ELSEVIER

Statistics & Probability Letters 30 (1996) 271–285

STATISTICS & PROBABILITY LETTERS

Asymptotics of generalized M-estimation of regression and scale with fixed carriers, in an approximately linear model

Douglas P. Wiens [1]

Department of Mathematical Sciences, University of Alberta, Edmonton, Alta, Canada T6G 2G1

Received August 1995; revised October 1995

Abstract

For the approximately linear model $Y_{i,n} = \tilde{\beta}_n^{\mathrm{T}}\mathbf{z}(\mathbf{x}_i) + n^{-1/2}f_n(\mathbf{x}_i) + \varepsilon_i$, with i.i.d. errors ε_i and fixed carriers $\mathbf{z}(\mathbf{x}_i)$, we establish the asymptotic normality of a generalized M-estimator of regression/scale. The estimator minimizes a weighted Huber–Dutter loss function. The function $f_n(\mathbf{x})$ contributes a bias term to the asymptotic normal distribution; apart from this term the estimator is $\sqrt{n}$-equivalent to the estimator obtained assuming the response to be exactly linear. Several estimate/design combinations are compared, in a simulation study.

AMS classifications: Primary 62J05, 62F35; Secondary 62F11, 62F12

Keywords: Generalized M-estimation; Robust regression; Bounded influence; Approximately linear regression response

1. Introduction

In this paper we establish the asymptotic normality of generalized ("bounded influence") M-estimators of regression and scale, assuming the carriers to be fixed rather than random, and without assuming the linear model to be exactly correct for finite samples.

Silvapullé (1985) proved the asymptotic normality of ordinary M-estimators in the fixed carriers case, assuming the linear model $y_i = \mathbf{x}_i^{\mathrm{T}}\beta + \varepsilon_i$ to be correct. Maronna and Yohai (1981) established the asymptotic normality of GM estimates with random carriers, and without the assumption of exact linearity. There is thus a gap in the literature, to which this paper is aimed.

Our model of approximate linearity is as follows. Suppose that one observes $\{(Y_{i,n}, \mathbf{x}_i)\colon i = 1,\dots,n\}$ obeying

$$Y_{i,n} = g_n(\mathbf{x}_i) + \varepsilon_i,$$

where $\varepsilon_1,\dots,\varepsilon_n$ are i.i.d. errors, $\mathbf{x}_1,\dots,\mathbf{x}_n$ are fixed design points chosen within a design space $S \subseteq \mathbb{R}^q$, and g_n is some scalar valued function. This function is thought to be approximately linear:

$$g_n(\mathbf{x}) \approx \beta_n^{\mathrm{T}}\mathbf{z}(\mathbf{x})$$

[1] Research supported by the Natural Sciences and Engineering Research Council of Canada.

SSDI 0167-7152(95)00230-8

for some $p \times 1$ vector $\mathbf{z}(\mathbf{x})$ of regressors. The "correct" vector $\tilde{\beta}_n$ of regression parameters may be defined by

$$\min_{\beta_n} \int_S \{g_n(\mathbf{x}) - \beta_n^{\mathrm{T}}\mathbf{z}(\mathbf{x})\}^2\, \mathrm{d}\mathbf{x} = \int_S \{g_n(\mathbf{x}) - \tilde{\beta}_n^{\mathrm{T}}\mathbf{z}(\mathbf{x})\}^2\, \mathrm{d}\mathbf{x}. \tag{1.1}$$

Then if we introduce $f_n(\mathbf{x}) = \sqrt{n}\{g_n(\mathbf{x}) - \tilde{\beta}_n^{\mathrm{T}}\mathbf{z}(\mathbf{x})\}$ we obtain

$$Y_{i,n} = \tilde{\beta}_n^{\mathrm{T}}\mathbf{z}(\mathbf{x}_i) + n^{-1/2}f_n(\mathbf{x}_i) + \varepsilon_i, \qquad i = 1,\ldots,n; \tag{1.2}$$

$$\int_S \mathbf{z}(\mathbf{x})f_n(\mathbf{x})\,\mathrm{d}\mathbf{x} = \mathbf{0}. \tag{1.3}$$

We will impose conditions below which ensure that $f_n(\mathbf{x})$ is well behaved as $n \to \infty$; in this sense the conjecture of linearity of $g_n(\mathbf{x})$ is asymptotically correct.

Condition (1.1) was introduced only as one possible way of motivating the model – our derivations assume only (1.2). Even without (1.1), (1.3) may be assumed without loss of generality, and it then ensures that $\boldsymbol{\beta}_n$ is well defined as long as the matrix $\int_S \mathbf{z}(\mathbf{x})\mathbf{z}^{\mathrm{T}}(\mathbf{x})\,\mathrm{d}\mathbf{x}$ is non-singular.

The estimate $\hat{\boldsymbol{\theta}}_n = (\hat{\boldsymbol{\beta}}_n^{\mathrm{T}}, \hat{\sigma}_n)^{\mathrm{T}}$ is the nth member of any sequence satisfying

$$n^{-1/2}\sum_{i=1}^{n} \mathbf{U}_i(\hat{\boldsymbol{\theta}}_n) \xrightarrow{\mathrm{pr}} \mathbf{0},$$

where

$$\mathbf{U}_i(\boldsymbol{\theta}) = \begin{pmatrix} \psi_i\left(\dfrac{Y_{i,n} - \boldsymbol{\beta}^{\mathrm{T}}\mathbf{z}(\mathbf{x}_i))}{\sigma}\right)\mathbf{z}(\mathbf{x}_i) \\ \chi_i\left(\dfrac{Y_{i,n} - \boldsymbol{\beta}^{\mathrm{T}}\mathbf{z}(\mathbf{x}_i)}{\sigma}\right) - \dfrac{A_n}{n} \end{pmatrix}$$

for a bounded sequence of positive constants $n^{-1}A_n$. The function $\psi_i = \psi_{i,n}$ is weakly increasing and bounded, and satisfies further conditions given in Section 2. Boundedness is assumed for mathematical convenience and for robustness; our final results clearly hold for the least squares estimator, with the unbounded and non-robust $\psi(x) = x$. Following Huber and Dutter (1974) and Huber (1981) we take $\chi_i(r) = \int_0^r t\,\mathrm{d}\psi_i(t)$. Examples are given by

$$\psi_i(r) = w_n(\mathbf{x}_i)\psi(r/v_n(\mathbf{x}_i)) \tag{1.4}$$

for a sufficiently smooth function ψ and functions $w_n(\mathbf{x}), v_n(\mathbf{x})$ weighting the independent variables. With $v_n(\mathbf{x}) \equiv 1$, (1.4) describes a Mallows-type GM estimate (Hill, 1977). Schweppe (Merrill and Schweppe, 1971) proposed $v_n(\mathbf{x}) \equiv w_n(\mathbf{x})$. Corresponding to these two proposals, one-step estimates for exactly linear models have been investigated by Simpson et al. (1992) and Coakley and Hettmansperger (1993).

For a sequence $\{\tilde{\sigma}_n\}$ defined in Assumption (B) below, we set

$$\mathbf{V}_i = \begin{pmatrix} \psi_i\left(\dfrac{\varepsilon_i}{\tilde{\sigma}_n}\right)\mathbf{z}(\mathbf{x}_i) \\ \chi_i\left(\dfrac{\varepsilon_i}{\tilde{\sigma}_n}\right) - \dfrac{A_n}{n} \end{pmatrix},$$

$$Q_n = n^{-1}\sum_{i=1}^{n} \mathrm{cov}[\mathbf{V}_i],$$

$$\mathbf{b}_n = (n\tilde{\sigma}_n)^{-1}\sum_{i=1}^{n}\begin{pmatrix} E\left[\psi_i'\left(\dfrac{\varepsilon}{\tilde{\sigma}_n}\right)\right]\mathbf{z}(\mathbf{x}_i)f_n(\mathbf{x}_i) \\ E\left[\left(\dfrac{\varepsilon}{\tilde{\sigma}_n}\right)\psi_i'\left(\dfrac{\varepsilon}{\tilde{\sigma}_n}\right)\right]f_n(\mathbf{x}_i)\end{pmatrix}, \tag{1.5}$$

$$M_n = (n\tilde{\sigma}_n)^{-1}\sum_{i=1}^{n}\begin{pmatrix} E\left[\psi_i'\left(\dfrac{\varepsilon}{\tilde{\sigma}_n}\right)\right]\mathbf{z}(\mathbf{x}_i)\mathbf{z}^{\mathrm{T}}(\mathbf{x}_i) & E\left[\left(\dfrac{\varepsilon}{\tilde{\sigma}_n}\right)\psi_i'\left(\dfrac{\varepsilon}{\tilde{\sigma}_n}\right)\right]\mathbf{z}(\mathbf{x}_i) \\ E\left[\left(\dfrac{\varepsilon}{\tilde{\sigma}_n}\right)\psi_i'\left(\dfrac{\varepsilon}{\tilde{\sigma}_n}\right)\right]\mathbf{z}^{\mathrm{T}}(\mathbf{x}_i) & E\left[\left(\dfrac{\varepsilon}{\tilde{\sigma}_n}\right)^2\psi_i'\left(\dfrac{\varepsilon}{\tilde{\sigma}_n}\right)\right]\end{pmatrix}.$$

Put $\tilde{\boldsymbol{\theta}}_n = (\tilde{\boldsymbol{\beta}}_n^{\mathrm{T}}, \tilde{\sigma}_n)^{\mathrm{T}}$. In Section 3 we prove:

Theorem 1. *Under assumptions* (A)–(E) *of Section* 2 *below, we have*
(a) *Asymptotic expansion*:

$$\sqrt{n}(\hat{\boldsymbol{\theta}}_n - \tilde{\boldsymbol{\theta}}_n) = M_n^{-1}\mathbf{b}_n + M_n^{-1}n^{-1/2}\sum_{i=1}^{n}\mathbf{V}_i + o_p(1).$$

(b) *Asymptotic normality*:

$\sqrt{n}(\hat{\boldsymbol{\theta}}_n - \tilde{\boldsymbol{\theta}}_n)$ *is* $AN(M_n^{-1}\mathbf{b}_n, M_n^{-1}Q_nM_n^{-1})$.

Our assumptions are detailed in Section 2. The results of a simulation study are reported in Section 4.

2. Preliminaries

Define

$$\Theta = \{\boldsymbol{\theta} = (\boldsymbol{\beta}^{\mathrm{T}}, \sigma)^{\mathrm{T}}:\ \boldsymbol{\beta} \in \mathbb{R}^p,\ \sigma > 0\}.$$

For $\boldsymbol{\theta} \in \Theta$ define

$$r_{i,n}(\boldsymbol{\theta}) = (Y_{i,n} - \boldsymbol{\beta}^{\mathrm{T}}\mathbf{z}(\mathbf{x}_i))/\sigma.$$

Let $\psi_i(r)$ be a weakly increasing function with two continuous derivatives, satisfying conditions stated below. (For asymptotic results under weaker smoothness conditions, see He and Wang (1995) and He and Shao (1995)). For $\rho_i(r) = \int_0^r \psi_i(t)\,\mathrm{d}t$, define

$$L_n(\boldsymbol{\theta}) = \sigma\sum_{i=1}^{n}(\rho_i(r_{i,n}(\boldsymbol{\theta})) + n^{-1}A_n).$$

The gradient and Hessian of $L_n(\boldsymbol{\theta})$ are

$$L_n'(\boldsymbol{\theta}) = -\sum_{i=1}^{n}\begin{pmatrix}\psi_i(r_{i,n}(\boldsymbol{\theta}))\mathbf{z}(\mathbf{x}_i) \\ \chi_i(r_{i,n}(\boldsymbol{\theta})) - n^{-1}A_n\end{pmatrix},$$

$$L_n''(\boldsymbol{\theta}) = \sigma^{-1}\sum_{i=1}^{n}\psi_i'(r_{i,n}(\boldsymbol{\theta}))\begin{pmatrix}\mathbf{z}(\mathbf{x}_i) \\ r_{i,n}(\boldsymbol{\theta})\end{pmatrix}\begin{pmatrix}\mathbf{z}(\mathbf{x}_i) \\ r_{i,n}(\boldsymbol{\theta})\end{pmatrix}^{\mathrm{T}},$$

so that L_n is convex. Assume

(A) The function $\rho_i(r)$ has three continuous derivatives, and $E[\rho_i(r_{i,n}(\boldsymbol{\theta}))]$ may be differentiated under the integral sign three times with respect to $\boldsymbol{\theta}$.

Define

$$\overline{L}_n(\boldsymbol{\theta}) = E[L_n(\boldsymbol{\theta})],$$

$$\mathbf{U}_i(\boldsymbol{\theta}) = \frac{-\partial}{\partial \boldsymbol{\theta}} \sigma\{\rho_i(r_{i,n}(\boldsymbol{\theta})) + n^{-1}A_n\},$$

$$T_n(\boldsymbol{\theta}) = -n^{-1/2} \sum_{i=1}^{n} \mathbf{U}_i(\boldsymbol{\theta}) = n^{-1/2} L_n'(\boldsymbol{\theta}),$$

$$\overline{T}_n(\boldsymbol{\theta}) = E[T_n(\boldsymbol{\theta})] = n^{-1/2}\, \overline{L}_n'(\boldsymbol{\theta})$$

$$B_n(\boldsymbol{\theta}) = n^{-1} L_n''(\boldsymbol{\theta}),$$

$$\overline{B}_n(\boldsymbol{\theta}) = n^{-1}\, \overline{L}_n''(\boldsymbol{\theta}),$$

$$= (n\sigma)^{-1} \sum_{i=1}^{n} \begin{pmatrix} E[\psi_i'(r_{i,n}(\boldsymbol{\theta}))]\mathbf{z}(\mathbf{x}_i)\mathbf{z}(\mathbf{x}_i)^{\mathrm{T}} & E[r_{i,n}(\boldsymbol{\theta})\psi_i'(r_{i,n}(\boldsymbol{\theta}))]\mathbf{z}(\mathbf{x}_i) \\ E[r_{i,n}(\boldsymbol{\theta})\psi_i'(r_{i,n}(\boldsymbol{\theta}))]\mathbf{z}(\mathbf{x}_i)^{\mathrm{T}} & E[r_{i,n}^2(\boldsymbol{\theta})\psi_i'(r_{i,n}(\boldsymbol{\theta}))] \end{pmatrix},$$

$$C_n(\boldsymbol{\theta}) = \operatorname{cov}[T_n(\boldsymbol{\theta})] = n^{-1} \sum_{i=1}^{n} \operatorname{cov}[\mathbf{U}_i(\boldsymbol{\theta})].$$

Let $\{\hat{\boldsymbol{\theta}}_n\}, \{\overline{\boldsymbol{\theta}}_n\}$ be sequences asymptotically minimizing $n^{-1/2}L_n(\boldsymbol{\theta})$, $n^{-1/2}\overline{L}_n(\boldsymbol{\theta})$, respectively, so that

$$T_n(\hat{\boldsymbol{\theta}}_n) \xrightarrow{\text{pr}} \mathbf{0}, \qquad \overline{T}_n(\overline{\boldsymbol{\theta}}_n) \to \mathbf{0}. \tag{2.1}$$

Assume

(B) The errors are centred so as to satisfy the condition

$$n^{-1/2} \sum_{i=1}^{n} E\left[\psi_i\left(\frac{\varepsilon}{\tilde{\sigma}_n}\right)\right] \mathbf{z}(\mathbf{x}_i) \to \mathbf{0},$$

where $\{\tilde{\sigma}_n\}$ is a sequence of positive constants satisfying

$$n^{-1/2} \sum_{i=1}^{n} \left\{ E\left[\chi_i\left(\frac{\varepsilon}{\tilde{\sigma}_n}\right)\right] - n^{-1}A_n \right\} \to 0.$$

See Carroll and Welsch (1988) for a discussion related to Assumption (B). Similar to Proposition 1 in Silvapullé (1985) we have that a sufficient condition for $\{\tilde{\sigma}_n\}$ to exist and be bounded away from zero as $n \to \infty$ is that the largest jump η in the error distribution satisfy the following assumption.

(C) $\eta < \liminf_{n\to\infty}(1 - (n^{-1}A_n/v_n))$, where

$$v_n = \min\left(n^{-1} \sum_{i=1}^{n} \chi_i(\infty), n^{-1} \sum_{i=1}^{n} \chi_i(-\infty)\right).$$

Apart from the term $\mathbf{b}_n$ defined in Section 1, $\hat{\boldsymbol{\theta}}_n$ turns out to be $\sqrt{n}$-equivalent to the estimator obtained when $f_n \equiv 0$. Thus, define

$$\tilde{Y}_{i,n} = Y_{i,n} - n^{-1/2} f_n(\mathbf{x}_i) = \tilde{\boldsymbol{\beta}}_n \mathbf{z}(\mathbf{x}_i) + \varepsilon_i,$$

$$\tilde{\boldsymbol{\theta}}_n = (\tilde{\boldsymbol{\beta}}_n^{\mathrm{T}},\ \tilde{\sigma}_n)^{\mathrm{T}},\ \tilde{r}_{i,n}(\boldsymbol{\theta}) = (\tilde{Y}_{i,n} - \boldsymbol{\beta}^{\mathrm{T}}\mathbf{z}(\mathbf{x}_i))/\sigma,$$

$$\tilde{L}_n(\boldsymbol{\theta}) = \sigma\left\{\sum_{i=1}^{n}(E[\rho_i(\tilde{r}_{i,n}(\boldsymbol{\theta}))] + n^{-1}A_n)\right\},$$

$$\tilde{T}_n(\boldsymbol{\theta}) = n^{-1/2}\tilde{L}_n'(\boldsymbol{\theta}), \qquad \tilde{B}_n(\boldsymbol{\theta}) = n^{-1}\tilde{L}_n''(\boldsymbol{\theta}),$$

$$\tilde{C}_n(\boldsymbol{\theta}) = n^{-1}\sum_{i=1}^{n} \operatorname{cov}\left[\begin{pmatrix} \psi_i(\tilde{r}_{i,n}(\boldsymbol{\theta}))\mathbf{z}(\mathbf{x}_i) \\ \chi_i(\tilde{r}_{i,n}(\theta)) - n^{-1}A_n \end{pmatrix}\right].$$

By (A) and (B),

$$\tilde{T}_n(\tilde{\boldsymbol{\theta}}_n) \to \mathbf{0}. \tag{2.2}$$

Assume

(D) The eigenvalues of

$$M_n := \tilde{B}_n(\tilde{\boldsymbol{\theta}}_n) \quad \text{and} \quad Q_n := \tilde{C}_n(\tilde{\boldsymbol{\theta}}_n)$$

are bounded above, and away from zero, as $n \to \infty$.

Now define

$$d_{ijk} = \sup_r \left| r^k \frac{\mathrm{d}^j}{\mathrm{d}r^j}\psi_i(r)\right|.$$

Assume:

(E) (i) The following are O(1) as $n \to \infty$:

$$n^{-1}\sum_{i=1}^{n} f_n^2(\mathbf{x}_i), \qquad n^{-1}\sum_{i=1}^{n} d_{i00}^2\|\mathbf{z}(\mathbf{x}_i)\|^2, \qquad n^{-1}\sum_{i=1}^{n} d_{i01}^2,$$

$$n^{-1}\sum_{i=1}^{n} d_{i12}, \qquad n^{-1}\sum_{i=1}^{n} d_{i12}\|\mathbf{z}(\mathbf{x}_i)\|^2, \qquad n^{-1}\sum_{i=1}^{n} d_{i23}\|\mathbf{z}(\mathbf{x})\|^2,$$

$$n^{-1}\sum_{i=1}^{n} d_{ijk}, \qquad n^{-1}\sum_{i=1}^{n} d_{ijk}\|\mathbf{z}(\mathbf{x}_i)\|^3, \qquad n^{-1}\sum_{i=1}^{n} d_{ijk} f_n^2(\mathbf{x}_i),$$

$$n^{-1}\sum_{i=1}^{n} d_{ijk}\|\mathbf{z}(\mathbf{x}_i)\| f_n^2(\mathbf{x}_i), \qquad n^{-1}\sum_{i=1}^{n} d_{ijk}\|\mathbf{z}(\mathbf{x}_i)\|^2 |f_n(\mathbf{x}_i)| \quad \text{for } 0 \leqslant k \leqslant j,\ j = 1, 2.$$

(ii) The following are o(1) as $n\to\infty$:

$$n^{-1/2}\max_{1\leqslant i\leqslant n} d_{ijk},\quad j=0,1,2,\ k\geqslant 0,\ |k-j|\leqslant 1;$$

$$n^{-1/2}\max_{1\leqslant i\leqslant n} d_{ijk}\|\mathbf{z}(\mathbf{x}_i)\|,\quad (j,k)=(0,0),(0,1),(1,0),(1,1).$$

For ψ_i as at (1.4) we have the simpler

(E′) (i) $r^k\frac{d^j}{dr^j}\psi(r)$ is bounded for $0\leqslant k\leqslant j+1\leqslant 3$.
(ii) The following are O(1) as $n\to\infty$:

$$n^{-1}\sum_{i=1}^{n} f_n^2(\mathbf{x}_i),\qquad n^{-1}\sum_{i=1}^{n} w_n^2(\mathbf{x}_i)\|\mathbf{z}(\mathbf{x}_i)\|^2,\qquad n^{-1}\sum_{i=1}^{n} w_n^2(\mathbf{x}_i)v_n^2(\mathbf{x}_i),\qquad n^{-1}\sum_{i=1}^{n} w_n(\mathbf{x}_i)v_n(\mathbf{x}_i)\|\mathbf{z}(\mathbf{x}_i)\|^2,$$

and for $\ell=0,1,2$ and $u_{n,\ell}(\mathbf{x}_i):=w_n(\mathbf{x}_i)/v_n^{\ell}(\mathbf{x}_i)$,

$$n^{-1}\sum_{i=1}^{n} u_{n,\ell}(\mathbf{x}_i),\qquad n^{-1}\sum_{i=1}^{n} u_{n,\ell}(\mathbf{x}_i)\|\mathbf{z}(\mathbf{x}_i)\|^3,\qquad n^{-1}\sum_{i=1}^{n} u_{n,\ell}(\mathbf{x}_i)f_n^2(\mathbf{x}_i),$$

$$n^{-1}\sum_{i=1}^{n} u_{n,\ell}(\mathbf{x}_i)\|\mathbf{z}(\mathbf{x}_i)\|f_n^2(\mathbf{x}_i),\qquad n^{-1}\sum_{i=1}^{n} u_{n,\ell}(\mathbf{x}_i)\|\mathbf{z}(\mathbf{x}_i)\|^2|f_n(\mathbf{x}_i)|.$$

(iii) The following are o(1) as $n\to\infty$, for $\ell=-1,0,1$:

$$n^{-1/2}\max_{1\leqslant i\leqslant n} u_{n,\ell}(\mathbf{x}_i),\qquad n^{-1/2}\max_{1\leqslant i\leqslant n} u_{n,\ell}(\mathbf{x}_i)\|\mathbf{z}(\mathbf{x}_i)\|.$$

3. Derivations

Before proving Theorem 1, we establish a number of lemmas. Throughout this section we write $\mathbf{z}_i$ for $\mathbf{z}(\mathbf{x}_i)$.

Lemma 1. *As* $n\to\infty$, $\overline{B}_n(\tilde{\boldsymbol\theta}_n)-\tilde{B}_n(\tilde{\boldsymbol\theta}_n)\to 0$.

Proof. Note that

$$r_{i,n}(\tilde{\boldsymbol\theta}_n)=\tilde{r}_{i,n}(\tilde{\boldsymbol\theta}_n)+n^{-1/2}\tilde\sigma_n^{-1}f_n(\mathbf{x}_i).$$

For $k,\ell\leqslant p$ and $\lambda_i\in[0,1]$ we then have, by Taylor's Theorem,

$$\begin{aligned}|\overline{B}_n(\tilde{\boldsymbol\theta}_n)-\tilde{B}_n(\tilde{\boldsymbol\theta}_n)|_{k,\ell}&=(n\tilde\sigma_n)^{-1}\left|\sum_{i=1}^{n}E[\psi_i'(r_{i,n}(\tilde{\boldsymbol\theta}_n))-\psi_i'(\tilde{r}_{i,n}(\tilde{\boldsymbol\theta}_n))]z_{ik}z_{i\ell}\right|\\&=(n\tilde\sigma_n)^{-1}\left|\sum_{i=1}^{n}E[\psi_i''(\tilde{r}_{i,n}(\tilde{\boldsymbol\theta}_n)+n^{-1/2}\tilde\sigma_n^{-1}\lambda_i f_n(\mathbf{x}_i))]f_n(\mathbf{x}_i)z_{ik}z_{i\ell}/\sqrt{n}\tilde\sigma_n\right|\\&\leqslant n^{-3/2}\tilde\sigma_n^{-2}\sum_{i=1}^{n}d_{i20}|f_n(\mathbf{x}_i)||z_{ik}||z_{i\ell}|.\end{aligned}$$

Thus

$$\sum_{k,\ell=1}^{p}|\overline{B}_n(\tilde{\boldsymbol\theta}_n)-\tilde{B}_n(\tilde{\boldsymbol\theta}_n)|_{k,\ell}\leqslant pn^{-3/2}\tilde\sigma_n^{-2}\sum_{i=1}^{n}d_{i20}|f_n(\mathbf{x}_i)|\|\mathbf{z}_i\|^2,$$

which is o(1) by assumptions (C) and (E). Similarly,

$$\sum_{\ell=1}^{p} |\overline{B}_n(\tilde{\boldsymbol{\theta}}_n) - \tilde{B}_n(\tilde{\boldsymbol{\theta}}_n)|_{p+1,\ell} \leqslant p^{1/2} n^{-3/2} \tilde{\sigma}_n^{-2} \sum_{i=1}^{n} d_{i21} |f_n(\mathbf{x}_i)| \|\mathbf{z}_i\|$$

and

$$|\overline{B}_n(\tilde{\boldsymbol{\theta}}_n) - \tilde{B}_n(\tilde{\boldsymbol{\theta}}_n)|_{p+1,p+1} \leqslant n^{-3/2} \tilde{\sigma}_n^{-2} \sum_{i=1}^{n} d_{i22} |f_n(\mathbf{x}_i)|$$

are both o(1) as $n \to \infty$. □

The linearizations presented below in Lemmas 2 and 5 are similar to that in Theorem 1 of Silvapullé (1985).

Lemma 2. *$\overline{T}_n(\tilde{\boldsymbol{\theta}}_n + n^{-1/2}\boldsymbol{\gamma}) = \overline{T}_n(\tilde{\boldsymbol{\theta}}_n) + \tilde{B}_n(\tilde{\boldsymbol{\theta}}_n)\boldsymbol{\gamma} + p_n(\boldsymbol{\gamma})$, where $\sup\{\|p_n(\boldsymbol{\gamma})\| : \|\boldsymbol{\gamma}\| \leqslant K\} \to 0$ as $n \to \infty$, for any $K > 0$.*

Proof. By Taylor's theorem,

$$\overline{T}_n(\tilde{\boldsymbol{\theta}}_n + n^{-1/2}\boldsymbol{\gamma}) = \overline{T}_n(\tilde{\boldsymbol{\theta}}_n) + \overline{B}_n(\tilde{\boldsymbol{\theta}}_n)\boldsymbol{\gamma} + n^{-3/2}\overline{R}_n(\boldsymbol{\gamma}, \boldsymbol{\lambda}),$$

where $\overline{R}_n(\boldsymbol{\gamma}, \boldsymbol{\lambda})$ has ℓth element

$$\overline{R}_{n,\ell}(\boldsymbol{\gamma}, \lambda_\ell) = \boldsymbol{\gamma}^{\mathrm{T}} \left[\frac{\partial^2}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^{\mathrm{T}}} \overline{L}'_{n,\ell}(\tilde{\boldsymbol{\theta}}_n + n^{-1/2}\lambda_\ell\boldsymbol{\gamma}) \right] \boldsymbol{\gamma}$$

and $\lambda_\ell \in [0,1]$. For $\|\boldsymbol{\gamma}\| \leqslant K$ we have

$$|\overline{R}_{n,\ell}(\boldsymbol{\gamma}, \lambda_\ell)| \leqslant K^2 \sum_{a,b=1}^{p+1} \left| \frac{\partial^2}{\partial\theta_a\partial\theta_b} \overline{L}'_{n,\ell}(\boldsymbol{\theta}_*) \right|_{\boldsymbol{\theta}_* = \tilde{\boldsymbol{\theta}}_n + n^{-1/2}\lambda_\ell\boldsymbol{\gamma}}$$

and so it suffices to show that this last term is $o(n^{3/2})$ and to then apply Lemma 1.

Write r_i for $r_{i,n}(\boldsymbol{\theta}_*)$. Then the terms of $(\partial^2/\partial\theta_a\partial\theta_b)L'_{n,\ell}(\boldsymbol{\theta}_*)$ are given by:

	$\ell \leqslant p$	$\ell = p+1$
$a, b \leqslant p$	$-\sigma_*^{-2} \sum_1^n E[\psi_i''(r_i)] z_{ia} z_{ib} z_{i\ell}$	$-\sigma_*^{-2} \sum_1^n E[\psi_i'(r_i) + r_i\psi_i''(r_i)] z_{ia} z_{ib}$
$a = p+1, b \leqslant p$	$-\sigma_*^{-2} \sum_1^n E[\psi_i'(r_i) + r_i\psi_i''(r_i)] z_{i\ell} z_{ib}$	$-\sigma_*^{-2} \sum_1^n E[2r_i\psi_i'(r_i) + r_i^2\psi_i''(r_i)] z_{ib}$
$a = b = p+1$	$-\sigma_*^{-2} \sum_1^n E[r_i\psi_i'(r_i) + r_i^2\psi_i''(r_i] z_{il}$	$-\sigma_*^{-2} \sum_1^n E[r_i^2(3\psi_i'(r_i) + r_i\psi_i''(r_i))]$

Elementary bounds then give

$$n^{-3/2}|\overline{R}_{n,\ell}(\boldsymbol{\gamma}, \lambda_\ell)| \leqslant n^{-3/2} K^2 \sigma_*^{-2} \sum_{i=1}^{n} \{ p^{3/2} d_{i20} \|\mathbf{z}_i\|^3 + 3p(d_{i10} + d_{i21}) \|\mathbf{z}_i\|^2 + \sqrt{p}(5d_{i11} + 3d_{i22}) \|\mathbf{z}_i\| + 3d_{i12} + d_{i23} \},$$

which is o(1) by (C) and (E). □

Lemma 3. (a) $\tilde{\gamma}_n := \sqrt{n}(\overline{\boldsymbol{\theta}}_n - \tilde{\boldsymbol{\theta}}_n)$ *is* O(1) *as* $n \to \infty$.
(b) $\overline{\sigma}_n$ *is bounded away from zero as* $n \to \infty$.

Proof. The proof of (a) is based on the expansion of Lemma 2. It is very similar to that on p. 1493 of Silvapullé (1985) and so is omitted. By (a), $\overline{\sigma}_n - \tilde{\sigma}_n \to 0$ so that (b) then follows upon applying (C). □

Lemma 4. (a) $B_n(\overline{\boldsymbol{\theta}}_n) - \overline{B}_n(\overline{\boldsymbol{\theta}}_n) \xrightarrow{\text{pr}} 0$.
(b) $\overline{B}_n(\overline{\boldsymbol{\theta}}_n) - \overline{B}_n(\tilde{\boldsymbol{\theta}}_n) \to 0$.
(c) $B_n(\overline{\boldsymbol{\theta}}_n) - \tilde{B}_n(\tilde{\boldsymbol{\theta}}_n) \xrightarrow{\text{pr}} 0$.
(d) $C_n(\overline{\boldsymbol{\theta}}_n) - \tilde{C}_n(\overline{\boldsymbol{\theta}}_n) \to 0$.
(e) *The eigenvalues of* $\overline{B}_n(\overline{\boldsymbol{\theta}}_n)$ *and* $C_n(\overline{\boldsymbol{\theta}}_n)$ *are bounded above, and away from zero, as* $n \to \infty$.

Proof. (a) Put $\|B\| = (\text{tr } BB^{\mathrm{T}})^{1/2}, \mathbf{c}_i = (\mathbf{z}_i^{\mathrm{T}}, r_{i,n}(\overline{\boldsymbol{\theta}}_n))^{\mathrm{T}}$. Then

$$E[\|B_n(\overline{\boldsymbol{\theta}}_n) - \overline{B}_n(\overline{\boldsymbol{\theta}}_n)\|^2] = \sum_{k,\ell=1}^{p+1} \text{VAR}[B_n(\overline{\boldsymbol{\theta}}_n)_{k,\ell}] \leqslant n^{-2}\overline{\sigma}_n^{-2} \sum_{k,\ell=1}^{p+1} \sum_{i=1}^{n} E[\{\psi_i'(r_{i,n}(\overline{\boldsymbol{\theta}}_n))c_{ik}c_{i\ell}\}^2]$$
$$= n^{-2}\overline{\sigma}_n^{-2} \sum_{i=1}^{n} E[\{\psi_i'(r_{i,n}(\overline{\boldsymbol{\theta}}_n))\}^2\{r_{i,n}^2(\overline{\boldsymbol{\theta}}_n) + \|\mathbf{z}_i\|^2\}^2]$$
$$\leqslant n^{-2}\overline{\sigma}_n^{-2} \sum_{i=1}^{n} (d_{i12}^2 + 2d_{i11}^2\|\mathbf{z}_i\|^2 + d_{i10}^2\|\mathbf{z}_i\|^4).$$

This last term is $o_p(1)$ by (E) and Lemma 3(b).
(b) Write $\overline{\boldsymbol{\theta}}_n = \tilde{\boldsymbol{\theta}}_n + n^{-1/2}\tilde{\gamma}_n$, where $\tilde{\gamma}_n$ is bounded by Lemma 3(a). Then

$$(\overline{B}_n(\overline{\boldsymbol{\theta}}_n) - \overline{B}_n(\tilde{\boldsymbol{\theta}}_n))_{k,\ell} = \left[\frac{\partial}{\partial \boldsymbol{\theta}} \overline{B}_{n,k,\ell}(\tilde{\boldsymbol{\theta}}_n + n^{-1/2}\lambda_{n,k,\ell}\,\tilde{\gamma}_n)\right]\tilde{\gamma}_n \Big/ \sqrt{n},$$

where $\lambda_{n,k,\ell} \in [0,1]$. Now proceed as in Lemma 2.
(c) Immediate from (a), (b) and Lemma 1.
(d) First note that

$$\|C_n(\overline{\boldsymbol{\theta}}_n) - \tilde{C}_n(\overline{\boldsymbol{\theta}}_n)\| \leqslant n^{-1}\sum_{i=1}^{n}\Big\{\|\mathbf{z}_i\|^2|\text{VAR}[\psi_i(r_{i,n}(\overline{\boldsymbol{\theta}}_n))] - \text{VAR}[\psi_i(\tilde{r}_{i,n}(\overline{\boldsymbol{\theta}}_n))]|$$
$$+2\|\mathbf{z}_i\|\,|\text{cov}[\psi_i(r_{i,n}(\overline{\boldsymbol{\theta}}_n)), \chi_i(r_{i,n}(\overline{\boldsymbol{\theta}}_n))] - \text{cov}[\psi_i(\tilde{r}_{i,n}(\overline{\boldsymbol{\theta}}_n)), \chi_i(\tilde{r}_{i,n}(\overline{\boldsymbol{\theta}}_n))]|$$
$$+|\text{VAR}[\chi_i(r_{i,n}(\overline{\boldsymbol{\theta}}_n))] - \text{VAR}[\chi_i(\tilde{r}_{i,n}(\overline{\boldsymbol{\theta}}_n))]|\Big\}. \quad (3.1)$$

Write

$$r_{i,n}(\overline{\boldsymbol{\theta}}_n) = \tilde{r}_{i,n}(\overline{\boldsymbol{\theta}}_n) + n^{-1/2}t_{in},$$

where $t_{in} = f_n(\mathbf{x}_i)/\overline{\sigma}_n$. We then find that the first difference of variances in (3.1) may be bounded by

$$4d_{i00}d_{i10}t_{in}/\sqrt{n} + 2d_{i10}^2t_{in}^2/n,$$

that the difference of covariances may be bounded by

$$2(d_{i10}d_{i01} + d_{i00}d_{i11}t_{in})/\sqrt{n} + 2d_{i10}d_{i11}t_{in}^2/n$$

and that the final difference of variances may be bounded by

$$4d_{i01}d_{i11}t_{in}/\sqrt{n} + 2d_{i11}^2t_{in}^2/n.$$

By (E) and Lemma 3(b), $C_n(\overline{\boldsymbol{\theta}}_n) - \tilde{C}_n(\overline{\boldsymbol{\theta}}_n) \to 0$. Now

$$[\tilde{C}_n(\overline{\boldsymbol{\theta}}_n) - \tilde{C}_n(\tilde{\boldsymbol{\theta}}_n)]_{k,\ell} = \left[\frac{\partial}{\partial\boldsymbol{\theta}}\tilde{C}_{n,k,\ell}(\tilde{\boldsymbol{\theta}}_n + n^{-1/2}\lambda_{n,k,\ell}\tilde{\boldsymbol{\gamma}}_n)\right]\tilde{\boldsymbol{\gamma}}_n\Big/\sqrt{n}$$

is shown in a similar fashion to be $o_p(1)$.

(e) Immediate from (a)–(d), using (D). □

Lemma 5. *$T_n(\overline{\boldsymbol{\theta}}_n + n^{-1/2}\boldsymbol{\gamma}) = T_n(\overline{\boldsymbol{\theta}}_n) + \overline{B}_n(\overline{\boldsymbol{\theta}}_n)\boldsymbol{\gamma} + q_n(\boldsymbol{\gamma})$, where $\sup\{\|q_n(\boldsymbol{\gamma})\| : \|\boldsymbol{\gamma}\| \leqslant K\} \to 0$ a.s. as $n \to \infty$, for any $K > 0$.*

Proof. Similar to the proof of Lemma 2. Expand by Taylor's theorem, then use Lemma 3(b), Lemma 4(a) and (E). □

Lemma 6. (a) *$\overline{\boldsymbol{\gamma}}_n := \sqrt{n}(\hat{\boldsymbol{\theta}}_n - \overline{\boldsymbol{\theta}}_n)$ is $O_p(1)$ as $n \to \infty$.*

(b) *$\tilde{\boldsymbol{\gamma}}_n - M_n^{-1}\mathbf{b}_n \to \mathbf{0}$ as $n \to \infty$.*

Proof. (a) The proof is similar to that of Lemma 3(a), but is based on Lemma 5 rather than Lemma 2.

(b) In Lemma 2, replace $\boldsymbol{\gamma}$ by $\tilde{\boldsymbol{\gamma}}_n$ to obtain

$$\overline{T}_n(\overline{\boldsymbol{\theta}}_n) = \overline{T}_n(\tilde{\boldsymbol{\theta}}_n) + \tilde{B}_n(\tilde{\boldsymbol{\theta}}_n)\tilde{\boldsymbol{\gamma}}_n + p_n(\tilde{\boldsymbol{\gamma}}_n).$$

By (2.1) and Lemma 3(a),

$$\tilde{\boldsymbol{\gamma}}_n + \tilde{B}_n^{-1}(\tilde{\boldsymbol{\theta}}_n)\overline{T}_n(\tilde{\boldsymbol{\theta}}_n) \to \mathbf{0}. \tag{3.2}$$

Now expand $\psi_i(r_{i,n}(\tilde{\boldsymbol{\theta}}_n)) = \psi_i(\varepsilon_i/\tilde{\sigma}_n + s_{in})$, where $s_{in} = n^{-1/2}f_n(\mathbf{x}_i)/\tilde{\sigma}_n$, around $s_{in} = 0$ and substitute into the definition of $\overline{T}_n(\tilde{\theta}_n)$ to get

$$\overline{T}_n(\tilde{\theta}_n) = \tilde{T}_n(\tilde{\boldsymbol{\theta}}_n) - \mathbf{b}_n - n^{-3/2}\tilde{R}_n(\boldsymbol{\lambda}, \boldsymbol{\mu}, \mathbf{s})/2\tilde{\sigma}_n^{\,2}, \tag{3.3}$$

where

$$\tilde{R}_n(\boldsymbol{\lambda}, \boldsymbol{\mu}, \mathbf{s}) = \sum_{i=1}^{n} \begin{pmatrix} E[\psi_i''(\frac{\varepsilon_i}{\tilde{\sigma}_n} + \lambda_{in}s_{in})]\mathbf{z}_i \\ E[\chi_i''(\frac{\varepsilon_i}{\tilde{\sigma}_n} + \mu_{in}s_{in})] \end{pmatrix} f_n^2(\mathbf{x}_i)$$

and $\lambda_{in}, \mu_{in} \in [0,1]$. By (E), $n^{-3/2}\tilde{R}_n(\boldsymbol{\lambda}, \boldsymbol{\mu}, \mathbf{s}) \to \mathbf{0}$. This, together with (2.2) and (3.3) gives

$$\overline{T}_n(\tilde{\boldsymbol{\theta}}_n) = -\mathbf{b}_n + o(1)$$

which, with (3.2), completes the proof. □

Lemma 7. *$T_n(\overline{\boldsymbol{\theta}}_n) = -n^{-1/2}\sum_{i=1}^{n}\mathbf{V}_i + o_p(1)$ as $n \to \infty$.*

Proof. Define

$$t_{in} = f_n(\mathbf{x}_i) - \sqrt{n}(\bar{\boldsymbol{\beta}}_n - \tilde{\boldsymbol{\beta}}_n)^{\mathrm{T}}\mathbf{z}_i, \qquad s_n = \sqrt{n}(\bar{\sigma}_n - \tilde{\sigma}_n),$$
$$u_{in} = t_{in} - \varepsilon_i s_n/\tilde{\sigma}_n = f_n(\mathbf{x}_i) - (\mathbf{z}_i^{\mathrm{T}}, \varepsilon_i/\tilde{\sigma}_n)\tilde{\boldsymbol{\gamma}}_n.$$

Then

$$-T_n(\bar{\boldsymbol{\theta}}_n) - n^{-1/2}\sum_{i=1}^n \mathbf{V}_i = n^{-1/2}\sum_{i=1}^n \begin{pmatrix} (\psi_i((Y_{i,n} - \bar{\boldsymbol{\beta}}_n^{\mathrm{T}}\mathbf{z}_i)/\bar{\sigma}_n) - \psi_i(\varepsilon_i/\tilde{\sigma}_n))\mathbf{z}_i \\ \chi_i((Y_{i,n} - \bar{\boldsymbol{\beta}}_n^{\mathrm{T}}\mathbf{z}_i)/\bar{\sigma}_n)) - \chi_i(\varepsilon_i/\tilde{\sigma}_n) \end{pmatrix}$$

$$= n^{-1/2}\sum_{i=1}^n \begin{pmatrix} (\psi_i((\varepsilon_i + n^{-1/2}t_{in})/(\tilde{\sigma}_n + n^{-1/2}s_n)) - \psi_i(\varepsilon_i/\tilde{\sigma}_n))\mathbf{z}_i \\ \chi_i((\varepsilon_i + n^{-1/2}t_{in})/(\tilde{\sigma}_n + n^{-1/2}s_n)) - \chi_i(\varepsilon_i/\tilde{\sigma}_n) \end{pmatrix}$$

$$= (n\tilde{\sigma}_n)^{-1}\sum_{i=1}^n \begin{pmatrix} \psi_i'(\varepsilon_i/\tilde{\sigma}_n)\mathbf{z}_i \\ \chi_i'(\varepsilon_i/\tilde{\sigma}_n) \end{pmatrix} u_{in} + 2n^{-3/2}D_n(\boldsymbol{\lambda}, s_n, \mathbf{t}_n), \tag{3.4}$$

where $D_n(\boldsymbol{\lambda}, s_n, \mathbf{t}_n)$ has ℓth element defined as follows.

$$1 \leqslant \ell \leqslant p: \quad D_{n,\ell}(\boldsymbol{\lambda}, s_n, \mathbf{t}_n) = \sum_{i=1}^n (t_{in}, s_n)H(\psi_i, \boldsymbol{\lambda}_\ell, s_n, \mathbf{t}_n)\begin{pmatrix} t_{in} \\ s_n \end{pmatrix}\mathbf{z}_i,$$

$$D_{n,p+1}(\boldsymbol{\lambda}, s_n, \mathbf{t}_n) = \sum_{i=1}^n (t_{in}, s_n)H(\chi_i, \boldsymbol{\lambda}_\ell, s_n, \mathbf{t}_n)\begin{pmatrix} t_{in} \\ s_n \end{pmatrix};$$

$$\boldsymbol{\lambda}_\ell = (\lambda_{\ell 1}, \lambda_{\ell 2}), \lambda_{\ell j} \in [0,1];$$

$$H(\psi_i, \boldsymbol{\lambda}, s_n, \mathbf{t}_n) = (\tilde{\sigma}_n + n^{-1/2}\lambda_{\ell 2}s_n)^{-2}\begin{pmatrix} \psi_i''(r_i) & -r_i\psi_i''(r_i) - \psi_i'(r_i) \\ -r_i\psi_i''(r_i) - \psi_i'(r_i) & r_i^2\psi_i''(r_i) + 2r_i\psi_i'(r_i) \end{pmatrix},$$

$$r_i = (\varepsilon_i + n^{-1/2}\lambda_{\ell 1}t_{in})/(\tilde{\sigma}_n + n^{-1/2}\lambda_{\ell 2}s_n).$$

It now suffices to show that:

(i) $\operatorname{cov}\left[(n\tilde{\sigma}_n)^{-1}\sum_{i=1}^n \begin{pmatrix} \psi_i'(\varepsilon_i/\tilde{\sigma}_n)\mathbf{z}_i \\ \chi_i'(\varepsilon_i/\tilde{\sigma}_n) \end{pmatrix} u_{i,n}\right] \to 0,$

(ii) $n^{-3/2}D_{n,\ell}(\boldsymbol{\lambda}, s_n, \mathbf{t}_n) \xrightarrow{\mathrm{pr}} 0$ for each ℓ.

Then by (3.4),

$$-T_n(\bar{\boldsymbol{\theta}}_n) - n^{-1/2}\sum_{i=1}^n \mathbf{V}_i = E\left[(n\tilde{\sigma}_n)^{-1}\sum_{i=1}^n \begin{pmatrix} \psi_i'(\varepsilon_i/\tilde{\sigma}_n)\mathbf{z}_i \\ \chi_i'(\varepsilon_i/\tilde{\sigma}_n) \end{pmatrix} u_{i,n}\right] + o_p(1) = \mathbf{b}_n - M_n\tilde{\boldsymbol{\gamma}}_n + o_p(1)$$

and this last term is o(1) by Lemma 6(b).

For (i) the inequality $\operatorname{Var}[X - Y] \leqslant 2E[X^2 + Y^2]$ may be used to bound the trace of the covariance matrix by

$$2n^{-2}\tilde{\sigma}_n^{-2}\sum_{i=1}^n \{d_{i10}^2\|\mathbf{z}_i\|^2 f_n^2(\mathbf{x}_i) + \|\tilde{\boldsymbol{\gamma}}_n\|^2(\|z_i\|^4 d_{i10}^2 + \|z_i\|^2 d_{i11}^2) + d_{i11}^2 f_n^2(\mathbf{x}_i) + \|\tilde{\boldsymbol{\gamma}}_n\|^2(\|\mathbf{z}_i\|^2 d_{i11}^2 + d_{i12}^2)\},$$

which is o(1) by (E).

For (ii), denote $H(\psi_i, \boldsymbol{\lambda}_\ell, s_n, \mathbf{t}_n)$ by H_i. For sufficiently large n we have

$$\|H_i\| \leqslant 2\tilde{\sigma}_n^{-2} \left\| \begin{pmatrix} d_{i20} & d_{i21}+d_{i10} \\ d_{i21}+d_{i10} & d_{i22}+2d_{i11} \end{pmatrix} \right\| \leqslant 4\tilde{\sigma}_n^{-2}(d_{i20}+d_{i21}+d_{i10}+d_{i22}+2d_{i11}).$$

Then for $1 \leqslant \ell \leqslant p$,

$$\begin{aligned} n^{-3/2} D_{n,\ell}(\boldsymbol{\lambda}, s_n, \mathbf{t}_n) &\leqslant n^{-3/2} \sum_{i=1}^n (t_{in}^2 + s_n^2)\|H_i\|\|\mathbf{z}_i\| \\ &\leqslant 2n^{-3/2} \sum_{i=1}^n (f_n^2(\mathbf{x}_i) + \|\tilde{\boldsymbol{\gamma}}_n\|^2(1+\|\mathbf{z}_i\|^2))\|H_i\|\|\mathbf{z}_i\|), \end{aligned}$$

which is o(1) by (E). The case $\ell = p+1$ is similar. □

Proof of Theorem 1. (a) By Lemmas 5 and 6 with $\boldsymbol{\gamma} = \bar{\boldsymbol{\gamma}}_n$, then by Lemma 4 (c) and (2.1),

$$\bar{\boldsymbol{\gamma}}_n = (\overline{B}_n(\bar{\boldsymbol{\theta}}_n))^{-1}(T_n(\hat{\boldsymbol{\theta}}_n) - T_n(\bar{\boldsymbol{\theta}}_n) - q_n(\bar{\boldsymbol{\gamma}}_n)) = -M_n^{-1} T_n(\bar{\boldsymbol{\theta}}_n) + o_p(1).$$

Now write $\sqrt{n}(\hat{\boldsymbol{\theta}}_n - \tilde{\boldsymbol{\theta}}_n) = \bar{\boldsymbol{\gamma}}_n + \tilde{\boldsymbol{\gamma}}_n$ and apply Lemmas 6(b) and 7.

(b) We must show that

$$Q_n^{-1/2} n^{-1/2} \sum_{i=1}^n \mathbf{V}_i \xrightarrow{d} N(\mathbf{0}, I).$$

Denote the smallest eigenvalue of a matrix by $ch_{\min}$. Note that for any $\boldsymbol{\lambda} \in \mathbb{R}^{p+1}$ with $\|\boldsymbol{\lambda}\| = 1$ we have

$$\begin{aligned} n^{-3/2} \sum_{i=1}^n E[|\boldsymbol{\lambda}^{\mathrm{T}} Q_n^{-1/2}(\mathbf{V}_i - E[\mathbf{V}_i])|^3] &\leqslant n^{-3/2}(ch_{\min}(Q_n))^{-3/2} \sum_{i=1}^n E[\|\mathbf{V}_i - E[\mathbf{V}_i]\|^3] \\ &\leqslant 2n^{-3/2}(ch_{\min}(Q_n))^{-3/2} \sum_{i=1}^n (d_{i00}^3 \|\mathbf{z}_i\|^3 + d_{i01}^3), \end{aligned}$$

which is o(1) by (D) and (E). This implies the Lindeberg–Feller condition (see, e.g. Serfling (1980, p. 30) with $X_i = \mathbf{t}^{\mathrm{T}}\mathbf{V}_i$ for any $\mathbf{t}$; then apply the Cramér–Wold theorem) and so

$$Q_n^{-1/2} \left\{ n^{-1/2} \sum_{i=1}^n \mathbf{V}_i - E\left[n^{-1/2} \sum_{i=1}^n \mathbf{V}_i \right] \right\} \xrightarrow{d} N(\mathbf{0}, I).$$

This, with (B), completes the proof. □

4. Simulations

We have simulated samples satisfying

$$y_i = 1 + x_i + n^{-1/2} f(x_i) + \varepsilon_i, \quad i = 1, \ldots, n, \; -\tfrac{1}{2} \leqslant x_i \leqslant \tfrac{1}{2},$$

where:

(i) $n = 20$ or $n = 40$.

(ii) 80% of the ε_i were $N(0,1)$, the remaining 20% were $N(0, \sigma^2 = 25)$.

(iii) $f(x) = \sqrt{5}(6x^2 - 0.5)$. Then (1.3) holds and

$$\int_{-1/2}^{1/2} f^2(x)\,\mathrm{d}x = 1. \tag{4.1}$$

(iv) The x_i were chosen according to one of three designs:
(a) "Classical" (C): $\frac{1}{2}n$ designs points at each of $\pm\frac{1}{2}$.
(b) "Uniform" (U): $x_i = \{(i-0.5)/n\} - 0.5$.
(c) "Minimax" (M): Huber (1975) derived a design density which minimizes the maximum integrated mean squared error (IMSE) of the fitted values, when the estimates are obtained by least squares. The maximum is evaluated subject to (1.2) and (1.3), and is attained at a quadratic f as in (iii) above, if the errors are homoscedastic. When as well $\sigma^2 = 1$ and (4.1) holds, the minimax density is given by

$$m_0(x) = a + 12(1-a)x^2, \quad \text{with } a = 0.68217.$$

We have taken $x_i = M_0^{-1}((i-0.5)/n)$, where M_0 is the corresponding distribution function.
For these designs, assumptions (E′) (ii), (iii) are easily verified.
(v) Three types of estimates were computed:
(a) W2: a Schweppe estimate – as at (1.4) with $w_n(x_i) = v_n(x_i) = \sqrt{1-h_{ii}}$, where $\{h_{ii}\}$ is the diagonal of the hat matrix. See Handshin et al. (1975) and Huber (1983).
(b) W1: a Mallows estimate – as at (1.4) with $v_n(x_i) = 1$, $w_n(x_i) = \sqrt{1-h_{ii}}$.
(c) W0: a Huber estimate – as at (1.4) with $w_n(x_i) = v_n(x_i) = 1$.

In each case we used Huber's $\psi(x) = \text{sign }(x)\cdot\min(k,|x|)$ with $k = 1.5$. Although this ψ does not satisfy all of our assumptions, it can be approximated arbitrarily closely by sufficiently smooth functions and hence serves as a test of the sensitivity of Theorem 1 to these assumptions. Corresponding to this ψ,

$$\chi_i(r) = 0.5(v(x_i)/w(x_i))\psi_i^2(r) = 0.5v(x_i)w(x_i)\psi^2(r/v(x_i)).$$

In analogy with classical least squares, and for Fisher consistency at the normal distribution, we took

$$A_n = ((n-p)/n)\sum_i E_\Phi[\chi_i(\varepsilon)].$$

The computing algorithm consisted of alternating between a regression step – weighted least squares with weights $u_i = \psi_i(r_i)/r_i$ – and a scale step:

$$\sigma^2 := \sigma^2 \sum_i \chi_i(r_i)/A_n$$

until convergence was reached. For each set of n random errors, all nine design/estimate combinations were computed. This was repeated 1000 times, for each value of n.

The assumption of a symmetric error distribution implies (B), and that the regression and scale estimates are asymptotically independent. With Z representing the $n \times p$ design matrix we then have

$$Q_n = Z^{\mathrm{T}} D_Q Z/n \oplus d_Q/n,$$

$$M_n = Z^{\mathrm{T}} D_M Z/n \oplus d_M/n,$$

where D_Q and D_M are diagonal matrices with

$$D_{Q,ii} = E\left[\psi_i^2\left(\frac{\varepsilon}{\tilde{\sigma}_n}\right)\right], \qquad D_{M,ii} = E\left[\psi_i'\left(\frac{\varepsilon}{\tilde{\sigma}_n}\right)\right],$$

and

$$d_Q = \sum_i E\left[\left(\chi_i\left(\frac{\varepsilon}{\tilde{\sigma}_n}\right) - n^{-1}A_n\right)^2\right],$$

$$d_M = \sum_i E\left[\left(\frac{\varepsilon}{\tilde{\sigma}_n}\right)^2 \psi_i'\left(\frac{\varepsilon}{\tilde{\sigma}_n}\right)\right].$$

After computing W2, these quantities were estimated from the residuals $r_i = r_i(\hat{\boldsymbol{\theta}})$ using

$$\widehat{D}_{Q,ii} = \psi_i^2(r_i), \qquad \widehat{D}_{M,ii} = \psi_i'(r_i),$$

$$\hat{d}_Q/n = \underset{i}{\text{var}}\,(\chi_i(r_i)), \qquad \hat{d}_M/n = \underset{i}{\text{aver}}\,(r_i^2\psi_i'(r_i)).$$

Here, var denotes the usual sample variance. For W1 and W0 we used

$$\widehat{D}_{Q,ii} = w_n^2(x_i)\left(\sum_j \psi^2(r_j)/(n-p)\right), \qquad \widehat{D}_{M,ii} = w_n(x_i)\cdot \underset{j}{\text{aver}}\,(\psi'(r_j)),$$

$$\hat{d}_Q/n = 0.25\,\underset{i}{\text{aver}}\,w_n^2(x_i)\cdot \underset{i}{\text{var}}\,\psi^2(r_i), \qquad \hat{d}_M/n = \underset{i}{\text{aver}}\,w_n(x_i)\cdot \underset{i}{\text{aver}}(r_i^2\psi'(r_i)).$$

Some numerical results are presented in Table 1. There, the term *bias*$(\hat{\theta})$ refers to the simulated-sample estimate of $E[\hat{\theta}] - \tilde{\theta}$, viz. $\sum_{j=1}^{1000}(\hat{\theta}_j - 1)/1000$. The *var. ratio* for an estimate $\hat{\theta}$ is the ratio of two estimates of the variance of $\hat{\theta}$. The numerator of this ratio is obtained from the average of the 1000 estimates of the asymptotic covariance matrix. The denominator is obtained from the sample variances of the 1000 simulated values of $\hat{\theta}$. The quantity $n \cdot IMSE$ is an estimate of

$$n\int_{-1/2}^{1/2} E[\{\hat{y}(x) - E[y(x)]\}^2]\,\mathrm{d}x = n\cdot \text{VAR}[\hat{\beta}_0] + (n/12)\cdot \text{VAR}[\hat{\beta}_1] + n\cdot bias^2(\hat{\beta}_0) + (n/12)\cdot bias^2(\hat{\beta}_1) + \int f^2(x)\,\mathrm{d}x,$$

obtained from the means and variances of the 1000 simulated values, and (4.1).

For all three designs the estimators perform quite similarly. For C the three regression/scale estimates are in fact identical – the methods differ only in the manner in which the asymptotic covariance matrix is estimated. Note also the huge, but not unexpected, biases of the intercept estimates for design C. For M and U, both W1 and W2 are slightly superior to W0 in terms of bias, somewhat more so with respect to IMSE.

In all cases, the estimated variance of the scale estimate was too large. The var. ratios for the intercept estimate tended to be closer to unity when W2 was used, while those for the slope tended to be farther from unity. An attempt was made to correct the estimated covariance matrices of the regression estimates, by multiplying them by K^2, where $K = 1+\{p(1-m)/mn,\}$ and $m = \text{aver}(\psi_i'(r_i))$ for W2, $= \text{aver}(\psi'(r_i))$ for W1 and W0. (See Huber, 1981, p. 173 ff.). This resulted in slight improvements in the estimates of the variances of the slope estimates, when $n = 40$. In all other cases, it tended to move the var. ratios farther away from unity. Similarly, a suggestion of Simpson et al. (1992) – multiply the covariances by $\{U/(U - p)\}$, where U is the number of non-zero regression weights u_i – proved to be deleterious in all but one of the 18 cases.

The empirical quantiles of $\hat{y}(1/4) = \hat{\beta}_0 + 0.25\hat{\beta}_1$ were compared with those of the Student's t distribution on $0.6n$ degrees of freedom. This follows a suggestion in Field (1982). For $n = 20$ there is very close agreement, while for $n = 40$ the degrees of freedom could perhaps be increased. The Q–Q plots are available from the author.

Table 4.1
Biases, variance ratios and integrated MSE values based on 1000 simulated samples, for each of three designs and three types of estimates

		$n = 20$			$n = 40$		
Design ↓	Estimate→	W0	W1	W2	W0	W1	W2
M	$n \cdot \text{bias}^2$ (intercept)	0.054	0.045	0.052	0.049	0.045	0.048
	$n \cdot \text{bias}^2$ (slope)	0.054	0.054	0.052	0.005	0.005	0.005
	var. ratio (intercept)	1.053	1.055	1.021	1.022	1.023	1.006
	var. ratio (slope)	1.042	1.045	1.053	0.979	0.980	0.978
	var. ratio (scale)	1.569	1.579	1.598	1.504	1.508	1.509
	$n \cdot$ IMSE	5.199	5.184	5.174	4.919	4.913	4.908
U	$n \cdot \text{bias}^2$ (intercept)	0.003	0.006	0.003	0.004	0.006	0.004
	$n \cdot \text{bias}^2$ (slope)	0.068	0.066	0.062	0.008	0.007	0.008
	var. ratio (intercept)	1.050	1.053	1.018	1.021	1.022	1.005
	var. ratio (slope)	1.033	1.039	1.071	0.975	0.976	0.985
	var. ratio (scale)	1.554	1.578	1.593	1.499	1.498	1.508
	$n \cdot$ IMSE	5.624	5.617	5.585	5.324	5.323	5.308
C	$n \cdot \text{bias}^2$ (intercept)	4.791	4.791	4.791	4.731	4.731	4.731
	$n \cdot \text{bias}^2$ (slope)	0.047	0.047	0.047	0.000	0.000	0.000
	var. ratio (intercept)	1.032	1.032	1.003	1.010	1.010	0.994
	var. ratio (slope)	1.021	1.020	0.991	0.985	0.985	0.970
	var. ratio (scale)	1.634	1.634	1.634	1.552	1.552	1.552
	$n \cdot$ IMSE	8.801	8.801	8.801	8.530	8.530	8.530

A question deserving further investigation, and suggested by these comparisons, concerns the interplay between the choice of design points, and the choice of weights $w_n(\mathbf{x}_i)$ for W2 and W1. The derivation of suitable correction factors for the variance/covariance estimates, and especially for the variance of the scale estimate, is another area requiring more research.

References

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

Coakley, W.C. and T.P. Hettmansperger (1993), A bounded influence, high breakdown, efficient regression estimator, *J. Amer. Statistic. Assoc.* **88**, 872–880.

Field, C.A. (1982), Small sample asymptotic expansions for multivariate M-estimates, *Ann. Statist.* **10**, 671–689.

Handshin, E., F.C. Schweppe, J. Kohlas and A. Fiechter (1975), Bad data analysis for power system state elimination (with discussion). *IEEE Trans. Power App. Syst.* **PAS-94**, 329–337.

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

He, X. and Q.-M. Shao (1995), A strong Bahadur representation theorem of M-estimators, Preprint.

Hill, R.W. (1977), Robust regression when there are outliers in the carriers, Ph.D. Thesis, Harvard University, Cambridge, MA.

Huber, P.J. (1975), Robustness and designs, in: J.N. Srivastava, ed., *A Survey of Statistical Design and Linear Models* (North Holland, Amsterdam).

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

Huber, P.J. (1983), Minimax aspects of bounded-influence regression (with discussion), *J. Amer. Statist. Assoc.* **78**, 1333–1346.

Huber, P.J. and R. Dutter (1974), Numerical solutions of robust regression problems, in: G. Brickmann, ed., COMPSTAT 1974 (Physika Verlag, Wein) 165–172.

Maronna, R.A. and V.J. Yohai (1981), Asymptotic behaviour of general M-estimators for regression and scale with random carriers, *Z. Wahrsch. verw. Gebiete* **58**, 7–20.

Merrill, H.M. and F.C. Schweppe (1971), Bad data suppression in power system static state estimation, *IEEE Trans. Power App. Syst.* **PAS-90**, 2718–2725.

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

Silvapullé, M.J. (1985), Asymptotic behavior of robust estimators of regression and scale parameters with fixed carriers, *Ann. Statist.* **13**, 1490–1497.

Simpson, D.G., D. Ruppert, and R.J. Carroll (1992), On one-step GM estimates and stability of inferences in linear regression, *J. Amer. Statist. Assoc.* **87**, 439–450.