The Annals of Statistics
1994, Vol. 22, No. 3, 1478–1501

ASYMPTOTICS OF THE REPEATED MEDIAN SLOPE ESTIMATOR

By Ola Hössjer,[1] Peter J. Rousseeuw and Christophe Croux

Lund Institute of Technology and University of Antwerp

The influence function is determined for (twice) repeated median estimators with arbitrary kernel functions, and more generally in the case where the two medians are replaced by a general class of estimators. Asymptotic normality is then established for the repeated median estimator of the slope parameter in simple linear regression. In this case the influence function is bounded. For bivariate Gaussian data the efficiency becomes $4/\pi^2 \approx 40.5\%$, which is the square of the efficiency of the univariate median. The asymptotic results are compared with finite-sample efficiencies. It turns out that the convergence to the asymptotic behavior is extremely slow.

1. Introduction. Consider the simple linear regression model

$$y_i = \alpha + \beta x_i + e_i, \qquad i = 1, \dots, n, \tag{1.1}$$

where $\mathbf{z}_i = (x_i, y_i)$ is the observed vector and e_i represents noise. We assume that the random vectors (x_i, e_i) are i.i.d., and that x_i and e_i are mutually independent with distributions G and F, respectively. Many estimates of the slope parameter β are based on the pairwise slopes $h(\mathbf{z}_i, \mathbf{z}_j) = (y_j - y_i)/(x_j - x_i)$ when $x_i \neq x_j$, and $h(\mathbf{z}_i, \mathbf{z}_j) = 0$ when $x_i = x_j$. For instance, the least squares estimator $\widehat{\beta}_{\mathrm{LS}}$ may be written as a weighted average,

$$\widehat{\beta}_{\mathrm{LS}} = \frac{\sum_{i<j} w_{ij} h(\mathbf{z}_i, \mathbf{z}_j)}{\sum_{i<j} w_{ij}}, \tag{1.2}$$

with weights $w_{ij} = (x_i - x_j)^2$. In a data set with $n = 5$ observations, Boscovich (1757) computed the unweighted average of the 10 pairwise slopes, as well as a 10% trimmed mean given by the average of 8 of these slopes [for a more complete historical discussion see Stigler (1986)]. The estimator of Theil (1950) and Sen (1968) is the median of all pairwise slopes. Frees (1991) gives a survey of these and related estimators.

Another estimator, the repeated median,

$$\widehat{\beta}_n = \operatorname*{med}_{i} \operatorname*{med}_{j,\, j \neq i} h(\mathbf{z}_i, \mathbf{z}_j), \tag{1.3}$$

was proposed by Siegel (1982). He showed that when all x_i are distinct (an event with probability 1 if G is continuous), $\widehat{\beta}_n$ has a finite-sample breakdown point

Received March 1992; revised March 1993.

[1]Work supported by the Swedish Board for Technical Development under contract 88-01359P.

AMS 1991 *subject classifications.* 62F35, 62J05.

Key words and phrases. Asymptotic normality, influence function, repeated median, simple linear regression.

$\varepsilon_n^* = [n/2]/n$, that is, if fewer than $[n/2]$ vectors $\mathbf{z}_i$ are changed, the estimate remains bounded. This is the maximal possible value of ε_n^* for any regression equivariant estimator [Rousseeuw (1984)] and it yields an asymptotic breakdown point of 0.5. [A regression equivariant estimator is one which satisfies

$$\widehat{\beta}_n\Big(\big\{(x_i, y_i + c + dx_i)\big\}\Big) = \widehat{\beta}_n\Big(\big\{(x_i, y_i)\big\}\Big) + d,$$

for any c and d.] Siegel also showed that $\widehat{\beta}_n$ is a Fisher consistent estimate of β.

The purpose of this paper is to derive the influence function (Section 2) and to prove asymptotic normality (Section 3) of the repeated median slope, given some regularity conditions on F and G. These findings are compared with Monte Carlo variances in Section 4. In Section 5, we discuss some possible extensions.

The influence function is actually determined quite generally, for an arbitrary kernel function $h(\mathbf{z}_1, \mathbf{z}_2)$, and with the two medians in (1.3) replaced by arbitrary estimators T_1 and T_2. However, a strict proof of asymptotic normality is given only for the repeated median, and the kernel function corresponding to the pairwise slope. With $\widehat{\beta}_n$ indicating the estimate for sample size n, our main result (Theorem 3.1) is that

$$\sqrt{n}(\widehat{\beta}_n - \beta) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} \mathrm{IF}\,(\mathbf{z}_i) + o_p(1) \to_d N\big(0, \sigma^2\big) \quad \text{as } n \to \infty, \tag{1.4}$$

where the influence function is given by

$$\mathrm{IF}\,(x,y) = \frac{\mathrm{sgn}\Big(\big[y - \alpha - F^{-1}(0.5) - \beta G^{-1}(0.5)\big]\big/\big[x - G^{-1}(0.5)\big] - \beta\Big)}{2f\big(F^{-1}(0.5)\big)E_G\Big(\big|X - G^{-1}(0.5)\big|\Big)} \tag{1.5}$$

and

$$\begin{aligned}\sigma^2 &= \int \mathrm{IF}\,(x, \alpha + \beta x + e)^2\, dK(x,e)\\ &= \frac{1}{4f\big(F^{-1}(0.5)\big)^2\Big(E_G\big|X - G^{-1}(0.5)\big|\Big)^2},\end{aligned} \tag{1.6}$$

with $K = G \times F$. [Formula (1.4) is linked to Hampel's (1974) definition of the influence function by means of a von Mises expansion.] We see from (1.5) that the influence function is bounded, giving another illustration of the robustness of the repeated median. Actually, it follows from (1.4) and the Bahadur approximation of sample medians by a sum of i.i.d. variables that

$$\widehat{\beta}_n - \widehat{\beta}_{\mathrm{MED}} = o_p\big(n^{-1/2}\big), \tag{1.7}$$

where

$$\widehat{\beta}_{\mathrm{MED}} = \operatorname*{med}_i \left\{\frac{y_i - \alpha - F^{-1}(0.5) - \beta G^{-1}(0.5)}{x_i - G^{-1}(0.5)}\right\}, \tag{1.8}$$

which in general is not computable, since α and β are unknown. The influence function for $\widehat{\beta}_{\mathrm{MED}}$ is also given by (1.5). In the special case of simple linear regression through the origin ($\alpha = 0$), when $F \sim N(0, V)$ for some $V > 0$ and G is symmetric, $\widehat{\beta}_{\mathrm{MED}} = \mathrm{med}(y_i/x_i)$, and this estimator has minimal gross-error sensitivity

$$\gamma^* = \sup_{\mathbf{z}} |\mathrm{IF}(\mathbf{z})| \tag{1.9}$$

within a large class of estimators including all GM-estimators [cf. Ronchetti and Rousseeuw (1985), Hampel, Ronchetti, Rousseeuw and Stahel (1986), Section 6.3, and He and Simpson (1993)].

For bivariate Gaussian data, the asymptotic efficiency of $\widehat{\beta}_n$ becomes $4/\pi^2 \approx 40.5\%$. However, the finite-sample efficiencies vary between 53% and 62% for sample sizes between 20 and 40,000 (see Section 4). The Theil–Sen estimator (obtained by taking the median of all pairwise slopes) has a much higher efficiency of 91.5%, but a lower breakdown point of $1 - 1/\sqrt{2} \approx 29\%$ and a higher gross-error sensitivity. The L_1-estimator also has a higher asymptotic efficiency (in fact, $2/\pi \approx 63.7\%$) at bivariate Gaussian data, but an unbounded influence function and a 0% breakdown point.

Our results are restricted to simple linear regression. Repeated medians can also be used for estimating the slope parameters in multiple linear regression, using kernel functions with more than two arguments [Siegel (1982)]. However, these estimators are not affine equivariant when the number of slope parameters is two or more, that is, they do not transform properly under affine transformations of the carriers. The asymptotic properties of the repeated median estimator in higher dimensions form an interesting area for future research.

2. Influence functions. In this section we give a heuristic derivation of the influence function, in order to motivate the results of the next section, even though the setup is more general here.

Given a Euclidean space $\mathcal{X}$, define the kernel function $h: \mathcal{X} \times \mathcal{X} \to \mathbb{R}$. Assume also that $\mathbf{z}_1, \dots, \mathbf{z}_n$ are i.i.d. observations from $\mathcal{X}$ with common distribution K. Let T_1 and T_2 be two estimators that may be written as functionals of the empirical distribution. For each $\mathbf{z}$, put $H(\mathbf{z}) = T_1(L_{\mathbf{z}})$, where $L_{\mathbf{z}} = \mathcal{L}_K(h(\mathbf{z}, \mathbf{Z}))$ and let $\theta = T_2(L)$, where $L = \mathcal{L}_K(H(\mathbf{Z}))$, be the functional that we want to estimate. In order to estimate θ we first estimate $H(\mathbf{z}_i)$ by $\widehat{H}(\mathbf{z}_i) = T_1(L_{\mathbf{z}_i, n-1})$, where $L_{\mathbf{z}_i, n-1}$ is the empirical distribution formed by $\{h(\mathbf{z}_i, \mathbf{z}_j);\ j \neq i,\ i \text{ fixed}\}$. Then set

$$\widehat{\theta}_n = T_2(L_n), \tag{2.1}$$

where L_n is the empirical distribution formed by $\widehat{H}(\mathbf{z}_1), \dots, \widehat{H}(\mathbf{z}_n)$. Note that $\widehat{\theta}_n$ reduces to a U-statistic if both T_1 and T_2 are sample means, and to a repeated median estimator if both T_1 and T_2 are sample medians.

Assume next that T_1 is differentiable at $L_{\mathbf{z}}$ for all $\mathbf{z}$ and that T_2 is differentiable at L, and introduce the influence functions $\mathrm{IF}_1(\mathbf{z}_1, \mathbf{z}_2) = \mathrm{IF}(h(\mathbf{z}_1, \mathbf{z}_2),$

$T_1, L_{\mathbf{z}_1})$ and $\mathrm{IF}_2(x) = \mathrm{IF}(x, T_2, L)$. Provided that IF_2 is differentiable, the estimate $\widehat{\theta}_n$ may be expanded as

$$
\begin{aligned}
\widehat{\theta}_n - \theta &= \frac{1}{n}\sum_{i=1}^{n} \mathrm{IF}_2\big(\widehat{H}(\mathbf{z}_i)\big) + R \\
&= \frac{1}{n}\sum_{i=1}^{n} \mathrm{IF}_2\left(H(\mathbf{z}_i) + \frac{1}{n-1}\sum_{j,\, j \neq i} \mathrm{IF}_1(\mathbf{z}_i, \mathbf{z}_j) + R_i\right) + R \qquad (2.2)\\
&= \frac{1}{n}\sum_{i=1}^{n} \mathrm{IF}_2\big(H(\mathbf{z}_i)\big) + \frac{1}{n(n-1)}\sum_{i,\, j,\, i \neq j} \widetilde{h}(\mathbf{z}_i, \mathbf{z}_j) + \widetilde{R},
\end{aligned}
$$

where $\widetilde{h}(\mathbf{z}_1, \mathbf{z}_2) = \mathrm{IF}_2'(H(\mathbf{z}_1))\mathrm{IF}_1(\mathbf{z}_1, \mathbf{z}_2)$. (If T_1 and T_2 are both sample means, we have $\widetilde{R} \equiv 0$.) When the remainder term $\widetilde{R}$ is $o_p(n^{-1/2})$ (which has to be determined for each case separately) and the kernel of the U-statistic in (2.2) is square integrable, that is, if $E_{K \times K}\widetilde{h}(\mathbf{Z}_1, \mathbf{Z}_2)^2 < \infty$, we may use the method of projection of a U-statistic [cf. Serfling (1980), Section 5.3] to obtain

$$
\widehat{\theta}_n - \theta = \frac{1}{n}\sum_{i=1}^{n} \mathrm{IF}(\mathbf{z}_i) + o_p\big(n^{-1/2}\big), \qquad (2.3)
$$

where

$$
\mathrm{IF}(\mathbf{z}) = \mathrm{IF}_2\big(H(\mathbf{z})\big) + E_K\Big[\mathrm{IF}_2'\big(H(\mathbf{Z})\big)\mathrm{IF}_1(\mathbf{Z}, \mathbf{z})\Big]. \qquad (2.4)
$$

The central limit theorem then yields

$$
\sqrt{n}(\widehat{\theta}_n - \theta) \to_d N\big(0, \sigma^2\big), \qquad (2.5)
$$

where

$$
\sigma^2 = E_K \mathrm{IF}(\mathbf{Z})^2. \qquad (2.6)
$$

Suppose now that both T_1 and T_2 are medians, so that $\widehat{\theta}_n$ corresponds to a repeated median. For uniqueness, we define the median as the right-continuous inverse of the corresponding distribution function throughout the paper, so that

$$
H(\mathbf{z}) = L_{\mathbf{z}}^{-1}(0.5) = \inf\big\{x;\, L_{\mathbf{z}}(x) > 0.5\big\}, \qquad (2.7)
$$

and

$$
\theta = L^{-1}(0.5) = \inf\big\{x;\, L(x) > 0.5\big\}. \qquad (2.8)
$$

Similarly, sample medians are defined as the right-continuous inverse of the empirical distribution formed by the sample, that is, the observation with rank

$[n/2] + 1$. The influence functions are given by

$$\mathrm{IF}_1(\mathbf{z}_1, \mathbf{z}_2) = \frac{\operatorname{sgn}\big(h(\mathbf{z}_1, \mathbf{z}_2) - H(\mathbf{z}_1)\big)}{2l_{\mathbf{z}_1}\big(H(\mathbf{z}_1)\big)}$$

and

$$\mathrm{IF}_2(x) = \frac{\operatorname{sgn}(x - \theta)}{2l(\theta)},$$

where $l_{\mathbf{z}} = L'_{\mathbf{z}}$ and $l = L'$. Since IF'_2 is difficult to interpret directly in (2.4), we rather replace T_2 by an M-estimator T_2^{ε}, based on a score function

$$\psi_\varepsilon(x) = \begin{cases} \operatorname{sgn}(x), & |x| > \varepsilon, \\ x/\varepsilon, & |x| \le \varepsilon, \end{cases} \tag{2.9}$$

and then we let $\varepsilon \to 0+$. Setting $\theta_\varepsilon = T_2^\varepsilon(L)$, formula (2.4) for the influence function becomes

$$\mathrm{IF}^\varepsilon(\mathbf{z}) = \frac{\varepsilon\psi_\varepsilon\big(H(\mathbf{z}) - \theta_\varepsilon\big)}{L\big\{[\theta_\varepsilon - \varepsilon, \theta_\varepsilon + \varepsilon]\big\}} + E_{K_\varepsilon}\frac{\operatorname{sgn}\big(h(\mathbf{Z}, \mathbf{z}) - H(\mathbf{Z})\big)}{2l_{\mathbf{Z}}\big(H(\mathbf{Z})\big)}, \tag{2.10}$$

with K_ε the conditional distribution of $\mathbf{Z} \sim K$, given that $H(\mathbf{Z}) \in [\theta_\varepsilon - \varepsilon, \theta_\varepsilon + \varepsilon]$. Of course, it has to be shown for each separate case that the remainder term $\widetilde{R}$ in (2.2) is negligible and that $\widetilde{h}$ is square integrable. Let us give some simple conditions for this to hold (these conditions can be weakened at the cost of more technical arguments). Assume that $L\{[\theta_\varepsilon - \varepsilon, \theta_\varepsilon + \varepsilon]\} > 0$ and that $l_{\mathbf{z}}(H(\mathbf{z}))$ is lower bounded away from zero on the support of K_ε. Then $\mathrm{IF}'_2(\cdot)$ and $\mathrm{IF}_1(\cdot, \cdot)$ are bounded on $\mathbb{R}$ and $\operatorname{supp}(K_\varepsilon) \times \mathbb{R}$, respectively. This implies that $\widetilde{h}(\cdot, \cdot)$ is bounded and, in particular, square integrable. In order to handle $\widetilde{R}$, set

$$S_i = \frac{1}{n-1}\sum_{j, j \ne i} \mathrm{IF}_1(\mathbf{z}_i, \mathbf{z}_j),$$

assume that for some $\frac{1}{4} < \alpha < \frac{1}{2}$ it holds that

$$\max_i |S_i| = o_p(n^{-\alpha}), \tag{2.11}$$

$$\max_i |R_i| = o_p\big(n^{-1/2}\big), \tag{2.12}$$

that $R = o_p(n^{-1/2})$ and, finally, that L has a bounded density in neighborhoods of $-\varepsilon$ and ε. Then the first-order Taylor approximation in (2.2) holds whenever $|S_i + R_i| < n^{-\alpha}$ and $\big||H(\mathbf{z}_i)| - \varepsilon\big| \ge n^{-\alpha}$. Therefore, with probability tending to 1 as $n \to \infty$,

$$\begin{aligned} \big|\widetilde{R}\big| &\le |R| + \frac{1}{n}\left|\sum_{i=1}^n \mathrm{IF}'_2(H(\mathbf{z}_i))R_i\right| + \frac{\|\mathrm{IF}'_2\|_\infty}{n}\sum_{i=1}^n I\Big(\big||H(\mathbf{z}_i)| - \varepsilon\big| < n^{-\alpha}\Big)|S_i + R_i| \\ &= |R| + o_p\big(n^{-1/2}\big) + O_p(n^{-\alpha})n^{-\alpha} = o_p\big(n^{-1/2}\big). \end{aligned}$$

Of all the conditions given above, the imposed positive lower bound on $l_{\mathbf{z}}(H(\mathbf{z}))$ is the most restrictive.

If now $\varepsilon \to 0+$ implies that $\theta_\varepsilon \to \theta, L\{[\theta_\varepsilon - \varepsilon, \theta_\varepsilon + \varepsilon]\}/\varepsilon \to 2l(\theta)$ and $K_\varepsilon \to_d K_0$, for some distribution K_0, and if the appropriate uniform integrability conditions are satisfied for the second term in (2.10) as $\varepsilon \to 0+$, it follows that

$$\text{(2.13)} \qquad \mathrm{IF}^{\varepsilon}(\mathbf{z}) \to \mathrm{IF}(\mathbf{z}) = \frac{\operatorname{sgn}\big(H(\mathbf{z}) - \theta\big)}{2l(\theta)} + E_{K_0} \frac{\operatorname{sgn}\big(h(\mathbf{Z}, \mathbf{z}) - H(\mathbf{Z})\big)}{2l_{\mathbf{Z}}\big(H(\mathbf{Z})\big)}.$$

To be more precise, the following two conditions justify the limit in (2.13): Let $\mathbf{Z}_\varepsilon$ be a random variable with distribution K_ε. Then, suppose that

$$\text{(2.14)} \qquad \big\{\mathrm{IF}_1(\mathbf{Z}_\varepsilon, \mathbf{z})\big\}_{0 \le \varepsilon \le \varepsilon_0} \quad \text{is uniformly integrable,}$$

for some $\varepsilon_0 > 0$, and that

$$\text{(2.15)} \qquad P(\mathbf{Z}_0 \in C_{\mathbf{z}}) = 1,$$

where $C_{\mathbf{z}} = \{\mathbf{z}'; \mathrm{IF}_1(\cdot, \mathbf{z}) \text{ is continuous at } \mathbf{z}'\}$ [Billingsley (1968), Theorems 5.1 and 5.4].

Let us now specialize further to estimation of the slope parameter β in (1.1), that is, $\mathcal{X} = \mathbb{R}^2$, $\mathbf{z} = (x, y)$ and $h(\mathbf{z}_1, \mathbf{z}_2) = (y_2 - y_1)/(x_2 - x_1)$, as in Section 1. It follows from Theorem A.1(i) that the slope β is actually given by (2.8), that is,

$$\text{(2.16)} \qquad \beta = \theta = \operatorname*{med}_{Z_1 \sim K} \operatorname*{med}_{Z_2 \sim K} \frac{Y_2 - Y_1}{X_2 - X_1}.$$

Because of regression equivariance, we assume w.l.o.g. in the rest of the paper that $\alpha = \beta = F^{-1}(0.5) = G^{-1}(0.5) = 0$. Then, under the regularity conditions (F) and (G) in Section 3, $\operatorname{sgn}(H(\mathbf{z})) = \operatorname{sgn}(xy)$ [Theorem A.1(i)] and $l(0) = \infty$ (Theorem B.1), which implies that the first term in (2.13) vanishes. It may also be seen from the results in Appendix B that K_0 equals the Dirac measure at $(0, 0)$. The reason for this is that the set $\{\mathbf{z}; |H(\mathbf{z})| \le \varepsilon\}$ looks roughly like [cf. (B.4)]

$$\Big\{\mathbf{z}; 2\big|\big(G(x) - 0.5\big)\big(F(y) - 0.5\big)\big| \le f(y) E_G|X - x|\varepsilon\Big\},$$

and in particular, around the origin, like

$$\big\{\mathbf{z}; 2g(0)|xy| \le E_G|X|\varepsilon\big\}.$$

This implies that, given any $d > 0$ and $\Omega_d = [-d, d] \times [-d, d]$,

$$\text{(2.17)} \qquad P\Big(\Omega_d \cap \big\{\mathbf{z};\ |H(\mathbf{z})| \le \varepsilon\big\}\Big) \asymp \varepsilon \log\left(\frac{1}{\varepsilon}\right), \quad \text{as } \varepsilon \to 0+,$$

while

$$\text{(2.18)} \qquad P\Big(\Omega_d^c \cap \big\{\mathbf{z};\ |H(\mathbf{z})| \le \varepsilon\big\}\Big) = O(\varepsilon), \quad \text{as } \varepsilon \to 0+.$$

If now either the error distribution F or the carrier distribution G is symmetric, it is not hard to see that L is also a symmetric distribution, and therefore $\theta_\varepsilon = 0$. Hence, in this special case K_ε is the conditional distribution of $\mathbf{Z} \sim K$ on the set $\{\mathbf{z};\ |H(\mathbf{z})| \le \varepsilon\}$, and so by (2.17)–(2.18) it converges weakly to δ_0 as $\varepsilon \to 0+$. In the general case L need not be a symmetric distribution, but $L^{-1}(0.5) = 0$ and hence $|\theta_\varepsilon| \le \varepsilon$. This in turn implies that (2.17)–(2.18) remain valid when $|H(\mathbf{z})| \le \varepsilon$ is replaced by $H(\mathbf{z}) \in [\theta_\varepsilon - \varepsilon, \theta_\varepsilon + \varepsilon]$, so $K_0 = \delta_0$ even in the general case. In fact, it is quite surprising that K_0 is supported on a small subset of $\{\mathbf{z};\ |H(\mathbf{z})| = 0\} = \{\mathbf{z}; xy = 0\}$. Summarizing, the influence function in (2.13) becomes [cf. (1.5)]

$$\text{(2.19)} \qquad \mathrm{IF}(\mathbf{z}) = \frac{\operatorname{sgn}\big(h(\mathbf{0}, \mathbf{z})\big)}{2l_0(\mathbf{0})} = \frac{\operatorname{sgn}(xy)}{2f(0)E_G|X|},$$

where the last equality follows from (A.4). Actually, the fact that K_0 is a one-point distribution simplifies the expression for the influence function a lot. Observe that our reasoning to obtain (2.19) is so far based on just plugging in the slope kernel expressions for K_0, H, h and $l(\theta)$ into (2.13). Our argument could be made rigorous by checking which of the conditions imposed above are valid for the slope kernel. However, we will show by different methods in Section 3 that $\widehat{\beta}_n$ is asymptotically normal, with the influence function given by (2.19).

3. Asymptotic normality of the slope estimator. We assume the following regularity conditions:

(F) The error distribution F is absolutely continuous, $F^{-1}(0.5) = 0$, the density f is bounded ($\|f\|_\infty < \infty$), strictly positive and Lipschitz continuous of order η, that is, $\sup_{y_1 \ne y_2} |f(y_2) - f(y_1)|/|y_2 - y_1|^\eta = \|f\|_\eta < \infty$, where $\eta > 0$. [Actually, the facts that f is positive, Lipschitz continuous and integrates to 1 imply that $\lim_{|x| \to \infty} f(x) = 0$ and, in particular, that f is bounded. We will assume w.l.o.g. that $\eta < 0.5$ in the following, since this will simplify some formulas, for instance, in Lemma 3.3 and (A.10).]

(G) The distribution G of the carriers is continuous, $G^{-1}(0.5) = 0$, and G has a positive and continuous density g around 0 with $g(0) > 0$. Moreover, $E_G|X|^{1+\eta} < \infty$, where η is the same number as in (F).

The main result of the paper is:

THEOREM 3.1. *Suppose that* $\beta = 0$ *in* (1.1), *with the error and carrier distributions satisfying conditions* (F) *and* (G), *respectively*. *Then*

$$\text{(3.1)} \qquad \sqrt{n}\widehat{\beta}_n = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \mathrm{IF}(\mathbf{z}_i) + o_p(1) \to_d N\big(0, \sigma^2\big) \quad \text{as } n \to \infty,$$

where IF *is given by* (2.19) *and*

$$\text{(3.2)} \qquad \sigma = \frac{1}{2f(0)E_G|X|}.$$

The theorem is proved through a series of lemmas. In all of these lemmas, we will tacitly assume the same regularity conditions as in Theorem 3.1. First we introduce some notation. We fix $0 < \gamma < \frac{1}{4}$, set

$$\varepsilon_n = \frac{(\log n)^{1/2+\gamma}}{n^{1/2}}, \tag{3.3}$$

$$\delta_n = \frac{1}{(\log n)^{\gamma+1/\eta}} \tag{3.4}$$

and divide the plane into three regions according to

$$\begin{aligned} A_1 &= \{\mathbf{z};\ |H(\mathbf{z})| \le \varepsilon_n, |\mathbf{z}| \le \delta_n\},\\ A_2 &= \{\mathbf{z};\ |H(\mathbf{z})| \le \varepsilon_n, |\mathbf{z}| > \delta_n\},\\ A_3 &= \{\mathbf{z};\ 0.5 - \rho'\varepsilon_n < L_{\mathbf{z}}(\varepsilon_n) < 0.5,\ |y| > 1\}\\ &\qquad \cup \{\mathbf{z};\ 0.5 < L_{\mathbf{z}}(-\varepsilon_n) < 0.5 + \rho'\varepsilon_n,\ |y| > 1\},\\ A_4 &= \{\mathbf{z};\ |H(\mathbf{z})| > \varepsilon_n\} - A_3, \end{aligned} \tag{3.5}$$

where by $|\mathbf{z}|$ we mean (say) the L^∞-norm $\max(|x|, |y|)$, and ρ' is a positive constant whose value will be defined in the proof of Lemma 3.6. In order to analyze the asymptotic behaviour of $\widehat{\beta}_n$ [cf. (1.3) with $h(\cdot,\cdot)$ the pairwise slope kernel function], we introduce two other statistics. Let

$$\overline{\beta}_n = \operatorname*{med}_i \{H(\mathbf{z}_i) + \xi\} = \operatorname*{med}_i H(\mathbf{z}_i) + \xi, \tag{3.6}$$

where

$$\xi = \frac{1}{n-1} \sum_{i=1}^{n} \mathrm{IF}_1(0, \mathbf{z}_i), \tag{3.7}$$

and

$$\widetilde{\beta}_n = \operatorname*{med}_i \{H(\mathbf{z}_i) + \xi + W_i\}, \tag{3.8}$$

with

$$W_i = \begin{cases} 0, & \mathbf{z}_i \in A_4,\\ \widehat{H}(\mathbf{z}_i) - H(\mathbf{z}_i) - \xi, & \mathbf{z}_i \notin A_4, \end{cases} \tag{3.9}$$

and

$$\begin{aligned} \widehat{H}(\mathbf{z}_i) &= \operatorname*{med}_{j,\, j \ne i} h(\mathbf{z}_i, \mathbf{z}_j)\\ &= H(\mathbf{z}_i) + \frac{1}{n-1} \sum_{j,\, j\ne i} \mathrm{IF}_1(\mathbf{z}_i, \mathbf{z}_j) + R_i \overset{\triangle}{=} H(\mathbf{z}_i) + S_i + R_i. \end{aligned} \tag{3.10}$$

The idea of the proof is that taking the median of all $\widehat{H}(\mathbf{z}_i)$ is asymptotically equivalent to taking the median of all $H(\mathbf{z}_i) + \xi$, as in (3.6). With probability tending to 1, both $\widehat{H}(\mathbf{z}_i)$ and $H(\mathbf{z}_i) + \xi$ are too far away from 0 for all $\mathbf{z}_i \in A_4$ to interfere with any of the two medians (Lemma 3.6). The remaining, “interesting,” observations $\mathbf{z}_i \notin A_4$ give values of $\widehat{H}(\mathbf{z}_i)$ close to 0. Among these observations, $\widehat{H}(\mathbf{z}_i) \approx H(\mathbf{z}_i) + \xi$ when $\mathbf{z}_i \in A_1$ (Lemmas 3.2–3.3). In addition, the number of observations from A_2 and A_3 becomes negligible in comparison with the number of observations from A_1 [Lemma 3.4; cf. also (2.17)–(2.18)], so the approximation above is valid for a majority of the "interesting" observations. Finally, $\overline{\beta}_n$ is asymptotically equivalent to ξ (Lemma 3.1), which is what we want to prove. This is because of (3.12), which corresponds to the fact that $l(0) = \infty$.

LEMMA 3.1. *Let* $\overline{\beta}_n$ *be given by* (3.6) *and* IF *by* (2.19), *then*

$$\overline{\beta}_n = \frac{1}{n}\sum_{i=1}^{n} \mathrm{IF}(\mathbf{z}_i) + o_p\left(n^{-1/2}\right). \tag{3.11}$$

PROOF. Since $\mathrm{IF}_1(0, \mathbf{z}) = \mathrm{sgn}(xy)/(2l_0(0))$ and $l_0(0) = f(0)E_G|X|$ according to (A.6), it follows that $\mathrm{IF}_1(0, \mathbf{z}) = \mathrm{IF}(\mathbf{z})$. It therefore suffices to show that

$$\beta_n^* = \operatorname*{med}_i H(\mathbf{z}_i) = o_p\left(n^{-1/2}\right). \tag{3.12}$$

Given $\mathbf{z}_i$, let u_i have a uniform distribution on $[L(H(\mathbf{z}_i)-), L(H(\mathbf{z}_i))]$, independently for each i. Then $u_1, \dots, u_n$ is an i.i.d. sample from a uniform distribution on $[0, 1]$. Denote by $\{H(\mathbf{z})_{(i)}\}$ and $\{u_{(i)}\}$ the ordered samples. It then follows from Theorem B.1 (with C_1 denoting the same constant as there) that, for large enough n and any $\varepsilon > 0$,

$$\begin{aligned}
&P\left(|\beta_n^*| > \frac{\varepsilon}{\sqrt{n}}\right)\\
&\quad = P\left(\left|H(\mathbf{z})_{([n/2+1])}\right| \geq \frac{\varepsilon}{\sqrt{n}}\right)\\
&\quad \leq P\left(u_{([n/2+1])} \leq L\left(-\frac{\varepsilon}{\sqrt{n}}\right) \text{ or } u_{([n/2+1])} \geq L\left(\frac{\varepsilon}{\sqrt{n}}-\right)\right)\\
&\quad \leq P\left(u_{([n/2+1])} \leq L\left(-\frac{\varepsilon}{2\sqrt{n}}-\right) \text{ or } u_{([n/2+1])} \geq L\left(\frac{\varepsilon}{2\sqrt{n}}\right)\right)\\
&\quad \leq P\left(\left|u_{([n/2+1])} - \frac{1}{2}\right| \geq C_1 \frac{\varepsilon/2}{\sqrt{n}} \log\left(\frac{\sqrt{n}}{\varepsilon/2}\right)\right) \to 0 \quad \text{as } n \to \infty,
\end{aligned} \tag{3.13}$$

since $|u_{([n/2+1])} - \frac{1}{2}| = O_p(n^{-1/2})$. $\square$

In the next two lemmas, we show that $|W_i| = |S_i + R_i - \xi|$ is small when $\mathbf{z}_i \in A_1$. We introduce

$$\overline{S} = \max_{\mathbf{z}_i \in A_1} |S_i - \xi|. \tag{3.14}$$

Since $\mathbf{z}_i$ is close to $(0,0)$ when $\mathbf{z}_i \in A_1$, we expect this quantity to be small.

LEMMA 3.2. *As $n \to \infty$, the quantity $\overline{S}$ of* (3.14) *satisfies*

$$\overline{S} = O_p\left(\frac{\delta_n^\eta \log n}{n^{1/2}}\right) = O_p\left(\frac{1}{(\log n)^{\gamma\eta} n^{1/2}}\right). \tag{3.15}$$

PROOF. We first observe that

$$\overline{S} \le \max_{i, |\mathbf{z}_i| \le \delta_n} |S_i - \xi|$$

since $|\mathbf{z}_i| \le \delta_n$ for each $\mathbf{z}_i \in A_1$, and that

$$S_i - \xi = -\frac{\mathrm{IF}(0, \mathbf{z}_i)}{n-1} + \frac{1}{n-1} \sum_{j, j \ne i} \big(\mathrm{IF}(\mathbf{z}_i, \mathbf{z}_j) - \mathrm{IF}(0, \mathbf{z}_j)\big).$$

Hence

$$\overline{S} \le \frac{1}{2(n-1) l_0(0)} + \max_{1 \le i \le n} |\Delta(\mathbf{z}_i)|, \tag{3.16}$$

where

$$\begin{aligned}\Delta(\mathbf{z}_i) &= \frac{I\big(|\mathbf{z}_i| \le \delta_n\big)}{n-1} \sum_{j, j \ne i} \big(\mathrm{IF}(\mathbf{z}_i, \mathbf{z}_j) - \mathrm{IF}(0, \mathbf{z}_j)\big) \\ &\triangleq \frac{I\big(|\mathbf{z}_i| \le \delta_n\big)}{n-1} \sum_{j, j \ne i} Y_{ij}.\end{aligned}$$

With $\mathbf{z}_i$ fixed, $\Delta(\mathbf{z}_i) \equiv 0$ when $|\mathbf{z}_i| > \delta_n$, and if $|\mathbf{z}_i| \le \delta_n$, all $Y_{ij}, j \ne i$, are i.i.d. with zero mean. Suppose in the rest of the proof that n is so large that $\delta_n \le d$, where $0 < d \le 1$ is chosen so small that Theorem A.1 holds with this choice of d, and also that G has a bounded density on $[-d, d]$. Then

$$P\big(|Y_{ij}| \le M\big) = 1, \tag{3.17}$$

where

$$M = \frac{1}{2 l_0(0)} + \frac{1}{2 \inf_{\mathbf{z}, |\mathbf{z}| \le d} l_{\mathbf{z}}\big(H(\mathbf{z})\big)} \le \frac{1}{2 l_0(0)} + \frac{1}{2\underline{l}} < \infty,$$

where the last inequality follows from (A.8), with $\varepsilon = 0$. Now introduce the region $B_{\mathbf{z}} = \{\mathbf{z}';\, h(\mathbf{z}, \mathbf{z}') > H(\mathbf{z})\}$. Then if $|\mathbf{z}_i| \le \delta_n$,

$$\begin{aligned}|Y_{ij}| \le{}& \frac{1}{l_0(0)} I\big(\mathbf{z}_j \in B_{\mathbf{z}_i} \triangle B_0\big) \\ &+ \frac{1}{2}\left|\left(\frac{1}{l_{\mathbf{z}_i}\big(H(\mathbf{z}_i)\big)} - \frac{1}{l_0(0)}\right) \mathrm{sgn}\big(h(\mathbf{z}_i, \mathbf{z}_j) - H(\mathbf{z}_i)\big)\right|,\end{aligned}$$

and hence

$$
\begin{aligned}
E\left(Y_{ij}^2 \mid \mathbf{z}_i\right) &\leq \frac{2}{l_0(0)^2} P_K\left(\mathbf{Z} \in B_{\mathbf{z}_i} \triangle B_0\right) + \frac{1}{2}\left(\frac{1}{l_{\mathbf{z}_i}(H(\mathbf{z}_i))} - \frac{1}{l_0(0)}\right)^2 \\
&\leq \frac{2}{l_0(0)^2} P_K\left(\mathbf{Z} \in B_{\mathbf{z}_i} \triangle B_0\right) + \frac{1}{2\underline{l}^4}\left(l_{\mathbf{z}_i}(H(\mathbf{z}_i)) - l_0(0)\right)^2 \qquad (3.18) \\
&\leq C\left(|\mathbf{z}_i| + |\mathbf{z}_i|^{2\eta}\right) \leq 2C|\mathbf{z}_i|^{2\eta},
\end{aligned}
$$

for some constant $C > 0$. The last inequality in (3.18) holds since $|\mathbf{z}_i| \leq \delta_n \leq 1$ and $0 < \eta < 0.5$, and the second-last inequality follows from (A.10) and the fact that

$$
\begin{aligned}
P_K\left(\mathbf{Z} \in B_{\mathbf{z}_i} \triangle B_0\right) &\leq P_K\Big(\operatorname{sgn}\big(h(0, \mathbf{Z})\big) \neq \operatorname{sgn}\big(h(\mathbf{z}_i, \mathbf{Z})\big)\Big) \\
&\quad + P_K\left(|h(\mathbf{z}_i, \mathbf{Z})| \leq |H(\mathbf{z}_i)|\right) \\
&\leq |G(x_i) - G(0)| + |F(y_i) - F(0)| + \Big|L_{\mathbf{z}_i}\Big\{\big[-H(\mathbf{z}_i), H(\mathbf{z}_i)\big]\Big\}\Big| \\
&\leq C|\mathbf{z}_i|,
\end{aligned}
$$

where the last inequality follows since both F and G have bounded densities on $[-d, d]$, $|H(\mathbf{z}_i)| \leq C'|\mathbf{z}_i|$ by Theorem A.1(i) and (iii) and the fact that $l(\mathbf{z}, t)$ is bounded on Ω_d [cf. (2.17)], since $L(\mathbf{z}, t)$ is a C^1-function on Ω_d according to Theorem A.1(ii).

It then follows from Bernstein's exponential inequality [see Pollard (1984), Appendix B], (3.17) and (3.18) that

$$
\begin{aligned}
P\Big(|\Delta(\mathbf{z}_i)| > t \,\big|\, \mathbf{z}_i\Big) &\leq 2\exp\left(-\frac{(n-1)^2 t^2}{4(n-1)C|\mathbf{z}_i|^{2\eta} + (2/3)M(n-1)t}\right) \\
&\leq 2\exp\left(-\frac{(n-1)^2 t^2}{4(n-1)C\delta_n^{2\eta} + (2/3)M(n-1)t}\right),
\end{aligned}
$$

with C the same constant as in (3.18). Since this inequality holds uniformly in $\mathbf{z}_i$ [remember that $\Delta(\mathbf{z}_i) \equiv 0$ when $|\mathbf{z}_i| > \delta_n$], we obtain

$$
\begin{aligned}
P\left(|\Delta(\mathbf{z}_i)| > \frac{\delta_n^\eta u}{n^{1/2}}\right) &\leq 2\exp\left(-\frac{(n-1)^2\left(\delta_n^{2\eta}/n\right)u^2}{4(n-1)C\delta_n^{2\eta} + (2/3)M(n-1)\left(\delta_n^\eta/n^{1/2}\right)u}\right) \\
&\leq 2\exp\left(-\frac{(1/2)u^2}{4C + u}\right),
\end{aligned}
$$

the last inequality holding for n large enough. This yields

$$
\begin{aligned}
&P\left(\max_{1\le i\le n}|\Delta(\mathbf{z}_i)| > \frac{\delta_n^\eta \log n}{n^{1/2}}v\right)\\
&\quad\le nP\left(|\Delta(\mathbf{z}_1)| > \frac{\delta_n^\eta \log n}{n^{1/2}}v\right)\\
&\quad\le 2n\exp\left(-\frac{(1/2)(\log n)^2 v^2}{4C + v\log n}\right)\\
&\quad\le 2n\exp\left(-\frac{(1/2)(\log n)^2 v^2}{2v\log n}\right) \to 0 \text{ if } n\to\infty \text{ and } v>4,
\end{aligned}
\tag{3.19}
$$

where again the last inequality in (3.19) holds for large enough n. The lemma now follows from (3.16) and (3.19). □

As for the remainder terms R_i, we have the following Bahadur representation result, the proof of which may be found in Hössjer, Rousseeuw and Croux (1992).

LEMMA 3.3. *With R_i as defined in* (3.10) *and* $0 < \eta < 0.5$ *from* (F),

$$
\overline{R} = \max_{\mathbf{z}_i\in A_1}|R_i| = O_p\left(\left(\frac{\log n}{n}\right)^{(1+\eta)/2}\right).
\tag{3.20}
$$

The next lemma controls the number of $\mathbf{z}_i$ in $A_2\cup A_3$.

LEMMA 3.4. *Let A_2 and A_3 be given by* (3.5). *Then*

$$
N = |\{i;\ \mathbf{z}_i\in A_2\cup A_3\}| = o_p\left(n^{1/2}(\log n)^{3/4}\right).
\tag{3.21}
$$

PROOF. Clearly

$$
N \sim \operatorname{Bin}(n, p_n),
\tag{3.22}
$$

where

$$
p_n = K\{A_2\} + K\{A_3\} = \sum_{i=1}^4 K\{A_{2i}\} + \sum_{i=1}^4 K\{A_{3i}\} \triangleq \sum_{i=1}^4 p_{ni}^{(2)} + \sum_{i=1}^4 p_{ni}^{(3)},
$$

A_{2i} is the intersection between A_2 and the i-th quadrant and A_{3i} the intersection between A_3 and the ith quadrant. The lemma will follow if we establish that, for $i = 1, \dots, 4$,

$$p_{ni}^{(2)} = O\left(\varepsilon_n \log\left(\frac{1}{\delta_n}\right)\right)$$

$$(3.23) \qquad = O\left(\frac{(\log n)^{1/2+\gamma}\log\left((\log n)^{\gamma+1/\eta}\right)}{n^{1/2}}\right)$$

$$= O\left(\frac{(\log n)^{1/2+\gamma'}}{n^{1/2}}\right),$$

where $\gamma < \gamma' < \frac{1}{4}$, and

$$(3.24) \qquad p_{ni}^{(3)} = O(\varepsilon_n).$$

We will only consider $i = 1$ (the other cases being similar). Formula (3.23) is established with similar reasoning as in (B.5)–(B.7). In order to prove (3.24), it follows as in the proof of Theorem B.1 that

$$\begin{aligned} A_{31} &= \left\{\mathbf{z};\, x > 0,\, y > 1,\, 2\big(G(x) - 0.5\big)\big(F(y) - 0.5\big) < L_{\mathbf{z}}(\varepsilon_n) - L_{\mathbf{z}}(0) + \rho'\varepsilon_n\right\} \\ &\subseteq \left\{\mathbf{z};\, x > 0,\, 2\big(G(x) - 0.5\big)\big(F(1) - 0.5\big) < \big(E_G|X - x|\,\|f\|_\infty + \rho'\big)\varepsilon_n\right\} \\ &\subseteq \left\{\mathbf{z};\, x > 0,\, G(x) - 0.5 < (C_1 + C_2 x)\varepsilon_n\right\} \\ &\subseteq \left\{\mathbf{z};\, 0 < x \le 1,\, G(x) - 0.5 < (C_1 + C_2)\varepsilon_n\right\} \\ &\qquad \cup \left\{\mathbf{z};\, x > 1,\, (C_1 + C_2 x)\varepsilon_n > G(1) - 0.5\right\}, \end{aligned}$$

for some positive constants C_1 and C_2. Hence,

$$p_{n1}^{(3)} \le (C_1 + C_2)\varepsilon_n + P_G\left(X > \frac{G(1) - 0.5}{C_2\varepsilon_n} - \frac{C_1}{C_2}\right) \le C'\varepsilon_n,$$

for some positive constant C'. □

LEMMA 3.5. *With $\overline{\beta}_n$ and $\widetilde{\beta}_n$ as defined by* (3.6) *and* (3.8),

$$(3.25) \qquad \overline{\beta}_n - \widetilde{\beta}_n = o_p\left(n^{-1/2}\right) \quad \text{as } n \to \infty.$$

PROOF. Set $H_{(i)} = H(\mathbf{z})_{(i)}$ (cf. Lemma 3.1) for short. It then follows from the definition of $\overline{\beta}_n$ and $\widetilde{\beta}_n$ and from (3.10) that

$$(3.26) \; |\overline{\beta}_n - \widetilde{\beta}_n| \le \overline{S} + \overline{R} + \max(H_{([n/2+1]+N)} - H_{([n/2+1])}, H_{([n/2+1])} - H_{([n/2+1]-N)}).$$

By Lemmas 3.2 and 3.3, the first two terms in (3.26) are $o_p(n^{-1/2})$. It thus remains to investigate the last term. We confine ourselves to $H_{([n/2+1]+N)} -$

$H_{([n/2+1])}$, since the treatment of $H_{([n/2+1])} - H_{([n/2+1]-N)}$ is similar. We first notice that

$$\begin{aligned}|H_{([n/2+1]+N)} - H_{([n/2+1])}| &\le |H_{([n/2+1])}| + |H_{([n/2+1]+N)}| \\ &= o_p\left(n^{-1/2}\right) + |H_{([n/2+1]+N)}|\end{aligned}$$

by (3.12). Let $\varepsilon > 0$ be arbitrary. Then

$$\begin{aligned}P\left(|H_{([n/2+1]+N)}| > \frac{\varepsilon}{\sqrt{n}}\right) &\le P\left(N \ge \sqrt{n}(\log n)^{3/4}\right) + P\left(H_{([n/2+1])} < -\frac{\varepsilon}{\sqrt{n}}\right) \\ &\quad + P\left(H_{([n/2+1]+\sqrt{n}(\log n)^{3/4})} > \frac{\varepsilon}{\sqrt{n}}\right) \\ &= o(1) + P\left(H_{([n/2+1]+\sqrt{n}(\log n)^{3/4})} > \frac{\varepsilon}{\sqrt{n}}\right)\end{aligned}$$

because of Lemma 3.4 and (3.12). Choose now ε' such that $0 < \varepsilon' < \varepsilon$. Then

$$\begin{aligned}&P\left(H_{([n/2+1]+\sqrt{n}(\log n)^{3/4})} > \frac{\varepsilon}{\sqrt{n}}\right) \\ &\quad \le P\left(H_{([n/2+1])} \ge \frac{\varepsilon'}{\sqrt{n}}\right) + P\left(H_{([n/2+1])} \le \frac{\varepsilon'}{\sqrt{n}}, H_{([n/2+1]+\sqrt{n}(\log n)^{3/4})} > \frac{\varepsilon}{\sqrt{n}}\right) \\ &\quad \le o(1) + P\left(|\widetilde{N}| \le \sqrt{n}(\log n)^{3/4}\right),\end{aligned}$$

where $\widetilde{N} = |\{i; \varepsilon'/\sqrt{n} < |H(\mathbf{z}_i)| \le \varepsilon/\sqrt{n}\}|$. However, $\widetilde{N} \sim \mathrm{Bin}(n, \widetilde{p}_n)$, where

$$\widetilde{p}_n = L\left(\frac{\varepsilon}{\sqrt{n}}\right) - L\left(\frac{\varepsilon'}{\sqrt{n}}\right) \ge C_1 \frac{\varepsilon}{\sqrt{n}} \log \frac{\sqrt{n}}{\varepsilon} - C_2 \frac{\varepsilon'}{\sqrt{n}} \log \frac{\sqrt{n}}{\varepsilon'} \ge \frac{C_1 \varepsilon (\log n)}{4\sqrt{n}},$$

where C_1 and C_2 are defined in (B.2), and the last inequality holds for large n, provided ε' is chosen small enough. Therefore,

$$P\left(\widetilde{N} \le \sqrt{n}(\log n)^{3/4}\right) \to 0 \quad \text{as } n \to \infty,$$

and hence we have proved that

$$H_{([n/2+1]+N)} - H_{([n/2+1])} = o_p\left(n^{-1/2}\right). \qquad \square$$

LEMMA 3.6. *Let* $\widehat{\beta}_n$ *and* $\widetilde{\beta}_n$ *be defined by* (1.3) *and* (3.8). *Then*

$$\widehat{\beta}_n - \widetilde{\beta}_n = o_p\left(n^{-1/2}\right). \tag{3.27}$$

PROOF. It suffices to show that for each $\varepsilon > 0$ there exists N such that

$$P(\widetilde{\beta}_n = \widehat{\beta}_n,\ n > N) \ge 1 - \varepsilon.$$

Let $0 < \rho < 1$ and subdivide A_4 into A_4^+ and A_4^- according to whether $H(\mathbf{z}) > \varepsilon_n$ or $H(\mathbf{z}) < -\varepsilon_n$. By definition, $\widetilde{\beta}_n = \widehat{\beta}_n$ if the following conditions are satisfied: $|\widetilde{\beta}_n| < (1-\rho)\varepsilon_n$, for all $\mathbf{z}_i \in A_4^+$ the quantities $H(\mathbf{z}_i) + \xi$ and $\widehat{H}(\mathbf{z}_i)$ both exceed $(1-\rho)\varepsilon_n$ and for all $\mathbf{z}_i \in A_4^-$ both $H(\mathbf{z}_i) + \xi$ and $\widehat{H}(\mathbf{z}_i)$ are smaller than $-(1-\rho)\varepsilon_n$. Therefore,

$$
\begin{aligned}
P(\widetilde{\beta}_n \neq \widehat{\beta}_n) \leq {} & P\big(|\widetilde{\beta}_n| \geq (1-\rho)\varepsilon_n\big) + P\big(|\xi| \geq \rho\varepsilon_n\big) \\
& + P\bigg(\min_{\mathbf{z}_i \in A_4^+} \widehat{H}(\mathbf{z}_i) \leq (1-\rho)\varepsilon_n\bigg) \\
& + P\bigg(\max_{\mathbf{z}_i \in A_4^-} \widehat{H}(\mathbf{z}_i) \geq -(1-\rho)\varepsilon_n\bigg).
\end{aligned} \tag{3.28}
$$

We want to show that the RHS of (3.28) tends to 0 as $n \to \infty$. We know from Lemmas 3.1 and 3.5 that $\widetilde{\beta}_n = O_p(n^{-1/2})$, and by the definition of ξ we also have $\xi = O_p(n^{-1/2})$. Hence, the first two terms on the RHS of (3.28) tend to zero as $n \to \infty$. Since the last two terms are similar, we will only study the third. We first show that

$$\inf_{\mathbf{z} \in A_4^+} L_{\mathbf{z}}^{-1}(0.5 - \rho'\varepsilon_n) \geq (1-\rho)\varepsilon_n. \tag{3.29}$$

If $\mathbf{z} \in A_4^+$, then $H(\mathbf{z}) > \varepsilon_n$ and either

$$L_{\mathbf{z}}(\varepsilon_n) \leq 0.5 - \rho'\varepsilon_n \tag{3.30}$$

or

$$|y| \leq 1. \tag{3.31}$$

If (3.30) holds,

$$L_{\mathbf{z}}^{-1}(0.5 - \rho'\varepsilon_n) \geq \varepsilon_n > (1-\rho)\varepsilon_n, \tag{3.32}$$

so it remains to consider those $\mathbf{z} \in A_4^+$ for which (3.31) holds. For any such $\mathbf{z}$, and if n is large enough, we claim that

$$|x| \leq \frac{2}{\varepsilon_n}. \tag{3.33}$$

Suppose, for instance, that $\mathbf{z}$ belongs to the first quadrant. Since $0 \leq y \leq 1$ for any $\mathbf{z}$ satisfying (3.31), $x > 2/\varepsilon_n$ would imply that the line though $\mathbf{z}$ with slope ε_n intersected the x axis at a point with x-coordinate $> 1/\varepsilon_n$ and the y axis at a point with y-coordinate < -1. Hence,

$$L_{\mathbf{z}}(\varepsilon_n) \geq \frac{1}{2}G\bigg(\frac{1}{\varepsilon_n}\bigg) + \frac{1}{2}\big(0.5 - F(-1)\big) > \frac{1}{2},$$

for large enough n, that is, $H(\mathbf{z}) < \varepsilon_n$. Hence, (3.33) must hold if $\mathbf{z} \in A_4^+$ and

$|y| \le 1$. For any $\mathbf{z} \in A_4^+$ satisfying (3.31) we have

$$L_{\mathbf{z}}(\varepsilon_n) - L_{\mathbf{z}}\big((1-\rho)\varepsilon_n\big) \ge \rho\varepsilon_n \inf_{\substack{(1-\rho)\varepsilon_n \\ \le t \le \varepsilon_n}} l_{\mathbf{z}}(t)$$

$$\ge \rho\varepsilon_n \underline{f} \int_{-2/\varepsilon_n}^{2/\varepsilon_n} |x' - x|\, dG(x'), \tag{3.34}$$

where $\underline{f}$ is a lower bound for f on $[-6, 6]$. The last inequality in (3.34) follows from (A.6) and the fact that, for any line through $\mathbf{z}$ with slope t, $|t| \le \varepsilon_n$, those points with x-coordinate in $[-2/\varepsilon_n, 2/\varepsilon_n]$ have y-coordinates in $[-6, 6]$ because of (3.33). It is not hard to see that the integral in (3.34) can be lower bounded by some positive constant I when $\varepsilon_n \le 1$ (say), uniformly for all x. Hence, for any $\mathbf{z} \in A_4^+$ satisfying (3.31),

$$L_{\mathbf{z}}\big((1-\rho)\varepsilon_n\big) \le 0.5 - \underline{f} I \rho\varepsilon_n < 0.5 - \rho'\varepsilon_n, \tag{3.35}$$

if we choose ρ' so that $0 < \rho' < \underline{f} I \rho$. Formula (3.29) now follows from (3.32) and (3.35). For ease of notation, set $\underline{H}(\mathbf{z}) = L_{\mathbf{z}}^{-1}(0.5 - \rho'\varepsilon_n)$. Our next objective is to show that

$$\max_{\mathbf{z}_i \in A_4^+} \big|L_{\mathbf{z}_i, n-1}\big(\underline{H}(\mathbf{z}_i)\big) - L_{\mathbf{z}_i}\big(\underline{H}(\mathbf{z}_i)\big)\big| = O_p\big((\log n)^{1/2} n^{-1/2}\big). \tag{3.36}$$

Actually, (3.36) is a consequence of Hoeffding's exponential inequality [cf. Pollard (1984), Appendix B], which in our case implies (after first conditioning on $\mathbf{z}_i$) that

$$P\Big(\big|L_{\mathbf{z}_i, n-1}\big(\underline{H}(\mathbf{z}_i)\big) - L_{\mathbf{z}_i}\big(\underline{H}(\mathbf{z}_i)\big)\big| \ge t(n-1)^{-1/2}\Big) \le 2\exp\big(-2t^2\big).$$

By definition $L_{\mathbf{z}_i}(\underline{H}(\mathbf{z}_i)) = 0.5 - \rho'\varepsilon_n$, so it follows from (3.3) and (3.36) that with probability tending to 1,

$$\max_{\mathbf{z}_i \in A_4^+} L_{\mathbf{z}_i, n-1}\big(\underline{H}(\mathbf{z}_i)\big) < 0.5.$$

Hence, because of (3.29),

$$\min_{\mathbf{z}_i \in A_4^+} \widehat{H}(\mathbf{z}_i) \ge \min_{\mathbf{z}_i \in A_4^+} \underline{H}(\mathbf{z}_i) \ge (1-\rho)\varepsilon_n,$$

with probability tending to 1. This shows that the third term of (3.28) goes to 0. $\square$

It is now easy to complete the proof of Theorem 3.1.

PROOF OF THEOREM 3.1. The result follows from Lemmas 3.1, 3.5 and 3.6 together with the central limit theorem and Slutsky's lemma. $\square$

TABLE 1
Simulation results of the repeated median slope estimator, applied to bivariate Gaussian data

n	Bias	n-fold variance	Finite-sample efficiency
10	−0.0035	2.615	38.2%
20	0.0009	1.880	53.2%
40	−0.0006	1.670	59.9%
60	0.0015	1.666	60.0%
100	−0.0004	1.628	61.4%
200	0.0007	1.627	61.5%
300	−0.0002	1.655	60.4%
500	0.0009	1.644	60.8%
800	0.0010	1.620	61.7%
1000	0.0004	1.673	59.8%
2000	0.0005	1.825	54.8%
3000	−0.0002	1.801	55.5%
5000	−0.0012	1.816	55.1%
10000	0.0006	1.747	57.2%
20000	−0.0002	1.848	54.1%
40000	−0.0003	1.861	53.7%
∞	0.0000	2.467	40.5%

4. Finite-sample efficiencies. Theorem 3.1 confirms that the asymptotic variance of the RM slope estimator is given by the expected square of its IF. Therefore, when both G and F equal the standard Gaussian distribution we obtain the asymptotic variance $\pi^2/4 \approx 2.467$ and the corresponding asymptotic efficiency $4/\pi^2 \approx 40.5\%$.

In order to check whether this asymptotic variance provides a good approximation to the variance of the RM slope at finite samples, we carried out a Monte Carlo experiment. For each n in Table 1 we generated $m = 10{,}000$ samples of size n and computed the corresponding slope estimates $\widehat{\beta}_n^{(k)}$ for $k = 1, \dots, m$. Table 1 lists the bias

$$\operatorname*{average}_{k=1,\dots,m} \left(\widehat{\beta}_n^{(k)} - \beta\right),$$

where the true β equals 0 by construction. It also gives the n-fold variance

$$n \operatorname*{variance}_{k=1,\dots,m} \widehat{\beta}_n^{(k)}$$

which should converge (as n tends to ∞) to 2.467. The last column of Table 1 gives the corresponding finite-sample efficiency (in the sense of the information inequality).

The Gaussian variables in the simulation were generated by means of the Box–Muller transform. For $n \leq 5000$, the naive algorithm for the RM slope was used, with computation time $O(n^2)$. These results were also confirmed with the fast algorithm described in Rousseeuw, Netanyahu and Mount (1993), needing only $O(n \log^2 n)$ time. The results for $n \geq 10{,}000$ could only be obtained with the fast algorithm. The n-fold variances in the table have a standard error of approximately 0.025, and that of the finite-sample efficiencies is slightly less than 1%.

In addition to computing the average estimated value and the n-fold variance for each n, we also made Gaussian Q–Q plots of the set $\{\widehat{\beta}_n^{(k)}, k = 1, \ldots, m\}$ of estimated slopes. From these it does appear that the sampling distribution of the estimator $\widehat{\beta}_n$ is approximately Gaussian.

The first three lines of Table 1 confirm the Monte Carlo results of Siegel [(1982), page 243] and Johnstone and Velleman [(1985), page 1051], who found that for $n \leq 40$ the finite-sample efficiencies are increasing with n. In the next lines of the table, we see that the efficiencies stay around 60%–61% for n up to about 1000, after which they slowly decrease. For n around 40,000, we obtain 54%. A way to explain these high finite-sample efficiencies is by looking at the proof of the asymptotic normality, in which the remainder term tends to zero at a very slow rate. The underlying cause for this is the slow convergence of K_ε to K_0. As a consequence, unusually large samples are needed before the finite-sample efficiency comes close to its asymptotic limit of 40.5%.

In conclusion, the RM slope estimator performs much better at finite samples than would be expected from its asymptotics. More information on the finite-sample behavior of this estimator can be found in Rousseeuw, Croux and Hössjer (1994), including data-based approximations to the influence function and a numerical study of the function H defined in the beginning of Section 2.

REMARK. The efficiency of the RM method could still be increased by replacing the outer median in (1.3) by an M-estimator. In the notation of Section 2, T_1 remains the median whereas T_2 becomes an M-estimator. If T_2 has a 50% breakdown point, so will the resulting slope estimator. We carried out a small simulation for n between 10 and 200 with the same setup as in Table 1, using a scale-equivariant one-step Huber estimator with bending constant 1.5 for T_2. The resulting Monte Carlo variances were roughly 12% lower than those of the plain RM slope.

5. Weaker assumptions on the carrier distribution. Our assumptions on the carrier distribution G in Theorem 3.1 are quite restrictive, and we will now discuss what happens when these conditions are relaxed. First of all, (3.1) still holds if

$$C_1|x|^\tau \leq g(x) \leq C_2|x|^\tau \tag{5.1}$$

holds in a neighborhood of 0 for some $\tau \geq 0$ and $C_1, C_2 > 0$. The reason is that the number of observations in A_1 still dominates the number of observations in $A_2 \cup A_3$, and Lemma 3.2 can also be pushed through with small changes. However, if there exist $a < 0 < b$ such that $G\{(a,b)\} = 0$ and G has a density to the right of b and to the left of a such that $g(b+), g(a-) > 0$, then K_0 has a two-point distribution concentrated at $(a, 0)$ and $(b, 0)$. The techniques of Theorem 3.1 cover only the case when K_0 is a one-point distribution, so a separate proof is needed to verify that (2.19) still holds.

Another extension is to allow G to have point masses. In this case we have $x_i = x_j$ for some $i \neq j$ with positive probability and we define

$$\widehat{H}(\mathbf{z}_i) = \operatorname*{med}_{j, x_j \neq x_i} h(\mathbf{z}_i, \mathbf{z}_j),$$

which leads to

$$H(\mathbf{z}) = \operatorname{med}\big(h(\mathbf{z}, \mathbf{Z}) \,|\, X \neq x\big) \tag{5.2}$$

with $\mathbf{z} = (x, y)$ and $\mathbf{Z} = (X, Y)$, and, after some calculations,

$$\operatorname{sgn}\big(H(\mathbf{z})\big) = \operatorname{sgn}\left(\left(F(y) - \frac{1}{2}\right)\left(\frac{G(x-) + G(x)}{2} - \frac{1}{2}\right)\right). \tag{5.3}$$

Note that (5.2) and (5.3) agree with (2.7) and (A.3) when $G(\{x\}) = 0$ [assuming $F(0) = G(0) = 0.5$ in (A.3)]. The proof of Theorem 3.1 goes through with only technical changes when all point masses of G are outside $(G^{-1}(0.5 - \delta), G^{-1}(0.5 + \delta))$ for some $\delta > 0$. This is because $\widehat{H}(\mathbf{z}_i)$ and $H(\mathbf{z}_i)$ are unchanged for all "interesting" data points $\mathbf{z}_i \in A_1$ provided n is so large that $[-\delta_n, \delta_n]$ [cf. (3.4)] contains no point masses of G.

If G has a point mass at its median the situation changes. Assume, for instance, $G(0-) = 0.5 - \delta'$ and $G(0) = 0.5 + \delta''$, with $\delta', \delta'' > 0$. If $\delta' = \delta''$ it follows from (5.3) that $\{\mathbf{z}; H(\mathbf{z}) = 0\} = \{\mathbf{z}; xy = 0\}$. It is easy to see that the points along the y axis will soon dominate the set $\{\mathbf{z}; |H(\mathbf{z})| \leq \varepsilon\}$ as $\varepsilon \to 0$, so that $K_0 = \delta_0 \times F$. In particular, the support of K_0 becomes the whole y axis. If on the other hand $\delta' \neq \delta''$ it again follows from (5.3) that $\{\mathbf{z}; H(\mathbf{z}) = 0\} = \{\mathbf{z}; y = 0\}$ and the support of K_0 becomes the whole x axis. (In this case a more refined analysis is needed to find the exact form of K_0.) Observe, however, that formula (2.19) is no longer valid when $\operatorname{supp}(K_0)$ contains points where G has a point mass, because the analysis in Section 2 requires that the double sum in (2.2) is taken over all $i \neq j$ such that $\mathrm{IF}_2'(H(\mathbf{z}_i)) \neq 0$. Therefore, a separate formula has to be worked out for the influence function when G has a point mass at its median.

A further generalization is to fixed design, that is, suppose $x_1, \dots, x_n$ are all fixed. This implies that $\{\mathbf{z}_i\}$ are independent but not identically distributed random variables. We conjecture that if the empirical distribution

$$G_n = \frac{1}{n}\sum_{i=1}^{n} \delta_{x_i}$$

converges weakly to some distribution G, the influence function of the estimator becomes the same as for a random design with carrier distribution G. The proof of Theorem 3.1 made use of Bahadur representation theorems (Lemma 3.3) and exponential inequalities (Lemmas 3.2 and 3.6) for independent and identically distributed random variables. In the fixed-design case one has to use similar theorems for independent and nonidentically distributed random variables.

APPENDIX A

In this appendix we establish a number of properties of the distribution $L_{\mathbf{z}}$ introduced in Section 2. Its distribution function may be written

$$L_{\mathbf{z}}(t) = P_K\big(h(\mathbf{z}, \mathbf{Z}) \le t\big) = L(\mathbf{z}, t). \tag{A.1}$$

For the slope kernel function this becomes:

$$\begin{aligned} L_{\mathbf{z}}(t) = &\int_{-\infty}^{x} \big(1 - F(y + t(x' - x)\big)\, dG(x') \\ &+ \int_{x}^{\infty} F\big(y + t(x' - x)\big)\, dG(x'). \end{aligned} \tag{A.2}$$

We then have the following theorem.

THEOREM A.1. *Suppose that* $\alpha = \beta = 0$ *in* (1.1), *with the error and carrier distributions satisfying conditions* (F) *and* (G) *of Section* 3.

(i) *The function* $H(\mathbf{z})$ *then satisfies*

$$\operatorname{sgn}\big(H(\mathbf{z})\big) = \operatorname{sgn}(xy), \tag{A.3}$$

and

$$L^{-1}(0.5) = \operatorname*{med}_{\mathbf{Z} \sim K} H(\mathbf{Z}) = 0. \tag{A.4}$$

(ii) *Moreover*, *there exists* $d > 0$ *such that* $L(\mathbf{z}, t)$ *is a* C^1-*function on* $\Omega_d \times \mathbb{R}$ [*cf*. (2.17)], *with*

$$\begin{aligned} \frac{\partial L(\mathbf{z}, t)}{\partial x} &= t \int_{-\infty}^{x} f\big(y + t(x' - x)\big)\, dG(x') \\ &\quad - t \int_{x}^{\infty} f\big(y + t(x' - x)\big)\, dG(x') + \big(1 - 2F(y)\big) g(x), \\ \frac{\partial L(\mathbf{z}, t)}{\partial y} &= - \int_{-\infty}^{x} f\big(y + t(x' - x)\big)\, dG(x') \\ &\quad + \int_{x}^{\infty} f\big(y + t(x' - x)\big)\, dG(x') \end{aligned} \tag{A.5}$$

and

$$\frac{\partial L(\mathbf{z}, t)}{\partial t} = l(\mathbf{z}, t) = l_{\mathbf{z}}(t) = \int_{-\infty}^{\infty} |x' - x| f\big(y + t(x' - x)\big)\, dG(x'), \tag{A.6}$$

and, in particular, $l(0,0) = f(0)E_G|X|$. *Moreover, for each* $\mathbf{z} \in \mathbb{R}^2$, (A.5)–(A.6) *hold;* $l_{\mathbf{z}}(\cdot)$ *is the density of* $L_{\mathbf{z}}(\cdot)$; *and* $l_{\mathbf{z}}(t) > 0$ *for all* t.

(iii) *The function* H *is* C^1*-differentiable on* Ω_d, *with*

$$\left(\frac{\partial H(\mathbf{z})}{\partial x}, \frac{\partial H(\mathbf{z})}{\partial y}\right) = -\left(\frac{\partial L(\mathbf{z},t)/\partial x}{\partial L(\mathbf{z},t)/\partial t}, \frac{\partial L(\mathbf{z},t)/\partial y}{\partial L(\mathbf{z},t)/\partial t}\right)\Bigg|_{t=H(\mathbf{z})}. \tag{A.7}$$

(iv) *The density function* $l_{\mathbf{z}}(H(\mathbf{z}))$ *is continuous on* Ω_d.

(v) *If* $0 \le \varepsilon < 0.5$ *and* $I_\varepsilon = [0.5-\varepsilon, 0.5+\varepsilon]$, *there exist positive numbers* $\underline{l}$ *and* $\overline{L}$ *such that*

$$\underline{l} = \inf_{(\mathbf{z},u)\in\Omega_d\times I_\varepsilon} \left|l_{\mathbf{z}}\left(L_{\mathbf{z}}^{-1}(u)\right)\right| > 0 \tag{A.8}$$

and

$$\overline{L} = \sup_{\mathbf{z}\in\Omega_d}\sup_{t_1\ne t_2} \frac{|l_{\mathbf{z}}(t_2) - l_{\mathbf{z}}(t_1)|}{|t_2-t_1|^\eta} < \infty. \tag{A.9}$$

(vi) *If* η *is defined as in* (F), *there exists a positive constant* C *such that*

$$\left|l_{\mathbf{z}}\left(H(\mathbf{z})\right) - l_0(0)\right| \le C|\mathbf{z}|^\eta \tag{A.10}$$

as soon as $|\mathbf{z}| \le d$.

For the proof we refer to Hössjer, Rousseeuw and Croux (1992).

APPENDIX B

In this Appendix we consider the distribution

$$L(t) = P_K\left(H(\mathbf{z}) \le t\right), \tag{B.1}$$

and we formalize the statement, made in Section 2, that the density of L is infinite at 0:

THEOREM B.1. *Under the same assumptions as in Theorem* A.1, *there exist positive constants* C_1 *and* C_2 *such that, for small enough* $\varepsilon > 0$, *it holds that*

$$\begin{aligned} C_1\varepsilon\log\frac{1}{\varepsilon} &\le L\left\{[0,\varepsilon]\right\} \le C_2\varepsilon\log\frac{1}{\varepsilon}, \\ C_1\varepsilon\log\frac{1}{\varepsilon} &\le L\left\{[-\varepsilon,0]\right\} \le C_2\varepsilon\log\frac{1}{\varepsilon}. \end{aligned} \tag{B.2}$$

PROOF. Given $\varepsilon > 0$, let $A_\varepsilon = \{\mathbf{z}; |H(\mathbf{z})| \le \varepsilon\}$ and let $A_{\varepsilon i}$ be the intersection between A_ε and the ith quadrant. Since $\mathrm{sgn}(H(\mathbf{z})) = \mathrm{sgn}(xy)$, it suffices to show that

$$\text{(B.3)} \qquad C_1\varepsilon \log\frac{1}{\varepsilon} \le K\{A_{\varepsilon i}\} \le \frac{1}{2}C_2\varepsilon\log\frac{1}{\varepsilon} \quad \text{for } i = 1,\dots,4,$$

with C_1 and C_2 the same positive constants as in (B.2). We confine ourselves to $A_{\varepsilon 1}$ (the other cases being similar). Suppose therefore in the rest of the proof that $\mathbf{z}$ lies in the first quadrant. Then $H(\mathbf{z}) > 0$ by Theorem A.1(i), and $H(\mathbf{z}) \le \varepsilon$ is equivalent to $L(\mathbf{z},\varepsilon) \ge 0.5$, since $l_{\mathbf{z}}(\cdot) > 0$ by Theorem A.1(ii). Moreover, since

$$\begin{aligned} L(\mathbf{z},0) &= \big(1 - G(x)\big)F(y) + G(x)\big(1 - F(y)\big) \\ &= 0.5 - 2\big(G(x) - 0.5\big)\big(F(y) - 0.5\big), \end{aligned}$$

it follows that

$$\text{(B.4)} \quad A_{\varepsilon 1} \subseteq \Big\{\mathbf{z};\, x,y > 0,\, 2\big(G(x) - 0.5\big)\big(F(y) - 0.5\big) \le L(\mathbf{z},\varepsilon) - L(\mathbf{z},0)\Big\}.$$

Choose $d > 0$ so small that G has a density lower-bounded by $\underline{g} > 0$ and upper-bounded by $\overline{g} < \infty$ on $[-2d, 2d]$. Let also $\underline{f} > 0$ be a lower bound for f on $[-2d, 2d]$. From (B.4) we obtain [since $l(\mathbf{z},t) \le \|f\|_\infty E_G|X - x|$ by (A.6)],

$$\begin{aligned} A_{\varepsilon i} &\subseteq \Big\{\, \mathbf{z};\, x,y > 0,\, 2\big(G(x) - 0.5\big)\big(F(y) - 0.5\big) \le \|f\|_\infty E_G|X - x|\varepsilon\Big\} \\ &\subseteq \Big\{\, \mathbf{z};\, 0 < x,y < d,\, 2\underline{g}\underline{f}xy \le \|f\|_\infty E_G|X - d|\varepsilon\Big\} \\ &\quad \cup \Big\{\mathbf{z};\, 0 < x < d, y \ge d,\, 2\underline{g}\big(F(d) - 0.5\big)x \le \|f\|_\infty E_G|X - d|\varepsilon\Big\} \\ &\quad \cup \Big\{\mathbf{z};\, x \ge d,\, 0 \le 2\big(G(d) - 0.5\big)\big(F(y) - 0.5\big) \le \|f\|_\infty\big(E_G|X| + x\big)\varepsilon\Big\} \\ &\triangleq \widetilde{A}_1 \cup \widetilde{A}_2 \cup \widetilde{A}_3. \end{aligned}$$

Clearly, with $C = \|f\|_\infty E_G|X - d|/(2\underline{g}\underline{f})$, and ε so small that $\varepsilon \le d^2/C$,

$$\text{(B.5)} \qquad \begin{aligned} K\{\widetilde{A}_1\} &\le \int_0^d F\bigg\{\bigg[0, d \wedge \frac{C\varepsilon}{x}\bigg]\bigg\}g(x)\,dx \\ &\le \|f\|_\infty \int_0^d \bigg(d \wedge \frac{C\varepsilon}{x}\bigg)g(x)\,dx \\ &\le \|f\|_\infty d\int_0^{C\varepsilon/d} g(x)\,dx + \|f\|_\infty C\varepsilon\int_{C\varepsilon/d}^d \frac{1}{x}g(x)\,dx \\ &\le \|f\|_\infty\overline{g}C\varepsilon + \|f\|_\infty\overline{g}C\varepsilon\log\frac{d^2}{C\varepsilon} \le C'\varepsilon\,\log\frac{1}{\varepsilon}. \end{aligned}$$

If instead $C = \|f\|_\infty E_G|X - d|/(2(F(d) - 0.5)\underline{g})$, then

$$\text{(B.6)} \qquad K\{\widetilde{A}_2\} \le G\big\{[0, C\varepsilon]\big\} \le \overline{g}C\varepsilon.$$

Next, we can find constants $C_3, C_4 > 0$ such that $0 \leq F(y) - 0.5 \leq (C_3 + C_4 x)\varepsilon$ whenever $\mathbf{z} \in \widetilde{A}_3$, and hence

$$K\{\widetilde{A}_3\} \leq \int_d^{\infty} (C_3 + C_4 x)\varepsilon g(x)\,dx \leq \big(C_3 + C_4 E_G|X|\big)\varepsilon. \tag{B.7}$$

The second inequality in (B.3) now follows from (B.5)–(B.7). The first inequality is established in a similar manner. First, a positive lower bound for $l(\mathbf{z}, t)$ is obtained for $\mathbf{z} \in (0, d) \times (0, d)$ and $0 < t < \varepsilon$. Then one shows that $K\{A_{\varepsilon 1} \cap (0, d) \times (0, d)\} \geq C_1 \varepsilon \log(1/\varepsilon)$, for some constant $C_1 > 0$. More details can be found in Hössjer, Rousseeuw and Croux (1992). □

Acknowledgments. We would like to thank Bert van Zomeren and Nathan Netanyahu for running part of the simulations reported in Table 1. Professor R. W. Farebrother kindly informed us about the work of Boscovich (1757). We are also grateful to the Editor, an Associate Editor, and the referees for their helpful comments.

REFERENCES

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

BOSCOVICH, R. J. (1757). De litteraria expeditione per pontificiam ditionem, et synopsis amplioris operis. *Bononiensi Scientiarum et Artum Instituto atque Academia Commentarii* **4** 353–396.

FREES, E. W. (1991). Trimmed slope estimates for simple linear regression. *J. Statist. Plann. Inference* **27** 203–221.

HAMPEL, F. R. (1974). The influence function and its role in robust estimation. *J. Amer. Statist. Assoc.* **69** 383–393.

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

HE, X. and SIMPSON, D. (1993). Lower bounds for contamination bias: Globally minimax versus locally linear estimation. *Ann. Statist.* **21** 314–337.

HÖSSJER, O., ROUSSEEUW, P. J. and CROUX, C. (1992). Influence function and asymptotic normality of the repeated median slope estimator. Report 1992:2, Dept. Mathematics, Uppsala Univ.

JOHNSTONE, I. M. and VELLEMAN, P. F. (1985). The resistant line and related regression methods. *J. Amer. Statist. Assoc.* **80** 1041–1054.

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

RONCHETTI, E. and ROUSSEEUW, P. J. (1985). Change-of-variance sensitivities in regression analysis. *Z. Wahrsch. Verw. Gebiete* **68** 503–519.

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

ROUSSEEUW, P. J., CROUX, C. and HÖSSJER, O. (1994). Sensitivity functions and numerical analysis of the repeated median slope. *Computational Statistics*. To appear.

ROUSSEEUW, P. J. and LEROY, A. (1987). *Robust Regression and Outlier Detection.* Wiley, New York.

ROUSSEEUW, P. J., NETANYAHU, N. S. and MOUNT, D. M. (1993). New statistical and computational results on the repeated median line. In *New Directions in Data Analysis and Robustness* (S. Morgenthaler, E. Ronchetti and W. A. Stahel, eds.) 177–194. Birkhäuser.

SEN, P. K. (1968). Estimates of the regression coefficient based on Kendall's tau. *J. Amer. Statist. Assoc.* **63** 1379–1389.

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

SIEGEL, A. F. (1982). Robust regression using repeated medians. *Biometrika* **69** 242–244.

STIGLER, S. (1986). *The History of Statistics: The Measurement of Uncertainty before 1900.* Harvard Univ. Press.

THEIL, H. (1950). A rank-invariant method of linear and polynomial regression analysis, I, II and III. *Koninklijke Nederlandse Akademie van Wetenschappen, Proceedings* **53** 386–392; 521–525; 1397–1412.

OLA HÖSSJER
DEPT. OF MATHEMATICAL STATISTICS
LUND INSTITUTE OF TECHNOLOGY
BOX 118
S-221 00 LUND
SWEDEN

PETER J. ROUSSEEUW
CHRISTOPHE CROUX
DEPARTMENT OF MATHEMATICS, UIA
UNIVERSITEITSPLEIN 1
B-2610 ANTWERPEN
BELGIUM