The Annals of Statistics
1989, Vol. 17, No. 1, 337–361

ON *M*-PROCESSES AND *M*-ESTIMATION[1]

BY A. H. WELSH

The Australian National University

We relate the asymptotic behavior of M-estimators of the regression parameter in a linear model in which the dimension of the regression parameter may increase with the sample size to the stochastic equicontinuity of an associated M-process. The approach synthesises a number of results for the dimensionally fixed regression model and then extends these results in a direct unified way. The resulting theorems require only mild conditions on the ψ-function and the underlying distribution function. In particular, the results do not require ψ to be smooth and hence can be applied to such estimators as the least absolute deviations estimator. We also treat one-step M-estimation.

1. Introduction. Suppose that for each n we observe $Y_1, \dots, Y_n$, where

$$Y_j = x_j'\theta + \sigma e_j, \qquad 1 \le j \le n, \tag{1.1}$$

with $\{x_j = (1, x_{j2}')\}$ a sequence of p-vectors, $\theta \in \mathbb{R}^p$ an unknown parameter, $\sigma > 0$ a nuisance parameter and $\{e_j\}$ a sequence of independent random variables with location zero, unit scale and common distribution function F. Although $\{x_j\}$ and p (and consequently θ) may depend on n, we suppress this and other dependences on n for notational simplicity. Since for each n we can center the components of each x_j, there is no loss of generality in supposing that $\bar{x} = n^{-1}\sum_{j=1}^n x_j = (1, 0, \dots, 0)'$. We will also assume that $X'X = \sum_{j=1}^n x_j x_j'$ is nonsingular for each n and put $z_j = (X'X)^{-1}x_j$, $1 \le j \le n$.

A rich class of estimators of the regression parameter θ can be represented as M-estimators. Let $\psi\colon \mathbb{R} \to \mathbb{R}$ be a given function, let $\tau \in \mathbb{R}$ be fixed and let $\{M_n(t, s)\}$ be defined by

$$M_n(t, s) = n^{1/2} \sum_{j=1}^n z_j \psi\big((1 + s)(\sigma e_j - \tau - x_j't)/\sigma\big), \qquad t \in \mathbb{R}^p,\ s \in \mathbb{R}.$$

Then an M-estimator $\hat{\theta}$ of θ [Relles (1968) and Huber (1973)] is a solution of

$$M_n\big(\hat{\theta} - \theta - \tau u_1, 0\big) = 0, \tag{1.2}$$

if σ is known or if $1 + s$ can be factored out, or more generally of

$$M_n\big(\hat{\theta} - \theta - \tau u_1, \hat{\sigma}^{-1}\sigma - 1\big) = 0, \tag{1.3}$$

Received June 1987; revised March 1988.

[1]This research was partially supported by National Science Foundation Grant DMS-86-01732. This manuscript was prepared using computer facilities supported in part by National Science Foundation Grants DMS-86-01732 and DMS-84-04941 to the Department of Statistics at The University of Chicago.

AMS 1980 *subject classifications*. 62G05, 62J05, 60F05.

Key words and phrases. Asymptotic linearity, large p asymptotics, M-estimators, one-step M-estimators, regression quantiles, robust estimation, stochastic equicontinuity.

where $u_1 = (1, 0, \dots, 0)' \in \mathbb{R}^p$ and $\hat{\sigma}$ is a location invariant and scale equivariant estimator of σ. The function ψ can be chosen to achieve asymptotic efficiency (maximum likelihood), robustness [Huber (1973)] or for convenience (least-squares or least absolute deviations). In the general asymmetric case, τ represents the asymptotic bias in estimating the intercept. The asymptotic properties of an M-estimator $\hat{\theta}$ are related to the properties of the centered process $\{M_n(t, s) - EM_n(t, s)\}$ associated with $\hat{\theta}$. We will call this process an M-process.

Many interesting inferential problems arising from the model (1.1) can be formulated in terms of fixed contrasts in the regression parameter of the form $a'\theta$, where $a \in \mathbb{R}^p$ satisfies $|a| = 1$. (Here $|\cdot|$ denotes the p-dimensional Euclidean norm.) Asymptotic inference procedures for such contrasts can be derived from the asymptotic distribution of the estimated contrasts $a'\hat{\theta}$. Another class of interesting inferential problems depends on quadratic forms in the slope estimates. Asymptotic inference procedures require the asymptotic distribution of these quadratic forms. In the classical approach, p is regarded as fixed [Huber (1973), Bickel (1975), Jurečková (1977) and Yohai and Maronna (1979)] but p may also be permitted to diverge to infinity with n [Huber (1973), Yohai and Maronna (1979), Portnoy (1984, 1985, 1986a), Antille and Milasević (1987), and Mammen (1987)]. Allowing p to diverge with n is a way of allowing the model to become more complicated as the sample size increases and, through the restrictions on the rate at which p can increase, suggests restrictions on the complexity of the model for each finite n; see Portnoy (1986a), page 1153.

Allowing p to increase with n, Huber (1973) obtained a normal approximation theorem for fixed contrasts under conditions which entail $p^3/n \to 0$, a result which was improved by Yohai and Maronna (1979) who required $p^{5/2}/n \to 0$. At the culmination of a remarkable series of papers, Portnoy (1986a) managed to weaken the condition on p to $p^{11/8}(\log n)^2/n \to 0$; see also Portnoy (1985) and Mammen (1987). Portnoy (1985) also obtained the asymptotic distribution of a quadratic form in the estimates when $(p \log n)^{3/2}/n \to 0$. While the results of Huber (1973) and Yohai and Maronna (1979) depend on essentially classical arguments, the results of Portnoy (1985, 1986a) are obtained by exploiting special features of the problem. Interestingly, Portnoy (1986b) showed that the best rate that can be hoped for using classical methods is $p^2/n \to 0$. In all of these references, ψ is assumed to have at least two smooth, bounded derivatives and σ is supposed known. It is desirable to weaken these conditions to obtain results for estimators such as the least absolute deviations estimator which has a discontinuous ψ-function and to reflect the fact that σ is not usually known. Antille and Milasević (1987) investigated a one-step analogue of the least absolute deviations estimator but their result is unsatisfactory because it requires moment conditions on $\{e_j\}$ and the very strong condition $p^6/n \to 0$.

For p fixed, the asymptotic distribution of $a'\hat{\theta}$ (or equivalently of $\hat{\theta}$) and of quadratic forms in $\hat{\theta}$ can be obtained without assuming that ψ is smooth and σ is known by invoking a stochastic equicontinuity argument: Essentially, this involves uniformly approximating $M_n(t, s)$ by $M_n(0, 0) - E\{M_n(t, s) - M_n(0, 0)\}$ and then expanding the integrated part of the approximating process. Bickel (1975), Jurečková (1977) and Jurečková and Sen (1987a, 1987b) have obtained suitable uniform approximations for $\{M_n(t, s)\}$. Bickel (1975) applied a stochastic equicontinuity argument to obtain the asymptotic distribution of one-step M-estimators and, realising that an M-estimator can be viewed as a one-step M-estimator using itself as the initial estimator, Jurečková (1977) and Yohai and Maronna (1979) invoked the stochastic equicontinuity to obtain the asymptotic distribution of a consistent M-estimator. While Yohai and Maronna (1979) gave a separate argument to show that $\hat{\theta}$ is consistent, Jurečková (1977) managed to deduce consistency from the approximating process in the stochastic equicontinuity argument. Welsh (1986) used the uniform approximation argument to investigate the problem of estimating $\hat{\sigma}$. Thus the stochastic equicontinuity argument neatly unifies the arguments required for a full treatment of the asymptotic behaviour of one-step M-estimators and M-estimators.

In this article we extend the stochastic equicontinuity argument to allow p to diverge with n. The stochastic equicontinuity argument then enables us to obtain new results for one-step M-estimators, M-estimators and robust scale estimators. One-step M-estimators are of interest both as simple approximations to M-estimators and as a means of overcoming nonuniqueness problems associated with M-estimators based on nonmonotone ψ-functions. The present results extend those of Bickel (1975) in a natural way. The results for both one-step M-estimators and M-estimators allow ψ to be discontinuous and σ to be unknown. In consequence, the results apply to the regression quantile estimators of Koenker and Bassett (1978) and in particular to the least absolute deviations estimator. The least absolute deviations estimator is an attractive initial estimator since it does not require a concomitant scale estimator. Finally, the stochastic equicontinuity results for M-processes with discontinuous ψ-functions enable us to extend the results of Welsh (1986) on scale estimation to allow p to diverge with n. Indeed, the results of Welsh (1986) hold when p diverges with n provided that F has a bounded second derivative in appropriate neighbourhoods. This article therefore presents an aesthetically and pedagogically pleasing synthesis of fixed p results, a unified extension of these results to permit p to diverge with n and a synthesis of the two types of results.

In contrast to the work of Portnoy (1984, 1985, 1986a), our objective is to impose the weakest possible conditions on ψ and F rather than on p. Nonetheless, for contrast estimation, if ψ has a bounded derivative, we require

$$p^2(\log n)^{2+\gamma}/n \to 0,$$

$\gamma > 0$, a rate which is slower only than those of Portnoy (1984, 1985, 1986a) and Mammen (1987) while if ψ is discontinuous, we require

$$p^3(\log n)^2/n \to 0,$$

a surprisingly mild restriction in view of the result of Antille and Milasević (1987). To obtain the distribution of quadratic forms, we require slower growth rates on p. In particular, if ψ has a bounded derivative we require

$$p^3(\log n)^{2+\gamma}/n \to 0,$$

$\gamma > 0$, while if ψ is discontinuous, we require

$$p^4(\log n)^2/n \to 0.$$

We present the basic stochastic equicontinuity results in Section 3 after discussing the conditions they require in Section 2. We then apply the stochastic equicontinuity results to obtain results for one-step M-estimators (Section 4) and M-estimators (Section 5).

2. Conditions. The results in this article require conditions on ψ and F jointly, on $\{x_j\}$ and on p as a function of n. Of course, we will assume that the basic linear model (1.1) holds throughout the article.

We will require:

(C.1) The function ψ is a bounded function of bounded variation, that is, $\psi = \psi^+ - \psi^-$, where $|\psi^\pm| \le K \le \infty$ and $\psi^\pm$ is monotone increasing.

The requirement that ψ be of bounded variation is standard in M-estimation. In his work on the M-process with p fixed, Bickel (1975) did not require ψ to be bounded but this condition is required for the exponential inequalities we need to apply. The remaining conditions on ψ and F depend on whether σ is known (or can be factored out) or not. Let $\tau \in \mathbb{R}$ be a fixed quantity to be specified in each section. To establish stochastic equicontinuity when σ is known, we require:

(C.2) There is a $K < \infty$ such that, as $h \to 0$,

$$\int_{-\infty}^{\infty} \{\psi^\pm(x - \sigma^{-1}\tau + h) - \psi^\pm(x - \sigma^{-1}\tau - h)\}^2 \, dF(x) \le K|h|^\alpha,$$

where $0 < \alpha \le 2$;

and:

(C.3) There is a $K < \infty$ such that for some $\varepsilon > 0$,

$$\sup_{|q| \le \varepsilon} \sup_{|h| \le \varepsilon} \frac{1}{|h|} \int_{-\infty}^{\infty} \{\psi^\pm(x - \sigma^{-1}\tau + q + h) - \psi^\pm(x - \sigma^{-1}\tau + q)\} \, dF(x) \le K.$$

If σ is not known, we also require:

(C.4) There is a $K < \infty$ such that for some $\varepsilon > 0$,

$$\sup_{|r| \le \varepsilon} \sup_{|q| \le \varepsilon} \sup_{|h| \le \varepsilon} \frac{1}{q^2} \int_{-\infty}^{\infty} \{\psi^\pm((1 + r + q)(x - \sigma^{-1}\tau + h)) - \psi^\pm((1 + r)(x - \sigma^{-1}\tau + h))\}^2 \, dF(x) \le K;$$

and:

(C.5) There is a $K < \infty$ such that for some $\varepsilon > 0$,

$$\sup_{|r| \le \varepsilon} \sup_{|q| \le \varepsilon} \sup_{|h| \le \varepsilon} \frac{1}{|h|} \int_{-\infty}^{\infty} \{\psi^{\pm}((1+r)(x - \sigma^{-1}\tau + q + h)) - \psi^{\pm}((1+r)(x - \sigma^{-1}\tau + q - h))\} \, dF(x) \le K.$$

These conditions explicitly permit smoothness requirements to be traded off between ψ and F. In the Lipschitz condition, (C.2), $\alpha = 2$ if $\psi^{\pm}$ has a bounded derivative but otherwise the weaker $\alpha < 2$ applies; for $\alpha < 2$, the case $\alpha = 1$ is of particular interest. The growth condition on p will depend on the smoothness of ψ through the value of α. The above conditions are essentially those used by Bickel (1975) for the fixed p problem.

The stochastic equicontinuity results lead to asymptotic linearity results provided either:

(C.6) There exists a $\lambda^{\pm} \ne 0$ such that

$$\int \{\psi^{\pm}(x - \sigma^{-1}\tau + h) - \psi^{\pm}(x - \sigma^{-1}\tau)\} \, dF(x) = \sigma h \lambda^{\pm} + O(h^2);$$

or:

(C.7) There exist a $\lambda^{\pm} \ne 0$ and an $\eta^{\pm}$ such that

$$\int \{\psi^{\pm}((1+q)(x - \sigma^{-1}\tau + h)) - \psi^{\pm}(x - \sigma^{-1}\tau)\} \, dF(x) = \sigma h \lambda^{\pm} + \sigma q \eta^{\pm} + O(h^2 + |qh| + q^2),$$

depending on whether σ is known or not. In the sequel, let

$$\lambda = \lambda^{+} - \lambda^{-}, \qquad \eta = \eta^{+} - \eta^{-}.$$

If ψ is absolutely continuous with derivative ψ', $\lambda = \sigma^{-1}\int \psi'(x - \sigma^{-1}\tau) \, dF(x)$ and $\eta = \sigma^{-1}\int (x - \sigma^{-1}\tau)\psi'(x - \sigma^{-1}\tau) \, dF(x)$ so that $\eta = 0$ when F is symmetric about $\sigma^{-1}\tau$ and ψ is antisymmetric. These conditions are required by Bickel (1975) except that he imposes the additional condition that $\eta = 0$. To treat the one-step M-estimators and the M-estimators, we will set

$$d = \int \psi(x - \sigma^{-1}\tau)^2 \, dF(x).$$

Finally, in Section 5, we will also require:

(C.8) There is a $T(F) \in \mathbb{R}$ such that $\int \psi(x - \sigma^{-1}T(F)) \, dF(x) = 0$ and (C.1)–(C.7) to hold with $\tau = T(F)$. Clearly, $T(F) = 0$ if F is symmetric about the origin and ψ is antisymmetric.

The conditions on the sequence $\{x_j\}$ are intended to hold in probability if $\{x_j\}$ is a sample from an appropriate distribution. As noted in the Introduction, we suppose that $x_{j1} = 1$, $1 \le j \le n$, that we center the components of each x_j so that for each n, $\bar{x} = n^{-1}\sum_{j=1}^n x_j = (1, 0, \dots, 0)'$ and $X'X = \sum_{j=1}^n x_j x_j'$ is nonsingular. These conditions imply that

$$(X'X)^{-1} = \begin{pmatrix} n^{-1} & 0' \\ 0 & \left(\sum_{j=1}^n x_{j2}x_{j2}'\right)^{-1} \end{pmatrix},$$

where $\{x_j' = (1, x_{j2}')\}$. Centering the design ensures that it is possible to estimate the slope when the underlying distribution is asymmetric since any bias will appear only in the intercept [see Carroll and Welsh (1988)]. We will take

$$u_1 = (1, 0, \dots, 0)' \in \mathbb{R}^p$$

throughout the sequel. We will also require:

(C.9) There are constants $0 < C' < C < \infty$ such that the maximum and minimum eigenvalues of $X'X$ satisfy

$$\lambda_{\max}(X'X) \le nC$$

and

$$\lambda_{\min}(X'X) \ge nC';$$

(C.10) There is a $C < \infty$ such that

$$\max_{1 \le j \le n} |x_j'(X'X)^{-1}x_j| \le (p/n)C;$$

and:

(C.11) There is a $C < \infty$ such that for each $a \in \mathbb{R}^p$ satisfying $|a| = 1$, $\max_{1 \le j \le n}|a'(X'X)^{-1}x_j| \le n^{-1}C$ and $\max_{1 \le j \le n}\|(X'X)^{-1}x_j\| \le n^{-1}C$, where $\|\cdot\|$ is the maximum norm.

As noted by Portnoy (1985), it follows from (C.9) and (C.10) that there is a $C < \infty$ such that

$$\max_{1 \le j \le n} |x_j|^2 \le pC \quad \text{and} \quad \sum_{j=1}^n |x_j|^2 \le npC.$$

We will frequently use the fact that

$$\begin{aligned} \sup_{|t| \le (p/n)^{1/2}B} \sum_{j=1}^n (x_j't)^2 &= \sup_{|t| \le (p/n)^{1/2}B} |t|^2 \frac{t'X'Xt}{|t|^2} \\ &\le (p/n)B^2\lambda_{\max}(X'X) \\ &\le pB^2C, \end{aligned}$$

by (C.9), and hence that

$$\sup_{|t| \leq (p/n)^{1/2}B} \sum_{j=1}^{n} |x_j't| \leq \sup_{|t| \leq (p/n)^{1/2}B} n^{1/2}\left(\sum_{j=1}^{n} (x_j't)^2\right)^{1/2}$$
$$\leq n^{1/2}p^{1/2}BC^{1/2}.$$

The above conditions are slightly simpler and weaker than those imposed by Portnoy (1985, 1986a).

It is useful to note that if we reparametrise the model (1.1) so that we have

$$(2.1) \qquad Y_j = w_j'\beta + \sigma e_j, \qquad 1 \leq j \leq n,$$

where $w_j = n^{1/2}(X'X)^{-1/2}x_j$, $1 \leq j \leq n$, and $\beta = n^{-1/2}(X'X)^{1/2}\theta$, and conditions (C.9), (C.10) and

(C.11′) There is a $C < \infty$ such that for each $a \in \mathbb{R}^p$ satisfying $|a| = 1$,

$$\max_{1 \leq j \leq n} |a'(X'X)^{-1/2}x_j| \leq n^{-1/2}C \quad \text{and} \quad \max_{1 \leq j \leq n} \|(X'X)^{-1/2}x_j\| \leq n^{-1/2}C$$

hold, then $\{w_j\}$ satisfies conditions (C.9)–(C.11). Hence the results of Sections 3–5 hold for the reparametrised model with condition (C.11) replaced by (C.11′).

Conditions (C.11) and (C.11′) will usually fail to hold for balanced ANOVA problems. Indeed, as noted by Portnoy (1984), for the one-way design with p cells and n/p observations per cell, we find that the squared Euclidean norm of the centered estimator is of order p^2/n rather than p/n. Thus the results of Section 3 cannot apply to this problem. This is in accordance with the finding of Portnoy (1984) that the regression problem and the balanced ANOVA problem should be treated separately. It turns out that if we modify the approach of the present article for the balanced ANOVA problem then we require more restrictive growth conditions on p than for the regression problem. We will not pursue the ANOVA problem further in this article.

The conditions we impose on p are a function of α in (C.2). We will require:

(C.12) $\phi_n \to 0$, where

$$\phi_n^2 = \begin{cases} (p/n)^{\alpha/2}p \log n, & \text{if } 0 < \alpha < 2, \\ p^2(\log n)^{2+\gamma}/n, \quad \gamma > 0, & \text{if } \alpha = 2. \end{cases}$$

The most important cases arise when $\alpha = 1$ or 2. As noted in the Introduction, for $\alpha = 2$, this condition is stronger than those of Portnoy (1985, 1986a) but weaker than those of Huber (1973) and Yohai and Maronna (1979) and, for $\alpha = 1$, it is weaker than that of Antille and Milasević (1987). It is interesting to note that these rates are simply p times the fixed p remainder rate evaluated at

n/p rather than n. Since the fixed p rates are known to be good, these rates seem quite reasonable.

3. Stochastic equicontinuity and asymptotic linearity results. For fixed p, conditions under which an M-process with a possibly discontinuous ψ-function is stochastically equicontinuous have been given by Bickel (1975) and Jurečková (1977) (without bounds on the rate of convergence) and by Jurečková and Sen (1987a, 1987b) (with bounds on the rate of convergence). Both the contiguity arguments of Jurečková (1977) and the induction argument of Jurečková and Sen (1987a) seem difficult to apply when $p \to \infty$. Our arguments are based on direct approximations as in Bickel (1975) and exponential inequalities. Although our methods are based on ideas used in the theory of empirical processes [see, for example, Pollard (1984), Chapters 2 and 7], a straight application of the results in Pollard (1984) leads to inferior results to those obtained below. Indeed, the present results are a substantial improvement over the result of Antille and Milasević (1987) which was obtained by applying empirical process results.

We will need to establish the stochastic equicontinuity of M-processes with respect to two norms. We will need a stochastic equicontinuity result for a contrast process to find the asymptotic distribution of contrast estimators and we will need a stochastic equicontinuity result in terms of the maximum norm, $\| \ \|$, of the M-process to establish a consistency result for M-estimators and to find the asymptotic distribution of quadratic forms. For p fixed, the two results are equivalent.

For the simple cases that σ is known or $(1 + s)$ can be factored out, we have the following stochastic equicontinuity result.

THEOREM 3.1. *Suppose that conditions* (C.1)–(C.3) *and* (C.9)–(C.12) *hold. Then for any* $a \in \mathbb{R}^p$ *satisfying* $|a| = 1$,

$$\sup_{|t| \le (p/n)^{1/2}B} |a'\{M_n(t,0) - EM_n(t,0) - M_n(0,0) + EM_n(0,0)\}| = O_p(\phi_n) \tag{3.1}$$

and

$$\sup_{|t| \le (p/n)^{1/2}B} \|M_n(t,0) - EM_n(t,0) - M_n(0,0) + EM_n(0,0)\| = O_p(\phi_n), \tag{3.2}$$

for each fixed $B < \infty$, *where* ϕ_n *is defined in condition* (C.12).

PROOF. Without loss of generality, suppose that $\sigma = 1$, $\tau = 0$ and ψ is nondecreasing. Cover the ball $\{|t| \le (p/n)^{1/2}B\}$ with cubes $C = \{C(t_k)\}$, where $C(t_k)$ is a cube containing t_k with sides of length $(p/n^5)^{1/2}B$ so that $N = \text{card}(C) = (2n^2)^p$, $|t_k| \le (p/n)^{1/2}B$ and for $t \in C(t_k)$, $|t - t_k| \le (p/n^{5/2})B = \beta_n$, say.

Since ψ is monotone, we have the upper bound

$$\begin{aligned}
&\sup_{|t|\le (p/n)^{1/2}B} |a'\{M_n(t,0) - EM_n(t,0) - M_n(0,0) + EM_n(0,0)\}| \\
&\quad\le \max_{1\le k\le N} |a'\{M_n(t_k,0) - EM_n(t_k,0) - M_n(0,0) + EM_n(0,0)\}| \\
&\qquad + \max_{1\le k\le N} \Bigg| n^{1/2} \sum_{j=1}^{n} |a'z_j| \{\psi(e_j - x_j't_k + |x_j|\beta_n) - E\psi(e - x_j't_k + |x_j|\beta_n) \\
&\qquad\qquad -\psi(e_j - x_j't_k) + E\psi(e - x_j't_k)\}\Bigg| \qquad (3.3)\\
&\qquad + \max_{1\le k\le N} n^{1/2} \sum_{j=1}^{n} |a'z_j| E\{\psi(e - x_j't_k + |x_j|\beta_n) - \psi(e - x_j't_k - |x_j|\beta_n)\}.
\end{aligned}$$

Now, by (C.3),

$$\begin{aligned}
&\max_{1\le k\le N} n^{1/2} \sum_{j=1}^{n} |a'z_j| E\{\psi(e - x_j't_k + |x_j|\beta_n) - \psi(e - x_j't_k - |x_j|\beta_n)\} \\
&\quad\le K\beta_n n^{1/2} \sum_{j=1}^{n} |a'z_j|\,|x_j| \\
&\quad\le (p/n^{5/2}) BKn^{1/2} \max_{1\le j\le n} |a'z_j| \sum_{j=1}^{n} |x_j| \\
&\quad\le (p/n^2) BKCn^{-1} \sum_{j=1}^{n} |x_j| \\
&\quad\le (p/n)^2 BKC^2,
\end{aligned}$$

by (C.9) and (C.10).

Next, consider the first term in (3.3). By (C.2),

$$\begin{aligned}
&n \sum_{j=1}^{n} (a'z_j)^2 E\{\psi(e - x_j't_k) - E\psi(e - x_j't_k) - \psi(e) + E\psi(e)\}^2 \\
&\quad\le n \sum_{j=1}^{n} (a'z_j)^2 E\{\psi(e - x_j't_k) - \psi(e)\}^2 \\
&\quad\le Kn \sum_{j=1}^{n} (a'z_j)^2 |x_j't_k|^{\alpha} \\
&\quad\le KCn^{-1} \sum_{j=1}^{n} |x_j't_k|^{\alpha} \\
&\quad\le (p/n)^{\alpha/2} B^2 KC(\log n)^{\gamma},
\end{aligned}$$

where $\gamma = 0$ if $\alpha < 2$ and $\gamma > 0$ if $\alpha = 2$, by (C.9) and (C.10). Also, ψ is bounded, so by Bernstein's inequality,

$$P\Big[\max_{1\le k\le N}|a'\{M_n(t_k,0)-EM_n(t_k,0)-M_n(0,0)+EM_n(0,0)\}|>4(B^2KC)^{1/2}\phi_n\Big]$$
$$\le 2N\exp\Big[-16B^2KC\phi_n^2/\Big\{B^2KC(p/n)^{\alpha/2}(\log n)^{\gamma}+4(B^2KC)^{1/2}n^{-1/2}\phi_n\Big\}\Big]$$
$$\le 2N\exp(-4p\log n)\quad(\text{for } n \text{ large enough})$$
$$\le 2\exp(-p\log n).$$

Since $\beta_n=(p/n^{5/2})B\le(p/n)^{\alpha/2}B^2KC$, the above argument can be used to show that the remaining term in (3.3) is $O_p(\phi_n)$ so (3.1) is obtained.

With only minor modifications, the above argument also yields (3.2). □

Before extending Theorem 3.1 to allow for unknown scale, we prove a preliminary lemma which plays the role of the tightness arguments used in Bickel (1975).

LEMMA 3.1. *Suppose that conditions* (C.1), (C.4) *and* (C.9)–(C.12) *hold. Let* $\{t_1,\dots,t_N\}$ *be a set of points,* $t_k\in\mathbb{R}^p$ *such that* $|t_k|\le(p/n)^{1/2}B$, *for some* $B<\infty$, $1\le k\le N=(2n^2)^p$. *Then for any* $a\in\mathbb{R}^p$ *satisfying* $|a|=1$,

$$\max_{1\le k\le N}\sup_{|s|\le n^{-1/2}/2}|a'\{M_n(t_k,s)-EM_n(t_k,s)-M_n(t_k,0)+EM_n(t_k,0)\}|$$
$$=O_p(\nu_n)$$

and

$$\max_{1\le k\le N}\sup_{|s|\le n^{-1/2}/2}\|M_n(t_k,s)-EM_n(t_k,s)-M_n(t_k,0)+EM_n(t_k,0)\|=O_p(\nu_n),$$

where $\nu_n^2=p\log n/n^{1/2}$.

PROOF. As in the proof of Theorem 3.1, suppose that $\sigma=1$, $\tau=0$ and ψ is nondecreasing. Let $\phi^+(u)=\phi(u)I(u\ge0)$ and $\phi^-(u)=\psi(u)I(u<0)$ so $\psi(u)=\phi^+(u)+\phi^-(u)$ and

$$M_n(t_k,s)-EM_n(t_k,s)-M_n(t_k,0)+EM_n(t_k,0)$$
$$=n^{1/2}\sum_{j=1}^n z_j\Big\{\phi^+\big((1+s)(e_j-x_j't_k)\big)-E\phi^+\big((1+s)(e-x_j't_k)\big)$$
$$-\phi^+(e_j-x_j't_k)+E\phi^+(e-x_j't_k)\Big\}$$
$$+n^{1/2}\sum_{j=1}^n z_j\Big\{\phi^-\big((1+s)(e_j-x_j't_k)\big)-E\phi^-\big((1+s)(e-x_j't_k)\big)$$
$$-\phi^-(e_j-x_j't_k)+E\phi^-(e-x_j't_k)\Big\}.$$

Let $\{b_n\}$ be any sequence of positive integers such that $b_n \sim n^{1/4}$. Then put $s_l = n^{-1/2}(b_n^{-1}l - 1/2)$, $l = 0, 1, \dots, b_n$, so that

$$\max_{1 \le k \le N} \sup_{|s| \le n^{-1/2}/2} \Big| n^{1/2} \sum_{j=1}^{n} a'z_j \big\{ \phi^+\big((1+s)(e_j - x_j't_k)\big) - E\phi^+\big((1+s)(e - x_j't_k)\big) - \phi^+(e_j - x_j't_k) + E\phi^+(e - x_j't_k) \big\} \Big|$$

$$\begin{aligned}
&\le \max_{1 \le k \le N} \max_{1 \le l \le b_n} \Big| n^{1/2} \sum_{j=1}^{n} a'z_j \big\{ \phi^+\big((1+s_l)(e_j - x_j't_k)\big) - E\phi^+\big((1+s_l)(e - x_j't_k)\big) \\
&\qquad - \phi^+(e_j - x_j't_k) + E\phi^+(e - x_j't_k) \big\} \Big| \\
&\quad + \max_{1 \le k \le N} \max_{1 \le l \le b_n} \Big| n^{1/2} \sum_{j=1}^{n} |a'z_j| \big\{ \phi^+\big((1+s_l)(e_j - x_j't_k)\big) - E\phi^+\big((1+s_l)(e - x_j't_k)\big) \\
&\qquad - \phi^+\big((1+s_{l-1})(e_j - x_j't_k)\big) + E\phi^+\big((1+s_{l-1})(e - x_j't_k)\big) \big\} \Big| \\
&\quad + \max_{1 \le k \le N} \max_{1 \le l \le b_n} 2n^{1/2} \sum_{j=1}^{n} |a'z_j| E\big\{ \phi^+\big((1+s_l)(e - x_j't_k)\big) - \phi^+\big((1+s_{l-1})(e - x_j't_k)\big) \big\}.
\end{aligned} \tag{3.4}$$

Now, the last term in (3.4) is bounded above by

$$\begin{aligned}
&\max_{1 \le k \le N} \max_{1 \le l \le b_n} 2n^{1/2} \sum_{j=1}^{n} |a'z_j| \Big[E\big\{ \phi^+\big((1+s_l)(e - x_j't_k)\big) - \phi^+\big((1+s_{l-1})(e - x_j't_k)\big) \big\}^2 \Big]^{1/2} \\
&\le \max_{1 \le k \le N} \max_{1 \le l \le b_n} 2n^{1/2} \sum_{j=1}^{n} |a'z_j| \Big[E\big\{ \psi\big((1+s_l)(e - x_j't_k)\big) - \psi\big((1+s_{l-1})(e - x_j't_k)\big) \big\}^2 \Big]^{1/2} \\
&\le K^{1/2} b_n^{-1} \sum_{j=1}^{n} |a'z_j| \\
&= O(n^{-1/4}),
\end{aligned}$$

by (C.4) and (C.11). Also, by (C.4),

$$n \sum_{j=1}^{n} (a'z_j)^2 E\{\phi^+((1+s_l)(e-x_j't_k)) - \phi^+(e-x_j't_k)\}^2$$
$$\le n \sum_{j=1}^{n} (a'z_j)^2 E\{\psi((1+s_l)(e-x_j't_k)) - \psi(e-x_j't_k)\}^2$$
$$\le n^{-1/2}K,$$

so that by Bernstein's inequality, the first term in (3.4) is $O_p(\nu_n)$. A similar argument shows that the second term in (3.4) is also $O_p(\nu_n)$.

With only minor modifications to the above argument, we can replace ϕ^+ by ϕ^-. Then combining the results for ϕ^+ and ϕ^-, we obtain the first part of the lemma.

Similar arguments yield the second statement in the lemma. □

We can now prove a stochastic equicontinuity result which is applicable when σ is unknown.

THEOREM 3.2. *Suppose that conditions* (C.1)–(C.5) *and* (C.9)–(C.12) *hold. Then for any* $a \in \mathbb{R}^p$ *satisfying* $|a| = 1$,

$$\sup_{|t| \le (p/n)^{1/2}B} \sup_{|s| \le n^{-1/2}/2} |a'\{M_n(t,s) - EM_n(t,s) - M_n(0,0) + EM_n(0,0)\}| = O_p(\phi_n) \tag{3.5}$$

and

$$\sup_{|t| \le (p/n)^{1/2}B} \sup_{|s| \le n^{-1/2}/2} \|M_n(t,s) - EM_n(t,s) - M_n(0,0) + EM_n(0,0)\| = O_p(\phi_n), \tag{3.6}$$

for each fixed $B < \infty$, *where* ϕ_n *is defined in condition* (C.12).

PROOF. As in the proof of Theorem 3.1, suppose that $\sigma = 1$, $\tau = 0$ and ψ is nondecreasing. Also, cover the ball $\{|t| \le (p/n)^{1/2}B\}$ with the cubes C defined in Theorem 3.1. Then

$$\begin{aligned} &\sup_{|t| \le (p/n)^{1/2}B} \sup_{|s| \le n^{-1/2}/2} |a'\{M_n(t,s) - EM_n(t,s) - M_n(0,0) + EM_n(0,0)\}| \\ &\quad \le \max_{1 \le k \le N} |a'\{M_n(t_k,0) - EM_n(t_k,0) - M_n(0,0) + EM_n(0,0)\}| \\ &\qquad + \max_{1 \le k \le N} \sup_{|s| \le n^{-1/2}/2} |a'\{M_n(t_k,s) - EM_n(t_k,s) - M_n(t_k\, 0) + EM_n(t_k,0)\}| \\ &\qquad + \max_{1 \le k \le N} \sup_{t \in C(t_k)} \sup_{|s| \le n^{-1/2}/2} |a'\{M_n(t,s) - EM_n(t,s) - M_n(t_k,s) + EM_n(t_k,s)\}|. \end{aligned} \tag{3.7}$$

The first term is treated in the proof of Theorem 3.1 and the second term is treated by Lemma 3.1 so it remains to consider the third term.

Now, since ψ is monotone,

$$\max_{1\le k\le N}\sup_{t\in C(t_k)}\sup_{|s|\le n^{-1/2}/2}|a'\{M_n(t,s)-EM_n(t,s)-M_n(t_k,s)+EM_n(t_k,s)\}|$$

$$\begin{aligned}\le \max_{1\le k\le N}\sup_{|s|\le n^{-1/2}/2}\Bigg|n^{1/2}\sum_{j=1}^{n}|a'z_j|\Big\{&\psi\big((1+s)(e_j-x_j't_k+|x_j|\beta_n)\big)\\ &-E\psi\big((1+s)(e-x_j't_k+|x_j|\beta_n)\big)\\ &-\psi(e_j-x_j't_k+|x_j|\beta_n)\\ &+E\psi(e-x_j't_k+|x_j|\beta_n)\Big\}\Bigg|\\ +\max_{1\le k\le N}\sup_{|s|\le n^{-1/2}/2}\Bigg|n^{1/2}\sum_{j=1}^{n}|a'z_j|\Big\{&\psi\big((1+s)(e_j-x_j't_k)\big)\\ &-E\psi\big((1+s)(e-x_j't_k)\big)\\ &-\psi(e_j-x_j't_k)+E\psi(e-x_j't_k)\Big\}\Bigg|\\ +\max_{1\le k\le N}\Bigg|n^{1/2}\sum_{j=1}^{n}|a'z_j|\Big\{&\psi(e_j-x_j't_k+|x_j|\beta_n)\\ &-E\psi(e-x_j't_k+|x_j|\beta_n)-\psi(e_j-x_j't_k)\\ &+E\psi(e-x_j't_k)\Big\}\Bigg|\\ +\max_{1\le k\le N}\sup_{|s|\le n^{-1/2}/2}n^{1/2}\sum_{j=1}^{n}|a'z_j|E\Big\{&\psi\big((1+s)(e-x_j't_k+|x_j|\beta_n)\big)\\ &-\psi\big((1+s)(e-x_j't_k-|x_j|\beta_n)\big)\Big\},\end{aligned}$$

where $\beta_n=(p/n^{5/2})B$. The first term can be treated by a slight variation of Lemma 3.1, the second term can be treated by Lemma 3.1 and the last two terms can be treated as in the proof of Theorem 3.1.

A similar argument yields (3.6). □

In order to obtain asymptotic linearity results, we need to expand $E(M_n(t,0)-M_n(0,0)\}$ or $E\{M_n(t,s)-M_n(0,0)\}$ depending on whether σ is known or not. If σ is known and (C.6) holds,

$$|Ea'\{M_n(t,0) - M_n(0,0) - n^{1/2}t\lambda\}|$$
$$= \left| n^{1/2} \sum_{j=1}^{n} a'z_j E\{\psi(e_1 - (\tau + x_j't)/\sigma) - \psi(e_1 - \sigma^{-1}\tau) - x_j't\lambda\} \right|$$
$$\le Kn^{1/2} \sum_{j=1}^{n} |a'z_j|(x_j't)^2$$
$$\le KCn^{-1/2} \sum_{j=1}^{n} (x_j't)^2$$
$$\le (p/n^{1/2})KC^2,$$

by (C.9) and (C.11). Similarly,

$$\|E\{M_n(t,0) - M_n(0,0) - n^{1/2}t\lambda\}\| \le (p/n^{1/2})KC^2.$$

Combining the above expansions with Theorem 3.1, we have the following asymptotic linearity result.

THEOREM 3.3. *Suppose that conditions* (C.1)–(C.3), (C.6) *and* (C.9)–(C.12) *hold. Then*

$$\sup_{|t| \le (p/n)^{1/2}B} |a'\{M_n(t,0) - M_n(0,0) + n^{1/2}t\lambda\}| = O_p(\phi_n)$$

and

$$\sup_{|t| \le (p/n)^{1/2}B} \|M_n(t,0) - M_n(0,0) + n^{1/2}t\lambda\| = O_p(\phi_n),$$

for each fixed $B < \infty$, *where* ϕ_n *is defined in condition* (C.12).

Similar arguments lead to the asymptotic linearity result for the unknown σ case.

THEOREM 3.4. *Suppose that conditions* (C.1)–(C.5), (C.7) *and* (C.9)–(C.12) *hold. Then*

$$\sup_{|t| \le (p/n)^{1/2}B} \sup_{|s| \le n^{-1/2}/2} |a'\{M_n(t,s) - M_n(0,0) + n^{1/2}t\lambda - n^{1/2}s\sigma\eta u_1\}|$$
$$= O_p(\phi_n)$$

and

$$\sup_{|t| \le (p/n)^{1/2}B} \sup_{|s| \le n^{-1/2}/2} \|M_n(t,s) - M_n(0,0) + n^{1/2}t\lambda - n^{1/2}s\sigma\eta u_1\| = O_p(\phi_n),$$

for each fixed $B < \infty$, *where* ϕ_n *is defined in condition* (C.12).

4. One-step M-estimators. To obtain an M-estimator of θ, we need to solve the estimating equations (1.2) or (1.3) and then, if there are multiple solutions, we need to identify a particular solution to be the estimator of θ. Consequently, it is important to investigate the properties of the solution sequence obtained from a specified algorithm. The estimating equations are often solved by an iterative Newton–Raphson procedure so that a simple algorithm for obtaining a unique approximate solution to (1.2) or (1.3) is to take a predetermined number of Newton–Raphson steps from an initial estimator. The resulting estimator is a finite sequence of one-step estimators and its analysis can be reduced to that of one-step estimators. While one-step M-estimators may arise as convenient practical approximations to M-estimators, they can have theoretical advantages over M-estimators. Le Cam (1956) has shown that one-step estimators are no less efficient and can even be more efficient than their fully iterated counterparts.

For the fixed p problem, Bickel (1975) considered two types of one-step estimators. Let θ^* be an initial estimator of θ and let $r_j = Y_j - x_j'\theta^*$, $1 \le j \le n$, denote the residuals from θ^*. Also let $\hat{\sigma}$ be a location invariant and scale equivariant estimator of σ. If ψ is absolutely continuous with derivative ψ', a type I one-step M-estimator is defined by

$$\tilde{\theta} = \theta^* + \left\{ \sum_{j=1}^{n} x_j x_j' \psi'(r_j/\hat{\sigma}) \right\}^{-1} \hat{\sigma} \sum_{j=1}^{n} x_j \psi(r_j/\hat{\sigma}).$$

Alternatively, whether ψ is smooth or not, a type II one-step M-estimator is defined by

$$\tilde{\theta} = \theta^* + \hat{\lambda}^{-1}(X'X)^{-1} \sum_{j=1}^{n} x_j \psi(r_j/\hat{\sigma}),$$

where $\hat{\lambda}$ is an estimator of $\lambda = \lambda^+ - \lambda^-$ defined in conditions (C.6) and (C.7). Of course, if σ is known, we put $\hat{\sigma} = \sigma$.

The results below are conveniently stated in terms of the conditions presented in Section 2 and the following additional conditions:

(I.1) There is an estimator θ^* such that $|\theta^* - \theta - \tau_0 u_1| = O_p((p/n)^{1/2})$ and $a'(\theta^* - \theta - \tau_0 u_1) = O_p(n^{-1/2})$ for some fixed $\tau_0 \in \mathbb{R}$ and $a \in \mathbb{R}^p$ satisfying $|a| = 1$;

(I.2) There is a location invariant and scale equivariant estimator $\hat{\sigma}$ such that $\hat{\sigma} - \sigma = O_p(n^{-1/2})$; and

(I.3) There is an estimator $\hat{\lambda}$ such that $\hat{\lambda} - \lambda = O_p(n^{-1/2})$.

To approximate the distribution of a quadratic form in the slope estimates, we will replace (I.1) by:

(I.1′) There is an estimator θ^* such that $|(X'X)^{1/2}(\theta^* - \theta - \tau_0 u_1)| = O_p(p^{1/2})$ for some fixed $\tau_0 \in \mathbb{R}$.

These conditions are essentially required by Bickel (1975). The choice of $\hat{\lambda}$ depends on the precise nature of λ and so cannot be pursued in any generality.

However, if ψ is absolutely continuous, we may take $\hat{\lambda} = n^{-1}\Sigma_{j=1}^{n}\psi'(r_j/\hat{\sigma})/\hat{\sigma}$ and impose further conditions on ψ to ensure that (I.3) holds with $\lambda = E\psi'(e - \sigma^{-1}\tau_0)/\sigma$. If $E\psi(e - \sigma^{-1}\tau_0) = 0$, we can weaken (I.3) to require only that $\hat{\lambda} \to_P \lambda$. The results of Section 3 and the arguments of Welsh (1986) yield conditions under which the median absolute deviation from the median and the semi-interquartile range satisfy (I.2). We will exhibit estimators satisfying (I.1) in Section 5. Estimators satisfying (I.1) under conditions different from those given in Section 2 may be found in Huber (1973), Yohai and Maronna (1979) and Portnoy (1985, 1986a). Estimators satisfying (I.1′) may be obtained from the results of Section 5 by applying the arguments of Sections 3 and 5 to the reparametrised model (2.1).

We begin by considering the type II contrast estimator. The following generalisation of Theorem 4.1 of Bickel (1975) establishes conditions under which the type II estimator has an asymptotic normal distribution.

THEOREM 4.1. *Suppose that conditions* (C.1)–(C.7), (C.9)–(C.12) *and* (I.1)–(I.3) *hold with* $\tau = \tau_0$. *Then if* $\tilde{\theta}$ *is a type II one-step M-estimator and* $a \in \mathbb{R}^p$ *satisfies* $|a| = 1$,

$$\Big[a'\big(\tilde{\theta} - \theta - \{\tau_0 + \lambda^{-1}E\psi(e - \sigma^{-1}\tau_0)\}u_1\big) - (n^{1/2}\lambda)^{-1}a'\{M_n(0,0) - EM_n(0,0)\} + \lambda^{-1}\eta(\hat{\sigma} - \sigma)a'u_1 + \lambda^{-2}E\psi(e - \sigma^{-1}\tau_0)(\hat{\lambda} - \lambda)a'u_1\Big]\Big/\{a'(X'X)^{-1}a\}^{1/2} \to_P 0.$$

Consequently, if either $\eta = 0$ *and* $E\psi(e - \sigma^{-1}\tau_0) = 0$ *or* $a_1 = 0$,

$$a'(\tilde{\theta} - \theta - \tau_0 u_1)\lambda/\{a'(X'X)^{-1}ad\}^{1/2} \to_{\mathscr{D}} N(0,1).$$

If σ *is known, conditions* (C.4), (C.5), (C.7) *and* (I.2) *can be omitted.*

PROOF. Notice that with $\tau = \tau_0$,

$$\begin{aligned}
n^{1/2}a'(\tilde{\theta} - \theta - \tau_0 u_1)
&= n^{1/2}a'(\theta^* - \theta - \tau_0 u_1) + n^{1/2}\hat{\lambda}^{-1}a'(X'X)^{-1}\sum_{j=1}^{n} x_j\psi(r_j/\hat{\sigma})\\
&= n^{1/2}a'(\theta^* - \theta - \tau_0 u_1) + \hat{\lambda}^{-1}a'M_n(\theta^* - \theta - \tau_0 u_1, \sigma\hat{\sigma}^{-1} - 1)\\
&= \hat{\lambda}^{-1}(\hat{\lambda} - \lambda)n^{1/2}a'(\theta^* - \theta - \tau_0 u_1)\\
&\quad + \hat{\lambda}^{-1}a'\{\lambda n^{1/2}(\theta^* - \theta - \tau_0 u_1) + \hat{\lambda}^{-1}a'M_n(\theta^* - \theta - \tau_0 u_1, \sigma\hat{\sigma}^{-1} - 1)\}\\
&= \hat{\lambda}^{-1}\{a'M_n(0,0) + \sigma\eta n^{1/2}(\sigma\hat{\sigma}^{-1} - 1)a'u_1\} + o_p(1)\\
&= \lambda^{-1}a'\{M_n(0,0) - EM_n(0,0)\} - \lambda^{-1}\eta n^{1/2}(\hat{\sigma} - \sigma)a'u_1\\
&\quad - \lambda^{-2}E\psi(e - \sigma^{-1}\tau_0)n^{1/2}(\hat{\lambda} - \lambda)a'u_1\\
&\quad + n^{1/2}\lambda^{-1}E\psi(e - \sigma^{-1}\tau_0)a'u_1 + o_p(1),
\end{aligned}$$

by (I.1)–(I.3) and Theorem 3.4. Now

$$\max_{1\le j\le n} |a'z_j| / \left\{a'(X'X)^{-1}a\right\}^{1/2} \le \{\lambda_{\max}(X'X)\}^{1/2} \max_{1\le j\le n} |a'z_j| \to 0,$$

by (C.9) and (C.11), and

$$\sum_{j=1}^{n} (a'z_j)^2 / a'(X'X)^{-1}a = 1,$$

so

$$\begin{aligned} a'\{M_n(0,0) - EM_n(0,0)\} / \left\{na'(X'X)^{-1}a\right\}^{1/2} \\ = \sum_{j=1}^{n} a'z_j\left\{\psi\left(e_j - \sigma^{-1}\tau_0\right) - E\psi\left(e - \sigma^{-1}\tau_0\right)\right\} / \left\{a'(X'X)^{-1}a\right\}^{1/2} \\ \to_{\mathscr{D}} N(0, d), \end{aligned}$$

by the central limit theorem. The result is then obtained from Slutsky's theorem. □

The estimator considered by Antille and Milasević (1987) is a type II M-estimator so Theorem 4.1 improves on their result.

The following result yields conditions under which a quadratic form in the slope estimate has an asymptotic normal distribution and hence the distribution of the quadratic form may be approximated by a chi-squared distribution with $p - 1$ degrees of freedom.

THEOREM 4.2. *Suppose that conditions* (C.1)–(C.7), (C.9), (C.10), (C.11′), (I.1′), (I.2) *and* (I.3) *hold with* $\tau = \tau_0$. *Then if* $\tilde{\theta}$ *is a type II one-step M-estimator and* $p\phi_n^2 \to 0$ *with* ϕ_n *defined in* (C.12),

$$\left\{\lambda^2(\tilde{\theta}_2 - \theta_2)'X_2'X_2(\tilde{\theta}_2 - \theta_2)/d - (p-1)\right\} / \{2(p-1)\}^{1/2} \to_{\mathscr{D}} N(0,1),$$

where $\tilde{\theta}_2$ *and* θ_2 *denote the slope estimator and parameter, respectively, and* $X_2'X_2 = \sum_{j=1}^{n} x_{j2}x_{j2}'$ *with* $x_j' = (1, x_{j2}')$, $1 \le j \le n$. *If* σ *is known, conditions* (C.4), (C.5), (C.7) *and* (I.2) *can be omitted.*

PROOF. With $\tau = \tau_0$, we have in an obvious notation that

$$\begin{aligned} \hat{\lambda}(X_2'X_2)^{1/2}(\tilde{\theta}_2 - \theta_2) = n^{-1/2}(X_2'X_2)^{1/2}M_{n2}(0,0) \\ +(\hat{\lambda} - \lambda)(X_2'X_2)^{1/2}(\theta_2^* - \theta_2) + R_{n2}, \end{aligned}$$

where

$$\begin{aligned} R_{n2} = n^{-1/2}(X_2'X_2)^{1/2}\left\{M_{n2}\left(\theta^* - \theta - \tau_0 u_1, \sigma\hat{\sigma}^{-1} - 1\right) \right. \\ \left. -M_{n2}(0,0) + n^{1/2}\lambda(\theta_2^* - \theta_2)\right\}. \end{aligned}$$

We then take the squared Euclidean norm of both sides and multiply out the right-hand side. By Theorem 4.1 of Portnoy (1985) (see the note following

Portnoy's Theorem 4.1),

$$\left\{|n^{-1/2}(X_2'X_2)^{1/2}M_{n2}(0,0)|^2/d - (p-1)\right\}/\{2(p-1)\}^{1/2} \to_{\mathscr{D}} N(0,1)$$

so the result will follow if the remaining five terms divided by $(p-1)^{1/2}$ converge in probability to zero. Since $|n^{-1/2}(X_2'X_2)^{1/2}M_{n2}(0,0)| = O_p((p-1)^{1/2})$, $|(X_2'X_2)^{1/2}(\theta_2^* - \theta_2)| = O_p(p^{1/2})$ by (I.1′) and $\hat{\lambda} - \lambda = O_p(n^{-1/2})$ by (I.3), the result will follow if we can show that $|R_{n2}| \to_{\mathscr{P}} 0$. But

$$|R_{n2}| \le p^{1/2}\|R_{n2}\| = O_p\left(p^{1/2}\phi_n\right),$$

by Theorem 3.4 applied to the reparametrised model (2.1) and the result is obtained. □

The arguments of Theorems 4.1 and 4.2 and hence analogous results apply to type I estimators under additional smoothness conditions on ψ, but we will not pursue the matter.

5. *M*-estimators. M-estimators are of interest both in their own right and as potential initial estimators for the one-step M-estimators. In practice, M-estimators which arise as the unique solution of the equations (1.2) or (1.3) are of particular importance in both contexts. The least absolute deviations estimator has the further advantage of not requiring a concomitant scale estimator and hence is a useful initial estimator. The results below establish that these estimators can be used to make inferences about the regression parameter or as initial estimators but are not restricted to these estimators.

We derive the asymptotic distribution of a contrast of an M-estimator $\hat{\theta}$ from the asymptotic linearity results in Section 3. The argument is straightforward once we establish that $\hat{\theta}$ satisfies the consistency condition $|\hat{\theta} - \theta - T(F)u_1| = O_p((p/n)^{1/2})$. We will actually use the asymptotic linearity results to establish that the equations (1.2) or (1.3) admit a solution which satisfies the consistency condition. We establish the consistency condition for M-estimators based on continuous but not necessarily monotone ψ-functions and for M-estimators based on monotone but not necessarily continuous ψ-functions. For the first case, we extend an argument of Portnoy (1984) while for the second we extend an argument of Jurečková (1977). The results apply to the regression quantile estimators of Koenker and Bassett (1978) and hence in particular to the least absolute deviations estimator.

We begin by considering M-estimators associated with continuous but not necessarily monotone ψ-functions.

THEOREM 5.1. *Suppose that conditions* (C.1)–(C.12) *and* (I.2) *hold with* $\tau = T(F)$. *Then if* ψ *is continuous, there is a solution* $\hat{\theta}$ *of* (1.3) *such that*

$$|\hat{\theta} - \theta - T(F)u_1| = O_p\left((p/n)^{1/2}\right).$$

If σ *is known, conditions* (C.4), (C.5), (C.7) *and* (I.2) *can be omitted.*

PROOF. Let $\mathscr{B} = \{t \in \mathbb{R}^p: |t| = (p/n)^{1/2}B\}$ for some B. Then, arguing as in the proof of Theorem 3.2 of Portnoy (1984), it is enough to show that $t'M_n(t, \sigma\hat{\sigma}^{-1} - 1) < 0$ for $t \in \mathscr{B}$ in probability. Let $\varepsilon > 0$ be given. Then for $L > 0$,

$$\begin{aligned}
&P\{t'M_n(t, \sigma\hat{\sigma}^{-1} - 1) < 0 \text{ for all } t \in \mathscr{B}\} \\
&\quad\geq P\{t'M_n(t, \sigma\hat{\sigma}^{-1} - 1) < -(p/n^{1/2})L \text{ for all } t \in \mathscr{B}\} \\
&\quad\geq P\{t'M_n(0,0) - n^{1/2}|t|^2\lambda - n^{1/2}(\hat{\sigma} - \sigma)\eta t'u_1 \\
&\qquad\qquad < -(p/n^{1/2})2L \text{ for all } t \in \mathscr{B}\} \\
&\qquad - P\{t'M_n(t, \sigma\hat{\sigma}^{-1} - 1) - t'M_n(0,0) \\
&\qquad\qquad + n^{1/2}|t|^2\lambda + n^{1/2}(\hat{\sigma} - \sigma)\eta t'u_1 > (p/n^{1/2})L \text{ for all } t \in \mathscr{B}\} \\
&\quad\geq P\{t'M_n(0,0) - n^{1/2}(\hat{\sigma} - \sigma)\eta t'u_1 < (p/n^{1/2})(B^2\lambda - 2L) \text{ for all } T \in \mathscr{B}\} \\
&\qquad - P\{|M_n(t, \sigma\hat{\sigma}^{-1} - 1) - M_n(0,0) \\
&\qquad\qquad - n^{1/2}t\lambda + n^{1/2}(\hat{\sigma} - \sigma)\eta u_1| > p^{1/2}L/B \text{ for all } t \in \mathscr{B}\} \\
&\quad\geq P\{|M_n(0,0) - n^{1/2}(\hat{\sigma} - \sigma)\eta u_1| < p^{1/2}(B\lambda - 2L/B)\} \\
&\qquad - P\{\|M_n(t, \sigma\hat{\sigma}^{-1} - 1) - M_n(0,0) \\
&\qquad\qquad + n^{1/2}t\lambda + n^{1/2}(\hat{\sigma} - \sigma)\eta u_1\| > L/B \text{ for all } t \in \mathscr{B}\} \\
&\quad\geq P\{|M_n(0,0)| < \tfrac{1}{2}p^{1/2}(B\lambda - 2L/B)\} \\
&\qquad - P\{|n^{1/2}(\hat{\sigma} - \sigma)\eta| > \tfrac{1}{2}p^{1/2}(B\lambda - 2L/B)\} \\
&\qquad - P\Big\{\sup_{|t| \leq (p/n)^{1/2}B} \|M_n(t, \sigma\hat{\sigma}^{-1} - 1) - M_n(0,0) \\
&\qquad\qquad + n^{1/2}t\lambda + n^{1/2}(\hat{\sigma} - \sigma)\eta u_1\| > L/B\Big\}.
\end{aligned}$$

But

$$\begin{aligned}
E|M_n(0,0)|^2 &\leq 2dn \sum_{j=1}^{n} z_j'z_j \\
&\leq 2dn \operatorname{trace}\{(X'X)^{-1}\} \\
&\leq 2dnp\{\lambda_{\min}(X'X)\}^{-1} \\
&\leq 2pd/C'
\end{aligned}$$

so that for B large enough, by Chebyshev's inequality,

$$P\{|M_n(0,0)| < \tfrac{1}{2}p^{1/2}(B\lambda - 2L/B)\} \geq 1 - \varepsilon/3.$$

Also, for L and n large enough,

$$P\Big\{\sup_{|t|\le(p/n)^{1/2}B}\|M_n(t,\sigma\hat\sigma^{-1}-1)-M_n(0,0)$$

$$+n^{1/2}t\lambda/\sigma+n^{1/2}(\hat\sigma-\sigma)\eta u_1\|>L/B\Big\}$$

$$\le P\Big\{\sup_{|t|\le(p/n)^{1/2}B}\sup_{|s|\le n^{-1/2}/2}\|M_n(t,s)-M_n(0,0)$$

$$+n^{1/2}t\lambda/\sigma-n^{1/2}s\sigma\eta u_1\|>L/2B\Big\}$$

$$+P\{|\sigma\hat\sigma^{-1}-1|>n^{-1/2}/2\}+P\{|n^{1/2}(\sigma-\hat\sigma)(\sigma^{-1}-\hat\sigma^{-1})\sigma\eta|>L/2B\}$$

$$<\varepsilon/3$$

and

$$P\{|n^{1/2}(\hat\sigma-\sigma)\eta|>\tfrac{1}{2}p^{1/2}(B\lambda-2L/B)\}<\varepsilon/3,$$

by Theorem 3.4 and (I.2), and the result is obtained. □

The next result applies to M-estimators associated with monotone but not necessarily continuous ψ-functions. The argument below is based on the elegant argument of Jurečková (1977).

THEOREM 5.2. *Suppose that conditions* (C.1)–(C.12) *and* (I.2) *hold with* $\tau=T(F)$. *Then if* $\hat\theta$ *satisfies* (1.3) *and* ψ *is monotone,*

$$|\hat\theta-\theta-T(F)u_1|=O_p\big((p/n)^{1/2}\big).$$

If σ *is known, conditions* (C.4), (C.5), (C.7) *and* (I.2) *can be omitted.*

PROOF. Without loss of generality, suppose that ψ is nondecreasing. Then for B, $L>0$ and $\tau=T(F)$,

$$P\{|\hat\theta-\theta-T(F)u_1|\ge(p/n)^{1/2}B\}$$

$$\le P\Big\{\inf_{|t|\ge(p/n)^{1/2}B}|M_n(t,\sigma\hat\sigma^{-1}-1)|<p^{1/2}L\Big\}$$

$$+P\Big\{|\hat\theta-\theta-T(F)u_1|\ge(p/n)^{1/2}B,$$

$$\inf_{|t|\ge(p/n)^{1/2}B}|M_n(t,\sigma\hat\sigma^{-1}-1)|\ge p^{1/2}L\Big\}$$

$$=P\Big\{\inf_{|t|\ge(p/n)^{1/2}B}|M_n(t,\sigma\hat\sigma^{-1}-1)|<p^{1/2}L\Big\}.$$

If $|t|\ge(p/n)^{1/2}B$, put $v=(p/n)^{1/2}Bt/|t|$ so that $|v|=(p/n)^{1/2}B$ and $t=rv$

with $r = |t|/(p/n)^{1/2}B \geq 1$. Then

$$\begin{aligned}|M_n(t, \sigma\hat{\sigma}^{-1} - 1)| &\geq -t'M_n(t, \sigma\hat{\sigma}^{-1} - 1)/|t| \\ &= -v'M_n(rv, \sigma\hat{\sigma}^{-1} - 1)/(p/n)^{1/2}B \\ &\geq -v'M_n(v, \sigma\hat{\sigma}^{-1} - 1)/(p/n)^{1/2}B,\end{aligned}$$

as $-v'M_n(rv, \sigma\hat{\sigma}^{-1} - 1)$ is nondecreasing in r. Consequently,

$$\begin{aligned}
&P\Big\{\inf_{|t| \geq (p/n)^{1/2}B} |M_n(t, \sigma\hat{\sigma}^{-1} - 1)| < p^{1/2}L\Big\} \\
&\quad\leq P\Big\{\inf_{|v| = (p/n)^{1/2}B} - v'M_n(v, \sigma\hat{\sigma}^{-1} - 1) < (p/n^{1/2})BL\Big\} \\
&\quad\leq P\Big\{\inf_{|v| = (p/n)^{1/2}B} - v'M_n(0,0) + n^{1/2}|v|^2\lambda \\
&\qquad\qquad + n^{1/2}(\hat{\sigma} - \sigma)\eta v'u_1 < 2(p/n^{1/2})BL\Big\} \\
&\qquad + P\Big\{\inf_{|v| = (p/n)^{1/2}B} - v'M_n(v, \sigma\hat{\sigma}^{-1} - 1) < (p/n^{1/2})BL, \\
&\qquad\qquad \inf_{|v| = (p/n)^{1/2}B} - v'M_n(0,0) + n^{1/2}|v|^2\lambda \\
&\qquad\qquad\quad + n^{1/2}(\hat{\sigma} - \sigma)\eta v'u_1 > 2(p/n^{1/2})BL\Big\} \\
&\quad\leq P\Big\{-(p/n)^{1/2}B|M_n(0,0) - n^{1/2}(\hat{\sigma} - \sigma)\eta u_1| \\
&\qquad\qquad + (p/n^{1/2})B^2\lambda < 2(p/n^{1/2})BL\Big\} \\
&\qquad + P\Big\{\sup_{|v| = (p/n)^{1/2}B} v'M_n(v, \sigma\hat{\sigma}^{-1} - 1) - v'M_n(0,0) \\
&\qquad\qquad + n^{1/2}|v|^2\lambda + n^{1/2}(\hat{\sigma} - \sigma)\eta v'u_1 > (p/n)^{1/2}BL\Big\} \\
&\quad\leq P\big\{|M_n(0,0) - n^{1/2}(\hat{\sigma} - \sigma)\eta u_1| > p^{1/2}(B\lambda - 2L)\big\} \\
&\qquad + P\Big\{\sup_{|v| = (p/n)^{1/2}B} |M_n(v, \sigma\hat{\sigma}^{-1} - 1) - M_n(0,0) \\
&\qquad\qquad + n^{1/2}v\lambda + n^{1/2}(\hat{\sigma} - \sigma)\eta u_1| > p^{1/2}L\Big\} \\
&\quad\leq P\big\{|M_n(0,0)| > \tfrac{1}{2}p^{1/2}(B\lambda - 2L)\big\} \\
&\qquad + P\big\{|n^{1/2}(\hat{\sigma} - \sigma)\eta| > \tfrac{1}{2}p^{1/2}(B\lambda - 2L)\big\} \\
&\qquad + P\Big\{\sup_{|v| = (p/n)^{1/2}B} \|M_n(v, \sigma\hat{\sigma}^{-1} - 1) - M_n(0,0) \\
&\qquad\qquad + n^{1/2}v\lambda + n^{1/2}(\hat{\sigma} - \sigma)\eta u_1\| > L\Big\}.
\end{aligned}$$

The proof may be concluded as in the proof of Theorem 5.1. □

Applying Theorem 3.4 as in the proof of Theorem 4.1, we immediately obtain the following result.

THEOREM 5.3. *Suppose that conditions* (C.1)–(C.12) *and* (I.2) *hold with* $\tau = T(F)$. *Then if* ψ *is continuous or monotone and* $a \in \mathbb{R}^p$ *satisfies* $|a| = 1$, *there is a solution* $\hat{\theta}$ *of* (1.3) *such that*

$$\frac{a'\left(\hat{\theta} - \theta - T(F)u_1\right) - \left(n^{1/2}\lambda\right)^{-1} a'M_n(0,0) - \lambda^{-1}\eta(\hat{\sigma} - \sigma)a'u_1}{\left\{a'(X'X)^{-1}a\right\}^{1/2}} \to_P 0.$$

Consequently, if either $\eta = 0$ *or* $a_1 = 0$,

$$a'\left(\hat{\theta} - \theta - T(F)u_1\right)\lambda / \left\{a'(X'X)^{-1}ad\right\}^{1/2} \to_{\mathscr{D}} N(0,1).$$

Moreover, if $\tilde{\theta}$ *is a type II one-step M-estimator based on* ψ *and in addition* (I.1) *and* (I.3) *hold with* $\tau_0 = T(F)$,

$$a'(\hat{\theta} - \tilde{\theta}) / \left\{a'(X'X)^{-1}a\right\}^{1/2} \to_P 0.$$

If σ *is known, conditions* (C.4), (C.5), (C.7) *and* (I.2) *can be omitted.*

By a similar argument to that used to prove Theorem 4.2, we obtain the following result on quadratic forms.

THEOREM 5.4. *Suppose that conditions* (C.1)–(C.10), (C.11′) *and* (I.2) *hold with* $\tau = T(F)$. *Then if* ψ *is continuous or monotone and* $p\phi_n^2 \to 0$ *with* ϕ_n *defined in* (C.12) *there is a solution* $\hat{\theta}$ *of* (1.3) *such that*

$$\left\{\lambda^2\left(\hat{\theta}_2 - \theta_2\right)' X_2'X_2\left(\hat{\theta}_2 - \theta_2\right)/d - (p-1)\right\} / \{2(p-1)\}^{1/2} \to_{\mathscr{D}} N(0,1),$$

where $\hat{\theta}_2$ *and* θ_2 *denote the slope estimator and parameter, respectively, and* $X_2'X_2 = \sum_{j=1}^n x_{j2}x_{j2}'$ *with* $x_j' = (1, x_{j2}')$, $1 \le j \le n$. *If* σ *is known, conditions* (C.4), (C.5), (C.7) *and* (I.2) *can be omitted.*

PROOF. We apply either Theorem 5.2 or 5.4 to the reparametrised model (2.1) to show that $|(X'X)^{1/2}(\hat{\theta} - \theta - T(F)u_1)| = O_p(p^{1/2})$. Then we use the fact that $\hat{\theta}$ satisfies (1.3) to write in an obvious notation,

$$\lambda\left(X_2'X_2\right)^{1/2}\left(\hat{\theta}_2 - \theta_2\right) = n^{-1/2}\left(X_2'X_2\right)^{1/2} M_{n2}(0,0) + R_{n2},$$

where

$$R_{n2} = n^{-1/2}\left(X_2'X_2\right)^{1/2}\left\{M_{n2}\left(\hat{\theta} - \theta - T(F)u_1, \sigma\hat{\sigma}^{-1} - 1\right) - M_{n2}(0,0) + n^{1/2}\lambda\left(\hat{\theta}_2 - \theta_2\right)\right\}.$$

The proof is completed by arguing as in the proof of Theorem 4.2. □

It is useful for applications of Theorems 5.3 and 5.4 to note that they continue to hold if with $\tau = T(F)$,

$$p^{-1/2}|M_n(\hat{\theta} - \theta - T(F)u_1, \sigma\hat{\sigma}^{-1} - 1)| \le \|M_n(\hat{\theta} - \theta - T(F)u_1, \sigma\hat{\sigma}^{-1} - 1)\| \to_P 0$$

and

$$a'M_n(\hat{\theta} - \theta - T(F)u_1, \sigma\hat{\sigma}^{-1} - 1) \to_P 0$$

replace the condition that $\hat{\theta}$ satisfies (1.3), that is, $\hat{\theta}$ only has to satisfy (1.3) asymptotically.

We conclude this section with an application of the above results to the regression quantile estimators of Koenker and Bassett (1978). Let $0 < q < 1$ be fixed and define

$$\psi(x) = q - I(x < F^{-1}(q)), \qquad x \in \mathbb{R},$$

where F^{-1} is the usual left continuous inverse of F. Koenker and Bassett (1978) defined any value $\hat{\theta}(q)$ of θ that minimises

$$\sum_{j=1}^{n} (Y_j - x_j'\theta)\{q - I(Y_j - x_j'\theta < 0)\}$$

to be a qth regression quantile. Of course, $\hat{\theta}(1/2)$ is the least absolute deviations estimator of θ. The following result extends the results of Koenker and Bassett (1978) and Ruppert and Carroll (1980) to allow p to diverge with n. The proof is an extension of the proof of Theorem 4 of Ruppert and Carroll (1980).

THEOREM 5.5. *Let $0 < q < 1$ be fixed. Then suppose that F has two bounded derivatives in a neighbourhood of $F^{-1}(q)$ and that $F'(F^{-1}(q)) > 0$. Also suppose that conditions* (C.9)–(C.11) *hold and $p^3(\log n)^2/n \to 0$. Then if $a \in \mathbb{R}^p$ satisfies $|a| = 1$,*

$$\frac{a'(\hat{\theta}(q) - \theta(q)) + \sum_{j=1}^{n} a'z_j\{I(e_j \le F^{-1}(q)) - q\}/F'(F^{-1}(q))}{\{a'(X'X)^{-1}aq(1-q)\}^{1/2}} \to_P 0,$$

where $\theta(q) = \theta + F^{-1}(q)u_1$. Consequently,

$$\frac{a'\{\hat{\theta}(q) - \theta(q)\}F'(F^{-1}(q))}{\{a'(X'X)^{-1}aq(1-q)\}^{1/2}} \to_{\mathscr{D}} N(0,1).$$

PROOF. It is straightforward to show that the conditions of the theorem ensure that conditions (C.1)–(C.3), (C.6) and (C.8) hold for $\psi(x) = q - I(x < F^{-1}(q))$ with $\lambda = F'(F^{-1}(q))$ and $d = q(1-q)$. Consequently, the result will follow from Theorem 5.4 if we can show that $\hat{\theta}(q)$ satisfies

$$a'M_n(\hat{\theta}(q) - \theta(q), 0) \to_P 0$$

and

$$\|M_n(\hat{\theta}(q) - \theta(q), 0)\| \to_P 0.$$

For $u \in \mathbb{R}$ and $a \in \mathbb{R}^p$ satisfying $|a| = 1$, let

$$G(u, a) = \sum_{j=1}^{n} \left(Y_j - x_j'\hat{\theta}(q) - n^{1/2}a'z_j u\right)\left\{q - I\left(Y_j - x_j'\hat{\theta}(q) - n^{1/2}a'z_j u < 0\right)\right\}$$

and let $H(u, a)$ be the derivative from the right (with respect to u) of $G(u, a)$ so

$$H(u, a) = n^{1/2} \sum_{j=1}^{n} a'z_j \psi\left(Y_j - x_j'\hat{\theta}(q) - n^{1/2}a'z_j u + F^{-1}(q)\right).$$

Now $H(u, a)$ is nondecreasing in u so for $\varepsilon > 0$,

$$H(-\varepsilon, a) \le H(0, a) \le H(\varepsilon, a).$$

But $G(u, a)$ achieves its minimum at $u = 0$ so $H(-\varepsilon, a) < 0$ and $H(\varepsilon, a) > 0$ and we have

$$|H(0, a)| \le H(\varepsilon, a) - H(-\varepsilon, a).$$

Letting $\varepsilon \to 0$,

$$\begin{aligned} |H(0, a)| &\le n^{1/2} \sum_{j=1}^{n} |a'z_j| I\left(Y_j - x_j'\hat{\theta}(q) = 0\right) \\ &\le \left(p/n^{1/2}\right)C, \quad \text{almost surely,} \end{aligned}$$

by (C.11) and the fact that $\sum_{j=1}^{n} I(Y_j - x_j'\hat{\theta}(q) = 0) = p$ almost surely. Hence

$$a'M_n\left(\hat{\theta}(q) - \theta(q), 0\right) = H(0, a) \to 0, \quad \text{almost surely.}$$

Finally, notice that if $\{a_1, \ldots, a_p\}$ is the standard basis in $\mathbb{R}^p$, we have that

$$|H(0, a_i)| \le n^{1/2} \sum_{j=1}^{n} |z_{ji}| I\left(Y_j - x_j'\hat{\theta}(q) = 0\right)$$

so

$$\begin{aligned} \left\|M_n\left(\hat{\theta}(q) - \theta(q), 0\right)\right\| &= \max_{1 \le i \le p} |H(0, a_i)| \\ &\le \max_{1 \le i \le p} n^{1/2} \sum_{j=1}^{n} |z_{ji}| I\left(Y_j - x_j'\hat{\theta}(q) = 0\right) \\ &\le \left(p/n^{1/2}\right)C \quad \text{(almost surely)} \\ &\to 0 \end{aligned}$$

and the result is obtained. □

Finally, it is straightforward to obtain an approximation to the distribution of $|(X'X)^{1/2}\{\hat{\theta}(q) - \theta(q)\}|^2$.

Acknowledgments. I am grateful to David Pollard and Stephen Portnoy for their helpful conversations and encouragement and to a referee and Associate Editor for helpful suggestions including the inclusion of the results on quadratic forms.

REFERENCES

ANTILLE, A. and MILASEVIĆ, P. (1987). On linear regression with many parameters. Preprint.

BICKEL, P. J. (1975). One-step Huber estimates in the linear model. *J. Amer. Statist. Assoc.* **70** 428–433.

CARROLL, R. J. and WELSH, A. H. (1988). A note on asymmetry and robustness in linear regression. *Amer. Statist.* **42** 285–287.

HUBER, P. J. (1973). Robust regression: Asymptotics, conjectures and Monte Carlo. *Ann. Statist.* **1** 799–821.

JUREČKOVÁ, J. (1977). Asymptotic relations of M-estimates and R-estimates in linear regression model. *Ann. Statist.* **5** 464–472.

JUREČKOVÁ, J. and SEN, P. K. (1987a). An extension of Billingsley's theorem to higher dimensional M-processes. *Kybernetika* **23** 382–387.

JUREČKOVÁ, J. and SEN, P. K. (1987b). A second order asymptotic distributional representation of M-estimators with discontinuous score functions. *Ann. Probab.* **15** 814–823.

KOENKER, R. and BASSETT, G. (1978). Regression quantiles. *Econometrica* **46** 33–50.

LE CAM, L. (1956). On the asymptotic theory of estimation and testing hypotheses. *Proc. Third Berkeley Symp. Math. Statist. Probab.* **1** 129–156. Univ. California Press.

MAMMEN, E. (1987). Dimension asymptotics for robust regression with applications to the bootstrap. Preprint.

POLLARD, D. (1984). *Convergence of Stochastic Processes.* Springer, New York.

PORTNOY, S. (1984). Asymptotic behavior of M-estimators of p regression parameters when p^2/n is large. I. Consistency. *Ann. Statist.* **12** 1298–1309.

PORTNOY, S. (1985). Asymptotic behavior of M-estimators of p regression parameters when p^2/n is large. II. Normal approximation. *Ann. Statist.* **13** 1403–1417.

PORTNOY, S. (1986a). Asymptotic behavior of the empiric distribution of M-estimated residuals from a regression model with many parameters. *Ann. Statist.* **14** 1152–1170.

PORTNOY, S. (1986b). On the CLT in $\mathbb{R}^p$ when $p \to \infty$. *Probab. Theory Related Fields* **73** 571–584.

RELLES, D. (1968). Robust regression by modified least squares. Ph.D. dissertation, Yale Univ.

RUPPERT, D. and CARROLL, R. J. (1980). Trimmed least squares estimation in the linear model. *J. Amer. Statist. Assoc.* **75** 828–838.

WELSH, A. H. (1986). Bahadur representations for robust scale estimators based on regression residuals. *Ann. Statist.* **14** 1246–1251.

YOHAI, V. J. and MARONNA, R. A. (1979). Asymptotic behavior of M-estimators for the linear model. *Ann. Statist.* **7** 258–268.

DEPARTMENT OF STATISTICS
THE FACULTIES
AUSTRALIAN NATIONAL UNIVERSITY
GPO BOX 4
CANBERRA, ACT 2601
AUSTRALIA