The Annals of Statistics
1996, Vol. 24, No. 6, 2608–2630

A GENERAL BAHADUR REPRESENTATION OF *M*-ESTIMATORS AND ITS APPLICATION TO LINEAR REGRESSION WITH NONSTOCHASTIC DESIGNS

BY XUMING HE[1] AND QI-MAN SHAO[2]

University of Illinois and University of Oregon

We obtain strong Bahadur representations for a general class of M-estimators that satisfies $\sum_i \psi(x_i, \theta) = o(\delta_n)$, where the $\mathbf{x}_i$'s are independent but not necessarily identically distributed random variables. The results apply readily to M-estimators of regression with nonstochastic designs. More specifically, we consider the minimum L_p distance estimators, bounded influence GM-estimators and regression quantiles. Under appropriate design conditions, the error rates obtained for the first-order approximations are sharp in these cases. We also provide weaker and more easily verifiable conditions that suffice for an error rate that is suboptimal but strong enough for deriving the asymptotic distribution of M-estimators in a wide variety of problems.

1. Introduction. Bahadur representations are often useful to study the asymptotic properties of statistical estimators. Typically, an estimator is approximated by a sum of independent variables with a higher-order remainder; see Bahadur (1966) for some beginning work. The first-order terms may be used to measure the influence of a single observation or to derive the asymptotic distribution of the estimator. The asymptotic joint distribution of multiple statistics may also be obtained from individual Bahadur representations. Furthermore, a good error bound for the representation provides a quick guide to how good the linear approximation can be. Because of the wide applicability of M-estimators in parametric estimation, a number of authors have obtained Bahadur-type representations for M-estimators in their respective applications. For example, Carroll (1978) and Martinsek (1989) obtained strong representations for location and regression M-estimators with preliminary scale estimates. Recent work includes, among many others, Portnoy and Koenker (1989) for regression quantiles, He and Wang (1995) for multivariate location and scatter estimation, Babu (1989), Pollard (1991) and Arcones (1996a) for the least absolute deviation regression.

Some recent studies focused on M-estimators in general parametric problems. Niemiro (1992) and Bai, Rao and Wu (1992) considered a class of M-estimators defined by minimization of a convex objective function. Bose

Received July 1995; revised September 1996.
[1]Supported in part by National Security Agency Grant MDA904-96-1-0011 and University of Illinois Research Board.
[2]Supported in part by a National University of Singapore Research Project while teaching in Singapore.
AMS 1991 *subject classifications*. Primary 62F12, 62J05; secondary 60F15, 62F35.
Key words and phrases. Asymptotic approximation, Bahadur representation, M-estimator, linear regression, minimum L_p-distance estimators.

(1996) extended this approach to estimators defined through minimization of a U-statistic. A technical report of the authors, He and Shao (1994), dealt with those defined through scoring equations. A related development that can be found in Jurečková (1985) and Jurečková and Sen (1987) was to find the second-order asymptotics in weak representations. For example, it was known that the remainder term of the first-order approximation is of the order of $O_p(n^{-1})$ when the score function is twice differentiable, but it becomes $O_p(n^{-3/4})$ if the score function has jump discontinuities. The present paper shows that an almost sure Bahadur representation with good and often sharp error bounds can be obtained for a general class of M-estimators of independent observations.

Consider a sequence of variables $\{x_i,\ i = 1, \dots, n\}$ that are independent but not necessarily identically distributed. Suppose that there exists θ_0 such that $\sum_i E\psi(x_i, \theta_0) = 0$ for some score function ψ. We work with any M-estimator $\hat{\theta}_n$ of θ_0 which satisfies

$$\sum_{i=1}^{n} \psi(x_i, \widehat{\theta}_n) = o(\delta_n),$$

for some sequence δ_n.

Most authors have been concerned with i.i.d. variables. The relaxation of the assumption of identical distributions here allows this formulation to include regression models with nonstochastic designs. Representations for regression estimators with nonstochastic designs are often of more practical interest and harder to derive. In Section 2, we obtain a strong Bahadur representation for $\hat{\theta}_n$ by substantially strengthening the results of Huber (1967). It is developed for M-type estimation problems in a rather general setting. Specific applications to linear regression models are given in Section 3, where we show how the general theorems work for M-estimators with varying degrees of smoothness of their score functions. For example, the exact error bound is obtained for the minimum L_p distance estimators. The same can be done for regression quantiles and generalized M-estimators.

The reason we choose the minimum L_p distance estimators to illustrate the utility of our general representations is that such estimators are routinely used and analyzed not only in statistics but also in numerical analysis and computational geometry. There is a vast amount of literature for the algorithmic aspect of the L_p-approximation [see, e.g., Watson (1980) and Späth (1991) and the references therein]. Under appropriate conditions on the design, the error rates in the Bahadur representation are on the order of $((\log\log n)/n)^{1/4+p/2}$ for $1 \le p < 3/2$, and $(\log\log n)/n$ for $p > 3/2$. The same bounds obtained by Arcones (1996a) for random designs may be derived as a direct consequence of these results.

In some, if not many, statistical applications, a slightly suboptimal error rate would serve almost all purposes. The asymptotic distribution of an M-estimator often follows from an error rate of $o_p(n^{-1/2})$. A slight modification of our proofs shows that such results can be obtained under weaker conditions than required by Huber (1967). Our results may also facilitate asymptotic analyses for redescending M-estimators, one-step M-estimators and rank-based estimators [see, e.g., Hössjer (1994)].

The proofs of the general theorems and some of their generalizations are provided in Section 4, but detailed calculations used in the linear models of Section 3 are given in Section 5.

2. Bahadur representation: general results. In this section, we consider M-estimators in a general parametric framework. More specific applications to linear regression models will be discussed later.

Let $x_1, x_2, \dots, x_n$ be independent observations from probability distributions $F_{i,\theta}$, $i = 1, 2, \dots, n$, with a common unknown parameter $\theta \in \Theta$, an open subset of R^m, $m \geq 1$. We consider an M-estimator $\widehat{\theta}_n$ that satisfies

$$\sum_{i=1}^{n} \psi(x_i, \widehat{\theta}_n) = o(\delta_n), \tag{2.1}$$

where δ_n is a sequence of positive numbers. Various forms of δ_n have been used in the literature. For example, Huber (1967) considered $\delta_n = n$ for his consistency theorem and $\delta_n = \sqrt{n}$ for asymptotic normality. He and Wang (1995) used $\delta_n = \sqrt{n \log\log n}$ to establish the law of the iterated logarithm for $\widehat{\theta}_n$. In most cases, the left-hand side of (2.1) is actually equal to zero, but the least absolute deviation of linear regression is one important exception.

To fix notation, define

$$\Lambda_n(\theta) = \sum_{i=1}^{n} E\psi(x_i, \theta)$$

and

$$u(x, \theta, d) = \sup_{|\tau-\theta| \leq d} |\psi(x, \tau) - \psi(x, \theta)|, \tag{2.2}$$

where $|\cdot|$ is taken to be the sup norm: $|\theta| = \max(|\theta_1|, \dots, |\theta_m|)$, and the expectations are taken at the underlying distributions of x_i's. We aim for a generally good error rate in the Bahadur representation of $\widehat{\theta}_n$, and work under the following set of conditions (not all conditions are needed for each theorem; for explanations of what roles each condition plays, see Remark 2.1):

(B1) For each fixed $\theta \in \Theta$, $\psi(x, \theta)$ is Borel measurable.

(B2) There exists $\theta_0 \in \Theta$ such that $\Lambda_n(\theta_0) = 0$ and $|\hat{\theta}_n - \theta_0| \to 0$ almost surely as $n \to \infty$.

(B3) There exist $r > 0$, $d_0 > 0$ and a sequence of positive numbers $\{a_i,\ i \geq 1\}$ such that $Eu^2(x_i, \theta, d) \leq a_i^2 d^r$ for $|\theta - \theta_0| \leq d_0$ and $d \leq d_0$.

(B4) $A_{2n} = O(A_n)$, where $A_n = \sum_{i=1}^{n} a_i^2$.

(B5) There exist $0 < \beta \le \alpha$ and $\beta_1 > 0$ such that

$$Eu^{2+\alpha}(x_i, \theta_0, d) \le a_i^{2+\beta_1} d^{(2+\beta)r/2} \quad \text{for } d \le d_0,$$
$$\sum_{i=1}^{n} a_i^{2+\beta_1} = O(A_n^{(2+\beta)/2} (\log n)^{-7-2\alpha}).$$

(B6) There exists a sequence of positive numbers $\{s_n,\ n \ge 1\}$ with $s_n \to \infty$ and $s_n \le A_n$ such that

$$\limsup_{n\to\infty} \frac{|\sum_{i=1}^{n} \psi(x_i, \theta_0)|}{(s_n \log\log n)^{1/2}} \le 2 \quad \text{a.s.}$$

(B7) $|\Lambda_n(\widehat{\theta}_n)| \ge c_n |\widehat{\theta}_n - \theta_0|$ for some positive numbers c_n.

(B8) There exist a nonsingular matrix D_n and positive numbers $\{b_n\}$ such that

$$|\Lambda_n(\widehat{\theta}_n) - D_n(\widehat{\theta}_n - \theta_0)| \le b_n \quad \text{a.s.}$$

THEOREM 2.1. *Under conditions* (B1)–(B7), *we have*

$$\widehat{\theta}_n - \theta_0 = O\left(\frac{s_n^{1/2}}{c_n} (\log\log n)^{1/2} \left(1 + \left(\frac{A_n}{s_n}\right)^{1/2} \left(\frac{A_n}{c_n^2} \log\log n\right)^{r/4}\right)\right) \tag{2.3}$$

for any sequence $\widehat{\theta}_n$ satisfying (2.1) *with* $\delta_n = O((s_n \log\log n)^{1/2})$. *If, in addition*, (B8) *is satisfied, then $\widehat{\theta}_n$ has the following almost sure representation*:

$$\widehat{\theta}_n - \theta_0 = -\sum_{i=1}^{n} D_n^{-1} \psi(x_i, \theta_0) + O(R_{n,1}) + O(R_{n,2}), \tag{2.4}$$

where

$$R_{n,1} = |D_n^{-1}| \, (A_n^{1/2} |\widehat{\theta}_n - \theta_0|^{r/2} + 1)(\log\log n)^{1/2}$$

and

$$R_{n,2} = |D_n^{-1}|(b_n + \delta_n).$$

The first remainder term $R_{n,1}$ comes from the linearization given in Lemma 4.1. Usually this dominates the error rate in the representation. The second remainder $R_{n,2}$ is due to other approximations. In typical applications, especially when each x_i has the same distribution, the following corollary would give a simpler remainder term.

COROLLARY 2.1. *Suppose that* (B1)–(B8) *are satisfied with* $n/c_n = O(1)$, $|D_n^{-1}| = O(n^{-1})$, $A_n = O(n)$ *and* $b_n = O(n^{1/2-r/4}(\log\log n)^{r/2})$. *Then, any sequence $\hat{\theta}_n$ satisfying* (2.1) *with* $\delta_n = O(n^{1/2-r/4})$ *has the following almost sure representation*:

$$\widehat{\theta}_n - \theta_0 = -D_n^{-1} \sum_{i=1}^{n} \psi(x_i, \theta_0) + O\big(n^{-(1/2+r/4)} (\log\log n)^{1/2+r/4}\big). \tag{2.5}$$

The results apply to any consistent estimators as assumed in (B2). In some cases, the verification of strong consistency could be highly nontrivial. General conditions on consistency are not discussed in the present paper, but we refer to Huber (1967), Haberman (1989) and Liese and Vajda (1994) among others. On the other hand, consistency is often the first step in the asymptotic analysis and can be found in existing literature for most estimators in use. We also note that, with some nonessential modifications in the proof of Lemma 4.1, the results of Theorem 2.1 remain valid if the strong consistency in (B2) is weakened to the following:

(B2′) There exists $\theta_0 \in \Theta$ such that $\Lambda_n(\theta_0) = 0$ and $\log(|\hat{\theta}_n - \theta_0|)/\log(n)$ is bounded almost surely.

In fact, (B2′) together with other conditions implies strong consistency.

REMARK 2.1. The Borel measurability of ψ in condition (B1) is to ensure measurability of all quantities used in Section 4. Huber (1967) assumed some form of separability for the same purpose. The major conditions of Theorem 2.1 are (B3) and (B5) as they are used to obtain a linear expansion of $\Lambda_n(\widehat{\theta}_n)$ in Lemma 4.1. Condition (B7) is used to ensure its invertibility in order to obtain the rate of convergence of $\widehat{\theta}_n$ in (2.3). Condition (B6) is usually a result of some known law of the iterated logarithm such as those in Wittmann (1987) and Chen (1993). It is not critical to the representation but simplifies the error rate expression in (2.4). We require the growth rate of A_n to be limited by (B4), but this is not essential (see Lemma 4.6). The quantity b_n in (B8) is usually obtainable by Taylor expansion and by (2.3).

The most demanding condition here is (B5). In fact, to get a looser error bound in the representation that is still strong enough to imply asymptotic normality, this condition is not needed at all.

COROLLARY 2.2. *Assume conditions* (B1), (B2) *and* (B3) *with* $A_n = \sum_{i=1}^n a_i^2 = O(n)$. *If, in a neighborhood of* θ_0, $\Lambda_n(\theta)$ *has a nonsingular derivative* $D_n(\theta)$ *such that* $|D_n^{-1}(\theta_0)| = O(1/n)$ *and* $|D_n(\theta) - D_n(\theta_0)| \le \kappa n|\theta - \theta_0|$ *for some constant* κ, *then for any sequence* $\widehat{\theta}_n$ *satisfying* (2.1) *with* $\delta_n = O(n^{1/2-r/4})$,

$$\widehat{\theta}_n - \theta_0 = -D_n^{-1}(\theta_0)\sum_{i=1}^n \psi(x_i, \theta_0) + O\big(n^{-(1/2+r/4)}(\log n)^3\big).$$

The result of Corollary 2.2 holds true even if (B3) is generalized to $Eu^2(x_i, \theta, d) \le a_i^2 d^r |\log d|$. Therefore, our conditions are generally weaker than those employed by Huber (1967) for asymptotic normality in the special case of i.i.d. samples. Corollary 2.2 is convenient for verifying asymptotic normality of M-estimators. The proof is similar to what is given in Section 4, and more details may be obtained in He and Shao (1994), an unpublished technical report of the authors. On the other hand, if ψ as a function of θ

satisfies a Hölder condition of some order, the verification of (B5) becomes straightforward with $\beta_1 = \beta$.

Although we are working with a consistent estimator $\widehat{\theta}_n$, the rate of convergence (2.3), which is typically $O((s_n \log\log n)^{1/2} c_n^{-1})$, follows from those conditions automatically. If, however, a rate of convergence is already known, an alternative condition (B5′) may be used in lieu of (B5), (B6) and (B7) [in some applications, (B5′) is much easier to verify]:

(B5′) For some decreasing sequence of positive numbers d_n such that $d_n = O(d_{2n}) = o(1)$, $\max_{1\le i\le n} u(x_i, \theta_0, d_n) = O(A_n^{1/2} d_n^{r/2} (\log n)^{-2})$ a.s.

THEOREM 2.2. *Under conditions* (B1)–(B4), (B5′) *and* (B8), *any sequence* $\widehat{\theta}_n$ *satisfying* (2.1) *and* $|\widehat{\theta}_n - \theta_0| \le d_n = o(1)$ *almost surely has the following representation*:

$$\widehat{\theta}_n - \theta_0 = -\sum_{i=1}^{n} D_n^{-1} \psi(x_i, \theta_0) + O(R_n^*) \quad a.s., \tag{2.6}$$

where $R_n^* = |D_n^{-1}|(\delta_n + b_n + A_n^{1/2} d_n^{r/2} (\log\log n)^{1/2} + (\log\log n)^{1/2})$.

Some other generalizations of Theorem 2.1 are given briefly at the end of Section 4. However, we conclude this section with a quick comparison with the results of Niemiro (1992), who also obtained strong Bahadur representations of M-estimators.

Niemiro (1992) considers a class of M-estimators defined by minimization of a convex objective function. The convexity leads to some simplifications, but excludes redescending M-estimators often used in robust statistics. The smoothness condition used in Niemiro (1992) is similar in nature to (B3) used in the present paper. It is sometimes easier to check than our condition (B5), but the resulting error rates are not as sharp. Finally, the assumption of i.i.d. observations in Niemiro (1992) is not appropriate for applications in the regression problems with non-stochastic designs.

3. Applications to linear models. One important application of Theorems 2.1 and 2.2 is in the area of linear models. We consider the usual regression model

$$y_i = z_i'\theta + e_i,$$

where the z_i's are nonstochastic design points in R^m, and the e_i's are independent error variables with common probability density function f. Without loss of generality, we assume that the true parameter $\theta_0 = 0$. We now consider three different classes of M-estimators that are commonly used in the literature. As in Section 2, the consistency of these estimators is not explicitly discussed here but can be found in the existing literature.

3.1. *M-estimators with smooth score functions.* We first consider the simplest case, where $\widehat{\theta}_n$ is defined through

$$\sum_{i=1}^{n} \phi(y_i - z_i'\theta) z_i = 0, \tag{3.1}$$

where ϕ is Lipschitz. Results on consistency and asymptotic normality were given in Yohai and Maronna (1979). To illustrate the application of Theorem 2.1, we shall not seek the weakest possible assumptions at the cost of clarity.

Let $Q_n = \sum_{i=1}^n z_i z_i'$. To use Theorem 2.1, we identify $x_i = (y_i, z_i)$ and $\psi(x_i, \theta) = \phi(y_i - z_i'\theta) z_i$. Note that part of x_i has a degenerate point-mass distribution, but this is allowed in Section 2.

THEOREM 3.1. *If the following conditions* (C1)–(C3) *are satisfied, then*

$$\widehat{\theta}_n = -(\gamma Q_n)^{-1} \sum_{i=1}^{n} \phi(e_i) z_i + O\left(\frac{\log\log n}{n}\right) \quad a.s.:$$

(C1) *both* ϕ *and* f' *are Lipschitz*;

(C2) $E\phi(e) = 0$, $\gamma = \int_{-\infty}^{\infty} \phi(x) f'(x)\, dx \neq 0$ *and* $E\phi^{2+\varepsilon}(e) < \infty$ *for some* $\varepsilon > 0$;

(C3) $n^{-1} Q_n \to Q$ *for some positive definite matrix* Q *and* $\sum_{i=1}^n |z_i|^{4+\varepsilon} = O(n)$ *for some* $\varepsilon > 0$.

PROOF OF THEOREM 3.1. Since f' is Lipschitz and $\sum_{i=1}^n |z_i|^3 = O(n)$, the derivative of $\Lambda_n(\theta)$ equals $\sum_{i=1}^n \int_{-\infty}^{\infty} \phi(u) f'(u + z_i'\theta)\, du\, z_i z_i'$ and is Lipschitz in θ. It then follows that, for sufficiently large n, $|\Lambda_n(\widehat{\theta}_n)| \geq (1/2)\gamma n \lambda_{\min}(Q)|\widehat{\theta}_n|$, where $\lambda_{\min}(Q)$ is the smallest eigenvalue of Q, and (B7) is satisfied with $c_n = (1/2)\gamma n \lambda_{\min}(Q)$. Similarly, (B8) holds with $D_n = \gamma Q_n$, $|D_n^{-1}| = O(n^{-1})$ and $b_n = O(n)$. By the standard LIL, (B6) is satisfied with $s_n = O(\sum_{i=1}^n |z_i|^2) = O(n)$. Furthermore, because ϕ is Lipschitz, (B5) holds with $A_n = O(\sum_{i=1}^n |z_i|^4) = O(n)$ and $r = 2$. The result follows from Corollary 2.1. $\square$

3.2. *Minimum L_p distance estimators.* The minimum L_p distance estimators obtained through minimization of $\sum_{i=1}^n |y_i - z_i'\theta|^p$ for some $p \geq 1$ are automatically scale equivariant. The convexity of the objective function ensures the existence of a solution. Results on consistency and asymptotic normality of the L_p regression estimators are covered in Bai, Rao and Wu (1992), and the Bahadur representation with the exact error rate was obtained by Arcones (1996a) in the case of random designs. Under appropriate design conditions, our Theorem 2.1 gives the same exact second-order error rate, but for the more general case of deterministic designs. We first consider the case of $p > 1$, where $\widehat{\theta}_n$ solves (3.1) with

$$\phi(e) = |e|^{p-1} \operatorname{sgn}(e) \tag{3.2}$$

and the case of special interest with $p = 1$ is considered in Section 3.3.

THEOREM 3.2. *If $1 < p \le 2$ and $p \ne 3/2$, the minimum L_p distance estimator satisfies*

$$\hat{\theta}_n = \{(p-1)E|e|^{p-2}Q_n\}^{-1}\sum_{i=1}^n \phi(e_i)z_i + O\big((\log\log n/n)^{\min\{1,\,1/4+p/2\}}\big) \quad a.s., \tag{3.3}$$

provided that (C3) *and the following conditions* (C4) *and* (C5) *are satisfied*:

(C4) $\sum_{i=1}^n E(|e-z_i'\theta|^p - |e|^p)$ *has a unique minimum at* $\theta = 0$;
(C5) f *is bounded and* $E|e-t|^{p-2}$ *is Lipschitz in* t.

REMARK 3.1. It can be seen from the proof that the second part of condition (C3) can be weakened to $\sum_{i=1}^n |z_i|^{2p+1+\varepsilon} = O(n)$ for $1 < p < 3/2$. The representation (3.3) also holds for any $p > 2$ if we have $\sum_{i=1}^n |z_i|^{2p+\varepsilon} = O(n)$.

REMARK 3.2. The special case of $p = 3/2$ is not directly covered by Theorem 2.1, but we have $Eu^2(x_i, \theta, d) \le a_i^2 d^2 |\log d|$ for (B3). The proof of Theorem 2.1 can be slightly modified to yield the same representation as (3.3) with the remainder in the order of $(\log n)^{1/2}(\log\log n/n)$ (see Lemma 4.5). The fact that $p = 3/2$ serves as a break point in the second-order error rate has been observed by earlier authors in some restricted settings [see Niemiro (1992) and Arcones (1996a)].

3.3. *The least absolute deviation regression.* The minimum L_1 distance estimator is probably the most widely used regression estimator outside the least squares universe [see Bloomfield and Steiger (1983)]. It is also of special interest here, as it does not satisfy the requirement (B5) with $r = 1$. Instead, we can verify the alternative condition (B5′) to have

THEOREM 3.3. *The least absolute deviation* (*LAD*) *estimator satisfies*

$$\hat{\theta}_n = \{2f(0)Q_n\}^{-1}\sum_{i=1}^n \operatorname{sgn}(e_i)z_i + O\left(\left(\sum_{i=1}^n \frac{|z_i|^3}{n}\right)^{1/2}\left(\frac{\log\log n}{n}\right)^{3/4}\right) \quad a.s., \tag{3.4}$$

provided that (i) e *has zero median and a bounded density function* f *with* $f(0) > 0$, (ii) $n^{-1}Q_n \to Q$ *for a nonsingular matrix* Q, (iii) $\max_{i\le n}|z_i| = O(n^{1/4}(\log n)^{-2})$ *and* (iv) $f(y) - f(0) = O(|y|^{1/2})$ *as* $y \to 0$.

Note that the LAD estimator does not necessarily solve (3.1) with $\phi(e) = \operatorname{sgn}(e)$. However, condition (iii) ensures that (2.1) holds with $\delta_n = n^{1/4}$.

Theorem 3.3 does not assume differentiability of f at zero. Under a weaker condition on $\max_{i\le n}|z_i|$, Babu (1989) also obtained a strong representation for the LAD estimator, but the remainder term there is suboptimal. A more careful study indicates that the order of the remainder term depends not only on the magnitude of the design points, but also on the behavior of f near 0.

REMARK 3.3. We have assumed here and in condition (C3) that the matrix Q_n is on the order of n. This corresponds to the root-n consistency for each component of $\widehat{\theta}_n$. In this case, our error bound in (3.4) is known to be optimal. In more general designs (which are allowed in Section 2), the components of $\widehat{\theta}_n$ may converge at different rates. A representation can be derived directly from Theorem 2.1 for $Q_n^{1/2}\widehat{\theta}_n$ as in Babu (1989) and Pollard (1991), the error bound, however, may not be optimal unless the transformations from θ to $Q_n^{1/2}\theta$ and z_i to $Q_n^{-1/2}z_i$ are used. It is not clear, however, whether and how Theorem 2.1 should be adjusted at a general level (as in Section 2) to yield the best possible error rates. The sharp error rate for the least absolute deviation regression estimator with more general designs (in terms of Q_n) has recently been obtained by Arcones (1996b), and we shall not pursue it further in the present paper.

REMARK 3.4. With some routine modifications, a representation similar to (3.4) can be derived for regression quantiles of Koenker and Bassett (1978). We omit the details.

3.4. *Generalized M-estimators.* An M-estimator defined via a minimization of $\sum_{i=1}^n \rho(y_i - z_i'\theta)$ for some convex objective function ρ may become less sensitive to outlying response values, but can still be unduly influenced by leverage points. This is reflected partially in the use of design conditions in Theorems 3.1–3.3. Hampel, Ronchetti, Rousseeuw and Stahel (1986) studied the asymptotic behavior of the generalized M-estimators which solve

$$\sum_{i=1}^n \phi(y_i - z_i'\theta)z_i w(z_i) = 0$$

for some weight function w. If $\sup_z |z|^2 w(z) < \infty$, $E|\phi(e)|^{2+\varepsilon} < \infty$ for some $\varepsilon > 0$ and ϕ is Lipschitz as in Section 3.1, then, for *any* design, conditions (C1) and (C2) imply that any consistent sequence of the GM-estimator admits the following representation:

$$\widehat{\theta}_n = -\left\{\gamma \sum_{i=1}^n w(z_i)z_i z_i'\right\}^{-1} \sum_{i=1}^n \phi(e_i)z_i w(z_i) + O\left(\frac{\log\log n}{n}\right) \quad \text{a.s.}$$

4. Proofs and generalizations of Theorems 2.1 and 2.2. The basic idea of our proof takes root in Huber (1967), but a refinement of Huber's Lemma 3 is essential for better control of error rates.

Let $\{x_i,\ i \geq 1\}$ be independent random vectors. Define $\lambda_i(\theta) = E\psi(x_i, \theta)$ and

$$Z_n(\tau, \theta) = \left|\sum_{i=1}^n \big(\psi(x_i, \tau) - \psi(x_i, \theta) - \lambda_i(\tau) + \lambda_i(\theta)\big)\right|. \tag{4.1}$$

The following plays the central role in the proof of Theorem 2.1. For convenience, in the proofs of this and next section, we shall use a generic constant

K which may vary from line to line. It may depend on the fixed quantities such as m and ψ, but not on the sample size n.

LEMMA 4.1. *Assume conditions* (B1) *and* (B3)–(B5). *We have*

$$\limsup_{n\to\infty} \sup_{|\tau-\theta_0|\le d_0} \frac{Z_n(\tau,\theta_0)}{(A_n|\tau-\theta_0|^r+1)^{1/2}(\log\log n)^{1/2}} \le C \quad a.s., \tag{4.2}$$

for some constant $C < \infty$.

PROOF. Observe that, for any random vectors $\{Y_n = (Y_{n,1},\dots,Y_{n,m})\}_{n=1}^{\infty}$ of dimension m,

$$P(|Y_n| \ge C,\ \text{infinitely often}) \le \sum_{i=1}^{m} P(|Y_{n,i}| \ge C,\ \text{infinitely often}).$$

Therefore, we can assume, without loss of generality, that ψ is real-valued, $\theta_0 = 0$ and $d_0 = 1$. For simplicity, write

$$\eta_i(\tau) = \psi(x_i,\tau) - \psi(x_i,0) - \lambda_i(\tau) + \lambda_i(0), \qquad Z_n(\tau) := Z_n(\tau,0). \tag{4.3}$$

Since $A_{2n} = O(A_n)$, there exists $K_1 > 1$ such that

$$A_{2n} \le K_1 A_n \quad \text{and} \quad A_n \le n^{K_1} \quad \text{for } n = 2,3,\dots. \tag{4.4}$$

It is easy to see that

$$\begin{aligned}
&\limsup_{n\to\infty} \sup_{|\tau|\le 1} \frac{Z_n(\tau)}{(A_n|\tau|^r+1)^{1/2}(\log\log n)^{1/2}} \\
&\quad\le \limsup_{k\to\infty} \max_{2^{k-1}\le n\le 2^k} \sup_{|\tau|\le 1} \frac{Z_n(\tau)}{(A_n|\tau|^r+1)^{1/2}(\log\log n)^{1/2}} \\
&\quad\le \limsup_{k\to\infty} \max_{n\le 2^k} \sup_{|\tau|\le 1} \frac{Z_n(\tau)}{(A_{2^{k-1}}|\tau|^r+1)^{1/2}(\log\log 2^{k-1})^{1/2}} \\
&\quad\le K_1 \limsup_{k\to\infty} \max_{n\le 2^k} \sup_{|\tau|\le 1} \frac{Z_n(\tau)}{(A_{2^k}|\tau|^r+1)^{1/2}\log^{1/2} k}.
\end{aligned} \tag{4.5}$$

Let

$$\mathscr{B}_k = \left\{ \max_{n\le 2^k} \sup_{|\tau|\le 1} \frac{|\eta_n(\tau)|}{(A_{2^k}|\tau|^r+1)^{1/2}\, k^{-2}} \le 1 \right\}. \tag{4.6}$$

By Lemma 4.2, $P(\mathscr{B}_k^c,\ \text{infinitely often}) = 0$. It suffices to show that

$$\sum_{k\ge 1} P(\mathscr{O}_k \cap \mathscr{B}_k) < \infty, \tag{4.7}$$

where

$$\mathscr{O}_k = \left\{ \max_{n\le 2^k} \sup_{|\tau|\le 1} \frac{Z_n(\tau)}{\left(A_{2^k}|\tau|^r+1\right)^{1/2}(\log k)^{1/2}} \ge 3C \right\}, \tag{4.8}$$

for some constant C.

To prove (4.7), we decompose the cube $\{\tau\colon |\tau| \le 1\}$ in the definition of $\mathscr{O}_k$ into smaller cubes. The Freedman exponential inequality [Freedman (1975)] is evoked to bound the probabilities on each subcube followed by a chaining argument to extend the bound to the probability in (4.7).

Before we proceed, we state the following variant of the Freedman inequality for easy reference.

FREEDMAN INEQUALITY. If $\{\zeta_i,\ 1 \le i \le n\}$ are independent random variables with $E\zeta_i \le 0$ and $\zeta_i \le a$ for each $1 \le i \le n$, then, for any $x > 0$,

$$P\biggl(\max_{1\le i\le n} \sum_{j=1}^{i} \zeta_j \ge x\biggr) \le \exp\biggl(-\frac{x^2}{2(a\,x + \sum_{i=1}^n E\zeta_i^2)}\biggr)$$
$$\le \exp\biggl(-\frac{x}{4\,a}\biggr) + \exp\biggl(-\frac{x^2}{4\sum_{i=1}^n E\zeta_i^2}\biggr).$$

Let

$$\delta := \delta_k = 1/2^{[8kK_1/r]}, \qquad M := M_k = 1/\delta = 2^{[8kK_1/r]}, \tag{4.9}$$

where K_1 is defined as in (4.4), and $[x]$ denotes the integer part of x. We use $:=$ in the proofs to match the symbols on both sides in an obvious way (in order to simplify writing). Consider the concentric cubes

$$\mathscr{C}_l = \{\tau\colon |\tau| \le l\delta\}, \qquad l = 1, 2, \dots, M. \tag{4.10}$$

Subdivide the difference $\mathscr{C}_{l+1} \setminus \mathscr{C}_l$ into smaller cubes with edges of length δ. For each value of l there are $m_l = (2(l+1))^m - (2l)^m$ such small cubes, which are denoted by $\mathscr{C}_l^j$, $j = 1, 2, \dots, m_l$. Let c_l^j be the center of $\mathscr{C}_l^j$. Then we have $|c_l^j| = (l + 1/2)\delta$ and, for $x > 1$,

$$\begin{aligned}
&\biggl\{\max_{n\le 2^k} \sup_{|\tau|\le 1} \frac{Z_n(\tau)}{(A_{2^k}|\tau|^r + 1)^{1/2}} \ge 3x\biggr\}\\
&\quad\subset \biggl\{\sup_{\tau\in\mathscr{C}_1} \max_{n\le 2^k} \frac{Z_n(\tau)}{(A_{2^k}|\tau|^r + 1)^{1/2}} \ge 3x\biggr\}\\
&\qquad\cup \bigcup_{1\le l<M} \biggl\{\sup_{\tau\in\mathscr{C}_{l+1}\setminus\mathscr{C}_l} \max_{n\le 2^k} \frac{Z_n(\tau)}{(A_{2^k}|\tau|^r + 1)^{1/2}} \ge 3x\biggr\}\\
&\quad\subset \biggl\{\sum_{i=1}^{2^k} \sup_{|\tau|\le\delta} |\eta_i(\tau)| \ge 3x\biggr\} \cup \bigcup_{1\le l<M} \bigcup_{j\le m_l} \biggl\{\sup_{\tau\in\mathscr{C}_l^j} \max_{n\le 2^k} \frac{Z_n(\tau)}{(A_{2^k}(l\delta)^r + 1)^{1/2}} \ge 3x\biggr\}\\
&\quad\subset \biggl\{\sum_{i=1}^{2^k} \sup_{|\tau|\le\delta} |\eta_i(\tau)| \ge 1\biggr\} \cup \bigcup_{1\le l<M} \bigcup_{j\le m_l} \biggl\{\sum_{i=1}^{2^k} \sup_{\tau\in\mathscr{C}_l^j} \frac{|\eta_i(\tau) - \eta_i(c_l^j)|}{(A_{2^k}(l\delta)^r + 1)^{1/2}} \ge 2x\biggr\}\\
&\qquad\cup \bigcup_{1\le l<M} \bigcup_{j\le m_l} \biggl\{\max_{n\le 2^k} \frac{Z_n(c_l^j)}{(A_{2^k}(l\delta)^r + 1)^{1/2}} \ge x\biggr\}.
\end{aligned}$$

Therefore

$$
\begin{aligned}
P(\mathscr{O}_k \cap \mathscr{B}_k) &\le P\biggl(\sum_{i=1}^{2^k} \sup_{|\tau|\le\delta} |\eta_i(\tau)| \ge 1\biggr) \\
&\quad + \sum_{1\le l<M} \sum_{j\le m_l} P\biggl(\sum_{i=1}^{2^k} \sup_{\tau\in\mathscr{C}_l^j} \frac{|\eta_i(\tau)-\eta_i(c_l^j)|}{(A_{2^k}(l\delta)^r+1)^{1/2}} \ge 2C(\log k)^{1/2}, \mathscr{B}_k\biggr) \qquad (4.11)\\
&\quad + P\biggl(\bigcup_{1\le l<M} \bigcup_{j\le m_l} \biggl\{\max_{n\le 2^k} \frac{Z_n(c_l^j)}{(A_{2^k}(l\delta)^r+1)^{1/2}} \ge C(\log k)^{1/2}, \mathscr{B}_k\biggr\}\biggr) \\
&= P_k^{(1)} + P_k^{(2)} + P_k^{(3)},
\end{aligned}
$$

where $P_k^{(1)}$, $P_k^{(2)}$ and $P_k^{(3)}$ denote the three terms on the right-hand side of (4.11), respectively. It remains to show that each $P_k^{(i)}$, $i = 1, 2, 3$ is summable over k.

From (B3), (4.4) and (4.9) it follows that

$$
\begin{aligned}
\sum_{i=1}^{2^k} E \sup_{|\tau|\le\delta} |\eta_i(\tau)| &\le \sum_{i=1}^{2^k} 2(Eu^2(x_i, 0, \delta))^{1/2} \\
&\le 2^{1+k/2}\biggl(\sum_{i=1}^{2^k} Eu^2(x_i, 0, \delta)\biggr)^{1/2} \\
&\le 2^k (A_{2^k}\,\delta^r)^{1/2} \le 2^{-k}.
\end{aligned}
$$

Similarly,

$$
\sum_{i=1}^{2^k} E \sup_{\tau\in\mathscr{C}_l^j} |\eta_i(\tau) - \eta_i(c_l^j)| \le 2^{-k}. \qquad (4.12)
$$

By Chebyshev's inequality,

$$
\sum_{k\ge 1} P_k^{(1)} \le \sum_{k\ge 1} \sum_{i=1}^{2^k} E \sup_{|\tau|\le\delta} |\eta_i(\tau)| \le \sum_{k\ge 1} 2^{-k} < \infty.
$$

To estimate $P_k^{(2)}$, we shall use for convenience

$$
\pi_i := \pi_{i,l,j} = \sup_{\tau\in\mathscr{C}_l^j} |\eta_i(\tau) - \eta_i(c_l^j)|, \qquad G_{k,l} = (A_{2^k}(l\delta)^r + 1)^{1/2}.
$$

Note that $\mathscr{B}_k$ implies $\pi_i \le 2^{1+r} G_{k,l}\, k^{-2}$. By (4.12) and the Freedman inequality, we have

$$
\begin{aligned}
&P\biggl(\sum_{i=1}^{2^k} \sup_{\tau \in \mathscr{C}_l^j} \frac{|\eta_i(\tau) - \eta_i(c_l^j)|}{(A_{2^k}\,(l\delta)^r + 1)^{1/2}} \ge 2C(\log k)^{1/2}, \mathscr{B}_k \biggr) \\
&\quad \le P\biggl(\sum_{i=1}^{2^k} \pi_i I_{\{\pi_i \le 2^{1+r} G_{k,l} k^{-2}\}} \ge 2CG_{k,l}\,(\log k)^{1/2} \biggr) \\
&\quad \le P\biggl(\sum_{i=1}^{2^k} \bigl\{ \pi_i I_{\{\pi_i \le 2^{1+r} G_{k,l} k^{-2}\}} - E\pi_i I_{\{\pi_i \le 2^{1+r} G_{k,l} k^{-2}\}} \bigr\} \ge CG_{k,l}\,(\log k)^{1/2} \biggr) \\
&\quad \le \exp\biggl(-\frac{CG_{k,l}(\log k)^{1/2}}{2^{3+r} G_{k,l} k^{-2}} \biggr) + \exp\biggl(-\frac{(CG_{k,l}(\log k)^{1/2})^2}{4\sum_{i=1}^{2^k} E\pi_i^2 I_{\{\pi_i \le 2^{1+r} G_{k,l} k^{-2}\}}} \biggr) \qquad (4.13)\\
&\quad \le \exp\biggl(-\frac{Ck^2(\log k)^{1/2}}{2^{3+r}} \biggr) + \exp\biggl(-\frac{(CG_{k,l}(\log k)^{1/2})^2}{2^{3+r} G_{k,l} k^{-2} \sum_{i=1}^{2^k} E\pi_i} \biggr) \\
&\quad \le \exp\biggl(-\frac{Ck^2(\log k)^{1/2}}{2^{3+r}} \biggr) + \exp\biggl(-\frac{C^2 G_{k,l} \log k}{2^{3+r} k^{-2}} \biggr) \\
&\quad \le K \exp\biggl(-\frac{Ck^2(\log k)^{1/2}}{2^{3+r}} \biggr).
\end{aligned}
$$

Also, by the choice of $M = M_k$, we have, for any $C > 1$,

$$
\begin{aligned}
\sum_{k \ge 1} P_k^{(2)} &\le K \sum_{k\ge 1} \sum_{1 \le l < M} \sum_{j \le m_l} \exp\biggl(-\frac{Ck^2(\log k)^{1/2}}{2^{3+r}} \biggr) \\
&\le K \sum_{k \ge 1} (2M)^m \exp\biggl(-\frac{Ck^2(\log k)^{1/2}}{16} \biggr) < \infty.
\end{aligned} \tag{4.14}
$$

By Lemma 4.3, there exists $C < \infty$ such that $P_k^{(3)}$ is also summable over k. The proof of Lemma 4.1 is complete. $\square$

We shall now prove Lemmas 4.2 and 4.3. The former is rather straightforward, and the latter is much more technical but based on similar ideas used above.

LEMMA 4.2. *Under the assumptions of Lemma* 4.1, *we have*

$$
\sum_{k \ge 1} P\biggl(\max_{n \le 2^k} \sup_{|\tau| \le 1} \frac{|\eta_n(\tau)|}{(A_{2^k} |\tau|^r + 1)^{1/2}\, k^{-2}} \ge 1 \biggr) < \infty.
$$

PROOF. Using the same notation $\mathscr{B}_k$ as in (4.6), we have

$$
P(\mathscr{B}_k^c) \le \sum_{n \le 2^k} P\biggl(\sup_{|\tau| \le 1} \frac{|\eta_n(\tau)|}{(A_{2^k} |\tau|^r + 1)^{1/2}\, k^{-2}} \ge 1 \biggr) := \sum_{n \le 2^k} p_{n,k}.
$$

Let $I_k = \{l\colon -(\log_2 A_{2^k})/r \le l \le 0\}$ be an index set. From (B5) it follows that

$$
\begin{aligned}
p_{n,k} &\le P\biggl(\sup_{|\tau|\le A_{2^k}^{-1/r}} |\eta_n(\tau)| \ge k^{-2}\biggr) + P\biggl(\sup_{A_{2^k}^{-1/r}\le|\tau|\le 1} \frac{|\eta_n(\tau)|}{(A_{2^k}|\tau|^r)^{1/2}} \ge k^{-2}\biggr)\\
&\le k^{4+2\alpha} E \sup_{|\tau|\le A_{2^k}^{-1/r}} |\eta_n(\tau)|^{2+\alpha} + \sum_{l\in I_k} P\biggl(\sup_{2^{l-1}\le|\tau|\le 2^l} |\eta_n(\tau)| \ge k^{-2}(A_{2^k}\, 2^{(l-1)r})^{1/2}\biggr)\\
&\le 8k^{4+2\alpha}\biggl\{Eu^{2+\alpha}(x_n, 0, A_{2^k}^{-1/r}) + \sum_{l\in I_k} (A_{2^k}\, 2^{(l-1)r})^{-(2+\alpha)/2} Eu^{2+\alpha}(x_n, 0, 2^l)\biggr\}\\
&\le 8k^{4+2\alpha}\biggl\{a_n^{2+\beta_1}\, A_{2^k}^{-(2+\beta)/2} + \sum_{l\in I_k} (A_{2^k} 2^{(l-1)r})^{-(2+\alpha)/2} a_n^{2+\beta_1}\, 2^{lr(2+\beta)/2}\biggr\}\\
&\le K k^{4+2\alpha} a_n^{2+\beta_1}\, A_{2^k}^{-(2+\beta)/2}\biggl\{1 + A_{2^k}^{-(\alpha-\beta)/2} \sum_{l\in I_k} 2^{lr(\beta-\alpha)/2}\biggr\}\\
&\le K k^{4+2\alpha} a_n^{2+\beta_1} A_{2^k}^{-(2+\beta)/2}\{1+k\}\\
&\le K\, k^{5+2\alpha}\, A_{2^k}^{-(2+\beta)/2}\, a_n^{2+\beta_1}.
\end{aligned}
$$

Therefore, $P(\mathscr{B}_k^c) \le K k^{-2}$, from which the lemma follows immediately. □

LEMMA 4.3. *Under the assumptions of Lemma* 4.1, $\sum_k P_k^{(3)} < \infty$ *for some constant* C, *where* $P_k^{(3)}$ *is defined in the proof of Lemma* 4.1.

PROOF. We shall continue to use the partition and the associated notation used around (4.10). Furthermore, let

$$N = (\log_2 M) - 1 = [8kK_1/r] - 1 \quad \text{and} \quad \mathscr{Q}_u = \bigl\{c_l^j\colon j \le m_l,\ l < 2^{u+1}\bigr\}$$

for $u = 0, 1, \dots, N$. Note that the subscript k for N is suppressed in our notation here and that $\mathscr{Q}_u$ is actually the set of centers of the small cubes with edges of length δ in $\mathscr{C}_{2^{u+1}}$. It is easy to see that

$$
\begin{aligned}
&\bigcup_{1\le l<M} \bigcup_{j\le m_l} \biggl\{\max_{n\le 2^k} \frac{Z_n(c_l^j)}{(A_{2^k}(l\delta)^r + 1)^{1/2}} \ge C\log^{1/2} k\biggr\}\\
&\quad\subset \bigcup_{0\le u\le N} \bigcup_{2^u\le l<2^{u+1}} \bigcup_{j\le m_l} \biggl\{\max_{n\le 2^k} \frac{Z_n(c_l^j)}{(A_{2^k}(2^u\delta)^r + 1)^{1/2}} \ge C\log^{1/2} k\biggr\} \qquad (4.15)\\
&\quad\subset \bigcup_{0\le u\le N} \biggl\{\max_{\tau\in\mathscr{Q}_u} \max_{n\le 2^k} \frac{Z_n(\tau)}{(A_{2^k}(2^u\delta)^r + 1)^{1/2}} \ge C\log^{1/2} k\biggr\}.
\end{aligned}
$$

For $0 \le p \le u+1$, subdivide $\mathscr{C}_{2^{u+1}}$ into smaller cubes with edges of length $2^p\delta$. There are $m_{u,p} := (2^{u+2-p})^m$ such subcubes, which will be called $\mathscr{C}_{u,p}^j$, whose centers are denoted by $c_{u,p}^j$, $j = 1, 2, \dots, m_{u,p}$. Let

$$\mathscr{Q}_{u,p} = \bigl\{c_{u,p}^j\colon j = 1, 2, \dots, m_{u,p}\bigr\}.$$

Please note the difference between $\mathscr{Q}_u$ and the $\mathscr{Q}_{u,p}$'s. For any $\tau \in \mathscr{Q}_u$ and for each $0 \le p \le u+1$, there is a unique $\mathscr{C}_{u,p}^{j_{p,\tau}}$ which contains τ. When $p = 0$, we have $\tau = c_0^{j_{0,\tau}}$ for some $j_{0,\tau}$. Moreover, $\mathscr{C}_{u,p}^{j_{p,\tau}}$ is increasing in p, and

$$\left|c_{u,p}^{j_{p,\tau}} - c_{u,p+1}^{j_{p+1,\tau}}\right| = 2^{p-1}\delta. \tag{4.16}$$

Hence,

$$Z_n(\tau) = Z_n\big(c_{u,u+1}^{j_{u+1,\tau}}\big) + \sum_{0\le p\le u}\big(Z_n\big(c_{u,p}^{j_{p,\tau}}\big) - Z_n\big(c_{u,p+1}^{j_{p+1,\tau}}\big)\big),$$

and

$$\begin{aligned}
\max_{\tau\in\mathscr{Q}_u} Z_n(\tau) &\le \max_{\tau\in\mathscr{Q}_u} Z_n\big(c_{u,u+1}^{j_{u+1,\tau}}\big) + \sum_{0\le p\le u}\max_{\tau\in\mathscr{Q}_u}\big|Z_n\big(c_{u,p}^{j_{p,\tau}}\big) - Z_n\big(c_{u,p+1}^{j_{p+1,\tau}}\big)\big| \\
&\le \max_{\tau\in\mathscr{Q}_{u,u+1}} Z_n(\tau) + \sum_{0\le p\le u}\max_{\tau\in\mathscr{Q}_{u,p}}\big|Z_n\big(c_{u,p}^{j_{p,\tau}}\big) - Z_n\big(c_{u,p+1}^{j_{p+1,\tau}}\big)\big| \qquad (4.17)\\
&\le \max_{\tau\in\mathscr{Q}_{u,u+1}} Z_n(\tau) + \sum_{0\le p\le u}\max_{\tau\in\mathscr{Q}_{u,p}}\left|\sum_{i=1}^n \eta_i\big(c_{u,p}^{j_{p,\tau}}\big) - \eta_i\big(c_{u,p+1}^{j_{p+1,\tau}}\big)\right|.
\end{aligned}$$

For further notational convenience, let

$$\chi_{k,u,p} = k^{-1}(A_{2^k}(2^u\delta)^r + 1)^{1/2} + (u+2-p)\big(A_{2^k}(2^p\delta)^r\big)^{1/2}.$$

It is easy to show that

$$\sum_{0\le p\le u+1}\chi_{k,u,p} \le K(A_{2^k}(2^u\delta)^r + 1)^{1/2}.$$

Thus,

$$\begin{aligned}
&\left\{\max_{\tau\in\mathscr{Q}_u}\max_{n\le 2^k}\frac{Z_n(\tau)}{\big(A_{2^k}(2^u\delta)^r + 1\big)^{1/2}} \ge C\log^{1/2} k\right\} \\
&\quad\subset \bigcup_{\tau\in\mathscr{Q}_{u,u+1}}\left\{\max_{n\le 2^k} Z_n(\tau) \ge C^{1/2}\chi_{k,u,u+1}\log^{1/2} k\right\} \qquad (4.18)\\
&\quad\cup\bigcup_{0\le p\le u}\bigcup_{\tau\in\mathscr{Q}_{u,p}}\left\{\max_{n\le 2^k}\left|\sum_{i=1}^n \eta_i\big(c_{u,p}^{j_{p,\tau}}\big) - \eta_i\big(c_{u,p+1}^{j_{p+1,\tau}}\big)\right| \ge C^{1/2}\chi_{k,u,p}\log^{1/2} k\right\},
\end{aligned}$$

provided that C is sufficiently large.

For $0 \le p \le u$ and $\tau \in \mathscr{Q}_{u,p}$, the event $\mathscr{B}_k$ implies that

$$\big|\eta_i\big(c_{u,p}^{j_{p,\tau}}\big) - \eta_i\big(c_{u,p+1}^{j_{p+1,\tau}}\big)\big| \le 4^{1+r}k^{-2}(A_{2^k}(2^u\delta)^r + 1)^{1/2}.$$

Similar to the proof of (4.13), we have, by the Freedman inequality,

$$\begin{aligned}
&P\left(\max_{n\le 2^k}\left|\sum_{i=1}^n \eta_i\big(c_{u,p}^{j_{p,\tau}}\big) - \eta_i\big(c_{u,p+1}^{j_{p+1,\tau}}\big)\right| \ge C^{1/2}\chi_{k,u,p}\log^{1/2} k,\ \mathscr{B}_k\right) \qquad (4.19)\\
&\quad\le 2\exp\left(-\frac{C^{1/2}k\log^{1/2} k}{4^{2+r}}\right) + 2\exp\left(-\frac{C(u+2-p)^2\log k}{16}\right).
\end{aligned}$$

Similarly, we obtain, for $\tau \in \mathscr{Q}_{u,u+1}$,

$$P\Big(\max_{n\le 2^k} Z_n(\tau) \ge C^{1/2}\, c_{k,u,u+1}\log^{1/2} k,\, \mathscr{B}_k\Big) \tag{4.20}$$
$$\le K\exp\Big(-\frac{C^{1/2}\, k\log^{1/2} k}{4^{2+r}}\Big) + K\exp\Big(-\frac{C\log k}{16}\Big),$$

for some constant K. We conclude from (4.15) and (4.18)–(4.20) that

$$\begin{aligned} P_k^{(3)} \le K\sum_{0\le u\le N}\Big\{&\exp\Big(-\frac{C^{1/2}\, k\log^{1/2} k}{4^{2+r}}\Big) + \exp\Big(-\frac{C\log k}{16}\Big) \\ &+ \sum_{0\le p\le u}(2^{u+2-p})^m\Big\{\exp\Big(-\frac{C^{1/2}\, k\log^{1/2} k}{16}\Big) \\ &\qquad + \exp\Big(-\frac{C(u+2-p)^2\log k}{4^{2+r}}\Big)\Big\}\Big\} \\ \le Kk^{-2},& \end{aligned}$$

for sufficiently large k. Therefore $P_k^{(3)}$ is summable over k. □

Theorem 2.1 follows directly from the following lemma.

LEMMA 4.4. *Under conditions* (B1)–(B7) *we have, for any* $\widehat{\theta}_n$ *satisfying* (2.1) *with* $\delta_n = O((s_n\log\log n)^{1/2})$,

$$\widehat{\theta}_n - \theta_0 = O\big((s_n^{1/2}c_n^{-1} + A_n^{1/2+r/4}c_n^{-1-r/2}(\log\log n)^{r/4})(\log\log n)^{1/2}\big) \tag{4.21}$$

and

$$\sum_{i=1}^n \psi(x_i,\theta_0) + \sum_{i=1}^n \lambda_i(\widehat{\theta}_n) = O(R_n) \quad a.s., \tag{4.22}$$

where $R_n = A_n^{1/2}|\widehat{\theta}_n - \theta_0|^{r/2}(\log\log n)^{1/2} + (\log\log n)^{1/2} + \delta_n$.

PROOF. Without loss of generality, assume $\theta_0 = 0$ and $d_0 = 1$. Let Ω' be the set of ω such that

$$\limsup_{n\to\infty}|\widehat{\theta}_n| \le (2C)^{-2/r},$$

$$\limsup_{n\to\infty}\frac{|\sum_{i=1}^n \psi(x_i,0)|}{s_n^{1/2}(\log\log n)^{1/2}} \le 2.5,$$

$$\limsup_{n\to\infty}\sup_{|\tau|\le 1}\frac{|\sum_{i=1}^n\{\psi(x_i,\tau) - \psi(x_i,0) - \lambda_i(\tau)\}|}{(A_n|\tau|^r + 1)^{1/2}(\log\log n)^{1/2}} \le C,$$

where C is the same constant as in (4.2).

By (B2), (B6) and Lemma 4.1, $P(\Omega') = 1$. We will show that (4.22) holds on Ω'. Note that (B5) implies $A_n \to \infty$ and that, for each $\omega \in \Omega'$ and for sufficiently large n,

$$
\begin{aligned}
&\left|\sum_{i=1}^n \psi(x_i, 0) + \sum_{i=1}^n \lambda_i(\widehat{\theta}_n)\right| \\
&\qquad \le \left|\sum_{i=1}^n \psi(x_i, \widehat{\theta}_n)\right| + \left|\sum_{i=1}^n (\psi(x_i, 0) - \psi(x_i, \widehat{\theta}_n) + \lambda_i(\widehat{\theta}_n))\right| \\
&\qquad \le \left|\sum_{i=1}^n \psi(x_i, \widehat{\theta}_n)\right| + (A_n|\widehat{\theta}_n|^r + 1)^{1/2}(\log\log n)^{1/2} \\
&\qquad\qquad \times \sup_{|\tau|\le 1} \frac{|\sum_{i=1}^n \{\psi(x_i, 0) - \psi(x_i, \tau) + \lambda_i(\tau)\}|}{(A_n|\tau|^r + 1)^{1/2}(\log\log n)^{1/2}} \\
&\qquad \le CA_n^{1/2}|\widehat{\theta}_n|^{r/2}(\log\log n)^{1/2} + C(\log\log n)^{1/2} + o(\delta_n) \\
&\qquad \le \frac{1}{4}A_n^{1/2}(\log\log n)^{1/2} + C(\log\log n)^{1/2} + o(\delta_n) \\
&\qquad \le \frac{1}{2}A_n^{1/2}(\log\log n)^{1/2}.
\end{aligned}
\tag{4.23}
$$

By (B7), we have

$$
c_n|\widehat{\theta}_n| \le \left|\sum_{i=1}^n \psi(x_i, 0)\right| + \left|\sum_{i=1}^n \psi(x_i, 0) + \sum_{i=1}^n \lambda_i(\widehat{\theta}_n)\right| \le 3(A_n \log\log n)^{1/2},
$$

for sufficiently large n. Using

$$
|\widehat{\theta}_n| \le 3c_n^{-1}(A_n \log\log n)^{1/2} \tag{4.24}
$$

in (4.23), we obtain

$$
\begin{aligned}
&\left|\sum_{i=1}^n \psi(x_i, 0) + \sum_{i=1}^n \lambda_i(\widehat{\theta}_n)\right| \\
&\qquad \le 3CA_n^{1/2}\big((A_n \log\log n)^{1/2}/c_n\big)^{r/2}(\log\log n)^{1/2} + C(\log\log n)^{1/2} + o(\delta_n).
\end{aligned}
$$

Therefore,

$$
c_n|\widehat{\theta}_n| \le 3(\log\log n)^{1/2}\big(s_n^{1/2} + CA_n^{1/2}((A_n \log\log n)^{1/2}/c_n)^{r/2} + C\big) + o(\delta_n)
$$

from which Lemma 4.4 follows. $\square$

Lemma 4.1 can be generalized in several ways. Two useful ones are dealt with in Lemmas 4.5 and 4.6. One is to remove condition (B4), and the other is to generalize the conditions (B3) and (B5).

LEMMA 4.5. *Suppose that there exist* $\sigma(d)$, $0 < \beta \le \alpha \le 1$, $0 \le \beta_1 < \infty$ *and a sequence of positive numbers* $\{a_i,\ i \ge 1\}$ *such that*

$$Eu^2(x_i, \theta, d) \le a_i^2 \sigma(d) \quad \text{for } |\theta - \theta_0| + d \le 2d_0,$$

$$Eu^{2+\alpha}(x_i, \theta_0, d) \le a_i^{2+\beta_1} \sigma(d)^{(2+\beta)/2} \quad \text{for } d \le d_0$$

and

$$\sum_{i \le n} a_i^{2+\beta_1} = O\big(A_n^{(2+\beta)/2} (\log n)^{-7-2\alpha}\big),$$

where $A_n = \sum_{i=1}^n a_i^2$. *Assume that there are* $r > 0$ *and* $L > 0$ *such that* $\sigma(d)/d^r$ *is nondecreasing and* $\sigma(2d) \le L\sigma(d)$. *Then we have*

$$\limsup_{n\to\infty} \sup_{|\tau-\theta_0| \le d_0} \frac{Z_n(\tau, \theta_0)}{(A_n \sigma(|\tau - \theta_0|) + 1)^{1/2} (\log\log(n + A_n))^{1/2}} \le C \quad a.s.,$$

for some constant $C < \infty$.

The proof of Lemma 4.5 is similar to that of Lemma 4.1, and Theorem 2.1 can also be modified under the weaker conditions. We omit the details. One can also use (B5′) to replace (B4) and (B5).

LEMMA 4.6. *Assume that* (B1), (B3) *and* (B5′) *are satisfied. Then we have*

$$\limsup_{n\to\infty} \sup_{|\tau-\theta_0| \le d_n} \frac{Z_n(\tau, \theta_0)}{(A_n d_n^r + 1)^{1/2} (\log\log(n + A_n))^{1/2}} \le C \quad a.s.,$$

for some constant $C < \infty$.

The proof of Lemma 4.6 is also a variant of that of Lemma 4.1. The following lemma is analogous to Lemma 4.4.

LEMMA 4.7. *Under the conditions of Theorem* 2.2, *for any sequence* $\widehat{\theta}_n$ *satisfying* (2.1) *and* $|\widehat{\theta}_n - \theta_0| \le d_n$ *almost surely*,

$$\sum_{i=1}^n \psi(x_i, \theta_0) + \sum_{i=1}^n \lambda_i(\widehat{\theta}_n) = O(\delta_n + A_n^{1/2} d_n^{r/2} (\log\log n)^{1/2}) \quad a.s.$$

Theorem 2.2 is a direct consequence of the preceding lemma.

5. Proofs of Theorems 3.2 and 3.3. Theorems 3.2 and 3.3 are direct applications of Theorems 2.1 and 2.2, respectively. We give some nontrivial details in both verifications.

PROOF OF THEOREM 3.2. To verify conditions (B3) and (B7), we have, by direct calculations,

$$\frac{d\Lambda_n(\theta)}{d\theta} = (p-1)\sum_{i=1}^{n} z_i\, z_i' E|e - z_i'\theta|^{p-2},$$

which is Lipschitz in θ under the assumptions of (C3) and (C5). We now turn to condition (B5). The same proof would validate (B3), whereas the remaining conditions of Theorem 2.1 are easy to check.

It suffices to show that, for sufficiently small $\alpha \ge 0$ and for every $i \ge 1$,

$$Eu^{2+\alpha}(x_i, \theta, d) \le K\,|z_i|^{\min(2p+1,\,4)+\varepsilon_1} d^{\min(2p-1,\,2)(2+\alpha\min(2(p-1),\,1))/2} \tag{5.1}$$

for some constant $\varepsilon_1 \in [0, \varepsilon]$. As can be seen from the proof below, one may take ε_1 to be 2α if $2/3 < p \le 2$ and $p\alpha$ if $1 \le p < 3/2$.

Recall that $\phi(e) = \operatorname{sgn}(e)|e|^{p-1}$. First note that, for any $s, t \in R$ and $1 \le p \le 2$,

$$|\phi(t+s) - \phi(t)| \le 5\min(|s||t|^{p-2}, |s|^{p-1})$$

[cf. Arcones (1996a), Lemma 4]. Therefore

$$\begin{aligned} u(x_i, \theta, d) &= \sup_{|\tau-\theta|\le d} \big|\big(\phi(y_i - z_i'\theta + z_i'(\tau-\theta)) - \phi(y_i - z_i'\theta)\big)z_i\big| \\ &\le 5 \sup_{|\tau-\theta|\le d} \min\big(|z_i'(\tau-\theta)|\,|y_i - z_i'\theta|^{p-2}|z_i|, |z_i'(\tau-\theta)|^{p-1}|z_i|\big) \\ &\le 5\min\big(|z_i|^2\, d\,|y_i - z_i'\theta|^{p-2}, d^{p-1}|z_i|^p\big). \end{aligned} \tag{5.2}$$

We complete the proof for two cases.

Case 1. $3/2 < p \le 2$. Let $0 \le \alpha < \min((2p-3)/2, \varepsilon/2)$. Since $0 \ge (p-2)(2+\alpha) > -1$ and f is bounded, we have

$$\begin{aligned} Eu^{2+\alpha}(x_i, \theta, d) &\le 5^{2+\alpha} d^{2+\alpha}|z_i|^{4+2\alpha} E|y_i - z_i'\theta|^{(p-2)(2+\alpha)} \\ &\le K\, d^{2+\alpha}|z_i|^{4+2\alpha}. \end{aligned}$$

Case 2. $1 \le p < 3/2$. Let $0 \le \alpha < \varepsilon/2$, and let

$$u_1 = |z_i|^2\, d\,|y_i - z_i'\theta|^{p-2}\, I_{\{|y_i - z_i'\theta| \ge |z_i| d\}} \quad \text{and} \quad u_2 = d^{p-1}|z_i|^p\, I_{\{|y_i - z_i'\theta| \le |z_i| d\}}.$$

It is easy to see that

$$\begin{aligned}
Eu_1^{2+\alpha} &\le (|z_i|^2 d)^{2+\alpha}\Big\{ I_{\{|z_i|d\le 1\}}\Big(E|y_i - z_i'\theta|^{(p-2)(2+\alpha)} I_{\{|y_i - z_i'\theta|>1\}} \\
&\qquad\qquad + E|y_i - z_i'\theta|^{(p-2)(2+\alpha)} I_{\{|z_i|d\le|y_i-z_i'\theta|\le 1\}}\Big) \\
&\qquad\qquad + I_{\{|z_i|d>1\}} E|y_i - z_i'\theta|^{(p-2)(2+\alpha)} I_{\{|y_i-z_i'\theta|\ge|z_i|d\}}\Big\} \\
&\le (|z_i|^2 d)^{2+\alpha}\Big\{ I_{\{|z_i|d\le 1\}}\Big(1 + \int_{|z_i|d\le|y_i|\le 1} |y_i|^{(p-2)(2+\alpha)} f(y + z_i'\theta)\,dy\Big) \\
&\qquad\qquad + I_{\{|z_i|d>1\}}(|z_i|d)^{(p-2)(2+\alpha)}\Big\} \\
&\le K(|z_i|^2 d)^{2+\alpha}\Big\{ I_{\{|z_i|d\le 1\}}\big(1 + (|z_i|d)^{(p-2)(2+\alpha)+1}\big) \\
&\qquad\qquad + I_{\{|z_i|d>1\}}(|z_i|d)^{(p-2)(2+\alpha)}\Big\} \\
&\le K\Big\{ |z_i|^{p(2+\alpha)+1} d^{(p-1)(2+\alpha)+1} + I_{\{|z_i|d>1\}}|z_i|^{p(2+\alpha)} d^{(p-1)(2+\alpha)}\Big\} \\
&\le K\,|z_i|^{2p+1+p\alpha} d^{(2p-1)(2+2\alpha(p-1))/2},
\end{aligned}$$

and

$$\begin{aligned}
Eu_2^{2+\alpha} &= d^{(p-1)(2+\alpha)}|z_i|^{p(2+\alpha)} P\big(|y_i - z_i'\theta| \le |z_i|d\big) \\
&\le K d^{(p-1)(2+\alpha)}|z_i|^{p(2+\alpha)}|z_i|d \\
&= K\,|z_i|^{2p+1+p\alpha} d^{(2p-1)(2+2\alpha(p-1))/2}.
\end{aligned}$$

Therefore, (5.1) follows from (5.2). □

REMARK 5.1. When $p = 3/2$, following the proof of Case 2 above, we have

$$Eu^2(x_i, \theta, d) \le K(1 + |z_i|^{4+\varepsilon}) d^2 |\log d|.$$

The Bahadur representation in this case follows from Lemma 4.5 and a corresponding modification of Theorem 2.1.

PROOF OF THEOREM 3.3. We first note that a corollary of Babu (1989) implies

$$\limsup_{n\to\infty} (n/\log\log n)^{1/2} |\widehat{\theta}_n| \le K \tag{5.3}$$

for some constant K. Arguments similar to those used in the proof of Theorem 3.2 show that (B3) holds with $r = 1$ and $A_n = K\sum_{i=1}^n |z_i|^3$. Since

$n^{-1}\sum_{i=1}^n z_i z_i' \to Q$, we have $\sum_{i=1}^n |z_i|^2 = \mathrm{trace}(\sum_{i=1}^n z_i z_i') = O(n)$. Furthermore,

$$\begin{aligned}\Lambda_n(\widehat{\theta}_n) &= -2\sum_{i=1}^n z_i \int_0^{z_i'\widehat{\theta}_n} f(y)\,dy \\ &= -2\sum_{i=1}^n z_i \left(\int_0^{z_i'\widehat{\theta}_n} f(0)\,dy + \int_0^{z_i'\widehat{\theta}_n} (f(y)-f(0))\,dy\right) \\ &= -2f(0)Q_n\,\widehat{\theta}_n - 2\sum_{i=1}^n z_i \int_0^{z_i'\widehat{\theta}_n} (f(y)-f(0))\,dy.\end{aligned}$$

Therefore, by (5.3) and condition (iv),

$$\begin{aligned}|\Lambda_n(\widehat{\theta}_n) + 2f(0)Q_n\,\widehat{\theta}_n| &\le \sum_{i=1}^n |z_i| \left|\int_0^{z_i'\widehat{\theta}_n} (f(y)-f(0))\,dy\right| \\ &= O\left(\sum_{i=1}^n |z_i| \int_0^{|z_i'\widehat{\theta}_n|} y^{1/2}\,dy\right) \\ &= O\left(\left(\frac{\log\log n}{n}\right)^{3/4} \sum_{i=1}^n |z_i|^{5/2}\right) \quad \text{a.s.}\end{aligned}$$

On the other hand, by the Lyapunov inequality

$$\begin{aligned}\sum_{i=1}^n |z_i|^{5/2} &\le \left(\sum_{i=1}^n |z_i|^2\right)^{1/2} \left(\sum_{i=1}^n |z_i|^3\right)^{1/2} \\ &= O\left(\left(\sum_{i=1}^n |z_i|^3\right)^{1/2} n^{1/2}\right).\end{aligned}$$

Thus, we have

$$|\Lambda_n(\widehat{\theta}_n) + 2f(0)Q_n\,\widehat{\theta}_n| = O\left(\left(\sum_{i=1}^n |z_i|^3\right)^{1/2} n^{1/2}((\log\log n)/n)^{3/4}\right) \quad \text{a.s.}$$

This means that (B8) is satisfied with

$$b_n = O\left(\left(n\sum_{i=1}^n |z_i|^3\right)^{1/2} ((\log\log n)/n)^{3/4}\right)$$

and $D_n = -2f(0)Q_n$. Condition (B5$'$) is clearly satisfied by assumption (iii), and other conditions of Theorem 2.2 are easy to verify with $n/c_n = O(1)$ and $s_n = O(n)$. $\square$

Note that condition (B8) would be easier to check under the stronger condition on the design used in Theorem 3.2. In Theorem 3.3, we allow for the cases where $n^{-1}\sum_i |z_i|^3$ is unbounded in n.

Acknowledgments. We thank two referees and an Associate Editor for their truly helpful suggestions which led to a much improved presentation among other things.

REFERENCES

ARCONES, M. A. (1996a). The Bahadur–Kiefer representation of L_p regression estimators. *Econometric Theory* **12** 257–283.

ARCONES, M. A. (1996b). Second order representations of the least absolute deviation regression estimator. Unpublished manuscript.

BABU, G. J. (1989). Strong representation for LAD estimators in linear models. *Probab. Theory Related Fields* **83** 547–558.

BAHADUR, R. R. (1966). A note on quantiles in large samples. *Ann. Math. Statist.* **37** 577–581.

BAI, Z. D., RAO, C. R. and WU, Y. (1992). M-estimation of multivariate linear regression parameters under a convex discrepancy function. *Statist. Sinica* **2** 237–254.

BLOOMFIELD, P. and STEIGER, W. L. (1983). *Least Absolute Deviation: Theory, Applications and Algorithms*. Birkhäuser, Boston.

BOSE, A. (1996). Bahadur representation of M-estimates based on U functionals. Technical Report 96-2, Dept. Statistics, Purdue Univ.

CARROLL, R. J. (1978). On almost sure expansions for M-estimates. *Ann. Statist.* **6** 314–318.

CHEN, X. (1993). On the law of the iterated logarithm for independent Banach space valued random variables. *Ann. Probab.* **21** 1991–2011.

FREEDMAN, D. (1975). On tail probability for martingales. *Ann. Probab.* **3** 100–118.

HABERMAN, S. J. (1989). Convexity and estimation. *Ann. Statist.* **17** 1631–1661.

HAMPEL, F. R., RONCHETTI, E. M., ROUSSEEUW, P. J. and STAHEL, W. A. (1986). *Robust Statistics: The Approach Based on Influence Functions*. Wiley, New York.

HE, X. and SHAO, Q. M. (1994). A strong Bahadur representation theorem of M-estimators. Research Report 590, Dept. Mathematics, National Univ. Singapore.

HE, X. and WANG, G. (1995). Law of the iterated logarithm and invariance principle for M-estimators. *Proc. Amer. Math. Soc.* **123** 563–573.

HÖSSJER, O. (1994). Rank-based estimates in the linear model with high breakdown point. *J. Amer. Statist. Assoc.* **89** 149–158.

HUBER, P. J. (1967). The behavior of maximum likelihood estimates under nonstandard conditions. *Proc. Fifth Berkeley Symp. Math. Statist. Probab.* **1** 221–233. Univ. California Press, Berkeley.

JUREČKOVÁ, J. (1985). Representation of M-estimators with the second-order asymptotic distribution. *Statist. Decisions* **3** 263–276.

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

KIEFER, J. (1967). On Bahadur representation of sample quantiles. *Ann. Math. Statist.* **38** 1323–1342.

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

LIESE, F. and VAJDA, I. (1994). Consistency of M-estimates in general regression models. *J. Multivariate Anal.* **50** 93–114.

MARTINSEK, A. (1989). Almost sure expansion for M-estimators and S-estimators in regression. Technical Report 25, Dept. Statistics, Univ. Illinois.

NIEMIRO, W. (1992). Asymptotics for M-estimators defined by convex minimization. *Ann. Statist.* **20** 1514–1533.

POLLARD, D. (1991). Asymptotics for least absolute deviation regression estimators. *Econometric Theory* **7** 186–199.

PORTNOY, S. and KOENKER, R. (1989). Adaptive L-estimation for linear models. *Ann. Statist.* **17** 362–381.

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

SPÄTH, H. (1991). *Mathematical Algorithms for Linear Regression*. Academic Press, New York.

WATSON, G. A. (1980). *Approximation Theory and Numerical Methods*. Wiley, New York.
WELSH, A. (1989). On M-process and M-estimation. *Ann. Statist.* **17** 337–362.
WITTMANN, R. (1987). Sufficient moment and truncated moment conditions for the law of the iterated logarithm. *Probab. Theory Related Fields* **75** 509–530.
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
UNIVERSITY OF ILLINOIS
725 S. WRIGHT
CHAMPAIGN, ILLINOIS 61820
E-MAIL: x-he@uiuc.edu

DEPARTMENT OF MATHEMATICS
UNIVERSITY OF OREGON
EUGENE, OREGON 97403
E-MAIL: shao@math.uoregon.edu