The Annals of Statistics
1992, Vol. 20, No. 3, 1548–1570

BOOTSTRAPPING *M*-ESTIMATORS OF A MULTIPLE LINEAR REGRESSION PARAMETER

BY SOUMENDRA NATH LAHIRI

Iowa State University

Consider a multiple linear regression model $Y_i = x_i'\beta + \varepsilon_i$, where the ε_i's are independent random variables with common distribution F and the x_i's are known design vectors. Let $\bar{\beta}_n$ be the M-estimator of β corresponding to a score function ψ. Under some conditions on F, ψ and the x_i's, two-term Edgeworth expansions for the distributions of standardized and studentized $\bar{\beta}_n$ are obtained. Furthermore, it is shown that the bootstrap method is second order correct in the studentized case when the bootstrap samples are drawn from some suitable weighted empirical distribution or from the ordinary empirical distribution of the residuals.

1. Introduction. Consider the following multiple linear regression model:

$$Y_i = x_i'\beta + \varepsilon_i, \qquad i = 1, \dots, n, \tag{1.1}$$

where $\varepsilon_1, \dots, \varepsilon_n$ are independent and identically distributed (iid) random variables with common distribution F, $x_1, \dots, x_n$ are known, nonrandom design vectors and β is the $p \times 1$ vector of parameters. In (1.1) and throughout the paper, A' denotes the transpose of a matrix A. Let ψ be a real-valued function defined on the real line. Then an M-estimator $\bar{\beta}_n$ of β corresponding to ψ is defined as a solution of the vector equation

$$\sum_{i=1}^{n} x_i \psi(Y_i - x_i't) = 0. \tag{1.2}$$

It is well known [cf. Huber (1980)] that under some conditions on F, ψ and the x_i's, $\Delta_n \equiv (\sum_{i=1}^{n} x_i x_i')^{1/2}(\bar{\beta}_n - \beta)$ has a limiting p-variate normal distribution with mean 0 and dispersion matrix $[E\psi^2(\varepsilon_1)/(E\psi'(\varepsilon_1))^2]I_p$, where I_p denotes the identity matrix of order p. However, Edgeworth expansions for these estimators are not very well studied in the present setup. For normalized, real-valued linear functions of the least squares estimator $\hat{\beta}_n$ of β, Navidi (1989) obtained a two-term Edgeworth expansion. Expansion for the least squares estimator $\hat{\beta}_n$ itself is recently obtained by Qumsiyeh (1990a). Because of the special structure of the score function [viz. $\psi(x) \equiv x$, $x \in \mathbb{R}^p$], the estimator $\hat{\beta}_n$ can be expressed as a vector of linear combinations of the iid random variables $\varepsilon_1, \dots, \varepsilon_n$. The expansion for $\hat{\beta}_n$ is then proved by applying the methods of Bhattacharya and Ranga Rao (1986) and Bhattacharya and Ghosh (1978). As for general M-estimators of β, Lahiri (1989b) considers the regression model (1.1) for the special case $p = 1$ and obtains Edgeworth

Received January 1989; revised September 1991.
AMS 1980 *subject classifications*. Primary 62G05; secondary 62E20.
Key words and phrases. Bootstrap, Edgeworth expansion, M-estimator, regression.

expansions for general M-estimators under weaker growth conditions on the design points. Moreover, it allows score functions that are not necessarily smooth.

This paper considers the standardized and the studentized versions of M-estimators and derives Edgeworth expansions in both cases. The proofs are based on the techniques developed by Bhattacharya and Ghosh (1978) and require a number of modifications to deal with the non-iid structure of the present model. Using the smoothness of the score function ψ and the definition of $\bar{\beta}_n$, first a sufficiently close stochastic approximation T_n (say) is derived for the standardized statistic Δ_n. The components of the approximating random vector T_n are smooth functions of normalized sums of certain independent random vectors. The derivation of this result suggests a general method for obtaining stochastic approximations in such contexts and, by itself, could be of independent interest. The Edgeworth expansion for T_n (and hence, for Δ_n) is then proved by adapting the techniques for Bhattacharya and Ghosh (1978) and Bhattacharya (1985) to the present setup. In carrying out the second step, the major difficulties arise from the lack of uniformity in the form of T_n for different n's. The approximating statistics $\{T_n\}$ are actually obtained through a sequence of transformations on the row sums of a triangular array of independent random vectors. Skovgaard (1981) gives some extensions of the Bhattacharya and Ghosh (1978) result for deriving expansions in such cases. However, the sequence of transformations defining T_n is not amenable to the analysis of the above papers.

Next, consider bootstrapping $\bar{\beta}_n$. Let F_n denote an estimator of the error distribution F based on the residuals $\bar{\varepsilon}_i = Y_i - x_i'\bar{\beta}_n$, $1 \le i \le n$. Let $\varepsilon_1^*, \dots, \varepsilon_n^*$ be a random sample from F_n and define $Y_i^* = x_i'\bar{\beta}_n + \varepsilon_i^*$, $1 \le i \le n$. In accordance with the original model, one may define the bootstrapped estimator β_n^* as a solution of

$$\sum_{i=1}^{n} x_i \psi(Y_i^* - x_i't) = 0. \tag{1.3}$$

But this does not work in general, as shown in Lahiri (1989b). For $p = 1$, $F_n =$ the empirical distribution of $\bar{\varepsilon}_1, \dots, \bar{\varepsilon}_n$, it gives an example which shows that almost surely, $(\beta_n^* - \bar{\beta}_n)$ is asymptotically normal with some random mean, and the random mean itself converges weakly to the standard normal distribution. Along the same line one can show that the situation hardly improves if, instead, F_n is taken to be the empirical distribution of the centered residuals. Hence, one has to modify the naive bootstrap procedure as described above. For the uniparameter case, certain modifications are proposed and shown to be second order correct in Lahiri (1989b). It follows from the above paper that one must remove the inherent asymptotic bias of the bootstrapped M-estimator by requiring F_n and ψ to satisfy

$$E_n \psi(\varepsilon_1^*) = 0, \tag{1.4}$$

where E_n denotes the (bootstrap) expectation under F_n. This can be achieved either by choosing the resampling distribution suitably or, alternatively, by modifying (1.3) according to Shorack (1982). In this paper we consider one case under each type of modification. First, consider the situation where for some j, $1 \leq j \leq p$, the jth component of all x_i's are of the same sign (i.e., either all positive or all negative). For example, this holds if the regression model (1.1) has an intercept. Choosing F_n to be a suitable weighted empirical distribution with weights depending on the x_{ij}'s, it is shown that the corresponding bootstrap procedure provides a better approximation than the usual normal approximation. See Section 2 for details.

In the second modification, F_n is taken to be the ordinary empirical distribution of residuals but the bootstrapped estimator β_n^* is defined as a solution of the modified equation

$$\sum_{i=1}^{n} x_i\big(\psi(Y_i^* - x_i't) - E_n\psi(\varepsilon_1^*)\big) = 0. \tag{1.5}$$

Note that, in this case, (1.4) is satisfied with ψ replaced by $(\psi - E_n\psi(\varepsilon_1^*))$. This modification has been originally proposed by Shorack (1982) for bootstrapping M-estimators of multiple linear regression parameters. The results of Shorack (1982) show that this version of bootstrap procedure yields valid approximations, provided the number of parameters p increases with the sample size n at the rate $o(n^{1/3})$. In this context, some interesting first-order results have been proved by Mammen (1989). There it has been shown that the distributions of linear functions $a'\Delta_n$ ($a \in \mathbb{R}^p$, $\|a\| = 1$) can be approximated by the above procedure even in some cases where the limiting distribution of the unbootstrapped statistic is not known. Results on bootstrapping least squares estimators have been obtained by Freedman (1981), Bickel and Freedman (1983) and Qumsiyeh (1990b).

For fixed $p \geq 1$, here it will be shown that both the modifications described above are second order correct for the studentized M-estimator.

The rest of the paper is organized as follows. Section 2 gives the results on Edgeworth expansions and bootstrap approximations. Section 3 contains the proofs.

2. Assumptions and main results. For stating the results of this section, we need to develop some notation. All throughout the paper, the dependence on n will be suppressed in the notation, unless it is demanded by clarity. Let $D \equiv D_n = (\sum_{i=1}^n x_i x_i')^{-1/2}$ and $d_i \equiv d_{in} = Dx_i$, $1 \leq i \leq n$. Write $q = p(p+1)/2$. For each $d_i = (d_{i1}, \dots, d_{ip})'$, define a $q \times 1$ vector c_i $(\equiv c_{in})$ by $c_i' = (d_{i1}^2, d_{i1}d_{i2}, \dots, d_{i1}d_{ip}; d_{i2}^2, d_{i2}d_{i3}, \dots, d_{i2}d_{ip}; \dots; d_{ip}^2)$. Note that for any constants $u_1, \dots, u_n \in \mathbb{R}$, $\sum_{i=1}^n u_i c_i = 0 \Leftrightarrow \sum_{i=1}^n u_i d_{ij} d_{im} = 0$ for all $1 \leq j, m \leq p \Leftrightarrow \sum_{i=1}^n u_i d_i d_i' = 0 \Leftrightarrow \sum_{i=1}^n \alpha_i x_i x_i' = 0$. Hence $\{c_1, \dots, c_n\}$ are linearly independent $\Leftrightarrow \{x_i x_i', 1 \leq i \leq n\}$ are linearly independent. As a result, $r_n \equiv$ the rank of $\sum_{i=1}^n c_i c_i'$ is nondecreasing in n. Let $r = \max\{r_n: n \geq 1\}$. Then $1 \leq r \leq q$ and $r_n = r$ for all large n. Without loss of generality (w.l.o.g.), assume that $r_n = r$ for all $n \geq q$. Note that r may not be equal to q. [An important example, where $r < q$, is provided by the one-way-lay-out model. In this case, $x_i = (1, 0, \dots, 0)'$, $1 \le i \le n_1$, $x_i = (0, 1, \dots, 0)'$, $n_1 < i \le n_1 + n_2, \dots, x_i = (0, \dots, 0, 1)'$, $n - n_p < i \le n$ for some $n_1 + \cdots + n_p = n$ and hence, $r_n = p$ for all $n \ge p$.] For any $1 \le r \le q$, the spectral decomposition of the real symmetric matrix $\sum_{i=1}^n c_i c_i'$ yields a $q \times q$ nonsingular matrix $B \equiv B(n)$ of rank r such that

$$B\left(\sum_{i=1}^{n} c_i c_i'\right)B' = \begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix}. \tag{2.1}$$

Partition B as $B' = [\overline{B}_1' \colon \overline{B}_2']$, where $\overline{B}_1$ is of order $r \times q$. Define the $r \times 1$ vectors b_i by

$$b_i = \overline{B}_1 c_i, \qquad 1 \le i \le n. \tag{2.2}$$

Note that $\sum_{i=1}^n b_i b_i' = \overline{B}_1(\sum_{i=1}^n c_i c_i')\overline{B}_1' = I_r$ and $\sum_{i=1}^n d_i d_i' = I_p$. Let $\gamma_n = (\sum_{i=1}^n \|d_i\|^6)^{1/4} + (\sum_{i=1}^n \|b_i\|^4)^{1/2}$, $n \ge 1$. For $\delta > 0$, define the set $A_n(\delta)$ by $A_n(\delta) = \{i\colon 1 \le i \le n,\ (d_i't_1)^2 + (b_i't_2)^2 > \delta\gamma_n^2$ for all $t_1 \in \mathbb{R}^p$, $t_2 \in \mathbb{R}^r$ with $\|t_1\|^2 + \|t_2\|^2 = 1\}$. Let $K_n(\delta) = |A_n(\delta)|$, where for any set A, $|A|$ denotes the size of A. For any positive definite (p.d.) Λ, let Φ_Λ denote the normal distribution with mean 0 and dispersion matrix Λ. Write ϕ_Λ for the Lebesgue density of Φ_Λ. For notational simplicity, set $\Phi_\Lambda = \Phi$ and $\phi_\Lambda = \phi$, if $\Lambda = I_p$. For any real-valued function $h\colon \mathbb{R}^k \to \mathbb{R}$, $k \ge 1$, let $\|h\|_\infty$ denote its supremum norm and let h', h'', h''' respectively denote the first, the second and the third derivatives of h, when $k = 1$. Unless otherwise specified, all the limits are taken as $n \to \infty$.

Next, we state the assumptions needed for deriving the Edgeworth expansion of Δ_n. These assumptions are made on three different quantities, viz. the score function ψ, the error distribution F and the design vectors $(x_1, \dots, x_n)$. Here we shall list them down in the same order.

Conditions on ψ:

(C.1) ψ is twice differentiable and the second derivative ψ'' satisfies a Lipschitz condition of order α for some $0 < 2\alpha \le 1$.

Conditions on F:

(C.2) $E|\psi(\varepsilon_1)|^3 + E|\psi'(\varepsilon_1)|^3 + E|\psi''(\varepsilon_1)|^2 < \infty$.
(C.3) $E\psi(\varepsilon_1) = 0$, and $\sigma^2 \equiv E\psi^2(\varepsilon_1)/(E\psi'(\varepsilon_1))^2 \in (0, \infty)$.
(C.4) A maximal linearly independent [as elements of the vector space $L^2(F)$] subset of the random variables $\{1, \psi(\varepsilon_1), \psi'(\varepsilon_1)\}$ satisfies the Cramér condition. Recall that a $\mathbb{R}^k$-valued random vector Y with characteristic function $\hat{f}(t)$ satisfies the Cramér condition if $\limsup_{\|t\| \to \infty} |\hat{f}(t)| < 1$, where $\|t\|$ denotes the usual Euclidean norm of a vector $t \in \mathbb{R}^k$, $k \ge 1$.

Conditions on the x_i's:

(C.5) $(\sum_{i=1}^n x_i x_i')$ is invertible for some $n \ge p$.
(C.6) $\gamma_n \equiv (\sum_{i=1}^n \|d_i\|^6)^{1/4} + (\sum_{i=1}^n \|b_i\|^4)^{1/2} = o(1)$.
(C.7) There exists a $\delta > 0$ such that $(-\log \gamma_n)/K_n(\delta) = o(1)$.

Some of the above conditions are fairly standard in the literature while some others are not. For proving the asymptotic normality of $\bar{\beta}_n$, one typically assumes conditions (C.3) and (C.5). For this, another necessary condition on the design points $(x_1, \ldots, x_n)$ is that $\max\{\|d_i\|: i = 1, \ldots, n\} = o(1)$. Using the fact $\sum_{i=1}^n d_i d_i' = I_p$, it is easy to show that this is equivalent to $(\sum_{i=1}^n \|d_i\|^6)^{1/4} = o(1)$. Similarly, the relation $\sum_{i=1}^n b_i b_i' = I_r$ implies that $\max\{\|b_i\|: 1 \le i \le n\} = o(1)$ if and only if $(\sum_{i=1}^n \|b_i\|^4)^{1/2} = o(1)$. The restriction $2\alpha \le 1$ in (C.1) is imposed only for notational simplicity. The conclusions of all the theorems remain valid without this restriction. The Cramér condition on the joint distribution of $(\psi(\varepsilon_1), \psi'(\varepsilon_1))$ in (C.4) can be relaxed in some cases. See Remark 2.3. Condition (C.7) is relatively uncommon in the literature and possibly needs some clarification. It requires only a very small fraction [e.g., $O((\log n)^{1+\alpha}/n)$ for some $\alpha > 0$ if $\gamma_n = O(n^{-1/2})$] of the n design vectors not to cluster in some lower-dimensional hyperplane. A somewhat similar condition has been used in Lahiri (1989b) for the uniparameter case. See the proposition following Remark 2.4 for a sufficient condition.

We are now ready to state the first theorem.

THEOREM 2.1. *Let conditions* (C.1), (C.2), (C.3), (C.5) *and* (C.6) *hold.*

(a) *Then there exists a sequence of statistics* $\{\bar{\beta}_n\}$ *and constants* $C_1 > 0$ *and* $C_2 > 0$ *such that*

$$P\big(\bar{\beta}_n \text{ solves } (1.2) \text{ and } \|\Delta_n\| < C_1 \nu_n\big) > 1 - C_2 \nu_n^{-3}\Bigg(\sum_{i=1}^n \|d_i\|^6\Bigg)^{1/4}$$

for all $n \ge C_1$, *where* $\nu_n \equiv -\log(\sum_{i=1}^n \|d_i\|^6)$.

(b) *If, in addition,* (C.4) *and* (C.7) *hold, then for the sequence* $\{\bar{\beta}_n\}$ *of part* (a), *there exist polynomials* $a_n(F, \cdot)$, $n \ge 1$, *such that*

$$\sup_{B \in \mathscr{B}} \left| P\big((\sigma D)^{-1}(\bar{\beta}_n - \beta) \in B\big) - \int_B (1 + a_n(F, x))\phi(x)\,dx \right| = o(\gamma_n),$$

where $\mathscr{B}$ *is a class of Borel subsets of* $\mathbb{R}^p$ *satisfying*

$$\sup_{B \in \mathscr{B}} \Phi\big((\partial B)^{\varepsilon}\big) = O(\varepsilon) \quad \text{as } \varepsilon \downarrow 0, \tag{2.3}$$

$\sigma^2 = E\psi^2(\varepsilon_1)/(E\psi'(\varepsilon_1))^2$, *and the coefficients of* $a_n(F, \cdot)$ *are continuous functions of the finite moments of* $\psi(\varepsilon_1)$, $\psi'(\varepsilon_1)$ *and* $\psi''(\varepsilon_1)$. *Furthermore,* $\|a_n(F, \cdot)\phi(\cdot)\|_\infty = O(\gamma_n)$.

Next, we define the studentized M-estimator. Note that the asymptotic dispersion matrix of Δ_n is given by $\sigma^2 I_p$, where $\sigma^2 = s^2\tau^{-2}$, $\tau = E\psi'(\varepsilon_1)$ and $s^2 = E\psi^2(\varepsilon_1)$. Hence, a natural estimator of σ^2 is $\hat{\sigma}_n^2 = s_n^2 \tau_n^{-2}$, where $s_n^2 =$

$n^{-1}\sum_{i=1}^{n}\psi^2(\bar{\varepsilon}_i)$ and $\tau_n = n^{-1}\sum_{i=1}^{n}\psi'(\bar{\varepsilon}_i)$. We need the following condition to derive an Edgeworth expansion for the studentized statistic $(\hat{\sigma}_n D)^{-1}(\bar{\beta}_n - \beta)$.

(C.8)(i) A maximal linearly independent subset L of $\{1, \psi(\varepsilon_1), \psi'(\varepsilon_1), \psi^2(\varepsilon_1)\}$ satisfies the Cramér condition and the elements in it have finite third moments.

(ii) If $\psi'(\varepsilon_1)$ lies in the maximal set, conditions (C.6) and (C.7) hold with the b_i's replaced by $\tilde{b}_i$'s, where $\tilde{b}_i$ is defined as in (2.1), starting with $\tilde{c}_i = [c_i' : n^{-1}]'$, $1 \le i \le n$.

For stating the next result, define $\bar{\bar{\gamma}}_n = (\sum_{i=1}^{n}\|d_i\|^6)^{1/4} + (\sum_{i=1}^{n}\|\tilde{b}_i\|^4)^{1/2}$. Let $\bar{\gamma}_n = \bar{\bar{\gamma}}_n$ or γ_n according as $\psi'(\varepsilon_1) \in L$ or $\notin L$ in (C.8)(i). With this, we have the following result on the Edgeworth expansion of the studentized M-estimator.

THEOREM 2.2. *In addition to the hypotheses of Theorem* 2.1(a) *and* (b), *assume that condition* (C.8) *holds and that* $\nu_n = O(n^t)$ *for some* $0 < 8t < 1$. *Then, for the sequence* $\{\bar{\beta}_n\}$ *of Theorem* 2.1(a), *there exist polynomials* $\bar{a}_n(F, \cdot)$ *such that*

$$\sup_{B \in \mathscr{B}} \left| P\Big((\hat{\sigma}_n D)^{-1}(\bar{\beta}_n - \beta) \in B\Big) - \int_B (1 + \bar{a}_n(F, x))\phi(x)\,dx \right|$$
$$= o\big(\bar{\gamma}_n + n^{-1/2}\big)$$

for every class $\mathscr{B}$ *of Borel subsets of* $\mathbb{R}^p$ *satisfying* (2.3). *The coefficients of* $\bar{a}_n(F, \cdot)$ *are continuous functions of the finite moments of* $(\psi(\varepsilon_1), \psi'(\varepsilon_1))$, *and* $\|\bar{a}_n(F, \cdot)\phi(\cdot)\|_\infty = O(\bar{\gamma}_n + n^{-1/2})$.

REMARK 2.1. Instead of $\hat{\sigma}_n$, one may actually use some other estimators of σ for studentizing $\bar{\beta}_n$. In fact, if one uses the "less natural" estimator $\tilde{\sigma}_n = s_n/\tilde{\tau}_n$, where $\tilde{\tau}_n = \sum_{i=1}^{n} d_{ij}^2 \psi'(\bar{\varepsilon}_i)$ for some $1 \le j \le p$, an Edgeworth expansion for the statistics $(\tilde{\sigma}_n D)^{-1}(\bar{\beta}_n - \beta)$ can be derived under conditions (C.1), (C.3), (C.5)–(C.7) and (C.8)(i) only. In this case, condition (C.8)(ii) is not at all necessary.

REMARK 2.2. Theorems 2.1 and 2.2 can be extended in a straightforward way to get higher-order Edgeworth expansions. Under additional smoothness conditions on ψ, one may follow the steps in the first part of the proof of Theorem 2.1 to derive a closer stochastic approximation for Δ_n and use it to obtain more terms in the Edgeworth expansion. In that case, one needs to impose the Cramér condition on the joint distribution of a maximal linearly independent subset of the random variables $\{1, \psi(\varepsilon_1), \psi'(\varepsilon_1), \dots, \psi^{(k)}(\varepsilon_1)\}$ for some $k \ge 2$, depending on the order of expansion.

REMARK 2.3. The Cramér condition (C.4) in Theorem 2.1 can be replaced by a strong nonlatticeness condition on the corresponding maximal set of random variables, provided the design vectors satisfy some additional growth conditions. Without any such conditions on the x_i's, the strong nonlatticeness condition ensures an error bound $o(\max\{\|b_i\|, \|d_i\|: 1 \le i \le n\})$ only, which can be coarser than $o(\gamma_n)$ in some situations [cf. Remark 2.4 of Lahiri (1989b)]. A similar comment applies also for Theorem 2.2.

REMARK 2.4. Though condition (C.7) may look somewhat restrictive at first glance, it is satisfied under some simple conditions on γ_n. If γ_n^2 is asymptotically equivalent to n^{-1}, then condition (C.7) is satisfied. More precisely, one has the following proposition.

PROPOSITION. *If* $\limsup_{n\to\infty} n\gamma_n^2 < \infty$, *then condition* (C.7) *holds.*

Next, we consider the performance of the bootstrap approximation under the modifications mentioned in Section 1. For the first modification, suppose that there is a j, $1 \le j \le p$, such that the jth component x_{ij} of x_i, $i \ge 1$, are all of the same sign. Let $p_n = \sum_{i=1}^n |x_{ij}|$. Let F_{jn} denote the weighted empirical distribution putting mass $|x_{ij}|/p_n$ at the ith residual $\bar{\varepsilon}_i$, $i = 1, \dots, n$. Define β_n^* by (1.3) and $\hat{\sigma}_n^*$ by replacing the role of F by F_{jn} in the definitions of $\hat{\sigma}_n$ respectively. Let P_n denote the bootstrap probability, given $Y_1, \dots, Y_n$. Then we have the following theorem.

THEOREM 2.3. *Assume that the conditions of Theorem* 2.2 *hold. If, in addition,* $p_n^{-2}(\sum_{i=1}^n x_{ij}^2) = o((\log n)^{-2})$, *then there exist constants* $C_1 > 0$, $C_2 > 0$ *and a sequence of Borel sets* $A_{1n} \subseteq \mathbb{R}^n$, *such that* $P((\varepsilon_1, \dots, \varepsilon_n) \in A_{1n}) \to 1$ *as* $n \to \infty$, *and given* $(\varepsilon_1, \dots, \varepsilon_n) \in A_{1n}$, $n \ge C_1$,

(a) *there exists a random vector* β_n^*, *depending on* $(\varepsilon_1^*, \dots, \varepsilon_n^*)$, *such that*

$$P_n\Big(\beta_n^* \text{ solves } (1.3) \text{ and } \big\|D^{-1}\big(\beta_n^* - \bar{\beta}\big)\big\| < C_1\nu_n\Big) > 1 - C_2\bar{\gamma}_n\nu_n^{-3},$$

$$\text{(b)} \quad \sup_{B\in\mathscr{B}} \Big|P_n\Big(\big(\hat{\sigma}_n^* D\big)^{-1}\big(\beta_n^* - \bar{\beta}_n\big) \in B\Big) - P\Big(\big(\hat{\sigma}_n D\big)^{-1}\big(\bar{\beta}_n - \beta\big) \in B\Big)\Big|$$

$$\le C_2\delta_n\big(\bar{\gamma}_n + n^{-1/2}\big),$$

where the random variables $\delta_n \equiv \delta_n(\varepsilon_1, \dots, \varepsilon_n)$ *tend to* 0 *in probability as* $n \to \infty$ *and* $\mathscr{B}$ *satisfies* (2.3).

For the other modification, again, we have a similar result. Take the resampling distribution F_n to be the ordinary empirical distribution of the residuals $\bar{\varepsilon}_i = y_i - x_i'\bar{\beta}_n$, $i = 1, \dots, n$. Define β_n^* as a solution of (1.5). Then we have the following result.

THEOREM 2.4. *Assume that the conditions of Theorem 2.2 hold. Then there exist constants* $C_1, C_2 > 0$ *and a sequence of Borel sets* $A_{1n} \subseteq \mathbb{R}^n$, *such that* $P((\varepsilon_1, \dots, \varepsilon_n) \in A_{1n}) \to 1$ *as* $n \to \infty$, *and given* $(\varepsilon_1, \dots, \varepsilon_n) \in A_{1n}$, $n \geq C_1$,

(a) *there exists a random vector* β_n^*, *depending on* $(\varepsilon_1^*, \dots, \varepsilon_n^*)$, *such that*

$$P_n\Big(\beta_n^* \text{ solves (1.5) and } \big\|D^{-1}\big(\beta_n^* - \bar{\beta}_n\big)\big\| < C_1\nu_n\Big) > 1 - C_2\bar{\gamma}_n\nu_n^{-3},$$

(b)
$$\sup_{B \in \mathscr{B}} \Big| P_n\Big(\big(\hat{\sigma}_n^* D\big)^{-1}\big(\beta_n^* - \bar{\beta}_n\big) \in B\Big) - P\Big(\big(\hat{\sigma}_n D\big)^{-1}\big(\bar{\beta}_n - \beta\big) \in B\Big)\Big| \leq C_2\delta_n\big(\bar{\gamma}_n + n^{-1/2}\big),$$

where $\delta_n \equiv \delta_n(\varepsilon_1, \dots, \varepsilon_n)$ *goes to* 0 *in probability as* $n \to \infty$ *and* $\mathscr{B}$ *satisfies* (2.3).

REMARK 2.5. Note that both these modifications coincide when the original model (1.1) has an intercept (so that $x_{i1} = 1$ for all $i \geq 1$) and F_n is taken to be the weighted empirical distribution based on the weights x_{i1}, $i = 1, \dots, n$. In this case one obtains the same conclusion from both Theorem 2.3 and Theorem 2.4.

REMARK 2.6. As pointed out by a referee, Theorems 2.3 and 2.4 lose much of their significance unless the rate of convergence of $P((\varepsilon_1, \dots, \varepsilon_n) \in A_{1n})$ to 1 is fast enough. Following the steps in the proofs of Theorems 2.3 and 2.4, one can show that $P((\varepsilon_1, \dots, \varepsilon_n) \in A_{1n}) = 1 - o(n^{-1/2} + (\bar{\gamma})^a + \bar{\eta}_n)$ for any $0 < a < 4/7$, and any sequence $\{\bar{\eta}_n\}$ satisfying

$$(\bar{\eta}_n)^{-1}\Big[\max\Big\{E\big|\psi(\varepsilon_1)^j\psi'(\varepsilon_1)^m\big|I\Big(\tilde{w}_n\big|\psi(\varepsilon_1)^j\psi'(\varepsilon_1)^m\big|^2 > 1\Big): 0 \leq j, m \leq j + 2m \leq 6\Big\} + E\psi''(\varepsilon_1)^2 I\big(\tilde{w}_n\psi''(\varepsilon_1)^4 > 1\big)\Big] = o(1),$$

where $\tilde{w}_n = \max\{n^{-1} + p_n^{-1}|x_{ij}|: 1 \leq i \leq n\}$. The bounds on a result from estimating the difference $E_n\psi(\varepsilon_1^*)^6 - E\psi(\varepsilon_1)^6$. Here one needs to show that $\sum_{i=1}^n w_i|\psi(\varepsilon_i)^5\psi'(\varepsilon_i)|\,|d_i'\Delta_n| = o_P(1)$ for some weights w_i, $1 \leq i \leq n$, under the condition $E|\psi(\varepsilon_1)^5\psi'(\varepsilon_1)|^{6/7} \leq (E\psi(\varepsilon_1)^6)^{5/7}(E|\psi'(\varepsilon_1)|^3)^{2/7} < \infty$. Similarly, the bound on η_n derives from the proofs of assertions like "$E\psi(\varepsilon_1)^6 < \infty \Rightarrow \sum_{i=1}^n w_i(\psi(\varepsilon_i)^6 - E\psi(\varepsilon_1)^6) = o_P(1)$." Under stronger moment conditions, one can easily modify the relevant steps in the proofs to establish faster convergence rates for $P((\varepsilon_1, \dots, \varepsilon_n) \in A_{1n})$. In particular, if ψ, ψ' and ψ'' are bounded, then $P((\varepsilon_1, \dots, \varepsilon_n) \in A_{1n}) = 1 - O(\exp(-C[n + (\tilde{w}_n)^{-1} + (\bar{\gamma})^{-1}]))$ for some $C > 0$.

Almost sure validity of the bootstrap procedures can be established under stronger conditions on the model (1.1). It is obvious that in this case, one needs at least the almost sure existence of $\bar{\beta}_n$. In fact, the existence of a solution $\bar{\beta}_n$ of (1.2) satisfying $\|\Delta_n\| < C\nu_n$ can be proved using the law of iterated logarithms for weighted sums of iid random variables [cf. Theorem 10.5 of Petrov (1975)] and Brouwer's fixed point theorem as in the proof of

Proposition 2.1 of Lahiri (1989a). However, if we are ready to assume this, then we have the following result.

THEOREM 2.5. *Let the conditions of Theorem* 2.2 *hold and* $\max\{\|d_i\|^2: 1 \le i \le n\} = O(n^{-1})$. *Assume that a solution* $\bar{\beta}_n$ *of* (1.2) *exists and* $\|\Delta_n\| = O(\nu_n)$ *almost surely.*

(I) *Suppose that* F_n *is taken to be the ordinary empirical distribution of* $\bar{\varepsilon}_i$, $i = 1, \dots, n$, *and* β_n^* *is defined as a solution of* (1.5).

(a) *Then, for almost all sample points* ω, *there exists a positive integer* $N(\omega)$ *such that for all* $n \ge N(\omega)$ *and for some constants* $C_1 > 0$, $C_2 > 0$,

$$P_n\left(\beta_n^* \text{ solves } (1.5) \text{ and } \left\|D^{-1}\left(\beta_n^* - \bar{\beta}_n\right)\right\| < C_1\nu_n\right) > 1 - C_2\bar{\gamma}_n\nu_n^{-3}.$$

(b) *For the solution* β_n^* *of part* (a) *and class* $\mathscr{B}$ *of Borel subsets of* $\mathbb{R}^p$ *satisfying* (2.3),

$$\sup_{B\in\mathscr{B}}\left|P_n\left(\left(\hat{\sigma}_n^*D\right)^{-1}\left(\beta_n^* - \bar{\beta}_n\right) \in B\right) - P\left(\hat{\sigma}_nD\right)^{-1}\left(\bar{\beta}_n - \beta\right) \in B\right)\right|$$
$$= o\left(\bar{\gamma}_n + n^{-1/2}\right) \quad a.s.$$

(II) *Suppose that the resampling distribution* F_n *is chosen to be* F_{jn} *and* $\max\{p_n^{-1}|x_{ij}|: 1 \le i \le n\} = O(n^{-1})$. *Then assertions* (a) *and* (b) *of part* (I) *hold for* β_n^* *defined by* (1.3).

REMARK 2.7. It can be shown that the above modifications of the bootstrap remain second order correct for the normalized M-estimator $(\sigma D)^{-1}(\bar{\beta}_n - \beta)$ under weaker conditions. More specifically, conclusions similar to Theorems 2.3, 2.4 and 2.5 hold for $(\sigma D)^{-1}(\bar{\beta}_n - \beta)$ provided the hypotheses of the respective theorems are satisfied with "conditions of Theorem 2.2" replaced by "conditions of Theorem 2.1(b)." See Lahiri (1990) for details.

3. Proofs. Before proving the theorems, we will state and prove some lemmas. Throughout this section, $C, C_1, C_2, \dots$ will denote generic constants that do not depend on the variables like n, x, and so on. For a nonnegative integral vector $\alpha = (\alpha_1, \dots, \alpha_k)'$ and a function $f = (f_1, \dots, f_k)$: $\mathbb{R}^k \to \mathbb{R}^k$, $k \ge 1$, write $|\alpha| = \alpha_1 + \cdots + \alpha_k$, $\alpha! = \alpha_1! \cdots \alpha_k!$, $f^\alpha = (f_1^{\alpha_1}) \cdots (f_k^{\alpha_k})$, $D^\alpha f_1 = D_1^{\alpha_1} \cdots D_k^{\alpha_k} f_1$, where $D_j f_1$ denotes the partial derivative of f_1 with respect to the jth argument $1 \le j \le k$. For $t = (t_1, \dots, t_k)' \in \mathbb{R}^k$ and α as above, define $t^\alpha = t_1^{\alpha_1} \cdots t_k^{\alpha_k}$. Some of the symbols used in this section are *local* to the lemmas and may denote different quantities elsewhere in the paper.

LEMMA 3.1. *Let* $\{Y_{in} = (Y_{i1n}, \dots, Y_{imn})', 1 \le i \le n\}_{n\ge1}$ *be a row iid triangular array of random vectors with* $EY_{1n} = 0$ *and* $E\|Y_{1n}\|^3 < K < \infty$ *for all* $n \ge 1$. *For* $1 \le j \le m$, $1 \le i \le n$ *and* $n \ge 1$, *let* $a_{ij} \equiv a_{ijn}$ *be a* $p_j \times 1$

vector of constants satisfying $\sum_{i=1}^{n} a_{ij} a'_{ij} = I_{p_j}$, $\tilde{\gamma}_n = \sum_{j=1}^{m} (\sum_{i=1}^{n} \|a_{ij}\|^4)^{1/2} = o(1)$, *and let* $-\log \tilde{\gamma}_n = o(|(1 \le i \le n: \|(a'_{i1} t_1, \dots, a'_{im} t_m)'\| > \tilde{C} \tilde{\gamma}_n$ *for all* $\|(t'_1, \dots, t'_m)'\| = 1)|)$ *for some constant* $\tilde{C} > 0$. *Define*

$$\tilde{U}_{in} = (a_{i1} Y_{i1n}, \dots, a_{im} Y_{imn})', \quad V_n = \operatorname{Disp}\left(\sum_{i=1}^{n} \tilde{U}_{in} \right), \quad U_{in} = V_n^{-1/2} \tilde{U}_{in},$$

for $1 \le i \le n$, *and* $\tilde{\alpha}_n = E\|Y_{1n}\|^3 I(\|Y_{1n}\|^2 > \lambda \tilde{\gamma}_n)$, *where* λ *is a real number satisfying* $0 < \lambda < \liminf_{n \to \infty} \lambda_n$, $\lambda_n \equiv$ *the smallest eigenvalue of* Σ_n, *and* $\Sigma_n = \operatorname{Disp}(Y_{1n})$. *Suppose that* $\tilde{\alpha}_n = o(1)$ *and* $\limsup_{n \to \infty} \sup\{|E \exp(it'Y_{1n})|: \delta \le \|t\| \le \kappa_n\} < 1$ *for all* $\delta > 0$ *and for some* $\{\kappa_n\}$ *satisfying* $\max\{\|a_{ij}\| / \tilde{\gamma}_n: 1 \le i \le n, 1 \le j \le m\} = o(\kappa_n)$. *Then, for any Borel set* B *of* $\mathbb{R}^k$,

$$\left| P\left(\sum_{i=1}^{n} U_{in} \in B \right) - \int_B \xi_n(y)\, dy \right| \le C\left[\tilde{\alpha}_n \tilde{\gamma}_n + \tilde{\gamma}_n^2 + \Phi\left((\partial B)^{e_n} \right) \right],$$

where $k = \sum_{j=1}^{m} p_j$, $e_n = o(\tilde{\gamma}_n)$, $\xi_n(y) = (1 + \sum_{|\nu|=3} (\chi_{\nu,n}/\nu!) p_\nu(y)) \phi(y)$, $p_\nu(\cdot)$ *is a polynomial and* $\chi_{\nu,n} = \sum_{i=1}^{n} (\nu$*th cumulant of* $U_{in})$.

PROOF. Lemma 3.1 can be proved along the line of the proof of Theorem 20.8 of Bhattacharya and Ranga Rao (1986) [hereafter, referred to as BR(86)]. Here we briefly mention the modifications required to obtain the above remainder term under the hypothesis of Lemma 3.1.

Define $\tilde{\gamma}_{1n} = \max\{\|a_{ij}\|: 1 \le i \le n, 1 \le j \le m\}$, $S_n = \sum_{i=1}^{n} U_{in}$, $Z_{in} = U_{in} I(\|U_{in}\| \le 1) - EU_{in} I(\|U_{in}\| \le 1)$, $a_n = \sum_{i=1}^{n} EU_{in} I(\|U_{in}\| \le 1)$, $S''_n = \sum_{i=1}^{n} Z_{in}$, $B_n = (\operatorname{Disp} S''_n)^{-1/2}$, $g_{in}(t) = Ee^{\sqrt{-1}\, t' Z_{in}}$, and $\bar{\chi}_{\nu,n} = \sum_{i=1}^{n} (\nu$th cumulant of $Z_{in})$, $1 \le i \le n$. Let

$$\bar{\xi}_{s,n}(x) = \phi_{B_n^{-2}}(x) + \sum_{j=1}^{s-3} \sum_{i=1}^{j} \frac{1}{i!} \sum_{(l_1, \dots, l_i)} \sum_{(\nu_1, \dots, \nu_i)} \frac{\bar{\chi}_{\nu_1!} \cdots \bar{\chi}_{\nu_i!}}{\nu_1! \cdots \nu_i!} (-D)^{\nu} \phi_{B_n^{-2}}(x),$$

where $s \ge 3$, $\sum_{m=1}^{i} l_m = j$, $0 \le l_m \le j$ and ν_i's are nonnegative integral k-vectors such that $|\nu_m| = l_m + 2$ and $\nu = \nu_1 + \cdots + \nu_i$. With $i = (-1)^{1/2}$, write

$$\hat{H}_n(t) = \prod_{j=1}^{n} Ee^{it'Z_{jn}} - \int e^{it'x} \bar{\xi}_{k+4,n}(x)\, dx.$$

As in the proof of Theorem 20.8 of BR(86), it can be shown that there exists a constant $a > 0$ such that for any $b > 0$ and any Borel set B,

$$\begin{aligned} &\left| P(S_n \in B) - \int_B \xi_n(y)\, dy \right| \\ (3.1) \qquad &\le \left| P(S''_n \in B - a_n) - \int_{B - a_n} \bar{\xi}_{k+4,n}(y)\, dy \right| + C\left[\tilde{\alpha}_n \tilde{\gamma}_n + \tilde{\gamma}_n^2 \right] \\ &\le C\left[\max_{|\alpha| \le k+4} \int_{\|t\| < a/b} \left| D^\alpha \hat{H}_n(t) \right| dt + \tilde{\alpha}_n \tilde{\gamma}_n + \tilde{\gamma}_n^2 + \Phi\left((\partial B)^{2b} \right) \right]. \end{aligned}$$

Next, we obtain a bound on the first term above. W.l.o.g. assume that $\lambda_n \geq \lambda > 0$ for all $n \geq 1$. For $i = 1, \dots, n$, define the $k \times m$ matrix M_{in} by the relation $\tilde{U}_{in} = M_{in} Y_{in}$. Then $\Sigma_{i=1}^{n} M_{in} M'_{in} = I_k$. Hence, for any $y \in \mathbb{R}^k$,

$$(3.2) \quad y' V_n Y = \sum_{i=1}^{n} y'[M_{in}] \Sigma_n [M'_{in}] y \geq \sum_{i=1}^{n} \lambda_n y'[M_{in}][M'_{in}] y' \geq \lambda y' y.$$

From (3.2), it follows that

$$(3.3) \qquad \|V_n^{-1/2}\| < \lambda^{-1/2} < \infty.$$

Since $\sup\{E\|Y_{1n}\|^3 \colon n \geq 1\} \leq K$, there exists $a_0 > 0$ such that $|Ee^{i(u'Y_{1n})} - 1| < 1/4$ for all $n \geq 1$, and for all $\|u\| < a_0$, $u \in \mathbb{R}^m$. From this and (3.3), it follows that for all $t \in \mathbb{R}^k$ with $\|t\| < \lambda^{1/2} a_0 / \tilde{\gamma}_{1n}$, and $n \geq 1$,

$$(3.4) \qquad |Ee^{it'U_{jn}} - 1| < 1/4.$$

Also, note that for all $n \geq 1$,

$$(3.5) \qquad \begin{gathered} E\|Z_{jn} - U_{jn}\| \leq 2E\left\{\|U_{jn}\| I(\|U_{jn}\| > 1)\right\} \leq 2E\left\{\|U_{jn}\|^3 I(\|U_{jn}\| > 1)\right\}, \\ \sum_{j=1}^{n} E\|Z_{jn} - U_{jn}\| \leq 2\tilde{\alpha}_n \tilde{\gamma}_n, \qquad \tilde{\gamma}_n^2 \leq k \tilde{\gamma}_{1n}^2 \leq k \tilde{\gamma}_n, \\ \left\| \mathrm{Disp}\left(\sum_{j=1}^{n} Z_{jn} \right) - I \right\| \leq \sum_{i=1}^{n} \|\mathrm{Disp}(Z_{jn}) - \mathrm{Disp}(U_{jn})\| \leq C \tilde{\alpha}_n \tilde{\gamma}_n. \end{gathered}$$

As a result, $(\Sigma_{j=1}^{n} \mathrm{Disp}(Z_{jn}))^{-1/2} = B_n$ exists whenever $C\tilde{\alpha}_n \tilde{\gamma}_n < 1$. Hence, for all such n, $(I + B_n^{-1})$ is p.d., $\|(I + B_n^{-1})^{-1}\| < 1$, and by (3.5),

$$(3.6) \quad \|B_n - I\| \leq \|B_n\| \left\|(I + B_n^{-1})^{-1}\right\| \|B_n^{-2} - I\| \leq C\tilde{\alpha}_n \tilde{\gamma}_n / (1 - C\tilde{\alpha}_n \tilde{\gamma}_n).$$

Next, note that for $s = k + 4$,

$$(3.7) \qquad \left(\sum_{j=1}^{n} E\|B_n Z_{jn}\|^s \right) \leq C\tilde{\alpha}_n \tilde{\gamma}_n / (1 - C\tilde{\alpha}_n \tilde{\gamma}_n)^{s/2}.$$

Using (3.4), (3.6) and (3.7), one can conclude from Theorem 9.9 of BR(86) that there exist constants $C > 0$ and $C_1 > 0$ such that for all n with $2C\tilde{\alpha}_n \tilde{\gamma}_n < 1$ and $|\alpha| \leq s$,

$$(3.8) \qquad \left| D^\alpha \hat{H}_n(t) \right| \leq C\tilde{\alpha}_n \tilde{\gamma}_n \left(\|t\|^{s - |\alpha|} + \|t\|^{3(s-2) + |\alpha|} \right) e^{-\|t\|^2/4},$$

whenever $\tilde{\gamma}_{1n} \|t\|^{s-2} < C_1$, $t \in \mathbb{R}^k$.

Let $\rho_n \equiv (C_1 / \tilde{\gamma}_{1n})^{1/(s-2)}$. From (3.8), it follows that there exists a positive integer N (depending only on $\tilde{\gamma}_n$) such that for all $n \geq N$,

$$(3.9) \qquad \max_{|\alpha| \leq s} \int_{\{|t| \leq \rho_n\}} \left| D^\alpha \hat{H}_n(t) \right| dt \leq C\tilde{\alpha}_n \tilde{\gamma}_n.$$

Next, use (3.5) above and Lemma 14.3 of BR(86) to conclude that

$$\max_{|\alpha| \le s} \int_{\{\rho_n \le \|t\| \le C_0/\tilde{\gamma}_n\}} \left| D^\alpha \prod_{j=1}^{n} g_{jn}(t) \right| dt \le C \tilde{\gamma}_n^2,$$

where $C_0 = \lambda^{3/2}/16kK$. To complete the proof [as in the case of Theorem 20.8 of BR(86)], it is now enough to show that given any $C_2 > 0$, there exists $C_3 \ge 1$, such that for all $n \ge C_3$,

$$\max\left\{ \left| D^\alpha \prod_{j=1}^{n} g_{jn}(t) \right| : |\alpha| \le s,\, C_0 \le \tilde{\gamma}_n \|t\| \le C_2 \right\} \le C_3 \tilde{\gamma}_n^{2+k}.$$

But this follows from (3.3), the assumption on the a_{ij}'s and the condition that

$$\limsup_{n \to \infty} \sup\left\{ |E \exp(it' Y_{1n})| : \tilde{C} C_0 \le \|t\| \le \kappa_n \right\} < 1. \qquad \square$$

LEMMA 3.2. *Let $\{M_{0n}\}_{n \ge 1}, \{M_{in}\}_{n \ge 1}$, $i = 1, \dots, p$, be $(p+1)$ sequences of matrices such that for each $n \ge 1$, M_{0n} is of order $p \times (p + r)$, and M_{in}, $1 \le i \le p$, are of order $(p + r) \times (p + r)$, $p \ge 1$, $r \ge 1$. Let $k = p + r$, $\overline{M}_{0n} = [0\colon I_r]_{r \times k}$ and $\tilde{M}_{0n} = [M'_{0n}\colon \overline{M}'_{0n}]'$. Define the functions $g_n\colon \mathbb{R}^k \to \mathbb{R}^p$ by $g_n(x) = M_{0n}x + (x'M_{1n}x, \dots, x'M_{pn}x)'$, $x \in \mathbb{R}^k$, $n \ge 1$. Assume that:*

(i) *the hypotheses of Lemma* 3.1 *hold with $m = 2$, $p_1 = p$, $p_2 = r$,*
(ii) $\max\{\|M_{in}\|\colon 1 \le i \le p\} = O(\tilde{\gamma}_n)$, *where $\tilde{\gamma}_n$ is as in Lemma* 3.1,
(iii) $\|M_{0n}\| = O(1)$, *and* $\liminf_{n\to\infty} \mathrm{Inf}\{\|\tilde{M}_{0n}u\|\colon \|u\| = 1,\ u \in \mathbb{R}^k\} \ge \rho$ *for some constant $\rho > 0$.*

Then, for any class $\mathscr{B}$ of Borel sets in $\mathbb{R}^p$ satisfying (2.3),

$$\sup_{B \in \mathscr{B}} \left| P(g_n(S_n) \in B) - \int_B \tilde{\xi}_n(x)\, dx \right| = o(\tilde{\gamma}_n),$$

where $S_n = \sum_{i=1}^n U_{in}$, $\tilde{\xi}_n(\cdot) = (1 + \tilde{a}_n(F_0, \cdot))\phi_{\tilde{D}_n}(\cdot)$, $\tilde{D}_n = (\overline{M}_{0n}\overline{M}'_{0n})$ and $\tilde{a}_n(F_0, \cdot)$ is a polynomial. Moreover, $\|a_n(F_0, \cdot)\phi_{\tilde{D}_n}(\cdot)\| = O(\tilde{\gamma}_n)$ and the coefficients of $a_n(F_0, \cdot)$ are continuous functions of $E(Y_1)^\alpha$, $|\alpha| \le 3$.

PROOF. Define the sets $L_n = \{x \in \mathbb{R}^k\colon \|x\| < -\log \tilde{\gamma}_n\}$, $n \ge 1$. Then, from Lemma 3.1, it follows that for any Borel set $B \subseteq \mathbb{R}^p$,

$$P(g_n(S_n) \in B) = \int'_{\{g_n^{-1}B\} \cap L_n} \xi_n(y)\, dy + r_n(B), \tag{3.10}$$

where $|r_n(B)| \le C[\tilde{\alpha}_n \tilde{\gamma}_n + \tilde{\gamma}_n^2 + \Phi((\partial B)^{e_n})]$ and $e_n = o(\tilde{\gamma}_n)$. Now define a function $G_n\colon \mathbb{R}^k \to \mathbb{R}^k$, by $G_n(x) = \binom{g_n(x)}{\tilde{x}}$, where $\tilde{x}$ denotes the last r components of $x \in \mathbb{R}^k$. Note that by condition (iii), grad $G_n(0) = \tilde{M}_{0n}$ is nonsingular. By the usual inverse function theorem, this is enough to guarantee that G_n is one to one (with a differentiable inverse) on some neighborhood N_n of 0. However, this is not strong enough for our purpose. For any $n \ge 1$, to transform the integral on the r.h.s. of (3.10), we at least need G_n to be one to one over all of

L_n. Since N_n also changes with n, it is clear that a more precise definition of N_n is necessary. Using some refinements of Theorems 13.2, 13.4 and 13.6 of Apostol (1974), one can show [cf. Lemmas 3.5 and 3.6 of Lahiri (1990)] that there exist constants $C_1 > 0$, $C_2 > 0$ such that for all $n \geq C_1^{-1}$, G_n has an infinitely differentiable inverse on $N_n \equiv \{x \in \mathbb{R}^k\colon \|x\|\tilde{\gamma}_n \leq \rho^2 C_1\} \supseteq L_n$ and

$$G_n^{-1}(x) = \tilde{M}_{0n}^{-1}x + \sum_{|\nu|=2} (\nu!)^{-1}(D^\nu f_i(0))x^\nu + R_{4n}(x), \tag{3.11}$$

where $G_n^{-1} = (f_1, \dots, f_k)'$ and uniformly in $x \in N_n$, $|R_{4n}(x)| \leq C_2\tilde{\gamma}_n^2(1 + \|x\|^{2k})$.

Now substituting $x = G_n(y)$ in the integral on the r.h.s. of (3.10) and using (3.11), one can complete the proof of Lemma 3.2 as in Lahiri (1989a) or Bhattacharya (1985). □

LEMMA 3.3. *Assume that the conditions of Theorem* 2.1 *hold. Let* $\eta_n^4 \equiv \sum_{i=1}^n \|d_i\|^6$, $A_n \equiv \sum_{i=1}^n d_i d_i' \psi'(\varepsilon_i)$, $n \geq 1$, $A \equiv \tau I_p$ *and* $K \equiv 1 + E|\psi(\varepsilon_1)|^3 + E|\psi'(\varepsilon_1)|^3 + E|\psi''(\varepsilon_1)|^2$. *Then there exists a constant* $C > 0$ *such that*

$$P\left(\left\|\sum_{i=1}^n d_i\psi(\varepsilon_i)\right\| > \nu_n\right) \leq C\left[K\nu_n^{-3}\eta_n + \exp(-C\nu_n^2/K)\right],$$

$$P(\|A_n - A\| > \nu_n\eta_n) \leq C\left[K\nu_n^{-3}\eta_n + \exp(-C\nu_n^2/K)\right]$$

and

$$P\left(\left|\sum_{i=1}^n d_{ij}d_{il}d_{im}(\psi''(\varepsilon_i) - E\psi''(\varepsilon_i))\right| > (\nu_n\eta_n)^{3/2}\right)$$
$$\leq CK\nu_n^{-3}\eta_n \quad \textit{for all } 1 \leq j, l, m \leq p.$$

PROOF. The first two inequalities follow from Fuk and Nagaev (1971), and the last one is a direct consequence of Chebyshev's inequality. □

PROOF OF THEOREM 2.1(a). With $\Delta = (\sum_{i=1}^n x_i x_i')^{1/2}(t - \beta)$, Taylor's expansion of (1.2) gives

$$0 = \sum_{i=1}^n d_i\psi(\varepsilon_i) - \sum_{i=1}^n d_i(d_i'\Delta)\psi'(\varepsilon_i) + \sum_{i=1}^n d_i(d_i'\Delta)^2\psi''(u_i)/2,$$

where u_i is a point between ε_i and $\varepsilon_i - d_i'\Delta$, $1 \leq i \leq n$. Next, note that if $\|A_n - A\| \leq |\tau|/2$, then A_n is invertible and $\|A_n^{-1}\| \leq 2/|\tau|$. In that case one can rewrite the above equation as

$$\Delta = A_n^{-1}\left[\sum_{i=1}^n d_i\psi(\varepsilon_i) + \sum_{i=1}^n d_i(d_i'\Delta)^2\psi''(u_i)/2\right]. \tag{3.12}$$

By Lemma 3.3 and the Lipschitz property of ψ'', it follows that there exist constants $C_1 > 0$, $C_2 > 0$ and a positive integer $N \geq 1$ (depending only on ψ'' and ν_n) such that for all $n \geq N$, outside a set of probability

$C_2\nu_n^{-3}(\sum_{i=1}^n\|d_i\|^6)^{1/4}$, the r.h.s. of (3.12) is less than $C_1\nu_n$, whenever $\|\Delta\| < C_1\nu_n$. Therefore, by Brouwer's fixed point theorem, it follows that on this set there exists $\Delta = \Delta_n$ satisfying (3.12) and $\|\Delta_n\| \le C_1\nu_n$. This proves part (a) of Theorem 2.1. □

PROOF OF THEOREM 2.1(b). Let $l_n = -\log\gamma_n$ and $\theta_n = \sum_{i=1}^n d_i\psi(\varepsilon_i)$, $n \ge 1$. As indicated in Section 1, the proof of part (b) is divided into two steps. First, the stochastic approximation is obtained and then the Edgeworth expansion.

Since Δ_n satisfies (3.12), we can write

$$\Delta_n = \tau^{-1}\theta_n + R_{1n}, \tag{3.13}$$

where, by condition (C.1),

$$\begin{aligned}|R_{1n}| &\le \|A_n^{-1} - A^{-1}\|\,\|\theta_n\| + \tfrac{1}{2}\left\|\sum_{i=1}^n d_i(d_i'\Delta_n)^2\psi''(u_i)\right\|\\ &\le \|A - A_n\|\,\|A_n^{-1}\|\,\|A^{-1}\|\,\|\theta_n\|\\ &\quad + \|\Delta_n\|^2\sum_{i=1}^n\|d_i\|^3\big(|\psi''(\varepsilon_i)| + C(\|d_i\|\,\|\Delta_n\|)^\alpha\big).\end{aligned}$$

Hence, by part (a) of the theorem and Lemma 3.3, there exists a constant $C_3 > 0$ such that for all $n \ge N$,

$$P\big(\|R_{1n}\| > C_3\eta_n\nu_n^2\big) < C_3\eta_n\nu_n^{-3}. \tag{3.14}$$

Now, note that $A_n^{-1} = A^{-1} - A^{-1}(A_n - A)A^{-1} + A^{-1}(A_n - A)A^{-1}(A_n - A)A_n^{-1}$. Therefore, using (3.13), one can write

$$\sum_{i=1}^n d_i(d_i'\Delta_n)^2\psi''(u_i) = \sum_{i=1}^n d_i(d_i'\theta_n/\tau)^2\psi''(\varepsilon_i) + R_{2n}, \tag{3.15}$$

where for all $n \ge N$ (w.l.o.g.) and for some constant $C_4 > 0$, as in (3.14),

$$P\big(\|R_{2n}\| > C_4\eta_n^{1+\alpha}\nu_n^{2+\alpha}\big) \le C_4\eta_n\nu_n^{-3}. \tag{3.16}$$

Next, observe that for $1 \le j \le p$,

$$\begin{aligned}&\left|\sum_{i=1}^n d_{ij}(d_i'\theta_n)^2(\psi''(\varepsilon_i) - E\psi''(\varepsilon_i))\right|\\ &\qquad \le p\|\theta_n\|^2\max\left\{\left|\sum_{i=1}^n d_{ij}d_{ik}d_{im}(\psi''(\varepsilon_i) - E\psi''(\varepsilon_i))\right| : 1 \le k, m \le p\right\}.\end{aligned}$$

Hence, from Lemma 3.3 it follows that for all $n \ge N$ (w.l.o.g.),

$$\begin{aligned}&P\left(\left\|\sum_{i=1}^n d_i(d_i'\theta_n)^2(\psi''(\varepsilon_i) - E\psi''(\varepsilon_i))\right\| > C_5\nu_n^2(\eta_n\nu_n)^{3/2}\right)\\ &\qquad \le C_5\eta_n\nu_n^{-3}.\end{aligned} \tag{3.17}$$

Using (3.12) and (3.14)–(3.17), one can write

$$(3.18)\quad \Delta_n = \tau^{-1}\theta_n - \tau^{-2}(A_n - A)\theta_n + \frac{\tau^{-3}}{2}\left(\sum_{i=1}^{n} d_i(d_i'\theta_n)^2 E\psi''(\varepsilon_i)\right) + R_n,$$

where for all $n \ge N$, the remainder term R_n satisfies

$$(3.19)\qquad P(\|R_n\| > C_6\eta_n^{1+\alpha}\nu_n^{2+\alpha}) \le C_6\eta_n\nu_n^{-3}.$$

This gives the stochastic expansion for Δ_n.

Next, we derive the Edgeworth expansion for Δ_n. Define

$$T_n = \tau^{-1}\theta_n - \tau^{-2}(A_n - A)\theta_n + \left(\sum_{i=1}^{n} d_i(d_i'\theta_n)^2\right)E\psi''(\varepsilon_1)/(2\tau^3)$$

and let T_{jn} denote the jth component of T_n, $j = 1, 2, \ldots, p$. Write $Y_{1i} = \psi(\varepsilon_i)$, $Y_{2i} = \psi'(\varepsilon_i) - \tau$, $i \ge 1$.

It is clear that each T_{jn} is a polynomial in the variables $\Sigma_{i=1}^n d_i Y_{1i}, \Sigma_{i=1}^n c_i Y_{2i}$. So one may try to obtain the Edgeworth expansion for T_n from an expansion for $\Sigma_{i=1}^n d_i Y_{1i}, \Sigma_{i=1}^n c_i Y_{2i}$. The main difficulty in implementing this arises from the fact that the dispersion matrix of $\Sigma_{i=1}^n c_i Y_{2i}$ may be singular, and hence may require a new *singular* linear transformation for each $n \ge 1$.

Let $B \equiv B(n)$ be the $q \times q$ nonsingular matrix defined by (2.1). Note that $\Sigma_{i=1}^n \|\overline{B}_2 c_i\|^2 = \mathrm{tr}\Sigma_{i=1}^n \overline{B} c_i c_i' \overline{B}' = 0$. Hence, by (2.2), $Bc_i = ((\overline{B}_1 c_i)' : (\overline{B}_2 c_i)')' = (b_i' : 0)'$ for all $1 \le i \le n$. Therefore, $c_i = B_1 b_i$, $1 \le i \le n$, where the $q \times r$ matrix B_1 consists of the first r columns of B^{-1}.

Next, we express T_n in terms of $\Sigma_{i=1}^n d_i Y_{1i}$ and $\Sigma_{i=1}^n b_i Y_{2i}$. Denoting $\Sigma_{i=1}^n c_i Y_{2i}$ by $\tilde{Z}$, one can show [as in Lemma 3.4 of Lahiri (1990)] that for $1 \le j \le p$, the jth component of the second term of T_n can be written as $\tau^{-2}\tilde{Z}'\tilde{E}_j\theta_n$, where $\tilde{E}_j$ is a $q \times p$ matrix, and $\|\tilde{E}_j\| \le q$. Let

$$V = \sum_{i=1}^{n} \mathrm{Disp}\begin{pmatrix} d_i Y_{1i} \\ b_i Y_{2i} \end{pmatrix}, \qquad \tilde{X}_i = V^{-1/2}\begin{pmatrix} d_i Y_{1i} \\ b_i Y_{2i} \end{pmatrix}, \qquad 1 \le i \le n.$$

With V_1 of dimension $p \times (p + r)$, partition $V^{1/2}$ as $V^{1/2} = [V_1' : V_2']'$. Then, for $1 \le j \le p$, one can write

$$\tilde{Z}'\tilde{E}_j\theta_n = \left(\sum_{i=1}^{n} b_i Y_{2i}\right)' B_1\tilde{E}_j\left(\sum_{i=1}^{n} d_i Y_{1i}\right) = S_n' V_2' B_1' \tilde{E}_j V_1 S_n,$$

where $S_n = \Sigma_{i=1}^n \tilde{X}_i$. Hence, writing $\tilde{P}_i = V_1' d_i d_i' V_1$, one has

$$(3.20)\quad T_n = V_1 S_n + \tau^{-2}\begin{pmatrix} \vdots \\ S_n' V_2' B_1' \tilde{E}_j V_1 S_n \\ \vdots \end{pmatrix} + (2\tau^3)^{-1} E\psi''(\varepsilon_1)\sum_{i=1}^{n} d_i S_n' \tilde{P}_i S_n.$$

Note that $V_1 V_1' = \sigma^2 I_p$, $\|\Sigma_{i=1}^n d_i \tilde{P}_i\| = O(\gamma_n)$, $\|V_2' B_1' \tilde{E}_j V_1\| = O(\gamma_n)$, $1 \le j \le p$, and $(\Sigma_{i=1}^n \|d_i\|^4)^{1/2} \le p(\Sigma_{i=1}^n \|d_i\|^6)^{1/4}$. Hence, part (b) of the theorem now follows from inequality (3.19), Lemma 3.1 (with $m = 2$, $a_{i1} = d_i$, $a_{i2} = b_i$, $U_i = \tilde{X}_i$, $1 \le i \le n$) and Lemma 3.2. □

PROOF OF THEOREM 2.2. The stochastic approximation for the studentized statistics is derived by using the approximation T_n above. W.l.o.g. assume that $1, \psi(\varepsilon_1), \psi'(\varepsilon_1), (\psi(\varepsilon_1))^2$ are linearly independent. For $n \ge 1$, let $\bar{l}_n = -\log \bar{\gamma}_n$, $\tau_{1n} = (1/n)\sum_{i=1}^n \psi'(\varepsilon_i)$, and $s_{1n}^2 = (1/n)\sum_{i=1}^n \psi^2(\varepsilon_i)$. For real numbers u_i with $|u_i| \le 1$, $1 \le i \le n$, using inequality (3.5), the smoothness of ψ and Hölder's inequality, one gets

$$\begin{aligned} &\left\| \frac{1}{n} \sum_{i=1}^n d_i(\psi''(\varepsilon_i + u_i d_i'\Delta_n) - \psi''(\varepsilon_i)) \right\| \\ &\quad \le C\left(\sum_{i=1}^n \|d_i\|^{2+\alpha} \right)^{(1+\alpha)/(2+\alpha)} n^{-1+1/(2+\alpha)} \|\Delta_n\|^{\alpha}. \end{aligned} \tag{3.21}$$

Next, using (3.21), Theorem 2.1(a), and Chebyshev's inequality, one can show that

$$\tau_n^2 - \tau_{1n}^2 = 2E\psi''(\varepsilon_1) n^{-1}\left(\sum_{i=1}^n d_i'\theta_n \right) + R_{6n},$$

where the remainder term R_{6n} satisfies

$$\begin{aligned} &P\Big(|R_{6n}| > Cn^{-1/2}\big((\nu_n)^{-2} + (\nu_n \log n)^{-1}\big)\Big) \\ &\quad < C\Big[n^{-1/2}(\log n)^{-3} + \bar{\gamma}_n(\bar{l}_n)^{-3}\Big] \end{aligned} \tag{3.22}$$

for some $C > 0$. Next, write

$$\begin{aligned} W_{1i} &= \psi(\varepsilon_i)\psi'(\varepsilon_i), \quad W_{2i} = (\psi'(\varepsilon_i))^2 + |\psi''(\varepsilon_i)\psi(\varepsilon_i)| + |\psi(\varepsilon_i)|, \\ W_{3i} &= |\psi'(\varepsilon_i)\psi''(\varepsilon_i)| + |\psi'(\varepsilon_i)| + (\psi''(\varepsilon_i))^2 + 1, \qquad i \ge 1. \end{aligned}$$

Using (3.13) and Taylor's expansion, one can conclude that

$$s_n^2 - s_{1n}^2 = 2n^{-1}\left(\sum_{i=1}^n d_i'\theta_n \right)(\tau^{-1}EW_{11}) + R_{7n}, \tag{3.23}$$

where on the set $J_n \equiv \{\|\Delta_n\| \, \|d_i\| \le 1 \text{ for all } 1 \le i \le n\}$, the remainder R_{7n} satisfies

$$\begin{aligned} |R_{7n}| \le Cn^{-1}\Bigg(&\left\| \sum_{i=1}^n d_i(W_{1i} - EW_{1i}) \right\| \|\Delta_n\| + \left| \sum_{i=1}^n d_i' R_{1n} EW_{11} \right| \\ &+ \sum_{i=1}^n \|d_i\|^2 \|\Delta_n\|^2 W_{2i} + \sum_{i=1}^n \|d_i\|^3 \|\Delta_n\|^3 W_{3i} \Bigg). \end{aligned}$$

Note that by Hölder's inequality, $E\{|W_{11}|^2 + |W_{21}|^{3/2} + |W_{31}|\} < \infty$. Let $S_{1n} = \sum_{i=1}^n \|d_i\|(W_{1i} - EW_{1i})$, $S_{2n} = \sum_{i=1}^n \|d_i\|^2(W_{2i} I(W_{2i} \le n^{3/2}) - EW_{2i} I(W_{2i} \le n^{3/2}))$ and $S_{3n} = \sum_{i=1}^n \|d_i\|^3 W_{3i}$, $n \ge 1$. By (3.14), Theorem 2.1(a), Corollary

4.4 of Fuk and Nagaev (1971), and Chebyshev's inequality, it follows that

$$
\begin{aligned}
&P\Big(|R_{7n}| > Cn^{-1/2}\big[(\nu_n)^{-2} + (\nu_n \log n)^{-1}\big]\Big)\\
&\quad \le \sum_{j=1}^{3} P\Big(|S_{jn}| > Cn^{1/2}(\nu_n \log n)^{-1}(\nu_n)^{-j}\Big) + nP\big(W_{21} > n^{3/2}\big)\\
&\qquad + P\Big(\|R_{1n}\| > C(\nu_n)^{-2}\Big) + C\bar{\gamma}_n(\nu_n)^{-3}\\
&\quad \le Cn^{-1}(\nu_n \log n)^2\Bigg[(\nu_n)^2\big(EW_{11}^2\big) + n^{3/4}(\nu_n)^4\Bigg(\sum_{i=1}^{n}\|d_i\|^4\Bigg)E|W_{21}|^{3/2}\Bigg]\\
&\qquad + Cn^{-1/2}(\log n)(\nu_n)^4\Bigg(\sum_{i=1}^{n}\|d_i\|^3\Bigg)EW_{3i} + n^{-5/4}E|W_{21}|^{3/2} \qquad (3.24)\\
&\qquad + C\bar{\gamma}_n(\nu_n)^{-3}\\
&\quad \le Cn^{-1}(\log n)^2\Big[(\nu_n)^4 + n^{3/4}(\nu_n)^6(\bar{\gamma}_n)^2 + n^{3/4}(\nu_n)^4(\bar{\gamma}_n)^{3/2}\Big]\\
&\qquad + C\bar{\gamma}_n(\nu_n)^{-3}\\
&\quad = o\Big(n^{-1/2}(\log n)^{-3}\Big) + O\Big(\bar{\gamma}_n\big(\bar{l}_n\big)^{-3}\Big).
\end{aligned}
$$

The last inequality follows from the fact that

$$n^{-1}\sum_{i=1}^{n}\|d_i\|^3 \le \Bigg(n^{-1}\sum_{i=1}^{n}\|d_i\|^4\Bigg)^{3/4} \le Cn^{-3/4}(\bar{\gamma}_n)^{3/2}.$$

Now using (3.18), (3.21) and (3.23), expand the studentized statistic stochastically as

$$
\begin{aligned}
\frac{|\tau_n|}{s_n}\Delta_n &= \sigma^{-1}\Delta_n + \Bigg(\frac{|\tau_n|}{s_n} - \sigma^{-1}\Bigg)\Delta_n\\
&= \sigma^{-1}T_n + \big(2|\tau|s^3\big)^{-1}\theta_n\Big[\big(\tau_n^2 - \tau^2\big)s^2 + \tau^2\big(s^2 - s_n^2\big)\Big] + R_{8n}\\
&= \sigma^{-1}T_n + \big(2|\tau|s^3\big)^{-1}\theta_n\Bigg[2\tau(\tau_{1n} - \tau)s^2 + \tau^2\big(s^2 - s_{1n}^2\big)\\
&\qquad + 2n^{-1}\Bigg(\sum_{i=1}^{n} d_i'\theta_n'\Bigg)\Big(s^2E\psi''(\varepsilon_1) - \tau E\psi(\varepsilon_1)\psi'(\varepsilon_1)\Big)\Bigg] + R_{9n}\\
&= T_{2n} + R_{9n} \quad \text{(say)},
\end{aligned}
$$

where, by (3.19), (3.22) and (3.24), the remainder terms R_{8n} and R_{9n} satisfy

$$
\begin{aligned}
&P\Big(\|R_{8n}\| + \|R_{9n}\| > Cn^{-1/2}\big\{(\nu_n)^{-1} + (\log n)^{-1}\big\}\Big)\\
&\qquad < C\Big[\bar{\gamma}_n\big(\bar{l}_n\big)^{-3} + n^{-1/2}(\log n)^{-3}\Big]
\end{aligned}
$$

for some $C > 0$. Now the Edgeworth expansion for T_{2n} can be derived by using Lemma 3.1 [with $m = 3$, $a_{i1} = d_i$, $a_{i2} = \tilde{b}_i$, $a_{i3} = n^{-1/2}$, $Y_{1i} = \psi(\varepsilon_i)$, $Y_{2i} = \psi'(\varepsilon_i) - E\psi'(\varepsilon_i)$, $Y_{3i} = \psi^2(\varepsilon_i) - E\psi^2(\varepsilon_i)$, $Y_i = (Y_{1i}, Y_{2i}, Y_{3i})'$, $1 \le i \le n$, and so on], Lemma 3.2 and the arguments similar to those used in the proof of Theorem 2.1(b). □

PROOF OF THE PROPOSITION. Let $\gamma_{1n} = \max\{\|b_i\|, \|d_i\|: i = 1, 2, \dots, n\}$. Since $\sum_{i=1}^n d_i d_i' = I_p$ and $\sum_{i=1}^n b_i b_i' = I_r$, it follows that

$$\operatorname{Inf}\left\{\sum_{i=1}^n (d_i' t)^2 : t \in \mathbb{R}^p, \|t\| = 1\right\} = 1$$

and

$$\operatorname{Inf}\left\{\sum_{i=1}^n (b_i' t)^2 : t \in \mathbb{R}^r, \|t\| = 1\right\} = 1.$$

Consequently,

$$\operatorname{Inf}\left\{\sum_{i=1}^n (b_i' t_1)^2 + (d_i' t_2)^2 : \|t_1\|^2 + \|t_2\|^2 = 1, t_1 \in \mathbb{R}^r, t_2 \in \mathbb{R}^p\right\} \ge 1.$$

Hence, for any $0 < \delta < 1$ and $t_1 \in \mathbb{R}^r$, $t_2 \in \mathbb{R}^p$ with $\|t_1\|^2 + \|t_2\|^2 = 1$, the definition of the set $A_n(\delta)$ gives

$$\begin{aligned} 1 &\le \sum_{i=1}^n \left[(b_i' t_1)^2 + (d_i' t_2)^2\right] \\ &= \sum{}' + \sum{}''\left[(b_i' t_1)^2 + (d_i' t_2)^2\right], \quad \text{where } \sum{}' \text{ extends over all } i \in A_n(\delta), \\ &\le K_n(\delta)\gamma_{1n}^2 + (n - K_n(\delta))\delta\gamma_n^2 \le \delta n \gamma_n^2 + (\gamma_{1n}^2 - \delta\gamma_n^2)K_n(\delta). \end{aligned}$$

By the hypothesis of the proposition, $C = \limsup_{n\to\infty} n\gamma_n^2 < \infty$. Also, as in (3.5), $\gamma_{1n}^2 \le \gamma_n$. Hence, it follows that for all large n and for $\delta = (2 + 2C)^{-1}$, $K_n(\delta) \ge (1 - \delta(C + 1))/\gamma_n \ge (2\gamma_n)^{-1}$. Therefore, $-\log \gamma_n / K_n(\delta) = O(\gamma_n |\log \gamma_n|) = o(1)$ as $n \to \infty$, implying condition (C.7). □

PROOF OF THEOREM 2.3. W.l.o.g. assume that $1, \psi(\varepsilon_1), \psi'(\varepsilon_1), \psi^2(\varepsilon_1)$, are linearly independent, and that $\varepsilon_1, \varepsilon_2, \dots$ are defined on the product space $(\mathbb{R}^\infty, \mathscr{R}^\infty)$ such that ε_j is the jth coordinate variable on $(\mathbb{R}^\infty, \mathscr{R}^\infty)$. For notational convenience, we will identify any set $A \in \mathscr{R}^n \equiv$ the Borel σ-field on $\mathbb{R}^n$, with the corresponding cylinder set in $\mathscr{R}^\infty$. Let A_{2n} be the set for which the conclusions of Lemma 3.3 hold. Then, for $(\varepsilon_1, \dots, \varepsilon_n) \in A_{2n}$, $\bar{\beta}_n$ exists and $\|\Delta_n\| \le \nu_n$. Next, we show that the bootstrap moments converge in probability to the corresponding moments of $\{\psi(\varepsilon_1), \psi'(\varepsilon_1), \psi^2(\varepsilon_1)\}$ and $\psi''(\varepsilon_1)$. To that end, write $w_i = |x_{ij}|/p_n$, $1 \le i \le n$. For $a > 0$, let $Y_1(a), Y_2(a), \dots$ denote a generic sequence of iid random variables satisfying $E|Y_1(a)|^a < \infty$. For $(\varepsilon_1, \dots, \varepsilon_n) \in A_{2n}$, using Taylor's expansion and Hölder's inequality, one can show that for all $j, m \in \Gamma \equiv \{(j_1, m_1): 0 \le j_1, m_1 \le j_1 + 2m_1 \le 6\}$,

$$
\begin{aligned}
&\left| E_n\left[\psi(\varepsilon_1^*)^j \psi'(\varepsilon_1^*)^m\right] - \sum_{i=1}^n w_i\left(\psi(\varepsilon_i)^j \psi'(\varepsilon_i)^m\right)\right| \\
(3.25)\qquad &\le C\sum_{i=1}^n w_i \max\Big(|d_i'\Delta_n|^b |Y_i(a)|: (a,b) \\
&\qquad = \left(\tfrac{6}{7},1\right), \left(\tfrac{3}{4},2\right), \left(\tfrac{1}{2},6\right), \left(\tfrac{1}{3},7\right), \left(\tfrac{2}{3},3\right), \left(\tfrac{5}{6},\tfrac{6}{5}\right)\Big)
\end{aligned}
$$

for some $\{Y_i(a)\}$'s. Next, for any $\rho > 0$, $b > 0$, $0 < u < b$ and $0 < a < 1$, with $Z_i(a) = |Y_i(a)| I(w_i \|d_i\|^u |Y_i(a)| \le 1)$, one has

$$
\begin{aligned}
&P\left(\sum_{i=1}^n w_i \|d_i\|^b |Y_i(a)| > \left[2\nu_n^{-\rho} + \sum_{i=1}^n w_i \|d_i\|^b EZ_i(a)\right]\right) \\
&\quad \le P\left(\left|\sum_{i=1}^n w_i \|d_i\|^b (Z_i(a) - EZ_i(a))\right| > \nu_n^{-\rho}\right) + \sum_{i=1}^n P(|Y_i(a)| \ne Z_i(a)) \\
&\quad \le \nu_n^{2\rho} \sum_{i=1}^n w_i^2 \|d_i\|^{2b} EZ_i(a)^2 + \sum_{i=1}^n w_i^a \|d_i\|^{au} E|Y_i(a)|^a \\
&\quad \le C\left(\sum_{i=1}^n w_i^a \|d_i\|^{au}\right) \le C\left(\sum_{i=1}^n \|d_i\|^{au/(1-a)}\right)^{(1-a)}.
\end{aligned}
$$

Note that for each pair (a, b) appearing on the r.h.s. of (3.25),

$$\sum_{i=1}^n w_i \|d_i\|^\alpha EZ_i(a) = o(\nu_n^{-\rho}),$$

$$\left(\sum_{i=1}^n \|d_i\|^{ab/(1-a)}\right)^{(1-a)} = O\left((\bar{\gamma}_n)^{ab+2a-2}\right)$$

and $(ab + 2a - 2) \ge 4/7$. Hence, choosing $0 < u < b$ suitably for each of these pairs, one can show that

$$P\left(\left|E_n\left[\psi(\varepsilon_1^*)^j (\psi'(\varepsilon_1^*))^m\right] - \sum_{i=1}^n w_i \psi(\varepsilon_i)^j \psi'(\varepsilon_i)^m\right| > C\nu_n^{-1}\right) < C(\bar{\gamma}_n)^a$$

for every $0 < a < 4/7$ and $(j, m) \in \Gamma$.

Next, note that by Hölder's inequality, $\{E[\psi(\varepsilon_1)^j \psi'(\varepsilon_1)^m]: (j, :m) \in \Gamma\}$ = the cross-product moments of $\{\psi(\varepsilon_1), \psi'(\varepsilon_1), \psi(\varepsilon_1)^2\}$ of order less than or equal to 3 are finite. Using an argument similar to the above, one can show that for every $(j, m) \in \Gamma$,

$$P\left(\left|\sum_{i=1}^n w_i \psi(\varepsilon_i)^j \psi'(\varepsilon_i)^m - E\left[\psi(\varepsilon_1)^j \psi'(\varepsilon_1)^m\right]\right| > 3\bar{\delta}_n\right) < C\bar{\delta}_n,$$

where $\overline{w}_n \equiv \max\{w_i: 1 \le i \le n\} = o(1)$ and

$$(\bar{\delta}_n)^3 \equiv (\overline{w}_n)^{1/2} + E\psi''(\varepsilon_1)^2 I(\overline{w}_n \psi''(\varepsilon_1)^2 > 1) + \max\Big\{E|\psi(\varepsilon_1)^j \psi'(\varepsilon_1)^m| I\big(\overline{w}_n|\psi(\varepsilon_1)^j\psi'(\varepsilon_1)^m|^2 > 1\big): (j,m) \in \Gamma\Big\}.$$

Hence, $E_n[\psi(\varepsilon_1^*)^j \psi'(\varepsilon_1^*)^m]$, and by similar arguments, $E_n(\psi''(\varepsilon_1^*))^i$ converge in probability to the corresponding population moments. Let $A_{3n} = \{(\varepsilon_1, \dots, \varepsilon_n)$: $|E_n\psi(\varepsilon_1^*)^j\psi'(\varepsilon_1^*)^m - E\psi(\varepsilon_1)^j\psi'(\varepsilon_1)^m| < C(\nu_n^{-1} + \bar{\delta}_n)$, $|E_n\psi''(\varepsilon_1^*)^i - E\psi''(\varepsilon_1)^i| < C(\nu_n^{-1}) + \bar{\delta}_n$: $(j,m) \in \Gamma$, $1 \le i \le 2\}$ for some large $C > 0$. Then

$$P((\varepsilon_1, \dots, \varepsilon_n) \in A_{3n}) = 1 - O(\bar{\delta}_n + (\bar{\gamma}_n)^{1/2}).$$

Next, we check the convergence of $E_n \exp(it'Z_1^*)$, $t \in \mathbb{R}^3$, where $Z_1^* = (\psi(\varepsilon_1^*), \psi'(\varepsilon_1^*), \psi(\varepsilon_1^*)^2)'$. Note that $(\bar{\gamma}_n)^{-1/2} \max\{\|d_i\|: 1 \le i \le n\} \le (\bar{\gamma}_n)^{-1/2}(\sum_{i=1}^n \|d_i\|^6)^{1/6} \le (\bar{\gamma}_n)^{1/6}$. Hence,

$$\begin{aligned} &P\Big(\sup\big\{|E_n \exp(it'Z_1^*) - f_n(t)|: \|t\| < (\bar{\gamma}_n)^{-1/2}\bar{l}_n\big\} > (\bar{l}_n)^{-1}\Big) \\ &\quad \le P\left(\sum_{i=1}^n w_i|d_i\Delta_n|(\bar{\gamma}_n)^{-1/2}\bar{l}_n|Y_i(1)| > (\bar{l}_n)^{-1}\right) \\ &\quad \le P\left(\sum_{i=1}^n w_i|Y_i(1)| > \bar{l}_n\right) + C\bar{\gamma}_n(\bar{l}_n)^{-3} \\ &\quad \le C\big((\bar{\delta}_n)^3 + \bar{\gamma}_n(\bar{l}_n)^{-3}\big), \end{aligned} \tag{3.26}$$

where $f_n(t) = \sum_{j=1}^n w_j \exp(it'Z_j)$ and $Z_j = (\psi(\varepsilon_j), \psi'(\varepsilon_j), \psi(\varepsilon_j)^2)'$, $j \ge 1$. Now using Theorem 2 of Hoeffding (1963) and a discretizing argument as in the proof of Lemma 4.2 of Babu and Singh (1984), one can show that

$$\begin{aligned} &P\Big(\sup\big\{|f_n(t) - E\exp(it'Z_1)|: \|t\|^2 < n\big\} > C(\log\log n)^{-1}\Big) \\ &\quad \le P\left(\sum_{i=1}^n w_i\|Z_i\| > Cn(\log\log n)^{-1}\right) \\ &\qquad + P\Big(\max\big\{|f_n(t) - E\exp(it'Z_1)|: nt \in Z^3, \|t\|^2 < n\big\} > C(\log\log n)^{-1}\Big) \\ &\quad \le CE\|Z_1\|^2 n^{-2} + Cn^5 \exp\left(-C(\log\log n)^{-2}\Big/\sum_{i=1}^n w_i^2\right) \\ &\quad \le Cn^{-2}. \end{aligned} \tag{3.27}$$

Let

$$A_{4n} = \Big\{(\varepsilon_1, \dots, \varepsilon_n): \sup\big\{|E_n \exp(it'Z_1^*) - E\exp(it'Z_1)|: \|t\| < \min\{n^{1/2}, (\bar{\gamma}_n)^{-1/2}\bar{l}_n\}\big\} \le C(\log\log n)^{-1} + (\bar{l}_n)^{-1}\Big\}.$$

Take $A_{1n} = A_{2n} \cap A_{3n} \cap A_{4n}$. Then $P((\varepsilon_1, \dots, \varepsilon_n) \in A_{1n}) > 1 - O(\bar{\delta}_n + (\bar{\gamma}_n)^{1/2} + n^{-2})$. For $(\varepsilon_1, \dots, \varepsilon_n) \in A_{1n}$, one can retrace the proof of Theorem 2.1(a) to prove part (a). As for part (b), note that $\tilde{\gamma}_n \equiv (\sum_{i=1}^n \|d_i\|^4)^{1/4} + (\sum_{i=1}^n \|\tilde{b}_i\|^4)^{1/2} + n^{-1/2} \le \bar{\gamma}_n + n^{-1/2}$ implies $-\tilde{\gamma}_{1n}(\tilde{\gamma}_n)^{-1} \log \tilde{\gamma}_n \le -(\tilde{\gamma}_n)^{-1/2} \log \tilde{\gamma}_n \le \min\{n^{1/2}, (\bar{\gamma}_n)^{-1/2} \bar{l}_n\}$. Hence, the arguments in the proofs of Theorems 2.1(b) and 2.2 and Lemmas 3.1 and 3.2 entail

$$\sup_{B \in \mathscr{B}} \left| P_n\Big((\hat{\sigma}_n^* D)^{-1} (\beta_n^* - \bar{\beta}_n) \in B \Big) - \int_B \xi_n^*(x)\, dx \right| = o\big(\bar{\gamma}_n + n^{-1/2}\big),$$

where

$$\xi_n^*(\cdot) = (1 + \bar{a}_n(F_n; \cdot))\phi(\cdot).$$

Hence, part (b) follows by comparing this with the expansion for the unbootstrapped statistic in Theorem 2.2. □

Proof of Theorem 2.4. Similar to the proof of Theorem 2.3. □

Proof of Theorem 2.5. Here we outline the proof of part (II) only. Specializing the arguments of part (II) with $x_{1j} = 1$ for all $i \ge 1$, one can construct a proof of part (I).

First, we show that

$$(3.28)\quad \begin{aligned} &E_n\Big[\psi(\varepsilon_1^*)^j \psi'(\varepsilon_1^*)^m\Big] \to E\Big[\psi(\varepsilon_1)^j \psi'(\varepsilon_1)^m\Big], &&\text{a.s. for all } i, j \in \Gamma,\\ &E_n\Big[\psi''(\varepsilon_1^*)^j\Big] \to E\Big[\psi''(\varepsilon_1)^j\Big], &&\text{a.s. for } j \le 2. \end{aligned}$$

Note that for any $a > 0$, $E|Y_1(a)|^a < \infty$ implies $|Y_n(a)| = o(n^{1/a})$ a.s. Define $Z_n(a) = Y_n(a) I(|Y_n(a)| < n^{1/a})$, $n \ge 1$, $a > 0$. Then $Y_n(a) = Z_n(a)$, eventually, a.s. by Hölder's inequality, for any $a > 0$ and $b > 0$,

$$\begin{aligned} &E\left(\sum_{i=1}^n w_i \|d_i\|^b (Z_i(a) - EZ_i(a)) \right)^4 \\ &\quad \le Cn^{(4-a)/a} \sum_{i=1}^n w_i^4 \|d_i\|^{4b} + C\left(n^{(2-a)/a} \sum_{i=1}^n w_i^2 \|d_i\|^{4b} \right)^2 \\ &\quad \le Cn^{-4-2b+(4/a)}. \end{aligned}$$

Hence, using the Borel–Cantelli lemma and the fact that $\|\Delta_n\| = O(\nu_n)$ a.s., one can show that $\sum_{i=1}^n w_i |d_i' \Delta_n|^b |Y_i(a)| = \sum_{i=1}^n w_i |d_i' \Delta_n|^b |Z_i(a)| = o(1)$ a.s. for every (a, b) appearing in the r.h.s. of (3.25). To prove the almost sure convergence of the bootstrap moments, it is now enough to show that $\sum_{i=1}^n w_i (Y_i(1) - EY_i(1)) = o(1)$ a.s. W.l.o.g. assume that $EY_1(1) = 0$. Since $\max\{w_i\colon 1 \le i \le n\} = O(n^{-1})$,

$$\begin{aligned} &\sum_{n=1}^\infty p_n^{-2} \operatorname{var}\big(|x_{nj}| Z_n(1) \big) \\ &\quad \le \sum_{n=1}^\infty n^{-2} E|Y_1(1)|^2 I\big(|Y_1(1)| < n\big) \le 2 + E|Y_1(1)| < \infty. \end{aligned}$$

Hence, by standard arguments, $\sum_{n=1}^{\infty} p_n^{-1}|x_{nj}|(Z_n(1) - EZ_n(1)) = O(1)$ a.s., so that $\sum_{i=1}^{n} w_i(Z_n(1) - EZ_n(1)) = o(1)$ a.s. Consequently $\sum_{i=1}^{n} w_i Y_i(1) = o(1)$ a.s.

Next, using (3.26), (3.27) and the Borel–Cantelli lemma, one gets

$$\begin{aligned} &\sup\left\{ |E_n \exp(it'Z_1') - E\exp(it'Z_1)| : \|t\| < \min\left\{n^{1/2}, (\bar{\gamma}_n)^{-1/2}\bar{l}_n\right\}\right\} \\ &\quad \le \left(\sum_{i=1}^{n} w_i |d_i^* \Delta_n| (\bar{\gamma}_n)^{-1/2} \bar{l}_n |Y_i(1)| \right) + O\left((\log\log n)^{-1} \right) \quad \text{a.s.} \qquad (3.29) \\ &\quad = o(1) \quad \text{a.s.} \end{aligned}$$

Now fix a sample point for which (3.28) and (3.29) hold. Then, repeating the arguments in the proofs of Theorems 2.2 and 2.3, one can complete the proof of Theorem 2.5. □

Acknowledgments. The author would like to thank an anonymous referee for a number of constructive suggestions, and for pointing out an oversight in the proof of Lemma 3.1. Special thanks are due to Professor H. L. Koul for his constant encouragement during the preparation of this manuscript.

REFERENCES

APOSTOL, T. M. (1974). *Mathematical Analysis*, 2nd ed. Addison-Wesley, Reading, Mass.

BABU, G. J. and SINGH, K. (1984). On one term Edgeworth correction by Efron's bootstrap. *Sankhyā Ser. A* **46** 219–232.

BHATTACHARYA, R. N. (1985). Some recent results on Cramér–Edgeworth expansions with applications. In *Multivariate Analysis VI* (P. R. Krishnaiah, ed.) 57–75. North-Holland, Amsterdam.

BHATTACHARYA, R. N. and GHOSH, J. K. (1978). On the validity of formal Edgeworth expansion. *Ann. Statist.* **6** 434–451.

BHATTACHARYA, R. N. and RANGA RAO, R. (1986). *Normal Approximations and Asymptotic Expansions*. Krieger, Malabar, Fla.

BICKEL, P. J. and FREEDMAN, D. (1983). Bootstrapping regression models with many parameters. In *A Festschrift for E. L. Lehmann* (P. J. Bickel, K. A. Docksum and J. L. Hodges, Jr., eds.) 28–48. Wadsworth, Belmont, Calif.

FREEDMAN, D. (1981). Bootstrapping regression models. *Ann. Statist.* **9** 1218–1228.

FUK, D. K. and NAGAEV, S. V. (1971). Probability inequalities for sums of independent random variables. *Theory Probabl. Appl.* **16** 643–660.

HOEFFDING, W. (1963). Probability inequalities for sums of independent bounded random variables. *J. Amer. Statist. Assoc.* **58** 13–30.

HUBER, P. (1980). *Robust Statistics*. Wiley, New York.

LAHIRI, S. N. (1989a). On bootstrapping M-estimators. *Sankhyā Ser. A.* To appear.

LAHIRI, S. N. (1989b). Bootstrap approximation and Edgeworth expansion for the distributions of M-estimators of a regression parameter. Preprint 89-36, Dept. Statistics, Iowa State Univ.

LAHIRI, S. N. (1990). Bootstrapping M-estimators of a multiple linear regression parameter. Preprint 90-5, Dept. Statistics, Iowa State Univ.

MAMMEN, E. (1989). Asymptotics with increasing dimension for robust regression with applications to the bootstrap. *Ann. Statist.* **17** 382–400.

NAVIDI, W. (1989). Edgeworth expansions for bootstrapping regression models. *Ann. Statist.* **17** 1472–1478.

PETROV, V. V. (1975). *Sums of Independent Random Variables*. Springer, New York.

QUMSIYEH, M. B. (1990a). Edgeworth expansion in regression models. *J. Multivariate Anal.* **35** 86–101.

QUMSIYEH, M. B. (1990b). Bootstrapping and empirical Edgeworth expansions in multiple linear regression models. Preprint, Dept. Mathematics, Indiana Univ.

SHORACK, G. (1982). Bootstrapping robust regression. *Comm. Statist. Theory Methods* **11** 961–972.

SKOVGAARD, IB. M. (1981). Transformation of an Edgeworth expansion by a sequence of smooth functions. *Scand. J. Statist.* **8** 207–217.

DEPARTMENT OF STATISTICS
SNEDECOR HALL
IOWA STATE UNIVERSITY
AMES, IOWA 50011