CHAPTER 4

M, R AND SOME SCALE ESTIMATORS

4.1. INTRODUCTION

In the last three decades statistics has seen the emergence and consolidation of many competitors of the Least Square estimator of β of (1.1.1). The most prominent are the so-called M- and R- estimators. The class of M-estimators was introduced by Huber (1973) and its computational aspects and some robustness properties are available in Huber (1981). The class of R-estimators is based on the ideas of Hodges and Lehmann (1963) and has been developed by Adichie (1967), Jurečková (1971) and Jaeckel (1972).

One of the attractive features of these estimators is that they are robust against certain outliers in errors. All of these estimators are translation invariant, whereas only R-estimators are scale invariant.

Our purpose here is to illustrate the usefulness of the results of Chapter 2 in deriving the asymptotic distributions of these estimators under a fairly general class of heteroscedastic errors. Section 4.2a gives the asymptotic distributions of M-estimators while those of R-estimators are given in Section 4.4. Among other things, the results obtained enable one to study their qualitative robustness against an array of non-identical error d.f.'s converging to a fixed error d.f. The sufficient conditions given here are fairly general for the underlying score functions and the design variables.

Efron (1979) introduced a general resampling procedure, called the bootstrap, for estimating the distribution of a pivotal statistic. Singh (1981) showed that the bootstrap estimate B_n is *second order accurate*, i.e., provides more accurate approximation to the sampling distribution G_n of the standardized sample mean than the usual normal approximation in the sense that $\sup\{|G_n(x) - B_n(x)|;\ x\in\mathbb{R}\}$ tends to zero at a faster rate than that of the square-root of n. This kind of result holds more generally as noted by Babu and Singh (1983, 1984).

Section 4.2b discusses similar results pertaining to a class of M-estimators of β when the errors in (1.1.1) are i.i.d.. It is noted that Shorack's (1982) modified bootstrap estimator and the one obtained by resampling the residuals according to a w.e.p. are second order accurate.

In an attempt to make M-estimators scale invariant one often needs a preliminary robust scale estimator. Two such estimators are the *MAD* (median of absolute deviations of residuals) and the *MASD* (median of absolute symmetrized deviations of residuals). The asymptotic distributions of these estimators under heteroscedastic errors appear in Section 4.3.

In carrying out the analysis of variance of an experimental design or a linear model based on ranks one needs an estimator of the asymptotic variance of certain rank statistics, see, e.g., Hettmansperger (1984). These variances involve the functional. $Q(f) = \int f\, d\varphi(F)$ where φ is a known function, F a common error d.f. having a density f. Some estimators of Q(f) under (1.1.1) are presented in Section 4.5. Again, the results of Chapter 2 are found useful in proving their consistency.

4.2. M-ESTIMATORS

4.2a. First Order Approximations: Asymptotic Normality

This subsection contains the asymptotic distributions of M-estimators of $\boldsymbol{\beta}$ when the errors in (1.1.1) are heteroscedastic. The following subsection 4.2b gives some results on the bootstrap approximations to these distributions.

Let the model (1.1.1) hold. Let ψ be a nondecreasing function from $\mathbb{R}$ to $\mathbb{R}$. The corresponding M-estimator $\hat{\boldsymbol{\Delta}}$ of $\boldsymbol{\beta}$ is defined to be a zero of the M-score $\int \psi(y)\, \mathbf{V}(dy, \mathbf{t})$, where $\mathbf{V}$ is defined at (1.1.2). Our objective is to investigate the asymptotic behavior of $\mathbf{A}^{-1}(\hat{\boldsymbol{\Delta}} - \boldsymbol{\beta})$ when the errors in (1.1.1) are heteroscedastic. Our method is still the usual one, v.i.z., to obtain the expansion of the M-score uniformly in $\mathbf{t} \in \{\mathbf{t};\ \|\mathbf{A}^{-1}(\mathbf{t}-\boldsymbol{\beta})\| \le B\}$, $0 < B < \infty$, to observe that there is a zero of the M-score, $\hat{\boldsymbol{\Delta}}$, in this set and then to apply this expansion to obtain the approximation for $\mathbf{A}^{-1}(\hat{\boldsymbol{\Delta}} - \boldsymbol{\beta})$ in terms of the given M-score at the true $\boldsymbol{\beta}$. To make all this precise, we need to standardize the M-score. For that reason we need some more notation. Let

$$\text{(1)} \qquad \Lambda^*(y) := \operatorname{diag}(f_{n1}(y), \ldots., f_{nn}(y)), \qquad y \in \mathbb{R},$$

$$\mathcal{C} := \mathbf{A}\mathbf{X}' \int \Lambda^*(y)\, d\psi(y)\, \mathbf{X}\mathbf{A},$$

$$\mathbf{T}(\psi, \mathbf{t}) := -\mathcal{C}^{-1}\mathbf{A} \int \psi(y)\mathbf{V}(dy, \mathbf{t}),$$

$$\overline{\mathbf{T}}(\psi, \mathbf{t}) := \mathbf{A}^{-1}(\mathbf{t} - \boldsymbol{\beta}) - \mathbf{T}(\psi, \boldsymbol{\beta}), \qquad \mathbf{t} \in \mathbb{R}^p.$$

An approximation to $\hat{\boldsymbol{\Delta}}$ is given by the zero $\overline{\boldsymbol{\Delta}}$ of $\overline{\mathbf{T}}(\psi, \mathbf{t})$, v.i.z.,

$$\text{(2)} \qquad \mathbf{A}^{-1}(\overline{\boldsymbol{\Delta}} - \boldsymbol{\beta}) = \mathbf{T}(\psi, \boldsymbol{\beta}).$$

A basic result needed to make this precise is the a.u.l. of $\mathbf{T}(\psi, \mathbf{t})$ in $\mathbf{A}^{-1}(\mathbf{t} - \boldsymbol{\beta})$. Often such a result is obtained under some smooth conditions on ψ and under i.i.d. errors. Theorem 4.2a.1 below gives such a result for a general nondecreasing right continuous bounded ψ and for fairly general independent heteroscedastic errors.

Theorem 4.2a.1. *Let* $\{(\mathbf{x}_{ni}', Y_{ni}),\ 1 \le i \le n\}$, β, $\{F_{ni},\ 1 \le i \le n\}$ *be as in the model* (1.1.1) *satisfying all conditions of Theorem* 2.3.3. *In addition, assume the following*:

(3) $\quad \psi \in \Psi := \{\psi\colon \mathbb{R}$ *to* $\mathbb{R}$, $\psi \in \mathbb{D}\mathcal{I}(\mathbb{R})$, *bounded with* $|\psi|_{tv} \le k < \infty\}$.

(4) $\quad \limsup_n \|C^{-1}\|_\infty < \infty.$

Then, $\forall\ 0 < B < \infty$,

(5) $$\sup_{\psi,\mathbf{u}} \|\mathbf{T}(\psi, \beta + \mathbf{A}\mathbf{u}) - \overline{\mathbf{T}}(\psi, \beta + \mathbf{A}\mathbf{u})\| = o_p(1).$$

where the supremum is taken over all $\psi \in \Psi$ *and* $\|\mathbf{u}\| \le B$.

Proof: Rewrite, after integration by parts,

$$\mathbf{T}(\psi, \mathbf{t}) - \overline{\mathbf{T}}(\psi, \mathbf{t}) = \int C^{-1}\mathbf{A}\,[\mathbf{V}(y, \mathbf{t}) - \mathbf{V}(y, \beta) - \Gamma_1(H(y))\mathbf{A}^{-1}(\mathbf{t} - \beta)]d\psi(y).$$

Now (5) readily follows from this and (2.3.37). □

In order to use this theorem, we must be able to argue that $\|\mathbf{A}^{-1}(\hat{\Delta} - \beta)\| = O_p(1)$. To that effect, define

$$\mu_i := E\,\psi(e_i), \quad \tau_i^2 = \mathrm{Var}\,\psi(e_i), \quad 1 \le i \le n,$$

$$\mathbf{b}_n := E\,\mathbf{T}(\psi, \beta) = -C^{-1}\mathbf{A}\,\Sigma_i\,\mathbf{x}_i\,\mu_i$$

and observe that

$$E\|\mathbf{A}^{-1}(\overline{\Delta} - \beta) - \mathbf{b}_n\|^2 = C^{-1}\,\Sigma_i\,\mathbf{x}_i'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_i\,\tau_i^2\,C^{-1} = O(1),$$

by (3), (4) and the fact that $\Sigma_i\,\mathbf{x}_i'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_i \equiv p < \infty$. Hence by the Markov inequality, $\forall\ \epsilon > 0\ \exists\ 0 < K_\epsilon < \infty \ni$

$$P(\|\mathbf{A}^{-1}(\bar{\Delta} - \beta) - \mathbf{b}_n\| \le K_\epsilon) \ge 1 - \epsilon, \text{ for all } n \ge 1.$$

Thus, assuming that

(6) $$\Sigma_i\,\mathbf{x}_i\,\mu_i = \mathbf{0},$$

and arguing via Brouwer's fixed point theorem as in Huber (1981,p 169), one concludes, in view (5), that $\forall\ \epsilon > 0\ \exists\ N_\epsilon$ and K_ϵ such that

(7) $$P(\|A^{-1}(\hat{\Delta}-\beta)\| \le K\epsilon) \ge 1-\epsilon, \quad \forall\ n \ge N\epsilon.$$

Now, a routine application of (5) enables one to conclude that

$$0 = T(\psi, \hat{\Delta}) = A^{-1}(\hat{\Delta}-\beta) - T(\psi, \beta) + o_p(1),$$

i.e.,

(8) $$A^{-1}(\hat{\Delta}-\beta) = T(\psi, \beta) + o_p(1).$$

Note that, under (6), with $T_0(\psi, \beta) = C\,T(\psi, \beta)$,

$$E\, T_0(\psi, \beta)\, T_0'(\psi, \beta) = A\, \Sigma_i\, x_i' x_i\, \tau_i^2\, A = AX'\mathcal{T} XA$$

where $\mathcal{T} = \text{diag}\,(\tau_1^2, \ldots, \tau_n^2)$. Moreover, for any $\lambda \in \mathbb{R}^p$,

$$\lambda' T_0(\psi, \beta) = \Sigma_i \sum_{j=1}^{p} \lambda_j d_{ij} \psi(e_i) = \Sigma_i\, \lambda' A\, x_i'\, \psi(e_i)$$

where $\{d_{ij}\}$ are as in (2.3.32). In view of (2.3.33) and (2.3.34), (NX) and (6) imply that $\lambda' T_0(\psi, \beta)$ is asymptotically normally distributed with mean 0 and the asymptotic variance $\lambda' AX'\mathcal{T} XA\,\lambda$. Thus by the Cramer-Wold device [Theorem 7.7, p 49, Billingsley (1968)], (4) and (8),

(9) $$\Sigma^{-1/2}A^{-1}(\hat{\Delta}-\beta) \xrightarrow[d]{} N(0, I_{p\times p}),\ \Sigma := C^{-1}AX'\mathcal{T} XAC^{-1}.$$

We summarize the above discussion as a

Proposition 4.2a.1. *Suppose that the d.f.'s* $\{F_{ni}\}$ *of the errors and the design matrix* X *of* (1.1.1) *satisfy* (4), (6) *and the assumptions of Theorem* 2.3.3 *including that* H *is strictly increasing for each* $n \ge 1$. *Then* (9) *holds.* □

Now, consider the case of the *i.i.d. errors* in (1.1.1) with $F_{ni} \equiv F$. Then,

(10) $$\tau_i^2 = \int \psi^2\, dF - \left(\int \psi dF\right)^2 = \tau^2,\ \text{(say)}, \qquad 1 \le i \le n,$$
$$C = \left(\int f\, d\psi\right) I_{p\times p}, \qquad \Sigma = \left(\int f\, d\psi\right)^{-2} \tau^2\, I_{p\times p}.$$

Consequently (4) is equivalent to requiring $\int f\, d\psi > 0$. Next, observe that (6) becomes

(6*) $$\Sigma_i\, x_i \int \psi\, dF = 0.$$

Obviously, this is satisfied if either $\Sigma_i \mathbf{x}_i = \mathbf{0}$, i.e., if $\mathbf{X}$ is a centered design matrix or if $\int \psi \, dF = 0$, the often assumed condition. The former excludes the possibility of the presence of the location parameter in (1.1.1). Thus to summarize, we have

Proposition 4.2a.2. *Suppose that in* (1.1.1), $F_{ni} \equiv F$. *In addition, assume that* $\mathbf{X}$ *and* F *satisfy* (NX), (F1), (F2), (6*) *and that* $\int f \, d\psi > 0$. *Then,*

$$\mathbf{A}^{-1}(\hat{\boldsymbol{\Delta}} - \boldsymbol{\beta}) \underset{d}{\longrightarrow} N_p(0, \tau^2/(\int f \, d\psi)^2 \, \mathbf{I}_{p \times p}). \qquad \square$$

Condition (6*) suggests another way of defining M-estimators of $\boldsymbol{\beta}$ in (1.1.1) in the case of the *i.i.d. errors.* Let

$$(11) \qquad \bar{x}_{nj} := n^{-1} \sum_{i=1}^{n} x_{nij}, \quad 1 \le j \le p; \qquad \bar{\mathbf{x}}_n' := (\bar{x}_{n1}, \ldots, \bar{x}_{np}),$$

$$\bar{\mathbf{X}}' := [\,\bar{\mathbf{x}}_n, \ldots, \bar{\mathbf{x}}_n]_{n \times p}, \qquad \mathbf{X}_c := \mathbf{X} - \bar{\mathbf{X}}.$$

Assume

$$\text{(NX1)} \qquad (\mathbf{X}_c'\mathbf{X}_c)^{-1} \quad \text{exists for all } n \ge p,$$

$$\max_i \mathbf{x}_{ni}' (\mathbf{X}_c'\mathbf{X}_c)^{-1} \mathbf{x}_{ni} = o(1).$$

Let

$$(12) \qquad \mathbf{T}^*(\psi, \mathbf{t}) := \mathbf{A}_1 \sum_i (\mathbf{x}_{ni} - \bar{\mathbf{x}}_n) \, \psi(Y_i - \mathbf{x}_{ni}'\mathbf{t}), \qquad \mathbf{t} \in \mathbb{R}^p,$$

$$\mathbf{A}_1 := (\mathbf{X}_c'\mathbf{X}_c)^{-1/2}.$$

Define an M-estimator $\boldsymbol{\Delta}^*$ as a solution $\mathbf{t}$ of

$$(13) \qquad \mathbf{T}^*(\psi, \mathbf{t}) = \mathbf{0}.$$

Apply Corollary 2.3.1 p times, j^{th} time with d_{ni} = i^{th} element of the j^{th} column of $\mathbf{X}_c\mathbf{A}_1$, $1 \le i \le n$, $1 \le j \le p$, to conclude an analogue of (5) above, v.i.z.,

$$(5^*) \qquad \sup_{\psi, \|\mathbf{A}_1^{-1}(\mathbf{t}-\boldsymbol{\beta})\| \le B} \|\mathbf{T}^*(\psi, \mathbf{t}) - \bar{\mathbf{T}}^*(\psi, \mathbf{t})\| = o_p(1)$$

where

$$\bar{\mathbf{T}}^*(\psi, \mathbf{t}) := \mathbf{A}_1^{-1}(\mathbf{t} - \boldsymbol{\beta}) - (\int f \, d\psi)^{-1} \, \mathbf{T}^*(\psi, \boldsymbol{\beta}).$$

The proof of (5*) is exactly similar to that of (5) with appropriate modifications of replacing $\mathbf{X}$ by $\mathbf{X}_c$ and $\mathbf{A}$ by $\mathbf{A}_1$ and using $F_{ni} \equiv F$ in the discussion there.

Now, clearly, $F_{ni} \equiv F$ implies that $\mathrm{E}\,\mathbf{T}^*(\psi, \boldsymbol{\beta}) = \mathbf{0}$,

$$\mathrm{E}\|\mathbf{T}^*(\psi, \boldsymbol{\beta})\|^2 = \Sigma_i (\mathbf{x}_i - \bar{\mathbf{x}})' \mathbf{A}_1' \mathbf{A}_1 (\mathbf{x}_i - \bar{\mathbf{x}})\, \tau^2 = O(1).$$

Hence, $\|\mathbf{T}^*(\psi, \boldsymbol{\beta})\| = O_p(1)$. If $\overline{\boldsymbol{\Delta}}^*$ is zero of $\overline{\mathbf{T}}^*(\psi, .)$, then

$$\mathbf{A}_1^{-1}(\overline{\boldsymbol{\Delta}}^* - \boldsymbol{\beta}) = \left(\int f\, d\psi\right)^{-1} \mathbf{T}^*(\psi, \boldsymbol{\beta}).$$

Argue, as for (7), (8) and (9) to conclude the following

Proposition 4.2a.3. *Suppose that in* (1.1.1), $F_{ni} \equiv F$. *In addition, assume that* $\mathbf{X}$ *and* F *satisfy* **(NX1)**, **(F1)** *and* **(F2)**. *Then,*

$$\mathbf{A}_1^{-1}(\boldsymbol{\Delta}^* - \boldsymbol{\beta}) \underset{d}{\rightarrow} N_p(\mathbf{0}, \tau^2/(\textstyle\int f\, d\psi)^2\, \mathbf{I}_{p\times p}),$$

where $\boldsymbol{\Delta}^*$ *is as in* (13). □

Remark 4.2a.2. Note that the Proposition 4.2a.3 does not require the condition $\int \psi\, dF = 0$. An advantage of this is that $\boldsymbol{\Delta}^*$ can be used as a preliminary estimator when constructing adaptive estimators of $\boldsymbol{\beta}$. An *adaptive* estimator is one that achieves the Hájek – Le Cam (Hájek 1972, Le Cam 1972) lower bound over a large class of error distributions. Often, a minimal condition required to construct an adaptive estimator of $\boldsymbol{\beta}$ is that F have finite Fisher information, i.e., that F satisfy (3.2.a) of Theorem 3.2.3. See, e.g., Bickel (1982), Fabian and Hannan (1982) and Koul and Susarla (1983). Recall, from Remark 3.2.2, that this implies **(F1)**.

On the other hand, the condition **(NX1)** does not allow for any location term in the linear regression model. □

So far we have been dealing with the linear regression model with known scale. Now consider the model (2.3.38) where γ is an unknown scale parameter. Let s be an $n^{1/2}$ – consistent estimator of γ, i.e.,

$$(14) \qquad |n^{1/2}(s - \gamma)\gamma^{-1}| = O_p(1).$$

Define an M-estimator $\hat{\boldsymbol{\Delta}}_1$ of $\boldsymbol{\beta}$ as a solution $\mathbf{t}$ of

$$(15) \qquad \Sigma\, \mathbf{x}_i\, \psi((Y_i - \mathbf{x}_i'\mathbf{t})s^{-1}) = \mathbf{0} \quad \text{or} \quad \int \psi(y)\, \mathbf{V}(s dy, \mathbf{t}) = \mathbf{0}.$$

To keep exposition simple, now we shall not exhibit ψ in some of the functions defined below. Define, for an $\alpha > 0$, $\mathbf{t} \in \mathbb{R}^p$,

$$(16)\qquad \mathbf{S}(\alpha, \mathbf{t}) := -\mathbf{A} \int \psi(y)\, \mathbf{V}(\alpha dy, \mathbf{t}),$$

$$\overline{\mathbf{S}}(\alpha, \mathbf{t}) := \mathbf{A}^{-1}(\mathbf{t}-\boldsymbol{\beta})\gamma^{-1} + \mathcal{C}^{-1}\mathcal{C}_1 n^{1/2}(\alpha-\gamma)\gamma^{-1} - \mathcal{C}^{-1}\mathbf{S}(\gamma, \boldsymbol{\beta}),$$

where

$$\mathcal{C}_1 := n^{-1/2}\mathbf{A}\mathbf{X}' \int y\mathbf{f}(y)\, d\psi(y), \quad \mathbf{f}'(y) := (f_1(y), \ldots, f_n(y)),$$

and where $\mathcal{C}$ is as in (1) above. Note that by **(NX)**, **(F1)**, **(F3)**, and (3),

$$(17)\qquad \|\mathcal{C}_1\| = O(1).$$

The following theorem is a direct consequence of Theorem 2.3.4. In it $N_1 := \{(\alpha, \mathbf{t})\colon \alpha > 0, \mathbf{t}\in\mathbb{R}^p, \|\mathbf{A}^{-1}(\mathbf{t}-\boldsymbol{\beta})\| \le B\gamma, |n^{1/2}(\alpha-\gamma)| \le b\gamma\}$, $0<b, B<\infty$.

Theorem 4.2a.2. *Let* $\{(\mathbf{x}_{ni}', Y_{ni}),\ 1 \le i \le n\}$, $\boldsymbol{\beta}$, γ, $\{F_{ni},\ 1 \le i \le n\}$ *be as in* (2.3.38) *satisfying all the conditions of Theorem* 2.3.4. *Moreover, assume* (3) *and* (4) *hold. Then, for every* $0 < b, B < \infty$ *fixed,*

$$(18)\qquad \sup \|\mathbf{S}(\alpha, \mathbf{t}) - \overline{\mathbf{S}}(\alpha, \mathbf{t})\| = o_p(1).$$

where the supremum is taken over all $\psi \in \Psi$, *and* $(\alpha, \mathbf{t}')' \in N_1$. □

Now argue as in the proof of the Proposition 4.2a.1 to conclude

Proposition 4.2a.3. *Suppose that the design matrix* $\mathbf{X}$ *and* d.f.'s $\{F_{ni}\}$ *of* $\{\epsilon_{ni}\}$ *in* (2.3.38) *satisfy* (5), (6) *and the assumptions of Theorem* 2.3.4 *including that* H *is strictly increasing for each* $n \ge 1$. *In addition assume that there exists an estimate* s *of* γ *satisfying* (14). *Then*

$$(19)\qquad \mathbf{A}^{-1}(\hat{\boldsymbol{\Delta}}_1 - \boldsymbol{\beta})\gamma^{-1} = \mathcal{C}^{-1}\mathbf{S}(\gamma, \boldsymbol{\beta}) - \mathcal{C}^{-1}\mathcal{C}_1 n^{1/2}(s-\gamma)\,\gamma^{-1} + o_p(1),$$

where $\hat{\boldsymbol{\Delta}}_1$ *now is a solution of* (15). □

Remark 4.2a.3. In (6), F_i is now the d.f. of ϵ_i, and not of $\gamma\epsilon_i$, $1 \le i \le n$. □

Remark 4.2a.4. ***Effect of symmetry on*** $\hat{\boldsymbol{\Delta}}_1$. As is clear from (19), in general the asymptotic distribution of $\hat{\boldsymbol{\Delta}}_1$ depends on s. However, suppose that

$$(20)\qquad d\psi(y) = -\,d\psi(-y), \quad f_i(y) \equiv f_i(-y), \quad 1 \le i \le n, \quad -\infty < y < +\infty.$$

Then $\int y\, f_i(y)\, d\psi(y) = 0$, $1 \le i \le n$, and, from (16), $\mathcal{C}_1 = \mathbf{0}$. Consequently, in this case,

$$\mathbf{A}^{-1}(\hat{\Delta}_1 - \beta) = \gamma C^{-1}\, \mathbf{S}(\gamma, \beta) + o_p(1).$$

Hence, with Σ as in (9), we obtain

$$\Sigma^{-1/2}\mathbf{A}^{-1}(\hat{\Delta}_1 - \beta)\, \gamma^{-1} \underset{d}{\rightarrow} N_p(\mathbf{0}, \mathbf{I}_{p\times p}). \tag{21}$$

Note that this asymptotic distribution differs from that of (9) only by the presence of γ^{-1}. In other words, in the case of symmetric errors $\{\epsilon_i\}$ and the skew symmetric score functions $\{\psi\}$, the asymptotic distribution of M-estimator of β of (2.3.38) with a preliminary $n^{1/2}$-consistent estimator of the scale parameter is the same as that of $\gamma^{-1}\times$M-estimator of β of (1.1.1). □

4.2b. Bootstrap Approximations

Before discussing the specific bootstrap approximations we shall describe the concept of Efron's bootstrap a bit more generally in the one sample setup.

Let $\xi_1, \xi_2, \ldots, \xi_n$ be n i.i.d. G r.v.'s, G_n be their empirical d.f. and $T_n = T_n(\xi_n, G)$ be a function of $\xi_n' := (\xi_1, \xi_2, \ldots, \xi_n)$ and G such that $T_n(\xi_n, G)$ is a r.v. for every G. Let $\zeta_1, \zeta_2, \ldots, \zeta_n$ denote i.i.d. G_n r.v.'s and $\zeta_n' := (\zeta_1, \zeta_2, \ldots, \zeta_n)$. The *bootstrap* d.f. B_n of $T_n(\xi_n, G)$ is the d.f. of $T_n(\zeta_n, G_n)$ under G_n. Efron (1979) showed, via numerical studies, that in several examples B_n provides better approximation to the d.f. Γ_n of $T_n(\xi_n, G)$ under G than the normal approximation. Singh (1981) substantiated this observation by proving that in the case of the standardized sample mean the bootstrap estimate B_n is *second order accurate*, i.e.,

$$\sup\{|\Gamma_n(x) - B_n(x)|;\, x \in \mathbb{R}\} = o(n^{-1/2}), \text{ a.s..} \tag{1}$$

Recall that the Edgeworth expansion or the Berry-Esseen bound gives that

$$\sup\{|\Gamma_n(x) - \Phi(x)|;\, x \in \mathbb{R}\} = O(n^{-1/2}),$$

where Φ is the d.f. of a N(0, 1) r.v. See, e.g., Feller (1966. Ch. XVI). Babu and Singh (1983, 1984), among others, pointed out that this phenomenon is shared by a large class of statistics. For further reading on bootstrapping we refer the reader to Efron (1982).

We now turn to the problem of bootstrapping M-estimators in a linear regression model. For the sake of clarity we shall restrict our attention to a *simple* linear regression model only. Our main purpose is to show how a certain weighted empirical sampling distribution naturally helps to overcome some inherent difficulties in defining bootstrap M-estimators. What follows is based on the work of Lahiri (1989). No proofs will be given as they involve intricate technicalities of the Edgeworth expansion for independent non-identically distributed r.v.'s.

Accordingly, assume that $\{e_i, i \geq 1\}$ are i.i.d F r.v.'s, $\{x_{ni}, i \geq 1\}$ are the known design points, $\{Y_{ni}, i \geq 1\}$ are observable r.v.'s such that for a $\beta \in \mathbb{R}$,

$$Y_{ni} = x_{ni}\beta + e_i, \qquad i \geq 1. \tag{2}$$

The score function ψ is assumed to satisfy

$$\int \psi \, dF = 0. \tag{3}$$

Let $\hat{\Delta}_n$ be an M-estimator obtained as a solution t of

$$\sum_{i=1}^{n} x_{ni}\, \psi(Y_{ni} - x_{ni}t) = 0 \tag{4}$$

and F_n be an estimator of F based on the residuals $\hat{e}_{ni} := Y_{ni} - x_{ni}\hat{\Delta}_n$, $1 \leq i \leq n$. Let $\{e^*_{ni}, 1 \leq i \leq n\}$ be i.i.d. F_n r.v.'s and define

$$Y^*_{ni} = x_{ni}\hat{\Delta}_n + e^*_{ni}, \qquad 1 \leq i \leq n. \tag{5}$$

The bootstrap M-estimator Δ^*_n is defined to be a solution t of

$$\sum_{i=1}^{n} x_{ni}\, \psi(Y^*_{ni} - x_{ni}t) = 0. \tag{6}$$

Recall, from the previous section, that in general (3) ensures the absence of the asymptotic bias in $\hat{\Delta}_n$. Analogously, to ensure the absence of the asymptotic bias in Δ^*_n, we need to have F_n such that

$$\int \psi \, dF_n = E_n\, \psi(e^*_{n1}) = 0, \tag{7}$$

where E_n is the expectation under F_n. In general, the choice of F_n that will satisfy (7) and at the same time be a reasonable estimator of F depends heavily on the forms of ψ and F. When bootstrapping the least square estimator of β, i.e., when $\psi(x) \equiv x$, Freedman (1981) ensures (7) by choosing F_n to be the empirical d.f. of the centered residuals $\{\hat{e}_{ni} - \hat{e}_{n\cdot}, 1 \leq i \leq n\}$, where $\hat{e}_{n\cdot} := n^{-1}\sum_{j=1}^{n} \hat{e}_{nj}$. In fact, he shows that if one does not center the residuals, the bootstrap distribution of the least squares estimator does not approximate the corresponding original distribution.

Clearly, the ordinary empirical d.f. $\hat{H}_n$ of the residuals $\{\hat{e}_{ni};\ 1\leq i\leq n\}$ does not ensure the validity of (7) for general designs and a general ψ. We are thus forced to look at appropriate modifications of the usual bootstrap. Here we describe two modifications. One chooses the resampling distribution appropriately and the other modifies the defining equation (6) *a la* Shorack (1982). Both provide the second order correct approximations to the distribution of standardized $\hat{\Delta}_n$.

Weighted Empirical Bootstrap:

Assume that the design points $\{x_{ni}\}$ are either all non–negative or all non–positive. Let $\omega_x := \sum_{i=1}^n |x_{ni}|$ be positive and define

$$F_{1n}(y) := \omega_x^{-1} \sum_{i=1}^n |x_{ni}| I(\hat{e}_{ni} \leq y), \qquad y\in\mathbb{R}. \tag{8}$$

Take the resampling distribution F_n to be F_{1n}. Then, clearly,

$$E_{1n}\ \psi(e^*_{n1}) = \omega_x^{-1} \sum_{i=1}^n |x_{ni}|\, \psi(\hat{e}_{ni}) = \operatorname{sign}(x_1)\ \omega_x^{-1} \sum_{i=1}^n x_{ni}\psi(Y_{ni} - x_{ni}\hat{\Delta}) = 0,$$

by the definition of $\hat{\Delta}_n$. That is, F_{1n} satisfies (7) for any ψ.

Modified Scores Bootstrap:

Let F_n be any resampling distribution based on the residuals. Define the bootstrap estimator Δ_{ns} to be a solution t of the equation

$$\sum_{i=1}^n x_{ni}\, [\psi(Y^*_{ni} - x_{ni}t) - E_n\ \psi(e^*_{n1})] = 0. \tag{9}$$

In other words the score function is now *a priori* centered under F_n and hence (7) holds for any F_n and any ψ.

We now describe the the second order correctness of these procedures. To that effect we need some more notation and assumptions. To begin with let $\tau_x^2 := \sum_{i=1}^n x_{ni}^2$ and define

$$m_x := \max\{|x_{ni}|;\ 1\leq i\leq n\},\quad b_{1x} := \sum_{i=1}^n x_{ni}^3/\tau_x^3,\quad b_x := \sum_{i=1}^n |x_{ni}^3|/\tau_x^3.$$

For a d.f. F and any sampling d.f. F_n, define

$$\gamma(x) := E\psi(e_1 - x), \qquad \varpi(x) = \sigma^2(x) := E\{\psi(e_1 - x) - \gamma(x)\}^2,$$

$$\varkappa_1(x) := E\{\psi(e_1 - x) - \gamma(x)\}^3, \qquad x \in \mathbb{R}.$$

$$\gamma_n(x) := E_n\psi(\overset{*}{e}_{n1} - x), \quad \varkappa_n(x) = \sigma_n^2(x) := E_n\{\psi(\overset{*}{e}_{n1} - x) - \gamma_n(x)\}^2,$$

$$\varkappa_{1n}(x) := E_n\{\psi(\overset{*}{e}_{n1} - x) - \gamma_n(x)\}^3, \qquad x \in \mathbb{R}.$$

$$A_n(c) := \{i:\ 1 \le i \le n,\ |x_{ni}| > c\tau_x\, b_x\}, \quad \kappa_n(c) := \# A_n(c), \quad c > 0.$$

For any real valued function g on $\mathbb{R}$, let $\dot{g}$, $\ddot{g}$ denote its first and second derivatives at 0 whenever they exist, respectively. Also, write γ_n, $\varkappa_n$ etc. for $\gamma_n(0)$, $\varkappa_n(0)$, etc. Finally, let $\alpha := -\dot{\gamma}/\sigma$, $\alpha_n := -\dot{\gamma}_n/\sigma_n$, and, define for $x \in \mathbb{R}$, $H_2(x) := x^2 - 1$, and

$$\mathscr{P}_n(x) := \Phi(x) - b_{1x}\,[\{\ddot{\gamma}_n/\sigma_n - \dot{\gamma}_n\, \dot{\varkappa}_n/\sigma_n^3\}(x^2/2\alpha_n^2) + (\varkappa_{1n}/6\sigma_n^3)\, H_2(x)]\ \varphi(x).$$

In the following theorems, a.s. means for almost all sequences $\{e_i;\ i \ge 1\}$ of i.i.d. F r.v.'s.

Theorem 4.2b.1. *Let the model* (2) *hold. In addition assume that* ψ *has uniformly continuous bounded second derivative and that the following hold:*

(a) $\tau_x^2 \longrightarrow \infty$. (b) $\alpha > 0$.

(c) *There exists a constant* $0 < c < 1$, *such that* $\ell n\ \tau_x = o(\kappa_n(c))$.

(d) $m_x\ \ell n\ \tau_x = o(\tau_x)$.

(e) *There exist constants* $\theta > 0$, $\delta > 0$ *and* $q < 1$ *such that*

$$\sup[\,|E\exp\{it\psi(e_1 - x)\}|:\ |x| < \delta,\ |t| > \theta] < q.$$

(f) $\forall\ \ \lambda > 0,\ \ \sum_{n=1}^{\infty} \exp(-\lambda\omega_x^2/\tau_x^2) < \infty$.

Then, with $\overset{*}{\Delta}_n$ *defined as a solution of* (6) *with* $F_n = F_{1n}$,

$$\sup_y |P_{1n}(\alpha_n\tau_x(\overset{*}{\Delta}_n - \hat{\Delta}_n) \le y) - \mathscr{P}_n(y)| = o(m_x/\tau_x),$$

$$\sup_y |P_{1n}(\alpha\tau_x(\hat{\Delta}_n - \beta) \le y) - P_{1n}(\tau_x(\overset{*}{\Delta}_n - \hat{\Delta}_n) \le y)| = o(m_x/\tau_x), \text{ a.s.},$$

where P_{1n} *denotes the bootstrap probability under* F_{1n}, *and where the supremum is over* $y \in \mathbb{R}$. □

Next we state the analogous result for Δ_{ns}.

Theorem 4.2b.2. *Suppose that all of the hypotheses of Theorem* 4.2b.1 *except* (f) *hold and that* Δ_{ns} *is defined as a solution of* (9) *with* $F_n = \hat{H}_n$, *the* *ordinary empirical of the residuals. Then,*

$$\sup_y |\hat{P}_n(\alpha_n \tau_x(\Delta_{ns} - \hat{\Delta}_n) \leq y) - \mathscr{F}_n(y)| = o(m_x/\tau_x),$$

$$\sup_y |\hat{P}_n(\alpha\tau_x(\hat{\Delta}_n - \beta) \leq y) - \hat{P}_n(\tau_x(\Delta_{ns} - \hat{\Delta}_n) \leq y)| = o(m_x/\tau_x), \text{ a.s.},$$

where $\hat{P}_n$ *denotes the bootstrap probability under* $\hat{H}_n$. □

The proofs of these theorems appear in Lahiri (1989) where he also discusses analogous results for a non–smooth ψ. In this case he chooses the sampling distribution to be a smooth estimator obtained from the kernel type density estimator. Lahiri (1991) gives extensions of the above theorems to multiple linear regression models.

Here we briefly comment about the assumptions (a) – (f). As is seen from the previous section, (a) and (b) are minimally required for the asmyptotic normality of M–estimators. Assumptions (c), (e) and (f) are required to carry out the Edgeworth expansions while (d) is slightly stronger than Noether's condition (NX) applied to (2). In particular, $x_i \equiv 1$ and $x_i \equiv i$ satisfy (a), (c), (d) and (f).

A sufficient condition for (e) to hold is that F have a positive density and ψ have a continuous positive derivative on an open interval in $\mathbb{R}$.

4.3. DISTRIBUTION OF SOME SCALE ESTIMATORS

Here we shall now discuss some robust scale estimators.

Definitions. An estimator $\hat{\beta}(X, Y)$ based on the design matrix X and the observation vector Y of β is said to be *location invariant* if

$$(1) \qquad \hat{\beta}(X, Y + Xb) = \hat{\beta}(X, Y) + b, \qquad \forall\ b \in \mathbb{R}^p.$$

It is said to be *scale invariant* if

$$(2) \qquad \hat{\beta}(X, aY) = a\hat{\beta}(X, Y), \qquad \forall\ a \in \mathbb{R},\ a \neq 0.$$

A scale estimator $s(X, Y)$ of a scale parameter γ is said to be *location invariant* if

$$(3) \qquad s(X, Y + Xb) = s(X, Y), \qquad \forall\ b \in \mathbb{R}^p.$$

It is said to be *scale invariant* if

$$(4) \qquad s(X, aY)) = |a|\ s(X,Y), \qquad \forall\ a \in \mathbb{R},\ a \neq 0.$$

Now observe that M-estimators $\hat{\Delta}$ and Δ^* of $\boldsymbol{\beta}$ of Section 4.2a are location invariant but not scale invariant. The estimators $\hat{\Delta}_1$, defined at (4.2a.13), are location and scale invariant whenever s satisfies (3) and (4). Note that if s does not satisfy (3) then $\hat{\Delta}_1$ need not be location invariant. Some of the candidates for s are

$$\text{(5)} \qquad s := \{(n-p)^{-1} \Sigma_i (Y_i - \mathbf{x}_i' \hat{\boldsymbol{\beta}})^2\}^{1/2},$$

$$s_1 := \text{med}\,\{|Y_i - \mathbf{x}_i' \hat{\boldsymbol{\beta}}|;\ 1 \le i \le n\},$$

$$s_2 := \text{med}\,\{|Y_i - Y_j - (\mathbf{x}_i - \mathbf{x}_j)' \hat{\boldsymbol{\beta}}|;\ 1 \le i < j \le n\},$$

where $\hat{\boldsymbol{\beta}}$ is a preliminary estimator of $\boldsymbol{\beta}$ satisfying (1) and (2).

Estimator s^2, with $\hat{\boldsymbol{\beta}}$ as the least square estimator, is the usual estimator of the error variance, assuming it exists. It is known to be non–robust against outliers in the errors. In robustness studies one needs scale estimators that are not sensitive to outliers in the errors. Estimator s_1 has been mentioned by Huber (1981, p. 175) as one such candidate. The asymptotic properties of s_1, s_2 will be discussed shortly. Here we just mention that each of these estimators estimates a different scale parameter, but that is not a point of concern if our goal is only to have location and scale invariant M-estimators of $\boldsymbol{\beta}$.

An alternative way of having location and scale invariant M-estimators of $\boldsymbol{\beta}$ is to use simultaneous M-estimation method for estimating $\boldsymbol{\beta}$ and γ of (2.3.38) as discussed in Huber (1981). We mention here, without giving details, that it is possible to study the asymptotic joint distribution of these estimators under heteroscedastic errors by using the results of Chapter 2.

We shall now study the asymptotic distributions of s_1 and s_2 under the model (1.1.1). With F_i denoting the d.f. of e_i, $H = n^{-1} \Sigma_i F_i$, let

$$\text{(6)} \qquad p_1(y) := H(y) - H(-y),$$

$$\text{(7)} \qquad p_2(y) := \int [H(y+x) - H(-y+x)]\, dH(x), \qquad y \ge 0.$$

Define γ_1 and γ_2 by the relations

$$\text{(8)} \qquad p_1(\gamma_1) = 1/2,$$

$$\text{(9)} \qquad p_2(\gamma_2) = 1/2.$$

Note that in the case $F_i \equiv F$, γ_1 is median of the distribution of $|e_1|$ and γ_2 is median of the distribution of $|e_1 - e_2|$. In general, γ_j, p_j, etc., depend on n, but we suppress this for the sake of convenience.

The asymptotic distribution of s_j is obtained by the usual method of connecting the event $\{s_j \le a\}$ with certain events based on certain empirical processes, as is done when studying the asymptotic distribution of the sample median, $j = 1, 2$. Accordingly, let

$$(10) \qquad S(y) := \sum_i I(|Y_i - \mathbf{x}_i'\hat{\beta}| \le y),$$
$$T(y) := \sum_{1 \le i \le j \le n}\sum I(|Y_i - Y_j - (\mathbf{x}_i - \mathbf{x}_j)'\hat{\beta}| \le y), \qquad y \ge 0.$$

Then, for an $a > 0$,

$$(11) \qquad \{s_1 \le a\} = \{S(a) \ge (n+1)2^{-1}\}, \qquad n \text{ odd},$$
$$\{S(a) \ge n2^{-1}\} \subseteq \{s_1 \le a\} \subseteq \{S(a) \ge n2^{-1} - 1\}, \qquad n \text{ even}.$$

Similarly, for an $a > 0$,

$$(12) \qquad \{s_2 \le a\} = \{T(a) \ge (N+1)2^{-1}\}, \qquad N := n(n-1)/2 \text{ odd},$$
$$\{T(a) \ge N2^{-1}\} \subseteq \{s_2 \le a\} \subseteq \{T(a) \ge N2^{-1} - 1\}, \qquad N \text{ even}.$$

Thus, to study the asymptotic distributions of s_j, $j = 1, 2$, it suffices to study those of $S(y)$ and $T(y)$, $y \ge 0$.

In what follows we shall be using the notation of Chapter 2 with the following modifications. As before, we shall write S_1^o, μ_1^o etc. for S_d^o, μ_d^o etc. of (2.3.1) whenever $d_{ni} \equiv n^{-1/2}$. Moreover, in (2.3.1), we shall take

$$(13) \qquad X_i = Y_i - \mathbf{x}_i'\beta = e_i, \quad \mathbf{c}_i = \mathbf{A}\mathbf{x}_i, \qquad 1 \le i \le n.$$

With these modifications, for all $n \ge 1$,

$$(14) \qquad S(y) = S_1^o(y, \mathbf{v}) - S_1^o(-y, \mathbf{v}) = n^{-1/2} \sum_i I(|e_i - \mathbf{c}_i'\mathbf{v}| \le y),$$
$$2n^{-1}T(y) = \int [S_1^o(y+x, \mathbf{v}) - S_1^o(-y+x, \mathbf{v})]\, S_1^o(dx, \mathbf{v}) - 1, \qquad y \ge 0,$$

with probability 1, where $\mathbf{v} = \mathbf{A}^{-1}(\hat{\beta} - \beta)$. Let

$$(15) \qquad \mu_1^o(y, \mathbf{u}) = \mu_1(H(y), \mathbf{u}), \quad Y_1^o(y, \mathbf{u}) = Y_1(H(y), \mathbf{u}), \qquad -\infty < y < \infty;$$
$$W(y, \mathbf{u}) = Y_1^o(y, \mathbf{u}) - Y_1^o(-y, \mathbf{u}),$$
$$K(y, \mathbf{u}) = \int [Y_1^o(y+x, \mathbf{u}) - Y_1^o(-y+x, \mathbf{u})]\, dH(x), \qquad y \ge 0, \ \mathbf{u} \in \mathbb{R}^p.$$

We shall write $W(y)$, $K(y)$ etc. for $W(y, 0)$, $K(y, 0)$ etc.

Theorem 4.3.1. *Assume that* (1.1.1) *holds with* $\mathbf{X}$ *and* $\{F_{ni}\}$ *satisfying* (NX) *and* (2.3.3). *Moreover, assume that* H *is strictly increasing for each* n *and that*

(16) $$\lim_{\delta\to 0} \limsup_n \sup_{0\le s\le 1-\delta} [H(H^{-1}(s+\delta)\pm\gamma_2) - H(H^{-1}(s)\pm\gamma_2)] = 0.$$

About $\{\hat{\beta}\}$ *assume that*

(17) $$\|A^{-1}(\hat{\beta}-\beta)\| = O_p(1).$$

Then, $\forall\ a \in \mathbb{R}$,

(18) $$P(n^{1/2}(s_1-\gamma_1) \le a\gamma_1)$$
$$= P\Big(W(\gamma_1) + n^{-1/2}\Sigma_i\, \mathbf{x}_i' A\, \{f_i(\gamma_1) - f_i(-\gamma_1)\}\cdot \mathbf{v}$$
$$\ge - a\cdot\gamma_1 n^{-1}\Sigma_i\, [f_i(\gamma_1) + f_i(-\gamma_1)]\Big) + o(1),$$

(19) $$P(n^{1/2}(s_2-\gamma_2) \le a\gamma_2)$$
$$= P\Big(2K(\gamma_2) + n^{-3/2}\sum_i\sum_j c_{ij} \int [f_i(\gamma_2+x) - f_i(-\gamma_2+x)]dF_j(x)\cdot\mathbf{v}$$
$$\ge -\gamma_2 a\, n^{-1} \sum_i \int [f_i(\gamma_2+x) + f_i(-\gamma_2+x)]\, dH(x)\Big) + o(1).$$

where $c_{ij} = (\mathbf{x}_i - \mathbf{x}_j)' A$, $1 \le i, j \le n$.

Proof. We shall give the proof of (19) only; that of (18) being similar and less involved. Fix an $a\in\mathbb{R}$ and let $Q_n(a)$ denote the left hand side of (19). Assume that n is large enough so that $a_n := (an^{-1/2} + 1)\gamma_2 > 0$. Then, by (12),

(20) $$Q_n(a) = P(T(a_n) \ge (N+1)/2), \qquad \text{N odd}\ \ (N := n(n-1)/2),$$
$$P(T(a_n) \ge N/2) \le Q_n(a) \le P(T(a_n) \ge N2^{-1}-1), \qquad \text{N even}.$$

It thus suffices to study $P(T(a_n) \ge N2^{-1} + b)$, $b \in \mathbb{R}$. Now, let

$$T_1(y) := n^{-1/2}(2n^{-1}T(y)+1) - n^{1/2}p_2(y), \qquad y \ge 0,$$
$$k_n := (N + 2b)\, n^{-3/2} + n^{-1/2} - n^{1/2}p_2(a_n).$$

Then, direct calculations show that

$$(21)\qquad P(T(a_n) \geq N2^{-1} + b) = P(T_1(a_n) \geq k_n).$$

We now *analyze* k_n: By (9),

$$k_n = -n^{1/2}\big(p_2(a_n) - p_2(\gamma_2)\big) + O(n^{-1/2}).$$

But

$$\begin{aligned} & n^{1/2}\big(p_2(a_n) - p_2(\gamma_2)\big) \\ & \quad = n^{1/2}\int [\{H(a_n+x) - H(\gamma_2+x)\} - \{H(-a_n+x) - H(-\gamma_2+x)\}]\, dH(x). \end{aligned}$$

By (2.3.3), the sequence of distributions $\{p_2\}$ is tight on $(\mathbb{R}, \mathscr{B})$, implying that $\gamma_2 = O(1)$, $n^{-1/2}\gamma_2 = o(1)$. Consequently, in view (2.3.3),

$$\begin{aligned} & n^{1/2}\int \{H(\pm a_n + x) - H(\pm\gamma_2 + x)\}\, dH(x) \\ & \qquad\qquad = a\gamma_2\, n^{-1}\Sigma_i \int f_i(\pm\gamma_2 + x)\, dH(x) + o(1), \end{aligned}$$

and

$$(22)\qquad k_n = -a\,\gamma_2 n^{-1}\,\Sigma_i \int [f_i(\gamma_2+x) + f_i(-\gamma_2+x)]dH(x) + o(1).$$

Next, we *approximate* $T_1(a_n)$ by a sum of independent r.v.'s. The proof is similar to the one used in approximating linear rank statistics of Section 3.4. From the definition of T_1 and (14),

$$\begin{aligned} (23)\qquad T_1(y) &= n^{-1/2}\int [S_1^o(y+x, \mathbf{v}) - S_1^o(-y+x, \mathbf{v})]\, S_1^o(dx, \mathbf{v}) - n^{1/2}p_2(y) \\ &= n^{-1/2}\int [Y_1^o(y+x, \mathbf{v}) - Y_1^o(-y+x, \mathbf{v})]\, S_1^o(dx, \mathbf{v}) \\ &\quad + n^{-1/2}\int [\mu_1^o(y+x, \mathbf{v}) - \mu_1^o(-y+x, \mathbf{v})]\, Y_1^o(dx, \mathbf{v}) \\ &\quad + n^{-1/2}\int [\mu_1^o(y+x, \mathbf{v}) - \mu_1^o(-y+x, \mathbf{v})]\, \mu_1^o(dx, \mathbf{v}) - n^{1/2}p_2(y) \\ &= E_1(y) + E_2(y) + E_3(y), \qquad \text{say.} \end{aligned}$$

But

$$(24)\quad E_3(y) := n^{-1/2}\int [\mu_1^o(y+x, v) - \mu_1^o(-y+x, v)]\, \mu_1^o(dx, v) - n^{1/2}p_2(y)$$
$$= n^{-3/2}\sum_i \sum_j \int \{F_i(y+x+c_{ij}'v) - F_i(-y+x+c_{ij}'v) - F_i(y+x) + F_i(-y+x)\}\, dF_j(x)$$
$$= n^{-3/2}\sum_i \sum_j c_{ij}'\, v \int [f_i(y+x) - f_i(-y+x)]\, dF_j(x) + \bar{o}_p(1),$$

by (2.3.3), (NX) and (17). In this proof, $\bar{o}_p(1)$ means $o_p(1)$ uniformly in $|y| \le k$, for every $0 < k < \infty$.

Integration by parts, (17), (2.3.25), H increasing and the fact that $\int n^{-1/2}\mu_1^o(dx, v) = 1$ yield that

$$(25)\quad E_2(y) := n^{-1/2}\int \{\mu_1^o(y+x, v) - \mu_1^o(-y+x, v)\}\, Y_1^o(dx, v)$$
$$= n^{-1/2}\int \{Y_1^o(y+x, v) - Y_1^o(-y+x, v)\}\, \mu_1^o(dx, v)$$
$$= \int \{Y_1^o(y+x) - Y_1^o(-y+x)\}\, dH(x) + \bar{o}_p(1).$$

Similarly,

$$(26)\quad E_1(y) = n^{-1/2}\int \{Y_1^o(y+x) - Y_1^o(-y+x)\}\, S_1^o(dx) + \bar{o}_p(1).$$

Now observe that $n^{-1/2}S_1^o = H_n$, the ordinary empirical d.f. of the errors $\{e_i\}$. Let

$$E_{11}(y) := \int \{Y_1^o(y+x) - Y_1^o(-y+x)\}\, d\big(H_n(x) - H(x)\big) = \mathcal{Z}(y) - \mathcal{Z}(-y),$$

where

$$\mathcal{Z}(\pm y) := \int Y_1^o(\pm y+x)\, d[H_n(x) - H(x)], \qquad y \ge 0.$$

We shall show that

$$(27)\qquad \mathcal{Z}(\pm a_n) = o_p(1).$$

But

$$|\mathcal{Z}(\pm a_n) - \mathcal{Z}(\pm\gamma_2)| = |\int [\mathcal{Y}_1(H(\pm a_n+x)) - \mathcal{Y}_1(H(\pm\gamma_2+x))]d\big(H_n(x) - H(x)\big)|.$$

$$\leq 2 \sup_{|y-z|\leq|a|n^{-1/2}\gamma} |\mathcal{Y}_1(H(y)) - \mathcal{Y}_1(H(z))|.$$

$$= o_p(1), \tag{28}$$

because of (2.3.3) and Corollary 2.3.1 applied with $d_{ni} \equiv n^{-1/2}$. Thus, to prove (27), it suffices to show that

$$\mathcal{Z}(\pm\gamma_2) = o_p(1). \tag{29}$$

But

$$|\mathcal{Z}(\pm\gamma_2)| = |\int_0^1 \big[\mathcal{Y}_1(H(\pm\gamma_2+H_n^{-1}(t))) - \mathcal{Y}_1(H(\pm\gamma_2+H^{-1}(t)))\big]dt|$$

$$\leq \sup_{0\leq t\leq 1} |\mathcal{Y}_1(H(\pm\gamma_2+H^{-1}(H\,H_n^{-1}(t)))) - \mathcal{Y}_1(H(\pm\gamma_2+H^{-1}(t)))|$$

$$= o_p(1),$$

by the assumption (16), Lemma 3.4.1 and Corollary 2.3.1 applied with $d_{ni} \equiv n^{-1/2}$. This proves (27). Consequently, from (26) and an argument like (28), it follows that

$$\begin{aligned} E_1(a_n) &= \int \{Y_1^o(a_n+x) - Y_1^o(-a_n+x)\}\, dH(x) + o_p(1) \\ &= \int \{Y_1^o(\gamma_2+x) - Y_1^o(-\gamma_2+x)\}\, dH(x) + o_p(1). \end{aligned} \tag{30}$$

From (23), (24), (25), (30) and the definition (15), we obtain

$$\begin{aligned} T_1(a_n) = 2K(\gamma_2) + n^{-3/2}\sum_i\sum_j c_{ij}'A \int \{f_i(\gamma_2+x) - f_i(-\gamma_2+x)\}\, dF_j(x)\cdot v \\ + o_p(1). \end{aligned} \tag{31}$$

Now, from the definition of k_n and (22), it follows that the $\lim k_n$ does not depend on b. Thus the limit of the l.h.s. of (21) is the same for $b = -1, 0, 1/2$, and, in view of (21), (22) and (31), it is given by the first term on the r.h.s. of (19). □

Remark 4.3.1. Observe that, in view of (8) and (9),

$$W(\gamma_1) = n^{-1/2}\Sigma_i \{I(|e_i| \leq \gamma_1) - 1/2\},$$

$$K(\gamma_2) = \int \{H(\gamma_2+x) - H(-\gamma_2+x)\}\, dY_1^o(x)$$
$$= n^{-1/2}\Sigma_i \{H(\gamma_2 + e_i) - H(-\gamma_2 + e_i) - 1/2\}.$$

Thus, $W(\gamma_1)$ and $K(\gamma_2)$ are the sums of bounded independent centered r.v.'s and by the L-F CLT one obtains

$$\sigma_1^{-1} W(\gamma_1) \underset{d}{\rightarrow} N(0, 1) \text{ and } \sigma_2^{-1} K(\gamma_2) \underset{d}{\rightarrow} N(0, 1), \tag{32}$$

where

$$\sigma_1^2 := \text{Var } W(\gamma_1) = n^{-1}\Sigma_i \{F_i(\gamma_1) - F_i(-\gamma_1)\}\{1 - F_i(\gamma_1) + F_i(-\gamma_1)\},$$
$$\sigma_2^2 := \text{Var } K(\gamma_2) = n^{-1}\Sigma_i \int [H(\gamma_2+x) - H(-\gamma_2+x)]^2\, dF_i(x) - (1/4).$$

Remark 4.3.2. If $\{F_i\}$ are all symmetric about zero, then from (32), (18) and (19), it follows that the asymptotic distribution of s_1 and s_2 does not depend on the initial estimator $\hat{\beta}$ of β. In fact, in this case we can deduce that

$$\tau_1^{-1} n^{1/2}(s_1 - \gamma_1)\gamma_1^{-1} \underset{d}{\rightarrow} N(0, 1), \tag{33}$$

$$\tau_2^{-1} n^{1/2}(s_2 - \gamma_2)\gamma_2^{-1} \underset{d}{\rightarrow} N(0, 1),$$

where

$$\tau_1^2 := \sigma_1^2 \{2\gamma_1 h(\gamma_1)\}^{-2}, \qquad h(x) := n^{-1} \Sigma_i f_i(x),$$
$$\tau_2^2 := \sigma_2^2 \{\gamma_2 \int h(\gamma_2+x)\, dH(x)\}^{-2}.$$

□

Remark 4.3.3. *i.i.d. case.* In the case $F_i \equiv F$, the asymptotic distribution of s_1 depends on $\hat{\beta}$ unless F is symmetric around zero. However, the asymptotic distribution of s_2 does not depend on $\hat{\beta}$. This is so because in this case the coefficient of $\mathbf{v}$ in (19) is

$$n^{-3/2} \sum_i \sum_j (\mathbf{x}_i - \mathbf{x}_j)' \mathbf{A} \int [f(\gamma_2+x) - f(-\gamma_2+x)]dF(x) = 0.$$

That the asymptotic distribution of s_2 is independent of $\hat{\beta}$ is not surprising because s_2 is essentially a symmetrized variant of s_1. We summarize this property of s_2 as

Corollary 4.3.1. *If in model* (1.1.1), $F_{ni} \equiv F$, F *satisfies* (F1), (F2) *and* $\mathbf{X}$ *satisfies* (NX) *then* $\tau_2^{-1} n^{1/2}(s_2 - \gamma_2) \underset{d}{\longrightarrow} N(0, 1)$, *where*

$$\tau_2^2 = \{\int [F(\gamma_2+x)-F(-\gamma_2+x)]^2 \, dF(x) - 1/4\} \cdot \{\int f(\gamma_2+x) \, dF(x)\}^{-2}. \quad \square$$

Note that γ_2 is now the median of the distribution of $|e_1 - e_2|$. Also, observe that the condition (16) now is equivalent to

$$\sup_{0 \le s \le 1-\delta} [P(F(e_1 - y) \le s+\delta) - P(F(e_1 - y) \le s)] \longrightarrow 0 \text{ as } \delta \longrightarrow 0, \forall\, y \in \mathbb{R},$$

which is implied by the assumptions on F. $\square$

4.4. R-ESTIMATORS OF $\boldsymbol{\beta}$.

Consider the model (1.1.1) and the vector of linear rank statistics

$$\text{(1)} \qquad \mathbf{T}(\mathbf{t}) := \mathbf{A}_1 \Sigma_i (\mathbf{x}_{ni} - \bar{\mathbf{x}}_n)\varphi(R_{it}/(n+1)), \qquad \mathbf{t} \in \mathbb{R}^p,$$

where $\mathbf{A}_1$ is as in (4.2a.12) and R_{it} is the rank of $Y_{ni} - \mathbf{x}_{ni}'\mathbf{t}$ among $\{Y_{nj} - \mathbf{x}_{nj}'\mathbf{t},\ 1 \le j \le n\}$.

One of the classes of R-estimators of $\boldsymbol{\beta}$ is defined by the relation

$$\text{(2)} \qquad \inf_{\mathbf{t}} |\mathbf{T}(\mathbf{t})|_1 = |\mathbf{T}(\hat{\boldsymbol{\beta}}_1)| = \sum_{j=1}^{p} |T_j(\mathbf{t})| = 0,$$

T_j being the jth component of $\mathbf{T}$ of (1). The estimators $\hat{\boldsymbol{\beta}}_1$ were initially studied by Adichie (1967) for the case $p = 1$ and by Jurečková (1971) for $p \ge 1$.

Another class of R-estimators can be defined by the relation

$$\text{(3)} \qquad \inf_{\mathbf{t}} \|\mathbf{T}(\mathbf{t})\| = \|\mathbf{T}(\hat{\boldsymbol{\beta}}_2)\|.$$

Yet another class of estimators, introduced by Jaeckel (1972), are defined by the relation

$$\text{(4)} \qquad \inf_{\mathbf{t}} \mathcal{J}(\mathbf{t}) = \mathcal{J}(\hat{\boldsymbol{\beta}}_3)$$

where

$$\text{(5)} \qquad \mathcal{J}(\mathbf{t}) := \Sigma_i (Y_{ni} - \mathbf{x}_{ni}'\mathbf{t})\, \varphi(R_{it}/(n+1)), \qquad \mathbf{t} \in \mathbb{R}^p.$$

Jaeckel (op. cit.) showed that for every observation vector $(Y_1, \ldots, Y_n)$ and for every $n \ge p$, $\Sigma_i \varphi(i/n+1) \equiv 0$ implies that $\mathcal{J}(\mathbf{t})$ is nonnegative, continuous and convex function of $\mathbf{t}$. If, in addition, $\mathbf{X}_c$ has the full rank p

then the set $\{\mathbf{t};\ \mathcal{J}(\mathbf{t})\le b\}$ is bounded for every $0 \le b < \infty$, where X_c is defined at (4.2a.11). Consequently, $\hat{\beta}_3$ exists.

Moreover, the almost everywhere derivative of $\mathcal{J}$, w.r.t. $\mathbf{t}$, is $-A_1^{-1}\mathbf{T}(\mathbf{t})$. Thus, at $\hat{\beta}_3$, $\mathbf{T}$ is nearly equal to zero and hence $\hat{\beta}_1$, $\hat{\beta}_2$, and $\hat{\beta}_3$ are essentially the same estimators. Jaeckel showed, using the a.u.l. property of $\mathrm{T}(\mathbf{t})$ due to Jurečková (1971), that indeed $\|A_1(\hat{\beta}_1 - \hat{\beta}_3)\| = o_p(1)$.

Here we shall discuss the asymptotic distribution of $\{\hat{\beta}_2\}$ under general heteroscedastic errors. The main tool is the a.u.l. Theorem 3.2.4. We shall also conclude that $\|A_1(\hat{\beta}_2 - \hat{\beta}_3)\| = o_p(1)$ under (1.1.1) with general independent errors.

To begin with note that $\mathbf{T}$ of (1) is a p–vector $(T_1, \ldots, T_p)'$ where $T_j(\mathbf{t})$ is a $T_d(\varphi, \mathbf{u})$ – statistic of (3.1.2) with

$$
\begin{aligned}
(6) \qquad & X_{ni} = Y_{ni} - \mathbf{x}_{ni}'\beta, \quad \mathbf{c}_{ni} = A_1(\mathbf{x}_{ni} - \bar{\mathbf{x}}_n), \quad \mathbf{u} = A_1^{-1}(\mathbf{t} - \beta), \\
& d_{ni} = \mathbf{a}_{(j)}'(\mathbf{x}_{ni} - \bar{\mathbf{x}}_n), \ 1 \le i \le n; \ \mathbf{a}_{(j)} = j^{\text{th}} \text{ column of } A_1, \ 1 \le j \le p.
\end{aligned}
$$

Thus specializing Theorem 3.2.4 to this case readily gives

Lemma 4.4.1. *Suppose that* (1.1.1) *holds with* F_{ni} *as a d.f. of* e_{ni}, $1 \le i \le n$. *In addition, assume that*

$$
(\mathrm{NX_c}) \qquad (X_c'X_c)^{-1} \textit{ exists for all } n \ge p,
$$

$$
\max_i (\mathbf{x}_{ni} - \bar{\mathbf{x}}_n)'(X_c'X_c)^{-1}(\mathbf{x}_{ni} - \bar{\mathbf{x}}_n) = o(1).
$$

About $\{F_{ni}\}$ *assume that* H *is strictly increasing for each* n *and that* (2.2.3b), (3.2.12), (3.2.35), (3.2.36) *hold and that*

$$
(7) \qquad \lim_{\delta\to 0} \limsup_n \sup_{0\le s\le 1-\delta} [L_j(s+\delta) - L_j(s)] = 0, \qquad j = 1, \ldots, p
$$

where

$$
L_j(s) := \Sigma_i \left(\mathbf{a}_{(j)}'(\mathbf{x}_{ni} - \bar{\mathbf{x}}_n)\right)^2 F_{ni}(H^{-1}(s)), \quad 0 \le s \le 1, \quad 1 \le j \le p.
$$

Then, for every $0 < B < \infty$,

$$
(8) \qquad \sup_{\varphi\in\mathscr{C}, \|A_1^{-1}(\mathbf{t}-\beta)\|\le B} \|\mathbf{T}(\mathbf{t}) - \mathbf{T}(\beta) + K_n A_1^{-1}(\mathbf{t}-\beta)\| = o_p(1)
$$

where

$$
K_n := A_1 \int_0^1 \sum_{i=1}^n (\mathbf{x}_{ni} - \tilde{\mathbf{x}}_n(s))(\mathbf{x}_{ni} - \bar{\mathbf{x}}_n)' q_{ni}(s)\, d\varphi(s)\, A_1
$$

$$\tilde{\mathbf{x}}_n(s) := n^{-1}\sum_{i=1}^{n} \mathbf{x}_{ni}\ell_{ni}(s),$$

$\ell_{ni}(s)$ *as in* (3.2.35) *and* $q_{ni}(s) := f_{ni}(H^{-1}(s)),\ i \le i \le n,\ 0 \le s \le 1.$ □

In order to prove the asymptotic normality of $\hat{\boldsymbol{\beta}}_2$, we need to show that $\|A_1^{-1}(\hat{\boldsymbol{\beta}}_2 - \boldsymbol{\beta})\| = O_p(1)$. To this effect let

$$\boldsymbol{\mu} := A_1 \Sigma_i (\mathbf{x}_{ni} - \bar{\mathbf{x}}_n) \int F_{ni}(H^{-1})\, d\varphi, \qquad \mathbf{S} := \mathbf{T}(\boldsymbol{\beta}) - \boldsymbol{\mu}.$$

Observe that the distribution of $(\hat{\boldsymbol{\beta}}_2 - \boldsymbol{\beta})$ does not depend on $\boldsymbol{\beta}$, even when $\{e_{ni}\}$ are not identically distributed.

Lemma 4.4.2. *In addition to the assumptions of Lemma* 4.4.1 *suppose that*

(9) $$\|\mathbf{S} + \boldsymbol{\mu}\| = O_p(1),$$

(10) $$\liminf_n \inf_{\|\boldsymbol{\theta}\|=1} |\boldsymbol{\theta}' K_n \boldsymbol{\theta}| \ge \alpha \ \textit{for an}\ \alpha > 0,$$

(11) $$K_n^{-1} \ \textit{exists for all}\ n \ge p, \quad \|K_n^{-1}\| = O(1).$$

Then, for every $\epsilon > 0,\ 0 < z < \infty$, *there exist a* $0 < b < \infty$ *and* N_ϵ *such that*

(12) $$P\Big(\inf_{\|\mathbf{u}\|>b} \|\mathbf{T}(A_1\mathbf{u} + \boldsymbol{\beta})\| \ge z\Big) \ge 1 - \epsilon, \qquad n \ge N_\epsilon.$$

Proof. Fix an $\epsilon > 0$, $0 < z < \infty$. Without loss of generality assume $\boldsymbol{\beta} = \mathbf{0}$. Observe that by the C–S inequality

$$\inf_{\|\mathbf{u}\|>b} \|\mathbf{T}(A_1\mathbf{u})\|^2 \ge \inf_{\|\boldsymbol{\theta}\|=1,\,|r|>b} (\boldsymbol{\theta}'\mathbf{T}(rA_1\boldsymbol{\theta}))^2.$$

Thus it suffices to prove that there exist a $0 < b < \infty$ and N_ϵ such that

(13) $$P\Big(\inf_{\|\boldsymbol{\theta}\|=1,\,|r|>b} (\boldsymbol{\theta}'\mathbf{T}(rA_1\boldsymbol{\theta}))^2 \ge z\Big) \ge 1 - \epsilon, \qquad n > N_\epsilon.$$

Let, for $\mathbf{t} \in \mathbb{R}^p$, $\hat{\mathbf{T}}(\mathbf{t}) := \mathbf{T}(\mathbf{0}) - K_n A_1^{-1}\mathbf{t}$, so that, by (8) for every $0 < B < \infty$,

(14) $$\sup_{\|\boldsymbol{\theta}\|=1,\,|r|\le B} |\boldsymbol{\theta}'\mathbf{T}(rA_1\boldsymbol{\theta}) - \boldsymbol{\theta}'\hat{\mathbf{T}}(rA_1\boldsymbol{\theta})| = o_p(1).$$

But

$$\theta' \hat{T}(rA_1\theta) = \theta'(S + \mu) - \theta' K_n \theta r.$$

By (9), there exist a K_ϵ and an $N_{1\epsilon}$ such that

$$P(|S + \mu| \le K_\epsilon) \ge 1 - \epsilon/2, \qquad n \ge N_{1\epsilon}.$$

Choose b to satisfy

$$b \ge (K_\epsilon + z^{1/2})\alpha^{-1}, \qquad \alpha \text{ as in (10).} \tag{15}$$

Then

$$\begin{aligned} P\Big(\inf_{\|\theta\|=1,\, |r|>b} (\theta' \hat{T}(rA_1 \theta))^2 \ge z\Big) &\\ \ge P\Big(\|S + \mu\| \le -z^{1/2} + b \inf_{\|\theta\|=1} |\theta' K_n \theta|\Big)& \\ \ge P(\|S + \mu\| \le K_\epsilon) \ge 1 - \epsilon/2, \quad \forall\, n \ge N_{1\epsilon}.& \end{aligned} \tag{16}$$

Therefore by (14) and (16) there exist N_ϵ and b as in (15) such that

$$P\Big(\inf_{\|\theta\|=1,\, |r|>b} (\theta' T(rA_1\theta))^2 \ge z\Big) \ge 1 - \epsilon, \qquad n \ge N_\epsilon. \tag{17}$$

But

$$\theta' T(rA_1\theta) = \theta' A_1 \Sigma_i (\mathbf{x}_i - \bar{\mathbf{x}})\, \varphi(R^*_{ir}/(n+1)) = \Sigma_i d_i\, \varphi(R^*_{ir}/(n+1)),$$

where $d_i = \theta' A_1 (\mathbf{x}_i - \bar{\mathbf{x}})$, R^*_{ir} is the rank of $Y_i - r(\mathbf{x}_i - \bar{\mathbf{x}})' A_1 \theta$. But such a linear rank statistic is nondecreasing in r, for every θ. See, e.g., Hájek (1969; Theorem 7E, Chapter II). This together with (17) enables one to conclude (13) and hence (12). □

Theorem 4.4.1. *Suppose that* (1.1.1) *holds and that the design matrix* $\mathbf{X}$ *and the error d.f.'s* $\{F_{ni}\}$ *satisfy the assumptions of Lemmas* 4.4.1 *and* 4.4.2 *above. Then*

$$A_1^{-1}(\hat{\beta}_2 - \beta) - K_n^{-1}\mu = K_n^{-1}S + o_p(1). \tag{18}$$

Proof. Follows from Lemmas 4.4.1 and 4.4.2. □

Remark 4.4.1. Arguing as in Jaeckal combined with an argument of Lemma 4.4.2, one can show that $\|A_1^{-1}(\hat{\beta}_2 - \hat{\beta}_3)\| = o_p(1)$. Consequently, under the conditions of Lemmas 4.4.1 and 4.4.2, $\hat{\beta}_2$ and the Jaeckel estimator $\hat{\beta}_3$ also satisfy (18). □

Remark 4.4.2. Consider the case when $F_{ni} \equiv F$, F a d.f. satisfying (**F1**), (**F2**). Then $\mu = 0$ and $S = T(\beta)$. Moreover, under (**NX**$_c$) all other assumptions of Lemmas 4.4.1 and 4.4.2 are *a priori* satisfied. Note that here

$$\ell_{ni} \equiv 1, \quad \tilde{x}_n(s) \equiv \bar{x}_n \quad \text{and} \quad K_n \equiv \int f\, d\varphi(F)\cdot I_{p\times p}.$$

Moreover, from Theorem 3.4.3 above, it follows that $S \xrightarrow[d]{} N_p(0, \sigma_\varphi^2 I_{p\times p})$, $\sigma_\varphi^2 = \int_0^1 \varphi^2(u)du - \left(\int_0^1 \varphi(u)du\right)^2$. We summarize the above discussion in

Corollary 4.4.1. *Suppose that* (1.1.1) *with* $F_{ni} \equiv F$ *holds. Suppose that* F *and* X *satisfy* (**F1**), (**F2**), and (**NX**$_c$). *In addition, suppose that* φ *is nondecreasing bounded on* $[0, 1]$ *and* $\int f\, d\varphi(F) > 0$. *Then*

$$A_1^{-1}(\hat{\beta}_2 - \beta) = \left\{\int f\, d\varphi(F)\right\}^{-1} T(\beta) + o_p(1). \tag{19}$$

Moreover,

$$A_1^{-1}(\hat{\beta}_2 - \beta) \xrightarrow[d]{} N(0, \tau^2 I_{p\times p}), \qquad \tau^2 = \sigma_\varphi^2 \left(\int f\, d\varphi(F)\right)^{-2}. \quad □$$

This result is quite general as far as the conditions on the design matrix X and F are concerned but not that general as far as the score function φ is concerned. □

Remark 4.4.2. ***Robustness against heteroscedastic gross errors.*** First , we give a working definition of qualitative robustness. Consider the model (1.1.1). Suppose that we have modeled the errors $\{e_{ni}, 1 \le i \le n\}$ to be i.i.d. F whereas their actual d.f.'s are $\{F_{ni}, 1 \le i \le n\}$. Let $P^n := \prod_{i=1}^n F$, $Q^n := \prod_{i=1}^n F_{ni}$ denote the corresponding product probability measures.

Definition 4.4.1. A sequence of estimators $\hat{\beta}$ is said to be *qualitatively robust* for β at F against Q^n if it is consistent for β under P^n and under those Q^n that satisfy $\mathcal{D}_n := \max_i \sup_y |F_{ni}(y) - F(y)| \to 0$.

The above definition is a variant of that of Hampel (1971). One could use the notions of weak convergence on product probability spaces to give a bit more general definition. For example we could insist that the Prohorov distance between Q^n and P^n should tend to zero instead of requiring $\mathcal{D}_n \longrightarrow 0$. We do not pursue this any further here.

The result (18) can be used to study the qualitative robustness of $\hat{\beta}_2$ against certain heteroscedastic errors. Consider, for example, the gross errors model where, for some $0 \le \delta_{ni} \le 1$, with $\max_i \delta_{ni} \longrightarrow 0$,

$$F_{ni} = (1 - \delta_{ni})\, F + \delta_{ni}\, G, \quad 1 \le i \le n,$$

and, where G is d.f. having a uniformly continuous a.e. positive density. If, in addition, $\{\delta_{ni}\}$ satisfy

$$\|A_1 \Sigma_i (\mathbf{x}_{ni} - \bar{\mathbf{x}}_n)\, \delta_{ni}\| = O(1), \tag{20}$$

then one can readily see that $\|K_n^{-1}\| = O(1)$ and $\|\mu\| = O(1)$. It follows from (18) that $\hat{\beta}_2$ is qualitatively robust against the above heteroscedastic gross errors at every F that has uniformly continuous a.e. positive density. Examples of δ_{ni} satisfying (20) would be

$$\delta_{ni} \equiv n^{-1/2} \quad \text{or} \quad \delta_{ni} = p^{-1/2}\, \|A_1 (\mathbf{x}_{ni} - \bar{\mathbf{x}}_n)\|, \quad 1 \le i \le n.$$

It may be argued that the latter choice of contaminating proportions $\{\delta_{ni}\}$ is more natural to linear regression than the former.

A similar remark is applicable to $\hat{\beta}_1$ and $\hat{\beta}_3$. □

4.5. ESTIMATION OF Q(f).

Consider the model (1.1.1) with $F_{ni} \equiv F$, where F is a d.f. with density f on $\mathbb{R}$. Define

$$Q(f) = \int f\, d\varphi(F) \tag{1}$$

where $\varphi \in \mathscr{C}$ of (3.2.1).

As is seen from Corollary 4.4.1, the parameter Q appears in the asymptotic variance of R-estimators. The complete rank analysis of the model (1.1.1) requires an estimate of Q. This estimate is used to standardize rank test statistics when carrying out the ANOVA of linear models using Jaeckal's dispersion $\mathcal{J}$ of (4.4.5). See, for example, Hettmansperger (1984) and references therein for the rank based ANOVA.

Lehmann (1963) and Sen (1966) give estimators of Q in the one and two sample location models. These estimators are given in terms of lengths of confidence intervals based on linear rank statistics. Koul (1971) extended these estimators to the multiple linear regression model (1.1.1). In this case these estimators are given in terms of Lebesgue measures of certain confidence regions based on ranks and are hard to compute for $p > 1$.

Cheng and Serfling (1981) discuss several estimators of Q when observations are i.i.d. F, i.e., when there are no nuisance parameters. Some of these estimators are obtained by replacing f by a kernel type density estimator and F by an empirical d.f. in Q. Scheweder (1975) discusses similar estimates of Q in the one sample location model.

In this section we discuss two types of estimators of Q. Both use a kernel type density estimator of f based on the residuals and the ordinary residual empirical d.f. to estimate F. The difference is in the way the window width and the kernel are chosen. In one the window width is partially based on the data and is of the order of square root of n and the kernel is the histogram type whereas in the other the kernel and the window width are arbitrary. It will be observed that the a.u.l. result about the residual empirical process of Corollary 2.3.5 is the basic tool needed to prove the consistency of these estimators.

We begin with the class of estimators where the window width is *partly based on the data.* Define

$$p(y) := \int [F(y+x) - F(-y+x)]\, d\varphi(F(x)), \qquad y \geq 0. \tag{2}$$

Since φ is a d.f., $p(y) \equiv P(|e - e^*| \leq y)$ where e, e^* are independent r.v.'s with respective d.f.'s F and $\varphi(F)$. Consequently, under **(F1)**, the density of p at 0 is $2Q$. This suggests that an estimate of Q can be obtained by estimating the slope of p at 0.

Recall the definition of the residual empirical process $H_n(y, \mathbf{t})$ from (1.2.1). Let $\hat{\boldsymbol{\beta}}$ be an estimator of $\boldsymbol{\beta}$ and define

$$\hat{H}_n(y) := H_n(y, \hat{\boldsymbol{\beta}}), \qquad y \in \mathbb{R}. \tag{3}$$

A natural estimator of p is obtained by substituting $\hat{H}_n$ for F in p, v.i.z.,

$$\hat{p}_n(y) := \int [\hat{H}_n(y+x) - \hat{H}_n(-y+x)]\, d\varphi(\hat{H}_n(x)), \qquad y \geq 0. \tag{4}$$

Let $-\infty = \hat{e}_{(0)} < \hat{e}_{(1)} \leq \hat{e}_{(2)} \leq \dots \leq \hat{e}_{(n)} < \hat{e}_{(n+1)} = \infty$ denote the ordered residuals $\{\hat{e}_i,\ 1 \leq i \leq n\}$, where $\hat{e}_i = Y_i - \mathbf{x}_i'\hat{\boldsymbol{\beta}}$, $1 \leq i \leq n$. Since $\varphi(\hat{H}_n)$ assigns mass $\{\varphi(j/n) - \varphi((j-1)/n))\}$ to each $\hat{e}_{(j)}$ and zero mass to each of the intervals $(\hat{e}_{(j-1)}, \hat{e}_{(j)})$, $1 \leq j \leq n+1$; it readily follows that $\forall\, y \in \mathbb{R}$,

$$\hat{p}_n(y) = \sum_{j=1}^{n}\{\varphi(j/n) - \varphi((j-1)/n)\}\,[\hat{H}_n(y+\hat{e}_{(j)}) - \hat{H}_n(-y+\hat{e}_{(j)})]$$

$$(5) \qquad = n^{-1}\sum_{j=1}^{n}\{\varphi(j/n) - \varphi((j-1)/n)\}\sum_{i=1}^{n} I(|\hat{e}_{(i)} - \hat{e}_{(j)}| \le y).$$

From (5) one sees that $\hat{p}_n(y)$ has the following interpretation. For each j, one first computes the proportion of $\{\hat{e}_{(i)}\}$ falling in the interval $[-y+\hat{e}_{(j)},\ y+\hat{e}_{(j)}]$ and then $\hat{p}_n(y)$ gives the weighted average of such proportions. Formula (5) is clearly suitable for computations.

Now, if $\{h_n\}$ is a sequence of positive numbers tending to zero, an estimator of Q is given by

$$\mathcal{Q}_n = \hat{p}_n(h_n)/2h_n.$$

This estimator can be viewed from the density estimation point of view also. Consider a kernel–type density estimator f_n of f based on the residuals $\{\hat{e}_i\}$:

$$f_n(x) := (2nh_n)^{-1}\sum_{i=1}^{n} I(|x - \hat{e}_i| \le h_n),$$

which uses the window $w_n(x) = (1/2)\cdot I(|x| \le h_n)$. Then a natural estimator of Q is

$$\int f_n\, d\varphi(\hat{H}_n) = \sum_{j=1}^{n}\{\varphi(\tfrac{j}{n}) - \varphi(\tfrac{j-1}{n})\}\, f_n(\hat{e}_{(j)}) = \mathcal{Q}_n.$$

Scheweder (1975) studied the asymptotic properties of this estimator in the one sample location model. Observe that in this case the estimator of Q does not depend on the estimator of the location parameter which makes it relatively easier to derive its asymptotic properties.

In $\mathcal{Q}_n$, there is an arbitrariness due to the choice of the window width h_n. Here we recommend that h_n be determined from the spread of the data as follows. Let $0 < \alpha < 1$, t_n^α be α^{th} quantile of $\hat{p}_n$ and define the estimator $\mathcal{Q}_n^\alpha$ of Q as

$$(6) \qquad \mathcal{Q}_n^\alpha := \hat{p}_n(n^{-1/2}t_n^\alpha)/(2n^{-1/2}t_n^\alpha).$$

The quantile t_n^α is an estimator of the α^{th} quantile t^α of p. Note that if $\varphi(s) \equiv s$, then t^α is the α^{th} quantile of the distribution of $|e_1 - e_2|$ and t_n^α is the α^{th} quantile of the empirical d.f. $\hat{p}_n$ of the r.v.'s $\{|\hat{e}_i - \hat{e}_j|,\ 1 \le i, j \le n\}$. Thus, e.g., $t_n^{.5} = s_2$ of (4.3.5). Similarly, if $\varphi(s) = I(s \ge 0)$

then t^{α} (t_n^{α}) is α^{th} quantile of the d.f. of $|e_1|$ (empirical d.f. of $|\hat{e}_i|$, $1 \le i \le n$). Again, here $t_n^{.5}$ would correspond to s_1 of (4.3.5). In any case, in general, t^{α} is a scale parameter in the sense of Bickel and Lehmann (1975).

The consistency of $\mathcal{Q}_n^{\alpha}$ is asserted in the following

Theorem 4.5.1. *Let* (1.1.1) *hold with* $F_{ni} \equiv F$. *In addition to* (**NX**), (**F1**) and (**F2**), *assume that* $\hat{\beta}$ *is an estimator of* β *satisfying* (4.3.17). *Then,*

$$\sup_{\varphi \in \mathscr{C}} |\mathcal{Q}_n^{\alpha} - Q(f)| = o_p(1). \tag{7}$$

The proof of (7) will be a consequence of the following *three* lemmas.

Lemma 4.5.1. *Under the assumptions of Theorem* 4.5.1, $\forall\ 0 \le a < \infty$,

$$\sup_{\varphi \in \mathscr{C}, 0 \le z \le a} |n^{1/2}\{\hat{p}_n(n^{-1/2}z) - p(n^{-1/2}z)\}| = o_p(1). \tag{8}$$

Consequently, $\forall\ 0 \le a < \infty$,

$$\sup_{\varphi \in \mathscr{C}, 0 \le z \le a} |n^{1/2}\hat{p}_n(n^{-1/2}z) - 2zQ(f)| = o_p(1). \tag{9}$$

Proof. We shall apply Corollary 2.3.5. Let $\mathbf{v} = \mathbf{A}^{-1}(\hat{\beta} - \beta)$, $\mathbf{b}_n' = n^{-1/2}\sum_i \mathbf{x}_{ni}' \mathbf{A}$. Then, from (2.3.46), (3) and (4.3.17), we obtain

$$\sup_{-\infty \le y \le \infty} |n^{1/2}\{\hat{H}_n(y) - H_n(y)\} - \mathbf{b}_n'\mathbf{v}\, f(y)| = o_p(1). \tag{10}$$

where

$$H_n(y) \equiv H_n(y, \beta) \equiv n^{-1}\sum_{i=1}^{n} I(e_{ni} \le y), \qquad y \in \mathbb{R}.$$

Also, we will use the notation of (2.3.1) with

$$d_{ni} \equiv n^{-1/2}, \quad X_{ni} = Y_{ni} - \mathbf{x}_{ni}'\beta, \quad F_{ni} \equiv F \ \text{ and } \ \mathbf{u} = \mathbf{0}. \tag{11}$$

Then $\mathcal{Y}_1(t, 0) \equiv n^{1/2}[H_n(F^{-1}(t)) - t]$, $0 \le t \le 1$. Write $\mathcal{Y}_1(\cdot)$ for $\mathcal{Y}_1(\cdot, 0)$.

Now, (10) and φ bounded imply that,

$$\begin{aligned} n^{1/2}\{\hat{p}_n(y) - p(y)\} &= n^{1/2}\int \{H_n(y+x) - H_n(-y+x)\}\, d\varphi(\hat{H}_n(x)) \\ &\qquad + b_n' \, v \int [f(y+x) - f(-y+x)]\, d\varphi(\hat{H}_n(x)) \\ &\qquad - n^{1/2}p(y) + \bar{o}_p(1) \\ &= R_{n1}(y) + R_{n2}(y) + R_{n3}(y) + \bar{o}_p(1), \qquad (12) \end{aligned}$$

where $\bar{o}_p(1)$ stands for a sequence of random processes that converge to zero, uniformly in $-\infty \le y \le \infty$, $\varphi \in \mathscr{C}$, in probability, and where

$$\begin{aligned} R_{n1}(y) &= \int \{Y_1(F(y+x)) - Y_1(F(-y+x))\}\, d\varphi(\hat{H}_n(x)) \\ R_{n2}(y) &= b_n' \, v \int [f(y+x) - f(-y+x)]\, d\varphi(\hat{H}_n(x)) \\ R_{n3}(y) &= n^{1/2}\{\int [F(y+x) - F(-y+x)]\, d\varphi(\hat{H}_n(x)) \\ &\qquad - \int [F(y+x) - F(-y+x)]\, d\varphi(F(x))\}, \qquad y \in \mathbb{R}. \end{aligned}$$

From (F1), (F2), the boundedness of φ, and the asymptotic continuity of Y_1, which follows from Corollary 2.2a.1, applied to the quantities given in (11), we obtain, with $k = 2a\,\|f\|_\infty$,

$$\sup_{0\le z\le a,\, \varphi\in \mathscr{C}} |R_{n1}(n^{-1/2}z)| \le \sup_{|t-s|\le kn^{-1/2}} |Y_1(t) - Y_1(s)| = o_p(1). \tag{13}$$

Again, (F1) and the boundedness of φ imply, in a routine fashion, that

$$\sup_{0\le z\le a,\, \varphi\in \mathscr{C}} |R_{n1}(n^{-1/2}z)| = o_p(1). \tag{14}$$

Now consider R_{n3}. By the MVT, (F1) and the boundedness of φ, the first term of $R_{n3}(n^{-1/2}z)$ can be written as

$$2z \int f(\xi_{xzn})\, d\varphi(\hat{H}_n(x)) = 2z \int f(x)\, d\varphi(\hat{H}_n(x)) + \bar{o}_p(1)$$

where $\{\xi_{xzn}\}$ are real numbers such that $|\xi_{xzn} - x| \le an^{-1/2}$. Do the same with the second integral and put the two together to obtain

$$\begin{aligned} R_{n3}(n^{-1/2}z) &= 2z\{\int f\, d\varphi(\hat{H}_n) - \int f\, d\varphi(F)\} + \bar{o}_p(1) \\ &= 2z\Big\{\int_0^1 [q(F\hat{H}_n^{-1}(t)) - q(t)]\, d\varphi(t)\Big\} + \bar{o}_p(1). \end{aligned}$$

But,

$$(15) \qquad \sup_{0\le t\le 1} |F\hat{H}_n^{-1}(t) - t| \le n^{-1} + \sup_y |\hat{H}_n(y) - F(y)| = o_p(1)$$

by (10) and the Glivenko-Cantelli Lemma. Hence, q being uniformly continuous, we obtain

$$\sup_{0\le z\le a,\, \varphi\in\mathscr{C}} |R_{n3}(n^{-1/2}z)| = o_p(1).$$

This together with (12) – (15) completes the proof of (8) whereas that of (9) follows from (8) and the fact that the uniform continuity of f implies that

$$\sup_{0\le z\le a,\, \varphi\in\mathscr{C}} |n^{1/2}p(n^{-1/2}z) - 2z\, Q(f)| \longrightarrow 0. \qquad \square$$

Lemma 4.5.2. *Under the assumptions of Theorem* 4.5.1, $\forall\ y \ge 0$,

$$\sup_{\varphi\in\mathscr{C}} |\hat{p}_n(y) - p(y)| = o_p(1).$$

Proof. Proceed as in the proof of the previous lemma to rewrite

$$\hat{p}_n(y) - p(y) = \Gamma_{n1}(y) + \Gamma_{n2}(y) + \Gamma_{n3}(y) + \bar{o}_p(1)$$

where $\Gamma_{nj} = n^{-1/2}R_{nj}$, $j = 1, 2, 3$, with R_{nj} defined at (12).

By Corollary 2.2a.2 applied to the quantities given at (10), $\|Y_1\|_\infty = O_p(1)$ and hence f, φ bounded trivially imply that

$$\sup_{\varphi\in\mathscr{C},\, y\ge 0} |\Gamma_{nj}(y)| = o_p(1), \qquad j = 1, 2.$$

Now, rewrite

$$\begin{aligned} \Gamma_{n3}(y) = \Big[\int F(y+x)\, d\varphi(\hat{H}_n(x)) - \int F(y+x)\, d\varphi(F(x))\Big] \\ - \Big[\int F(-y+x)\, d\varphi(\hat{H}_n(x)) - \int F(-y+x)\, d\varphi(F(x))\Big] \end{aligned}$$

$$= \Gamma_n(y) + \Gamma_n(-y), \qquad \text{say.}$$

But, $\forall\ y \in \mathbb{R}$,

$$\Gamma_n(y) = \int_0^1 \{F(y + F^{-1}(F\hat{H}_n^{-1}(t))) - F(y + F^{-1}(t))\}\, d\varphi(t) = o_p(1),$$

because of (15) and because, by **(F1)** and **(F2)**, $\forall\ y \geq 0$, $F(y + F^{-1}(t))$ is uniformly continuous function of $t \in [0, 1]$. □

Lemma 4.5.3. *Under the conditions of Theorem* 4.5.1, $\forall\ \epsilon > 0$,

$$P(|t_n^\alpha - t^\alpha| \leq \epsilon t^\alpha,\ \forall\ \varphi \in \mathscr{C}) \longrightarrow 1.$$

Proof: Observe that the event $[\hat{p}_n((1-\epsilon)t^\alpha) < \alpha \leq \hat{p}_n((1+\epsilon)t^\alpha)]$ implies the event $[(1-\epsilon)t^\alpha \leq t_n^\alpha \leq (1+\epsilon)t^\alpha]$. Hence, by two applications of Lemma 4.5.2, once with $y = (1+\epsilon)t^\alpha$, and once with $y = (1-\epsilon)t^\alpha$, we obtain that

$$\liminf_n P(|t_n^\alpha - t^\alpha| \leq \epsilon\, t^\alpha,\ \forall\ \varphi \in \mathscr{C})$$
$$\geq P\Big(p((1-\epsilon)t^\alpha) < \alpha \leq p((1+\epsilon)t^\alpha),\ \forall\ \varphi \in \mathscr{C}\Big) = 1. \qquad □$$

Proof of Theorem 4.5.1. Clearly, $\forall\ \varphi \in \mathscr{C}$,

$$|\mathcal{Q}_n^\alpha - Q(f)| = (2t_n^\alpha)^{-1}|n^{1/2}\,\hat{p}_n(n^{-1/2}\,t_n^\alpha) - 2t_n^\alpha\,Q(f)|.$$

By Lemma 4.5.3, $\forall\ \epsilon > 0$,

$$P(0 < t_n^\alpha \leq (1+\epsilon)t^\alpha,\ \forall\ \varphi \in \mathscr{C}) \longrightarrow 1.$$

Hence (7) follows from (9) applied with $a = (1+\epsilon)t^\alpha$, Lemma 5.4.3 and Slutsky's Theorem. □

Remark 4.5.1. The estimator $\mathcal{Q}_n^\alpha$ shifts the burden of choosing the window width to the choice of α. There does not seem to be an easy way to recommend a universal α. In an empirical study done in Koul, Sievers and McKean (1987) that investigated level and power of some rank tests in the linear regression setting, $\alpha = 0.8$ was found to be most desirable. □

Remark 4.5.2. It is an interesting theoretical exercise to see if, for some $0 < \delta < 1$, the processes $\{n^{1/2}(\mathcal{Q}_n^\alpha - Q(f)),\ \delta \leq \alpha \leq 1-\delta\}$ converge weakly to a Gaussian process. In the case $\varphi(t) \equiv t$, Thewarapperuma (1987) has proved, under (F1), (F2), (NX), and (4.3.17), that $\forall$ fixed $0 < \alpha < 1$, $n^{1/2}(\mathcal{Q}_n^\alpha - Q(f)) \underset{d}{\to} N(0, \sigma^2)$, where $\sigma^2 = 16\,\{\int f^3(x)dx - (\int f^2(x)dx)^2\}$. □

Remark 4.5.3. As mentioned earlier, $\{t_n^\alpha,\ \varphi \in \mathscr{C}\}$ provides a class of scale estimators for the class of scale parameters $\{t^\alpha,\ \varphi \in \mathscr{C}\}$. Recall that s_1 and s_2 of (4.3.5) are special cases of these estimators. The former is obtained by taking $\varphi(u) \equiv I(u \geq 0)$ and the latter by taking $\varphi(u) \equiv u$. For general interest we state a theorem below, giving asymptotic normality of these estimators. The details of proof are similar to those of Theorem 4.3.1. To state this theorem we need to introduce appropriately modified analogues of the entities defined at (4.3.15):

$$K_1(y) := \int [Y_1^o(y+x) - Y_1^o(-y+x)]\, d\varphi(F(x)),$$

$$K_2(y) := \int Y_1^o(x)\, [f(y+x) - f(-y+x)]\, \{f(x)\}^{-1}\, d\varphi(F(x)),$$

$$K(y) := K_1(y) - K_2(y), \qquad y \geq 0,$$

where Y_1^o is as (4.3.15) adapted to the i.i.d. errors setup. It is easy to check that $K(t^\alpha)$ is $n^{-1/2} \times$ {a sum of i.i.d. r.v.'s} with $E\,K(t^\alpha) \equiv 0$ and $0 < (\sigma^\alpha)^2 := Var(K(t^\alpha)) < \infty$, not depending on n.

Theorem 4.5.2. *In addition to the conditions of Theorem* 4.5.1, *assume that either* $\varphi(t) = I(t \geq u)$, $0 < u < 1$, *fixed or* φ *is uniformly differentiable on* $[0, 1]$. *Then,* $\forall\ 0 < \alpha < 1$,

$$n^{1/2}(t_n^\alpha - t^\alpha) \underset{d}{\to} N(0, (\nu^\alpha)^2),$$

where

$$(\nu^\alpha)^2 := (\sigma^\alpha)^2\{t^\alpha \int [f(t^\alpha+x) + f(-t^\alpha+x)]\, d\varphi(F(x))\}^{-2}. \qquad □$$

We now turn to the ***arbitrary window width and kernel–type estimators*** of Q. Accordingly, let K be a probability density on $\mathbb{R}$, h_n be a sequence of positive numbers and $\hat{\beta}$ and $\{\hat{e}_i\}$ be as before. Define

$$\hat{f}_n(x) := (nh_n)^{-1} \sum_{i=1}^{n} K\big((x - \hat{e}_i)/h_n\big),$$

$$f_n(x) := (nh_n)^{-1} \sum_{i=1}^{n} K\big((x - e_i)/h_n\big), \qquad x \in \mathbb{R},$$

$$\hat{\mathcal{Q}}_n := \int \hat{f}_n(x)\, d\varphi(\hat{H}_n(x)).$$

Theorem 4.5.3. *Assume that the model* (1.1.1) *with* $F_{ni} \equiv F$ *holds. In addition, assume that* (F1), (F2), (NX) *and* (4.3.17) *and the following hold:*

(i) $h_n > 0, \ h_n \longrightarrow 0, \ n^{1/2} h_n \longrightarrow \infty$.

(ii) K *is absolutely continuous with its* a.e. *derivative* $\dot{K}$ *satisfying* $\int |\dot{K}| < \infty$.

Then,

$$\sup_{\varphi \in \mathscr{C}} |\hat{\mathcal{Q}}_n - Q(f)| = o_p(1). \tag{16}$$

Proof. First we show $\hat{f}_n$ approximates f. This is done in several steps. To begin with, summation by parts shows that

$$\hat{f}_n(x) - f_n(x) = -h_n^{-1} \int [\hat{H}_n(x - h_n z) - H_n(x - h_n z)] \, \dot{K}(z) \, dz$$

so that

$$\|\hat{f}_n - f_n\|_\infty \le (n^{1/2} h_n)^{-1} \cdot \|n^{1/2}(\hat{H}_n - H_n)\|_\infty \cdot \int |\dot{K}|.$$

Hence, by (10) and the fact that $|b_n' v| = O_p(1)$ guaranteed by (4.3.17), it readily follows that

$$\|\hat{f}_n - f_n\|_\infty = O_p\big((n^{1/2} h_n)^{-1}\big) = o_p(1). \tag{17}$$

Now, let

$$\bar{f}_n(x) := h_n^{-1} \int K\big((x-y)/h_n\big) \, f(y) \, dy.$$

Note that integration by parts shows that

$$\bar{f}_n(x) = -h_n^{-1} \int \dot{K}(z) \, F(x - h_n z) \, dz$$

so that

$$\|f_n - \bar{f}_n\|_\infty \le (n^{1/2} h_n)^{-1} \cdot \|n^{1/2}[H_n - F]\|_\infty \cdot \int |\dot{K}| = o_p(1), \tag{18}$$

by (i) and by the fact that $\|n^{1/2}(H_n - F)\|_\infty = O_p(1)$. Moreover,

$$\|\bar{f}_n - f\|_\infty \le \sup_{|y-x| \le h_n} |f(y) - f(x)| = o(1), \qquad \text{by (F1).} \tag{19}$$

Now, consider the difference

$$\hat{Q}_n - Q(f) = \int (\hat{f}_n - f)\, d\varphi(\hat{H}_n) + \int f\, d[\varphi(\hat{H}_n) - \varphi(F)]$$

$$= D_{n1} + D_{n2}, \qquad \text{say.}$$

Let $q(t) = f(F^{-1}(t))$. Then

$$\sup_{\varphi \in \mathscr{C}} |D_{n2}| \le \sup_{0 \le t \le 1} |q(F(\hat{H}_n^{-1}(t))) - q(t)| = o_p(1)$$

by the uniform continuity of q and (15). Also,

$$\sup_{\varphi \in \mathscr{C}} |D_{n1}| \le \|\hat{f}_n - f\|_\infty = o_p(1)$$

by (17) – (19), thereby proving (16). □□