The Annals of Statistics
1989, Vol. 17, No. 1, 327–336

MINIMAX PROPERTIES OF *M*-, *R*- AND *L*-ESTIMATORS OF LOCATION IN LÉVY NEIGHBOURHOODS[1]

BY JOHN COLLINS AND DOUGLAS WIENS

University of Calgary and University of Alberta

In the context of Huber's theory of robust estimation of a location parameter, the literature on minimax properties of M-, R- and L-estimators is surveyed. New results are obtained for the model in which the unknown error distribution is assumed to lie in a Lévy neighbourhood of a symmetric distribution G: $\mathscr{P}_{\varepsilon,\delta}(G) = \{F | G(x-\delta) - \varepsilon \le F(x) \le G(x+\delta) + \varepsilon$ for all $x\}$. Under reasonably general conditions on G, the distribution F_0 in $\mathscr{P}_{\varepsilon,\delta}(G)$ which minimizes Fisher information for location is found. Huber's minimax property for M-estimators is shown to hold for R-estimators but to fail for L-estimators in Lévy neighbourhoods. The latter is proved by constructing a subneighbourhood of distributions $\mathscr{F}_0$, with $F_0 \in \mathscr{F}_0 \subset \mathscr{P}_{\varepsilon,\delta}(G)$, such that the asymptotic variance of the L-estimator which is asymptotically efficient at F_0 is minimized over $\mathscr{F}_0$ at F_0.

1. Introduction and summary. In the context of robust estimation of a location parameter, Huber (1964) found a general asymptotic minimax property for the class of M-estimators. In this section we survey the subsequent literature on the following two related problems: (1) finding the form of the minimax variance M-estimator corresponding to particular relevant models for the unknown neighbourhood of error distributions; and (2) ascertaining whether or not Huber's minimax variance property also holds for R-estimators and L-estimators in each such neighbourhood. We carry out programs (1) and (2) for the important Lévy neighbourhood model in Sections 2 and 3, respectively.

First we summarize Huber's minimax variance theory. Let $X_1, \dots, X_n$ be a random sample from a distribution $F(x-\theta)$, where θ is an unknown location parameter. Here F is an unknown member of a specified convex, vaguely compact neighbourhood, $\mathscr{F}$, of a fixed "ideal" distribution G which is symmetric about 0. Let $\mathscr{C}$ denote a class of estimators of θ—such as the M-estimators, R-estimators or L-estimators. If $\{T_n\}$ is a sequence of estimators in $\mathscr{C}$, then under mild regularity conditions, $n^{1/2}(T_n - \theta)$ converges in distribution to the normal law with mean 0 and variance $V(T, F)$.

Huber's (1964) minimax property for M-estimators is as follows. Let F_0 be the distribution which minimizes Fisher information for location,

$$I(F) = \int \left[(f')^2/f \right] dx, \quad \text{if } F \text{ has an absolutely continuous density } f,$$
$$= \infty, \quad \text{otherwise,}$$

Received April 1986; revised August 1987.

[1]Research supported by Natural Sciences and Engineering Research Council of Canada Grants A-4499 and A-8603.

AMS 1980 *subject classifications*. Primary 62F35; secondary 62F12.

Key words and phrases. Robust estimation, Lévy neighbourhood, M-, R- and L-estimators, minimax variance, minimum Fisher information.

over all F in $\mathscr{F}$. If T_0 denotes the M-estimator which is asymptotically efficient [i.e., $V(T_0, F_0) = 1/I(F_0)$] at F_0, then the minimum value of $\sup\{V(T, F)$: $F \in \mathscr{F}\}$ is $1/I(F_0)$, attained at T_0. Thus problem (1) reduces to the problem of finding the minimum information F_0 in $\mathscr{F}$. For other classes of estimators, however, whether or not the minimax property holds [problem (2)] depends on the neighbourhood $\mathscr{F}$ as well as upon the class $\mathscr{C}$ of estimators. For L- and R-estimators, problem (2) consists of determining whether or not

$$(1.1) \qquad \sup\{V(T_0, F)|F \in \mathscr{F}\} = 1/I(F_0),$$

where $T_0 \in \mathscr{C}$ is asymptotically efficient for F_0. The dual statement $\inf\{V(T, F_0)|T \in \mathscr{C}\} = 1/I(F_0)$ is, for each class of estimators, an easy consequence of the Cauchy–Schwarz inequality. See Section 4.7 of Huber (1981) for further background.

The neighbourhood model $\mathscr{F}$ which has been most thoroughly studied is the gross error or ε-contamination model $\mathscr{F} = \{F$: $F = (1 - \varepsilon)G + \varepsilon H$ for some distribution $H\}$, where ε is fixed, $0 < \varepsilon < 1$, and G is a fixed distribution symmetric about 0. Complete results have been obtained for problems (1) and (2) in the special case where G has a strongly unimodal density. See Huber (1964) or Example 5.2 on page 84 of Huber (1981) for details of the least informative F_0 and the corresponding minimax M-estimator in this case. Jaeckel (1971) proved that the minimax property holds for both L- and R-estimators when G has a strongly unimodal density. When the condition of strong unimodality of G is dropped, the results are not as complete. Collins and Wiens (1985) found least informative distributions in the ε-contamination model when G is quite general, but the corresponding question of whether the minimax property holds for L- and R-estimators is open and under investigation.

The only other neighbourhood model which has received extensive study is the Kolmogorov model $\mathscr{F} = \{F|G(x) - \varepsilon \le F(x) \le G(x) + \varepsilon$ for all $x\}$, where G is a fixed distribution symmetric about 0 and $\varepsilon > 0$ is fixed. The most complete results pertain to the special case where G is the standard normal distribution Φ. When $G = \Phi$, the least informative distribution was found by Huber (1964) for $\varepsilon < 0.0303$ and by Sacks and Ylvisaker (1972) for $\varepsilon \ge 0.0303$. In the same paper Sacks and Ylvisaker also discovered that the minimax property fails for the class of L-estimators when $\varepsilon > 0.07$. Collins (1983) showed that, for the class of R-estimators, the minimax property holds for all $\varepsilon \in (0, \frac{1}{2})$. For the case of G being a nonnormal distribution, Wiens (1985, 1986) obtained least informative distributions in Kolmogorov neighbourhoods of nonnormal G, subject to various regularity conditions.

In Sections 2 and 3, we study minimax variance properties of M-, L- and R-estimators when $\mathscr{F} = \mathscr{P}_{\varepsilon, \delta}(G)$, a Lévy neighbourhood of a distribution G:

$$(1.2) \qquad \mathscr{P}_{\varepsilon, \delta}(G) = \{F|G(x - \delta) - \varepsilon \le F(x) \le G(x + \delta) + \varepsilon \text{ for all } x\}.$$

Here ε and δ are assumed to be fixed, with $0 \le \varepsilon < \frac{1}{2}$ and $\delta \ge 0$; G is a fixed distribution symmetric about 0.

The Lévy model, discussed in Chapter 2 of Huber (1981), is an important neighbourhood structure in robust estimation theory. It is based on the "Lévy distance," which metrizes the weak topology [Theorem 3.3 of Huber (1981)].

From the point of view of practical application, the two-parameter family $\mathscr{P}_{\varepsilon,\delta}(G)$ allows wide flexibility in modelling the possible departures from G against which one wishes to protect. The choice $\delta = 0$ yields, of course, the Kolmogorov neighbourhood model as a special case. The choice $\varepsilon = 0$ yields a Lévy band about G whose width at x decreases to 0 as x approaches $\pm\infty$; this may be a more realistic model than the fixed-width Kolmogorov band.

In Section 2 the distribution F_0 is found which minimizes Fisher information over all F in $\mathscr{P}_{\varepsilon,\delta}(G)$. This is carried out, for all choices of ε and δ, under regularity conditions on G which are only slightly stronger than strong unimodality and which include the normal distribution and the logistic distribution as special cases. The minimum information F_0 is also found under some less restrictive conditions on G. The Cauchy and t-distributions are then included as special cases, although the solutions require restrictions on the choice of ε and δ. The minimum information distributions obtained are, not surprisingly, qualitatively similar to solutions previously obtained in the special case of Kolmogorov neighbourhoods by Huber (1964) and Sacks and Ylvisaker (1972) when $G = \Phi$ and by Wiens (1986) for more general G.

In Section 3, we investigate whether the minimax property also holds for the R- and L-estimators that are asymptotically efficient at the minimum information F_0 in $\mathscr{P}_{\varepsilon,\delta}(G)$. Under the conditions on G of Section 2, it is shown that the minimax property does hold for R-estimators but fails for L-estimators. The proof for R-estimators is a direct generalization of Collins' (1983) proof for the special case $G = \Phi$ and $\delta = 0$. But the proof of the failure of the minimax property for L-estimators is quite different from the proof of the special case $G = \Phi$, $\delta = 0$, $\varepsilon > 0.07$ given by Sacks and Ylvisaker (1972). Their method was to show that there is an $F_1 \in \mathscr{P}_{\varepsilon,0}(\Phi)$ for which $V(L_0, F_0) < V(L_0, F_1)$, where L_0 denotes the L-estimator which is asymptotically efficient at F_0. Their method entails numerical approximations that do not generalize easily. Our method requires no approximations: We show that there is a subset $\mathscr{F}_0 \subset \mathscr{P}_{\varepsilon,\delta}(G)$ over which $V(L_0, F)$ is nonconstant and attains its minimum value at F_0. The proof is based on a simple comparison of the influence curve of L_0 at F_0, and at other $F \in \mathscr{F}_0$.

In summary, results on the structure of minimum information distributions and on the minimax property for R- and L-estimators are now fairly complete for both the ε-contamination model and the Lévy neighbourhood model (including the special case of the Kolmogorov neighbourhood model). A conspicuous gap involves the investigation of the minimax property for L- and R-estimators in ε-contamination neighbourhoods of nonstrongly unimodal distributions. A further area of useful research is the extension of results on the minimax property to other neighbourhoods and to other classes of estimators besides M-, L- and R-estimators. An example of another class, $\mathscr{C}$, of location parameter estimators, large enough to contain an asymptotically efficient member corresponding to each F in $\mathscr{F}$, is the class of Cramér–von Mises estimators—see Boos (1981) or Parr and de Wet (1981) for details. Wiens (1987) has recently proved that the minimax property holds for $\mathscr{C}$ when $\mathscr{F}$ is an ε-contamination neighbourhood of a strongly unimodal distribution.

Another area for further research is to find general conditions on classes of estimators, $\mathscr{C}$, and on classes of distributions, $\mathscr{F}$, under which the minimax property holds or fails to hold. This general problem is posed in a paper by Sacks and Ylvisaker (1982), in which a neighbourhood $\mathscr{F}$ is constructed for which the minimax property fails for both the classes of L-estimators and R-estimators. Some progress has been made toward obtaining general answers by generalizing a method implicit in the proof of our Theorem 4.

2. Minimum information distributions in $\mathscr{P}_{\varepsilon,\delta}$. Throughout this paper $\mathscr{P}_{\varepsilon,\delta}(G)$ denotes a Lévy neighbourhood as defined by (1.2). We shall assume

ASSUMPTION A. The distribution function $G(x)$ is symmetric about 0 and proper ($G(\infty) = 1$), with an absolutely continuous density $g(x)$ and twice continuously differentiable (except possibly at 0) score function $\xi(x) = -g'(x)/g(x)$.

In Theorem 1 below, we shall as well assume

ASSUMPTION B. The function $J(\xi)(x) = 2\xi'(x) - \xi^2(x)$ is strictly decreasing on $(0, \infty)$ and $\xi(0^+) \geq 0$.

In Theorem 2, we assume either Assumption B or

ASSUMPTION C. (i) $\xi(x)$ is positive and $x\xi(x)$ is strictly increasing, on $(0, \infty)$, (ii) $\xi(x)/x$ is nonincreasing on $(0, \infty)$ and (iii) $\xi(x)$ has no local minima in $(\bar{A}, \infty)$, where $\bar{A}$ is defined by $\bar{A}\xi(\bar{A}) = 1$.

As in Lemma 1 of Wiens (1986), Assumption B implies that ξ is positive and strictly increasing on $(0, \infty)$, so that g is strongly unimodal.

Examples of distributions satisfying Assumption B are the logistic, normal and more generally those with densities $g_k(x)$ proportional to $\exp(-|x|^k/k)$, $1 < k \leq 2$. Some distributions satisfying Assumption C but not Assumption B are the Student's t and those with densities $g_k(x)$, $k \leq 1$.

The motivation behind Theorems 1 and 2 below is discussed in Wiens (1985, 1986), where they were proved for $\delta = 0$. Recall [Huber (1981)] that the necessary and sufficient condition for $F_0 \in \mathscr{P}_{\varepsilon,\delta}$ to minimize information there is

$$\int_{-\infty}^{\infty} J(\psi_0)(x)\, d(F - F_0)(x) \geq 0 \tag{2.1}$$

for all $F \in \mathscr{P}_{\varepsilon,\delta}$ with $I(F) < \infty$, where $\psi_0 = -f_0'/f_0$.

The proofs of Theorems 1 and 2, together with some tables of numerical values of the constants in the case $G = \Phi$, may be found in a technical report by Collins and Wiens (1986). The proofs consist of showing that, in each case, the exhibited F_0 exists, belongs to $\mathscr{P}_{\varepsilon,\delta}$ and satisfies (2.1).

THEOREM 1. *Make Assumptions A and B. Then there is a positive number ε_*, depending upon G, and a function $\delta_*(\varepsilon)$, such that for $0 \leq \varepsilon \leq \varepsilon_*$ and*

$0 \leq \delta \leq \delta_*(\varepsilon)$, *the minimum information* $F_0 \in \mathscr{P}_{\varepsilon,\delta}(G)$ *has density* f_0 *and score function* $\psi_0 = -f_0'/f_0$ *given by*

$$f_0(x) = f_0(-x) = \begin{cases} \dfrac{g(a-\delta)}{\cos^2(\lambda_1 a/2)} \cos^2 \dfrac{\lambda_1 x}{2}, & x \in [0, a], \\ g(x-\delta), & x \in (a, b), \\ g(b-\delta)\exp(-\lambda_2(x-b)), & x \in [b, \infty), \end{cases}$$

and

$$\psi_0(x) = -\psi_0(-x) = \begin{cases} \lambda_1 \tan \dfrac{\lambda_1 x}{2}, & x \in [0, a], \\ \xi(x-\delta), & x \in (a, b), \\ \lambda_2, & x \in [b, \infty), \end{cases}$$

where $\lambda_2 = \xi(b-\delta)$.

The three constants a, b and λ_1 ($b \geq a \geq \delta$, $\lambda_1 \geq 0$) are determined in terms of δ and ε by the conditions

(i) $$F_0(a) = G(a-\delta) - \varepsilon,$$

(ii) $$F_0(\infty) = 1,$$

(iii) $$\lambda_1 \tan \frac{\lambda_1 a}{2} = \xi(a-\delta).$$

The curve $\delta_*(\varepsilon)$ is decreasing from ∞ at $\varepsilon = 0$ to 0 at $\varepsilon = \varepsilon_*$ and is defined by (i)–(iii) together with $b = a$.

THEOREM 2. *Make Assumption A and either Assumption B or Assumption C. Then there is a positive number* ε_*, *depending upon* G, *such that for all* $\varepsilon \in [\varepsilon_*, \frac{1}{2}]$ *and all* $\delta \in [0, \infty)$, *the minimum information* $F_0 \in \mathscr{P}_{\varepsilon,\delta}(G)$ *has density and score function given by*

$$f_0(x) = f_0(-x) = \begin{cases} \dfrac{g(a-\delta)}{\cos^2(\lambda_1 a/2)} \cos^2 \dfrac{\lambda_1 x}{2}, & x \in [0, a], \\ g(a-\delta)\exp(-\lambda(x-a)), & x \in (a, \infty), \end{cases}$$

and

$$\psi_0(x) = -\psi_0(-x) = \begin{cases} \lambda_1 \tan \dfrac{\lambda_1 x}{2}, & x \in [0, a], \\ \lambda, & x \in (a, \infty), \end{cases}$$

where $\lambda = \lambda_1 \tan(\lambda_1 a/2)$. *The constants* a *and* λ_1 *are determined by* (i) $F_0(a) = G(a-\delta) - \varepsilon$ *and* (ii) $F_0(\infty) = 1$ *and satisfy as well* (iii′) $\lambda_1 \tan(\lambda_1 a/2) \leq \xi(a-\delta)$.

If Assumption B holds, then this ε_* coincides with that of Theorem 1. The solution is then also valid for $0 \le \varepsilon \le \varepsilon_*$, $\delta_*(\varepsilon) \le \delta < \infty$, where $\delta_*(\varepsilon)$ is as in Theorem 1.

COROLLARY 1. *Under Assumptions A and B, the minimum information $F_0 \in \mathscr{P}_{\varepsilon,\delta}$ is as described by Theorem 1, for $0 \le \varepsilon \le \varepsilon_*$, $0 \le \delta \le \delta_*(\varepsilon)$, and by Theorem 2, for all remaining $\varepsilon \le \frac{1}{2}$, $\delta < \infty$.*

REMARK 1. In both Theorems 1 and 2, the minimum information F_0 has the property that on each interval of x's for which $F_0(x)$ does not coincide with a boundary of the Lévy band, the corresponding $\psi_0(x)$ is a solution to a differential equation of form $J(\psi_0) = 2\psi_0' - \psi_0^2 \equiv$ constant. The minimum information distributions of Theorems 1 and 2 share the following common feature with the cases discussed in Section 1: namely, that Huber's variational condition (2.1) forces $J(\psi_0)$ to be constant on each interval where F_0 can vary freely.

REMARK 2. In the proof of each of Theorem 1 and Theorem 2, the easy part is the verification that the exhibited F_0 satisfies (2.1); the hard part is showing that F_0 lies in $\mathscr{P}_{\varepsilon,\delta}$. For this part, the assumptions on G (Assumption A, along with either Assumption B or Assumption C) are strongly required in the proof. Although these sufficient conditions on G may not be necessary, one can easily construct G's for which the assumptions are violated and the conclusions of the theorems fail.

REMARK 3. Corollary 1 applies to the logistic and normal distributions and more generally to those with densities $g_k(x)$, $1 < k \le 2$. For the Laplace distribution ($k = 1$) it applies as well, with $\varepsilon_* = 0$.

REMARK 4. For the special case $P_{\varepsilon,\varepsilon}(\Phi)$, the solution is as in Theorem 1 for $\varepsilon \le \varepsilon_0 = 0.02556$, and as in Theorem 2 for $\varepsilon \in [\varepsilon_0, \frac{1}{2}]$. For $\epsilon \le \varepsilon_0$, this was also proved by Kabatepe (1985), who as well correctly conjectured the form of the solution for $\varepsilon > \varepsilon_0$.

3. Minimax properties of *M*-, *R*- and *L*-estimators. Consider the M-, R- and L-estimators of θ which are asymptotically efficient at the minimum information F_0 in $\mathscr{P}_{\varepsilon,\delta}(G)$. Using the definitions and notation of Chapter 3 of Huber (1981), the efficient M-, R- and L-estimators have score functions

$$\psi_0(x) = -f_0'(x)/f_0(x),$$
$$J_0(u) = \psi_0(F_0^{-1}(u))$$

and

$$m_0(u) = \psi_0'[F_0^{-1}(u)]/I(F_0),$$

respectively. It follows from general theory (see the introductory remarks) that the minimum possible value (among all M-estimators of θ) of the supremum of the asymptotic variance as F ranges over $\mathscr{P}_{\varepsilon,\delta}$ is $1/I(F_0)$, attained by ψ_0 at F_0.

We now check whether this minimax property also holds for the R- and L-estimators which are asymptotically efficient at F_0. Throughout this section we shall use the usual formulas for the asymptotic variances of R- and L-estimators without discussion of the regularity conditions under which asymptotic normality holds. For such regularity conditions, see Huber (1981) or Serfling (1980).

Consider first the R-estimator with score function $J_0(u) = \psi_0(F_0^{-1}(u))$, $0 < u < 1$. Its asymptotic variance, under those distributions F in $\mathscr{P}_{\varepsilon,\delta}(G)$ with absolutely continuous density f, is given by

$$V(J_0, F) = \frac{\int J_0^2[F(x)]\, f(x)\, dx}{\left[-\int J_0[F(x)]\, f'(x)\, dx\right]^2}. \tag{3.1}$$

THEOREM 3. *Suppose that F_0 is the minimum information distribution in $\mathscr{P}_{\varepsilon,\delta}(G)$ which is either*: (i) *given by Theorem* 1 *under Assumptions A and B or* (ii) *given by Theorem* 2 *under Assumption A and either Assumption B or Assumption C. Then, with J_0 defined by $J_0(u) = \psi_0[F_0^{-1}(u)]$, $V(J_0, F)$ is maximized over $\mathscr{P}_{\varepsilon,\delta}(G)$ at F_0, so that* (1.1) *and the minimax property hold.*

PROOF. See the proof of Theorem 3 in Collins and Wiens (1986). We remark that the omitted proof is closely patterned after the proof of the special case $G = \Phi$, $\delta = 0$ on page 1193 of Collins (1983). □

Now consider the L-estimator with score function $m_0(u) = \psi_0'[F_0^{-1}(u)]/I(F_0)$ for $u \in (0,1)$. The asymptotic variance of this estimator under $F \in \mathscr{P}_{\varepsilon,\delta}(G)$ is

$$V(m_0, F) = \int IC^2(x; F)\, dF,$$

where the influence curve $IC(x; F)$ is given by

$$IC(x; F) = \int_{-\infty}^{x} m_0(F(y))\, dy - \int_{-\infty}^{\infty} [1 - F(y)]\, m_0(F(y))\, dy.$$

Note that $V(m_0, F)$ can be written as $E_F IC^2(X; F) = \mathrm{Var}_F IC(X; F)$, where X is a random variable with distribution F, since $E_F IC(X; F) = 0$ for all $F \in \mathscr{P}_{\varepsilon,\delta}(G)$. A useful alternative version is

$$IC(F^{-1}(u); F) = -\int_0^1 \{I[u \le t] - t\} m_0(t)\, dF^{-1}(t).$$

If F is continuous, we then have

$$V(m_0, F) = \mathrm{Var}_U\left[IC(F^{-1}(U); F)\right],$$

where U denotes a uniform random variable on $[0,1]$. Note also that $IC(F_0^{-1}(u); F_0) = \psi(F_0^{-1}(u))/I(F_0)$, with $V(m_0, F_0) = 1/I(F_0)$.

THEOREM 4. *Suppose that F_0 is the minimum information distribution in $\mathscr{P}_{\varepsilon,\delta}(G)$ under the conditions of either Theorem* 1 *or Theorem* 2. *Then with $m_0(u) = \psi_0'[F_0^{-1}(u)]/I(F_0)$, we have that*

$$\sup\{V(m_0, F) \colon F \in \mathscr{P}_{\varepsilon,\delta}(G)\} > V(m_0, F_0),$$

so that (1.1) *and the minimax property fail for L-estimators.*

PROOF. Under the conditions of either Theorem 1 or Theorem 2, define a subset $\mathscr{F}_0$ of $\mathscr{P}_{\varepsilon,\delta}(G)$ as follows:

$$\mathscr{F}_0 = \{F \in \mathscr{P}_{\varepsilon,\delta}(G) | F \text{ is continuous and } F(x) = F_0(x) \text{ whenever } |x| \geq a\}.$$

We will show that $V(m_0, F)$ is nonconstant on $\mathscr{F}_0$ and attains its *minimum* value there at F_0. The first part of the proof will be to show that, for all $F \in \mathscr{F}_0$,

$$\operatorname{Cov}[IC(F^{-1}(U); F), IC(F_0^{-1}(U); F_0)] = \operatorname{Var}[IC(F_0^{-1}(U); F_0)]. \tag{3.2}$$

Then (3.2) immediately implies that

$$V(m_0, F_0) = \rho_F^2 V(m_0, F), \tag{3.3}$$

where ρ_F is the correlation between $IC(F_0^{-1}(U); F_0)$ and $IC(F^{-1}(U); F)$. The second part, completing the proof of the theorem, will be to show that $\rho_F^2 = 1$ for an $F \in \mathscr{F}_0$ if and only if $F \equiv F_0$.

To show that (3.2) holds for all F in $\mathscr{F}_0$, we first set

$$\eta(u, t) = -\{I[u \leq t] - t\} m_0(t).$$

Then for $F \in \mathscr{F}_0$, we calculate that

$$\begin{aligned}
&[I(F_0)]^2\{\operatorname{Cov}[IC(F^{-1}(U); F), IC(F_0^{-1}(U); F_0)] - \operatorname{Var}[IC(F_0^{-1}(U); F_0)]\}\\
&\quad = I^2(F_0)\int_0^1 IC(F_0^{-1}(u); F_0)\{IC(F^{-1}(u); F) - IC(F_0^{-1}(u); F_0)\}\, du\\
&\quad = I^2(F_0)\int_0^1\int_0^1 \eta(u, s)\, dF_0^{-1}(s)\left\{\int_0^1 \eta(u, t)\, d(F^{-1}(t) - F_0^{-1}(t))\right\} du \qquad (3.4)\\
&\quad = I^2(F_0)\int_0^1\left\{\int_0^1 \eta(u, t)\int_0^1 \eta(u, s)\, dF_0^{-1}(s)\, du\right\} d(F^{-1}(t) - F_0^{-1}(t))\\
&\quad = I(F_0)\int_0^1\left\{\int_0^1 \eta(u, t)\psi_0(F_0^{-1}(u))\, du\right\} d(F^{-1}(t) - F_0^{-1}(t)).
\end{aligned}$$

The second-to-last step in (3.4) follows from Fubini's theorem. The change of variables $t = F_0(z)$ and $u = F(z)$ yields that (3.4) is equal to

$$\int_{-\infty}^{\infty} K(z)\, d(q_F^{-1}(z) - z),$$

where $q_F^{-1}(z) = F^{-1}(F_0(z))$ and

$$\begin{aligned} K(z) &= I(F_0)\int_{-\infty}^{\infty} \eta(F_0(x), F_0(z))\psi_0(x)\, dF_0(x) \\ &= \int_{-\infty}^{\infty} f_0'(x)\psi_0'(z)[I(x \le z) - F_0(z)]\, dx \\ &= \psi_0'(z) f_0(z). \end{aligned}$$

So to show that (3.2) holds for all $F \in \mathscr{F}_0$, it suffices to show that

$$\int_{-\infty}^{\infty} \psi_0'(z) f_0(z)\, d\big(q_F^{-1}(z) - z\big) = 0 \tag{3.5}$$

for all $F \in \mathscr{F}_0$. But (3.5) follows immediately from the fact that $\psi_0'(z) f_0(z) \equiv C_0$ for $|z| < a$ and that $q_F(z) \equiv z$ for $|z| \ge a$ by the definition of $\mathscr{F}_0$.

Now suppose that F is a member of $\mathscr{F}_0$ for which $\rho_F^2 = 1$. We need to show that this implies that $F \equiv F_0$. But $\rho_F^2 = 1$, together with $E[IC(F^{-1}(u);\, F)] = E[IC(F_0^{-1}(u);\, F_0)] = 0$ implies that $IC(F^{-1}(u);\, F) = IC(F_0^{-1}(u);\, F_0)$ a.e. $u \in [0, 1]$, or equivalently,

$$\begin{aligned} &\int_0^1 t m_0(t)\, d\big(F^{-1}(t) - F_0^{-1}(t)\big) \\ &\quad = \int_u^1 m_0(t)\, d\big(F^{-1}(t) - F_0^{-1}(t)\big) \quad \text{a.e. } u \in [0,1]. \end{aligned} \tag{3.6}$$

Letting $u \to 1$ shows that the right side of (3.6) is 0 a.e. $u \in [0, 1]$. The change of variable $t = F_0(z)$ then yields

$$\int_{F_0^{-1}(u)}^{\infty} \psi_0'(z)\, d\big(q_F^{-1}(z) - z\big) = 0 \quad \text{a.e. } u \in [0,1]. \tag{3.7}$$

But since $q_F^{-1}(z) \equiv z$ for $|z| \ge a$ and $\psi_0'(z) > 0$ for $|z| < a$, (3.7) forces $q_F^{-1}(z) \equiv z$ for $|z| < a$. Thus $F(z) = F_0(z)$ for all z, and this completes the proof of the theorem. □

REFERENCES

BOOS, D. D. (1981). Minimum distance estimators for location and goodness of fit. *J. Amer. Statist. Assoc.* **76** 663–670.

COLLINS, J. R. (1983). On the minimax property for R-estimators of location. *Ann. Statist.* **11** 1190–1195.

COLLINS, J. R. and WIENS, D. P. (1985). Minimax variance M-estimators in ε-contamination models. *Ann. Statist.* **13** 1078–1096.

COLLINS, J. R. and WIENS, D. (1986). Minimax properties of M-, R-, and L-estimators of location in Lévy neighbourhoods. Research Report DALTRS-86-2, Dalhousie Univ.

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

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

JAECKEL, L. A. (1971). Robust estimation of location: Symmetry and asymmetric contamination. *Ann. Math. Statist.* **42** 1024–1034.

KABATEPE, M. (1985). Robust detection and the Lévy distance. M.S. thesis, Rensselaer Polytechnic Institute, Troy, N.Y.

PARR, W. C. and DE WET, T. (1981). On minimum Cramér–von Mises-norm parameter estimation. *Comm. Statist. A —Theory Methods* **10** 1149–1166.

SACKS, J. and YLVISAKER, D. (1972). A note on Huber's robust estimation of a location parameter. *Ann. Math. Statist.* **43** 1068–1075.

SACKS, J. and YLVISAKER, D. (1982). L- and R-estimation and the minimax property. *Ann. Statist.* **10** 643–645.

SERFLING, R. J. (1980). *Approximation Theorems of Mathematical Statistics*. Wiley, New York.

WIENS, D. (1985). Minimax variance M-estimators of location in Kolmogorov neighbourhoods. Research Report DALTRS-85-6, Dalhousie Univ.

WIENS, D. (1986). Minimax variance M-estimators of location in Kolmogorov neighbourhoods. *Ann. Statist.* **13** 724–732.

WIENS, D. (1987). Robust weighted Cramér–von Mises estimators of location, with minimax variance in ε-contamination neighbourhoods. *Canad. J. Statist.* **15** 269–278.

DEPARTMENT OF MATHEMATICS
AND STATISTICS
UNIVERSITY OF CALGARY
2500 UNIVERSITY DRIVE N.W.
CALGARY, ALBERTA
CANADA T2N 1N4

DEPARTMENT OF STATISTICS AND
APPLIED PROBABILITY
UNIVERSITY OF ALBERTA
EDMONTON, ALBERTA
CANADA T6G 2G1