,0}$.
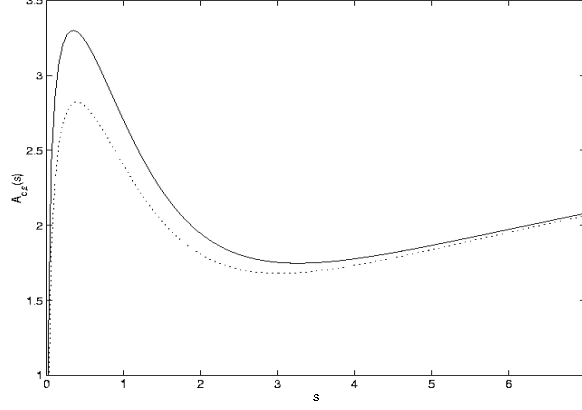
FIG. 5.   *Graph of $A_{c,\epsilon}(\sigma)$.*

d) if $\gamma < \sigma$, then $A_{c,\epsilon}(\gamma) - A_{c,\epsilon}(\sigma)$ is decreasing in ϵ.

Now, for $1/K < c \leq c_o$,

$$(7.43) \qquad \text{if } \sigma_{b,\epsilon} \leq \sigma_U(c,\epsilon), \text{ then } B_{CM}(\epsilon) \leq B_S(\epsilon).$$

since in this case $d_c(\epsilon) \leq 0$. So, to prove Theorem 6.2, it only needs to be shown that

$$(7.44) \qquad \text{if } \sigma_{b,\epsilon} > \sigma_U(c,\epsilon), \text{ then } A_{c,\epsilon}(\gamma_{b,\epsilon}) \leq A_{c,\epsilon}(\sigma_U(c,\epsilon))$$

since this implies $d_c(\epsilon) = 0$ and hence $B_{CM}(\epsilon) = B_S(\epsilon)$.

To show (7.44), first note that $\sigma_{b,0} \leq \sigma_M$ since $g(\sigma_{b,0}) = b \geq g(\sigma_M)$. Thus, since $\sigma_{b,\epsilon} \Uparrow$ and $\sigma_U(c,\epsilon) \Downarrow$ as ϵ increases and both are continuous, there exists an ϵ_b such that $\sigma_{b,\epsilon_b} = \sigma_U(c,\epsilon_b)$. For any $\epsilon \leq \epsilon_b$, it then follows that $\sigma_{b,\epsilon} \leq \sigma_{b,\epsilon_b} = \sigma_U(c,\epsilon_b)$, and so to show (7.44), it is only necessary to consider $\epsilon > \epsilon_b$.

For $\epsilon > \epsilon_b$, we have

$$A_{c,\epsilon_b}(\gamma_{b,\epsilon}) \leq A_{c,\epsilon_b}(\gamma_{b,\epsilon_b}) \leq A_{c,\epsilon_b}(\sigma_{b,\epsilon_b}) = A_{c,\epsilon_b}(\sigma_U(c,\epsilon_b)) \leq A_{c,\epsilon_b}(\sigma_U(c,\epsilon)).$$

The first inequality follows since $\gamma_{b,\epsilon_b} \leq \sigma_L(c, \epsilon_b)$, the second inequality follows from (7.41), and the third inequality follows from (b) since $\sigma_U(c, \epsilon_b) > \sigma_U(c, \epsilon) > \sigma_M$. Statement (7.44) then follows from (d) above. □

Proof of Remark 6.1. The remark for $c > c(\epsilon)$ and for $c \leq 1/K$ have already been established. If $c > 1/K$ and $g(\sigma_M) > b$, then $\gamma_{b,0} = \sigma_{b,0} > \sigma_M$. Now, if $\sigma_{b,0} \geq \sigma_U(c, 0)$, then since $\sigma_{b,\epsilon} \Uparrow$ and $\sigma_U(c, 0) \Downarrow$ as ϵ increases, it follows that $\sigma_{b,\epsilon} \geq \sigma_U(c, \epsilon)$ for all ϵ. This then implies $d_c(\epsilon) \geq 0$ and hence $B_{CM}(\epsilon) \geq B_S(\epsilon)$.

On the other hand, if $\sigma_M < \sigma_{b,0} < \sigma_U(c, \epsilon)$, then by continuity for small enough ϵ, $\sigma_M < \gamma_{b,\epsilon} < \sigma_{b,\epsilon} < \sigma_U(c, \epsilon)$. This implies $A_{c,\epsilon}(\gamma_{b,\epsilon}) < A_{c,\epsilon}(\sigma_{b,\epsilon})$, and so by (7.41), $c > c(\epsilon)$. □

Proof of Theorem 6.3. Note that, under the conditions of Theorem 6.2, $B_S(\epsilon) > B_{CM}(\epsilon)$ if and only if

$$(7.45) \qquad \sigma_{b,\epsilon} < \sigma_U(c, \epsilon) \text{ and } A_{c,\epsilon}(\sigma_{b,\epsilon}) > A_{c,\epsilon}(\sigma_U(c, \epsilon)).$$

So, to prove that an S-functional is inadmissible, one only needs to establish (7.45) for some ϵ. First, we will show that the condition $c_o = \lim_{\epsilon \to 0^+} c(\epsilon)$ implies that there exists ϵ such that (7.45) holds. Then, we will show that (6.29) is enough to guarantee $c_o = \lim_{\epsilon \to 0^+} c(\epsilon)$. Note that by using L'Hopital's rule one obtains

$$(7.46) \qquad c(0) = \lim_{\epsilon \to 0^+} c(\epsilon) = \frac{1}{\phi(\sigma_{b,0})}.$$

Also, note that

$$(7.47) \qquad c > c_1 = \frac{\log(\sigma_K/\sigma_{b,0})}{b - g(\sigma_M)} \Leftrightarrow A_{c,0}(\sigma_{b,0}) > A_{c,0}(\sigma_M).$$

Since $\sigma_{b,0} < \sigma_M$, this implies $c_1 \geq 1/K$ since otherwise $A_{c,0}(s)$ would be monotone in s. Now, for any $c > c_1$, we then have $\sigma_{b,0} < \sigma_M < \sigma_U(c, 0)$ and $A_{c,0}(\sigma_{b,0}) > A_{c,0}(\sigma_M) \geq A_{c,0}(\sigma_U(c, 0))$. By continuity, statement (7.45) then follows for small enough ϵ. Now,

we show that $c_1 \leq c(0)$. To show this, note that when $c = c(0)$, $\sigma_{b,0} = \sigma_L(c, 0)$ and so $A_{c,0}(\sigma_{b,0}) > A_{c,0}(\sigma_M)$. The first part of the proof then follows from (7.47).

Notice that the lower bound c_1 can be tighten by working with (7.45) directly. In general, it is difficult to use (7.45) to obtain a closed form expression, but it can be used for specific examples.

From (7.46), in the second part of the proof we need to show that (6.29) implies

$$(7.48) \qquad \epsilon\, c(\epsilon) \geq \epsilon/\phi(\sigma_{b,0}).$$

Since equality holds in (7.48) when $\epsilon = 0$, to show (7.48) it is sufficient to prove that the derivative of the left-hand side is never less than the derivative of the right-hand side, i.e. [see equations (7.42) and (7.46)]

$$(7.49) \qquad \frac{1}{(1-\epsilon)^2}\left\{\frac{1-b}{\phi(\sigma_{b,\epsilon})} + \frac{b}{\phi(\gamma_{b,\epsilon})}\right\} \geq \frac{1}{\phi(\sigma_{b,0})}.$$

Recall that we are assuming $g(\sigma_M) < b = g(\sigma_{b,0})$, or equivalently that $\sigma_{b,0} < \sigma_M$. This implies $\phi(\gamma_{b,\epsilon}) < \phi(\sigma_{b,0})$, and after some simple algebraic manipulations, we note that (7.49) holds if

$$(7.50) \qquad a_{b,\epsilon}\, \phi(\sigma_{b,\epsilon}) \leq \phi(\sigma_{b,0}),$$

where $a_{b,\epsilon} = [(1-\epsilon)^2 - b]/(1-b)$.

Since $\sigma_{b,\epsilon}$ is increasing in ϵ, then $\phi(\sigma_{b,\epsilon})$ is decreasing in ϵ whenever $\sigma_{b,\epsilon} \geq \sigma_M$, it follows that if (7.50) holds for $\sigma_{b,\epsilon} = \sigma_M$ then it holds for $\sigma_{b,\epsilon} \geq \sigma_M$. Thus, it is sufficient to show that (7.50) holds for $\sigma_{b,\epsilon} \leq \sigma_M$, or equivalently for

$$\epsilon \leq \epsilon_M \doteq \frac{b - g(\sigma_M)}{1 - g(\sigma_M)}\ .$$

Given that $g(s)$ is convex, $-g'(\sigma_{b,\epsilon}) \leq -g'(\sigma_{b,0})$, and so (7.50) holds if $a_{b,\epsilon}\, \sigma_{b,\epsilon} \leq \sigma_{b,0}$. Since $g(s)$ is also nonincreasing, this is equivalent to

$$(7.51) \qquad g(a_{b,\epsilon}\, \sigma_{b,\epsilon}) \geq b$$

Thus, the theorem is proven if (7.51) holds for $\epsilon \leq \epsilon_K$. By the convexity of $g(s)$, for $\epsilon \leq \epsilon_K$,

$$g(a_{b,\epsilon}\ \sigma_{b,\epsilon}) \geq g(\sigma_{b,\epsilon}) + (a_{b,\epsilon} - 1)\sigma_{b,\epsilon}g'(\sigma_{b,\epsilon})$$

$$= \tfrac{b-\epsilon}{1-\epsilon} + \tfrac{\epsilon(2-\epsilon)}{1-b}\phi(\sigma_{b,\epsilon}) \geq \tfrac{b-\epsilon}{1-\epsilon} + \tfrac{\epsilon(2-\epsilon)}{1-b}\phi(\sigma_{b,0})$$

The last term is $\geq b$ if and only if

$$(7.52) \qquad\qquad \phi(\sigma_{b,0}) \geq \frac{(1-b)^2}{(1-\epsilon)(2-\epsilon)}.$$

Notice that if (7.52) holds for $\epsilon = \epsilon_M$, then it holds for all $\epsilon \leq \epsilon_M$. With $\epsilon = \epsilon_M$, though, (7.52) corresponds to the bound (6.29). This completes the proof. $\square$

REFERENCES

[1] BERRENDERO, J.R. and ZAMAR, R. (2001). Maximum bias curves for robust regression with non-elliptical regressors. *Ann. Statist.*, **29**, 224-251.

[2] HAMPEL, P.R., RONCHETTI, E.M., ROUSSEEUW, P.J. and STAHEL, W.A. (1986). *Robust Statistics: The Apporach Based On Influence Function.*, Wiley, New York.

[3] HE, X. (1991). A local breakdown property of robust tests in linear regression. *J. Multivariate Anal.*, **38**, 293–305.

[4] HE, X. and SIMPSON, D.G. (1993). Lower bounds for contamination bias: globally minimax versus locally linear estimation. *Ann. Statist.*, **21**, 314–337.

[5] HENNIG, C. (1995). Efficient high breakdown point estimators in robust regression: which function to choose? *Statistics and Decisions*, **13**, 221-241.

[6] MARONNA, R., BUSTOS, O. and YOHAI, V.J. (1979). Bias- and efficiency-robustness of general M-estimators for regression with random carriers. In *Smoothing techniques for curve estimation*, Lecture Notes in Mathematics, 757, Springer, Berlin, 91–116.

[7] MARTIN, R.D., YOHAI, V.J. and ZAMAR, R.H. (1989). Min-Max Bias Robust Regression. *Ann. Statist.*, **17**, 1608-1630.

[8] MENDES, B.V.M. and TYLER, D.E. (1996). Constrained M–estimates for regression, In *Robust statistics, data analysis and computer intensive methods*, Lecture Notes in Statistics, 109, H. Rieder, ed., Springer-Verlag, 299–320.

[9] ROUSSEEUW, P.J. (1984). Least Median of Squares Regression. *J. Amer. Statist. Assoc.*, **79**, 871-880.

[10] ROUSSEEUW, P.J. and YOHAI, V.J. (1984). Robust regression by means of S-estimators. In *Robust and Nonlinear Time Series Analysis*, eds. J. Franke, W. Härdle and R.D. Martin. Lecture Notes in Statistics 26, Springer, New-York, 256–272.

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

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

[13] YOHAI, V.J. and ZAMAR, R.H. (1993). A minimax-bias property of the least α-quantile estimates. *Ann. Statist.* **21**, 1824–1842.

[14] YOHAI, V.J. and ZAMAR, R.H. (1997). Optimal locally robust M-estimates of regression. *J. Statist. Plann. Inference.*, **64**, 309–323.