

## Asymptotic Behavior of General $M$ -estimates for Regression and Scale with Random Carriers

Ricardo A. Maronna<sup>1</sup> and Victor J. Yohai<sup>2</sup>

<sup>1</sup> Universidad Nacional de La Plata, Argentina

<sup>2</sup> Universidad Nacional de Buenos Aires, Argentina  
and Centro de Estudios Macroeconómicos de Argentina

**Summary.** Let  $(\mathbf{x}_i, y_i)$  be a sequence of independent identically distributed random variables, where  $\mathbf{x}_i \in R^p$  and  $y_i \in R$ , and let  $\boldsymbol{\theta} \in R^p$  be an unknown vector such that  $y_i = \mathbf{x}_i' \boldsymbol{\theta} + u_i(*)$ , where  $u_i$  is independent of  $\mathbf{x}_i$  and has distribution function  $F(u/\sigma)$ , where  $\sigma > 0$  is an unknown parameter. This paper deals with a general class of  $M$ -estimates of regression and scale,  $(\boldsymbol{\theta}^*, \sigma^*)$ , defined as solutions of the system:  $\sum_i \phi(\mathbf{x}_i, r_i) \mathbf{x}_i = \mathbf{0}$ ,  $\sum_i \chi(|r_i|) = 0$ , where  $r_i = (y_i - \mathbf{x}_i' \boldsymbol{\theta}^*)/\sigma^*$ , with  $\phi: R^p \times R \rightarrow R$  and  $\chi: R \rightarrow R$ . This class contains estimators of  $(\boldsymbol{\theta}, \sigma)$  proposed by Huber, Mallows and Krasker and Welsch. The consistency and asymptotic normality of the general  $M$ -estimators are proved assuming general regularity conditions on  $\phi$  and  $\chi$  and assuming the joint distribution of  $(\mathbf{x}_i, y_i)$  to fulfill the model (\*) only approximately.

### 1. Introduction

Let  $(\mathbf{x}_i, y_i)$  be a sequence of independent identically distributed (i.i.d.) random variables, where  $\mathbf{x}_i \in R^p$  and  $y_i \in R$ , and let  $\boldsymbol{\theta} \in R^p$  be an unknown vector such that

$$y_i = \mathbf{x}_i' \boldsymbol{\theta} + u_i, \quad (1.1)$$

where  $u_i$  is independent of  $\mathbf{x}_i$  and has distribution function  $F(u/\sigma)$ , where  $\sigma > 0$  is an unknown scale parameter. It is well known that the classical least-squares estimator for  $\boldsymbol{\theta}$  is not robust, i.e. although it is optimal when  $F$  is normal, its efficiency for nonnormal  $F$  may be arbitrarily low, and its bias arbitrarily large, even for  $F$  arbitrarily close to the normal. Several classes of robust estimators have been proposed to deal with this problem. The most extensively studied ones are generalizations of maximum likelihood estimators, or " $M$ -estimators". This paper deals with the asymptotic behavior of a very general class of  $M$ -estimators of regression and scale,  $(\boldsymbol{\theta}^*, \sigma^*)$ , defined as solutions of the system

$$\sum_{i=1}^n \phi(\mathbf{x}_i, (y_i - \mathbf{x}_i' \boldsymbol{\theta})/\sigma) \mathbf{x}_i = \mathbf{0}, \quad (1.2)$$

$$\sum_{i=1}^n \chi(|y_i - \mathbf{x}_i' \boldsymbol{\theta}| / \sigma) = 0; \quad (1.3)$$

where  $\phi: R^p \times R \rightarrow R$ , and  $\chi: R \rightarrow R$ .

The best-known estimators of this type are those with  $\phi$  of the form

$$\phi(\mathbf{x}, u) = \psi(u) \quad (1.4)$$

for some function  $\psi$ . We shall call them “classical  $M$ -estimators” (Huber 1964 and 1973). They have robust efficiency of  $\psi$  is a properly chosen bounded function. An important example is the “Huber psi-function” defined as

$$\psi_k(u) = \text{sgn}(u) \min(|u|, k) \quad (1.5)$$

where  $k$  is a properly chosen parameter. Other possible choices for  $\psi$ , which yield greater efficiency for heavy-tailed  $F$  in (1.1), are the so-called “redescending” functions, i.e. functions such that  $\psi(u) \rightarrow 0$  when  $u \rightarrow \pm \infty$ . An important case is the “bisquare” (Beaton and Tukey, 1974) defined as  $\psi(u) = u(1 - u^2/k^2)^2 I(|u| \leq k)$  (where  $I$  is the indicator function and  $k$  is a parameter).

Note that the family (1.4) includes Maximum Likelihood Estimators (MLE) for model (1.1). An interesting case is the MLE for the Cauchy distribution, which is a redescending one.

The classical  $M$ -estimators are not “qualitatively robust” (Hampel 1971 and 1974) in the sense that they may have an arbitrarily large asymptotic bias when the joint distribution of  $(\mathbf{x}_i, y_i)$  follows model (1.1) only approximately. This is due to the fact that  $\phi(\mathbf{x}, u) \mathbf{x}$  is not bounded (Maronna, Bustos and Yohai, 1979). Several recent proposals attempt to correct this drawback. The family proposed by Mallows (1975) has the form

$$\phi(\mathbf{x}, u) = \pi(\mathbf{x}) \psi(u) \quad (1.6)$$

where  $\pi$  is a positive weight function. A choice for  $\pi$  is:  $\pi(\mathbf{x}) = \psi_k(|\mathbf{x}|) / |\mathbf{x}|$ , where  $|\mathbf{x}|$  is Euclidean norm. Denby and Martin (1979) have used these estimators to estimate autoregressive parameters, calling them “GM-estimates”.

Estimators of the form

$$\phi(\mathbf{x}, u) = \pi(\mathbf{x}) \psi(u / \pi(\mathbf{x})) \quad (1.7)$$

have been studied by Hampel (1977) and Krasker and Welsch (1979), for the particular case  $\psi = \psi_k$  and  $\pi(\mathbf{x}) = 1/|\mathbf{x}|$ . This estimator has important optimality properties within the family (1.2), as shown by these authors and by Maronna, Bustos and Yohai (1979). However, it is worth while to consider the general case (1.7).

As to the Eq. (1.3) for scale, a popular choice is  $\chi(u) = \psi_k(u)^2 - \beta$ , where  $\beta$  is a constant; this corresponds to Huber’s (1964) “Proposal 2”. Another method is Hampel’s (1971) “median deviation” defined by  $\chi(u) = \text{sgn}(u - c)$  where  $c$  is a constant.

As far as we know, all the results on consistency and/or asymptotic normality in the literature are concentrated in the case of classical  $M$ -estimators, with the  $\mathbf{x}_i$ ’s fixed (nonrandom) vectors. See (Relles, 1968), (Huber, 1973),

(Yohai and Maronna, 1979) and (Carroll and Ruppert, 1979). If the  $\mathbf{x}_i$ 's are random, and  $(\mathbf{x}_i, y_i)$  satisfy (1.1), then the consistency of classical  $M$ -estimators with  $\psi$  monotone and  $\sigma$  known is obtained in the Corollary to Theorem 2.1. of (Yohai and Maronna, 1979) assuming that  $P(\mathbf{x}'\boldsymbol{\theta}=0) < 1$  for all  $\boldsymbol{\theta} \neq 0$ . The asymptotic normality follows from Theorem 3.1 of that paper, assuming  $E|\mathbf{x}_i|^2 < \infty$ , by using a conditional argument. Consistency and asymptotic normality for  $\sigma$  known and  $\psi$  not necessarily monotone, may also be proved by applying the results of (Huber, 1967).

However, the former methods of proof do not extend to the situation we consider here, namely: the use of general  $\phi$ -functions, which include cases (1.6) and (1.7), simultaneous estimation of  $\sigma$ , and random  $\mathbf{x}_i$ 's which do not obey model (1.1) exactly.

The contents of the paper are the following. In Sect. 2 we state our general assumptions on  $\phi$  and  $\chi$ , prove the existence of solutions of (1.2)–(1.3) and give conditions for their uniqueness. The interest of the existence theorem is that consistency is easily derived from its method of proof. In Sect. 3 we prove consistency without assuming (1.1). In Sect. 4 asymptotic normality is shown, and in Sect. 5 the former results are extended to the case in which the  $\mathbf{x}$ 's are normalized by using some estimated scale matrix. Section 6 contains the proofs of some technical results used in Sect. 2.

*Notation.* Vectors and matrices will be respectively denoted by lower and upper case boldface letters. If  $\mathbf{M}$  is a matrix, its transpose is written as  $\mathbf{M}'$  and its norm as operator as  $\|\mathbf{M}\|$ ; the euclidean norm of vectors is written  $|\cdot|$ . If  $A$  is a set, its indicator function is denoted by  $I(A)$ , and its complement by  $A^c$ . In general,  $P$  will denote a probability measure on  $R^{p+1}$ , and  $E_P$  the respective expectation functional which will be written as  $E$  when this causes no confusion.

## 2. Existence and Uniqueness of Solutions

Although the existence of solutions is not a very interesting statistical problem in itself, the same method used to prove existence yields consistency, by using the same treatment of regression/scale problems as in (Maronna, 1976).

Let  $P$  be any distribution on  $R^{p+1}$ . We shall consider solutions  $(\boldsymbol{\theta}, \sigma)$  of the system

$$E_P \phi(\mathbf{x}, (y - \mathbf{x}'\boldsymbol{\theta})/\sigma) \mathbf{x} = \mathbf{0}, \tag{2.1}$$

$$E_P \chi(|y - \mathbf{x}'\boldsymbol{\theta}|/\sigma) = 0 \tag{2.2}$$

If  $P$  is the underlying distribution, then (2.1)–(2.2) define the parameters to be estimated; if  $P$  is the empirical distribution corresponding to a sample  $\{(\mathbf{x}_i, y_i), i = 1, \dots, n\}$ , then they are equivalent to (1.2)–(1.3), thus defining the estimators. Note that the parameter space is  $\Theta = R^p \times (0, \infty)$ . The following assumptions will be used throughout the paper.

(A1) For each  $\mathbf{x}$ ,  $\phi(\mathbf{x}, \cdot)$  is odd and uniformly continuous, and  $\phi(\mathbf{x}, u) \geq 0$  for  $u \geq 0$ .

- (A2) The function  $\mu(\mathbf{x}, u) = \phi(\mathbf{x}, u)/u$  is nonincreasing for  $u > 0$ , and there exists  $u_0 > 0$  such that  $\mu(\mathbf{x}, u_0) > 0$  for all  $\mathbf{x}$ .
- (A3)  $\chi$  is nondecreasing, continuous, and bounded. Let  $\chi(0) = -a$ ,  $\chi(+\infty) = b$ , with  $a, b \in (0, \infty)$ .
- (A4)  $\chi$  is strictly increasing in the interval  $\{u: \chi(u) < b\}$ .
- (A5) Let  $K(\mathbf{x}) = \sup_u |\phi(\mathbf{x}, u)|$ . Then  $E_p |\mathbf{x}| K(\mathbf{x}) \leq \infty$ .
- (A6)  $\sup \{P(\mathbf{x}'\boldsymbol{\theta} = 0): \boldsymbol{\theta} \neq \mathbf{0}\} < \chi(u_0)/(a + \chi(u_0))$ , with  $u_0$  defined in (A2).
- (A7)  $\sup \{P(\alpha y + \boldsymbol{\theta}' \mathbf{x} = 0): \alpha \in \mathbb{R}, \boldsymbol{\theta} \in \mathbb{R}^p, |\alpha| + |\boldsymbol{\theta}| \neq 0\} < b/(a + b)$ , with  $a, b$  as in (A3).

We shall briefly discuss the assumptions. For both families (1.6) and (1.7), (A1) is satisfied if  $\pi(\mathbf{x})$  is positive and  $\psi(u)$  is odd, bounded and uniformly continuous, and nonnegative for  $u > 0$ .

The first half of (A2) states that if one writes (2.1) as weighted least squares

$$E \mu(\mathbf{x}, r) r \mathbf{x} = \mathbf{0}, \quad \text{with } r = (y - \mathbf{x}'\boldsymbol{\theta})/\sigma$$

then the weights  $\mu(\mathbf{x}, r)$  should decrease with  $r$ , which is a reasonable requirement. It holds for (1.6) and (1.7) if  $\psi(u)$  is either monotonic and concave for  $u \geq 0$ , or if for some  $u_1 \geq 0$  it is concave on  $[0, u_1]$  and vanishes on  $(u_1, \infty)$  (as in the case of the bisquare). It is also easy to verify this assumption for the Cauchy Maximum Likelihood estimator and other redescending  $\psi$ 's.

The second part of (A2) is needed for technical reasons in the proofs. For (1.6) it holds for any  $\psi$ ; but for (1.7)  $\psi(u)$  must remain positive for  $u > 0$ , since  $u_0$  should not depend on  $\mathbf{x}$ .

The validity of (A5) for (1.6) and (1.7) is ensured, regardless of  $P$ , if the functions  $\psi$  and  $|\mathbf{x}| \pi(\mathbf{x})$  are bounded.

Assumption (A7) is needed to avoid solutions with  $\sigma = 0$ . It implies that the distribution of the "residual"  $y - \mathbf{x}'\boldsymbol{\theta}$  should not be too concentrated in any value. For Huber's Proposal 2, it implies that  $P(y - \mathbf{x}'\boldsymbol{\theta} = r_0) \leq 1 - \beta/k^2$  for any  $r_0$  and any  $\boldsymbol{\theta}$ , which is the assumption used in the hypothesis of the Proposition in p. 97 of (Huber 1964) for the location case.

Assumption (A6) implies that  $\mathbf{x}$  may not be concentrated on any subspace. For the proofs to work, the upper bound must be more stringent than in (A7).

The median deviation violates (A3) and (A1). It might be considered under stronger assumptions on  $P$ .

The main result of this section is the following:

**Theorem 2.1.** *If  $\phi$ ,  $\chi$  and  $P$  satisfy (A1) through (A7), then there exists a solution  $(\boldsymbol{\theta}, \sigma)$  of (2.1)–(2.2).*

The proof follows the same strategy as Theorem 2 of (Maronna 1976). Some lemmas will be first needed, the proofs of which are deferred to Sect. 6 to clarify the exposition. Since they will also be needed to prove consistency, we shall work not only with  $P$  but with all distributions  $Q$  in some "neighbourhood"  $W$  of  $P$ . Let  $\mathcal{W} = \mathcal{W}(P)$  the family of all sets  $W$  of the form

$$W = \{Q: E_Q |f_i| < \infty, |E_P f_i - E_Q f_i| < \varepsilon_i, i = 1, \dots, m\}. \quad (2.3)$$

where  $m$  is any natural number,  $f_1, \dots, f_m$  are any  $P$ -integrable functions on  $R^{p+1}$ , and  $\varepsilon_1, \dots, \varepsilon_m$  any positive numbers. If the  $f_i$ 's are restricted to be bounded and continuous, then the respective  $W$ 's generate the weak topology on probability measures, but no use will be made of this fact.

From now on it will be assumed without further notice that (A1)...(A7) hold.

**Lemma 2.1.** *For each fixed  $\theta \in R^p$  and all  $Q$  in some  $W_1 \in \mathcal{W}$ , there exists a unique solution  $\sigma = \sigma(\theta, Q)$  of the equation  $E_Q \chi(|y - \mathbf{x}'\theta|/\sigma) = 0$ .*

**Lemma 2.2.** *Let  $A_0$  be any positive number. There exist constants  $A_1, A_2, A_3, A_4 \in (0, \infty)$  and  $W_2 \in \mathcal{W}$  such that for all  $Q \in W_2$*

- i)  $\sigma(\theta, Q) \geq A_1$  for all  $\theta \in R^p$ ,
- ii)  $\sigma(\theta, Q)/|\theta| \leq A_2$  if  $|\theta| \geq A_0$ ,
- iii)  $\sigma(\theta, Q) \leq A_3$  if  $|\theta| \leq A_0$ ,
- iv)  $|\theta|/\sigma(\theta, Q) \leq A_4$  for all  $\theta$ .

**Corollary.**  $\sigma(\theta, P)$  is a continuous function of  $\theta$ .

**Lemma 2.3.** *Let  $B(\theta, Q) = : E_Q \phi(\mathbf{x}, (y - \mathbf{x}'\theta)/\sigma(\theta, Q)) \mathbf{x}'\theta/|\theta|$ . There exist  $A_5, A_6 \in (0, \infty)$  and  $W_3 \in \mathcal{W}$  such that*

$$\sup\{B(\theta, Q) : |\theta| \geq A_5, Q \in W_3\} \leq -A_6.$$

We comment briefly on these results. Lemma 1 allows one to work with  $\sigma$  as a function of  $\theta$ , and thus to concentrate on (2.1) only. Then parts (i) and (iii) of Lemma 2.2 imply that  $\sigma$  must remain bounded away from 0, and also away from  $\infty$  if  $\theta$  does. Parts (ii) and (iv) imply that  $\sigma(\theta)$  increases roughly linearly with  $|\theta|$  for large  $|\theta|$ , and thus that  $\sigma(\theta)$  is unbounded if  $\theta$  is so. Thus  $(\theta, \sigma)$  remains in a compact set if and only if  $\theta$  does. Finally Lemma 2.3 implies that there are no solutions outside a certain compact set.

*Proof of Theorem 2.1.* Define the function  $\mathbf{g} : R^p \rightarrow R^p$  as:

$$\mathbf{g}(\theta) = \theta + \gamma(|\theta| + 1) E_P \phi(\mathbf{x}, (y - \mathbf{x}'\theta)/\sigma(\theta, P)) \mathbf{x},$$

where  $\gamma$  is a positive constant to be later conveniently determined. If  $\theta$  is a fixed point of  $\mathbf{g}$  (i.e.  $\mathbf{g}(\theta) = \theta$ ), then  $(\theta, \sigma(\theta, P))$  is a solution of (2.1)–(2.2). By (A1), (A5), and the Corollary to Lemma (2.2),  $\mathbf{g}$  is continuous, so that we may apply Brower's Fixed Point Theorem (Dunford-Schwartz, 1958). That is, to prove the existence of a fixed point of  $\mathbf{g}$ , it suffices to show the existence of a positive constant  $C$  such that  $|\theta| \leq C$  implies  $|\mathbf{g}(\theta)| \leq C$ ; and to this end it suffices to show that, for some  $C'$  and  $\varepsilon > 0$ :  $|\mathbf{g}(\theta)|/|\theta| \leq 1 - \varepsilon$  for all  $|\theta| \geq C'$ . To obtain this, note that by the Cauchy-Schwartz inequality:

$$\begin{aligned} |\mathbf{g}(\theta)|^2/|\theta|^2 &\leq 1 + \gamma^2(1 + |\theta|^{-1})^2(EK(\mathbf{x})|\mathbf{x}|)^2 \\ &\quad + 2(1 + |\theta|^{-1})\gamma B(\theta, P), \end{aligned} \tag{2.6}$$

where  $B(\theta, P)$  is defined in Lemma (2.3). This Lemma implies that  $B(\theta, P) \leq -A_6$  if  $|\theta| \geq A_5$ , and thus by taking  $\gamma$  small enough, one can make the right side of (2.6) strictly less than one for  $|\theta| \geq A_5$ .

We have thus proved the existence of a solution, say  $(\theta_0, \sigma_0)$ . It seems difficult to prove uniqueness without imposing much stronger assumptions.

**Theorem 2.2.** *Let  $G_{\mathbf{x}}(r)$  be the conditional distribution of  $r = y - \mathbf{x}'\theta_0$  given  $\mathbf{x}$ . If  $G_{\mathbf{x}}$  is symmetric for all  $\mathbf{x}$ , then each of the assumptions (a) and (b) is sufficient for the uniqueness of  $(\theta_0, \sigma_0)$ .*

a) *For each  $\mathbf{x}$ ,  $\phi(\mathbf{x}, u)$  is nondecreasing and is strictly increasing for  $|u| \leq u_0$ , where  $u_0$  satisfies A6.*

b)  *$G_{\mathbf{x}}$  has a density  $g_{\mathbf{x}}(r)$  which is a decreasing function of  $|r|$ .*

*Proof.* Without loss of generality, take  $(\theta_0, \sigma_0) = (\mathbf{0}, 1)$ . Let  $(\theta, \sigma)$  be another solution, with  $\sigma = \sigma(\theta, P)$ . The assumed symmetry and (A1) imply that  $E\phi(\mathbf{x}, y/\sigma) \mathbf{x} = \mathbf{0}$ , and hence

$$E(\phi(\mathbf{x}, (y - \mathbf{x}'\theta)/\sigma) - \phi(\mathbf{x}, y/\sigma)) \mathbf{x}'\theta = \mathbf{0}.$$

If  $\theta \neq \mathbf{0}$ , then the assumed monotonicity implies:

$$1 \leq P(\mathbf{x}'\theta = 0) + P(|y - \mathbf{x}'\theta|/\sigma \geq u_0).$$

Since  $\sigma = \sigma(\theta, P)$ , it is not difficult to show, using (2.2), that the last term is  $\leq 1 - \chi(u_0)/(a + \chi(u_0))$ . This and (A6) complete the proof of (a).

The proof of (b) is similar to that of Theorem 3(i) in (Maronna, 1976). The details are left to the reader.

*Remark.* Note that even in the absence of uniqueness, all solutions must remain in a compact set, since Lemma 2.3 implies that, if  $(\theta, \sigma)$  is a solution, then  $B(\theta, P) = 0$ , and hence  $|\theta| \leq A_5$ . More generally, if  $(\theta_n, \sigma_n)$  is a sequence such that (2.2) holds and the left-hand side of (2.1) tends to zero, then  $|\theta_n| \leq A_5$  for large  $n$ . This may not hold for estimators with *fixed* scale - i.e. when  $\theta$  is defined as a solution of (2.1) alone, for a fixed  $\sigma$  - if  $\phi(x, \cdot)$  is not monotonic. The simplest case is location with a redescending  $\psi$ : here  $E\psi((y - \theta)/\sigma) \rightarrow 0$  for  $\theta \rightarrow \infty$ , and may even become  $= 0$  if  $\psi = 0$  outside some interval and  $P$  has compact support (as in the case of the bisquare when  $P$  is an empirical distribution).

### 3. Consistency

Since in robustness problems one avoids making "exact" statements about distributions, consistency will be proved for general  $P$ , without assuming (1.1).

**Theorem 3.1.** *Let  $P$  be a distribution satisfying (A5), (A6) and (A7) and such that (2.1)-(2.2) has a unique solution  $(\theta_0, \sigma_0)$ . Let  $(\mathbf{x}_i, y_i)$ ,  $i = 1, 2, \dots$ , be independent with distribution  $P$ , and let  $(\theta_n^*, \sigma_n^*)$  be any sequence such that*

$$\lim_{n \rightarrow \infty} n^{-1} \sum_{i=1}^n \phi(\mathbf{x}_i, (y_i - \mathbf{x}_i' \theta_n^*)/\sigma_n^*) \mathbf{x}_i = \mathbf{0} \quad \text{a.s.}, \quad (3.1)$$

$$\sum_{i=1}^n \chi(|y_i - \mathbf{x}_i' \theta_n^*|/\sigma_n^*) = \mathbf{0}. \quad (3.2)$$

Then  $\lim_{n \rightarrow \infty} (\theta_n^*, \sigma_n^*) = (\theta_0, \sigma_0)$  a.s.

*Proof.* It may be assumed without loss of generality that  $(\theta_0, \sigma_0) = (\mathbf{0}, 1)$ . By a classical argument (Huber, 1967, part B), it suffices to show the existence of a compact set  $C \subseteq \Theta$  such that the sequence  $(\theta_n^*, \sigma_n^*) \in C$  ultimately a.s.

Let  $P_n$  be the empirical distribution corresponding to the sample  $\{(\mathbf{x}_i, y_i), i = 1, \dots, n\}$ , so that  $\sigma_n^* = \sigma(\theta_n^*, P_n)$ . The Strong Law of Large Numbers implies that a.s. for large  $n$ ,  $P_n$  remains in the intersection of  $W_2$  and  $W_3$  defined in Lemmas (2.2) and (2.3). The latter implies that a.s.

$$\limsup_{n \rightarrow \infty} \sup \{B(\theta, P_n) : |\theta| \geq A_5\} < -A_6,$$

and (3.1) implies that  $\lim B(\theta_n^*, P_n) = 0$ ; so that  $\limsup |\theta_n^*|$  must be  $\leq A_5$  a.s. Parts (ii) and (iii) of Lemma (2.2) then imply that  $\limsup \sigma_n^* \leq \max(A_2, A_5, A_3)$ , and part (i) yields  $\liminf \sigma_n^* \geq A_1$  a.s. This completes the proof.

*Remark.* Theorem (3.1) is still true when the equality in (3.2) is replaced by a.s. equality in the limit when  $n \rightarrow \infty$ . The proof is based on a reformulation of the proofs of Lemmas (2.2) and (2.3), but the details require much more care.

#### 4. Asymptotic Normality

The asymptotic distribution of the estimators is easily derived by applying the results of Huber (1967). To simplify notation, define as  $\xi$  the  $(p+1)$ -dimensional vector with coordinates  $(\theta, \sigma)$ , and as  $\Psi(\mathbf{x}, y, \xi)$  the  $(p+1)$ -dimensional vector with coordinates  $\mathbf{x}\phi(\mathbf{x}, (y - \mathbf{x}'\theta)/\sigma)$  and  $\chi(|y - \mathbf{x}'\theta|/\sigma)$ . Let  $\{(\mathbf{x}_i, y_i)\}$  be a sample from the distribution  $P$ , and  $\xi_n^*$  be any sequence such that

$$p \lim_{n \rightarrow \infty} n^{-1/2} \sum_{i=1}^n \Psi(\mathbf{x}_i, y_i, \xi_n^*) = \mathbf{0}. \tag{4.1}$$

Assume that  $\xi_n^*$  is consistent; let  $\xi_0 = \lim \xi_n^*$ . Assume that  $\chi(\cdot)$  and  $\phi(\mathbf{x}, \cdot)$  are absolutely continuous. Let  $\mathbf{J}(\mathbf{x}, y, \xi)$  be the  $(p+1) \times (p+1)$ -matrix of derivatives of  $\Psi$  with respect to  $\theta$  and  $\sigma$ . Let  $\phi'(\mathbf{x}, u) = \partial \phi(\mathbf{x}, u) / \partial u$  and

$$\mathbf{C} = E_p \Psi(\mathbf{x}, y, \xi_0) \Psi(\mathbf{x}, y, \xi_0)' \quad \text{and} \quad \mathbf{D} = E_p \mathbf{J}(\mathbf{x}, y, \xi_0). \tag{4.2}$$

Let  $L(\mathbf{x}) = K(\mathbf{x})|\mathbf{x}|$ ,  $H_1(\mathbf{x}) = \sup_u |\phi'(\mathbf{x}, u)|$  and  $H_2(\mathbf{x}) = \sup_u |u\phi'(\mathbf{x}, u)|$

**Theorem 4.1.** Assume that (besides (A1) through (A7))  $\phi, \chi$  and  $P$  satisfy the following assumptions:

- (C1) The functions  $\chi'(t)$  and  $t\chi'(t)$  are bounded, and  $\chi'(0) = 0$ .
- (C2)  $EH_1(\mathbf{x})|\mathbf{x}|^2 L(\mathbf{x}) < \infty$  and  $EH_2(\mathbf{x})|\mathbf{x}| L(\mathbf{x}) < \infty$ .
- (C3)  $EH_1(\mathbf{x})|\mathbf{x}|^2 < \infty$  and  $EH_2(\mathbf{x})|\mathbf{x}| < \infty$ .
- (C4)  $E|\mathbf{x}| < \infty$ .
- (C5)  $\mathbf{D}$  is nonsingular.
- (C6)  $EL(\mathbf{x})^2 < \infty$ .

Then  $n^{1/2}(\xi_n^* - \xi_0)$  tends in law to a  $(p+1)$ -variate normal with covariance matrix  $\mathbf{V} = \mathbf{D}^{-1}\mathbf{C}(\mathbf{D}^{-1})$ .

*Proof.* Let  $\lambda(\xi) = E\Psi(\mathbf{x}, y, \xi)$ , and

$$U(\mathbf{x}, y, \xi, \delta) = \sup \{ |\Psi(\mathbf{x}, y, \xi) - \Psi(\mathbf{x}, y, \eta)| : |\eta - \xi| \leq \delta \}.$$

Then, according to (Huber, 1967, p. 227), it suffices to show that for certain constants  $c_1, c_2, c_3$  and  $\delta_0$ , and for  $|\xi - \xi_0| + \delta \leq \delta_0$ , we have: (i)  $|\lambda(\xi)| \geq c_1|\xi - \xi_0|$ ; (ii)  $EU(\mathbf{x}, y, \xi, \delta) \leq c_2\delta$ ; and (iii)  $EU^2(\mathbf{x}, y, \xi, \delta) \leq c_3\delta$ .

Part (i) stems from (C5). By (C3) one may interchange expectation and differentiation, hence (ii) and (iii) follow from (C1) through (C4) (which ensure the integrability of  $\mathbf{J}$  for all  $\xi$ ) and the Mean Value Theorem. Finally the existence of  $\mathbf{C}$  in (4.2) follows from (C6).

A more tractable expression for  $\mathbf{V}$  is obtainable only under further restrictions. Assume that  $P$  satisfies model (1.1). Without loss of generality take  $\xi_0 = (\mathbf{0}, 1)$ . Let  $\mathbf{M} = E\phi'(\mathbf{x}, y)\mathbf{x}\mathbf{x}'$  and  $\mathbf{N} = E\phi(\mathbf{x}, y)^2\mathbf{x}\mathbf{x}'$ . Assume that  $y$  has a symmetric distribution and  $\mathbf{M}$  is nonsingular. Then it follows from Theorem (4.1) that  $\theta^*$  and  $\sigma^*$  are asymptotically independent, and  $\theta^*$  has asymptotic covariance matrix  $\mathbf{V}_\theta = \mathbf{M}^{-1}\mathbf{N}\mathbf{M}^{-1}$ .

*Remark 4.1.* The inclusion of  $L(\mathbf{x})$  in (C2) is due to (iii). For the “qualitatively robust” estimates which are mainly of interest,  $L(\mathbf{x})$  is bounded and hence its inclusion is not restrictive. However, for classical  $M$ -estimators, (C2) implies that  $E|\mathbf{x}|^3 < \infty$ , which seems too strong a requirement. The difficulty is that our theorem depends heavily on Huber’s result, which is not adequate for regression problems. For example, in the case of the least squares estimator with fixed scale, assumption (iii) implies  $E|\mathbf{x}|^4 < \infty$ . It would be useful to modify Huber’s approach to treat regression problems more adequately.

Assume that the functions  $|\mathbf{x}|\pi(\mathbf{x})$ ,  $\psi(u)$ ,  $\psi'(u)$  and  $u\psi'(u)$  are bounded. Then estimators (1.7) satisfy (C2) and (C3) if  $E|\mathbf{x}|^2 < \infty$ , while for estimators (1.6), (C2) is implied by (C1).

In general, asymptotic normality is not true without some restrictions about the tails of  $|\mathbf{x}|$ , of which (C4) seems to be the lightest one. For estimators (1.7) with monotonic  $\psi$  under model (1.1), it is easily seen that if  $P(u=0) > 0$ , then the existence of the matrix  $\mathbf{M}$  defined above Remark 4.1 - and hence asymptotic normality - implies  $E|\mathbf{x}|^2 < \infty$ .

In effect:

$$\text{trace}(\mathbf{M}) = E\psi'(u/\pi(\mathbf{x}))|\mathbf{x}|^2 \geq \psi'(0)P(u=0)E|\mathbf{x}|^2. \quad (4.3)$$

This fact points out that the asymptotic normality of estimators (1.7) is less robust than that of estimators (1.6). The practical interest of asymptotic normality lies in the possibility of constructing approximate confidence intervals for  $\theta$  by using the normal approximation to the distribution of  $\theta^*$ , and estimating the matrices  $\mathbf{C}$  and  $\mathbf{D}$  - or  $\mathbf{M}$  and  $\mathbf{N}$  - from the sample. In the case (1.7) one would estimate  $\mathbf{M}$  by

$$\mathbf{M}^* = n^{-1} \sum_i \psi'(r_i/\pi(\mathbf{x}_i))\mathbf{x}_i\mathbf{x}_i',$$



where  $r_i = (y_i - \mathbf{x}'_i \boldsymbol{\theta}^*) / \sigma^*$ . Thus a small residual at a large  $\mathbf{x}_i$  will unduly inflate  $\mathbf{M}^*$ , and hence spoil the confidence intervals. Exploratory simulations performed by S. Morgenthaler indicate that in effect the confidence intervals for (1.7) are in general far from being “approximately distribution free”, while those for (1.6) are much more reliable.

### 5. Scaling the $\mathbf{x}$ 's

It is easy to see that classical  $M$ -estimators are equivariant under nonsingular linear transformations of the  $\mathbf{x}$ 's, but this is in general not true for the estimators (1.2)–(1.3). To obtain estimators invariant under some desired group of linear transformations, it is natural to replace (1.2) by

$$\sum_{i=1}^n \phi(\mathbf{S}_n^* \mathbf{x}_i, (y_i - \mathbf{x}'_i \boldsymbol{\theta}_n^*) / \sigma_n^*) \mathbf{x}_i = \mathbf{0}, \tag{5.1}$$

where  $\mathbf{S}_n^*$  is a  $p \times p$ -matrix which depends on the data. The matrix  $\mathbf{S}_n^*$  may be obtained from a “robust covariance” as in (Huber, 1977) or (Maronna, 1976). Krasker and Welsch (1979) study some matrices possessing certain optimal properties for regression. The former results will be extended to this situation by showing that if  $\mathbf{S}_n^*$  tends to a matrix  $\mathbf{S}$  when  $n \rightarrow \infty$ , then the asymptotic behavior of  $\boldsymbol{\theta}_n^*$  is the same as if  $\mathbf{S}_n^*$  were replaced by  $\mathbf{S}$  in (5.1).

Assume that  $E_P |\mathbf{x}| K(\mathbf{x}) < \infty$ , and that  $(\boldsymbol{\theta}_0, \sigma_0)$  is the only solution of

$$E_P \phi(\mathbf{S}\mathbf{x}, (y - \mathbf{x}'\boldsymbol{\theta}) / \sigma) \mathbf{x} = \mathbf{0}, \quad E_P \chi(|y - \mathbf{x}'\boldsymbol{\theta}| / \sigma) = 0.$$

Let  $(\boldsymbol{\theta}_n^*, \sigma_n^*)$  be a solution of (5.1)–(1.3). Then:

**Lemma 5.1.** *Let  $\mathbf{S}_n^* \rightarrow \mathbf{S}$  a.s., where  $\mathbf{S}_n^*$  and  $\mathbf{S}$  are nonsingular. Assume that for a constant  $c$  and all  $\mathbf{x}, \mathbf{z}$  and  $u$ :*

$$|\phi(\mathbf{x}, u) - \phi(\mathbf{z}, u)| \leq |\mathbf{z} - \mathbf{x}| c / \min(|\mathbf{x}|, |\mathbf{z}|)^2. \tag{5.2}$$

*Then  $(\boldsymbol{\theta}_n^*, \sigma_n^*) \rightarrow (\boldsymbol{\theta}_0, \sigma_0)$  a.s.*

*Proof.* According to Theorem (3.1), it suffices to show that

$$(1/n) \sum_{i=1}^n \phi(\mathbf{S}\mathbf{x}_i, (y_i - \mathbf{x}'_i \boldsymbol{\theta}_n^*) / \sigma_n^*) \mathbf{x}_i \rightarrow \mathbf{0} \text{ a.s.}$$

Define for  $\boldsymbol{\theta} \in R^p$ :

$$T_n(\boldsymbol{\theta}) = (1/n) \sum_{i=1}^n (\phi(\mathbf{S}_n^* \mathbf{x}_i, (y_i - \mathbf{x}'_i \boldsymbol{\theta}) / \sigma_n^*) - \phi(\mathbf{S}\mathbf{x}_i, (y_i - \mathbf{x}'_i \boldsymbol{\theta}) / \sigma_n^*)) \mathbf{x}_i \tag{5.3}$$

Then it suffices to show that  $T_n^*(\boldsymbol{\theta}_n^*) \rightarrow 0$  a.s. Since  $|\mathbf{S}\mathbf{x}| \geq |\mathbf{x}| \|\mathbf{S}^{-1}\|$ , (5.2) yields:

$$|T_n(\boldsymbol{\theta}_n^*)| \leq \|\mathbf{S}_n^* - \mathbf{S}\| \max(\|\mathbf{S}^{-1}\|, \|\mathbf{S}_n^{*-1}\|)^2 c \rightarrow 0 \text{ a.s.} \tag{5.4}$$

*Remark.* For Mallows estimators, (5.2) is satisfied if  $\psi$  is bounded and  $\pi$  is of the form

$$\pi(\mathbf{x}) = h(|\mathbf{x}|), \quad \text{with } |h'(t)| \leq c/t^2 \quad \text{for } t > 0. \quad (5.5)$$

This holds in particular for  $h(t) = \min(1, c/t)$ . For estimators (1.7), it suffices that  $\pi$  satisfies (5.5), and that  $\psi$  be of the form  $\psi(t) = t h(t)$ , with  $h$  as in (5.5); in particular for the Huber psi-function.

The treatment of asymptotic normality is more complicated. To simplify the problem, assume that  $\sigma$  is known and equal to one. Without loss of generality take  $\theta_0 = \mathbf{0}$ . Let  $\theta_n^*$  be a solution of (5.1) (with  $\sigma_n^* = 1$ ) and assume that  $p \lim \theta_n^* = \mathbf{0}$ . In  $R^{p+1}$  define the norm:

$$\|(\mathbf{x}, y)\| = \max(|\mathbf{x}|, |y|).$$

If  $A$  is a set in  $R^{p+1}$ , let for  $\delta > 0$ :

$$A(\delta) = \{(\mathbf{x}, y) : \|(\mathbf{x}, y) - (\mathbf{x}_0, y_0)\| < \delta \quad \text{for some } (\mathbf{x}_0, y_0) \in A\}.$$

**Lemma 5.2.** *Assume that the assumptions of Lemma (5.1) and (C6) of Theorem (4.1) are satisfied, that  $\mathbf{S}_n^*$  depends only on  $\mathbf{x}_1, \dots, \mathbf{x}_n$ , and  $\|\mathbf{S}_n^* - \mathbf{S}\| n^{1/2} = O_p(1)$ .*

*Assume also that the conditional distribution of  $y$  given  $\mathbf{x}$  is symmetric, and that for each  $\varepsilon > 0$ , there exist a set  $A_\varepsilon \subseteq R^{p+1}$ , and  $\delta = \delta(\varepsilon) > 0$ , such that  $P((\mathbf{S}\mathbf{x}, y) \in A) > 1 - \varepsilon$ , and the derivative  $d(\mathbf{x}, y) = \partial\phi(\mathbf{x}, y)/\partial\mathbf{x}\partial y$  exists and is bounded in every bounded set contained in  $A_{\varepsilon(\delta)}$ .*

*Then  $p \lim n^{1/2} T_n^*(\theta_n) = \mathbf{0}$ , where  $T_n^*$  is defined as in (5.3) but with  $\sigma_n^* = 1$ .*

*Proof.* Put  $n^{1/2} T_n(\theta_n^*) = \mathbf{r}_n + \mathbf{s}_n$ , where  $\mathbf{r}_n = n^{1/2} T_n(\mathbf{0})$ . First we prove that  $|\mathbf{r}_n| = o_p(1)$ . Call  $E_X$  and  $\text{cov}_V$  the conditional expectation and covariance matrix with respect to  $\mathbf{X} = \{\mathbf{x}_1, \dots, \mathbf{x}_n\}$ . Then a symmetry argument yields  $E_X \mathbf{r}_n = \mathbf{0}$ .

Since  $y_i$  and  $y_j$  are conditionally independent with respect to  $\mathbf{X}$  for  $i \neq j$ , this implies that, by (A1):

$$E_X \mathbf{r}_n \mathbf{r}_n' = (1/n) \sum_{i=1}^n E_X [\phi(\mathbf{S}_n^* \mathbf{x}_i, y_i) - \phi(\mathbf{S}\mathbf{x}_i, y_i)]^2 \mathbf{x}_i \mathbf{x}_i' \quad (5.5)$$

Reasoning as in (5.4), and taking the trace in (5.5), one obtains  $E_X |\mathbf{r}_n|^2 \leq \|\mathbf{S}_n^* - \mathbf{S}\|^2 O_p(1)$ , and this yields  $p \lim |\mathbf{r}_n| = 0$ .

Now we shall show that for any  $\varepsilon > 0$ ,  $|\mathbf{s}_n| \leq \varepsilon O_p(1)$  for large  $n$ . Let  $C$  be the set  $\{(\mathbf{x}, y) : (\mathbf{S}\mathbf{x}, y) \in A_\varepsilon, |\mathbf{x}| \leq k, |y| \leq k\}$  where  $k$  is chosen large enough that  $P(C) > 1 - \varepsilon$ .

Decompose  $\mathbf{s}_n = \mathbf{s}_n' + \mathbf{s}_n''$ , where  $\mathbf{s}_n'$  contains those terms with  $(\mathbf{x}_i, y_i) \in C$ . Then by bounding the terms in  $\mathbf{s}_n'$  using the Mean Value Theorem applied to  $d(\mathbf{x}, y)$ , one shows that  $\mathbf{s}_n' \rightarrow \mathbf{0}$ ; and applying (5.2):

$$|\mathbf{s}_n''| \leq n^{1/2} \|\mathbf{S}_n^* - \mathbf{S}\| O_p(1) \left[ n^{-1} \sum_{i=1}^n I((\mathbf{x}_i, y_i) \notin C) \right]$$

and the last expression tends to  $P(C^c) < \varepsilon$ . The details are left to the reader.

*Remark.* Let  $P$  be continuous. To apply this Lemma to Mallows estimators, it suffices that  $\pi$  be a piecewise differentiable function of  $|\mathbf{x}|$ , and that  $\psi$  be piecewise differentiable. For estimators (1.7) it suffices that  $\psi$  be twice piecewise differentiable.

## 6. Appendix

In this section we prove the Lemmas of Sect. 2. We begin with an auxiliary result.

**Lemma 6.1.** *Let  $P$  be any distribution on  $R^{p+1}$ . Let  $K$  be a compact set in  $R^m$ , and  $k(z)$  a  $P$ -integrable function on  $R^{p+1}$ . Suppose that  $g(\mathbf{z}, \tau)$  (with  $\mathbf{z} \in R^{p+1}$  and  $\tau \in K$ ) is continuous in  $\tau$  for each  $\mathbf{z}$ ,  $E_P |g(\mathbf{z}, \tau)| < \infty$  for  $\tau \in K$ , and  $g(\mathbf{z}, \tau) \geq k(\mathbf{z})$  for all  $\mathbf{z}$  and  $\tau \in K$ . Then, given  $\varepsilon > 0$ , there exists  $W \in \mathcal{W}$  (defined in (2.3)) such that for all  $Q \in W$ :*

$$\inf_{\tau \in K} E_Q g(\mathbf{z}, \tau) \geq \inf_{\tau \in K} E_P g(\mathbf{z}, \tau) - \varepsilon.$$

**Corollary 1.** *If  $h(\mathbf{z}, \tau)$  is continuous in  $\tau$  for each  $\mathbf{z}$ , and measurable in  $\mathbf{z}$  for each  $\tau \in K$ , then given  $\varepsilon > 0$ , there exist  $\delta > 0$  and  $W \in \mathcal{W}$  such that for all  $Q \in W$ :*

$$\sup_{\tau \in K} Q(\mathbf{z}: h(\mathbf{z}, \tau) \leq \delta) \leq \sup_{\tau \in K} P(\mathbf{z}: h(\mathbf{z}, \tau) \leq 0) + \varepsilon.$$

**Corollary 2.** *If  $P$  satisfies (A6) and (A7), then there exist positive  $c, d$  and  $e$ , and  $W_1 \in \mathcal{W}$ , such that for all  $Q \in W_1$*

$$\sup \{Q(|\mathbf{x}'\boldsymbol{\theta}| \leq e) : |\boldsymbol{\theta}| = 1\} \leq \chi(u_0)/(a + \chi(u_0)) - c, \quad (6.1)$$

$$\sup \{Q(|y - \mathbf{x}'\boldsymbol{\theta}| \leq d) : \boldsymbol{\theta} \in R^p\} \leq b/(a + b) - c, \quad (6.2)$$

where  $u_0$  is defined in (A2), and  $a$  and  $b$  in (A3).

The proof of Lemma (6.1) follows along the same lines as that of Lemma 1 in (Yohai, 1974). The proofs of the corollaries are left to the reader.

*Proof of Lemma 2.1.* For brevity put  $q_Q(\sigma, \boldsymbol{\theta}) = E_Q \chi(|y - \mathbf{x}'\boldsymbol{\theta}|/\sigma)$ , so that by definition

$$q_Q(\sigma(\boldsymbol{\theta}, Q), \boldsymbol{\theta}) = 0. \quad (6.3)$$

Let  $Q$  belong to the neighbourhood  $W_1$  in Corollary 2 above. Then (6.2) implies that  $\lim_{\sigma \rightarrow \infty} q_Q(\sigma, \boldsymbol{\theta}) < 0 < \lim_{\sigma \rightarrow 0} q_Q(\sigma, \boldsymbol{\theta})$ . Since  $q_Q(\cdot, \boldsymbol{\theta})$  is continuous by (A3), the Intermediate Value Theorem yields the existence of a solution. Uniqueness is proved by using (A4), (6.2) and the monotonicity of  $\chi$ .

*Proof of Lemma 2.2.* (i) Choose  $A_1$  such that  $\chi(d/A_1) = b - (a + b)c/2$ , where  $c$  and  $d$  are defined in (6.2). Partition the expectation in (6.3) into the sets:  $|y - \mathbf{x}'\boldsymbol{\theta}| \leq d$  and  $> d$ . Then by the monotonicity of  $\chi$ ,  $\sigma < A_1$  implies  $q_Q(\sigma, \boldsymbol{\theta}) > 0$  if  $Q \in W_1$  (with  $W_1$  as in the former Lemma).

(ii) Take  $A_2$  large enough that  $E_P \chi((|y|/A_0 + |\mathbf{x}|)/A_2) < -a/2$ . Take  $W' \in \mathcal{W}$  such that  $E_Q((|y|/A_0 + |\mathbf{x}|)/A_2) < -a/4$  for all  $Q \in W'$ . If  $|\boldsymbol{\theta}| \geq A_0$  and  $Q \in W' \cap W_1$ ,

and if  $\sigma(\theta, Q)/|\theta| > A_2$ , then (6.3) would yield

$$0 \leq E_Q \chi(|y|/\sigma(\theta, Q) + |\mathbf{x}| |\theta|/\sigma(\theta, Q)) < -a/4.$$

(iii) Take  $A_3$  such that  $E_P \chi((|y| + |\mathbf{x}| A_0)/A_3) < -a/2$ . The rest of the proof is like in (ii).

(iv) Recall the definitions of  $c$  and  $e$  in (6.2) and of  $A_1$  and  $W_1$  in (i).

Take  $k$  such that  $\chi(k) > b - \delta$ , where  $\delta > 0$  is to be determined later. Define the event  $U = \{e A_4 - |y|/A_1 > k\}$ . Then one may find  $A_4$  and  $W' \in \mathcal{W}$  such that  $Q(U) \geq 1 - c/2$  for all  $Q \in W'$ . It will be shown that if  $Q \in W' \cap W_1$  then  $|\theta|/\sigma(\theta, Q) > A_4$  contradicts (6.3). In effect, let  $V = \{\mathbf{x}: |\mathbf{x}' \theta|/|\theta| \geq e\}$ , so that  $Q(V) \geq a/(a+b) + c$  by (6.1). Partition the expectation in (6.3) into the sets  $U \cap V$ ,  $U \cap V^c$ , and  $U^c$ . Then recalling (i), (6.3) yields:  $0 \geq (\chi(k) + a)(Q(V) - c/2) - a$ , and this can be made positive by choosing  $\delta$  small enough.

Finally to prove the Corollary, note that  $q_p(\sigma, \theta)$  is continuous on  $\Theta$ , and that (i) and (iii) imply that  $\sigma(\theta, P)$  is bounded if  $|\theta|$  remains bounded. A simple argument shows then that  $\theta_n \rightarrow \theta$  implies  $\sigma(\theta_n, P) \rightarrow \sigma(\theta, P)$  for any sequence  $\{\theta_n\}$ .

Now to prove Lemma (2.3) an auxiliary result will be needed. Proceeding as in the proof of Lemma (2.1) it is easy to show that for each  $\tau \in R^p$ , there exists a unique solution  $C = C(\tau)$  of the equation  $E_P \chi(|\mathbf{x}' \tau| | C) = 0$ . Then:

**Lemma 6.2.** *For each  $\varepsilon > 0$ , there exist  $A^\varepsilon$  and  $W^\varepsilon \in \mathcal{W}$  such that for all  $Q \in W^\varepsilon$ ,  $|\theta| \geq A^\varepsilon$  implies*

$$-\varepsilon \leq |\theta|/\sigma(\theta, P) - C(\theta/|\theta|) \leq \varepsilon.$$

*Proof.* We shall find  $A^\varepsilon$  and  $W^\varepsilon$  to ensure  $|\theta|/\sigma(\theta, P) \leq C(\theta^0) + \varepsilon$ , where  $\theta^0 = \theta/|\theta|$ ; the other inequality is dealt with in the same way. Put  $S = \{\tau \in R^p: |\tau| = 1\}$ . A simplified version of the proof of the Corollary to Lemma (2.2) shows that  $C(\tau)$  is continuous for  $\tau \in S$ . Let  $\beta = \inf_{\tau \in S} E_P \chi(|\mathbf{x}' \tau| (C(\tau) + \varepsilon))$ . Then the continuity and uniqueness of  $C(\tau)$ , and the compactness of  $S$ , yield  $\beta > 0$ . By Lemma (6.1) there exists  $W' \in \mathcal{W}$  such that for  $Q \in W'$

$$\inf_{\tau \in S} E_Q \chi(|\mathbf{x}' \tau| (C(\tau) + \varepsilon)) > \beta/2. \quad (6.4)$$

Now note that Lemma (2.2) (iv) implies that, given any constant  $A$ ,  $1/\sigma(\theta, Q) \leq A_4/A$  for  $|\theta| > A$  and  $Q \in W_2$ . Since  $\chi$  is continuous, monotone and bounded by (A3), it is uniformly continuous, and it is easy to show the existence of  $A^\varepsilon$  and  $W'' \in \mathcal{W}$  such that for  $Q \in W''$  and  $|\theta| > A^\varepsilon$ :

$$|E_Q \chi(|y - \mathbf{x}' \theta|/\sigma(\theta, Q)) - E_Q \chi(|\mathbf{x}' \theta|/\sigma(\theta, Q))| < \beta/4. \quad (6.5)$$

Suppose that  $Q \in W^\varepsilon = W' \cap W'' \cap W_2$ ,  $|\theta| \geq A^\varepsilon$  and  $|\theta|/\sigma(\theta, Q) > C(\theta^0) + \varepsilon$ . Then (6.4) and (6.5) and the monotonicity of  $\chi$  would yield:  $E_Q \chi(|y - \mathbf{x}' \theta|/\sigma(\theta, Q)) \geq \beta/4$ , contradicting (6.3).

*Proof of Lemma 2.3.* Let for  $\varepsilon > 0$ :

$$\alpha_\varepsilon = \inf_{\tau \in S} E_P \phi(\mathbf{x}, |\mathbf{x}' \tau| (C(\tau) + \varepsilon)) |\mathbf{x}' \tau|;$$

where  $S = \{\tau: |\tau| = 1\}$ .

Note that the integrand is nonnegative by (A1). We shall first show the existence of  $\varepsilon > 0$  such that  $\alpha_\varepsilon > 0$ . Put for brevity:  $r(t) = \chi(t)/(a + \chi(t))$  (with  $a$  as in (A3)). From the definition of  $C(\tau)$  one derives

$$P(|\mathbf{x}' \tau | C(\tau) \leq t) \geq r(t) \quad \text{for all } t > 0. \quad (6.6)$$

Then by the continuity of  $C(\tau)$ , the compactness of  $S$ , and (A2), if  $\alpha_\varepsilon = 0$ , then for some  $\tau$  one must have:  $1 = P(\mathbf{x}' \tau = 0) + P(|\mathbf{x}' \tau | (C(\tau) + \varepsilon) > u_0)$ . Thus (6.1) and (6.6) imply

$$1 \leq r(u_0) - c + 1 - r(u_0 C(\tau)/(C(\tau) + \varepsilon)). \quad (6.7)$$

Now note that  $r(t)$  is uniformly continuous in a neighborhood of  $u_0$ , and that Lemma (6.2) together with Lemma (2.2) (ii) and (iv), imply

$$1/A_2 \leq C(\tau) \leq A_4 \quad \text{for } \tau \in S. \quad (6.8)$$

Hence there exists  $\varepsilon > 0$  such that  $r(u_0) - r(u_0 C(\tau)/(C(\tau) + \varepsilon)) < c/2$  for all  $\tau \in S$ .

The preceding reasoning implies that this  $\varepsilon$  makes  $\alpha_\varepsilon > 0$ .

Put  $\alpha = \alpha_\varepsilon$ . Lemma (6.1) yields the existence of  $W' \in \mathcal{W}$  such that for  $Q \in W'$ :

$$\inf_{\tau \in S} E_Q \phi(\mathbf{x}, |\mathbf{x}' \tau | (C(\tau) + \varepsilon)) |\mathbf{x}' \tau | > \alpha/2. \quad (6.9)$$

Now let  $J(\mathbf{x}, t) = \sup \{|\phi(\mathbf{x}, u) - \phi(\mathbf{x}, u')| : |u - u'| \leq t\}$ . Then (A1) implies that  $J(\mathbf{x}, t) \rightarrow 0$  when  $t \rightarrow 0$ , and hence by (A5), there exists  $k$  such that  $E_P J(\mathbf{x}, |y| A_4/k) |\mathbf{x}| < A_6/4$ , where  $A_6 = \alpha/(4(1 + \varepsilon A_2))$ , with  $A_4$  and  $A_2$  as in Lemma (2.2). Then there exists  $W'' \in \mathcal{W}$  such that  $E_Q J(\mathbf{x}, |y| A_4/k) |\mathbf{x}| < A_6/2$  for  $Q \in W''$ : and this implies that for  $|\theta| \geq k$  and  $Q \in W''$ :

$$|B(\theta, Q) + B_1| \leq A_6/2, \quad (6.10)$$

where  $B_1 = E_Q \phi(\mathbf{x}, |\mathbf{x}' \theta|/\sigma(\theta, Q)) |\mathbf{x}' \theta^0|$ .

Finally let  $A_5 = \max(k, A^\varepsilon, A_0)$ , and  $W_3 = W' \cap W'' \cap W_2 \cap W^\varepsilon$ ; where  $k$ ,  $W'$  and  $W''$  are defined above,  $A^\varepsilon$  and  $W^\varepsilon$  appear in Lemma (6.2) and  $A_0$  and  $W_2$  in Lemma (2.2). By Lemma (6.2),  $|\theta| > A_5$  and  $Q \in W_3$  imply  $|\theta|/\sigma(\theta, Q) \leq C(\theta^0) + \varepsilon$ ; and (A2) implies  $\phi(\mathbf{x}, s) \geq \mu(\mathbf{x}, t) s = \phi(\mathbf{x}, t) s/t$  for  $0 < s < t$ . In consequence:

$$B_1 \geq E_Q \phi(\mathbf{x}, |\mathbf{x}' \theta^0| (C(\theta^0) + \varepsilon)) |\mathbf{x}' \theta^0| |\theta|/(\sigma(\theta, Q) (C(\theta^0) + \varepsilon)). \quad (6.11)$$

Now (6.9), Lemma (2.2) (ii) and (6.8) applied to (6.11) yield:  $B_1 \geq (\alpha/2) (1/A_2) (1/A_2 + \varepsilon)^{-1} = 2A_6$ ; and this, together with (6.10), yields  $B(\theta, Q) \leq -A_6$ .

## References

- Beaton, A.E., Tukey, J.W.: The fitting of power series, meaning polynomials, illustrated on bandspectroscopic data. *Technometrics* **16**, 147-185 (1974)
- Carroll, R.J., Ruppert, D.: On the asymptotic normality of robust regression estimators. Mimeo Series No. 1242, Institute of Statistics, Chapel Hill, North Carolina (1979)

- Denby, L., Martin, R.D.: Robust estimation of the first-order autoregressive parameter. *J. Amer. Statist. Assoc.* **74**, 140–146 (1979)
- Dunford, N., Schwartz, J.: *Linear Operators*. Vol. 1. New York: Interscience 1958
- Hampel, F.R.: A general qualitative definition of robustness. *Ann. Math. Statist.* **42**, 1887–1896 (1971)
- Hampel, F.R.: The influence curve and its role in robust estimation. *J. Amer. Statist. Assoc.* **69**, 383–394 (1974)
- Hampel, F.R.: Modern trends in the theory of robustness. Research Report 13, Fachgruppe für Statistik, ETH Zürich (1977)
- Huber, P.J.: Robust estimation of a location parameter. *Ann. Math. Statist.* **35**, 73–101 (1964)
- Huber, P.J.: The behavior of maximum likelihood estimates under nonstandard conditions. *Proc. Fifth Berkeley Sympos. Math. Statist. Probab.* **1**, 221–233. University of California Press (1967)
- Huber, P.J.: Robust regression: Asymptotics, conjectures, and Monte Carlo. *Ann. Statist.* **1**, 799–821 (1973)
- Huber, P.J.: Robust covariances. In: *Statistical Decision Theory and Related Topics II*, 165–191. New York: Academic Press 1977
- Krasker, W.S., Welsch, R.E.: Efficient bounded-influence regression estimation using alternative definitions of sensitivity. Technical Report No. 3. M.I.T. Center for Computational Research in Economics and Management Science. Cambridge, Massachusetts (1979)
- Mallows, C.: On some topics in robustness. Unpublished memorandum. Bell Telephone Laboratory, Murray Hill (1975)
- Maronna, R.A.: Robust  $M$ -estimators of multivariate location and scatter. *Ann. Statist.* **4**, 51–67 (1976)
- Maronna, R.A., Bustos, O., Yohai, V.J.: Bias- and efficiency-robustness of general  $M$ -estimators for regression with random carriers. *Smoothing Techniques for Curve Estimation* (T. Gasser and M. Rosenblatt, Editors) 91–116. *Lecture Notes in Mathematics* **757**. Berlin-Heidelberg-New York: Springer 1979
- Relles, D.: Robust regression by modified least squares. Ph.D. Thesis, Yale University (1968)
- Yohai, V.J.: Robust estimation in the linear model. *Ann. Statist.* **2**, 562–567 (1974)
- Yohai, V.J., Maronna, R.A.: Asymptotic behavior of  $M$ -estimates for the linear model. *Ann. Statist.* **7**, 258–268 (1979)

Received February 15, 1980