AD-A242 918
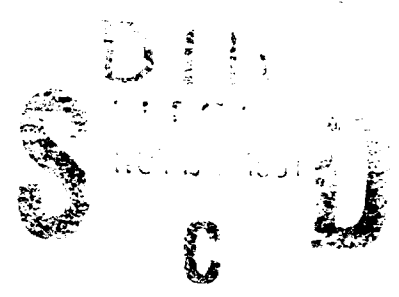

EFFICIENCY CONSTRAINED BIAS ROBUST ESTIMATION
OF LOCATION

by

R. Douglas Martin
Ruben H. Zamar

TECHNICAL REPORT No. 217

October 1991

Department of Statistics, GN-22

University of Washington

Seattle, Washington 98195 USA

91-16495

91 1125 095

Efficiency Constrained Bias

Robust Estimation of Location

R.D. Martin

Department of Statistics

University of Washington, Seattle

Ruben H. Zamar

Department of Statistics

University of British Columbia

October, 1991

ABSTRACT

In 1964 P. Huber established the following minimax bias robustness result for estimating the location μ in the ϵ-contamination family $F(x) = (1-\epsilon)\Phi[(x-\mu)/s] + \epsilon H(x)$ where Φ is the standard normal distribution and H is an arbitrary distribution function: The median minimizes the maximum asymptotic bias among all translation equivariant estimates of location. However, the median efficiency of $2/\pi$ at the Gaussian model may be unacceptably low in some applications. This motivates one to solve the following problem for the above ϵ-contamination family: Among all location M-estimates, find the one which minimizes the maximum asymptotic bias subject to a constraint on efficiency at the Gaussian model. This problem is the dual form analog of Hampel's optimality problem of minimizing the asymptotic variance at the nominal model (e.g. the Gaussian model) subject to a bound on the gross-error sensitivity. We solve the global problem completely for the case of a known scale parameter. The main conclusion is that Hampel's heuristic is essentially correct: the resulting M-estimate is based on a psi-function which is amazingly close, but not exactly equal, to the Huber/Hampel optimal ψ. It turns out that one pays only a relatively small price in terms of increase in maximal bias for increasing efficiency from 64% to the range $90\% - 95\%$. We also present a conjectured solution to the problem, based on heuristic arguments and numerical calculations, when the nuisance scale parameter is unknown.

Some key words: bias-robustness, minimax, efficiency.

1. INTRODUCTION

Consider the family of ϵ-contaminated Gaussian distribution functions

$$(1.1) \qquad \mathcal{F} = \left\{ F : F(x) = (1 - \epsilon)\Phi\left(\frac{x - \mu}{s}\right) + \epsilon H(x) \right\}$$

where $0 < \epsilon < .5$ is fixed, Φ is the standard normal distribution, and H is an arbitrary distribution. The main focus will be on estimation of the location parameter μ, with s being a nuisance scale parameter.

In this setup, where the contamination distribution may be asymmetric, all the "usual" robust estimates of μ will be biased asymptotically as well as in finite sample sizes for many F in the family $\mathcal{F}$. This problem was recognized by Huber(1964) in a brief section of his seminal paper on robust M-estimation. Huber's primary focus was on the restricted *symmetric* form of $\mathcal{F}$, where H is constrained to be any symmetric distribution, and for this family he obtained the asymptotic variance minimax M-estimate of μ. However, working with the full asymmetric family $\mathcal{F}$, Huber(1964) also proved the following result: Among all translation equivariant estimates of location, the sample median minimizes the maximum asymptotic bias over the family (1.1) with $s = 1$. His solution also holds with Φ replaced by certain other symmetric distributions, and for the class of all translation and scale equivariant estimates of location with s unknown.

The minimax bias robustness problem can be stated formally for a class $\mathcal{T}$ of location estimates and the family $\mathcal{F}$ given by (1.1) as follows. Assuming as usual that $\mathcal{T}$ contains only translation and scale equivariant estimates, one takes $\mu = 0$ and $s = 1$ without loss of generality. Let $T(F)$ be the asymptotic value of an estimate $T \in \mathcal{T}$, and let $b_T(\epsilon, F)$ be the asymptotic bias of T at F. Since $\mu = 0$, we have

$$b_T(\epsilon, F) = T(F).$$

Then the maximum asymptotic bias of T over $\mathcal{F}$ is

$$(1.2) \qquad B_T(\epsilon) = \sup_{F \in \mathcal{F}} b_T(\epsilon, F).$$

A minimax estimate T_ϵ^* is one which satisfies

$$(1.3) \qquad T_\epsilon^* = \operatorname*{argmin}_{T} B_T(\epsilon)$$

for each $\epsilon \in (0, .5)$. In general, any estimate which minimizes the maximum asymptotic bias with respect to specified classes of estimates and mixture distributions will be called a *bias-robust* estimate.

Curiously enough, the global problem of constructing bias robust estimates was ignored for many years following Huber (1964). Only quite recently do we find a number of results along these lines for problems such as minimum distance estimation (Donoho and Liu, 1988 a and b); *estimation of scale* (Martin and Zamar, 1989; Martin and Zamar, 1990), regression (Martin, Yohai and Zamar, 1989; Yohai, 1990), and covariance matrices (Maronna and Yohai, 1989). While Huber (1981) found that the bias robustness problem produced "a rather uneventful theory" in the case of estimating location, the results cited above indicate that this is not the case for other kinds of parameter estimation problems.

Of course, one criticism of bias robustness is that this kind of robustness might be achieved at the expense of a severe loss of efficiency at the central model, e.g., at Φ in (1.1). Indeed this is the case to some extent for the median as a bias robust estimate of location, and to a much more serious extent in the case of regression: Martin, Yohai, and Zamar (1989) show that among all M-estimates of regression based on bounded rho-functions, the bias robust estimate minimizes a quantile of the absolute residuals. This bias robust estimate has the same slow rate of convergence as the least median

of squared residuals (Rousseeuw, 1984; Rousseeuw and Leroy, 1988), which turns out to be a quite good approximation to the bias robust M-estimate of regression (Martin, Yohai and Zamar, 1989).

The slow rate of convergence of the bias robust estimate could be avoided by imposing an efficiency constraint at the central model, and this approach could lead to a useful tradeoff between Gaussian-case efficiency and bias control. Because of the relative simplicity of the location problem, we initiated our efforts to construct *efficiency-constrained bias robust estimates* on the location problem for the ϵ-contamination model (1.1). Thus our problem is to solve:

$$(1.4) \qquad\qquad T_\epsilon^* = \underset{T \in \mathcal{T}}{\operatorname{argmin}} \, B_T(\epsilon)$$

subject to $\mathrm{EFF}(T, \Phi) \leq e$, where $\mathrm{EFF}(T, \Phi)$ is the asymptotic efficiency of T at the standard normal distribution Φ, with T in the class $\mathcal{F}$ of M-estimates of location, and $e \in (0, 1)$ a prescribed efficiency.

We remark that this problem could equally well be stated in the dual form

$$(1.5) \qquad\qquad T_\epsilon^* = \underset{T \in \mathcal{T}}{\operatorname{argmin}} \, \mathrm{VAR}(T, \Phi)$$

subject to $B_T(\epsilon) \leq b$, where $\mathrm{VAR}(T, \Phi)$ is the asymptotic variance of T at Φ and $b > 0$ is a prescribed bound on the maximum bias. Stated in this form, it is clear that our problem of interest is a global form of Hampel's well-known local optimality problem (see Hampel, 1968, 1974 and Hampel et. al., 1986): Minimize the variance at the central model, subject to a bound on the gross-error sensitivity (GES). The latter provides, under regularity, a local linear approximation to the maximum bias of an estimate for small ϵ (see Section 2.3). Fortuitously, the technique of proof originally used by Hampel for his local optimality problem turned out to be a key ingredient in

establishing our global result.

2. MAXIMUM ASYMPTOTIC BIAS WITH NUISANCE SCALE

2.1 The Maximum Bias Functional

A location M-estimate T_n is a solution of

$$(2.1) \qquad \sum_1^n \psi\left(\frac{Y_i - T_n}{s_n}\right) = 0.$$

We work with the following assumptions

- (A1) ψ *is continuous, monotone, odd, and bounded.*

- (A2) s_n *is an estimate of scale whose almost sure limit* $\lim_{n\to\infty} s_n$ *defines a scale functional* $S(F)$ *for all* $F \in \mathcal{F}$ *with the boundedness property:* $0 < S(F) \leq S(F^\infty) = \bar{s} < \infty$ *for all* $F \in \mathcal{F}$ *where* $F^\infty = (1 - \epsilon)F_0 + \epsilon\delta_\infty$ *and* δ_∞ *is a point-mass at infinity.*

It is easy to check, using (2.2) below, that under the contamination model (1.1) the maximum asymptotic bias is unbounded for unbounded ψ. Thus the boundedness part of A1 entails no loss of generality. Formulas for computing $\bar{s}$ for the case of M-estimates of scale can be found in Martin and Zamar (1990).

Huber (1981), Section 3.2, Corollary 2.2 shows that under (A1) and (A2) T_n converges almost surely to a functional $T(F) = T(\psi, F)$ provided this functional is uniquely defined by the asymptotic estimating equation

$$(2.2) \qquad \int \psi[(y - T(F))/S(F)]dF(y) = 0.$$

It is not difficult to see that for our setup $T(F)$ is uniquely defined for all $F \in \mathcal{F}$. First

5

let

$$(2.3) \qquad g(t,s) = g_\psi(t,s) \;=\; -\int_\infty^\infty \psi[(y-t)/s]\varphi(y)\,dy, \quad s > 0$$

where $\varphi(y)$ in the standard normal density, and note that for $t > 0$, we have

$$(2.4) \qquad g(t,s) \;=\; s\int_0^\infty \psi(y)[\varphi(sy - t) - \varphi(sy + t)]\,dy.$$

For all $F \in \mathcal{F}$, the equation (2.2), with $T(F)$ replaced by t, can be written as

$$(2.5) \qquad -(1-\epsilon)g[t, S(F)] + \epsilon\int \psi[(y-t)/S(F)]dH(y) = 0.$$

The function $g(t,s)$ is strictly increasing and continuous in t, with $\lim_{t\to\infty} g(t,s) = \psi(\infty)$ and $\lim_{t\to-\infty} g(t,s) = -\psi(\infty)$. Therefore, for all $\epsilon < .5$, the left-hand-side is positive for sufficiently large positive t, and negative for sufficiently small negative t. It follows that the solution (in t) of this equation defines a unique functional $T(F)$.

Furthermore $g(t,s)$ is strictly decreasing in s for $s \in (0,\bar{s})$ with $\lim_{s\to 0} g(t,s) = [2\Phi(t) - 1]\psi(\infty)$. ¿From this and the above observations it is easy to verify that under (A1) and (A2) the maximum asymptotic bias $B_\psi(\epsilon)$ is achieved when H is the point mass δ_∞ at infinity. Thus $B_\psi(\epsilon)$ satisfies the equation

$$(2.6) \qquad -(1-\epsilon)g[B_\psi(\epsilon),\bar{s}] + \epsilon\psi(\infty) = 0.$$

We summarize this result as a lemma.

LEMMA 1: If (A1) and (A2) hold then

$$(2.7) \qquad B_\psi(\epsilon) = g_{\bar{s}}^{-1}\left[\psi(\infty)\frac{\epsilon}{1-\epsilon}\right]$$

where $g_{\bar{s}}^{-1}(\cdot)$ is the inverse of $g(\cdot,\bar{s})$ and $\bar{s} = S(F^\infty)$ is the maximum asymptotic bias of the scale estimate.

6

2.2 Relation with the Gross-Error-Sensitivity

The influence function of a location M-estimate of $T(F)$ with psi-function ι at F_0 and x (see Hampel 1968, 1974) is given by

$$IC_\psi(x) = \frac{\partial}{\partial \epsilon} T[(1-\epsilon)F_0 + \epsilon \delta_x]_{\epsilon=0} = \frac{\psi(x)}{g_{10}},$$

where δ_x is a point-mass at x,

$$g_{\alpha\beta}(t,s) = (\partial^\alpha/\partial t^\alpha)(\partial^\beta/\partial s^\beta)g(t,s) \quad , \quad \alpha, \beta = 1, 2, \ldots$$

and $g_{\alpha\beta} = g_{\alpha\beta}(0,1)$. The gross-error-sensitivity of T at F_0, also introduced by Hampel (1968), is given by

$$GES(\psi) = \sup_x |IC_\psi(x)| = \frac{\sup_x \psi(x)}{g_{10}}.$$

One expects that under sufficient regularity conditions the GES will provide a local linear approximation to $B_\psi(\epsilon)$ for ϵ near zero, that is, that the GES will be equal to the derivative $B'_\psi(0)$ of the maximal bias function $B_\psi(\epsilon)$ at the origin. The following lemma shows that in fact (A1) and (A2) provide sufficient regularity.

LEMMA 2: Under (A1) and (A2) $B'_\psi(0) = GES(\psi)$.

Proof: Follows from (2.6) by differentiation.

Figure 1 gives the maximum bias curves, along with the local linear approximations based on the GES, for the following two location estimators: (1) the median and (2) Huber M-estimate with with $c = 1.5$ using the shorth as scale (see comments at the end of Section 4). Notice that the linear approximation is better for the median than for the efficient Huber estimate.

2.3 The Unconstrained Bias Robust M-Estimate

For the moment assume that $\psi(\infty) = 1$, which entails no loss of generality. ¿From (2.4) and (2.6) and the monotonicity of $g(t, \bar{s})$ in t, it follows that if ψ_1 and ψ_2 satisfy

7

$g_{\psi_1}(t,\bar{s}) \geq g_{\psi_2}(t,\bar{s})$ for all t then $B_{\psi_1}(\epsilon) \leq B_{\psi_2}(\epsilon)$. Thus $B_\psi(\epsilon)$ can be minimized by maximizing the function $g_\psi(t,\bar{s})$ for each $t > 0$. Noting that $[\varphi(\bar{s}x - t) - \varphi(\bar{s}x + t)] \geq 0$ for all $x \geq 0$ and $t \geq 0$, one sees that the "sign" function $\psi(x) = \mathrm{sgn}(x)$ maximizes $g_\psi(t,\bar{s})$ with respect to ψ. Thus the sample median is the bias-robust M-estimate of location with minimax bias

$$B_{MED}(\epsilon) = \Phi^{-1}\left[0.5/(1-\epsilon)\right].$$

This is a particular case, for M-estimates, of the more general result obtained by Huber (1964) for the class of all translation equivariant estimates of location.

3. EFFICIENCY CONSTRAINED SOLUTION WITH SCALE KNOWN

In this section we find the efficiency constrained bias-robust M-estimate of location for the case where scale is known, taking $s = 1$ without loss of generality.

3.1 Candidate Solutions via Calculus of Variation

First, we use calculus of variation to give a heuristic derivation of the optimal psi-function ψ^*. In the next subsection we give a direct proof based on projection methods.

By definition, an efficiency constrained bias-robust M-estimate of location solves the following constrained minimization problem:

$$\inf_\psi B_\psi(\epsilon)$$

subject to

$$(3.1) \qquad V(\psi, \Phi) = .5\frac{\int_0^\infty \psi^2(x)\varphi(x)dx}{[\int_0^\infty \psi'(x)\varphi(x)dx]^2} \leq e^{-1}$$

where $V(\psi, \Phi)$ is the M-estimate asymptotic variance and e is the desired efficiency.

In view of (2.4), (2.6) and the strict monotonicity of $g(t,s)$, it suffices to find, for each $t > 0$, a score function ψ_t which maximizes the functional

$$(3.2) \qquad J_t(\psi) = \frac{1}{\psi(\infty)} \int_0^\infty \psi(x)[\varphi(x - t) - \varphi(x + t)]dx$$

subject to the given side constraint. Then, assuming that $J_t(\psi_t)$ is continuous and monotone increasing in t, the solution to the constraint optimization problem will be $\psi^* = \psi_{t_0}^*$ where $J_{t_0}(\psi_{t_0}^*) = \epsilon/(1 - \epsilon)$. Since the constraint in (3.1) is not an integral constraint and since the objective function $J_t(\psi)$ is not an integral on a finite interval, it is convenient to consider instead the following family of standard optimization problems:

Maximize $J_t(\psi)$ subject to the constraints

(B1) $\psi(0) = 0, \quad \psi(c) = M$.

(B2) $\int_0^c \psi^2(x)\varphi(x)dx + 2[1 - \Phi(c)]M^2 \leq e^{-1}$.

(B3) $2^{\frac{1}{2}} \int_0^c \psi'(x)\varphi(x)dx = 1$.

where c and M are constants such that (B1)-(B3) can be simultaneously satisfied for at least one ψ.

The functional $J_t(\psi)$ is linear, and hence convex, and the set of ψ−functions satisfying (B1)-(B3) is convex. Thus we have a convex optimization problem for each fixed (c, M). At first sight, a natural approach is to solve the convex optimization problem for each allowable pair (c, M) and then optimize over all allowable (c, M). The Lagrangian for the problem with (c, M) fixed is

$$G(x, \psi, \psi') = -\psi(x)[\varphi(x - t) - \varphi(x + t)] + \lambda_1\psi^2(x)\varphi(x) + \lambda_2\psi'(x)\varphi(x)$$

and, by convexity, a sufficient condition for optimality is that the Euler-Lagrange equation be satisfied:

$$G_\psi - \frac{\delta}{\delta x}G_{\psi'} = \varphi(x - t) - \varphi(x + t) + 2\lambda_1\psi(x)\varphi(x) + \lambda_2\varphi'(x) = 0.$$

9

Thus for fixed (c, M), the optimal psi-function is of the form

$$
\psi_t^*(x) = \begin{cases} \alpha_1 x + \alpha_2 \dfrac{\varphi(x-t) - \varphi(x+t)}{\varphi(x)}, & |x| \leq c \\ M & |x| > c \end{cases}
$$

Notice that $\alpha_1 = 1$ and $\alpha_2 = 0$ gives Huber's ψ-function and $\alpha_1 = 0$ and $\alpha_2 = 1$ gives a ψ-function which is proportional to a truncated hyperbolic sine function.

Unfortunately, as c and M vary we no longer have a convex optimization problem and we were unable to make this variational argument rigorous. In the remainder of this section we give a direct proof that a solution to the optimization problem actually exists and is of this form.

3.2 The Dual Problem

Let B be an achievable maximum bias, that is $B = B_\psi(\epsilon)$ in (2.7) for some ψ satisfying (A1) and (A2). Notice that if $B_{MED} = B_{MED}(\epsilon)$ is the maximum bias of the median then $B \geq B_{MED}$. Let Ψ_B be the set of psi-functions satisfying (A1), (A2) and

(C1) $2 \int_0^\infty \psi(x) x \varphi(x) dx = 1$

(C2) $(1 - \epsilon) g_\psi(B, 1) = \epsilon \psi(\infty)$.

Since the asymptotic variance $VAR(\psi, \phi)$ and bias $B_\psi(\epsilon)$ are invariant under multiplication of ψ by a constant, the condition C1 above is just a convenient standarization. Observe that given C1, $\psi(\infty)$ is the gross-error-sensitivity of the corresponding M-estimate, so we will write $GES(\psi) = \psi(\infty)$. Also notice that C1 implies

(3.3) $\qquad\qquad\qquad GES(\psi) \geq (2\pi)^{1/2} = GES(Median)$.

The dual optimization problem can now be stated:

Fix $B > B_{MED}$ and find $\psi^* \in \Psi_B$ which minimizes

(3.4) $\qquad\qquad\qquad J(\psi) = \int_0^\infty \psi^2(x) \varphi(x) \, dx$.

10

We need the following lemma, which states that for any fixed $B > B_{MED}$ and any nice unbounded, odd function $\theta(x)$, there exists a rescaled truncated version of $\theta(x)$ which is in Ψ_B. We will denote by $\theta_c(x)$ the truncation of $\theta(x)$ at c, that is,

$$\theta_c(x) = \begin{cases} \theta(x), & |x| \le c \\ \theta(c)sgn(c) & |x| > c. \end{cases}$$

LEMMA 3: Let $\theta(x)$ be differentiable, odd and monotone, with $\lim_{x\to\infty} \theta(x) = \infty$ and $E_\Phi|\theta(X)| < \infty$. For each $B > B_{MED}$ there exist c_0 and k_0 such that $k_0\theta_{c_0}(x) \in \Psi_B$.

<u>Proof</u>: For fixed $c > 0$, the function

$$\gamma(t,c) = \frac{1}{\theta(c)} \int_0^\infty \theta_c(x)[\varphi(x-t) - \varphi(x+t)]dx = -\frac{1}{\theta(c)} \int_{-\infty}^\infty \theta_c(x-t)\varphi(x)dx$$

is continuous and monotone increasing in t. By the dominated convergence theorem $\lim_{t\to 0} \gamma(t,c) = 0$ and $\lim_{t\to\infty} \gamma(t,c) = 1$ Thus, given $0 < c < \infty$ there exists $B(c)$ such that $\gamma[B(c),c] = \epsilon/(1-\epsilon)$. Since $B(c)$ is continuous and non-decreasing with $\lim_{c\to\infty} B(c) = \infty$ and $\lim_{c\to 0} B(c) = B_{MED}$ there exists c^* such that $B(c^*) = B$. The lemma follows now with $c_0 = c^*$ and $k_0 = [2\int_0^{c_0} \theta'(x)\varphi(x)dx]^{-1}$.

3.3 Solution of the Dual Problem

The dual problem we wish to solve is similar to Hampel's optimality problem of minimizing the variance at the central model subject to a bound on the gross error sensitivity. See Hampel (1968) and Hampel et. al. (1986). The difference of course is that we are replacing the bound on infinitesimal bias with a bound on the actual maximum bias. Nonetheless, Hampel's technique for obtaining a key inequality still provides an essential step in the solution given here.

11

For each $b > GES(Median) = b_0$ let $\Psi_{B,b}$ be the subset of all the $\psi \in \Psi_B$ with gross-error-sensitivity equal to b. Then, clearly $\Psi_B = \cup_{b>b_0}\Psi_{B,b}$. Also, let $\psi_1(x)$ and $\psi_2(x)$ be the truncated and scaled ψ-functions given by Lemma 3 corresponding to the identity function $I(x) = x$ and to the function

$$\Delta(x) = \frac{\varphi(x - B) - \varphi(x + B)}{\varphi(x)} = e^{-B^2/2}(e^{Bx} - e^{-Bx}),$$

which is proportional to the hyperbolic sine. The gross-error sensitivities are

$$GES_i = GES(\psi_i) = \psi_i(c_i) \quad , \quad i = 1, 2$$

where c_1 and c_2 are the corresponding trunction constants.

The following theorem shows that, in terms of $J(\psi)$, ψ_1 dominates all ψ in $\Psi_{B,b}$ with $b_0 < b \leq GES_1$ and ψ_2 dominates all ψ in $\Psi_{B,b}$ with $b \geq GES_2$.

THEOREM 1: Suppose that $\psi \in \Psi_B$ satisfies $GES(\psi) \leq GES_1$ or $GES(\psi) \geq GES_2$. Then $J(\psi) \geq \min\{J(\psi_1), J(\psi_2)\}$.

Proof: Assume first that $b_0 < b = GES(\psi) \leq GES_1$. Since $GES(I_c) = .5c[\Phi(c) - .5]^{-1}$ is continuous in c, tends to $b_0 = (\pi/2)^{1/2}$ as $c \to 0$ and tends to ∞ as $c \to \infty$, there exists $0 < c \leq c_1$ such that $GES(I_c) = b$. The inequality $J(\psi) \geq VAR(I_c, \Phi)$ now directly follows from Hampel's result of optimality of I_c among all functions which satisfy A1-A2 and have GES bounded above by b. (see for example Hampel et. al. 1986, Theorem 1, Section 2.4, p.117; see also Theorem 5, Section 2.5d, p. 135). Then, since $VAR(I_c, \Phi)$ is a decreasing function of c, one follows that

$$J(\psi) \geq VAR(I_c, \Phi) \geq VAR(I_{c_1}, \Phi) = J(\psi_1).$$

We now turn to the optimality of ψ_2 over $\cup_{b \geq GES_2}\Psi_{B,b}$. Assume that $b = GES(\psi) \geq GES_2 = b_2$ and let

$$\overline{\psi}(x) = \frac{\psi(x)}{b} \quad , \quad \overline{\psi}_2(x) = \frac{\psi_2(x)}{b_2}$$

12

$$\overline{\Delta}(x) = \frac{\Delta(x)}{\Delta(c_2)} \quad , \quad \overline{\beta} = \frac{\epsilon}{\Delta(c_2)(1-\epsilon)}.$$

Observe that $b = \psi(\infty)$ and so, using C2, one obtains $\int_0^\infty \overline{\psi}(x)\overline{\Delta}(x)\varphi(x)dx = \overline{\beta}$. A similar argument shows that $\int_0^\infty \overline{\psi}_2(x)\overline{\Delta}(x)\varphi(x)dx = \overline{\beta}$. Using the last two equations and the fact that $\overline{\Delta}(x) = \overline{\psi}_2(x)$ for $0 \le x \le c_2$, $\overline{\Delta}(x) > 1$ for $x > c_2$ and $\overline{\psi}(x) \le 1$ for $x \ge 0$ one follows that

$$
\begin{aligned}
\int_0^\infty \overline{\psi}^2(x)\varphi(x)dx + d &= \int_0^\infty [\overline{\psi}(x) - \overline{\Delta}(x)]^2\varphi(x)dx \\
&\ge \int_0^\infty [\overline{\psi}_2(x) - \overline{\Delta}(x)]^2\varphi(x)dx \\
&= \int_0^\infty \overline{\psi}_2^2(x)\varphi(x)dx + d,
\end{aligned}
$$

where $d = -2\overline{\beta} + \int_0^\infty \overline{\Delta}^2(x)\varphi(x)dx$ is a constant which does not depend on ψ. Therefore,

$$J(\psi) = \int_0^\infty \psi^2(x)\varphi(x)dx \ge (b/b_2)^2 \int_0^\infty \overline{\psi}_2^2(x)\varphi(x)dx \ge J(\psi_2),$$

completing the proof.

The following theorem is our main result.

THEOREM 2: For each $\psi \in \Psi_B$ there exist $c \in [c_1, c_2]$, $\alpha_1 \ge 0$, and $\alpha_2 \ge 0$ such that

$$
\phi_c(x) = \begin{cases}
\alpha_1 I(x) + \alpha_2 \Delta(x) & |x| \le c \\
[\alpha_1 I(c) + \alpha_2 \Delta(c)]sgn(x) & |x| \ge c
\end{cases}
$$

belongs to Ψ_B and $J(\psi) \ge J(\phi_c)$.

<u>Proof</u>: Let $\psi \in \Psi_B$ be such that

$$GES(\psi_1) < GES(\psi) < GES(\psi_2).$$

Note that by Theorem 1, we only need to consider this case because if $GES(\psi_2) \le GES(\psi_1)$ then the theorem trivially holds. To fix ideas, suppose that $c_1 > c_2$. The cases $c_1 < c_2$ and $c_1 = c_2$ can be handled in a similar way.

The function $\psi_c(x) = \alpha_1(c)I_c(x) + \alpha_2(c)\Delta_c(x)$ is in Ψ_B provided that $\alpha_1(c)$ and $\alpha_2(c)$ are non-negative and satisfy the equations

$$\alpha_1(c)A_{11} + \alpha_2(c)A_{12} = 0$$

$$\alpha_1(c)A_{21} + \alpha_2(c)A_{22} = 1$$

where

$$
\begin{aligned}
A_{11} &= \int_0^\infty I_c(x)\Delta(x)\varphi(x)\,dx - \beta c,\\
A_{12} &= \int_0^\infty \Delta_c(x)\Delta(x)\varphi(x)\,dx - \beta\Delta(c),\\
A_{21} &= 2[\Phi(c) - 0.5],\\
A_{22} &= 2\int_0^c \Delta_c'(x)\varphi(x)\,dx = 2B[\Phi(c+B) + \Phi(c-B) - 1].
\end{aligned}
$$

with $\beta = \epsilon/(1-\epsilon)$. Clearly, $A_{21} > 0$ and $A_{22} > 0$. Also $A_{12} > 0$ because

$$\int_0^\infty I_{c_1}(x)\Delta(x)\varphi(x)\,dx - \beta c_1 = 0.$$

and $c < c_1$ implies that $I_c(x)/c \geq I_{c_1}(x)/c_1$ for all $x \geq 0$ with strict inequality for $0 < x < c$. Analogously, using the fact that $c > c_2$ implies $\Delta_c(x)/c \leq \Delta_{c_2}(x)/c_2$, for all $x \geq 0$ with strict inequality for $0 < x < c$, together with $\int_0^\infty \Delta_{c_2}(x)\Delta(x)\varphi(x)\,dx - \beta\Delta(c_2) = 0$, one concludes that $A_{12} < 0$.

Therefore, for all $c_2 < c < c_1$ we have $\alpha_1(c) = -A_{11}/[A_{11}A_{22} - A_{12}A_{21}] > 0$ and $\alpha_2(c) = -\alpha_1(c)A_{11}/A_{12} > 0$. Moreover, since $\psi_{c_1}(x) = k_1 I_{c_1}(x)$ and $\psi_{c_2}(x) = k_2\Delta_{c_2}(x)$ and since $\text{GES}(\psi_c)$ is a continuous function of c, there exist $c_1 < c^* < c_2$ such that $\text{GES}(\psi) = \text{GES}(\psi_{c^*})$. Let

$$\theta(x) = \alpha_1(c^*)I(x) + \alpha_2(c^*)\Delta(x).$$

14

Notice that $\theta(x)$ is strictly increasing and that $\theta(x) > \psi(x)$ for all $x > c^*$ [because $\theta(c^*) = \alpha_1(c^*)I(c^*) + \alpha_2(c^*)\Delta(c^*) = \mathrm{GES}(\phi_{c^*}) = \mathrm{GES}(\psi) = \psi(\infty)$]. Thus

$$\int_0^\infty [\psi(x) - \theta(x)]^2 \varphi(x)\, dx \geq \int_0^\infty [\phi_{c^*}(x) - \theta(x)]^2 \varphi(x)\, dx.$$

The theorem follows now because

$$\int_0^\infty [\psi(x) - \theta(x)]^2 \varphi(x)\, dx = J(\psi) + d$$

and

$$\int_0^\infty [\phi_{c^*}(x) - \theta(x)]^2 \varphi(x)\, dx = J(\psi_{c^*}) + d,$$

with $d = \int_0^\infty \theta^2(x)\varphi(x)\, dx + 2[\alpha_1(c^*) - \alpha_2(c^*)\beta \mathrm{GES}(\psi)]$.

3.4 Numerical Results

The numerical calculation of the optimal ψ^* is done as follows. For a given value of ϵ ($\epsilon = 0.05$, say) the constants c_1 and c_2 are determined by solving the non-linear equations

$$(1 - \epsilon) \int_0^{c_1} x\Delta(x)\varphi(x)\, dx + (1 - \epsilon)c_1[\Phi(c_1 + B) - \Phi(c_1 - B)] - c_1\epsilon = 0$$

and

$$(1 - \epsilon)\left\{ \int_0^{c_2} \Delta^2(x)\varphi(x)\, dx + \Delta(c_2)\left[\Phi(c_2 + B) - \Phi(c_2 - B)\right] \right\} - \Delta(c_2)\epsilon = 0.$$

Our numerical results show that in general $c_1 > c_2$.

Next, using the fact that for all ψ, $\psi(x)$ and $k\psi(x)$ determine the same location estimate, it follows that the optimal ψ^* is a truncation at c of the function $\alpha x + (1 - \alpha)\Delta(x)$, for some $0 \leq \alpha \leq 1$ and $c_2 \leq c \leq c_1$.

This psi-function is denoted by $\psi_c(x)$ and for each c, the value of α is determined using $C2$, i.e., solving the linear equation

$$(1 - \epsilon) \int_0^c [\alpha x + (1 - \alpha)\Delta(x)]\, \Delta(x)\varphi(x)\, dx +$$

15

$$[\alpha c + (1 - \alpha)\Delta(c)] \{[\Phi(c - B) - \Phi(c + B)](1 - \epsilon) - \epsilon\} = 0.$$

Finally, the asymptotic variance of ψ_c, $VAR(\psi_c, \Phi)$ can be computed. This is done on a fine grid of values of c in $[c_2, c_1]$ $\{c^{(1)}, c^{(2)}, \ldots, c^{(m)}\}$ and c^* is approximated by the grid point that minimizes $VAR(\psi_c, \Phi)$. In general, as shown by Figure 2, the optimal $\psi^*(x)$ is very well approximated by the Huber's function $\psi_{c_1}(x)$ having the same asymptotic bias B, even for values of ϵ near the breakdown point one half.

Figure 3 shows the maximum bias curve of the optimal ψ^* for several efficiencies. Notice that a significant increase in efficiency can be obtained in exchange for a fairly small increase in bias.

Finally we present some numerical results which allow a direct comparison of our exact approach with that based on the GES linear approximation. Suppose that, for a given value of ϵ, we want to choose a robust location estimate according to the following criterion: among all the location estimates T which have a bias-deficiency of up to 10%, i.e. all T for which

$$\frac{B_T(\epsilon)}{B_{MED}(\epsilon)} - 1 = \frac{B_T(\epsilon)}{\Phi^{-1}\left[\frac{1}{2(1-\epsilon)}\right]} - 1 = 0.1$$

choose the T that minimizes the asymptotic variance under the Gaussian model. Note from Figure 2 that a bias deficiency of 10% corresponds to an efficiency constraint in the range 85%-90%.

According to the GES approach one would first approximate $B_T(\epsilon)$ by $\epsilon \cdot GES(T)$. By Hampel's optimality result we can restrict attention to M-estimates with Huber's ψ-functions ψ_c. Then we just need to choose the tuning constant c to achieve the 10% bias deficiency. Since $GES(\psi_c) = c/(2\Phi(c) - 1)$ the constant c is determined by the

non-linear equation

$$\frac{c\epsilon}{2\Phi(c) - 1} = 1.1\Phi^{-1}\left[\frac{1}{2(1 - \epsilon)}\right].$$

On the other hand, following the exact global approach one would set the maximum bias B equal to $1.1\Phi^{-1}[.5/(1 - \epsilon)]$ and choose the optimal ψ-function given by Theorem 2 for such value of B. In view of Figure 2 a good approximation for the optimal ψ can be obtained by restricting attention to ψ-functions of the Huber type. In this case the tuning constant c is determined by solving the equation

$$B_{\psi_c}(\epsilon) = 1.1\Phi^{-1}\left[\frac{1}{2(1 - \epsilon)}\right].$$

Table 1 gives the values of the constant c obtained by the GES and exact approaches as well as the corresponding bias deficiencies. Notice that the values of c given by the exact approach does not change much with ϵ (differences only occurred in the third decimal case) and that the resulting estimate is fairly efficient (eff = 0.85). On the other hand, the values of c given by the GES approximation varies considerably and tends to be disturbingly unconservative, particularly for moderate to large values of ϵ. For example, the actual bias deficiency of the estimates chosen according to the GES approach are 17% for $\epsilon = 0.05$, 22% for $\epsilon = 0.10$, 30% for $\epsilon = 0.15$ and 63% for $\epsilon = 0.30$, instead of the nominal 10%.

4. EFFICIENCY CONSTRAINED SOLUTION WITH SCALE UNKNOWN

When the scale is unknown the side constraint C2 must be replaced by C3

$$(1 - \epsilon)g(B, \bar{s}) = \epsilon\psi(\infty)$$

and consequently the function $\Delta(x)$ must be replaced by

$$\tilde{\Delta}(x) = \frac{\varphi(\overline{s}x - B) - \varphi(\overline{s}x + B)}{\varphi(x)}.$$

Unfortunately, $\tilde{\Delta}(x)$ is no longer monotone (notice that $\lim_{x \to \infty} \tilde{\Delta}(x) = 0$) and so Lemma 3 cannot be applied to ensure the existence of a scaled and truncated version of $\tilde{\Delta}(x)$ which satisfies C1 and C3. However, numerical calculations indicate that the result of Lemma 3 still holds in this case, that is, that there exist constants k. c_3 and c_3' (which depend on ϵ) such that the function

$$\psi_3(x) = \begin{cases} k\tilde{\Delta}(x), & |x| \le c_3 \ \text{ or } \ |x| \ge c_3' \\ k\tilde{\Delta}(c_3) & c_3 < |x| < c_3' \end{cases}$$

satisfies C1 and C3. It can also be proved, using a similar argument as in the proof of Theorem 1, that the function ψ_3 has the same property as the function ψ_2 in that theorem:

$$GES(\psi) \ge GES(\psi_3) \quad \text{implies} \quad VAR(\psi, \Phi) \ge VAR(\psi_3, \Phi)$$

for all $\psi \in \Psi$. Thus, as in the known-scale case, attention can be restricted to psi-functions $\phi_c(x)$ which are truncated and rescaled versions of

$$\phi(x) = \alpha x + (1 - \alpha)\tilde{\Delta}(x) \quad , \quad 0 \le \alpha \le 1$$

with truncation constant c between c_1 and c_3. Here c_1 is the value of the tuning constant of the Huber's psi-function satisfying C1 and C3, and its existence follows from Lemma 3.

If $0 \le \alpha \le 1$ is such that $\phi(x) \ge \phi(c)$ for all $x \ge c$ then

$$\phi_c(x) = \begin{cases} \phi(x), & |x| \le c \\ \phi(c) & |x| > c \end{cases}$$

18

On the other hand if α is such that $\phi(x)$ is not monotone, as in Figure 4, then

$$\phi_c(x) = \begin{cases} \phi(x), & |x| \leq c \ \text{or} \ c' \leq |x| \leq c'' \\ \phi(c) & c \leq |x| \leq c' \ \text{or} \ |x| \geq c'' \end{cases}$$

Unfortunately, the lack of monotonicity of $\tilde{\Delta}(x)$ makes the optimality problem much more involved and one must resort to a combination of analytical derivations and numerical calculations to obtain the optimal ψ^*. The main conclusion from our calculations is that, as in the known scale case, Huber's psi-function ψ_1 with the tuning constant c^* determined by the condition C3 is an excellent approximation to the optimal ψ^*.

Evidently, $c^* = c^*(\epsilon, \bar{s})$, depends on the fraction of contamination ϵ and the maximum value $\bar{s}$ of the asymptotic scale functional. In fact, it can be easily verified using the identity $\psi_c(x/s) = (1/s)\psi_{cs}(x)$ valid for Huber's psi-functions for all $s > 0$, that $c^* = c^*(\epsilon, \bar{s}) = c(\epsilon)/\bar{s}$, where $c(\epsilon)$ is the value of the tuning constant for the nearly optimal Huber's psi-function in the scale-known case. Thus, the tuning constant c^* for the nearly optimal Huber's psi-function is larger when the maximum asymptotic functional $\bar{s}$ of the scale estimate s_n is smaller. Since for Huber's psi-functions ψ_c the asymptotic variance $VAR(\psi_c, \Phi)$ is a decreasing function of c, it becomes evident that the degree of unconstrained bias-robustness of the scale estimate $\bar{s}$ will have an impact on the optimal bias-robust location estimate subject to an efficiency constraint. Therefore, according to the results in Martin and Zamar (1990), an appropriate choice of the scale estimate $\hat{s}_n$ is given by the shorth (see Andrews et. al. 1972) which is nearly optimal bias-robust among M-estimates of scale with breakdown-point equal to one half.

5 . CONCLUDING REMARKS

It has been correctly pointed out by an anonymous referee that the results of this paper amount to a rigorous analysis of the correctness of Hampel's heuristic approach in the location setup. Still, one might conjecture by analogy that, in the case of more complicated models, maximizing the efficiency under a constraint on the gross-error-sensitivity is almost the same as putting a constraint on the bias. However, the extension of our technique to more complicated models is by no means straightforward (if al all possible) and this matter deserves further study.

Another interesting issue brought up by an anonymous referee is that if one leaves the realm of M-estimates of location, there is not necessarily a pay-off between bias-robustness and efficiency at the central Gaussian model. This is shown by the following example (also provided by the referee): Let $h : R \to R$ be symmetric, differentiable and non-increasing on $[0, \infty)$ with $h(x) = 1$ for $0 \leq x \leq 1$ and $h(x) = 0$ for $x \geq 2$. Denote the mean and the median of a distribution F by $T_M(F)$ and $T_{MED}(F)$ respectively. For any $\delta > 0$ let

$$T_\delta(F) = T_{MED}(F) + [T_M(F) - T_{MED}(F)] \, h \left[\frac{T_M(F) - T_{MED}(F)}{\delta} \right] ,$$

where $0.\infty = 0$ and $T_M(F) = \infty$ if F doesn't have a finite mean. If the scale also has to be estimated, δ can be replaced by $\delta MAD(F)$.

Since

$$|T_\delta(F) - T_{MED}(F)| \leq C\delta \quad , \quad C = \sup_{x \geq 0} x h(x) \leq 2$$

the maximum bias of $T_\delta(F)$ exceeds that of the median by at most 2δ. If F is such that $T_M(F) - T_{MED}(F) = 0$ (e.g. F is symmetric, with finite mean) then

$$T_\delta(F_n) = T_M(F_n) \quad , \quad \text{a.s. } [F],$$

20

for n sufficiently large. Therefore, $T_\delta(F_n)$ has an asymptotic efficiency of 1 when $F = N(0,1)$ but an efficiency of 0 if F has infinite variance.

The following argument shows that there are some merits in considering the optimality problem on the class of M-estimates of location. It is well known that bias-robust M-estimates of location, that is, M-estimates with a bounded ψ, are variance-robust: the supremum of their asymptotic variance over *symmetric* ϵ-contamination neighborhoods of the Gaussian model is finite. This suggests that the efficient bias-robust M-estimate is not only relatively efficient at the central Gaussian model but also over a symmetric neighborhood of it. In fact, when the scale parameter is known one can easily verify that

$$\overline{AV}_\epsilon(\psi) = \sup_F AV(\psi, F) = \frac{(1 - \epsilon)\mathrm{E}_\Phi \psi^2(X) + \epsilon \psi^2(\infty)}{[(1 - \epsilon)\mathrm{E}_\Phi \psi'(X)]^2}.$$

Using this formula one can check that the supremum variance of the Huber's M-estimate of location with $c = 1.345$ (for instance) is fairly small: $\overline{AV} = 1.053$ when $\epsilon = 0$, $\overline{AV} = 1.257$ when $\epsilon = .05$ and $\overline{AV} = 1.795$ when $\epsilon = .15$.
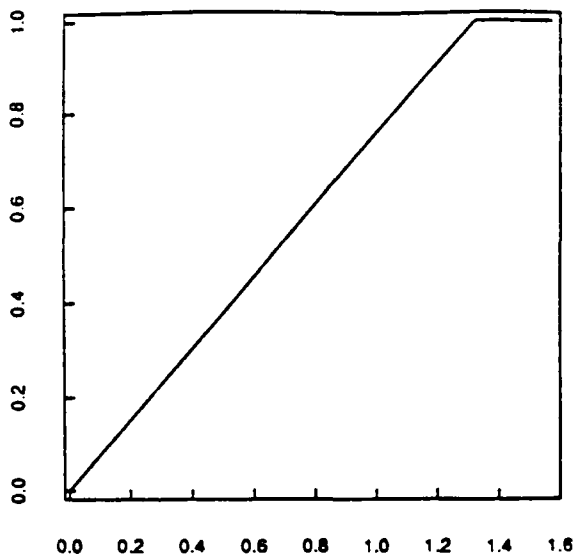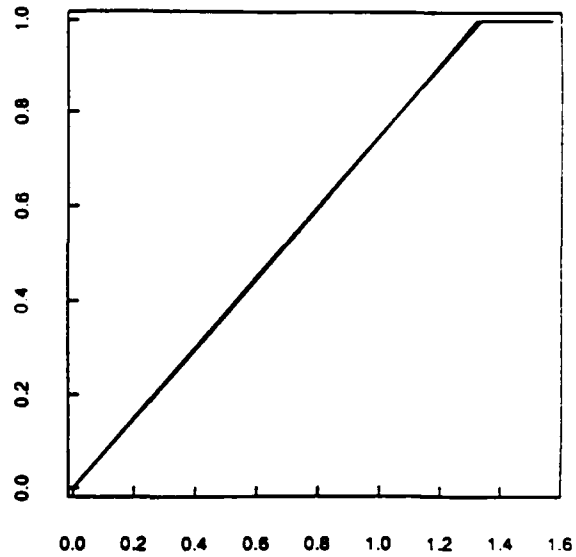
Figure 1: MAXIMUM BIAS CURVES AND LINEAR GES APPROXIMATION
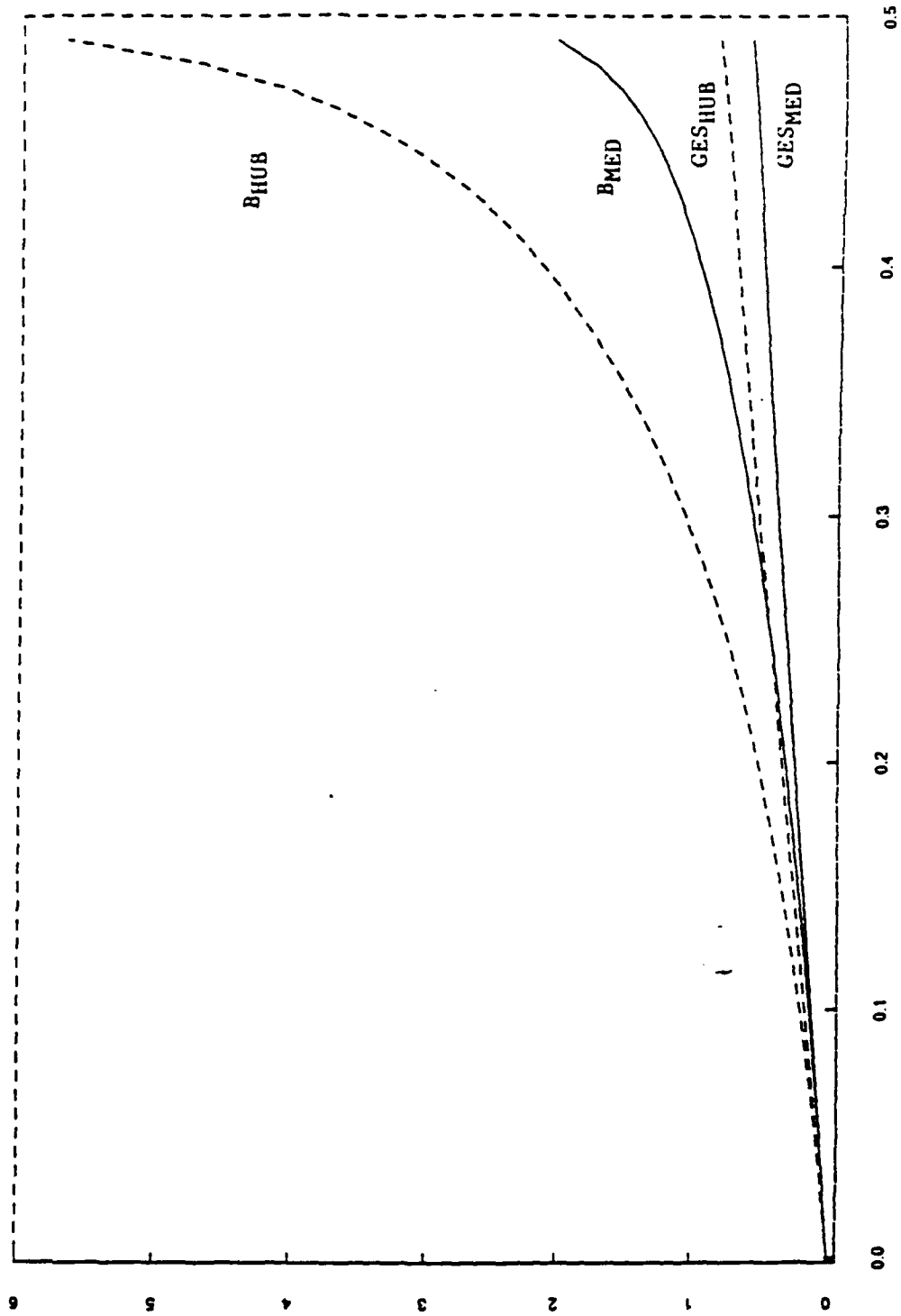
The dashed curves correspond to
the the maximum bias curve and GES
approximation for the efficient Huber estimate (c~1.5). The solid
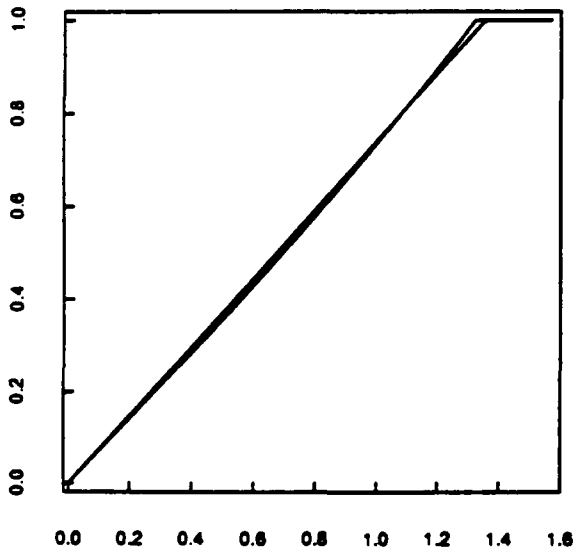curves correspond to the median.

Figure 2: OPTIMAL ι-FUNCTIONS AND CORRESPONDING NEARLY OPTIMAL HUBER APPROXIMATES
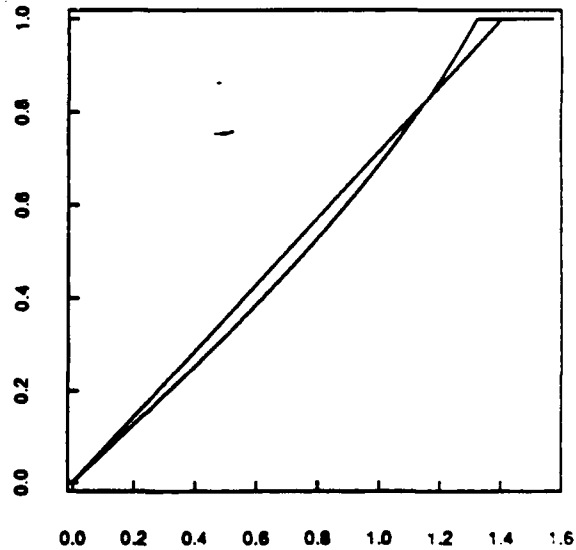

(a) EPSILON = .1, Bias = .175, ALPHA = .63
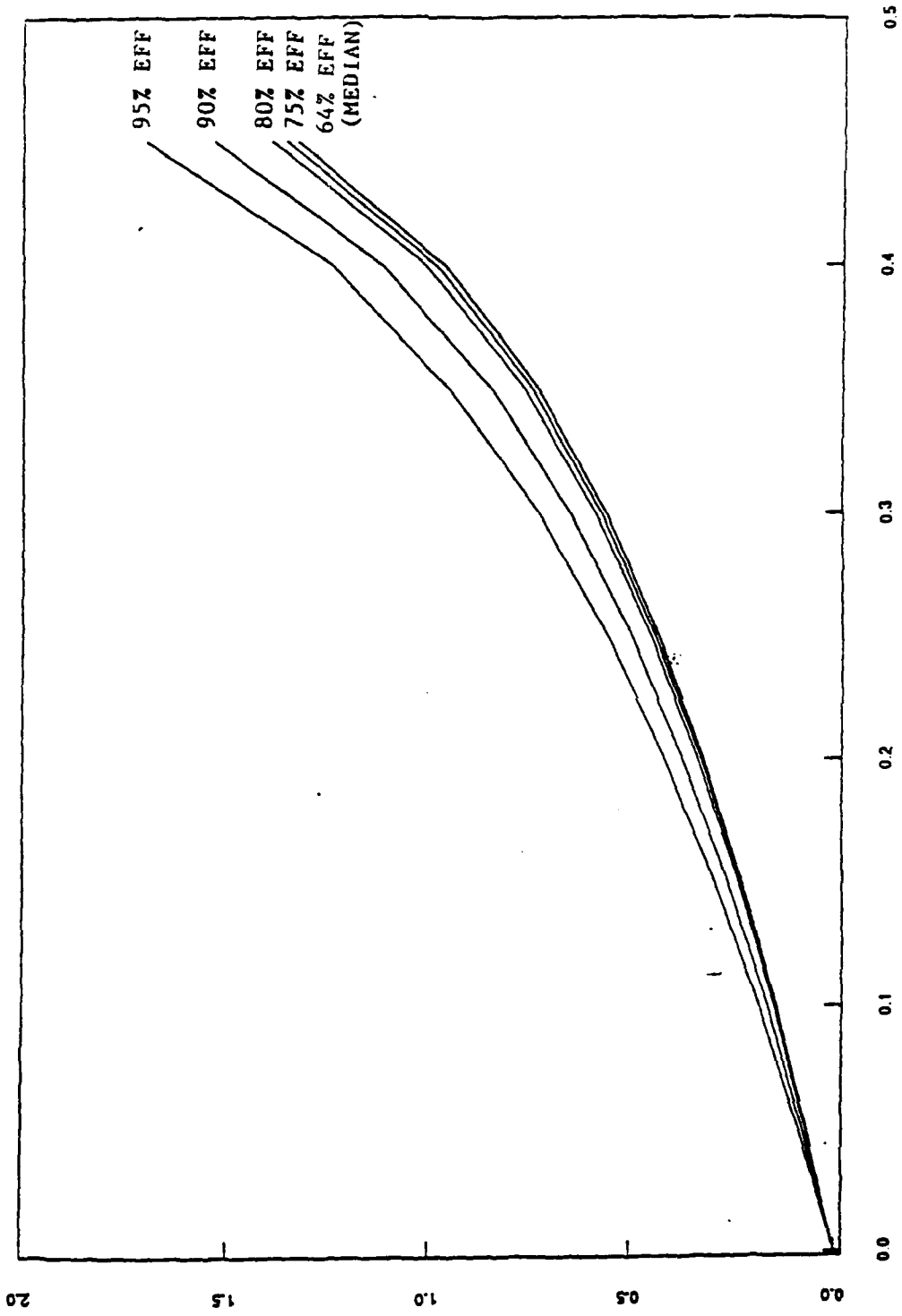
(b) EPSILON = .3, Bias = .71, ALPHA = .86
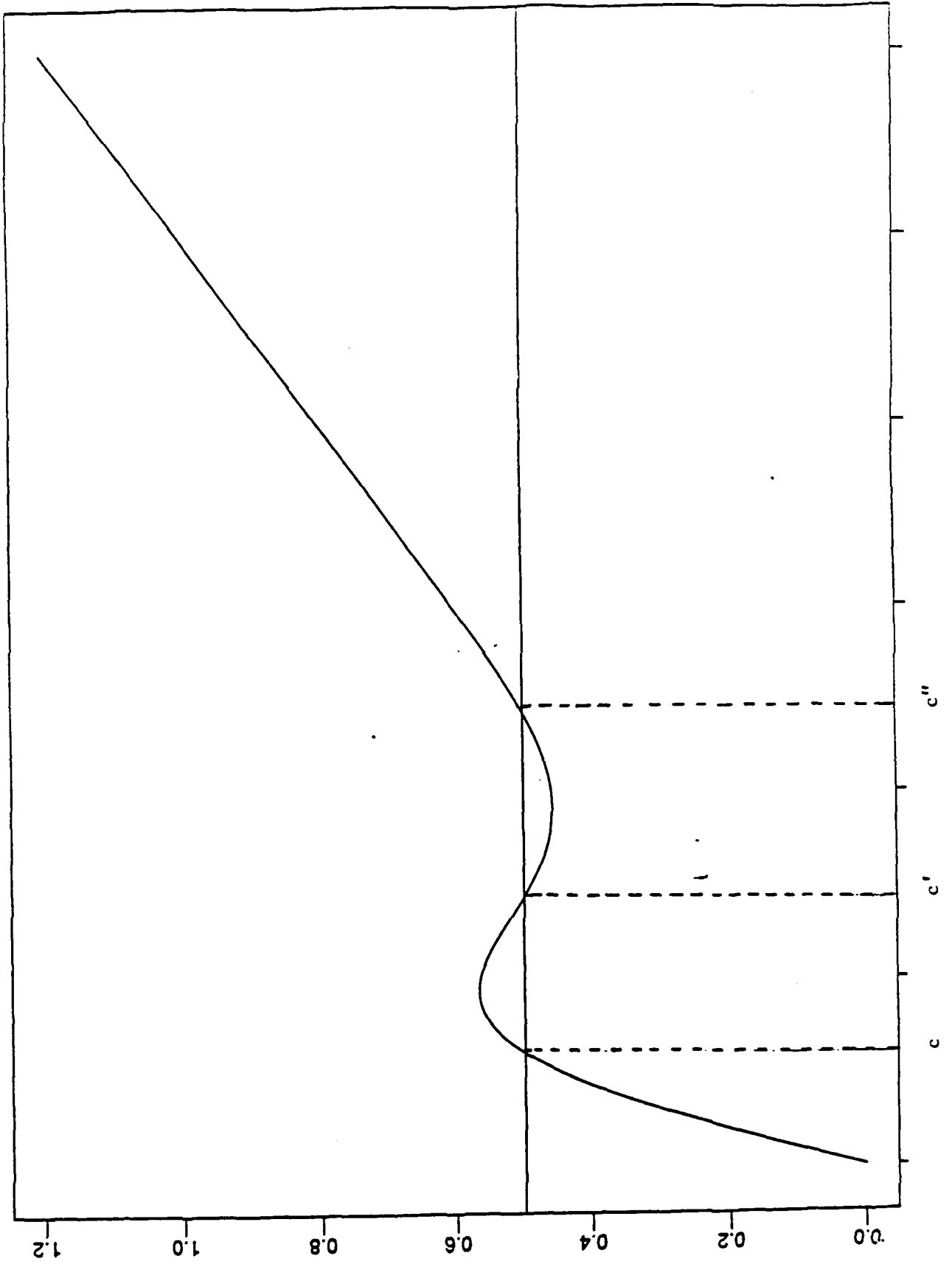
(c) EPSILON = .4, Bias = 1.21, ALPHA = .88

(d) EPSILON = .49, Bias = 2.58, ALPHA = .76

Figure 3: MAXIMUM BIAS CURVES OF SEVERAL BIAS-ROBUST ESTIMATES

The lowest curve corresponds to the median and the highest corresponds to the 95% efficient bias-robust estimate. The intermediate ones correspond to the 75%, 80% and 90% efficient bias-robust estimates.

Figure 4: REDESCENDING φ-FUNCTION

REFERENCES

Andrews, D.F., Bickel, P.J., Hampel, F.R., Roger, W.H. and Tukey, J.W. (1972) "Robust estimates of location: Survey and advances". Princeton University Press. Princeton. N.J.

Donoho, D.L. and Liu, R.C. (1988 a). "The automatic robustness of distance functionals". Ann. Statist., **2**, 552-586.

Donoho, D.L. and Liu, R.C. (1988 b). "Pathologies of some minimum distance estimators". Ann. Statist., **2**, 587-608.

Hampel, F.R. (1968). "Contributions to the theory of robust estimation". Ph.D. thesis, University of California, Berkeley.

Hampel, F.R. (1974). "The influence curve and its role in robust estimation". J. Am. Statist. Assoc., **69**, 383-393.

Hampel, F.R., Ronchetti, E.M., Rousseeuw, P.J. and Stahel, W.A. (1986). "Robust estatistics: The approach based on influence". Wiley, N.Y.

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

Huber, P.J. (1981). "Robust statistics". Wiley, N.Y.

Maronna, R.A. and Yohai, V.J. (1989). "The maximum bias of robust covariances". Comm. in Statist., Theory and Methods. To appear.

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

Martin, R.D. and Zamar, R.H. (1989). "Asymptotically min-max bias robust M-estimates of scale for positive random variables". J. Am. Statist. Assoc., **84**, 406-501.

Martin, R.D. and Zamar, R.H. (1990). "Bias robust estimates of scale when location

is unknown". Tech. Rep., Department of Statistics, University of Washington.

Rousseeuw, P.J. (1984). "Least median of squares regression". J. Am. 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, N.Y., 256-272

Yohai, V.J. (1990). "A new class of bias-robust regression estimates". Tech. Rep., Department of Statistics, University of Washington.

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER 217	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) Efficiency Constrained Bias Robust Estimation of Location		5. TYPE OF REPORT & PERIOD COVERED TR 10-1-90 to 9-30-93
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) R. Douglas Martin Ruben H. Zamar		8. CONTRACT OR GRANT NUMBER(s) N00014-91-J-1074
9. PERFORMING ORGANIZATION NAME AND ADDRESS Department of Statistics, GN-22 University of Washington Seattle, WA 98105		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS NR-661-003
11. CONTROLLING OFFICE NAME AND ADDRESS ONR Code N63374 1107 NE 45th Street Seattle, WA 98105		12. REPORT DATE October 1989
		13. NUMBER OF PAGES 32
14. MONITORING AGENCY NAME & ADDRESS(If different from Controlling Office)		15. SECURITY CLASS. (of this report) Unclassified
		15a. DECLASSIFICATION/ DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

APPROVED FOR PUBLIC RELEASE: DISTRIBUTION UNLIMITED.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Bias-robustness, minimax, efficiency

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

In 1964 P. Huber established the following minimax bias robustness result for estimating the location μ in the ϵ-contamination family $F(x) = (1 - \epsilon)\Phi[(x - \mu)/s] + \epsilon H(x)$ where Φ is the standard normal distribution and H is an arbitrary distribution function: The median minimizes the maximum asymptotic bias among all translation equivariant estimates of location. However, the median efficiency of $2/\pi$ at the Gaussian model may be unacceptably low in some applications. This motivates one to solve the following problem for the above ϵ-contamination family: Among all location M-estimates, find the one which minimizes the

DD FORM 1473 EDITION OF 1 NOV 65 IS OBSOLETE
1 JAN 73

S/N 0102-LF-014-6601

maximum asymptotic bias subject to a constraint on efficiency at the Gaussian model. This problem is the dual form analog of Hampel's optimality problem of minimizing the asymptotic variance at the nominal model (e.g. the Gaussian model) subject to a bound on the gross-error sensitivity. We solve the global problem completely for the case of a known scale parameter. The main conclusion is that Hampel's heuristic is essentially correct: the resulting M-estimate is based on a psi-function which is amazingly close, but not exactly equal, to the Huber/Hampel optimal ψ. It turns out that one pays only a relatively small price in terms of increase in maximal bias for increasing efficiency from 64% to the range $90\% - 95\%$. We also present a conjectured solution to the problem, based on heuristic arguments and numerical calculations, when the nuisance scale parameter is unknown.