JOURNAL OF MULTIVARIATE ANALYSIS 51, 148–177 (1994)

Asymptotics of Generalized S-Estimators

OLA HÖSSJER*, CHRISTOPHE CROUX†, AND PETER J. ROUSSEEUW‡

Lund Institute of Technology, Lund, Sweden;
and University of Antwerp, Antwerp, Belgium

An S-estimator of regression is obtained by minimizing an M-estimator of scale applied to the residuals r_i. On the other hand, a generalized S-estimator (or *GS-estimator*) minimizes an M-estimator of scale based on all pairwise differences $r_i - r_j$. Generalized S-estimators have similar robustness properties as S-estimators, including a high breakdown point. In this paper we prove asymptotic normality for the GS-esimator of the regression parameters, as well as for the accompanying scale estimator defined by the minimal value of the objective function. It turns out that the asymptotic efficiency can be much higher than that of S-estimators. For instance, by using a biweight ρ-function we obtain a GS-estimator with 50% breakdown point and 68.4% efficiency.

1. INTRODUCTION

In the linear model, the observations are generated according to

$$y_i = \beta_0' \mathbf{x}_i + \sigma e_i, \qquad i = 1, ..., n, \tag{1.1}$$

where the unknown regression parameter β_0 and the carriers $\mathbf{x}_i$ are p-dimensional column vectors, σ is the unknown scale parameter and e_i the error term.

Received March 10, 1993; revised October 26, 1993

AMS 1990 subject classifications: 62F35, 62J05.

Key words and phrases: Asymptotic normality, high breakdown point, linear model, outliers, robust estimator.

* Research Associate at Lund Institute of Technology, Department of Mathematical Statistics, Box 118, S-221 00 Lund. His research was partly carried out during postdoctoral studies at Cornell University, School of OR & IE, supported by the Swedish Natural Science Research Council, contract F-DP 6689-300.

† Research Assistant with the Belgian National Fund for Scientific Research.

‡ Professor at the Department of Mathematics and Computer Science, Universitaire Instelling Antwerpen (U.I.A.), Universiteitsplein 1, B-2610 Antwerp.

0047-259X/94 $6.00

Because real-life data often contain outliers, several robust regression techniques have been developed. S-estimators (Rousseeuw and Yohai, 1984) are defined as

$$\hat{\beta}_n = \underset{\beta}{\operatorname{argmin}}\ \hat{s}_n(\beta), \tag{1.2}$$

where $\hat{s}_n(\beta)$ is an M-estimator of scale given the equation

$$\frac{1}{n}\sum_{i=1}^{n} \rho\left(\frac{r_i(\beta)}{\hat{s}_n(\beta)}\right) = k \tag{1.3}$$

for the residuals $r_i(\beta) = y_i - \beta'\mathbf{x}_i$. Under appropriate conditions on ρ and k, it has been shown that S-estimators are robust and asymptotically normal (Rousseeuw and Yohai, 1984), but there is a tradeoff between breakdown point and efficiency. If a 50% breakdown point is imposed, the asymptotic gaussian efficiency of $\hat{\beta}_n$ is at most 33% (Hössjer, 1992).

S-estimators arose as generalizations of least median of squares (LMS) regression (Rousseeuw 1984), which can be obtained as a special case by putting $\rho(u) = I(|u| \geqslant c)$ and $k = 1/2$ which yields

$$\hat{\beta}_n = \underset{\beta}{\operatorname{argmin}}\ \underset{i}{\operatorname{med}}\ |r_i|. \tag{1.4}$$

However, this step function ρ does not satisfy the condition for asymptotic normality, and indeed the LMS estimator converges at the lower rate of $n^{1/3}$ to a nonnormal distribution (Rousseeuw, 1984; Davies, 1990).

Recently, the class of *generalized S-estimators* (or *GS-estimators*) was introduced (Croux *et al.*, 1994). They are also defined by a minimization as in (1.2), but now $\hat{s}_n(\beta)$ is an M-estimator of scale computed from $\{r_i(\beta) - r_j(\beta); i < j\}$. Essentially, $\hat{s}_n(\beta)$ is the solution of

$$\binom{n}{2}^{-1} \sum_{i<j} \rho\left(\frac{r_i(\beta) - r_j(\beta)}{\hat{s}_n(\beta)}\right) = k \tag{1.5}$$

(see (2.1) for the exact definition of $\hat{s}_n(\beta)$). The robustness aspects of GS-estimators have been investigated in Croux *et al.* (1994), where the breakdown point, the maxbias curve, and the influence function were derived. Also, an algorithm for computing GS-estimators was proposed there. It remains to determine their asymptotic behaviour, which is not easy. In the present paper the asymptotic normality of $\hat{\beta}_n$ is established, yielding asymptotic efficiencies that are much higher than those of S-estimators with the same breakdown.

A prototypical example of a GS-estimator is given by $\rho(u) = I(|u| \geqslant c)$ and $k = 3/4$. We will refer to the resulting $\hat{\beta}_n$ as the *least quartile difference* (LQD) regression, because the objective function $\hat{s}_n(\beta)$ is simply the first quartile of $\{|r_i(\beta)-r_j(\beta)|;\ 1 \leqslant i \leqslant j \leqslant n\}$. The latter scale estimator was introduced by Rousseeuw and Croux (1993). The LQD has a 50% breakdown point, and its role in the class of GS-estimators is similar to that of the LMS in the class of S-estimators. This time, however, the discontinuity of ρ does not prevent the estimator $\hat{\beta}_n$ from being asymptotically normal, due to the additional smoothness caused by the convolution $r_i - r_j$. In fact, it turns out that the gaussian efficiency of the LQD is as high as 67.1%. Because of these favorable properties, and the fact that its computation time is of the same order as the LMS, we can also use the LQD as first step in the two-stage procedure of Simpson *et al.* (1992).

Other regression estimators which combine high breakdown with high efficiency are MM-estimators (Yohai, 1987) and τ-estimators (Yohai and Zamar, 1988). The basic idea of MM-estimators is to start from an initial high-breakdown regression (for which the LMS or the LQD can be used), followed by a robust estimation of σ and a constrained M-type iteration with an efficient ρ-function. On the other hand, τ-estimators are based on the minimization of a two-step scale estimator based on the residuals, which uses an M-estimator. Therefore, τ-estimators do not have an explicit objective function, in contrast with the LQD. Note that τ-estimators can also be computed by resampling and used as initial regression estimators. Croux *et al.* (1994) provide a comparison between the maxbias curves of τ- and GS-estimators.

The paper will be organized as follows. In the next section we give an exact definition of GS-estimators. In Section 3 we prove the asymptotic normality of $\hat{\beta}_n$, and of the scale estimator $\hat{s}_n$ which is defined as the minimal value of the objective function. Some asymptotic efficiencies are computed in Section 4 for various ρ-functions. Finally, some of the more technical results are collected in Appendices A and B.

2. Definition of Generalized S-Estimators

The error terms e_i are assumed to be independent and identically distributed (i.i.d.) with common distribution $F(x) = P(e_i \leqslant x)$. The carriers $\mathbf{x}_i$ are assumed to be i.i.d. observations with distribution G, and independent of the errors e_i. We will denote by K the $(p+1)$-dimensional distribution $G \times F$ of the vectors $\mathbf{z}_i = (\mathbf{x}_i, e_i)$.

As for notation, we will denote L_p-norms of (column) vectors in $\mathbb{R}^p$ by $|\cdot|_p$, $1 \leqslant p \leqslant \infty$, with $p = 2$ as a default value, and L_p norms of random variables/functions by $\|\cdot\|_p$. Frequently in the calculations, constants will be denoted C (or e.g. $C(F, G)$ to indicate dependence on other quantities), and these may vary from line to line in the same equation, unless otherwise stated.

The following regularity conditions are imposed:

(X) The distribution of the carriers satisfies $E_G|\mathbf{X}|^3 < \infty$, $E_G(\mathbf{XX}') = \Sigma$ is positive definite, and $E_G(\mathbf{X}) = 0$.

(F) The error distribution has a unimodal denisty f, which is twice continuously differentiable with a bounded second derivative.

For each β we define a generalized M-estimator of scale $\hat{s}_n(\beta)$ (cf., Serfling, 1984) as

$$\inf\left\{s; \binom{n}{2}^{-1} \sum_{i<j} \rho\left(\frac{r_i(\beta) - r_j(\beta)}{s}\right) \leqslant k\right\}, \tag{2.1}$$

where $r_i(\beta) = y_i - \mathbf{x}_i'\beta$ are the residuals,

$$k = E_{F\times F}\rho(e_1 - e_2),$$

and finally,

(R) The function ρ is even, continuous at 0, nondecreasing and right continuous $\mathbb{R}^+$, and $\rho(0) = 0$, $\rho(\infty) < \infty$, and $\rho(c) = \rho(\infty)$ for some $c > 0$.

Remark 2.1. Observe that we allow ρ to have a countable number of discontinuities on the interval $(0, c]$. The regularity conditions on ρ do not include the case when ρ is unbounded, since we want to have a nonzero breakdown point (but see Remark 3.1).

The estimate of β_0 is defined as any value

$$\hat{\beta}_n \in \Gamma_n = \{\beta; \hat{s}_n(\beta) = \hat{s}_n\}, \tag{2.2}$$

where

$$\hat{s}_n = \min_{\beta} \hat{s}_n(\beta) \tag{2.3}$$

estimates σ.

3. Asymptotic Normality

We assume from now on w.l.o.g. that $\beta_0 = 0$ and $\sigma = 1$. For notational convenience, we write

$$D_n(\beta, s) = \binom{n}{2}^{-1} \sum_{i\leqslant j} \rho\left(\frac{r_i(\beta) - r_j(\beta)}{s}\right). \tag{3.1}$$

Since $\rho(x)$ is nondecreasing in $|x|$, it follows from (2.1), (2.3), and (3.1) that

$$\Gamma_n = \{\beta;\ D_n(\beta, \hat{s}_n(\beta)) \leqslant k\}. \tag{3.2}$$

Putting

$$\Delta_n(\beta, s) = D_n(\beta, s) - D_n(0, s) \tag{3.3}$$

we then obtain

$$\begin{aligned}\tilde{\Gamma}_n &= \{\beta;\ D_n(\beta, \hat{s}_n) = \min_\beta D_n(\beta, \hat{s}_n)\}\\ &= \{\beta;\ \Delta_n(\beta, \hat{s}_n) = \min_\beta \Delta_n(\beta, \hat{s}_n)\} \subseteq \Gamma_n.\end{aligned} \tag{3.4}$$

(And $\Gamma_n = \tilde{\Gamma}_n$ when ρ is continuous.) We will first prove asymptotic normality for any sequence $\{\tilde{\beta}_n \in \tilde{\Gamma}_n\}$, and then establish that an arbitrary sequence $\{\hat{\beta}_n \in \Gamma_n\}$ is asymptotically equivalent to $\{\tilde{\beta}_n\}$ (see the proof of Theorem 3.1).

It turns out to be of great importance to study the asymptotic behaviour of $\Delta_n(\beta, s)$ in a neighborhood of (0, 1). Actually, we will show in Lemma 3.1 that locally around (0, 1), Δ_n is asymptotically equivalent to a quadratic function in β. As a preliminary, we introduce some notation. Let

$$\bar{\rho}(x) = E_F \rho(x - e) \tag{3.5}$$

and

$$\bar{\psi}(x) = \bar{\rho}'(x) = E\psi(x - e),$$

where $\psi = \rho'$ may contain a countable number of delta functions. As a local neighborhood of (0, 1) we define

$$\Omega_n = \{(\beta, s);\ |s - 1| \leqslant n^{-\tau_1},\ |\beta|_\infty \leqslant n^{-\tau_2}\}, \tag{3.6}$$

where $0 < \tau_1 < \frac{1}{2}$ and $0 < \tau_2 < \frac{1}{4}$ are fixed numbers, satisfying $\tau_1 + \tau_2 > \frac{1}{2}$ (and hence $\tau_2 < \frac{1}{4} < \tau_1$).

LEMMA 3.1. *The function $\Delta_n(\beta, s)$ defined in (3.3) may be expanded as*

$$\Delta_n(\beta, s) = \mathbf{U}_n' \beta + \frac{1}{2} \beta' \mathbf{V}_n \beta + \mathrm{Rem}_n(\beta, s), \tag{3.7}$$

where

$$\mathbf{U}_n = -\frac{2}{n} \sum_i \bar{\psi}(e_i)\, \mathbf{x}_i, \tag{3.8}$$

$$\mathbf{V}_n = \frac{2}{n} \sum_i (\mathbf{x}_i \mathbf{x}_i' + \Sigma)\, \bar{\psi}'(e_i) - 2\Sigma E_F \bar{\psi}'(e), \tag{3.9}$$

and

$$\sup_{(\beta, s) \in \Omega_n} \frac{|\text{Rem}_n(\beta, s)|}{1/n + |\beta|^2} = o_p(1). \tag{3.10}$$

In order to make use of the remainder estimate (3.10) we need to establish that $(\tilde{\beta}_n, \hat{s}_n)$ stays in Ω_n with probability tending to 1.

LEMMA 3.2. *Let* $\tilde{\beta}_n \in \tilde{\Gamma}_n$. *Then*

$$|\tilde{\beta}_n| = o_p(n^{-\tau_2}).$$

LEMMA 3.3. *Let* $\hat{s}_n$ *be defined by* (2.3). *Then*

$$|\hat{s}_n - 1| = o_p(n^{-\tau_1}). \tag{3.11}$$

We are now ready to prove asymptotic normality for $\hat{\beta}_n$:

THEOREM 3.1. *Let* $\hat{\beta}_n$ *be any vector in* Γ_n, *and let* $\mathbf{U}_n$ *and* $\mathbf{V}_n$ *be defined by* (3.8) *and* (3.9) *respectively. Then*

$$\hat{\beta}_n = -\mathbf{V}_n^{-1}\mathbf{U}_n + o_p(n^{-1/2}) \tag{3.12}$$

and hence

$$n^{1/2}\hat{\beta}_n \xrightarrow{d} N\left(0, \frac{E_F\bar{\psi}(e)^2}{(E_F\bar{\psi}'(e))^2} \Sigma^{-1}\right). \tag{3.13}$$

Remark 3.1. It follows from (3.9) that $\hat{\beta}_n$ has the same asymptotic distribution as an M-estimator with score function $\bar{\psi}$. It is possible to prove Theorem 3.1 in the case when ρ is convex, $\psi = \rho'$ is square integrable, and $0 < E_F\bar{\psi}'(e) < \infty$ by showing that the solution $\hat{\beta}_n$ of $\partial D_n(\beta, \hat{s}_n)/\partial\beta = 0$ is asymptotically normally distributed with covariance matrix according to Theorem 3.1. The local minimum of $D_n(\cdot, s)$ thus obtained will then correspond to a global one, since $D_n(\cdot, s)$ (or $\Delta_n(\cdot, s)$) is a convex function of β. This includes least squares ($\rho(x) = x^2$) and Jaeckel's estimator with Wilcoxon scores $\rho(x) = |x|$ (cf. Jaeckel, 1972).

Proof of Theorem 3.1. We start by establishing (3.12)–(3.13) for a sequence $\tilde{\beta}_n \in \tilde{\Gamma}_n$. By the Central Limit Theorem,

$$n^{1/2}\mathbf{U}_n \xrightarrow{d} N(0, 4E_F\bar{\psi}(e)^2\,\Sigma),$$

and by the law of large numbers,

$$\mathbf{V}_n \xrightarrow{p} \mathbf{V} = 2\Sigma E_F\bar{\psi}'(e). \tag{3.14}$$

Hence, (3.13) follows from (3.11). It remains to prove (3.12). Let

$$\bar{\beta}_n = -\mathbf{V}_n^{-1}\mathbf{U}_n.$$

Given $\delta, \eta > 0$, we will show that for $n \geqslant n_0$,

$$P(|\tilde{\beta}_n - \bar{\beta}_n| > \delta n^{-1/2}) \leqslant \eta, \tag{3.15}$$

which will imply (3.12). Choose A so large that

$$P(|\bar{\beta}_n| > An^{-1/2}) \leqslant \frac{\eta}{4}$$

for all n. Let $\lambda > 0$ denote the smallest eigenvalue of $\mathbf{V}$, and choose ε so small that

$$\varepsilon \leqslant \frac{\lambda}{32} \tag{3.16}$$

and

$$\frac{16\varepsilon(1+2A^2)}{\lambda} \leqslant \delta^2. \tag{3.17}$$

In view of Lemmas 3.1–3.3 and (3.14), if n_0 is large enough, $n \geqslant n_0$ implies that

$$P((\tilde{\beta}_n, \hat{s}_n) \in \Omega_n) \geqslant 1 - \frac{\eta}{4},$$

$$P\left(\sup_{(\beta, s) \in \Omega_n} \frac{|\mathrm{Rem}_n(\beta, s)|}{\frac{1}{n} + |\beta|^2} > \varepsilon\right) \leqslant \frac{\eta}{4} \tag{3.18}$$

and finally

$$P\left(\mathbf{V}_n - \frac{1}{2}\mathbf{V} \text{ is positive definite}\right) \geqslant 1 - \frac{\eta}{4}.$$

Hence, it follows from (3.7) and (3.18) that with probability at least $1 - \eta$,

$$\begin{aligned}\Delta_n(\beta, \hat{s}_n) - \Delta_n(\bar{\beta}_n, \hat{s}_n) &= \frac{1}{2}(\beta - \bar{\beta}_n)' \, \mathbf{V}_n(\beta - \bar{\beta}_n) + \mathrm{Rem}_n(\beta, \hat{s}_n) - \mathrm{Rem}_n(\bar{\beta}_n, \hat{s}_n) \\ &\geqslant \frac{\lambda}{4} |\beta - \bar{\beta}_n|^2 - 2\varepsilon\left(\frac{1}{n} + |\beta|^2\right) \\ &\geqslant \left(\frac{\lambda}{4} - 4\varepsilon\right) |\beta - \bar{\beta}_n|^2 - 2\varepsilon\left(\frac{1}{n} + \frac{2A^2}{n}\right),\end{aligned}$$

which is greater than 0 as soon as

$$|\beta - \tilde{\beta}_n| \geqslant \sqrt{\frac{2\varepsilon(1+2A^2)}{\lambda/4 - 4\varepsilon}}\, n^{-1/2}. \tag{3.19}$$

Since the right hand side of (3.19) is dominated by $\delta n^{-1/2}$ because of (3.16) and (3.17), we have proved (3.15).

It remains to show that $\{\tilde{\beta}_n \in \tilde{\Gamma}_n\}$ is asymptotically equivalent to $\{\hat{\beta}_n \in \Gamma_n\}$. If not, we can find $\delta' > 0$ and $\eta' > 0$ such that

$$P\left(|\hat{\beta}_{n_i} - \tilde{\beta}_{n_i}| > \frac{\delta'}{\sqrt{n}}\right) > \eta'$$

for some subsequence $\{n_i\}$. By (3.2) and Lemma 3.1, this implies that

$$A_{n_i} = \left\{\beta;\, |\beta - \tilde{\beta}_{n_i}| \leqslant \frac{\delta'}{2\sqrt{n_i}}\right\} \subseteq \Gamma_{n_i}$$

with probability $> \eta'/2$ for all n_i large enough. However, this would imply that $\hat{s}_{ni}(\cdot)$ is constant on A_{n_i} with probability larger than $\eta'/2$. Since

$$\sup_{\beta \in \tilde{\Gamma}_n} (k - D_n(\beta, \hat{s}_n)) = o_p(n^{-1/2}),$$

by Lemma 3.1, 3.3 and 3.4 and the $\sqrt{n}$-consistency of $\tilde{\beta}_n$, this would imply that $D_{n_i}(\cdot, \hat{s}_{n_i})$ is essentially constant on A_{n_i} with probability exceeding $\eta'/2$, which contradicts Lemmas 3.1 and 3.3. ∎

Our next objective is to establish asymptotic normality for $\hat{s}_n$. For this we will need an asymptotic linearity result for $D_n(0, s)$ as a function of s. Let $\psi = \rho'$ and put

$$F^*(x) = P_{F \times F}(e_1 - e_2 \leqslant x).$$

LEMMA 3.4. *Let D_n be defined by* (3.1). *Then*

$$\sup_{s;\, |s-1| \leqslant n^{-r_1}} |D_n(0, s) - D_n(0, 1) + E_{F^*}(\psi(e)e)(s-1)| = o_p(n^{-1/2}). \tag{3.20}$$

LEMMA 3.5. *Let $\hat{s}_n(0)$ be defined as the solution of* (2.1), *with* $\beta = 0$. *Then*

$$n^{1/2}(\hat{s}_n(0) - 1) \xrightarrow{d} N\left(0, \frac{4\,\mathrm{Var}_F(\bar{\rho}(e))}{(E_{F^*}(e\psi(e)))^2}\right), \tag{3.21}$$

where $\bar{\rho}$ is defined in (3.5).

Proof. Since $D_n(\beta, s)$ is a U-statistic, it follows from e.g. Serfling (1980, Chap. 5), that

$$n^{1/2}(D_n(0,1)-k) \xrightarrow{d} N(0, 4\,\mathrm{Var}_F(\bar{\rho}(e))). \tag{3.22}$$

Put $B = E_{F^*}(e\psi(e))$ and let

$$R = \sup_{s;\,|s-1| \leqslant n^{-\eta}} |D_n(0,s) - D_n(0,1) + B(s-1)|.$$

Then $R = o_p(n^{-1/2})$ by Lemma 3.4. Since $D_n(0,s)$ is a decreasing function of s, it follows easily from (3.22) and Lemma 3.4 that $\hat{s}_n(0)$ is $n^{1/2}$-consistent. Since $D_n(0, \hat{s}_n(0)) \leqslant k$, it follows that

$$n^{1/2}(\hat{s}_n(0)-1) \geqslant n^{1/2}\frac{D_n(0,1)-k}{B} - n^{1/2}\frac{R}{B}$$

with probability tending to one. For any $\varepsilon > 0$,

$$D_n(0, \hat{s}_n(0) - \varepsilon n^{-1/2}) > k$$

according to (2.1). Hence, with probability tending to one,

$$n^{1/2}(\hat{s}_n(0) - \varepsilon n^{-1/2} - 1) \leqslant n^{1/2}\frac{D_n(0,1)-k}{B} + n^{1/2}\frac{R}{B},$$

simultaneously for all sufficiently small ε. The asserted asymptotic normily follows from (3.22) by letting $\varepsilon \to 0+$. ∎

THEOREM 3.2. *Let $\hat{s}_n$ and $\hat{s}_n(0)$ be as defined in* (2.3) *and* (2.1). *Then*

$$\hat{s}_n = \hat{s}_n(0) + o_p(n^{-1/2}), \tag{3.23}$$

and hence

$$n^{1/2}(\hat{s}_n - 1) \xrightarrow{d} N\left(0, \frac{4\,\mathrm{Var}_F(\bar{\rho}(e))}{(E_{F^*}(e\psi(e)))^2}\right). \tag{3.24}$$

Proof. It is clear that (3.24) follows from (3.23) and Lemma 3.5, so it remains to prove (3.23). By Lemma 3.4,

$$k - D_n(0, \hat{s}_n(0)) = o_p(n^{-1/2})$$

and as in the proof of Theorem 3.1 one shows that

$$\sup_{\beta \in \Gamma_n} (k - D_n(\beta, \hat{s}_n)) = o_p(n^{-1/2}).$$

It follows then from Lemma 3.3–3.5 that

$$\begin{aligned} B(\hat{s}_n(0)-\hat{s}_n) &= D_n(0,\hat{s}_n) - D(0,\hat{s}_n(0)) + o_p(n^{-1/2}) \\ &= D_n(0,\hat{s}_n) - D_n(\hat{\beta}_n,\hat{s}_n) + o_p(n^{-1/2}) \\ &= -\mathbf{U}_n'\hat{\beta}_n - \tfrac{1}{2}\hat{\beta}_n'\mathbf{V}_n\hat{\beta}_n - \mathrm{Rem}_n(\hat{\beta}_n,\hat{s}_n) + o_p(n^{-1/2}) \\ &= o_p(n^{-1/2}), \end{aligned}$$

where the last equality from Lemma 3.1, Lemma 3.3 and Theorem 3.1. ∎

4. Asymptotic Effienciencies

According to (3.13), the gaussian efficiency of a GS-estimator is given by

$$e = \frac{(\int \bar{\psi}'(y)\,d\Phi(y))^2}{\int \bar{\psi}(y)^2\,d\Phi(y)}. \tag{4.1}$$

For $\rho(x) = I(|x| \geqslant c)$ the efficiency is given by

$$\frac{\sqrt{3}}{4}\frac{c^2}{\exp(c^2/6) - \exp(-c^2/2)}. \tag{4.2}$$

For this estimator the objective function corresponds to the α-th quantile of the distances $|r_i - r_j|$, where in order for $\hat{s}_n$ to be consistent we must have

$$\alpha = 1 - E_{\Phi\times\Phi}\,\rho(e_1 - e_2) = P_{\Phi\times\Phi}(|e_1 - e_2| \leqslant c) = 2\Phi\left(\frac{c}{\sqrt{2}}\right) - 1. \tag{4.3}$$

We will denote this estimator as LQD(α), noting that the LQD estimator of Section 1 corresponds to $\alpha = 0.25$. In Fig. 1a we made a plot of the efficiency versus the quantile. Note that the efficiency gets very close to 1: the maximal value is attained for $\alpha = 0.8993$ where the efficiency is 97.73% and the breakdown point is $\varepsilon^* = 5.1\,\%$. Surprisingly, for $\alpha \to 0$ the efficiency does not tend to 0. The same thing was noted for the corresponding scale estimators (Rousseeuw and Croux, 1992, p. 80).

It is shown in Croux *et al.* (1994) that $\varepsilon^* = \min(\sqrt{\alpha}, 1 - \sqrt{\alpha}) \leqslant 1/2$. This means that for any value of ε^* there are always two possible values of α, but we will systematically use the larger of the two (hence, $\alpha \geqslant 0.25$). In Table I, we give for some values of ε^* the corresponding α, c, and efficiency.

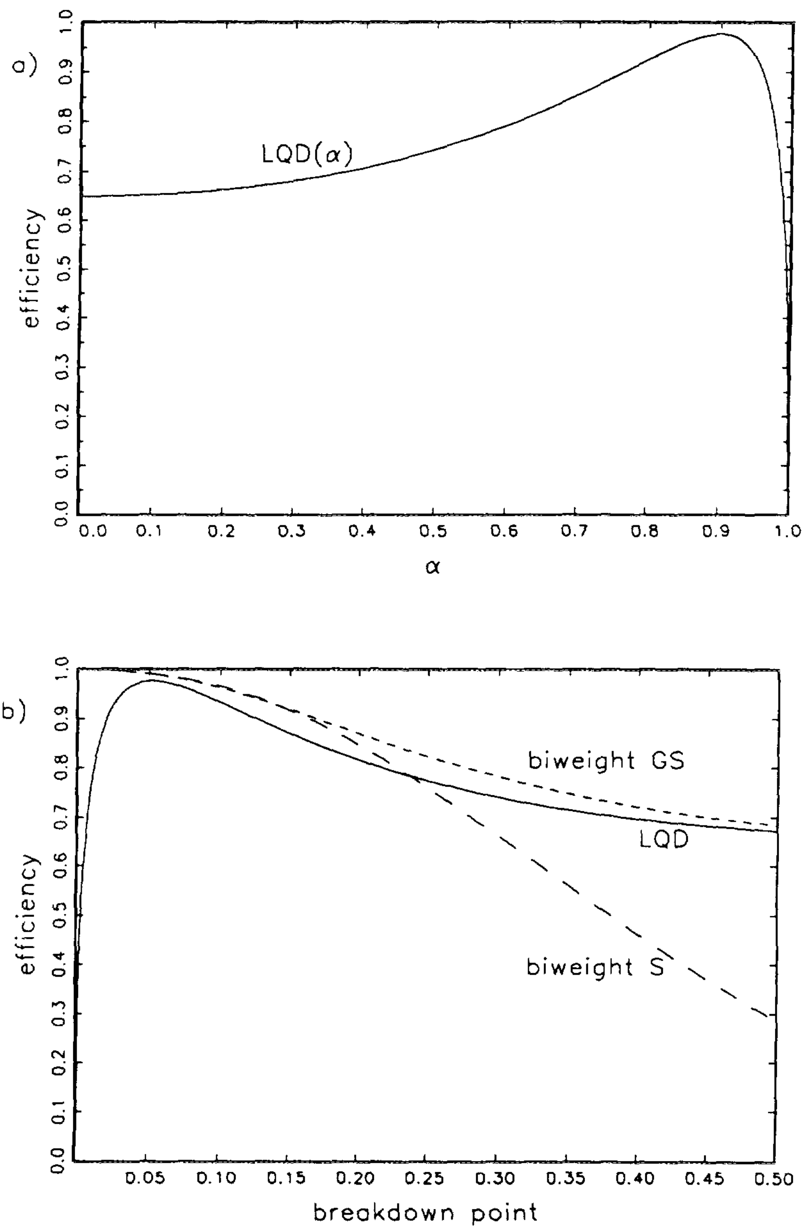

FIG. 1. Efficiency as a function of (a) the quantile α of the LQD(α) estimator; (b) the breakdown point of S- and GS-estimators using a biweight ρ-function.

TABLE I

Breakdown Point and Efficiency for GS-Estimators of Type LQD(α)

ε^*	α	c	e
0.5000	0.2500	0.4506	0.6714
0.4500	0.3025	0.5497	0.6819
0.4000	0.3600	0.6614	0.6962
0.3500	0.4225	0.7878	0.7152
0.3000	0.4900	0.9317	0.7403
0.2500	0.5625	1.0980	0.7731
0.2000	0.6400	1.2945	0.8157
0.1500	0.7225	1.5358	0.8699
0.1000	0.8100	1.8534	0.9336

Table II provides analogous results for GS-estimators using Tukey's biweight ρ-function:

$$\rho(x) = \begin{cases} \dfrac{x^2}{2} - \dfrac{x^4}{2c^2} + \dfrac{x^6}{6c^4} & \text{for } |x| \leqslant c \\ \dfrac{c^2}{6} & \text{for } |x| \geqslant c, \end{cases} \tag{4.4}$$

where the choice of the tuning constant c determines the breakdown point and the efficiency. By comparing Table II with the corresponding table for S-estimators (Rousseeuw and Yohai, 1984, p. 268) we see that GS-estimators are much more efficient.

TABLE II

Breakdown Point and Efficiency for GS-Estimators Using a Biweight ρ-Function

ε^*	k	c	e
0.5000	0.1240	0.9958	0.6837
0.4500	0.1733	1.2210	0.6998
0.4000	0.2335	1.4795	0.7209
0.3500	0.3047	1.7793	0.7480
0.3000	0.3867	2.1330	0.7819
0.2500	0.4786	2.5619	0.8228
0.2000	0.5787	3.1056	0.8697
0.1500	0.6843	3.8466	0.9192
0.1000	0.7921	5.0012	0.9636

In Figure 1b we plotted the efficiency versus the breakdown point for biweight GS-estimators, together with biweight S-estimators and LQD(α). Note that the biweight GS-estimator is always the most efficient of the three, and that LQD(α) performs almost as well (except for ε^* less than 5%, where LQD(α) is not recommended).

APPENDIX A. PROOF OF LEMMA 3.1

For ease of notation, let us introduce $\theta = (\beta, s)$ as a $(p+1)$-dimensional parameter. We then observe that $\Delta_n(\theta)$ introduced in (3.3) is a U-statistic,

$$\Delta_n(\theta) = \binom{n}{2}^{-1} \sum_{i<j} h(\theta; \mathbf{z}_i, \mathbf{z}_j), \tag{A.1}$$

where $\mathbf{z}_i = (\mathbf{x}_i, e_i)$ and the kernel function h is given by

$$h(\theta; \mathbf{z}_1, \mathbf{z}_2) = \rho\left(\frac{e_1 - e_2 - (\mathbf{x}_1 - \mathbf{x}_2)'\beta}{s}\right) - \rho\left(\frac{e_1 - e_2}{s}\right)$$

Next, we decompose (A.1) as

$$\Delta_n(\theta) = -h_0(\theta) + \frac{2}{n}\sum_{i=1}^{n} h_1(\theta; \mathbf{z}_i) + \binom{n}{2}^{-1} \sum_{i<j} h_2(\theta; \mathbf{z}_i, \mathbf{z}_j), \tag{A.2}$$

where

$$h_0(\theta) = E_{K\times K} h(\theta; \mathbf{Z}_1, \mathbf{Z}_2),$$
$$h_1(\theta; \mathbf{z}) = E_K h(\theta; \mathbf{z}, \mathbf{Z})$$

and

$$h_2(\theta; \mathbf{z}_1, \mathbf{z}_2) = h(\theta; \mathbf{z}_1, \mathbf{z}_2) - h_1(\theta; \mathbf{z}_1) - h_1(\theta; \mathbf{z}_2) + h_0(\theta).$$

We start by considering h_1. To this end, introduce $\bar{\rho}_s(x) = E_F \rho((x-e)/s)$ and

$$\bar{\psi}_s(x) = \bar{\rho}_s'(x), \tag{A.3}$$

so that $s = 1$ corresponds to $\bar{\rho}$ and $\bar{\psi}$ respectively (cf. (3.5)). After some calculations one finds

$$h_1(\theta; \mathbf{z}) = -\mathbf{x}'\beta\bar{\psi}(e) + \frac{1}{2}\beta'(\mathbf{x}\mathbf{x}' + \Sigma)\beta\bar{\psi}'(e) + \sum_{k=1}^{5} R^{1k}(\theta; \mathbf{z}), \tag{A.4}$$

where

$$R^{11}(\theta; \mathbf{z}) = E_G \bar{\rho}_s(e - \mathbf{x}'\beta + \mathbf{X}'\beta) - \bar{\rho}_s(e - \mathbf{x}'\beta) - \tfrac{1}{2}\beta'\Sigma\beta\bar{\psi}'_s(e - \mathbf{x}'\beta), \tag{A.5}$$

$$R^{12}(\theta; \mathbf{z}) = \bar{\rho}_s(e - \mathbf{x}'\beta) - \bar{\rho}_s(e) + \mathbf{x}'\beta\bar{\psi}_s(e) - \tfrac{1}{2}(\mathbf{x}'\beta)^2\,\bar{\psi}'_s(e), \tag{A.6}$$

$$R^{13}(\theta; \mathbf{z}) = \tfrac{1}{2}\beta'\Sigma\beta(\bar{\psi}'_s(e - \mathbf{x}'\beta) - \bar{\psi}'_s(e)), \tag{A.7}$$

$$R^{14}(\theta; \mathbf{z}) = \tfrac{1}{2}\beta'(\mathbf{x}\mathbf{x}' + \Sigma)\,\beta(\bar{\psi}'_s(e) - \bar{\psi}'(e)) \tag{A.8}$$

and

$$R^{15}(\theta; \mathbf{z}) = -\mathbf{x}'\beta(\bar{\psi}_s(e) - \bar{\psi}(e)). \tag{A.9}$$

By taking expectations in (A.4), and using the fact that $E_F\bar{\psi}(e) = 0$, one obtains

$$h_0(\theta) = \beta'\Sigma\beta E_F\bar{\psi}'(e) + R^0(\theta), \tag{A.10}$$

where

$$R^0(\theta) = \sum_{k=1}^{5} E_K R^{1k}(\theta; \mathbf{Z}). \tag{A.11}$$

If we now put

$$R_n^{1k}(\theta) = \frac{2}{n}\sum_{i=1}^{n} R^{1k}(\theta; \mathbf{z}_i) \tag{A.12}$$

and

$$R_n^2(\theta) = \binom{n}{2}^{-1} \sum_{i<j} h_2(\theta; \mathbf{z}_i, \mathbf{z}_j), \tag{A.13}$$

it follows from (A.1)–(A.2), (A.4), and (A.12)–(A.13) that (3.7) holds, with

$$\mathrm{Rem}_n(\theta) = -R^0(\theta) + \sum_{k=1}^{5} R_n^{1k}(\theta) + R_n^2(\theta). \tag{A.14}$$

We will prove Lemma 3.1 through a series of lemmas, starting with some preliminary properties of $\bar{\psi}_s$:

LEMMA A.1. *Let $\bar{\psi}_s$ be defined by (A.3). Then $\bar{\psi}_s$ is twice continuously differentiable,*

$$\|\bar{\psi}''_s\|_\infty \leqslant \|f''\|_\infty\, 2\rho(\infty), \tag{A.15}$$

$$|\bar{\psi}'_s(x) - \bar{\psi}'(x)| \leqslant 2\|f\|_\infty \int_0^\infty y\,d\rho(y)\,|s-1|, \tag{A.16}$$

and

$$E_F(\bar{\psi}_{s_1}(e) - \bar{\psi}_{s_2}(e))^2 \leqslant 4\|f'\|_\infty^2 \left(\int_0^\infty y d\rho(y)\right)^2 |s_1 - s_2|^2. \qquad (A.17)$$

Proof. Integration by parts shows that

$$\bar{\rho}_s(x) = \rho(\infty) + \int F(x - sy)\, d\rho(y). \qquad (A.18)$$

The regularity conditions on F allow us to differentiate three times under the integral sign in (A.18), obtaining

$$\bar{\psi}_s^{(k)}(x) = \int f^{(k)}(x - sy)\, d\rho(y),\ k = 0, 1, 2, \qquad (A.19)$$

with (k) denoting the kth derivative. Formulas (A.15)–(A.17), as well as the fact that $\bar{\psi}_s''$ is continuous, now easily follow from (A.19). ∎

LEMMA A.2. *Let $R^0(\theta)$ and $R^{1k}(\theta)$ be defined as in* (A.11) *and* (A.12). *Then there exists a nonnegative random variable W with finite expectation such that*

$$\left|\sum_{k=1}^{4} R_n^{1k}(\theta)\right| \leqslant W n^{-\tau_2} |\beta|^2 \qquad (A.20)$$

and

$$|R^0(\theta)| \leqslant \tfrac{1}{2} E(W)\, n^{-\tau_2} |\beta|^2 \qquad (A.21)$$

for all $\theta \in \Omega_n$. Moreover,

$$\sup_{\theta \in \Omega_n} |R_n^{15}(\theta)| = O_p(n^{-\tau_1 - \tau_2 - 1/2}) = o_p(n^{-1}). \qquad (A.22)$$

Proof. Since $E_G \mathbf{X} = 0$, we obtain from (A.5)–(A.8), (A.15), and a Taylor expansion that

$$|R^{11}(\theta; \mathbf{z})| \leqslant \tfrac{1}{6} \|\bar{\psi}_s''\|_\infty E_G |\mathbf{X}|^3 |\beta|^3 \leqslant \tfrac{1}{3} \rho(\infty) \|f''\|_\infty E_G |\mathbf{X}|^3 |\beta|^3, \qquad (A.23)$$

$$|R^{12}(\theta; \mathbf{z})| \leqslant \tfrac{1}{3} \rho(\infty) \|f''\|_\infty |\mathbf{x}|^3 |\beta|^3, \qquad (A.24)$$

$$|R^{13}(\theta; \mathbf{z})| \leqslant \rho(\infty) \|f''\|_\infty (E_G |\mathbf{X}|^2)\, |\mathbf{x}|\, |\beta|^3 \qquad (A.25)$$

and finally, from (A.16),

$$|R^{14}(\theta; \mathbf{z})| \leqslant \tfrac{1}{2}(|\mathbf{x}|^2 + E_G |\mathbf{X}|^2)\, |\bar{\psi}_s'(e) - \bar{\psi}'(e)|\, |\beta|^2$$

$$\leqslant (|\mathbf{x}|^2 + E_G |\mathbf{X}|^2) \|f''\|_\infty \int_0^\infty y d\rho(y)\, |s - 1|\, |\beta|^2. \qquad (A.26)$$

Formula (A.20) now follows from (A.12), (A.23)–(A.26) and the fact that $\tau_2 < \tau_1$. Next, (A.21) follows from (A.11) and taking expectations in (A.20). It remains to prove (A.22). Observe that

$$R_n^{15}(\theta) = \mathbf{T}_n(s)'\,\beta,$$

where $\mathbf{T}_n(s) = (T_{n1}(s), ..., T_{np}(s))'$,

$$T_{nj}(s) = -\frac{2}{n}\sum_{i=1}^{n} x_{ij}(\bar{\psi}_s(e_i) - \bar{\psi}(e_i))$$

and $\mathbf{x}_i = (x_{i1}, ..., x_{ip})'$. Obviously, it suffices to show

$$\sup_{|s-1| \leqslant n^{-\tau_1}} |T_{nj}(s)| = O_p(n^{-\tau_1 - 1/2}), \qquad j = 1, ..., p$$

and we consider w.l.o.g. only the case $j = 1$, $s \geqslant 1$. Define

$$\bar{T}_m = \max_{0 \leqslant l \leqslant m} |T_{n1}(s_{lm})|,$$

where

$$s_{lm} = 1 + \frac{ln^{-\tau_1}}{m}, \qquad l = 0, 1, ..., m.$$

We may regard

$$T_{n1}(s_{lm}) = \sum_{k=1}^{l} (T_{n1}(s_{km}) - T_{n1}(s_{(k-1)m})), \qquad l = 1, ..., m$$

as a sequence of partial sums, with $T_{n1}(s_{0m}) = 0$ and

$$E(T_{n1}(s_{lm}) - T_{n1}(s_{km}))^2 \leqslant \frac{16}{n} E_G|\mathbf{X}|^2\, \|f'\|_\infty^2 \left(\int_0^\infty y\,d\rho(y)\right)^2 (s_{lm} - s_{km})^2, \qquad 0 \leqslant k \leqslant l \leqslant m, \tag{A.27}$$

because of (A.17) and the fact that $E_F\bar{\psi}_s(e) = 0$. It now follows from (A.27) and Billingsley (1968), Theorem 12.2, that

$$P(\bar{T}_m \geqslant \lambda) \leqslant \frac{C}{\lambda^2}\frac{s_{mm}^2}{n} \leqslant \frac{C'}{\lambda^2} n^{-1-2\tau_1}, \tag{A.28}$$

where C and C' are constants (depending on F, ρ and G). We may also rewrite (A.28) as

$$P(\bar{T}_m \geqslant \lambda n^{-\tau_1 - 1/2}) \leqslant \frac{C'}{\lambda^2}. \tag{A.29}$$

Since the upper bound in (A.29) is independent of m and

$$\bar{T}_m \to \sup_{1 \leqslant s \leqslant 1+n^{-\tau_1}} |T_{n1}(s)| \quad \text{as} \quad m \to \infty,$$

it follows that

$$P\left(\sup_{1 \leqslant s \leqslant 1+n^{-\tau_1}} |T_{n1}(s)| \geqslant \lambda n^{-\tau_1 - 1/2} \right) \leqslant \frac{C'}{\lambda^2},$$

and we are done. ∎

We still have to show that $nR_n^2(\theta)$ is uniformly small in probability over Ω_n. Theorem 9 in Nolan and Pollard (1987) treats suprema of families of degenerate U-statistics. In order to apply this theorem we have to assume that the carrier distribution G has bounded support, which is too restrictive for our purposes. Instead, we will construct a direct proof. For this we first need the following lemma, which is reminiscent of Lemma B, p. 186, in Serfling (1980).

LEMMA A.3. *Let $W_1, ..., W_n$ be i.i.d. q-dimensional random vectors and let $g: \mathbb{R}^q \times \mathbb{R}^q \to \mathbb{R}$ be a symmetric kernel function such that $E_g(W_1, W_2) = 0$,*

$$\xi_1 = E_{W_2}(E_{W_1} g(W_1, W_2))^2 = 0,$$

$Eg(W_1, W_2)^2 < \infty$ and $\|g\|_\infty \leqslant M$. Define a U-statistic by

$$T_n = \binom{n}{2}^{-1} \sum_{i<j} g(W_i, W_j), \tag{A.30}$$

and let r be an even positive integer such that

$$E\,|g(W_1, W_2|^s \leqslant \kappa_n, \qquad s = 1, ..., r, \tag{A.30}$$

where $\{\kappa_n\}$ is a bounded sequence. Then there exists a constant C_r such that

$$E\,|T_n|^r \leqslant C_r M^r n^{-r} \max(\kappa_n, n^{-1})^{r/4}. \tag{A.31}$$

Proof. We may assume $M = 1$ w.l.o.g. (otherwise, replace g by g/M). We observe that

$$E\,|T_n|^r \leqslant \binom{n}{2}^{-r} \sum_{i} E \prod_{k=1}^{r} g(W_{i(k)}, W_{j(k)}), \tag{A.32}$$

where $\mathbf{i}$ denotes the multi-index $((i(1), j(1)), ..., (i(r), j(r)))$, so that there are $\binom{n}{2}^r = O(n^{2r})$ choices of $\mathbf{i}$. However, since $\xi_1 = 0$, only those multi-indices for which

$$|\{k; 1 \leqslant k \leqslant r, i(k) = m \text{ or } j(k) = m\}| \neq 1, \qquad m = 1, ..., n, \tag{A.33}$$

will give nonzero contributions to the sum in (A.32). We will make use of the functions

$$K(\mathbf{i}) = \text{maximal number of pairwise disjoint } (i(k), j(k)) \text{ in } \mathbf{i}$$

and

$$L(\mathbf{i}) = |\{i(k), j(k); k = 1, ..., r\}|.$$

Let I denote the set of those $\mathbf{i}$ satisfying (A.33). Then

$$1 \leqslant K(\mathbf{i}) \leqslant r \text{ and } 2 \leqslant L(\mathbf{i}) \leqslant r \qquad \text{for any } \mathbf{i} \in I. \tag{A.34}$$

Given a specific $\mathbf{i}$ with $K(\mathbf{i}) = K$, we suppose w.l.o.g. that

$$(i(k), j(k)) = (2k-1, 2k), \qquad k = 1, ..., K. \tag{A.35}$$

We then have, since $\|g\|_\infty \leqslant 1$,

$$\begin{aligned} \left| E \prod_{k=1}^{r} g(W_{i(k)}, W_{j(k)}) \right| &\leqslant E \prod_{k=1}^{K} |g(W_{2k-1}, W_{2k})| \\ &= \prod_{k=1}^{K} E|g(W_{2k-1}, W_{2k})| \leqslant \kappa_n^K. \end{aligned} \tag{A.36}$$

Let

$$J = \{(K, L); K = K(\mathbf{i}) \text{ and } L = L(\mathbf{i}) \text{ for some } \mathbf{i} \in I\}.$$

Then, if we combine (A.32) and (A.36) and use the fact that there are $O(n^L)$ multi-indices with $L(\mathbf{i}) = L$ (with a multiplicative constant only depending on r), we obtain

$$\begin{aligned} E|T_n|^r &\leqslant C_r n^{-2r} \sum_{\mathbf{i} \in I} \left| E \prod_{k=1}^{r} g(W_{i(k)}, W_{j(k)}) \right| \\ &\leqslant C_r' n^{-2r} \sum_{(K, L) \in J} n^L \kappa_n^K \\ &\leqslant C_r' n^{-r} \sum_{(K, L) \in J} \max\left(\frac{1}{n}, \kappa_n\right)^{r-L+K}. \end{aligned}$$

It thus remains to prove that (since $\{\kappa_n\}$ is a bounded sequence)

$$\min\{r - L + K;\ (K, L) \in J\} \geqslant \frac{r}{4}. \tag{A.37}$$

To this end, we fix **i** and suppose that (A.35) holds. Suppose also w.l.o.g. that

$$\{i(k), j(k);\ k = 1, ..., r\} = \{1, ..., L\},$$

where $L = L(\mathbf{i})$. Then, because of (A.33), there are at least $2(L - 2K)$ indices $i(k)$ or $j(k)$ (with $k > K$) belonging to $\{2K + 1, ..., L\}$. Moreover, by the definition of $K(\mathbf{i})$ and (A.35), at least one of the indices $i(k)$ and $j(k)$ must be less than or equal to $2K$. (Otherwise we would have $K(\mathbf{i}) > K$.) Hence there are also at least $2(L - 2K)$ indices $i(k)$ or $j(k)$ with $k > K$ that belong to $\{1, ..., 2K\}$. To summarize the discussion, we obtain a lower bound for the total number of indices ($= 2r$) in **i**,

$$2r \geqslant 2K + 2(L - 2K) + 2(L - 2K) = 4L - 6K$$

$$\Leftrightarrow r - L + K \geqslant L - 2K. \tag{A.38}$$

By (A.34), we can improve (A.38) to

$$r - L + K \geqslant \max(L - 2K, K, r - L + 1). \tag{A.39}$$

Suppose that the right hand side of (A.39) is less than $r/4$, so that $K < r/4$ and $L > 3r/4 + 1$. But then $L - 2K > (3r/4 + 1) - r/2 > r/4$, a contradiction, and thus we have proved (A.37). ∎

LEMMA A.4. *Let $R_n^2(\theta)$ be defined by* (A.13). *Then*

$$\sup_{\theta \in \Omega_n} |R_n^2(\theta)| = o_p\left(\frac{1}{n}\right).$$

Proof. We divide the parameter set Ω_n into a union of blocks, the centers of which form the lattice

$$\Gamma_n = \Big\{\theta_{\mathbf{i}} = 2\delta_n(i_1, ..., i_p, i_{p+1}) + (0, ..., 0, 1);$$

$$i_1, ..., i_{p+1} \in Z,\ |i_j| \leqslant \frac{n^{-\tau_2}}{\delta_n}, j = 1, ..., p \text{ and } |i_{p+1}| \leqslant \frac{n^{-\tau_1}}{\delta_n}\Big\}, \tag{A.41}$$

where $\mathbf{i} = (i_1, ..., i_{p+1})$ and $\delta_n = n^{-\tau}$, where $\tau > 1$. Around each lattice point $\theta_{\mathbf{i}}$, we form the block

$$\Omega_{n,\mathbf{i}} = \{\theta \in \Omega_n;\, |\theta - \theta_{\mathbf{i}}|_\infty \leqslant \delta_n\}, \tag{A.42}$$

so that $\Omega_n \subset \bigcup_{\mathbf{i}} \Omega_{n,\mathbf{i}}$. We first notice that the supremum in (A.40) may be upper bounded according to

$$\sup_{\theta \in \Omega_n} |R_n^2(\theta)| \leqslant \max_{\theta_{\mathbf{i}} \in \Gamma_n} |R_n^2(\theta_{\mathbf{i}})| + \max_{\theta_{\mathbf{i}} \in \Gamma_n} \bar{R}_n^2(\theta_{\mathbf{i}}), \tag{A.43}$$

where

$$\bar{R}_n^2(\theta_{\mathbf{i}}) = \sup_{\theta \in \Omega_{n,\mathbf{i}}} |R_n^2(\theta) - R_n^2(\theta_{\mathbf{i}})| \tag{A.44}$$

is an upper bound for the variation of $R_n^2(\theta)$ within each block.

Let $R_n^2(a;\theta)$ be defined as $R_n^2(\theta)$ in (A.13), but with $\rho_a(x) = I(|x| \geqslant a)$ in place of $\rho(x)$. Then $R_n^2(\theta)$ is a mixture of various $R_n^2(a;\theta)$ since (as is easily verified)

$$R_n^2(\theta) = \int_0^\infty R_n^2(a;\theta)\, d\rho(a).$$

An appeal to Minkowski's inequality gives for any $r \geqslant 1$,

$$\|R_n^2(\theta)\|_r \leqslant \int_0^\infty \|R_n^2(a;\theta)\|_r\, d\rho(a). \tag{A.45}$$

If $r \geqslant 1$ we also obtain

$$\begin{aligned}(E \max_{\theta_{\mathbf{i}} \in \Gamma_n} |R_n^2(\theta_{\mathbf{i}})|)^r &\leqslant E(\max_{\theta_{\mathbf{i}} \in \Gamma_n} |R_n^2(\theta_{\mathbf{i}})|)^r \leqslant E \max_{\theta_{\mathbf{i}} \in \Gamma_n} |R_n^2(\theta_{\mathbf{i}})|^r \\ &\leqslant E \sum_{\theta_{\mathbf{i}} \in \Gamma_n} |R_n^2(\theta_{\mathbf{i}})|^r \leqslant |\Gamma_n| \max_{\theta_{\mathbf{i}} \in \Gamma_n} E|R_n^2(\theta_{\mathbf{i}})|^r, \end{aligned} \tag{A.46}$$

by the convexity of $x \to |x|^r$ and Jensen's inequality. By taking $1/r$th powers on both sides of (A.46) and using (A.45) we obtain

$$\begin{aligned} E \max_{\theta_{\mathbf{i}} \in \Gamma_n} |R_n^2(\theta_{\mathbf{i}})| &\leqslant |\Gamma_n|^{1/r} \max_{\theta_{\mathbf{i}} \in \Gamma_n} \|R_n^2(\theta_{\mathbf{i}}\|_r \\ &\leqslant C n^{((\tau - \tau_2)p + (\tau - \tau_1))/r} \rho(\infty) \sup_{a,\, \theta_{\mathbf{i}}} \|R_n^2(a;\theta_{\mathbf{i}})\|_r, \end{aligned} \tag{A.47}$$

where the supremum on the last line of (A.47) ranges over $\theta_{\mathbf{i}} \in \Gamma_n$ and $0 < a \leqslant c$, where c is defined in (R). We will estimate the RHS of (A.47) by using Lemma A.3, and therefore we need to estimate the rth moments of the kernel functions involved. Define $h(a; \theta; \mathbf{z}_1, \mathbf{z}_2)$ as $h(\theta; \mathbf{z}_1, \mathbf{z}_2)$, but with $\rho(x)$ replaced by $\rho_a(x)$. Define in the same way $h_0(a; \theta)$, $h_1(a; \theta; \mathbf{z})$ and $h_2(a; \theta; \mathbf{z}_1, \mathbf{z}_2)$. By Minkowski's and Jensen's inequalities,

$$\begin{aligned}\|h_2(a; \theta; \mathbf{Z}_1, \mathbf{Z}_2)\|_r &\\ &\leqslant \|h(a; \theta; \mathbf{Z}_1, \mathbf{Z}_2\|_r + \|h_1(a; \theta; \mathbf{Z}_1)\|_r + \|h_1(a; \theta; \mathbf{Z}_2)\|_r + |h_0(a; \theta)| \\ &\leqslant 4\|h(a; \theta; \mathbf{Z}_1, \mathbf{Z}_2)\|_r.\end{aligned}$$

By definition,

$$h(a; \theta; \mathbf{z}_1, \mathbf{z}_2) = I(|e_1 - e_2 - (\mathbf{x}_1 - \mathbf{x}_2)' \beta| \geqslant as) - I(|e_1 - e_2| \geqslant as),$$

from which it follows that

$$|h(a; \theta; \mathbf{z}_1, \mathbf{z}_2)| \leqslant I(| \, |e_1 - e_2| - as| \leqslant |\mathbf{x}_1 - \mathbf{x}_2| \, |\beta|).$$

After taking expectations one obtains

$$E_{K \times K} |h(a; \theta; \mathbf{Z}_1, \mathbf{Z}_2|^r \leqslant 4\|f^*\|_\infty E_{G \times G} |\mathbf{X}_1 - \mathbf{X}_2| \, |\beta| \leqslant C(F, G) \, n^{-\tau_2},$$

where the constant $C(F, G)$ is independent of a and $\theta \in \bigcup_{\mathbf{i}} \Omega_{n, \mathbf{i}}$. Suppose now that r is a positive even integer. We may then apply Lemma A.3, with $g(\cdot, \cdot) = h_2(\theta_{\mathbf{i}}, \cdot, \cdot)$, $q = p + 1$, $\kappa_n = Cn^{-\tau_2}$ for some constant $C = C(r, F, G)$, and $M = 4$, to obtain (for n so large that $\kappa_n \geqslant n^{-1}$)

$$E|R_n^2(a; \theta_{\mathbf{i}})|^r \leqslant C(r, F, G) \, n^{-\tau} n^{-r\tau_2/4}. \tag{A.48}$$

Since the upper bound in (A.48) holds uniformly in a and $\theta_{\mathbf{i}} \in \Gamma_n$, we obtain from (A.47)

$$E \max_{\theta_{\mathbf{i}} \in \Gamma_n} |R_n^2(\theta_{\mathbf{i}})| \leqslant C(r, F, G) \, n^{((\tau - \tau_2) p + (\tau - \tau_1))/r} \rho(\infty) \, n^{-1 - \tau_2/4}. \tag{A.49}$$

By choosing r large enough in (A.49) we see that the RHS is $o(n^{-1})$. This completes the estimate of the first term in the RHS of (A.43). As for the second term, define first $\bar{R}_n^2(a; \theta_{\mathbf{i}})$ in the same way as $\bar{R}_n^2(\theta_{\mathbf{i}})$, but with $\rho(x)$ replaced by $\rho_a(x)$. Then clearly,

$$|\bar{R}_n^2(\theta_{\mathbf{i}})| \leqslant \int_0^\infty |\bar{R}_n^2(a; \theta_{\mathbf{i}}| \, d\rho(a). \tag{A.50}$$

Next, we define for $\theta_{\mathbf{i}} \in \Gamma_n$,

$$\bar{h}_2(a; \theta_{\mathbf{i}}; \mathbf{z}_1, \mathbf{z}_2) = \sup_{\theta \in \Omega_{n, \mathbf{i}}} |h_2(a; \theta; \mathbf{z}_1, \mathbf{z}_2) - h_2(a; \theta_{\mathbf{i}}; \mathbf{z}_1, \mathbf{z}_2)|,$$

so that

$$\bar{R}_n^2(a;\theta_i) \leqslant \binom{n}{2}^{-1} \sum_{i<j} \bar{h}_2(a;\theta_i;\mathbf{z}_i,\mathbf{z}_j). \tag{A.51}$$

Analogously, define $\bar{h}(a;\theta_i;\mathbf{z}_1,\mathbf{z}_2)$, $\bar{h}_1(a;\theta_i;\mathbf{z})$ and $\bar{h}_0(a;\theta_i)$. Then clearly,

$$\bar{h}_2(a;\theta_i;\mathbf{z}_1,\mathbf{z}_2) \leqslant \bar{h}(a;\theta_i;\mathbf{z}_1,\mathbf{z}_2) + \bar{h}_1(a;\theta_i;\mathbf{z}_1) + \bar{h}_1(a;\theta_i;\mathbf{z}_2) + \bar{h}_0(a;\theta_i). \tag{A.52}$$

In order to estimate $\bar{h}(a;\theta_i;\mathbf{z}_1,\mathbf{z}_2)$, put $\theta_i = (\beta_i, s_i)$, and observe that for any $\theta = (\beta, s) \in \Omega_{n,i}$,

$$\begin{aligned}
&|h(a;\theta;\mathbf{z}_1,\mathbf{z}_2) - h(a;\theta_i;\mathbf{z}_1,\mathbf{z}_2)| \\
&\quad \leqslant |I(|e_1 - e_2 - (\mathbf{x}_1-\mathbf{x}_2)'\beta| \geqslant as) - I(|e_1 - e_2 - (\mathbf{x}_1-\mathbf{x}_2)'\beta_i| \geqslant as_i)| \\
&\qquad + |I(|e_1-e_2| \geqslant as) - I(|e_1-e_2| \geqslant as_i)| \\
&\quad \leqslant I(||e_1 - e_2 - (\mathbf{x}_1-\mathbf{x}_2)'\beta_i| - as_i| \\
&\quad \leqslant (a + |\mathbf{x}_1-\mathbf{x}_2|_1)\,\delta_n) + I(||e_1-e_2| - as_i| \leqslant a\delta_n) \\
&\quad \triangleq M(a;\theta_i;\mathbf{z}_1,\mathbf{z}_2).
\end{aligned} \tag{A.53}$$

Hence, $M(a;\theta_i;\mathbf{z}_1,\mathbf{z}_2)$ is an upper bound for $\bar{h}(a;\theta_i;\mathbf{z}_1,\mathbf{z}_2)$. Actually, $M(a;\theta_i;\mathbf{z}_1,\mathbf{z}_2)$ may be interpreted as a kernel of a U-statistic, which leads us to the expansion

$$\begin{aligned}
M(a;\theta_i;\mathbf{z}_1,\mathbf{z}_2) = &-M_0(a;\theta_i) + M_1(a;\theta_i;\mathbf{z}_1) + M_1(a;\theta_i;\mathbf{z}_2) \\
&+ M_2(a;\theta_i;\mathbf{z}_1,\mathbf{z}_2),
\end{aligned} \tag{A.54}$$

where

$$M_1(a;\theta_i;\mathbf{z}) = E_K M(a;\theta_i;\mathbf{z},\mathbf{Z}) \leqslant 4\|f\|_\infty\,(2a + E_G|\mathbf{x}-\mathbf{X}|_1)\,\delta_n \tag{A.55}$$

and

$$M_0(a;\theta_i) = E_{K\times K} M(a;\theta_i;\mathbf{Z}_1,\mathbf{Z}_2) \leqslant 4\|f^*\|_\infty\,(2a + E_{G\times G}|\mathbf{X}_1-\mathbf{X}_2|_1)\,\delta_n.$$

It is not hard to see, by interchanging suprema and expectations, that

$$\bar{h}_1(a;\theta_i;\mathbf{z}_1) \leqslant M_1(a;\theta_i;\mathbf{z}_1) \tag{A.56}$$

and

$$\bar{h}_0(a;\theta_i) \leqslant M_0(a;\theta_i). \tag{A.57}$$

It now follows from (A.51)–(A.57) that we have

$$\begin{aligned}\bar{R}_n^2(a;\theta_i) &\leqslant \frac{4}{n}\sum_{j=1}^{n} M_1(a;\theta_i;\mathbf{z}_j) + \binom{n}{2}^{-1} \sum_{j<k} M_2(a;\theta_i;\mathbf{z}_j,\mathbf{z}_k)\\ &\leqslant C(F)\,\delta_n\left(c+\frac{1}{n}\sum_{j=1}^{n} E_G|\mathbf{x}_j-\mathbf{X}|\right)\\ &\quad+\binom{n}{2}^{-1}\sum_{j<k} M_2(a;\theta_i;\mathbf{z}_j,\mathbf{z}_k),\end{aligned} \tag{A.58}$$

where c is defined in (R). Denote for short, the last term in (A.58) by $M_{2n}(a;\theta_i)$. Then, because of (A.50),

$$\bar{R}_n^2(\theta_i) \leqslant C(F)\,\rho(\infty)\,\delta_n\left(c+\frac{1}{n}\sum_{j=1}^{n} E_G|\mathbf{x}_j-\mathbf{X}|\right)+\int_0^\infty |M_{2n}(a;\theta_i)|\,d\rho(a), \tag{A.59}$$

and hence, since the first term of the RHS in (A.59) is independent of θ_i, and $\delta_n = n^{-\tau} = o(n^{-1})$,

$$\max_{\theta_i\in\Gamma_n} \bar{R}_n^2(\theta_i) \leqslant o_p(n^{-1}) + \max_{\theta_i\in\Gamma_n} |M_{2n}(\theta_i)|, \tag{A.60}$$

where we have denoted the last term in (A.59) by $|M_{2n}(\theta_i)|$. In analogy with (A.46)–(A.47) we obtain

$$E\max_{\theta_i\in\Gamma_n} |M_{2n}(\theta_i)| \leqslant Cn^{((\tau-\tau_2)p+(\tau-\tau_1))/r}\rho(\infty)\sup_{a,\,\theta_i}\|M_{2n}(a;\theta_i)\|_r, \tag{A.61}$$

where the supremum ranges over $0<a\leqslant c$ and $\theta_i\in\Gamma_n$. We use Lemma A.3 in order to estimate $\|M_{2n}(a;\theta_i)\|_r$. In analogy with the estimates of $E|h_2(\theta_i;\mathbf{Z}_1,\mathbf{Z}_2)|^r$, one obtains from (A.53)

$$E|M_2(\theta_i;\mathbf{Z}_1,\mathbf{Z}_2|^s \leqslant C(c,r,F,G)\,\delta_n, \qquad s=1,\ldots,r, \tag{A.62}$$

uniformly for $0<a\leqslant c$, and $\theta_i\in\Gamma_n$. By letting r be a positive even integer, we obtain from (A.31), if n is so large that the RHS of (A.62) is exceeded by n^{-1},

$$\sup_{a,\,\theta_i}\|M_{2n}(a;\theta_i)\|_r \leqslant C(c,r,FD,G)\,n^{-5/4}. \tag{A.63}$$

It now follows from (A.60), (A.61), and (A.63), by choosing r large enough, that the second term of the RHS of (A.43) is also $o_p(n^{-1})$, and this completes the proof of the lemma. ∎

The claim of Lemma 3.1 now follows from (A.14), (A.20)–(A.22), and (A.40).

Appendix B. The proof of Lemmas 3.2, 3.3, and 3.4.

As with Lemma 3.1, we will start with a series of preliminary lemmas.

Lemma B.1. *Let*

$$m(\beta, s) = ED_n(\beta, s). \tag{B.1}$$

Then

$$m(t\beta, s) \geqslant m(\beta, s), \qquad \textit{for any} \quad t > 1, \tag{B.2}$$

and if $\varepsilon > 0$ *is chosen small enough, there exist numbers* $B_1, B_2 > 0$ *such that*

$$m(\beta, s) \geqslant m(0, s) - B_2(s-1) + B_1 |\beta|^2 = k - B_2(s-1) + B_1 |\beta|^2, \tag{B.3}$$

whenever $1 \leqslant s \leqslant 1 + \varepsilon$ *and* $|\beta| \leqslant \varepsilon$.

Proof. Let $D_n(a; \beta, s)$ and $m(a; \beta, s)$ be defined as $D_n(\beta, s)$ and $m(\beta, s)$, but with $\rho_a(x) = I(|x| \geqslant a)$ in place of $\rho(x)$. Then

$$D_n(\beta, s) = \int_0^\infty D_n(a; \beta, s)\, d\rho(a),$$

and after taking expectations,

$$m(\beta, s) = \int_0^\infty m(a; \beta, s)\, d\rho(a). \tag{B.4}$$

In view of (B.4), it suffices to prove (B.2) for each $m(a, \cdot, \cdot)$. First notice that

$$m(a; \beta, s) = E_{G \times G}(1 - F^*((\mathbf{X}_1 - \mathbf{X}_2)' \beta + as) + F^*((\mathbf{X}_1 - \mathbf{X}_2)' \beta - as)). \tag{B.5}$$

Clearly, f^* is symmetric. In addition, the unimodality of f implies that f^* is also unimodal (Dharmadhikari and Joag-Dev, 1988, Theorem 1.8). But this in turn implies that $1 - F^*(y + as) + F^*(y - as)$ is an increasing function of $|y|$, which proves (B.2) for $m(a; \cdot, \cdot)$, by conditioning on the value of $\mathbf{X}_1 - \mathbf{X}_2$. We now turn to (B.3). We differentiate $m(\beta, s)$ with respect to β to obtain

$$m_\beta(0, s) = \frac{\partial m(\beta, s)}{\partial \beta}\Big|_{\beta = 0} = 0 \tag{B.6}$$

and

$$m_{\beta\beta}(0, 1) = 2\Sigma E_F \bar{\psi}'(e).$$

The regularity conditions on G and F further imply (using dominated convergence) that $m_{\beta\beta}(\beta, s)$ is a continuous function of (β, s). Since $m_{\beta\beta}(0, 1)$ is positive definite, it follows from (B.6) that there exis numbers ε, $B_1 > 0$ such that

$$m(\beta, s) - m(0, s) \geqslant B_1 |\beta|^2, \tag{B.7}$$

whenever $|s-1|$ and $|\beta| \leqslant \varepsilon$. Since m is also continuously differentiable w.r.t. s, with

$$m_s(0, 1) = \frac{\partial m(\beta, s)}{\partial s}\Big|_{\theta=(0,1)} = -E_{F^*}(\psi(e)e) < 0,$$

we may choose ε, $B_2 > 0$ so that

$$m(0, s) \geqslant m(0,1) - B_2(s-1) \tag{B.8}$$

for all $1 \leqslant s \leqslant 1+\varepsilon$. Formula (B.3) now follws from (B.7) and (B.8). ∎

Now choose a sequence of s- and $|\beta|$-values

$$s_n = 1 + n^{-\tau'_1} \tag{B.9}$$

and

$$b_n = B_3 n^{-\tau'_2}, \tag{B.10}$$

where $\tau'_1 = 2\tau'_2$, $\tau_1 < \tau'_1 < \frac{1}{2}$, $\tau_2 < \tau'_2 < \frac{1}{4}$ (cf. (3.6)) and $B_3 = \sqrt{2B_2/B_1}$, with B_1 and B_2 the same constants as in (B.3). We then have:

LEMMA B.2. *Let $m(\beta, s)$ be given by* (B.1), *and s_n and b_n by* (B.9)–(B.10). *Then for large enough n,*

$$\inf_{\beta;\, |\beta| \geqslant b_n} m(\beta, s_n) \geqslant k + B_2 n^{-\tau'_1}, \tag{B.11}$$

with B_2 given by (B.8).

Proof. Suppose that $|\beta| \geqslant b_n$ and that n is so large that $|s_n - 1| \leqslant \varepsilon$ and $b_n \leqslant \varepsilon$, with ε as in (B.3). We then obtain from (B.2)–(B.3)

$$m(\beta, s_n) \geqslant m\left(\frac{b_n}{|\beta|}\beta, s_n\right) \geqslant k - B_2 n^{-\tau'_1} + B_1 B_3^2 n^{-2\tau'_2} = k + B_2 n^{-\tau'_1},$$

and the lemma is proved. ∎

LEMMA B.3. *Let $D_n(\beta, s)$, $m(\beta, s)$ and s_n be defined by* (3.1), (B.1) *and* (B.9) *respectively. Then*

$$\sup_{\beta \in \mathbb{R}^p} |D_n(\beta, s_n) - m(\beta, s_n)| = o_p(n^{-\tau_1'}). \tag{B.12}$$

Proof. Put $\tilde{h}(\beta; \mathbf{z}_1, \mathbf{z}_2) = \rho((e_1 - e_2 - (\mathbf{x}_1 - \mathbf{x}_2)' \beta)/s_n)$, $\tilde{h}_1(\beta; \mathbf{z}) = E_K \tilde{h}(\beta; \mathbf{z}, \mathbf{Z})$ and $\tilde{h}_2(\beta; \mathbf{z}_1, \mathbf{z}_2) = \tilde{h}(\beta; \mathbf{z}_1, \mathbf{z}_2) - \tilde{h}_1(\beta; \mathbf{z}_1) - \tilde{h}_1(\beta; \mathbf{z}_2) + E_K \tilde{h}_1(\beta; \mathbf{Z})$. Then note that

$$D_n(\beta, s_n) - m(\beta, s_n) = S_n(\beta) + T_n(\beta),$$

with

$$S_n(\beta) = \frac{2}{n} \sum_{i=1}^{n} (\tilde{h}_1(\beta; \mathbf{z}_i) - E_K \tilde{h}_1(\beta; \mathbf{Z}))$$

and

$$T_n(\beta) = \binom{n}{2}^{-1} \sum_{i<j} \tilde{h}_2(\beta; \mathbf{z}_i, \mathbf{z}_j).$$

It suffices to show that

$$\|S_n\| = o_p(n^{-\tau_1'}) \tag{B.13}$$

and

$$\|T_n\| = o_p(n^{-\tau_1'}), \tag{B.14}$$

with $\|\cdot\|$ denoting the supremum w.r.t. β. By construction, $\{T_n(\beta)\}_{\beta \in \mathbb{R}^p}$ is a family of degenerate U-statistics, and the kernels $\{\tilde{h}_2(\beta; \cdot, \cdot)\}$ have envelope $4\rho(\infty)$. (That is, $4\rho(\infty)$ uniformly majorizes all $\tilde{h}_2(\beta; \cdot, \cdot)$.) Formula (B.13) follows applying Theorem 9 of Nolan and Pollard (1987) to the kernels $\tilde{h}_2(\beta; \cdot, \cdot)/(4\rho(\infty))$ with the function $W(n, x)$ in the theorem equal to $(\log n)^{-2}$. The fact that $\{\tilde{h}_2(\beta; \cdot, \cdot)\}$ is a Euclidean class of kernels follows from Lemmas 16, 20 and 22 in Nolan and Pollard (1987), since ρ is of bounded variation.

It remains to prove (B.14). First one shows that $(\tilde{h}_1(\beta; \cdot)\}$ is a Euclidean class of functions in the same way as for $\tilde{h}_2$. Then apply Theorem 37 in Pollard (1984) applied to the kernels $\{\tilde{h}_1(\beta; \cdot,)\}/(2\rho(\infty))$, with $\alpha_n = n^{-1/2} \log n$ and $\delta_n = 1$. ∎

We are now ready to prove Lemma 3.2:

Proof of Lemma 3.2. It follows from Lemma B.2 and B.3 that

$$P(\inf_{\beta;\, |\beta| \geqslant b_n} |D_n(\beta, s_n)| > k) \to 1 \qquad \text{as} \quad n \to \infty,$$

with b_n as defined by (B.10). According to the definition of $\hat{s}_n(\beta)$ in (2.1) and the right continuity of ρ on $\mathbb{R}^+$ we have $D_n(\beta, \hat{s}_n(\beta)) \leqslant k$ for all β. Hence,

$$P(\inf_{\beta:\, |\beta| \geqslant b_n} \hat{s}_n(\beta) > s_n) \to 1. \tag{B.15}$$

By Lemma 3.5, $\hat{s}_n(0)$ is $n^{1/2}$-consistent, and hence by the definition of s_n,

$$P(\hat{s}_n(0) < \inf_{\beta:\, |\beta| \geqslant b_n} \hat{s}_n(\beta)) \to 1 \Rightarrow P(|\tilde{\beta}_n| < b_n) \to 1,$$

which proves the lemma, since $b_n = o(n^{-\tau_2})$. (Remember that $\tau_2' > \tau_2$.) ∎

Proof of Lemma 3.3. We note that

$$\hat{s}_n \leqslant \hat{s}_n(0) = 1 + O_p(n^{-1/2})$$

by (3.1) and Lemma 3.5, so it suffices to show that

$$P(\hat{s}_n \geqslant \tilde{s}_n) \to 1 \qquad \text{as} \quad n \to \infty,$$

where $\tilde{s}_n = 1 - n^{-\tau_1'}$. The proof of Lemma B.3 carries over to show that

$$\sup_{\beta \in \mathbb{R}^p} |D_n(\beta, \tilde{s}_n) - m(\beta, \tilde{s}_n)| = o_p(n^{-\tau_1'}). \tag{B.16}$$

Moreover, in analogy with (B.8) one shows that there exist $\varepsilon, B > 0$ such that

$$m(0, s) > k + B(1 - s) \qquad \text{for} \quad 1 - \varepsilon \leqslant s \leqslant 1. \tag{B.17}$$

Hence, if n is so large that $\tilde{s}_n \geqslant 1 - \varepsilon$, it follows from (B.2) and (B.17) that

$$\inf_{\beta \in \mathbb{R}^p} m(\beta, \tilde{s}_n) = m(0, \tilde{s}_n) > k + B(1 - \tilde{s}_n) = k + Bn^{-\tau_1'}. \tag{B.18}$$

In conjunction with (B.16), (B.18) implies that

$$P(\inf_{\beta \in \mathbb{R}^p} |D_n(\beta, \tilde{s}_n)| > k) \to 1,$$

which in turn implies

$$P(\hat{s}_n = \inf_{\beta} \hat{s}_n(\beta) > \tilde{s}_n) \to 1,$$

and this proves the lemma. ∎

Proof of Lemma 3.4. It follows from (F) and (R) that $m(0, s)$ is a twice continuously differentiable function of s. Since $m_s(0, 1) = -E_{F^*}(\psi(e)e)$ it follows from a Taylor expansion that

$$\sup_{|s-1| \leqslant n^{-\tau_1}} |m(0, s) - m(0, 1) + (s-1)\, E_{F^*}(\bar{\psi}(e)e)| = O(n^{-2\tau_1}) = o(n^{-1/2}).$$

Putting $\Xi_n(s) = D_n(0, s) - D_n(0, 1)$, it therefore remains to show that

$$\|\Xi(\cdot) - E\Xi(\cdot)\| \triangleq \sup_{|s-1| \leqslant n^{-\tau_1}} |\Xi_n(s) - E\Xi_n(s)| = o_p(n^{-1/2}). \tag{B.19}$$

We use a Hoeffding representation of $\Xi_n(s)$,

$$\Xi_n(s) = \binom{n}{2}^{-1} \sum_{i<j} \left(\rho\left(\frac{e_i - e_j}{s}\right) - \rho(e_i - e_j)\right) = \frac{1}{n!} \sum_{\pi} W_{n,\pi}(s), \tag{B.20}$$

where $\pi = (\pi(1), ..., \pi(n))$ is a permutation of $(1, ..., n)$ and

$$W_{n,\pi}(s) = \frac{1}{[n/2]} \sum_{i=1}^{[n/2]} \left(\rho\left(\frac{e_{\pi(2i-1)} - e_{\pi(2i)}}{s}\right) - \rho(e_{\pi(2i-1)} - e_{\pi(2i)})\right)$$

is an average of i.i.d. random variables. In particular, with π_0 the identity permutation and $\xi_i = e_{2i-1} - e_{2i}$, $i = 1, ..., [n/2]$, we have

$$W_{n,\pi_0}(s) = \frac{1}{[n/2]} \sum_{i=1}^{[n/2]} \left(\rho\left(\frac{\xi_i}{s}\right) - \rho(\xi_i)\right). \tag{B.21}$$

Define $W^1_{n,\pi_0}(a; s)$ by replacing $\rho(x)$ by $I(x \geqslant a)$ in (B.21), and similarly $W^2_{n,\pi_0}(a; s)$ by putting $I(x \leqslant -a)$ instead of $\rho(x)$. Then clearly,

$$W_{n,\pi_0}(s) = \int_0^{\infty} (W^1_{n,\pi_0}(a; s) + W^2_{n,\pi_0}(a; s))\, d\rho(a). \tag{B.22}$$

From (B.20), (B.22), and Minkowski's inequality we obtain

$$\begin{aligned} E\|\Xi(\cdot) - E\Xi(\cdot)\| &\leqslant E\| W_{n,\pi_0}(\cdot) - EW_{n,\pi_0}(\cdot)\| \\ &\leqslant 2\rho(\infty)\, E(\sup_{a,s,j} |W^j_{n,\pi_0}(a; s) - EW^j_{n,\pi_0}(a; s)|), \end{aligned} \tag{B.23}$$

where in (B.23), $0 < a \leqslant c$ (cf. (R)), $|s-1| \leqslant n^{-\tau_1}$ and $j = 1$ or 2. With $m = [n/2]$ and

$$F^*_m(x) = \frac{1}{m} \sum_{i=1}^{m} I(\xi_i \leqslant x)$$

the empirical distribution formed by $\xi_1, ..., \xi_m$, one observes that

$$W^1_{n, \pi_0}(a; s) = F^*_m(a) - F^*_m(as)$$

and

$$W^2_{n, \pi_0}(a; s) = F^*_m(-as) - F^*_m(-a).$$

Hence

$$\sup_{a, s, j} |W^j_{n, \pi_0}(a; s) - EW^j_{n, \pi_0}(a; s)| \leqslant \sup_{I \in \mathscr{F}} |F^*_m\{I\} - F_m\{I\}|, \qquad (B.24)$$

where $\mathscr{F}$ denotes the set of all intervals of length $\leqslant cn^{-\tau_1}$. In order to estimate the RHS of (B.24) we need a result concerning the oscillation of the empirical distribution function. From Reiss (1988, formula (6.3.2)) it follows that

$$E(\sup_{I \in \mathscr{F}} |F^*_m\{I\} - F_m\{I\}|) \leqslant Cn^{-((1+\tau_1)/2)} \log n, \qquad (B.25)$$

for some constant C (which depends on F^*). The lemma now follows from (B.19) and (B.23)-(B.25).

Acknowledgments

We thank the referees for valuable comments, especially for bringing to our attention the paper by Nolan and Pollard (1987) which made it possible to shorten the proof of Lemma B.3 substantially.

References

Billingsley, P. (1968). *Convergence of Probability Measures.* Wiley, New York.

Croux, C., Rousseeuw, P. J., and Hössjer, O. (1994). Generalized S-estimators. *J. Amer. Statist. Assoc.*, to appear.

Davies, L. (1990). The asymptotics of S-estimators in the linear regression model. *Ann. Statist.* **18** 1651–1675.

Dharmadhikari, S., and Joag-Dev, K. (1988). *Unimodality, Convexity, and Applications.* Academic Press, New York.

Hampel, F. R., Ronchetti, E. M., Rousseeuw, P. J., and Stahel, W. A. (1986). *Robust Statistics, the Approach Based on Influence Functions.* Wiley, New York.

Hössjer, O. (1992). On the optimality of S-estimators. *Statist. Prob. Lett.* **14** 413–419.

Jaeckel, L. A. (1972). Estimating regression coefficients by minimizing the dispersion of the residuals. *Ann. Math. Statist.* **43** 1449–1458.

Nolan, D., and Pollard, D. (1987). U-processes: Rates of convergence. *Ann. Statist.* **15** 780–799.

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

REISS, R.-D. (1988). Approximate Distributions of Order Statistics, with Applications to Nonparametric Statistics. Springer-Verlag, New York.

ROUSSEEUW, P. J. (1984). Least median of squares regression. *J. Amer. Statist. Assoc.* **79** 871–880.

ROUSSEEUW, P. J. AND CROUX, C. (1992). Explicit scale estimators with high breakdown point. In *L_1-Statistical Analysis and Related Methods*, (Y. Dodge, Ed.), pp. 77–92. North-Holland, Amsterdam.

ROUSSEEUW, P. J., AND CROUX, C. (1993). Alternatives to the median absolute deviation. *J. Amer. Statist. Assoc.* **88** 1273–1283.

ROUSSEEUW, P. AND YOHAI, V. (1984). Robust regression by means of S-estimators. In *Robust and Nonlinear Time Series Analysis*, (J. Franke, W. Härdle, and R. D. Martin, Eds.), pp. 256–272, Lecture Notes in Statistics, Vol. 26, Springer-Verlag, New York.

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

SERFLING, R. J. (1984). Generalized L-, M- and R-statistics. *Ann. Statist.* **12** 76–86.

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

YOHAI, V. Y. (1987). High breakdown-point and high efficiency robust estimates for regression, *Ann. Statist.* **15** 642–656.

YOHAI, V. Y. AND ZAMAR, R. H. (1988). High breakdown-point estimates of regression by means of the minimization of an efficient scale, *J. Amer. Statist. Assoc.* **83** 406–413.