*The Annals of Statistics*
2001, Vol. 29, No. 1, 194–223

# OPTIMAL ROBUST *M*-ESTIMATES OF LOCATION

BY RICARDO FRAIMAN[1], VÍCTOR J. YOHAI[2] AND RUBEN H. ZAMAR[3]

*University of San Andrés, University of Buenos Aires and CONICET and University of British Columbia*

We find optimal robust estimates for the location parameter of $n$ independent measurements from a common distribution $F$ that belongs to a contamination neighborhood of a normal distribution. We follow an asymptotic minimax approach similar to Huber's but work with full neighborhoods of the central parametric model including nonsymmetric distributions. Our optimal estimates minimize monotone functions of the estimate's asymptotic variance and bias, which include asymptotic approximations for the quantiles of the estimate's distribution. In particular, we obtain robust asymptotic confidence intervals of minimax length.

**1. Introduction.** In his seminal paper, Huber (1964) initiated two main approaches of modern robustness theory: variance robustness and bias robustness. He considered the location model and introduced the family of location $M$-estimates. In this context he found that the median minimizes the maximum asymptotic bias over contamination neighborhoods among translation equivariant estimates. He also found the $M$-estimate that minimizes the maximum asymptotic variance over contamination neighborhoods which include only symmetric distributions.

Hampel (1974) made a first step toward considering variance and bias simultaneously: he found the $M$-estimate that minimizes the asymptotic variance at the central model, subject to a bound on the gross error sensitivity (GES). Observe that the GES is the slope of the linear approximation (at zero) for the maximum asymptotic bias over $\varepsilon$-contamination neighborhoods. Martin and Zamar (1993) improved Hampel's result by minimizing the variance at the central model subject to a bound on the actual maximum asymptotic bias.

Hampel (1974) and Martin and Zamar (1993) take into account the effect of asymmetric contaminations on the bias of the estimate but ignore their effect on the variance. Samarov (1985) attempts to integrate asymptotic bias and variance under infinitesimal contaminations by minimizing a linear approximation to the maximum asymptotic mean squared error (in the regression set-up). Samarov used the linear approximation based on the influence and

Received December 1998; revised October 2000.

[1]Supported in part by Grants TX49 from University of Buenos Aires; PIP 4186/96 from CONICET; and PICT-97 03-00000-00576 from FONCYT.

[2]Supported in part by Grants TW92 from University of Buenos Aires; PIP 4186/96 from CONICET; and PICT-97 03-00000-00576 from FONCYT.

[3]Supported in part by the Natural Sciences and Engineering Research Council of Canada (NSERC).

*AMS* 2000 *subject classification*. Primary 62F35.

*Key words and phrases*. $M$-estimates, robust location, minimax intervals.

change of variance functions, which is valid only for vanishingly small $\varepsilon$. For other infinitesimal approaches, see Rieder (1994).

We wish to take into account the simultaneous effect of asymmetric contamination on the asymptotic bias and variance of a location $M$-estimate, $\hat{\mu}_n$, for nonvanishingly small fractions of contamination. To achieve this goal we consider general monotone functions of the asymptotic bias and variance of $\hat{\mu}_n$ and derive the corresponding minimax estimate over $\varepsilon$-contamination neighborhoods. In particular, our theory leads to robust confidence intervals of minimax length.

The rest of the paper is organized as follows. In Section 2 we introduce some definitions and technical background. We also give some asymptotic results and define a general minimax problem for location $M$-estimates which includes Huber's minimax bias problem and an extension of Huber's minimax variance problem as particular cases. In Section 3 we discuss the problem of robust confidence intervals and show that our general theory yields robust confidence interval of minimax length. In Section 4 we solve the general minimax problem posed in Section 2. In Section 5 we consider the case when the scale parameter is unknown. All the proofs are collected in the Appendix.

**2. Some definitions and technical background.** Suppose that $y_1, \dots, y_n$ are independent identically distributed random variables with common distribution $F$ in the contamination neighborhood

$$\mathscr{V}(F_0, \varepsilon) = \{F\colon F = (1-\varepsilon)F_0 + \varepsilon H, H \text{ arbitrary}\}, \tag{2.1}$$

where $F_0(y) = \Phi((y-\mu_0)/\sigma_0)$, $\Phi$ is the standard normal distribution and $0 < \varepsilon < 0.5$. We will first assume that the scale parameter $\sigma_0$ is known and consider location $M$-estimates, $\hat{\mu}_n$, satisfying the equation

$$\sum \psi\left(\frac{y_i - \hat{\mu}_n}{\sigma_0}\right) = 0, \tag{2.2}$$

where $\psi$ is an appropriate score function [see (A1)–(A3) below].

Let

$$\eta(t, F, \psi) = -\int_{-\infty}^{\infty} \psi\left(\frac{y-t}{\sigma_0}\right) dF(y). \tag{2.3}$$

Then under very general conditions [see Huber (1981)] $\hat{\mu}_n$ converges to the value $T(F, \psi)$ satisfying $\eta(T(F, \psi), F, \psi) = 0$.

We will consider the following assumptions on $\psi$:

A1. $\psi$ is continuous, non-decreasing, odd and bounded with $\lim_{y\to\infty} \psi(y) > 0$.

A2. $\psi$ has an uniformly continuous derivative $\psi'$.

A3. There exists a constant $c > 0$ such that (a) $\psi(y)$ is constant outside $(-c, c)$; (b) $\psi$ has a derivative $\psi'$ on $(-c, c)$; (c) $\psi$ has a left (right) derivative at $c$ $(-c)$ denoted $\psi'(c)(\psi'(-c))$; (d) the function $\psi'$ is continuous on $[-c, c]$.

Let $\mathscr{C}_1$ be the class of score functions $\psi$ that satisfy A1 and A2, let $\mathscr{C}_2$ be the class of score functions $\psi$ that satisfy A1 and A3 and set $\mathscr{C} = \mathscr{C}_1 \cup \mathscr{C}_2$.

Let $\psi \in \mathscr{C}$ and let $c$ be the constant introduced in A3 for functions in $\mathscr{C}_2$. If $\psi$ is differentiable everywhere set $c = \infty$. Define

$$v(F, \psi) = \sigma_0^2 \frac{\mathrm{E}_F \psi^2((y - T(F, \psi))/\sigma_0)}{[\mathrm{E}_F \psi'((y - T(F, \psi))/\sigma_0) I_{(-c,c)}((y - T(F, \psi))/\sigma_0))]^2}. \tag{2.4}$$

Notice that $v(F, \psi)$ is well defined for all $\psi \in \mathscr{C}$ and all $F \in \mathscr{V}(F_0, \varepsilon)$. Part (a) of the following lemma shows that $T(F, \psi)$ is uniquely defined on $\mathscr{V}(F_0, \varepsilon)$. Parts (b) and (c) show that $T(F, \psi)$ and $v(F, \psi)$ are bounded on $\mathscr{V}(F_0, \varepsilon)$.

LEMMA 1. *Assume that $\psi$ satisfies* A1 *and either* A2 *or* A3. *Then:*

(a) *The equation $\eta(t, F, \psi) = 0$ has an unique solution $T(F, \psi)$ for all $F \in \mathscr{V}(F_0, \varepsilon)$.*

(b) $\sup_{F \in \mathscr{V}(F_0, \varepsilon)} |T(F, \psi) - \mu_0| = \lim_{y \to \infty} T((1 - \varepsilon)F_0 + \varepsilon\delta_y, \psi) - \mu_0 < \infty$.

(c) $\sup_{F \in \mathscr{V}(F_0, \varepsilon)} |v(F, \psi)| < \infty$.

Part (a) of the following theorem shows that if $\psi \in \mathscr{C}_1$ then $\hat{\mu}_n$ is asymptotically normal with variance $v(F, \psi)$, uniformly on $\mathscr{V}(F_0, \varepsilon)$. Part (c) establishes a similar result for $\psi \in \mathscr{C}_2$, but now the uniformity holds on some restricted class $\mathscr{V}_k(F_0, \varepsilon)$ of distributions $F = (1 - \varepsilon)F_0 + \varepsilon H$, where $H$ has its support outside the interval $[-k, k]$. Part (b) shows that $v(F, \psi)$ has a sound interpretation even if $\hat{\mu}_n$ fails to be asymptotically normal.

THEOREM 1. *Let $\hat{\mu}_n$ be an $M$-estimate satisfying* (2.2).

(a) *Suppose that $\psi \in \mathscr{C}_1$. Then*

$$\lim_{n\to\infty} \sup_{F \in \mathscr{V}(F_0,\varepsilon)} \left| P_F\left( \frac{\sqrt{n}(\hat{\mu}_n - T(F, \psi))}{\sqrt{v(F, \psi)}} \le a \right) - \Phi(a) \right| = 0, \tag{2.5}$$

*uniformly in $a$.*

(b) *Suppose that $\psi \in \mathscr{C}_2$. Then*

$$\lim_{n\to\infty} \inf_{F \in \mathscr{V}(F_0,\varepsilon)} P_F\left( \left| \frac{\sqrt{n}(\hat{\mu}_n - T(F, \psi))}{\sqrt{v(F, \psi)}} \right| \le a \right) \ge 2\Phi(a) - 1, \tag{2.6}$$

*uniformly in $a$.*

(c) *Suppose that $\psi \in \mathscr{C}_2$. Let $\mathscr{V}_k(F_0, \varepsilon)$ be the set of distribution functions $F = (1 - \varepsilon)F_0 + \varepsilon H$, such that $H$ has its support outside the interval $[-k, k]$. Then there exists $0 < k < \infty$ such that*

$$\lim_{n\to\infty} \sup_{F \in \mathscr{V}_k(F_0,\varepsilon)} \left| P_F\left( \frac{\sqrt{n}(\hat{\mu}_n - T(F, \psi))}{\sqrt{v(F, \psi)}} \le a \right) - \Phi(a) \right| = 0,$$

*uniformly in $a$.*

GENERAL OPTIMALITY PROBLEM. The asymptotic bias of $\hat{\mu}_n$ for $F \in \mathscr{V}(F_0, \varepsilon)$ is given by

$$b(F, \psi) = T(F, \psi) - \mu_0. \tag{2.7}$$

One expects robust estimates to have relatively small asymptotic bias and variance as $F$ ranges over $\mathscr{V}(F_0, \varepsilon)$. Therefore, the overall performance of $\hat{\mu}_n$ can be assessed in terms of the *maximum asymptotic bias* (maxbias) $B_\psi(\varepsilon) = \sup_{F \in \mathscr{V}(F_0, \varepsilon)} |b(F, \psi)|$ and the *maximum asymptotic variance* (maxvariance) $V_\psi(\varepsilon) = \sup_{F \in \mathscr{V}(F_0, \varepsilon)} v(F, \psi)$. It can be shown [see Martin and Zamar (1993)] that $B_\psi(\varepsilon) = \lim_{y \to \infty} b((1-\varepsilon)F_0 + \varepsilon\delta_y, \psi)$ and $B_\psi(\varepsilon)$ satisfies the equation

$$\eta(t + \mu_0, F_0, \psi) = \frac{\varepsilon\psi(\infty)}{1-\varepsilon}. \tag{2.8}$$

The performance of $\hat{\mu}_n$ on $\mathscr{V}(F_0, \varepsilon)$ can be assessed by

$$L_g(\psi) = \sup_{F \in \mathscr{V}(F_0, \varepsilon)} g\left(b(F, \psi), \frac{v(F, \psi)}{n}\right), \tag{2.9}$$

where $g(x_1, x_2)$ has the following properties:

C1. $g$ is lower semicontinuous.
C2. $g$ is nonnegative.
C3. $g(-x_1, x_2) = g(x_1, x_2)$.
C4. $g$ is nondecreasing in $|x_1|$ and $x_2$.
C5. $g(\sigma x_1, \sigma^2 x_2) = \sigma^m g(x_1, x_2)$ for some $m > 0$.

These properties will imply that

$$L_g(\psi) = \sigma_0^m \sup_{F \in \mathscr{V}(\Phi, \varepsilon)} g\left(b(F, \psi), \frac{v(F, \psi)}{n}\right). \tag{2.10}$$

A score function $\psi^*$ minimizing $L_g(\psi)$ is called optimal. The function $L_g(\psi)$ as well as the optimal score function $\psi^*$ depend on $n$. However, for simplicity, this dependence will be omitted from the notations.

The special cases $g(x_1, x_2) = |x_1|$ and $g(x_1, x_2) = x_2$ were considered by Huber (1964). In fact, he showed that the median, that is, the $M$-estimate with score function $\psi(y) = \operatorname{sign}(y)$, minimizes $B_\psi(\varepsilon)$, for all $\varepsilon$, among all translation equivariant location estimates. Huber also considered the restricted maximum asymptotic variance

$$\widetilde{V}_\psi(\varepsilon) = \sup_{F \in \widetilde{\mathscr{V}}(F_0, \varepsilon)} v(F, \psi),$$

where

$$\widetilde{\mathscr{V}}(F_0, \varepsilon) = \{F\colon F = (1-\varepsilon)F_0 + \varepsilon H,\ H \text{ symmetric}\}, \tag{2.11}$$

and showed that the $M$-estimate with score function

$$\psi_c^H(y) = \max\{\min\{y, c\}, -c\} \tag{2.12}$$

minimizes $\widetilde{V}_\psi(\varepsilon)$ among all $M$-estimates satisfying certain mild regularity conditions.

More realistic loss functions take into account the combined effect of contamination on the asymptotic bias and variance of $\hat{\mu}_n$. For example,

$$g(x_1, x_2) = x_1^2 + x_2$$

gives a kind of asymptotic mean squared error. Another interesting family of $g$-functions is

$$g(x_1, x_2) = J_K(x_1) + x_2, \tag{2.13}$$

where $J_K(x_1) = 0$ if $x_1 \le K$ (for some $K > 0$) and $J_K(x_1) = \infty$ otherwise. The corresponding optimal estimates minimize the maximum asymptotic variance subject to a bound $K$ on the maximum asymptotic bias. An interesting choice for the function $g$ is given by (3.1) below, which yields asymptotic quantiles. This case is studied in detail in the next section.

The results of Theorem 4 in Section 3, with $g(x_1, x_2) = x_2$ [$K = \infty$ in (2.13)] improve Huber's (1964) minimax variance result as the maximum is now taken over unrestricted neighborhoods instead of only symmetric ones [see (2.1) and (2.11)]. For the same reason, taking $K < \infty$ we improve the result of Martin and Zamar (1993).

We close this section with a few remarks on the class $\mathscr{C}$ of $\psi$ functions and Theorem 1. As pointed out by Davies (1998), to have a more meaningful asymptotic optimality theory we need that the asymptotic normality of $\hat{\mu}_n$ hold uniformly over $\mathscr{V}(F_0, \varepsilon)$. Theorem 1(a) proves that for $\psi \in \mathscr{C}_1$ [see also Huber (1981) and Davies (1998)]. However, if we restrict attention to $\mathscr{C}_1$ our optimality problems do not have solution because, as we will see later on, the optimal $\psi$ belongs to $\mathscr{C}_2$. To remedy this problem we consider the larger class $\mathscr{C}$. But enlargement of class $\mathscr{C}$ brings along a new problem. If $\psi \in \mathscr{C}_2$ is nondifferentiable at its truncation point $c$ and $F = (1-\varepsilon)F_0 + \varepsilon H$, where $H$ places positive mass at $T(F, \psi) \pm \sigma_0 c$, then the estimate is not asymptotically normal. Theorem 1(b) shows that in such a case the asymptotic distribution of the estimate has tails thinner than those of a $N(T(F, \psi), v(F, \psi)/n)$, uniformly over $\mathscr{V}(F_0, \varepsilon)$. That is, $v(F, \psi)$ still provides in this case a normal probability bound for the estimation error. Theorem 1(c) shows that for $\psi \in \mathscr{C}_2$ the asymptotic normality of $\hat{\mu}_n$ holds uniformly for distributions $F$ which do not place positive mass inside a closed interval. This is particularly important because, as we will see later on, we will work with this type of distribution to obtain a lower bound for $L_g(\psi)$.

**3. Robust confidence intervals.** The need for robust confidence intervals is illustrated by the following small Monte Carlo simulation. We generated ten thousand normal samples of different sizes and containing various fractions of contamination. The contaminating distribution is a point mass distribution at $x = 4.0$. Similar results were found for other asymmetric, outlier generating, distributions. For each sample, we calculated the location $M$-estimate with Huber $\psi$-function (2.12) and the corresponding asymptotic 95

TABLE 1
*Percentage of coverage and average length for* 10,000 *asymptotic* 95% *confidence intervals based on Huber's location* $M$*-estimate with truncation constant* $c = 1.345$*

| $\varepsilon$ | $n$ | % of coverage | Average length |
|---|---|---|---|
| 0.05 | 20 | 92% | 0.91 |
| | 50 | 92% | 0.60 |
| | 100 | 88% | 0.44 |
| | 200 | 82% | 0.31 |
| 0.10 | 20 | 91% | 1.05 |
| | 50 | 84% | 0.68 |
| | 100 | 67% | 0.49 |
| | 200 | 39% | 0.35 |
| 0.15 | 20 | 88% | 1.19 |
| | 50 | 72% | 0.76 |
| | 100 | 35% | 0.56 |
| | 200 | 5% | 0.40 |
| 0.20 | 20 | 82% | 1.41 |
| | 50 | 45% | 0.92 |
| | 100 | 8% | 0.66 |
| | 200 | 0% | 0.47 |

*The simulated data are normal samples containing various fractions of point mass contamination at $x = 4.0$.

confidence interval based on the empirical asymptotic variance. The coverage and average length of these intervals are given in Table 1.

The poor coverage of these intervals based on highly robust $M$-estimates is because they ignore the important issue of asymptotic bias. See also Exhibit 4.2.2 on page 76 of Huber (1981).

Huber (1968) broke new ground on this problem by establishing a remarkable finite sample optimality result. He considered intervals of fixed length $2a$ and minimized the quantity

$$\max_{F \in \mathscr{V}(F_0,\varepsilon)} \max\{P_F(\mu_0 < \hat{\mu}_n - a),\, P_F(\mu_0 > \hat{\mu}_n + a)\}.$$

Notice that although Huber's objective function is not exactly equal to the maximum level

$$\max_{F \in \mathscr{V}(F_0,\varepsilon)} \{P_F(\mu_0 < \hat{\mu}_n - a) + P_F(\mu_0 > \hat{\mu}_n + a)\},$$

it is closely related to it. In principle, the value of $a$ could be varied to obtain the desired maximum level for each $n$ and $\varepsilon$. But the implementation of this idea is by no means straightforward. Huber–Carol (1970) proposed an asymptotic solution under the additional assumption that the fraction of contamination tends to zero at rate $1/\sqrt{n}$.

We wish to derive optimal asymptotic confidence intervals with a warranted maximum level over $\mathscr{V}(F_0, \varepsilon)$. In order to find robust confidence intervals, we propose to use the function $g(x_1, x_2)$ implicitly defined by

$$1 - \alpha = \Phi\left(\frac{g(x_1, x_2) - x_1}{\sqrt{x_2}}\right) + \Phi\left(\frac{g(x_1, x_2) + x_1}{\sqrt{x_2}}\right) - 1. \tag{3.1}$$

For this choice of $g$ we have that

$$q_{\alpha, n}(F, \psi) = g\left(b(F, \psi), \frac{v(F, \psi)}{n}\right) \tag{3.2}$$

[see (2.7) and (2.4)] is an asymptotic $(1-\alpha)$-quantile for $|\hat{\mu}_n - \mu_0|$. On the other hand, the finite sample $(1-\alpha)$th-quantile, $Q_{\alpha,n}(F, \psi)$, satisfies the equation

$$P_F(|\hat{\mu}_n - \mu_0| \le Q_{\alpha, n}(F, \psi)) = 1 - \alpha. \tag{3.3}$$

The following theorem establishes that the asymptotic quantile (3.2) constitutes a good (conservative) approximation for the finite sample quantile (3.3) uniformly on $\mathscr{V}(F_0, \varepsilon)$.

THEOREM 2. *Let* $0 < \alpha < 0.5$ *and* $0 < \varepsilon < 0.5$ *be fixed.*

(a) *Suppose that* $\psi \in \mathscr{C}_1$. *Then*

$$\sup_{F \in \mathscr{V}(F_0, \varepsilon)} |q_{\alpha,n}(F, \psi) - Q_{\alpha,n}(F, \psi)| = o\left(\frac{1}{\sqrt{n}}\right).$$

(b) *Suppose that* $\psi \in \mathscr{C}_2$. *Then* $q_{\alpha,n}(F, \psi)$ *are "conservative" asymptotic* $(1-\alpha)$*-quantiles of* $|\hat{\mu}_n - \mu_0|$. *More precisely,*

$$\sup_{F \in \mathscr{V}(F_0, \varepsilon)} [Q_{\alpha, n}(F, \psi) - q_{\alpha, n}(F, \psi)]^+ = o\left(\frac{1}{\sqrt{n}}\right),$$

*where* $[f]^+$ *is the positive part of* $f$.

(c) *Suppose that* $\psi \in \mathscr{C}_2$ *and let* $\mathscr{V}_k(F_0, \varepsilon)$ *be as defined in Theorem* 1(c). *Then there exists* $0 < k < \infty$ *such that*

$$\sup_{F \in \mathscr{V}_k(F_0, \varepsilon)} |q_{\alpha, n}(F, \psi) - Q_{\alpha, n}(F, \psi)| = o\left(\frac{1}{\sqrt{n}}\right).$$

Since $F$ is only assumed to be an unspecified member of $\mathscr{V}(F_0, \varepsilon)$, it seems appropriate to consider the conservative robust quantile

$$\bar{q}_{\alpha, n}(\psi) = \sup_{F \in \mathscr{V}(F_0, \varepsilon)} q_{\alpha, n}(F, \psi) = \sigma_0 \sup_{F \in \mathscr{V}(\Phi, \varepsilon)} q_{\alpha, n}(F, \psi) = L_g(\psi), \tag{3.4}$$

and the corresponding $1 - \alpha$ robust confidence interval $\hat{\mu}_n \pm \bar{q}_{\alpha,n}(\psi)$. The next theorem shows that, under mild regularity conditions, the coverage of intervals based on $\bar{q}_{\alpha, n}(\psi)$ is correct for all $\psi \in \mathscr{C}$, uniformly on $\mathscr{V}(F_0, \varepsilon)$.

THEOREM 3. *Let* $g(x_1, x_2)$ *and* $\bar{q}_{\alpha,n}(\psi)$ *be defined by* (3.1) *and* (3.4), *respectively. Then*:

(a) *If* $\psi \in \mathscr{C}_1$, *then*

$$\lim_{n\to\infty} \inf_{F\in\mathscr{V}(F_0,\,\varepsilon)} P_F(|\hat{\mu}_n - \mu_0| \le \bar{q}_{\alpha,n}(\psi)) = 1 - \alpha. \tag{3.5}$$

(b) *Assume now that* $\psi \in \mathscr{C}$ *and that* $\sup_{F\in\mathscr{V}(F_0,\,\varepsilon)} v(F, \psi) = \lim_{y\to\infty} v((1 - \varepsilon)F_0 + \varepsilon\delta_y, \psi) \equiv v(F^\infty, \psi)$, *where* $\delta_y$ *is the point-mass distribution at* $y$. *Then* (3.5) *holds and* $\bar{q}_{\alpha,n}(\psi) = g(B_\psi(\varepsilon), v(F^\infty, \psi)/n)$.

In view of (3.4) a natural goal at this point is to find the optimal score function $\psi^*$ which minimizes $L_g(\psi)$. The general theory in the next section achieves this goal.

**4. Minimaxity on $\mathscr{V}(F_0, \varepsilon)$.** The main result of this section is to find the function $\psi^*$ that minimizes $L_g(\psi)$, for a given $0 < \varepsilon < 0.5$. We will show that $\psi^*$ belongs to the family

$$\psi_{a,b,c,t}(y) = \begin{cases} a\lambda_t(-c) + b\delta_t(-c), & \text{if } \ y < -c, \\ a\lambda_t(y) + b\delta_t(y), & \text{if } \ |y| \le c, \\ a\lambda_t(c) + b\delta_t(c), & \text{if } \ y > c, \end{cases} \tag{4.1}$$

where

$$\lambda_t(y) = \frac{1 - e^{-2yt}}{1 + e^{-2yt}} \quad \text{and} \quad \delta_t(y) = y - t\lambda_t(y). \tag{4.2}$$

The function $\lambda_t(y)$ is strictly increasing with $\lim_{y\to\infty} \lambda_t(y) = 1$, for all $t > 0$. The function $\delta_t(y)$ is also strictly increasing for all $0 < t < 1$ and $\lim_{y\to\infty} \delta_t(y) = \infty$, for all $t > 0$.

For $\psi_1, \psi_2 \in \mathscr{C}$, let

$$\begin{aligned} \langle \psi_1, \psi_2 \rangle_t &= \int_{-\infty}^{\infty} \psi_1(y - t)\psi_2(y - t)\phi(y)\, dy \\ &= \int_0^{\infty} \psi_1(y)\psi_2(y)\Delta_t(y)\, dy, \end{aligned} \tag{4.3}$$

where $\phi = \Phi'$ and

$$\Delta_t(y) = \phi(y + t) + \phi(y - t). \tag{4.4}$$

After noticing that

$$\delta_t(y) = -\frac{\phi'(y + t) + \phi'(y - t)}{\phi(y + t) + \phi(y - t)} = -\frac{\phi'(y + t) + \phi'(y - t)}{\Delta_t(y)},$$

it is easy to show that

$$\langle \psi, \delta_t \rangle_t = -\int_{-\infty}^{\infty} \psi(y)\phi'(y + t)\, dy = (\partial/\partial t)\eta(t, \Phi, \psi) \tag{4.5}$$

[see (4.2) and (4.3)].

Let $F^y = (1-\varepsilon)F_0 + \varepsilon\delta_y$, and denote by $v(F^\infty, \psi) = \lim_{y\to\infty} v(F^y, \psi)$ and $b(F^\infty, \psi) = \lim_{y\to\infty} b(F^y, \psi)$. Similarly we define $b(\Phi^\infty, \psi)$ and $v(\Phi^\infty, \psi)$ replacing $F_0$ by $\Phi$. Thus by (2.4), (4.3) and (4.5) we have

$$v(F^\infty, \psi) = \sigma_0^2 \frac{(1-\varepsilon)\langle\psi, \psi\rangle_\mu + \varepsilon\psi^2(\infty)}{[(1-\varepsilon)\langle\psi, \delta_\mu\rangle_\mu]^2}, \tag{4.6}$$

where $\mu = b(\Phi^\infty, \psi)$.

Moreover, let $\bar{t}(\varepsilon)$ be implicitly defined by the equation (in $t$)

$$\langle\lambda_t, \lambda_t\rangle_t = \varepsilon/(1-\varepsilon). \tag{4.7}$$

REMARK 1. Since $\int_0^\infty \psi(y)[\phi(y-t) - \phi(y+t)]\,dy$ is strictly increasing in $t$ for all $\psi \in \mathscr{C}$, and $\lambda_t(y)$ is nondecreasing in $t$ for all $y > 0$,

$$\langle\lambda_t, \lambda_t\rangle_t = \int_0^\infty \lambda_t(y)\,[\phi(y-t) - \phi(y+t)]\,dy$$

is also strictly increasing in $t$. Since $\langle\lambda_t, \lambda_t\rangle_t \to 1$ as $t \to \infty$, (4.7) has a unique solution, $\bar{t}(\varepsilon)$. In addition, define

$$\bar{\mu}(\varepsilon) = \min\{\bar{t}(\varepsilon), 1\} \tag{4.8}$$

and

$$\underline{v}(\varepsilon) = \frac{1}{(1-\varepsilon)} + \frac{\varepsilon}{(1-\varepsilon)^2}\frac{1}{[\phi(\bar{\mu}(\varepsilon)) + \phi(0)]^2}. \tag{4.9}$$

As we will see in Lemma 7, if the bias of an $M$-estimate at a standard normal contaminated distribution exceeds $\bar{\mu}(\varepsilon)$ then the corresponding asymptotic variance is bounded below by $\underline{v}(\varepsilon)$.

The function

$$l_g(\psi) = g\left(b(F^\infty, \psi), \frac{v(F^\infty, \psi)}{n}\right) = \sigma^m g\left(b(\Phi^\infty, \psi), \frac{v(\Phi^\infty, \psi)}{n}\right) \tag{4.10}$$

plays an important role in our derivations. Notice that $L_g(\psi) \ge l_g(\psi)$.

Theorem 4 gives sufficient conditions for the existence of a minimax $M$-estimate with $\psi$-function in the family (4.1).

THEOREM 4. *Assume that $g$ satisfies* C1–C5. *Then there exists $\psi^* = \psi_{a^*, b^*, c^*, t^*}$ such that*

$$l_g(\psi_{a,b,c,t}) \ge l_g(\psi^*) \text{ for all } a \ge 0, b \ge 0, c \ge 0, t \ge 0,$$

*where $\psi_{a,b,c,t}$ is given by* (4.1) *and* (4.2). *If in addition*

$$g\left(\bar{\mu}(\varepsilon), \frac{\underline{v}(\varepsilon)}{n}\right) \ge l_g(\psi^*) \tag{4.11}$$

*and*

$$L_g(\psi^*) = l_g(\psi^*), \tag{4.12}$$

*then $L_g(\psi) \ge L_g(\psi^*)$, for all $\psi \in \mathscr{C}$.*

TABLE 2
*Smallest n for the validity of minimaxity property of minimum quantile*

| | Level | | |
|---|---|---|---|
| **ε** | **0.01** | **0.05** | **0.10** |
| 0.05 | 3 | 3 | 3 |
| 0.10 | 3 | 3 | 3 |
| 0.15 | 3 | 3 | 3 |
| 0.20 | 3 | 3 | 3 |
| 0.25 | 4 | 3 | 3 |
| 0.30 | 7 | 4 | 4 |
| 0.35 | 16 | 9 | 6 |

To illustrate the application of Theorem 4 we shall obtain $M$-estimates with minimax quantiles. First we find $\psi^*$ defined in Theorem 4. According to Lemmas 3, 4, 5 and 6 in the Appendix, it is only necessary to search for $(a^*, b^*, c^*, t^*)$ in the compact set $C = C_1 \cup C_2$, where $C_1$ and $C_2$ are defined as follows. Let $I(\mu)$ be the closed interval with endpoints $c_1(\mu)$ and $c_2(\mu)$ given by (A.42) and (A.43), respectively. Then

$$C_1 = \{(\tilde{a}(c), \tilde{b}(c), c, t)\colon c \in I(t),\ \underline{\mu}(\varepsilon) \le t \le \bar{\mu}(\varepsilon)\} \tag{4.13}$$

and

$$C_2 = \{(0, 1, c, t)\colon 0 \le c \le c_2(t), \underline{\mu}(\varepsilon) \le t \le \bar{\mu}(\varepsilon)\}, \tag{4.14}$$

where $\underline{\mu}(\varepsilon)$, $\tilde{a}(c)$ and $\tilde{b}(c)$ are given by (A.34), (A.45) and (A.46), in the Appendix.

We computed the optimal score functions $\psi^*$ for intervals of level 0.99, 0.95 and 0.90, for several values of $n$ and $\varepsilon$. For each $\varepsilon$ we numerically found the value $n(\varepsilon)$ such that $\psi^*$ satisfies (4.11) for all $n \ge n(\varepsilon)$. The values $n(\varepsilon)$ are shown in Table 2 for several values of $\varepsilon$.

In all considered cases, the function $\psi^*$ is extremely well approximated by the function $\psi^H_{c^*}$ given by (2.12) with the same truncation point $c^*$ as the optimal $\psi^*$. The corresponding maximum quantiles are also extremely close (identical up to the fourth decimal place).

It remains to check condition (4.12). Fortunately, this condition can be analytically verified for the functions $\psi^H_{c^*}$ which are optimal from a practical point of view. See Lemma 7 in the Appendix. We have also numerically verified this condition for $\psi^*$, for all the considered cases.

Table 3 displays the values of $c^*$ and $\bar{q}_{\alpha,n}(\psi^H_{c^*})$ for several values of $n$ and $\varepsilon$. As one expects, $c^*$ approaches zero (and the corresponding estimate approaches the median) as $n$ and $\varepsilon$ increase. Naturally, $\bar{q}_{\alpha,n}(\psi^H_{c^*})$ decreases with $n$ and increases with $\varepsilon$.

TABLE 3
*Maximum quantiles and truncation points for the optimal Huber-type score function when $\sigma$ is known*

| | | $\alpha = 0.01$ | | $\alpha = 0.05$ | | $\alpha = 0.10$ | |
|---|---|---|---|---|---|---|---|
| $n$ | $\varepsilon$ | $c^*$ | $q^*_{\alpha,n}(\psi^H_{c^*})$ | $c^*$ | $q^*_{\alpha,n}(\psi^H_{c^*})$ | $c^*$ | $q^*_{\alpha,n}(\psi^H_{c^*})$ |
| 20 | 0.05 | 1.174 | 0.681 | 1.158 | 0.519 | 1.151 | 0.436 |
| | 0.10 | 0.829 | 0.808 | 0.786 | 0.622 | 0.762 | 0.525 |
| | 0.15 | 0.641 | 0.956 | 0.583 | 0.747 | 0.547 | 0.637 |
| | 0.20 | 0.515 | 1.132 | 0.457 | 0.898 | 0.416 | 0.773 |
| | 0.25 | 0.420 | 1.349 | 0.366 | 1.083 | 0.328 | 0.941 |
| 40 | 0.05 | 1.064 | 0.500 | 1.031 | 0.382 | 1.015 | 0.322 |
| | 0.10 | 0.737 | 0.614 | 0.669 | 0.480 | 0.626 | 0.409 |
| | 0.15 | 0.568 | 0.745 | 0.496 | 0.595 | 0.444 | 0.515 |
| | 0.20 | 0.452 | 0.898 | 0.388 | 0.731 | 0.343 | 0.641 |
| | 0.25 | 0.365 | 1.083 | 0.310 | 0.894 | 0.270 | 0.793 |
| 100 | 0.05 | 0.908 | 0.342 | 0.838 | 0.265 | 0.798 | 0.224 |
| | 0.10 | 0.621 | 0.444 | 0.533 | 0.357 | 0.472 | 0.310 |
| | 0.15 | 0.468 | 0.557 | 0.392 | 0.461 | 0.340 | 0.410 |
| | 0.20 | 0.367 | 0.688 | 0.304 | 0.581 | 0.260 | 0.524 |
| | 0.25 | 0.292 | 0.846 | 0.238 | 0.726 | 0.202 | 0.661 |
| 500 | 0.05 | 0.654 | 0.193 | 0.552 | 0.157 | 0.481 | 0.137 |
| | 0.10 | 0.416 | 0.279 | 0.336 | 0.239 | 0.281 | 0.218 |
| | 0.15 | 0.297 | 0.375 | 0.235 | 0.331 | 0.196 | 0.308 |
| | 0.20 | 0.227 | 0.486 | 0.176 | 0.438 | 0.145 | 0.412 |
| | 0.25 | 0.175 | 0.619 | 0.135 | 0.564 | 0.110 | 0.535 |

Theorem 3 lends practical relevance to the above minimax results when the "loss function" $g$ is given by (3.1). In addition, let $\mathscr{V}_k(\Phi, \varepsilon)$ be as in Theorem 1(c). Then,

$$\begin{aligned}
\sup_{F\in\mathscr{V}(\Phi,\varepsilon)} Q_{\alpha,n}(F,\psi) &\ge \sup_{F\in\mathscr{V}_k(\Phi,\varepsilon)} Q_{\alpha,n}(F,\psi)\\
&\ge \sup_{F\in\mathscr{V}_k(\Phi,\varepsilon)} q_{\alpha,n}(F,\psi) - o(1/\sqrt{n}) \quad \text{by Theorem 2(c)}\\
&\ge q_{\alpha,n}(F^\infty,\psi) - o(1/\sqrt{n})\\
&\ge q_{\alpha,n}(F^\infty,\psi^*) - o(1/\sqrt{n}), \quad \text{by Theorem 4}\\
&= \sup_{F\in\mathscr{V}(\Phi,\varepsilon)} q_{\alpha,n}(F,\psi^*) - o(1/\sqrt{n}),\\
&\ge \sup_{F\in\mathscr{V}(\Phi,\varepsilon)} Q_{\alpha,n}(F,\psi^*) - o(1/\sqrt{n}), \quad \text{by Theorem 2(b).}
\end{aligned}$$

In words, the finite sample minimax quantiles (which would be very hard to derive) can only be marginally smaller than $q_{\alpha,n}(F^\infty, \psi^*)$. Moreover, one can expect that the optimal finite sample score function $\psi$ can be only marginally different from $\psi^*$. Finally, $\sup_{F\in\mathscr{V}(\Phi,\varepsilon)} Q_{\alpha,n}(F,\psi^*)$ can be well approximated by $q_{\alpha,n}(F^\infty,\psi^*)$, with an error of order $o(1/\sqrt{n})$. In summary, Theorems 2, 3 and 4 justify the use of the minimax interval $\hat{\mu}_n^* \pm q_{\alpha,n}(F^\infty, \psi^*)$, where $\hat{\mu}_n^*$ is given by equation (2.2) with $\psi = \psi^*$.

**5. Unknown scale.** In practice, the scale parameter $\sigma_0$ is usually unknown and must be robustly estimated from the data. In this case the estimating equation (2.2) becomes

$$\sum_{i=1}^{n} \psi\left(\frac{y_i - \hat{\mu}_n}{\hat{\sigma}_n}\right) = 0, \tag{5.1}$$

where $\hat{\sigma}_n$ is a robust estimate of $\sigma_0$. If the underlying distribution $F$ of the data is symmetric, under mild regularity conditions, the asymptotic distribution of $\hat{\mu}_n$ is normal with asymptotic variance given by (2.4) and $\sigma_0$ replaced by $S(F)$, where $S(F) = \lim_{n\to\infty} \hat{\sigma}_n$, almost surely $[F]$. On the other hand, if $F$ is asymmetric the asymptotic distribution is still normal (under suitable regularity conditions) but the asymptotic variance is much more involved and depends on the particular choice of $\hat{\sigma}_n$ [see Huber (1981), Sections 6.4 and 6.5, and Davies (1998)]. A convenient choice for $\hat{\sigma}_n$ is the scale $S$-estimate defined by Rousseeuw and Yohai (1984) as follows: given $y_1, \dots, y_n$, let $s_n(t)$ be defined by

$$\frac{1}{n}\sum_{i=1}^{n} \chi\left(\frac{y_i - t}{s_n(t)}\right) = \beta, \tag{5.2}$$

where $\chi$ is even, nondecreasing for positive values and bounded with $\chi(0) = 0$ and $\beta = \mathrm{E}_\Phi(\chi(u))$. Finally the scale estimate is defined by

$$\hat{\sigma}_n = \min_t \; s_n(t). \tag{5.3}$$

There is an associated location $S$-estimate $\tilde{\mu}_n$ defined by

$$\tilde{\mu}_n = \arg_t \min \; s_n(t). \tag{5.4}$$

As we will see below, this location estimate appears in the asymptotic variance of $\hat{\mu}_n$ through its asymptotic value, $\widetilde{T}(F)$. However, since $\hat{\mu}_n$ has superior robustness and efficiency properties, we will not use $\tilde{\mu}_n$ as a location estimate. The breakdown point $\varepsilon^*$ of both $\hat{\sigma}_n$ and $\hat{\mu}_n$ is given by $\min(\beta/\chi(\infty), 1 - \beta/\chi(\infty))$, and thus to achieve $\varepsilon^* \le 0.5$, one uses $\beta = \chi(\infty)\ \varepsilon^*$.

Salibian-Barrera (2000) shows that under regularity conditions, which include continuous differentiability of $\psi'$ and $\chi'$ there exists $\varepsilon_0$ such that the uniform asymptotic result of Theorem 1(a) holds for $\varepsilon < \varepsilon_0$ and with $v(F, \psi)$ replaced by

$$v(F, \psi, \chi) = \frac{S^2(F)}{B^2}\mathrm{E}_F\{\gamma^2(X)\}, \tag{5.5}$$

where

$$\gamma(X) = \psi\left(\frac{X - T(F, \psi)}{S(F)}\right) - A\left[\chi\left(\frac{X - \widetilde{T}(F)}{S(F)}\right) - \beta\right],$$

$$A = \frac{\mathrm{E}_F\left\{\psi'\big((X - T(F, \psi))/S(F)\big)\big((X - T(F, \psi))/S(F)\big)\right\}}{\mathrm{E}_F\left\{\chi'\big((X - \widetilde{T}(F))/S(F)\big)\big((X - \widetilde{T}(F))/S(F)\big)\right\}}$$

and

$$B = \mathrm{E}_F\left\{\psi'\left(\frac{X - T(F, \psi)}{S(F)}\right)\right\}.$$

Other uniform results for simultaneous location and scale $M$-estimates are given by Davies (1998).

Given the complexity of the expression for $v(F, \psi, \chi)$, it would be very difficult to solve a general minimax problem for the unknown $\sigma_0$ case. To obtain an approximate practical solution we introduce the following simplifications:

1. We restrict attention to the important case when $g(x_1, x_2)$ is the $\alpha$-quantile defined by (3.1).
2. We restrict our search for the optimal $\psi$ function to a family of twice differentiable approximations to the Huber functions given by (2.12). Notice that when $\sigma_0$ is known, the optimal minimax function cannot be practically distinguished from a Huber-type function. The transition from linearly increasing to constant is done in the interval $[d, 1]$, with $d = 0.8$. Other choices for $d$ between 0.7 and 0.9 produced similar results. The differentiable family is obtained as follows. First, define $\psi_1^D$ by

$$\psi_1^D(u) = \text{ sign } (u)\begin{cases} |u|, & \text{if } |u| \le 0.8, \\ p_4(|u|), & \text{if } |u|\, 0.8 < |u| \le 1, \\ p_4(1), & \text{if } |u| > 1, \end{cases}$$

   where $p_4(u) = 38.4 - 175.0u + 300.0u^2 - 225.0u^3 + 62.5u^4$, and the coefficients are chosen so that $\psi_1^D(u)$ is twice continuously differentiable. This function, which also turns out to be strictly monotone for $|u| < 1$, is plotted in Figure 1. Finally for any $c > 0$ put $\psi_c^D(u) = \psi_1^D(u/c)$.
3. The function $\chi$ appearing in (5.2) is taken from the Tukey's bisquare family $\chi_k^B(y) = \chi_1^B(y/k)$ with

$$\chi_1^B(u) = \begin{cases} u^2(3 - 3u^2 + u^4), & \text{if } |u| \le 1, \\ 1, & \text{if } |u| > 1, \end{cases} \tag{5.6}$$

   $k = 1.988$ and $\beta = 0.40$. This choice of $c$ which yields a 0.40 breakdown point for $S(F)$ and $0.15 \le \varepsilon_0 < 0.16$ was preferred to $c = 1.547$ which yields a 0.5 breakdown point for $S(F)$ and $0.10 \le \varepsilon_0 < 0.11$.
4. We restrict the contaminating distributions to the class of point mass distributions. When $\sigma_0$ is known, the least favorable contaminating distribution is in this class. We conjecture that this also holds when $\sigma$ is unknown.

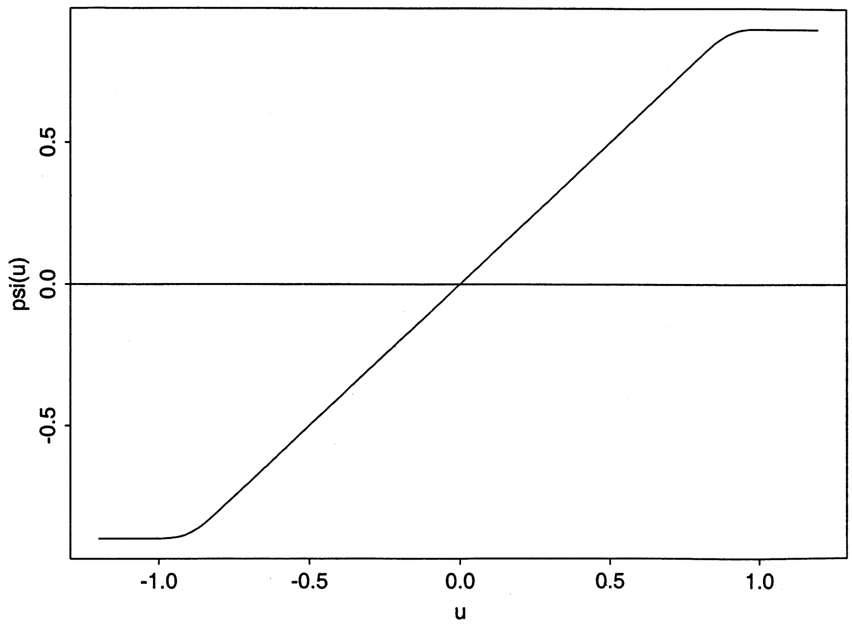


FIG. 1. *Differentiable approximations to Huber's score function.*

Thus, we proceed as follows:

1. Fix $\varepsilon, n$ and $\alpha$.
2. For given $y$ (contamination point) the corresponding asymptotic values of $\widetilde{T}(F_y)$ and $S(F_y)$ with $F_y = (1-\varepsilon)\Phi + \varepsilon\delta_y$ are numerically computed. More precisely, we numerically find the value $\widetilde{T}(F_y)$ which minimizes, in $t$, $S(F_y, t)$ defined by
$$\mathrm{E}_{F_y}\chi_k^B\left(\frac{X-t}{S(F_y, t)}\right) = 0.4$$
and set $S(F_y) = S(F_y, \widetilde{T}(F_y))$.
3. For each value of $c$ (truncation point of $\psi_c^D$) we calculate $T(F_y, \psi_c^D)$ by solving the equation
$$\mathrm{E}_{F_y}\psi_c^D\left(\frac{X - T(F_y, \psi)}{S(F_y, t)}\right) = 0.$$
4. Using $\widetilde{T}(F_y), S(F_y)$ and $T(F_y, \psi_c^D)$ we calculate $v(F_y, \psi_c^D, \chi_k^B)$ given by (5.5).
5. Using $T(F_y, \psi_c^D)$ and $v(F_y, \psi_c^D, \chi_k^B)$ we calculate the approximate $\alpha$-quantile $q_\alpha(F_y, \psi_c^D)$ given by (3.1), with $F = F_y$ and the given $n$.
6. Using a thin grid of values of $y$ we approximate the maximum value, $\bar{q}_\alpha(\psi_c^D)$, of $q_\alpha(F_y, \psi_c^D)$.

TABLE 4
*Maximum quantiles and truncation points for the optimal Huber-type score function when $\sigma$ is unknown*

| | | $\alpha = 0.01$ | | $\alpha = 0.05$ | | $\alpha = 0.10$ | |
|---|---|---|---|---|---|---|---|
| **n** | $\varepsilon$ | $c^*$ | $q^*_{\alpha,n}(\psi^H_{c^*})$ | $c^*$ | $q^*_{\alpha,n}(\psi^H_{c^*})$ | $c^*$ | $q^*_{\alpha,n}(\psi^H_{c^*})$ |
| 20 | 0.05 | 1.17 | 0.683 | 1.16 | 0.521 | 1.16 | 0.437 |
| | 0.10 | 0.73 | 0.811 | 0.70 | 0.624 | 0.68 | 0.527 |
| | 0.15 | 0.49 | 0.962 | 0.46 | 0.750 | 0.44 | 0.639 |
| | 0.20 | 0.36 | 1.139 | 0.33 | 0.902 | 0.31 | 0.776 |
| | 0.25 | 0.26 | 1.358 | 0.23 | 1.089 | 0.22 | 0.945 |
| 40 | 0.05 | 1.07 | 0.501 | 1.04 | 0.383 | 1.00 | 0.322 |
| | 0.10 | 0.66 | 0.616 | 0.60 | 0.481 | 0.56 | 0.410 |
| | 0.15 | 0.45 | 0.748 | 0.41 | 0.597 | 0.38 | 0.516 |
| | 0.20 | 0.33 | 0.902 | 0.30 | 0.733 | 0.27 | 0.643 |
| | 0.25 | 0.23 | 1.088 | 0.21 | 0.897 | 0.21 | 0.796 |
| 100 | 0.05 | 0.92 | 0.342 | 0.85 | 0.265 | 0.81 | 0.225 |
| | 0.10 | 0.56 | 0.444 | 0.49 | 0.357 | 0.44 | 0.311 |
| | 0.15 | 0.39 | 0.558 | 0.32 | 0.462 | 0.29 | 0.410 |
| | 0.20 | 0.28 | 0.690 | 0.21 | 0.583 | 0.21 | 0.525 |
| | 0.25 | 0.19 | 0.849 | 0.17 | 0.727 | 0.14 | 0.662 |
| 500 | 0.05 | 0.66 | 0.193 | 0.56 | 0.157 | 0.49 | 0.137 |
| | 0.10 | 0.39 | 0.279 | 0.35 | 0.239 | 0.28 | 0.218 |
| | 0.15 | 0.27 | 0.375 | 0.22 | 0.332 | 0.18 | 0.308 |
| | 0.20 | 0.19 | 0.487 | 0.15 | 0.438 | 0.13 | 0.412 |
| | 0.25 | 0.13 | 0.619 | 0.13 | 0.565 | 0.07 | 0.536 |

7. Again, using a thin grid of values of $c$ we minimize $\bar{q}_\alpha(\psi^D_c)$ to find the optimal truncation constant $c^* = c^*(\varepsilon, \alpha, n)$. The corresponding minimax quantile will be denoted by $\bar{q}^*(\varepsilon, \alpha, n)$.

Table 4 summarizes the results of our numerical calculations. Observe that the values of the minimax quantiles are very close to those in Table 2 which correspond to the known $\sigma_0$ case. This confirms the appropriateness of restricting the search to the family $\psi^D_c$, at least for point mass contaminations. The fact that $\sigma_0$ is unknown does not cause a sensible increase in the value of the minimax quantile because the self-adjustment of the truncation constant accounts for the possible overestimation of $\hat{\sigma}$.

In applications, the given quantiles must be multiplied by $\hat{\sigma}_n$, as illustrated in the examples below.

EXAMPLE 1. Table 5 gives Newcomb's measurements of the passage time of light as reported by Stigler (1977).

The data contains two possible outliers ($-44$ and $-2$). Newcomb was troubled by these two "unusual" measurements. Finally he deleted $-44$ but kept $-2$ for the calculation of his estimate of the speed of light in air. We take $\varepsilon = \alpha = 0.05$ for this illustration and calculate $c^*(0.05, 0.05, 66) = 0.92$ and

TABLE 5
*Newcomb's third series of measurements of the passage time of light* ($n = 66$)

| | | | | | |
|---|---|---|---|---|---|
| −44 | 23 | 25 | 27 | 29 | 32 |
| −2 | 23 | 26 | 28 | 29 | 33 |
| 16 | 24 | 26 | 28 | 30 | 33 |
| 16 | 24 | 26 | 28 | 30 | 34 |
| 19 | 24 | 26 | 28 | 30 | 36 |
| 20 | 24 | 26 | 28 | 31 | 36 |
| 21 | 24 | 27 | 28 | 31 | 36 |
| 21 | 25 | 27 | 28 | 32 | 36 |
| 22 | 25 | 27 | 29 | 32 | 37 |
| 22 | 25 | 27 | 29 | 32 | 39 |
| 23 | 25 | 27 | 29 | 32 | 40 |

The given values times 0.001 and plus 24.8 are Newcomb's original measurements, in millionth of a second.

$\bar{q}^*(0.05, 0.05, 66) = 0.31$. The robust estimates of location and scale [using (5.6)] are 27.32 and 4.98, respectively. The optimal robust confidence interval is $27.32 \pm (4.98 \times 0.31)$ or (25.78, 28.86). The corresponding Student's $t$ intervals are (23.57, 28.85) using all the data (25.74, 28.84) deleting $-44$ and (26.48, 29.02) deleting $-2$ and $-44$. The point $-44$ upsets the length and the lower end of Student's $t$ interval. The agreement between the robust interval and Student's $t$ interval using 65 points is consistent with Newcomb's decision of keeping $-2$ but deleting $-44$ for his final estimate.

EXAMPLE 2. This example is based on the artificial data set $X_i = 10 + \eta_i$ where the independent errors $\eta_i$ are N(0, 1), for $i = 1, \dots, 180$ and $N(2, 0.01)$ for $i = 181, \dots, 200$. The data are presented in Figure 2.

As in Example 1, we take $\alpha = 0.05$ and compute $c^*(0.05, 0.05, 200) = 0.72$ and $\bar{q}^*(0.05, 0.05, 200) = 0.21$, for $\varepsilon = 0.05$, and $c^*(0.10, 0.05, 200) = 0.42$ and $\bar{q}^*(0.05, 0.05, 200) = 0.3$, for $\varepsilon = 0.10$. The robust estimate of location is 10.16 for both values of $\varepsilon$. The initial scale estimate is 1.08. The optimal robust confidence intervals are (9.94, 10.38) for $\varepsilon = 0.05$ and (9.83, 10.48) for $\varepsilon = 0.10$. Student's $t$ interval using all the data is (10.07, 10.36). The results of this example are not atypical. In fact, in a small Monte Carlo simulation of 1000 similar data sets, 83% of the robust intervals for $\varepsilon = 0.05$, 97% of the robust intervals for $\varepsilon = 0.10$, 38% of Student's $t$ intervals using all the data and 34% of Student's $t$ intervals using the "clean data" (after hard rejection of outliers) included the true value 10.

## APPENDIX

Observe that (2.2) and (2.9) imply that the optimal function $\psi^*$ is independent of $\mu_0$ and $\sigma_0$. Then, in this Appendix we will assume, without loss of generality, that $\mu_0 = 0$ and $\sigma_0 = 1$. Consequently, from now on we will write $\Phi$ in place of $F_0$.

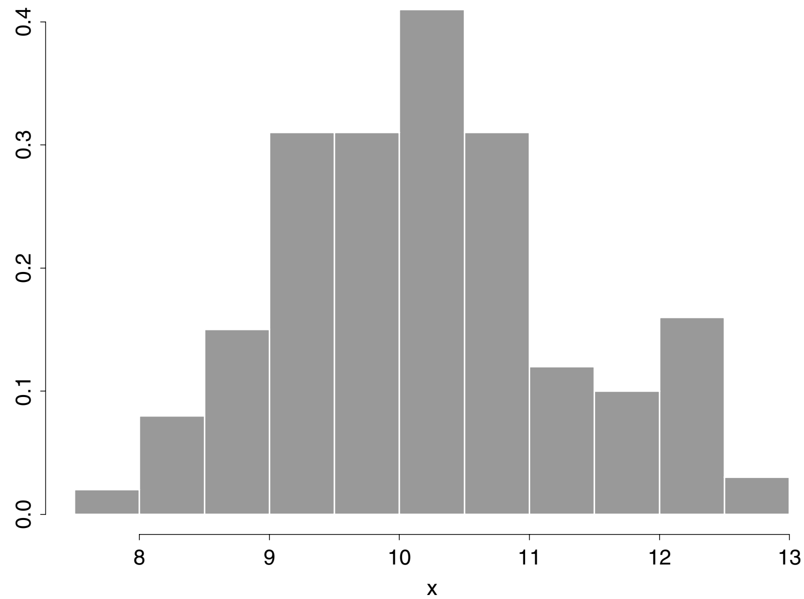


FIG. 2. *Artificial data.*

PROOF OF LEMMA 1. Without loss of generality we may assume that $\sup\psi = 1$. We start proving that

$$(\partial/\partial t)\eta(t, \Phi, \psi) > 0, \text{ for all } t. \tag{A.1}$$

For any $t$ we have

$$\begin{aligned}(\partial/\partial t)\eta(t, \Phi, \psi) &= -\int_{-\infty}^{\infty} \psi(y-t)\phi'(y)dy = \int_{-\infty}^{\infty} \psi(y-t)y\phi(y)\,dy \\ &= \int_0^{\infty} [\psi(y-t) + \psi(y+t)]y\phi(y)\,dy,\end{aligned}$$

and notice that $\psi(y-t)+\psi(y+t) \geq 0$ for all $y > 0$, and there exists $a$ such that the inequality is strict on $a < y$ because $\psi(y-t)+\psi(y+t) \to 2$ as $y \to \infty$.

To prove (a) notice that the existence of $T(F, \psi)$ follows from the continuity of $\eta(t, F, \psi)$ and the fact that, by the dominated convergence theorem, $\lim_{t\to\infty} \eta(t, F, \psi) = 1$ and $\lim_{t\to-\infty} \eta(t, F, \psi) = -1$. The uniqueness of $T(F, \psi)$ follows from the strict monotonicity of $\eta(t, F, \psi)$ [notice that $\eta(t, F, \psi) = (1-\varepsilon)\eta(t, \Phi, \psi)+\varepsilon\eta(t, H, \psi)$ is nondecreasing and $\eta(t, \Phi, \psi)$ is strictly increasing].

To prove (b) let $t_1$ and $t_2$ be the solutions to $(1-\varepsilon)\eta(t, \Phi, \psi) - \varepsilon = 0$ and $(1-\varepsilon)\eta(t, \Phi, \psi) + \varepsilon = 0$, respectively. Then we have $t_2 \leq T(F) \leq t_1$ because

$$\begin{aligned}(1-\varepsilon)\eta(t, \Phi, \psi) - \varepsilon &\leq \eta(t, F, \psi) \\ &\leq (1-\varepsilon)\eta(t, \Phi, \psi) + \varepsilon \quad \text{for all } F \in \mathscr{V}(\Phi, \varepsilon).\end{aligned}$$

To prove the "limit" part of (b) notice that

$$\begin{aligned}\lim_{t\to\infty} T((1-\varepsilon)\Phi+\varepsilon\delta_y) &= \lim_{t\to\infty} \eta^{-1}(\varepsilon\psi(y-t)/(1-\varepsilon), \Phi, \psi)\\ &= \eta^{-1}(\varepsilon/(1-\varepsilon), \Phi, \psi)\\ &= t_1,\end{aligned}$$

where $\eta^{-1}(u, \Phi, \psi)$ is the inverse of $\eta$ with respect to $t$.

Using (A.1) and part (b) we get

$$\inf_{F\in\mathscr{V}(\Phi,\varepsilon)} \left.\frac{\partial\eta(t, F, \psi)}{\partial t}\right|_{t=T(F,\psi)} \geq \inf_{t_2\leq t\leq t_1} \frac{\partial\eta(t, \Phi, \psi)}{\partial t} > 0.$$

Then (c) follows from the fact that $\psi$ is bounded. □

PROOF OF THEOREM 1. Consider first a fixed real number $a$ and define

$$t_n = t_n(F, \psi) = a\sqrt{\frac{v(F, \psi)}{n}} + T(F, \psi). \tag{A.2}$$

Using the monotonicity and continuity of $\psi$ one has

$$\begin{aligned}P\left(\frac{\sqrt{n}(\hat{\mu}_n - T(F,\psi))}{\sqrt{v(F,\psi)}} \leq a\right) &= P\left(\sum_{i=1}^{n} \psi(x_i - t_n) \leq 0\right)\\ &= P\left(\sum_{i=1}^{n} \frac{\psi(x_i-t_n)+\eta(t_n, F, \psi)}{\sqrt{n}\sigma(t_n, F, \psi)} \leq A_n(F,\psi)\right),\end{aligned} \tag{A.3}$$

where $\sigma^2(t, F, \psi) = \mathrm{Var}_F\{\psi(X-t)\}$ and

$$A_n(F, \psi) = \frac{\sqrt{n}\eta(t_n, F, \psi)}{\sigma(t_n, F, \psi)}. \tag{A.4}$$

Consider $\mathscr{F} \subset \mathscr{V}(\Phi, \varepsilon)$, then using (A.3) and the Berry–Essen version of the central limit theorem, to prove

$$\lim_{n\to\infty} \sup_{F\in\mathscr{F}} \left| P_F\left(\frac{\sqrt{n}(\hat{\mu}_n - T(F, \psi)}{\sqrt{v(F, \psi)}} \leq a\right) - \Phi(a)\right| = 0, \tag{A.5}$$

it is enough to show that

$$\limsup_{n\to\infty} \sup_{F\in\mathscr{V}(\Phi,\varepsilon)} \frac{\mathrm{E}_F|\psi(X-t_n)+\eta(t_n, F, \psi)|^3}{\sigma^3(t_n, F, \psi)} < \infty \tag{A.6}$$

and

$$\lim_{n\to\infty} \sup_{F\in\mathscr{F}} |A_n(F, \psi) - a| = 0. \tag{A.7}$$

Suppose that A1 holds, then since $\psi$ is bounded, to prove (A.6) it suffices to show that the denominator is uniformly bounded away from zero. In fact, by Chebyshev's inequality, for all $\delta > 0$,

$$\begin{aligned}\sigma^2(t_n, F, \psi) &\geq \delta^2 P_F(|\psi(X-t_n)+\eta(t_n, F, \psi)| \geq \delta)\\ &\geq \delta^2(1-\varepsilon)P_\Phi(|\psi(X-t_n)+\eta(t_n, F, \psi)| \geq \delta)\\ &\geq \delta^2(1-\varepsilon)\min\{P_\Phi(\psi(X-t_n) \leq -\delta), P_\Phi(\psi(X-t_n) \geq \delta)\}\\ &\geq \delta^2(1-\varepsilon)\min\{P_\Phi(\psi(X+\bar{t}) \leq -\delta), P_\Phi(\psi(X-\bar{t}) \geq \delta)\}\\ &> 0,\end{aligned} \tag{A.8}$$

where $\bar{t} = \bar{t}(\psi, \varepsilon) = \sup_{F\in\mathscr{V}(\Phi,\varepsilon)}|T(F, \psi)| + a\sup_{F\in\mathscr{V}(\Phi,\varepsilon)}\sqrt{v(F, \psi)}$. Notice that $\bar{t} < \infty$ by Lemma 1(b) and (c). This proves (A.6).

We will show now that

$$\lim_{n\to n}\sup_{F\in\mathscr{V}(\Phi,\varepsilon)}|\sigma^2(t_n, F, \psi) - \mathrm{E}_F(\psi^2(X - T(F, \psi)))| = 0. \tag{A.9}$$

We have

$$\sigma^2(t_n, F, \psi) = \mathrm{E}_F(\psi^2(X-t_n)) - \eta^2(t_n, F, \psi). \tag{A.10}$$

According to Lemma 1(c) we have $M = \sup_{F\in\mathscr{V}(\Phi,\varepsilon)}\sqrt{v(F, \psi)} < \infty$. Then

$$\sup_{F\in\mathscr{V}(\Phi,\varepsilon)}|t_n - T(F, \psi)| \leq \frac{aM}{\sqrt{n}}. \tag{A.11}$$

Let $\delta(u) = \sup_t|\psi^2(x-t-u) - \psi^2(x-t)|$. Then, since $\psi^2$ is uniformly continuous, we have $\lim_{u\to 0}\delta(u) = 0$. Then

$$\begin{aligned}&\lim_{n\to\infty}\sup_{F\in\mathscr{V}(\Phi,\varepsilon)}|\mathrm{E}_F(\psi^2(X-t_n)) - \mathrm{E}_F(\psi^2(X-T(F, \psi)))|\\ &\quad\leq \lim_{n\to\infty}\delta\left(\frac{aM}{\sqrt{n}}\right) = 0.\end{aligned} \tag{A.12}$$

Similarly we can prove

$$\begin{aligned}&\lim_{n\to\infty}\sup_{F\in\mathscr{V}(\Phi,\varepsilon)}\sup\eta(t_n, F, \psi)\\ &\quad= \lim_{n\to\infty}\sup_{F\in\mathscr{V}(\Phi,\varepsilon)}|\eta(t_n, F, \psi) - \eta(T(F, \psi), F, \psi)| = 0.\end{aligned} \tag{A.13}$$

Then (A.10), (A.12) and (A.13) imply (A.9).

Suppose now that A1 and A2 hold. To prove part (a) we have to show (A.7) with $\mathscr{F} = \mathscr{V}(\Phi, \varepsilon)$. Using the definition of $T(F, \psi)$ and the mean value theorem, we can write

$$\begin{aligned}A_n(F, \psi) &= \frac{\sqrt{n}}{\sigma(t_n, F, \psi)}(\eta(t_n, F, \psi) - \eta(T(F, \psi), F, \psi))\\ &= \frac{a\sqrt{v(F, \psi)}}{\sigma(t_n, F, \psi)}\mathrm{E}_F\psi'(X-\tilde{t}_n),\end{aligned}$$

where $\tilde{t}_n$ is between $T(F, \psi)$ and $t_n$. Then, since (A.9) holds, to prove (A.7) with $\mathscr{F} = \mathscr{V}(\Phi, \varepsilon)$, it suffices to show that

$$\lim_{n\to\infty} \sup_{F\in\mathscr{V}(\Phi,\varepsilon)} |\mathrm{E}_F\psi'(X-\tilde{t}_n) - \mathrm{E}_F\psi'(X-T(F,\psi))| = 0, \tag{A.14}$$

which is derived from the uniform continuity of $\psi'$ and (A.11). This proves part (a) of Theorem 1.

Suppose now that A1 and A3 hold. Because of Lemma 1(b) we can choose $k_0 > c + \sup_{F\in\mathscr{V}(\Phi,\varepsilon)} |T(F, \psi)|$. Since (A.6) holds, to prove (c) it will be enough to show that (A.7) holds with $\mathscr{F} = \mathscr{V}_{k_0}(\Phi, \varepsilon)$ and since (A.9) holds, it will be enough to show that if we define $C_n(F, \psi)$ by

$$\begin{aligned} C_n(F,\psi) &= \sqrt{n}(\eta(t_n,F,\psi) - \eta(T(F,\psi),F,\psi)) \\ &\quad - a\sqrt{v(F,\psi)}\mathrm{E}_F(I_{(-c,c)}(X-T(F,\psi))\psi'(X-T(F,\psi))), \end{aligned} \tag{A.15}$$

then

$$\sup_{F\in\mathscr{V}_k(\Phi,\varepsilon)} |C_n(F,\psi)| = 0.$$

Suppose this is not true; then there exists $\gamma > 0$ and a sequence $F_n \in \mathscr{V}_{k_0}(\Phi, \varepsilon)$, $n \geq 1$, such that

$$\lim_{n\to\infty} |C_n(F_n, \psi)| > \gamma \tag{A.16}$$

and without loss of generality we can assume that $\lim_{n\to\infty} T(F_n, \psi) = t_0$. Let $F_n = (1-\varepsilon)\Phi + \varepsilon H_n$, where $H_n$ assigns probability 0 to the interval $[-k_0, k_0]$. Then, since $\psi'(x) = 0$ outside $[-c, c]$, we have

$$\begin{aligned} &\lim_{n\to\infty} \sqrt{n}(-\mathrm{E}_{H_n}(\psi(X-t_n(F_n,\psi)) + \mathrm{E}_{H_n}(\psi(X-T(F_n,\psi)))) \\ &\quad - a\sqrt{v(F_n,\psi)}\mathrm{E}_{H_n}\psi'(X-T(F_n,\psi)) = 0. \end{aligned} \tag{A.17}$$

On the other hand, for all $x \neq (t_0 \pm c)$, (A.11) and the mean value theorem yields

$$\begin{aligned} &\lim_{n\to\infty} \sqrt{n}(-(\psi(x-t_n(F_n,\psi)) + \psi(x-T(F_n,\psi))) \\ &\quad -a\sqrt{v(F_n,\psi)}\psi'(x-T(F_n,\psi)) \\ &\quad = \lim_{n\to\infty} a\sqrt{v(F_n,\psi)}(\psi'(x-\tilde{t}_n(F_n,\psi)) - \psi'(x-T(F_n,\psi))) = 0, \end{aligned}$$

where $\tilde{t}_n(F_n, \psi)$ is between $T(F_n, \psi)$ and $t_n(F_n, \psi)$.

Using that $|\psi(x) - \psi(y)| \leq K|x-y|$, where $K = \sup \psi'$ we also have

$$\begin{aligned} &\left|\sqrt{n}(-(\psi(x-t_n(F_n,\psi)) + \psi(x-T(F_n,\psi))) - a\sqrt{v(F_n,\psi)}\psi'(x-T(F_n,\psi))\right| \\ &\quad \leq 2aK \sup_{F\in\mathscr{V}(\Phi,\varepsilon)} \sqrt{v(F,\psi)}. \end{aligned}$$

Then the dominated convergence theorem implies

$$\text{(A.18)}\qquad \begin{aligned}\lim_{n\to\infty}\sqrt{n}(-\mathrm{E}_\Phi(\psi(X-t_n(F_n))+\mathrm{E}_\Phi(\psi(X-T(F_n,\psi)))\\ -a\sqrt{v(F_n,\psi)}\mathrm{E}_\Phi\psi'(X-T(F_n,\psi))=0.\end{aligned}$$

Since (A.17) and (A.18) contradict (A.16), part (c) is proved.

Assume A1 and A3. Since (A.3) and (A.6) hold, to prove (b) it is enough to show that

$$\text{(A.19)}\qquad \liminf_{n\to\infty}\inf_{F\in\mathscr{V}(\Phi,\varepsilon)}A_n(F,\psi)\geq a\quad \text{if } a\geq 0$$

and

$$\text{(A.20)}\qquad \limsup_{n\to\infty}\sup_{F\in\mathscr{V}(\Phi,\varepsilon)}A_n(F,\psi)\leq a\quad \text{if } a<0.$$

The proofs of (A.19) and (A.20) are completely similar, so, we will only prove (A.19). Since (A.6) holds, to prove (A.19) it is enough to show that

$$\text{(A.21)}\qquad \liminf_{n\to\infty}\inf_{F\in\mathscr{V}(\Phi,\varepsilon)}C_n(F,\psi)\geq 0,$$

where $C_n(F,\psi)$ is defined in (A.16). Since $\psi$ is nondecreasing and $t_n\geq T(F,\psi)$, for all $\delta>0$ we have

$$\text{(A.22)}\qquad \begin{aligned}C_n(F,\psi)&=\sqrt{n}\mathrm{E}_F(-\psi(X-t_n)+\psi(X-T(F,\psi)))\\ &\geq\sqrt{n}(\mathrm{E}_F(I_{[-(c-\delta),(c-\delta)]}(X-T(F,\psi))(-\psi(X-t_n)\\ &\qquad\qquad +\psi(X-T(F,\psi))).\end{aligned}$$

Since $\sup_{F\in\mathscr{V}_k(\Phi,\varepsilon)}|T(F,\psi)-t_n|\to 0$ and $\psi$ is differentiable in $(-c,c)$, by the mean value theorem there exist $n_0$ such that for all $n\geq n_0$ and $F\in\mathscr{V}(\Phi,\varepsilon)$,

$$\begin{aligned}&\sqrt{n}(\mathrm{E}_F(I_{[-(c-\delta),\,(c-\delta)]}(X-T(F,\psi))(-\psi(X-t_n)+\psi(X-T(F,\psi)))\\ &\quad=av(F,\psi)\mathrm{E}_F(I_{[-(c-\delta),\,(c-\delta)]}(X-T(F,\psi))(\psi'(X-\tilde{t}_n)),\end{aligned}$$

where $\tilde{t}_n$ is between $T(F,\psi)$ and $t_n$. Using the uniform continuity of $\psi'$ and the fact that by Lemma 1(c) $\sup_{F\in\mathscr{V}_k(\Phi,\varepsilon)}v(F,\psi)<\infty$ we have that

$$\text{(A.23)}\qquad \begin{aligned}\sup_{F\in\mathscr{V}(\Phi,\varepsilon)}av(F,\psi)|\mathrm{E}_F(I_{[-(c-\delta),(c-\delta)]}(X-T(F,\psi))\\ \times(\psi'(X-\tilde{t}_n)-\psi'(X-T(F,\psi)))|=0\end{aligned}$$

and

$$\text{(A.24)}\quad \lim_{\delta\to 0^+}\sup_{F\in\mathscr{V}(\Phi,\varepsilon)}v(F,\psi)|\mathrm{E}_F(|I_{[-(c-\delta),\,(c-\delta)]}-I_{(-c,c)})|\psi'(X-T(F,\psi))=0.$$

(A.22)–(A.24) imply (A.21), and then part (b) is proved. □

PROOF OF THEOREM 2. (a) Using Theorem 1(a), (3.1), (3.2) and (3.3) we get

$$
\begin{aligned}
1-\alpha &= \Phi\left(\frac{\sqrt{n}(q_{\alpha,n}(F,\psi)-b(F,\psi))}{\sqrt{v(F,\psi)}}\right) \\
&\quad + \Phi\left(\frac{\sqrt{n}(q_{\alpha,n}(F,\psi)+b(F,\psi))}{\sqrt{v(F,\psi)}}\right) - 1 \\
&= P(|\hat{\mu}_n| \le Q_{\alpha,n}(F,\psi)) \\
&= P\left(-\frac{\sqrt{n}(Q_{\alpha,n}(F,\psi)+b(F,\psi))}{\sqrt{v(F,\psi)}}\right. \\
&\qquad \left. \le \frac{\sqrt{n}(\hat{\mu}_n - T(F,\psi))}{\sqrt{v(F,\psi)}} \le \frac{\sqrt{n}(Q_{\alpha,n}(F,\psi)-b(F,\psi))}{\sqrt{v(F,\psi)}}\right) \\
&= \Phi\left(\frac{\sqrt{n}(Q_{\alpha,n}(F,\psi)-b(F,\psi))}{\sqrt{v(F,\psi)}}\right) \\
&\quad + \Phi\left(\frac{\sqrt{n}(Q_{\alpha,n}(F,\psi)+b(F,\psi))}{\sqrt{v(F,\psi)}}\right) - 1 + R_{1,n}(F,\psi)
\end{aligned}
\tag{A.25}
$$

with

$$
\lim_{n\to\infty} \sup_{F\in\mathscr{V}(\Phi,\varepsilon)} |R_{1,n}(F,\psi)| = 0. \tag{A.26}
$$

Let

$$
h(t) = \Phi\left[\sqrt{n}(t-b(F,\psi))\Big/\sqrt{v(F,\psi)}\right] + \Phi\left[\sqrt{n}(t+b(F,\psi))\Big/\sqrt{v(F,\psi)}\right] - 1.
$$

Then

$$
\begin{aligned}
Q_{\alpha,n}(F,\psi) &= h^{-1}(1-\alpha-R_{1,n}(F,\psi)) \\
&= h^{-1}(1-\alpha) - (h^{-1})'(1-\alpha_n^*)R_{1,n}(F,\psi) \\
&= q_{\alpha,n}(F,\psi) - (h^{-1})'(1-\alpha_n^*)R_{1,n}(F,\psi),
\end{aligned}
$$

with $1-\alpha_n^*$ between $1-\alpha$ and $1-\alpha-R_{1,n}(F,\psi)$. Let $q_{\alpha,n}^* = h^{-1}(1-\alpha_n^*)$, $A_n = \sqrt{n}[h^{-1}(1-\alpha_n^*)-b(F,\psi)]/\sqrt{v(F,\psi)}$ and $B_n = \sqrt{n}[h^{-1}(1-\alpha_n^*)+b(F,\psi)]/\sqrt{v(F,\psi)}$. The result follows now because $(h^{-1})'(1-\alpha_n^*) = O(1/\sqrt{n})$. In fact,

$$
(h^{-1})'(1-\alpha_n^*) = \frac{1}{h'[q_{\alpha,n}^*]} = \frac{1}{\sqrt{n}}\frac{\sqrt{v(F,\psi)}}{[\phi(A_n)+\phi(B_n)]}
$$

and the second factor on the right side below is uniformly bounded. To see that, notice that, since $\phi(A_n)+\phi(B_n)-1 \to 1-\alpha$ uniformly on $\mathscr{V}(\Phi,\varepsilon)$, we have that $0 \le \min\{|A_n|,|B_n|\} \le K < \infty$, for large $n$, uniformly on $\mathscr{V}(\Phi,\varepsilon)$. Then $\phi(A_n)+\phi(B_n)$ is uniformly bounded away from 0. The proofs of parts (b) and (c) of the theorem are similar. □

PROOF OF THEOREM 3. (a) Let $A_n = -[\bar{q}_{\alpha,n}(\psi) + b(F,\psi)]/\sqrt{v(F,\psi)/n}$ and $B_n = [\bar{q}_{\alpha,n}(\psi) - b(F,\psi)]/\sqrt{v(F,\psi)/n}$. By Theorem 1(a) and the definition of $q_{\alpha,n}(F,\psi)$,

$$\begin{aligned} P_F(|\hat{\mu}_n| \leq \bar{q}_{\alpha,n}(\psi)) &= P_F\left(A_n \leq \sqrt{n}\frac{\hat{\mu}_n - T(F,\psi)}{\sqrt{v(F,\psi)}} \leq B_n\right) \\ &= \Phi(B_n) - \Phi(A_n) + r_n(F) \\ &\geq \Phi(b_n) - \Phi(a_n) + r_n(F) = 1 - \alpha + r_n(F), \end{aligned}$$

where $\sup_{F\in\mathscr{V}(\Phi,\varepsilon)} |r_n(F)| \to 0$, $a_n = -[q_{\alpha,n}(F,\psi) + b(F,\psi)]/\sqrt{v(F,\psi)/n}$ and $b_n = [q_{\alpha,n}(F,\psi) - b(F,\psi)]/\sqrt{v(F,\psi)/n}$. Then

$$\liminf_{n\to\infty} \inf_{F\in\mathscr{V}(\Phi,\varepsilon)} P_F(|\hat{\mu}_n| \leq \bar{q}_{\alpha,n}(\psi)) \geq 1 - \alpha. \tag{A.27}$$

On the other hand, for each $n$, let $\{F_{n,m}\}$ be a sequence in $\mathscr{V}(\Phi,\varepsilon)$ such that $\bar{q}_{\alpha,n}(\psi) = \lim_{m\to\infty} q_{\alpha,n}(F_{n,m},\psi)$. Let $A_{n,m}$, $B_{n,m}$, $a_{n,m}$ and $b_{n,m}$ be defined as the corresponding $A_n$, $B_n$, $a_n$ and $b_n$, with $F$ replaced by $F_{m,n}$. By Theorem 1(a) we have

$$\begin{aligned} \inf_{F\in\mathscr{V}(\Phi,\varepsilon)} P_F(|\hat{\mu}_n| \leq \bar{q}_{\alpha,n}(\psi)) &\leq P_{F_{n,m}}(|\hat{\mu}_n| \leq \bar{q}_{\alpha,n}(\psi)) \\ &= \Phi(B_{n,m}) - \Phi(A_{n,m}) + r_n(F_{n,m}). \end{aligned} \tag{A.28}$$

For each $n$ we can find $m_n$ such that $\Phi(B_{n,m_n}) - \Phi(A_{n,m_n}) - [\Phi(b_{n,m_n}) - \Phi(a_{n,m_n})] \leq 1/n$. Then

$$\begin{aligned} \Phi(B_{n,m_n}) - \Phi(A_{n,m_n}) &\leq (\Phi(b_{n,m_n}) - \Phi(a_{n,m_n})) + \frac{1}{n} \\ &= 1 - \alpha + \frac{1}{n}. \end{aligned} \tag{A.29}$$

From (A.28) and (A.29) we get

$$\limsup_{n\to\infty} \inf_{F\in\mathscr{V}(\Phi,\varepsilon)} P_F(|\hat{\mu}_n| \leq \bar{q}_{\alpha,n}(\psi)) \leq 1 - \alpha, \tag{A.30}$$

and (a) follows now from (A.27) and (A.30).

(b) We will first show that $g(x_1,x_2)$ is nondecreasing in $x_1$ and $x_2$. Differentiation of (3.1) gives

$$\frac{\partial}{\partial x_1} g(x_1,x_2) = \frac{\phi[(g(x_1,x_2)-x_1)/\sqrt{x_2/n}] - \phi[(g(x_1,x_2)+x_1)/\sqrt{x_2/n}]}{\phi[(g(x_1,x_2)-x_1)/\sqrt{x_2/n}] + \phi[(g(x_1,x_2)+x_1)/\sqrt{x_2/n}]} \tag{A.31}$$

and

$$\begin{aligned} \frac{\partial}{\partial x_2} g(x_1,x_2) &= \frac{1}{2x_2} \\ &\times \frac{\phi\left[(g(x_1,x_2)-x_1)/\sqrt{x_2/n}\right](g(x_1,x_2)-x_1) + \phi\left[(g(x_1,x_2)+x_1)/\sqrt{x_2/n}\right](g(x_1,x_2)+x_1)}{\phi\left[(g(x_1,x_2)-x_1)/\sqrt{x_2/n}\right] + \phi\left[(g(x_1,x_2)+x_1)/\sqrt{x_2/n}\right]}. \end{aligned} \tag{A.32}$$

Since the right side of (A.31) has the same sign as $x_1$ for all real $x_1$ and all $x_2 > 0$, $g(x_1, x_2)$ is nondecreasing in $|x_1|$. If $|x_1| > g(x_1, x_2)$ then $(g(x_1, x_2) - x_1)(g(x_1, x_2) + x_1) < 0$ and the right side of (3.1) is less than 0.5. Since the left side of (3.1) is larger than or equal to 0.5, we must have $|x_1| \le g(x_1, x_2)$. Hence, $g(x_1, x_2) + x_1 \ge 0$ and $g(x_1, x_2) - x_1 \ge 0$ for all real $x_1$ and all $x_2 > 0$. Therefore, the right side of (A.32) is nonnegative.

Equation (A.27) can be obtained for any $\psi \in \mathscr{C}$ in the same way as in the proof of part (a), using Theorem 1(b) instead of Theorem 1(a), if necessary. In a similar way, (A.30) can be proved as in part (a) using Theorem 1(c) instead of Theorem 1(a). For this, observe that the sequence $F_{n,m}$ can be taken of the form $(1-\varepsilon)\Phi + \varepsilon\delta_{y_m}$, with $y_m \to \infty$. This follows from the following facts: $g(x_1, x_2)$ is monotone in $|x_1|$ and $x_2$ by part (a), $B(\varepsilon) = \lim_{y\to\infty} b((1-\varepsilon)\Phi + \varepsilon\delta_y, \psi)$ [see Lemma 1(c)] and we are assuming that $\sup_{F\in\mathscr{V}(\Phi,\varepsilon)} v(F, \psi) = \lim_{y\to\infty} v((1-\varepsilon)\Phi + \varepsilon\delta_y, \psi)$.

The rest of the Appendix is devoted to the proof of Theorem 4.

Let $\varepsilon > 0$ be fixed. For each $\mu > 0$ and $K > 0$ let $\mathscr{C}_{\mu,K}$ be the family of score functions $\psi \in \mathscr{C}$ satisfying [see (4.3)]

B1. $\langle\psi, \lambda_\mu\rangle_\mu = \psi(\infty)\varepsilon/(1-\varepsilon)$.
B2. $\psi(\infty) = K$.
B3. $\langle\psi, \delta_\mu\rangle_\mu = 1$.

Since the asymptotic variance and bias of estimates with score functions $\psi$ and $k\psi$ $(k > 0)$ are the same (actually, the estimates themselves are the same) B3 entails no loss of generality.

REMARK 2. It is easy to verify that $\lambda_t(y) = [\phi(y-t) - \phi(y+t)]/\Delta_t(y)$ and so, by (2.3) and (4.3),

$$\eta(t, \Phi, \psi) = -\int_{-\infty}^{\infty} \psi(y-t)\phi(y)\,dy = \langle\psi, \lambda_t\rangle_t. \tag{A.33}$$

By (2.8), then, B1 is equivalent to $B_\psi(\varepsilon) = \mu$.

REMARK 3. By Huber (1964), the median, with maxbias

$$\underline{\mu} = \underline{\mu}(\varepsilon) = \Phi^{-1}\left(\frac{1}{2(1-\varepsilon)}\right), \tag{A.34}$$

minimizes the maxbias among all translation equivariant estimates. Therefore,

$$B_\psi(\varepsilon) \ge \underline{\mu} \qquad \text{for all } \psi \in \mathscr{C}. \tag{A.35}$$

REMARK 4. B3 implies that (see the proof of Lemma 7)

$$\psi(\infty) \ge [\phi(0) + \phi(\mu)]^{-1} \equiv K_0(\mu) \quad \text{for all } \psi \in \mathscr{C}_{\mu,K}. \tag{A.36}$$

By Remarks 2, 3 and 4, we can restrict attention to

$$\mathscr{C}^* = \bigcup_{\underline{\mu}<\mu} \bigcup_{K \geq K_0(\mu)} \mathscr{C}_{\mu, K}.$$

For technical reasons we will also consider the family

$$\mathscr{C}^{**} = \bigcup_{\underline{\mu}<\mu\leq\bar{\mu}} \bigcup_{K \geq K_0(\mu)} \mathscr{C}_{\mu, K}, \tag{A.37}$$

where $\bar{\mu}$ is given by (4.8).

Consider now the special score functions

$$\varphi_{c,\mu}(y) = \begin{cases} \lambda_\mu(y), & \text{if } |y| \leq c, \\ \lambda_\mu(c) \text{ sign } (y), & \text{if } |y| > c, \end{cases} \tag{A.38}$$

and

$$\gamma_{c,\mu}(y) = \begin{cases} \delta_\mu(y), & \text{if } |y| \leq c, \\ \delta_\mu(c) \text{ sign } (y), & \text{if } |y| > c. \end{cases} \tag{A.39}$$

LEMMA 2. *Let $\mu \in (\underline{\mu}(\varepsilon), \bar{\mu}(\varepsilon))$ be fixed. Then*:

(a) *There exists $c_1 = c_1(\mu)$ such that $b(\Phi^\infty, \varphi_{c_1,\mu}) = \mu$. Therefore, $\bar{\varphi}_{c_1,\mu} = \varphi_{c_1,\mu}/\langle \varphi_{c_1,\mu}, \delta_\mu \rangle_\mu \in \mathscr{C}_{\mu, K_1}$ with*

$$K_1(\mu) = \frac{\varphi_{c_1,\mu}(c_1)}{\langle \varphi_{c_1,\mu}, \delta_\mu \rangle_\mu}. \tag{A.40}$$

(b) *There exists $c_2 = c_2(\mu)$ such that $b(\Phi^\infty, \gamma_{c_2,\mu}) = \mu$. Therefore, $\bar{\gamma}_{c_2,\mu} = \gamma_{c_2,\mu}/\langle \gamma_{c_2,\mu}, \delta_\mu \rangle_\mu \in \mathscr{C}_{\mu, K_2}$ with*

$$K_2(\mu) = \frac{\gamma_{c_2,\mu}(c_2)}{\langle \gamma_{c_2,\mu}, \delta_\mu \rangle_\mu}. \tag{A.41}$$

PROOF. To prove part (a) it suffices to show that the equation (in $c$, with fixed $\mu$)

$$H_1(c) \equiv \frac{\langle \varphi_{c,\mu}, \lambda_\mu \rangle_\mu}{\varphi_{c,\mu}(c)} = \frac{\varepsilon}{(1-\varepsilon)} \tag{A.42}$$

has a (unique) solution. This follows because

$$\begin{aligned} \lim_{c\to 0} H_1(c) &= 2\Phi(u) - 1 > 2\Phi(\underline{\mu}(\varepsilon)) - 1 \\ &= 2\Phi[\Phi^{-1}(0.5/(1-\varepsilon))] - 1 = \varepsilon/(1-\varepsilon) \end{aligned}$$

and $\lim_{c\to\infty} H_1(c) = \langle \lambda_\mu, \lambda_\mu \rangle_\mu < \langle \lambda_{\bar{\mu}(\varepsilon)}, \lambda_{\bar{\mu}(\varepsilon)} \rangle_{\bar{\mu}(\varepsilon)} = \varepsilon/(1-\varepsilon)$. The last inequality follows by Remark 1. The proof of part (b) follows from a similar analysis of the equation

$$H_2(c) \equiv \frac{\langle \gamma_{c,\mu}, \lambda_\mu \rangle_\mu}{\gamma_{c,\mu}(c)} = \frac{\varepsilon}{(1-\varepsilon)}. \tag{A.43}$$

In fact, $\lim_{c\to 0} H_2(c) = 2\Phi(u) - 1 > \varepsilon/(1-\varepsilon)$ and $\lim_{c\to\infty} H_2(c) = 0$. $\square$

LEMMA 3. *Suppose that* $\psi_{a,b,c,\mu}$ *given by* (4.1) *belongs to* $\mathscr{C}_{\mu,K}$, *for some positive constants* $a$, $b$ *and* $c$ *and that* $g$ *satisfies* C1–C5. *Then* $l_g(\psi) \geq l_g(\psi_{a,b,c,\mu})$ *for all* $\psi \in \mathscr{C}_{\mu,K}$.

PROOF. Since $\psi$ and $\psi_{a,b,c,\mu}$ belong to $\mathscr{C}_{\mu,K}$ it suffices to show that

$$\langle \psi, \psi \rangle_\mu \geq \langle \psi_{a,b,c,\mu}, \psi_{a,b,c,\mu} \rangle_\mu.$$

Let

$$C(\mu, K) = a^2 \langle \delta_\mu, \delta_\mu \rangle_\mu + b^2 \langle \lambda_\mu, \lambda_\mu \rangle_\mu - 2a - 2bK\varepsilon/(1-\varepsilon) + 2ab\langle \delta_\mu, \lambda_\mu \rangle_\mu.$$

Since $\psi(y) \leq K$ on $[0, \infty)$, by the definition of $\psi_{a,b,c,\mu}$,

$$\begin{aligned}
\langle \psi, \psi \rangle_\mu + C(\mu, K) &= \int_0^\infty [\psi(y) - a\delta_\mu(y) - b\lambda_\mu(y)]^2 \Delta_\mu(y)\, dy \\
&\geq \int_0^\infty [\psi_{a,b,c,\mu}(y) - a\delta_\mu(y) - b\lambda_\mu(y)]^2 \Delta_\mu(y)\, dy \\
&= \langle \psi_{a,b,c,\mu}, \psi_{a,b,c,\mu} \rangle_\mu + C(\mu, K),
\end{aligned}$$

and this proves the lemma. $\square$

LEMMA 4. *Suppose that* $g$ *satisfies* C1–C5. *Let* $0 \leq \varepsilon < 0.5$ *and* $\mu \in (\underline{\mu}(\varepsilon), \bar{\mu}(\varepsilon))$ *be fixed. Let* $c_1 = c_1(\mu)$, $K_1 = K_1(\mu)$ *and* $\bar{\varphi}_{c_1,\mu}$ *be as given by Lemma* 2(a). *Then*

$$l_g(\psi) \geq l_g(\bar{\varphi}_{c_1,\mu}) \qquad \forall \psi \in \mathscr{C}_{\mu,K} \quad \text{with } K \geq K_1.$$

PROOF. To prove the lemma it suffices to show that

$$\langle \psi, \psi \rangle_\mu \geq \langle \bar{\varphi}_{c_1,\mu}, \bar{\varphi}_{c_1,\mu} \rangle_\mu.$$

Let $\hat{\psi}(y) = \psi(y)/\psi(\infty) = \psi(y)/K$ and $\hat{\varphi}_{c_1,\mu}(y) = \bar{\varphi}_{c_1,\mu}(y)/\bar{\varphi}_{c_1,\mu}(\infty) = \bar{\varphi}_{c_1,\mu}(y)/K_1$. Noticing that $\langle \hat{\psi}, \lambda_\mu \rangle_\mu = \langle \hat{\varphi}_{c_1,\mu}, \lambda_\mu \rangle_\mu = \varepsilon/(1-\varepsilon)$, that $\lambda_\mu(y)$ is strictly increasing in $y$ and using the definition of $\hat{\varphi}_{c_1,\mu}$, we have

$$\begin{aligned}
0 &\leq \int_0^\infty [\hat{\psi}(y) - \lambda_\mu(y)/\lambda_\mu(c_1)]^2 \Delta_\mu(y)\, dy \\
&\quad - \int_0^\infty [\hat{\varphi}_{c_1,\mu}(y) - \lambda_\mu(y)/\lambda_\mu(c_1)]^2 \Delta_\mu(y)\, dy \\
&= \langle \hat{\psi}, \hat{\psi} \rangle_\mu - \langle \hat{\varphi}_{c_1,\mu}, \hat{\varphi}_{c_1,\mu} \rangle_\mu.
\end{aligned}$$

Therefore, $\langle \psi, \psi \rangle_\mu \geq K^2 \langle \bar{\varphi}_{c_1,\mu}, \bar{\varphi}_{c_1,\mu} \rangle_\mu / K_1^2 \geq \langle \bar{\varphi}_{c_1,\mu}, \bar{\varphi}_{c_1,\mu} \rangle_\mu$. $\square$

LEMMA 5. *Suppose that* $g$ *satisfies* C1–C5. *Let* $0 \leq \varepsilon < 0.5$ *and* $\mu \in (\underline{\mu}(\varepsilon), \bar{\mu}(\varepsilon))$ *be fixed. Let* $c_2 = c_2(\mu)$ *and* $K_2 = K_2(\mu)$ *be the constants given by Lemma* 2(b). *Then for each* $\psi \in \mathscr{C}_{\mu,K}$ *with* $K \leq K_2$ *there exists* $c \leq c_2$ *such that*

$$l_g(\psi) \geq l_g(\bar{\gamma}_{c,\mu}).$$

PROOF. Let $\bar{\gamma}_{c,\mu}(y) = \gamma_{c,\mu}(y)/\langle\gamma_{c,\mu}, \delta_\mu\rangle_\mu$ and notice that $K_2 = \bar{\gamma}_{c_2,\mu}(\infty)$. It is not difficult to show that $\bar{\gamma}_{c,\mu}(c)$ is nondecreasing in $c$ and therefore there exists $c^* \le c_2$ such that $\bar{\gamma}_{c^*,\mu}(c^*) = K$. To simplify the notations, set $\gamma^* = \bar{\gamma}_{c^*,\mu}$. By definition, $\gamma^*$ satisfies B2 and B3. Furthermore, since $c^* \le c_2$, $\mu^* = b(\Phi^\infty, \gamma_{c^*}) \le \mu$. Therefore, it suffices to show that

$$\langle\psi, \psi\rangle_\mu \ge \langle\gamma^*, \gamma^*\rangle_{\mu^*}. \tag{A.44}$$

Noticing that $\langle\psi, \delta_\mu\rangle_\mu = \langle\gamma^*, \delta_\mu\rangle_\mu = 1$, $\delta_\mu(y)$ is strictly increasing in $y$ and using the definition of $\gamma^*$ and a reasoning similar to the proof of Lemma 3 gives

$$\langle\psi, \psi\rangle_\mu \ge \langle\gamma^*, \gamma^*\rangle_\mu$$

and the lemma follows from the monotonicity (in $\mu$) of $\langle\psi, \psi\rangle_\mu$. □

LEMMA 6. *Suppose that $g$ satisfies* C1–C5. *Let $0 \le \varepsilon < 0.5$ be fixed and let $\psi \in \mathscr{C}_{\mu,K}$, with $\mu \in (\underline{\mu}(\varepsilon), \bar{\mu}(\varepsilon))$ and $K$ between $K_1(\mu)$ and $K_2(\mu)$. Then there exists $c$ between $c_1(\underline{\mu})$ and $c_2(\mu)$, $a$ and $b$ such that $l_g(\psi_{a,b,c,\mu}) \le l_g(\psi)$. Moreover, $a = \tilde{a}(c)$ and $b = \tilde{b}(c)$ depend only on $c$ and are given by* (A.45) *and* (A.46).

PROOF. Since $v(\Phi^\infty, \psi) \ge 1$ and $g[\underline{\mu}(\varepsilon), (2(1-\varepsilon)\phi(\underline{\mu}(\varepsilon))^{-1}] = l_g(\text{sign})$, we only need to consider score functions with $b(\Phi^\infty, \psi) = \mu$ and $\underline{\mu}(\varepsilon) \le \mu < \bar{\mu}(\varepsilon) - \delta$. Let $K = \psi(\infty)$ and let $c_1 = c_1(\mu)$ and $c_2 = c_2(\mu)$ as defined by (A.42) and (A.43). In addition let $K_1 = K_1(\mu)$ and $K_2 = K_2(\mu)$ be given by (A.40) and (A.41). Let

$$A_{11}(c) = \int_0^c \delta_\mu(y)\delta_\mu(y)\Delta_\mu(y)\,dy + \delta_\mu(c)\int_c^\infty \delta_\mu(y)\Delta_\mu(y)\,dy,$$

$$A_{12}(c) = \int_0^c \lambda_\mu(y)\delta_\mu(y)\Delta_\mu(y)\,dy + \lambda_\mu(c)\int_c^\infty \delta_\mu(y)\Delta_\mu(y)\,dy,$$

$$A_{21}(c) = \int_0^c \delta_\mu(y)\lambda_\mu(y)\Delta_\mu(y)\,dy + \delta_\mu(c)\int_c^\infty \lambda_\mu(y)\Delta_\mu(y)\,dy - \delta_\mu(c)\varepsilon/(1-\varepsilon),$$

$$A_{22}(c) = \int_0^c \lambda_\mu(y)\lambda_\mu(y)\Delta_\mu(y)\,dy + \lambda_\mu(c)\int_c^\infty \lambda_\mu(y)\Delta_\mu(y)\,dy - \lambda_\mu(c)\varepsilon/(1-\varepsilon).$$

The function $\psi_{a,b,c,\mu}$ satisfies B1 and B3 if and only if there exist $a \ge 0$ and $b \ge 0$ satisfying

$$bA_{11}(c) + aA_{12}(c) = 1,$$
$$bA_{21}(c) + aA_{22}(c) = 0,$$

that is, if and only if

$$\tilde{a}(c) = \frac{-A_{21}(c)}{A_{11}(c)A_{22}(c) - A_{12}(c)A_{21}(c)} \ge 0, \tag{A.45}$$

$$\tilde{b}(c) = \frac{A_{22}(c)}{A_{11}(c)A_{22}(c) - A_{12}(c)A_{21}(c)} \ge 0. \tag{A.46}$$

Clearly $A_{11}(c)$ and $A_{12}(c)$ are positive. Suppose first that $c_1 < c_2$ since earlier truncations produce smaller maxbiases, $A_{21}(c) \le 0$ and $A_{22}(c) \ge 0$ and so $\tilde{a}(c) \ge 0$ and $\tilde{b}(c) \ge 0$. The cases $c_1 = c_2$, $c_2 < c_1$ can be analyzed in a similar way. Set $\psi_{\tilde{a}(c), \tilde{b}(c), c, \mu}(\infty) = K(c)$. Clearly, $K(c)$ is continuous and takes all the values between $K(c_1) = K_1$ and $K(c_2) = K_2$. This proves the lemma. □

LEMMA 7. *Let* $0 \le \varepsilon < 0.5$ *and suppose that* $b(\Phi^\infty, \psi) \ge \bar{\mu}(\varepsilon)$. *Then* $v(\Phi^\infty, \psi) \ge \underline{v}(\varepsilon)$, *where* $\underline{v}(\varepsilon)$ *is given by* (4.9).

PROOF. Let $\mu = b(\Phi^\infty, \psi)$. It is easy to see that

$$\langle \psi, \delta_\mu \rangle_\mu = -\int_{-\infty}^{\infty} \psi(x)\phi'(x+\mu)\,dx = \int_{-\infty}^{\infty} \psi(x)(x+\mu)\phi(x+\mu)\,dx.$$

Therefore, by the Cauchy–Schwarz inequality,

$$\begin{aligned}
\langle \psi, \delta_\mu \rangle_\mu^2 &= \left[\int_{-\infty}^{\infty} \psi(x)(x+\mu)\phi(x+\mu)\,dx\right]^2 \\
&\le \int_{-\infty}^{\infty} \psi^2(x)\phi(x+\mu)\,dx \int_{-\infty}^{\infty} (x+\mu)^2\phi(x+\mu)\,dx \qquad \text{(A.47)} \\
&= \int_{-\infty}^{\infty} \psi^2(x)\phi(x+\mu)\,dx = \langle \psi, \psi \rangle_\mu.
\end{aligned}$$

On the other hand,

$$\begin{aligned}
\langle \psi, \delta_\mu \rangle_\mu &= -\int_{0}^{\infty} \psi(x)[\phi'(x+\mu) + \phi'(x-\mu)]\,dx \\
&= \int_{0}^{\infty} \psi(x)(x+\mu)\phi(x+\mu)\,dx + \int_{0}^{\infty} \psi(x)(x-\mu)\phi(x-\mu)\,dx \\
&\le \int_{0}^{\infty} \psi(x)(x+\mu)\phi(x+\mu)\,dx + \int_{\mu}^{\infty} \psi(x)(x-\mu)\phi(x-\mu)\,dx \qquad \text{(A.48)} \\
&\le \psi(\infty)\left[\int_{\mu}^{\infty} x\phi(x)\,dx + \int_{0}^{\infty} x\phi(x)\,dx\right] \\
&= \psi(\infty)[\phi(\mu) + \phi(0)].
\end{aligned}$$

The lemma follows now from (4.6), (A.47) and (A.48). □

PROOF OF THEOREM 4. For each $\mu \in (\underline{\mu}(\varepsilon), \bar{\mu}(\varepsilon))$, let $I_2(\mu)$ be the closed interval with endpoints $K_1(\mu)$ and $K_2(\mu)$. Consider the set

$$\mathscr{C}^{***} = \bigcup_{\underline{\mu}<\mu<\bar{\mu}} \; \bigcup_{K \in I_2(\mu)} \mathscr{C}_{\mu, K},$$

and let $C = C_1 \cup C_2$, where $C_1$ and $C_2$ are given by (4.13) and (4.14), respectively. Since the function $l_g(\psi_{a,b,c,\mu})$ is lower semicontinuous and $C$ is compact, there exists $\psi^* = \psi_{a^*, b^*, c^*, \mu^*}$ such that

$$l_g(\psi_{a,b,c,\mu}) \ge l_g(\psi^*) \qquad \forall (a, b, c, \mu) \in C.$$

By Lemma 6,

$$l_g(\psi) \geq l_g(\psi^*) \quad \forall\ \psi \in \mathscr{C}^{***},$$

and by Lemmas 4 and 5,

$$l_g(\psi) \geq l_g(\psi^*) \quad \forall\ \psi \in \mathscr{C}^{**}.$$

By Lemma 7 and (4.11), if $b(\Phi^\infty, \psi) \geq \bar{\mu}(\varepsilon)$,

$$l_g(\psi) \geq g(\bar{\mu}(\varepsilon), \underline{v}(\varepsilon)) \geq l_g(\psi^*).$$

Therefore,

$$l_g(\psi) \geq l_g(\psi^*) \qquad \forall\ \psi \in \mathscr{C}^*.$$

Finally, by (4.12),

$$L_g(\psi) \geq l_g(\psi) \geq l_g(\psi^*) = L_g(\psi^*) \qquad \forall\ \psi \in \mathscr{C}^*.$$

LEMMA 8. *Let* $\psi_c^H$ *be given by* (2.12). *Then*

$$\sup_{F \in \mathscr{V}(\Phi,\varepsilon)} v(F, \psi_c^H) = v(\Phi^\infty, \psi_c^H).$$

PROOF. First notice that

$$\sup_{F \in \mathscr{V}(\Phi,\varepsilon)} |T(F, \psi_c^H)| \leq T(\Phi^\infty, \psi_c^H).$$

The contamination part in the numerator of $v(F, \psi_c^H)$ is bounded above by $\varepsilon c^2$ and the contamination part in the denominator of $v(F, \psi_c^H)$ is bounded below by 0. The lemma follows now because the normal parts in the numerator and denominator of $v(F, \psi_c^H)$ are monotone functions of $|t|$, with $\mathrm{E}_\Phi\{\psi_c^{H^2}(X-t)\}$ being nondecreasing in $|t|$ and $\mathrm{E}_\Phi\{\psi_c^{H'}(X-t)\} = \Phi(c+t) - \Phi(c-t)$ being nonincreasing in $|t|$. $\square$

## REFERENCES


DAVIES, P. L. (1998). On locally uniformly linearizable high breakdown location and scale functionals. *Ann. Statist.* **26** 1103–1125.

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

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

HUBER, P. J. (1968). Robust confidence limits. *Z. Wahrsch. Verw. Gebiete* **10** 269–278.

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

HUBER—CAROL, C. (1970). Etude asymptotique de tests robustes. Ph.D. thesis, Eidgen Technische Hochschule, Zürich.

MARTIN, R. D. and ZAMAR, R. H. (1993). Efficiency-constrained bias-robust estimates of location. *Ann. Statist.* **21** 338–354.

RIEDER, H. (1994). *Robust Asymptotic Statistics*. Springer, Berlin.

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

SALIBIAN-BARRERA, M. (2000). Contributions to the theory of robust inference. Ph.D. dissertation, Dept. Statistics, Univ. British Columbia, Vancouver, Canada. Available online at ftp://ftp.stat.ubc.ca/pub/matias/Thesis.

SAMAROV, A. M. (1985). Bounded-influence regression via local minimax mean squared error. *J. Amer. Statist. Assoc.* **80** 1032–1040.

STIGLER, S. M. (1977). Do robust estimators work with real data? *Ann. Statist.* **5** 1055–1098

R. FRAIMAN
DEPARTAMENTO DE MATEMÁTICAS
UNIVERSIDAD DE SAN ANDRÉS
VITO DUMAS ESQ. ARIAS
1644 VICTORIA
ARGENTINA
E-MAIL: rfraiman@udesa.edu.ar

V. J. YOHAI
DEPARTAMENTO DE MATEMÁTICAS
CIUDAD UNIVERSITARIA, PABELLÓN 1
1426 BUENOS AIRES
ARGENTINA
E-MAIL: vyohai@dm.uba.ar

R. ZAMAR
DEPARTMENT OF STATISTICS
UNIVERSITY OF BRITISH COLUMBIA
ROOM 333, 6356 AGRICULTURAL ROAD
VANCOUVER
CANADA V6T 1Z2
E-MAIL: ruben@stat.ubc.ca