The Repeated Median Intercept Estimator: Influence Function and Asymptotic Normality

OLA HÖSSJER

University of Lund, Lund, Sweden

AND

PETER J. ROUSSEEUW AND IDA RUTS

University of Antwerp, Antwerp, Belgium

Given the simple linear regression model $Y_i = \alpha + \beta X_i + e_i$ for $i = 1, ..., n$, we consider the repeated median estimator of the intercept α, defined as $\hat{\alpha}_n = \text{med}_i \, \text{med}_{j, j \neq i} (X_j Y_i - X_i Y_j)/(X_j - X_i)$. We determine the influence function and prove asymptotic normality for $\hat{\alpha}_n$ when the carriers X_i and error terms e_i are random. The resulting influence function is bounded, and is the same as if the intercept is estimated by the median of the residuals from a preliminary slope estimator. With bivariate gaussian data the efficiency becomes $2/\pi \approx 63.7\%$. The asymptotic results are compared with sensitivity functions and finite-sample efficiencies.

1. INTRODUCTION

Consider the simple linear regression model

$$Y_i = \alpha + \beta X_i + e_i, \qquad i = 1, ..., n, \tag{1.1}$$

where $\{\mathbf{Z}_i = (X_i, Y_i)\}$ are the observed vectors and $\{e_i\}$ represent 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 continuous distributions G and F for which $F(0) = \frac{1}{2} = G(0)$. When $X_1, ..., X_n$ are all distinct (an event with probability one), we may define the repeated median estimator of the intercept α,

$$\hat{\alpha}_n = \underset{i}{\text{med}} \underset{j, j \neq i}{\text{med}} \frac{X_j Y_i - X_i Y_j}{X_j - X_i}, \tag{1.2}$$

Received July 30, 1992; revised May 5, 1994.

AMS 1991 subject classifications: 62F35, 62J05.

Key words and phrases: asymptotic normality, influence function, intercept estimator, repeated median, simple linear regression.

0047-259X/95 $6.00

introduced by Siegel (1982). Given two points $\mathbf{z}_1 = (x_1, y_1)$ and $\mathbf{z}_2 = (x_2, y_2)$ with $x_1 \neq x_2$, the kernel function $h(\mathbf{z}_1, \mathbf{z}_2) = (x_2 y_1 - x_1 y_2)/(x_2 - x_1)$ gives the intersection between the y-axis and the line through $\mathbf{z}_1$ and $\mathbf{z}_2$. Siegel showed that $\hat{\alpha}_n$ has a 50% breakdown point, and is a Fisher consistent estimator of α (see Cox and Hinkley, 1974, p. 287) when F is a symmetric distribution. In this paper we will prove asymptotic normality (Section 3). The main result, Theorem 3.1, is that

$$\sqrt{n}(\hat{\alpha}_n - \alpha) \xrightarrow{d} N(0, (2f(0))^{-2}) \qquad \text{as} \quad n \to \infty. \tag{1.3}$$

The asymptotics will be compared with Monte Carlo results in Section 4.

By changing the kernel function in (1.2), we obtain repeated median estimators of other quantities. For instance, when $h(\mathbf{z}_1, \mathbf{z}_2) = (y_2 - y_1)/(x_2 - x_1)$ we obtain a repeated median estimator of the slope parameter β, as introduced by Siegel (1982). Influence function and asymptotic normality for this estimator were proved by Hössjer *et al.* 1994, henceforth denoted by HRC. The techniques used in this paper are based on those used in HRC.

In order to simplify expressions we will assume from now on that $\alpha = 0 = \beta$, which because of regression equivariance (cf. Rousseeuw and Leroy, 1987, p. 116) is no restriction.

2. Notation and Regularity Conditions

For a fixed $\mathbf{z}$ we define

$$L_{\mathbf{z}}(t) = P_K(h(\mathbf{z}, \mathbf{Z}) \leqslant t) \tag{2.1}$$

where $K = G \times F$ is the distribution of the random vector $\mathbf{Z}$, and $h(\mathbf{z}_1, \mathbf{z}_2) = (y_1 x_2 - y_2 x_1)/(x_2 - x_1)$ is the kernel function for the intercept estimator. Put

$$H(\mathbf{z}) = L_{\mathbf{z}}^{-1}(0.5) = \operatorname*{med}_{\mathbf{Z} \sim K} h(\mathbf{z}, \mathbf{Z}), \tag{2.2}$$

where

$$L_{\mathbf{z}}^{-1}(u) = \inf\{t; L_{\mathbf{z}}(t) > u\}$$

denotes the right continuous inverse of $L_{\mathbf{z}}$. (We use the right continuous inverse of the distribution function, evaluated at 0.5, as a definition of the median throughout the paper. The sample median then corresponds to inverting the empirical distribution function formed by the sample, which yields the observation with rank $[n/2]+1$.) Then define

$$L(t)=P_K(H(\mathbf{Z})\leqslant t). \tag{2.3}$$

Because L is symmetric (Lemma A.1),

$$L^{-1}(0.5)=\operatorname*{med}_{\mathbf{Z}\sim K} H(\mathbf{Z})=0. \tag{2.4}$$

Regarding $\mathbf{Z}_i$ as fixed, the finite sample counterpart of $L_{\mathbf{Z}_i}$ is

$$L_{\mathbf{Z}_i,n-1}(t)=\frac{1}{n-1}\sum_{j,j\neq i} I(h(\mathbf{Z}_i,\mathbf{Z}_j)\leqslant t), \tag{2.5}$$

and we also put

$$\hat{H}(\mathbf{Z}_i)=L^{-1}_{\mathbf{Z}_i,n-1}(0.5)=\operatorname*{med}_{j,j\neq i} h(\mathbf{Z}_i,\mathbf{Z}_j) \tag{2.6}$$

as an approximation of $H(\mathbf{Z}_i)$. With this notation,

$$\hat{\alpha}_n=\operatorname*{med}_i \hat{H}(\mathbf{Z}_i).$$

The following regularity conditions on the error distribution F and the carrier distribution G will be imposed:

(F) F has a bounded, strictly positive, and asymmetric density f, which is Lipschitz continuous of order η for some $0<\eta\leqslant 1$, that is,

$$\|f\|_\eta=\sup_{x\neq y}\frac{|f(x)-f(y)|}{|x-y|^\eta}<\infty.$$

(G) The carrier distribution G is symmetric and continuous and $E_G|X|^{1+\eta}<\infty$. In addition, either $|X|$ is bounded a.s. or

$$\limsup_{x\to\infty}\frac{E_G(XI(X\geqslant x))}{xP_G(X\geqslant x)}<\infty.$$

We will also use the notation of influence functions. Let T_n be a statistic such that $T_n\xrightarrow{P}T(K)$ as $n\to\infty$, where the asymptotic value $T(K)$ depends on the distribution K of the i.i.d. observations. Let $\delta_{\mathbf{z}}$ denote the point mass at $\mathbf{z}$. Then the influence function (Hampel *et al.*, 1986) of T at K,

$$IF(T,K,\mathbf{z})=IF(\mathbf{z})=\lim_{\varepsilon\to 0+}\frac{T((1-\varepsilon)K+\varepsilon\delta_{\mathbf{z}})}{\varepsilon},$$

measures the effect that an additional observation $\mathbf{z}$ has on T_n. By means of a von Mises expansion of the functional T, we obtain under certain regularity conditions

$$T_n = T(K) + \frac{1}{n}\sum_{i=1}^{n} IF(\mathbf{Z}_i) + o_p(n^{-1/2}). \tag{2.7}$$

Formula (2.7) implies, via the Central Limit Theorem, that

$$n^{1/2}(T_n - T(K)) \xrightarrow{d} N\left(0, \int IF(\mathbf{z})^2 \, dK(\mathbf{z})\right), \tag{2.8}$$

provided the integral in (2.8) is finite. The asymptotic linearity of the estimator as expressed in (2.7) can actually be used as an alternative definition of the influence function.

In Section 2 of HRC, a general expression for the influence function of repeated median estimators was given. For the intercept kernel, this formula simplifies to

$$\begin{aligned} IF(\mathbf{z}) &= \frac{1}{2f(F^{-1}(1/2))}\operatorname{sgn}\left(y - F^{-1}\left(\frac{1}{2}\right) - \alpha - \beta\left(x - G^{-1}\left(\frac{1}{2}\right)\right)\right) \\ &= \frac{1}{2f(0)}\operatorname{sgn}(y). \end{aligned} \tag{2.9}$$

A detailed derivation was given in Hössjer *et al.* (1992). In Section 3 we will prove that (2.7) holds for the function given in (2.9).

3. Asymptotic Normality

The main result of the paper is the following.

Theorem 3.1. *Let $\hat{\alpha}_n$ be the estimator of α defined in* (1.2). *Under the regularity conditions* (F) *and* (G), *it holds that*

$$\sqrt{n}\,\hat{\alpha}_n = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} IF(\mathbf{Z}_i) + o_p(1) \xrightarrow{d} N(0, (2f(0))^{-2}) \qquad \text{as} \quad n \to \infty \tag{3.1}$$

where $IF(\cdot)$ is defined by expression (2.9).

Theorem 3.1 will be proved through a series of lemmas. The technique of the proof is similar to the one used in HRC. We will introduce two statistics α_1 and α_2, which approximate $\hat{\alpha}_n$. Let γ and τ be positive numbers such that $\gamma+\tau<\frac{1}{4}$, and define the sequence of constants

$$\varepsilon_n=(\log n)^{1/2+\gamma}n^{-1/2}, \tag{3.2}$$

$$b_n=e^{-(\log n)^\tau}, \tag{3.3}$$

and a_n, which satisfies

$$1-G(a_n)=e^{-(\log n)^\tau}. \tag{3.4}$$

Next, we subdivide $\mathbb{R}^2$ into the regions

$$A_1=\{\mathbf{z}=(x,y);\ |H(\mathbf{z})|\leqslant\varepsilon_n,\ |x|\geqslant a_n,\ |y|\leqslant b_n\}, \tag{3.5}$$

$$A_2=\{\mathbf{z};\ |H(\mathbf{z})|\leqslant\varepsilon_n\}-A_1, \tag{3.6}$$

$$\begin{aligned}A_3=&\{\mathbf{z};\ 0.5-\rho'\varepsilon_n<L_{\mathbf{z}}(\varepsilon_n)<0.5,\ |x|<\delta \text{ or } y>1\}\\ &\cup\{\mathbf{z};\ 0.5<L_{\mathbf{z}}(-\varepsilon_n)<0.5+\rho'\varepsilon_n,\ |x|<\delta \text{ or } y<-1\},\end{aligned} \tag{3.7}$$

and

$$A_4=\{\mathbf{z};\ |H(\mathbf{z})|>\varepsilon_n\}-A_3 \tag{3.8}$$

where ρ' is a positive constant (whose value will be chosen in Lemma 3.3), and $\delta>0$ is chosen so that $0.5<G(\delta)<0.75$. Then introduce

$$\xi=\frac{1}{n-1}\sum_{i=1}^{n}IF(\mathbf{Z}_i). \tag{3.9}$$

The approximations of $\hat{\alpha}_n$ are now defined as

$$\alpha_1=\operatorname*{med}_i\,(H(\mathbf{Z}_i)+\xi) \tag{3.10}$$

and

$$\alpha_2=\operatorname*{med}_i\,(I(\mathbf{Z}_i\notin A_1)\,\hat{H}(\mathbf{Z}_i)+I(\mathbf{Z}_i\in A_1)(H(\mathbf{Z}_i)+\xi)). \tag{3.11}$$

The basic idea of the proof is that taking the median of all $\hat{H}(\mathbf{Z}_i)$ is asymptotically equivalent to taking the median of all $H(\mathbf{Z}_i)+\xi$. When $\mathbf{Z}_i\in A_1$, $H(\mathbf{Z}_i)+\xi$ is close to $\hat{H}(\mathbf{Z}_i)$; when $\mathbf{Z}_i\in A_4$, both $H(\mathbf{Z}_i)+\xi$ and $\hat{H}(\mathbf{Z}_i)$ are too far away to interfere with the corresponding medians; and finally, the number of observations falling into A_2 or A_3 is asymptotically negligible.

We first show that α_1 has the same asymptotic behaviour as claimed for $\hat{\alpha}_n$.

LEMMA 3.1. *Let* α_1 *and* $IF(\cdot)$ *be as defined in* (3.10) *and* (2.9). *Then*

$$\alpha_1 = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} IF(\mathbf{Z}_i) + o_p(1).$$

Proof. Since $\alpha_1 = \xi + \operatorname{med}_i H(\mathbf{Z}_i)$ it suffices to show that

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

But (3.12) is a consequence of Lemma B.2, which corresponds to the fact that $l(0) = \infty$, with $l(\cdot) = L'(\cdot)$ the density of $H(\mathbf{Z})$. This makes the median of all $H(\mathbf{Z}_i)$ superefficient. The proof of (3.12) is the same as the proof of Lemma 3.1 in HRC. ∎

LEMMA 3.2. *Let* α_1 *and* α_2 *be as defined in* (3.10)–(3.11). *Then*

$$\alpha_2 - \alpha_1 = o_p(n^{-1/2}).$$

Proof. We expand the median $\hat{H}(\mathbf{Z}_i)$ defined in (2.6) in terms of influence functions,

$$\hat{H}(\mathbf{Z}_i) = H(\mathbf{Z}_i) + \frac{1}{n-1} \sum_{j,\, j \neq i} IF(\mathbf{Z}_i, \mathbf{Z}_j) + R_i \triangleq H(\mathbf{Z}_i) + S_i + R_i,$$

with

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

and $l_{\mathbf{z}}(\cdot) = L'_{\mathbf{z}}(\cdot)$. Then introduce the random variables

$$\bar{S} = \max_{\mathbf{Z}_i \in A_1} |S_i - \xi|, \tag{3.13}$$

$$\bar{R} = \max_{\mathbf{Z}_i \in A_1} |R_i|, \tag{3.14}$$

and the discrete random variable

$$N = |\{i; \mathbf{Z}_i \in A_2 \cup A_3\}|. \tag{3.15}$$

By the definition of α_1 and α_2 we then have

$$|\alpha_2 - \alpha_1| \leqslant \bar{S} + \bar{R} + \max(H_{([n/2]+1+N)} - H_{([n/2]+1)}, \\ H_{([n/2]+1)} - H_{([n/2]+1-N)}), \tag{3.16}$$

where $H_{(1)}, ..., H_{(n)}$ denote the ordered $H(\mathbf{Z}_i)$. It follows from (A.4) and (A.5) in Appendix A, as in (HRC, Lemma 3.2), that

$$\bar{R} = O_p\left(\left(\frac{\log n}{n}\right)^{(1+\eta)/2}\right), \tag{3.17}$$

when $0 < \eta < 0.5$ (which we can assume w.l.o.g.). We will also prove below that

$$\bar{S} = o_p(n^{-1/2}). \tag{3.18}$$

According to Lemma B.3–B.4 we have $N \sim \text{Bin}(n, p_n)$ where

$$p_n = K\{A_2\} + K\{A_3\} = o((\log n)^{3/4} n^{-1/2}),$$

and hence

$$N = o_p((\log n)^{3/4} n^{1/2}). \tag{3.19}$$

It follows from (3.19) and Lemma B.2, in the same way as in the proof of Lemma 3.5 in HRC, that

$$\max(H_{([n/2]+1+N)} - H_{([n/2]+1)}, H_{([n/2]+1)} - H_{([n/2]+1-N)}) = o_p(n^{-1/2}). \tag{3.20}$$

The lemma now follows from (3.16)–(3.18) and (3.20). In order to prove (3.18), we observe that

$$S_i - \xi = -\frac{IF(\mathbf{Z}_i)}{n-1} + \frac{1}{n-1} \sum_{j, j \neq i} (IF(\mathbf{Z}_i, \mathbf{Z}_j) - IF(\mathbf{Z}_j)).$$

Hence

$$\bar{S} \leqslant \frac{1}{2(n-1) f(0)} + \max_{1 \leqslant i \leqslant n} |\Delta(\mathbf{Z}_i)|, \tag{3.21}$$

where

$$\Delta(\mathbf{Z}_i) = \frac{I(\mathbf{Z}_i \in A_1)}{n-1} \sum_{j, j \neq i} (IF(\mathbf{Z}_i, \mathbf{Z}_j) - IF(\mathbf{Z}_j)) \triangleq \frac{I(\mathbf{Z}_i \in A_1)}{n-1} \sum_{j, j \neq i} Y_{ij}.$$

With $\mathbf{Z}_i$ fixed, $\Delta(\mathbf{Z}_i) \equiv 0$ when $\mathbf{Z}_i \notin A_1$, and if $\mathbf{Z}_i \in A_1$, all $Y_{ij}, j \neq i$ are i.i.d. with zero mean. Now suppose that $n \geqslant n_0$, where n_0 is the same number as in Lemma A.2. Then

$$P(|Y_{ij}| \leqslant M) = 1, \tag{3.22}$$

where

$$M=\frac{1}{2f(0)}+\frac{1}{2\inf_{\mathbf{z}\in A_1} l_{\mathbf{z}}(H(\mathbf{z}))}\leqslant\frac{1}{2f(0)}+\frac{1}{2\underline{l}}<\infty,$$

and the last inequality follows from (A.4). Now introduce the regions $B_{\mathbf{z}}=\{\mathbf{z}';h(\mathbf{z},\mathbf{z}')>H(\mathbf{z})\}$ and $B=\{\mathbf{z};y>0\}$. Then, if $\mathbf{Z}_i\in A_1$,

$$\begin{aligned}|Y_{ij}|\leqslant{}&\frac{1}{f(0)}I(\mathbf{Z}_j\in B_{\mathbf{Z}_i}\Delta B)\\&+\frac{1}{2}\left|\left(\frac{1}{l_{\mathbf{Z}_i}(H(\mathbf{Z}_i))}-\frac{1}{f(0)}\right)\operatorname{sgn}(h(\mathbf{Z}_i,\mathbf{Z}_j)-H(\mathbf{Z}_i))\right|,\end{aligned}$$

and hence

$$\begin{aligned}E(Y_{ij}^2\mid\mathbf{Z}_i)&\leqslant\frac{2}{f(0)^2}P_K(\mathbf{Z}\in B_{\mathbf{Z}_i}\Delta B)+\frac{1}{2}\left(\frac{1}{l_{\mathbf{Z}_i}(H(\mathbf{Z}_i))}-\frac{1}{f(0)}\right)^2\\&\leqslant\frac{2}{f(0)^2}P_K(\mathbf{Z}\in B_{\mathbf{Z}_i}\Delta B)+\frac{1}{2\underline{l}^4}(l_{\mathbf{Z}_i}(H(\mathbf{Z}_i))-f(0))^2.\qquad(3.23)\end{aligned}$$

In order to estimate $P_K(\mathbf{Z}\in B_{\mathbf{z}}\Delta B)$ for $\mathbf{z}\in A_1$, the symmetry allows us to assume that $\mathbf{z}=(x,y)$ lies in the first quadrant (Lemma A.1), so that $x\geqslant a_n$ and $0\leqslant y\leqslant b_n$. Let $x'\to q_{t,\mathbf{z}}(x')$ denote the line through $(0,t)$ and $\mathbf{z}$. We then have

$$\begin{aligned}B_{\mathbf{z}}\Delta B\subseteq{}&\{\mathbf{z}';x'\geqslant x\}\cup\{\mathbf{z}';x'\leqslant x,0\leqslant y'\leqslant q_{H(\mathbf{z}),\mathbf{z}}(x')\}\\&\cup\{\mathbf{z}';x'\leqslant x,q_{H(\mathbf{z}),\mathbf{z}}(x')\leqslant y'\leqslant 0\}\triangleq B_1\cup B_2\cup B_3.\qquad(3.24)\end{aligned}$$

First,

$$K\{B_1\}\leqslant 1-G(a_n)\leqslant e^{-(\log n)^\tau}.\qquad(3.25)$$

According to (B.3), $H(\mathbf{z})\leqslant y$ for any $\mathbf{z}$ in the first quadrant. This implies that the line $q_{H(\mathbf{z}),\mathbf{z}}$ has nonnegative slope. Hence,

$$K\{B_2\}\leqslant P_K(0\leqslant Y\leqslant b_n)\leqslant\|f\|_\infty e^{-(\log n)^\tau}.\qquad(3.26)$$

Since $H(\mathbf{z})>0$ according to Lemma A.1, the line $q_{H(\mathbf{z}),\mathbf{z}}$ has slope $\leqslant y/x\leqslant b_n/a_n\leqslant b_n$ for large enough n (if $M>1$), and the intersection of $q_{H(\mathbf{z}),\mathbf{z}}$ and the x-axis has negative x-coordinate. Therefore, for large enough n,

$$\begin{aligned}K\{B_3\}&\leqslant P_K(X\leqslant 0,\ -b_nX\leqslant Y\leqslant 0)\leqslant\int_{-\infty}^{0}\|f\|_\infty b_n|x'|\,dG(x')\\&\leqslant\|f\|_\infty E_G|X|\,b_n=O(e^{-(\log n)^\tau}).\qquad(3.27)\end{aligned}$$

Formulas (3.24)–(3.27) imply that

$$P_K(\mathbf{Z} \in B_{\mathbf{Z}_i} \Delta B) = O(e^{-(\log n)^\tau}) \tag{3.28}$$

for $\mathbf{Z}_i \in A_1$. Our next objective is to estimate the last term of (3.23). For ease of notation, put $q_{H(\mathbf{z}),\mathbf{z}}(\cdot) = q(\cdot)$. Then, according to (A.7),

$$\begin{aligned} l_{\mathbf{z}}(H(\mathbf{z})) - f(0) &= \int_{-\infty}^{\infty} f(q(x')) \frac{|x'-x|}{x} \, dG(x') - f(0) \\ &= f(0) \int_{-\infty}^{\infty} \left(\frac{|x'-x|}{x} - 1 \right) dG(x') \\ &\quad + \int_{-\infty}^{\infty} (f(q(x')) - f(0)) \frac{|x'-x|}{x} \, dG(x') \overset{\Delta}{=} I_1 + I_2. \end{aligned} \tag{3.29}$$

The symmetry of G implies that

$$I_1 = 2f(0) \int_x^{\infty} \left(\frac{x'}{x} - 1 \right) dG(x'),$$

and hence we obtain from (G) that for some positive constant $C > 0$,

$$\begin{aligned} |I_1| &\leqslant 2f(0) \frac{E_G(XI(X \geqslant x))}{x} \leqslant C(1 - G(x)) \\ &\leqslant C(1 - G(a_n)) = O(e^{-(\log n)^\tau}). \end{aligned} \tag{3.30}$$

Next, by the Lipschitz continuity of f,

$$\begin{aligned} |I_2| &\leqslant \|f\|_\eta \int_{-\infty}^{\infty} |q(x')|^\eta \left| \frac{x'}{x} - 1 \right| dG(x') \\ &= \|f\|_\eta \int_{-\infty}^{\infty} \left| H(\mathbf{z}) + \frac{(y - H(\mathbf{z}))\, x'}{x} \right|^\eta \left| \frac{x'}{x} - 1 \right| dG(x') \\ &\leqslant \|f\|_\eta \int_{-\infty}^{\infty} \left(|H(\mathbf{z})|^\eta + \left| \frac{yx'}{x} \right|^\eta \right) \left| \frac{x'}{x} - 1 \right| dG(x') \\ &\leqslant \|f\|_\eta \left(\varepsilon_n^\eta \left(\frac{E_G\,|X|}{x} + 1 \right) + \left(\frac{y}{x} \right)^\eta E_G \left(|X|^\eta \left| \frac{X}{x} - 1 \right| \right) \right) \\ &\leqslant O(\varepsilon_n^\eta + b_n^\eta) = O(e^{-\eta(\log n)^\tau}), \end{aligned} \tag{3.31}$$

where the second to last inequality follows since $E_G\,|X|^{1+\eta} < \infty$ and $x \geqslant a_n$ is lower bounded away from 0. Since $0 < \eta < 0.5$ can be assumed w.l.o.g., it now follows from (3.29)–(3.31) that the last term in (3.23) can be bounded from above according to

$$\sup_{\mathbf{z}\in A_1} \frac{1}{2\underline{l}^4} (l_{\mathbf{z}}(H(\mathbf{z}))-f(0))^2 = O(e^{\ \ 2\eta(\log n)^\tau}). \tag{3.32}$$

Collecting the results, we see from (3.23), (3.28), and (3.32) that

$$\sup_{\mathbf{Z}_i\in A_1} E(Y_{ij}^2 \mid \mathbf{Z}_i) \triangleq \delta_n^2 = O(e^{-2\eta(\log n)^\tau}), \tag{3.33}$$

where we choose $\delta_n > 0$ in (3.33). As in the proof of (HRC, Lemma 3.2), we may now use (3.33) and an exponential inequality for sums of i.i.d. random variables due to Bernstein to show that

$$\bar{S} = O_p\left(\frac{\delta_n \log n}{n^{1/2}}\right) = o_p(n^{-1/2}).$$

This implies (3.18) and completes the proof of the lemma. ▌

LEMMA 3.3. *Let $\hat{\alpha}_n$ and α_2 be as defined in* (1.2) *and* (3.11). *Then*

$$\hat{\alpha}_n - \alpha_2 = o_p(n^{-1/2}). \tag{3.34}$$

Proof. Let $0<\rho<1$ and subdivide A_4 into A_4^+ and A_4^- according to whether $H(\mathbf{z}) > \varepsilon_n$ or $H(\mathbf{z}) < -\varepsilon_n$. Then

$$\begin{aligned} P(\alpha_2 \neq \hat{\alpha}_n) \leqslant{} & P(|\alpha_2| \geqslant (1-\rho)\,\varepsilon_n) + P(|\xi| \geqslant \rho\varepsilon_n) \\ & + P\left(\min_{\mathbf{Z}_i\in A_4^+} \hat{H}(\mathbf{Z}_i) \leqslant (1-\rho)\,\varepsilon_n\right) \\ & + P\left(\max_{\mathbf{Z}_i\in A_4^-} \hat{H}(\mathbf{Z}_i) \geqslant -(1-\rho)\,\varepsilon_n\right). \end{aligned} \tag{3.35}$$

We know from Lemmas 3.1, 3.2 that $\alpha_2 = O_p(n^{-1/2})$, and by the definition of ξ we also have $\xi = O_p(n^{-1/2})$. Hence, the first two terms on the RHS of (3.35) tend to zero as $n \to \infty$. Since the last two terms are treated in the same way, we confine ourselves to study the third one. Following the proof of Lemma 3.6 in HRC, it follows that the third term goes to zero, if we show that

$$\inf_{\mathbf{z}\in A_4^+} L_{\mathbf{z}}^{-1}(0.5-\rho'\varepsilon_n) \geqslant (1-\rho)\,\varepsilon_n. \tag{3.36}$$

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

$$L_{\mathbf{z}}(\varepsilon_n) \leqslant 0.5 - \rho'\varepsilon_n \tag{3.37}$$

or

$$|x| \geqslant \delta \qquad \text{and} \qquad 0 \leqslant y \leqslant 1. \tag{3.38}$$

If (3.37) holds, $L_{\mathbf{z}}^{-1}(0.5 - \rho'\varepsilon_n) \geqslant \varepsilon_n > (1-\rho)\,\varepsilon_n$, so it remains to consider those $\mathbf{z} \in A_4^+$ for which (3.38) holds. For any such $\mathbf{z}$ with $x > 0$ (the case $x < 0$ is treated similarly) we have

$$\begin{aligned} L_{\mathbf{z}}(\varepsilon_n) - L_{\mathbf{z}}((1-\rho)\,\varepsilon_n) &\geqslant \rho\varepsilon_n \lim_{\substack{(1-\rho)\varepsilon_n \\ \leqslant t \leqslant \varepsilon_n}} l_{\mathbf{z}}(t) \geqslant \rho\varepsilon_n \underline{f} \int_{-x}^{0} \frac{|x'-x|}{x}\, dG(x') \\ &\geqslant \rho \underline{f}(G(\delta) - 0.5)\,\varepsilon_n, \end{aligned} \tag{3.39}$$

where $\underline{f}$ is a lower bound for f on $[-2, 2]$. The second to last inequality in (3.39) follows from (A.7) and the fact that the line $q_{H(\mathbf{z}),\mathbf{z}}(\cdot)$ through $\mathbf{z}$ and $(0, H(\mathbf{z}))$ satisfies $|q_{H(\mathbf{z}),\mathbf{z}}(x')| \in [-2, 2]$ when $x' \in [-x, 0]$. If we now choose ρ' so that $0 < \rho' < \underline{f}(G(\delta) - 0.5)\rho$, it follows from (3.39) that

$$L_{\mathbf{z}}((1-\rho)\,\varepsilon_n) \leqslant 0.5 - \rho'\varepsilon_n \tag{3.40}$$

since $L(\varepsilon_n) \leqslant 0.5$. This proves (3.36), and the remainder of the proof is completely analogous to the proof of Lemma 3.6 in HRC. ▌

Theorem 3.1 now follows from Lemmas 3.1–3.3, Slutsky's Lemma, and the Central Limit Theorem.

4. Empirical Results

4.1. *Sensitivity Functions*

The influence function of the RM intercept is given by formula (2.9). Without loss of generality we have assumed that $\alpha = 0 = \beta$, meaning that the uncontaminated data are i.i.d. according to $K = G \times F$, where G is called the carrier distribution and F is the error distribution. If both G and F equal the standard Gaussian distribution Φ, we obtain

$$\mathrm{IF}(x, y) = \sqrt{\frac{\pi}{2}}\, \mathrm{sgn}(y) \tag{4.1}$$

which is plotted in Fig. 1. The IF is a step function which takes on only two values, and is therefore bounded.

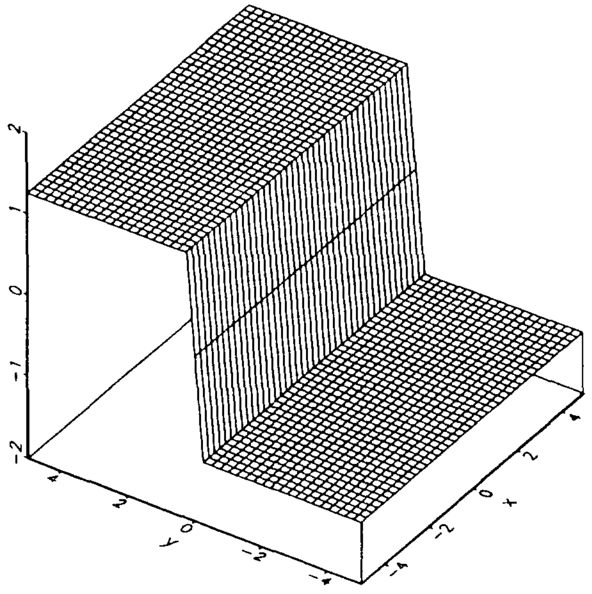

FIG. 1. Influence function $\mathrm{IF}(x, y)$ of the repeated median intercept.

For estimators operating on univariate data, Tukey (1970) proposed the *sensitivity curve* as a finite-sample approximation to the IF. However, the contamination now occurs in a point (x, y) with two coordinates. Therefore, Tukey's definition needs to be extended. Let us start with a sample $\{\mathbf{Z}_1, ..., \mathbf{Z}_n\}$ of the bivariate distribution $K = \Phi \times \Phi$. Then the sensitivity function is defined in each $\mathbf{z} = (x, y)$ by

$$\mathrm{SF}_n(\mathbf{z}; \hat{\alpha}, \mathbf{Z}) = n(\hat{\alpha}_{n+1}(\mathbf{Z}_1, ..., \mathbf{Z}_n, \mathbf{z}) - \hat{\alpha}_n(\mathbf{Z}_1, ..., \mathbf{Z}_n)). \tag{4.2}$$

Note that (4.2) depends not only on $\mathbf{z}$ but also on the intercept estimator $\hat{\alpha}$ as well as on the original sample $\mathbf{Z} = \{\mathbf{Z}_1, ..., \mathbf{Z}_n\}$. Also note that $\mathbf{z}$ is not restricted to the observations, but that it can be any point.

However, plotting such sensitivity functions was quite disappointing, because they were very "jumpy." This is because SF_n is based on a finite sample $\{\mathbf{Z}_1, ..., \mathbf{Z}_n\}$ from $K = \Phi \times \Phi$, which often provides a poor approximation to K. Such a random sample may be quite asymmetric, the median of the X_i may be nonzero, etc. One way to repair these "wiggles" in SF_n is to average the sensitivity over many samples, as proposed by Rousseeuw and Leroy (1987, p. 193). This yields the *averaged sensitivity function* given by

$$\mathrm{ASF}_n(x, y) = \underset{j=1,...,m}{\mathrm{average}}\ \mathrm{SF}_n(x, y; \hat{\alpha}, \mathbf{Z}^{(j)}), \tag{4.3}$$

where the $\mathbf{Z}^{(j)}$ (for $j = 1, \ldots, m$) are i.i.d. samples generated from K. The vertical size of the irregularities in the ASF will decrease roughly as $1/\sqrt{m}$ when m is increased, forcing a tradeoff between smoothness and computation time. (To improve the alignment of the SF_n in (4.3), one can prestandardize each $\mathbf{Z}^{(j)}$ by replacing all X_i by $X_i - \mathrm{med}_k\, X_k$ and all Y_i by $Y_i - \mathrm{med}_k\, Y_k$.)

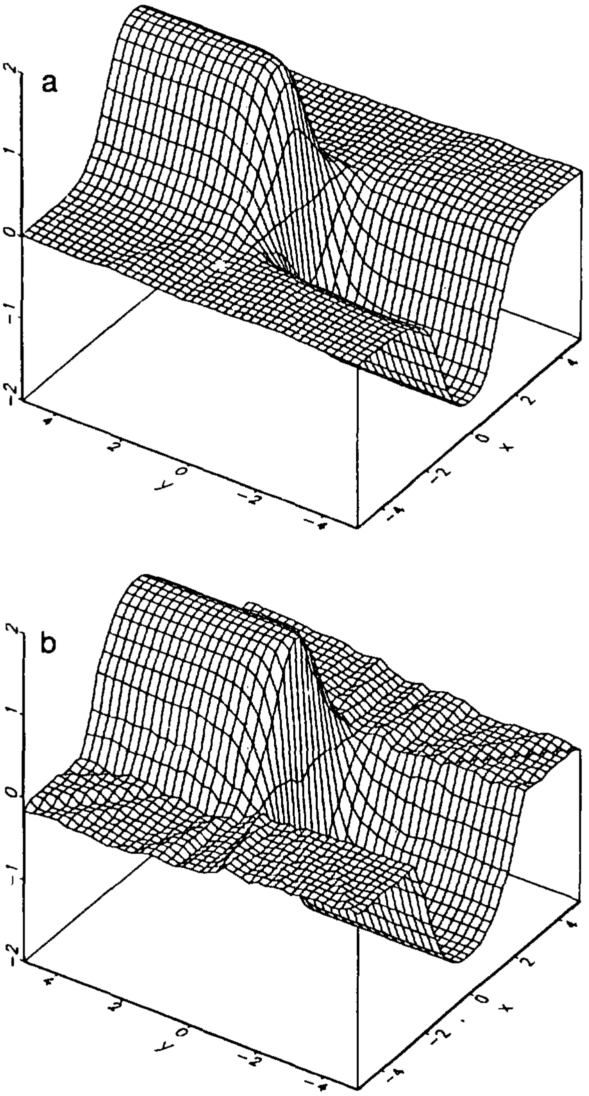

FIG. 2. (a) Averaged sensitivity function $\mathrm{ASF}_n(x, y)$ of RM intercept for $n = 20$ and $m = 1000$; (b) $\mathrm{ASF}_n(x, y)$ for $n = 100$ and $m = 200$.

Figure 2 shows two ASF_n surfaces, one for $n = 20$ and one for $n = 100$. At first sight, they look quite different from the asymptotic IF in Fig. 1. Still, they have positive values for $y > 0$, negative values for $y < 0$, and their upper and lower bounds are of the same order of magnitude as the IF. The main distinction is that the ASF_n depend on x, whereas the IF does not. For small $|x|$, we note that $\mathrm{ASF}_n(x, y) \approx \mathrm{IF}(x, y)$. For larger $|x|$, the value of $\mathrm{ASF}_n(x, y)$ is closer to zero because the relative effect of (x, y) on the intercept becomes smaller. We see that the ASF_n is slowly stretched out along the x-axis when n increases. This is how $\mathrm{ASF}_n(x, y)$ tends to $\mathrm{IF}(x, y)$ when $n \to \infty$.

4.2. *Monte Carlo Variances*

Theorem 3.1 confirms that the asymptotic variance of the RM intercept estimator is given by the expected square of its IF. Therefore, when both G and F are standard gaussian we obtain the asymptotic variance $\pi/2 \approx 1.571$ and the corresponding efficiency $2/\pi \approx 63.7\%$.

In order to check whether this asymptotic variance provides a good approximation to the variability at finite samples, we carried out a Monte Carlo experiment. For each n in Table I we generated $m = 1000$ samples of size n, and computed the corresponding intercepts $\hat{\alpha}_n^{(k)}$ for $k = 1, ..., m$. Table I lists the bias given by

$$\operatorname*{average}_{k=1,\ldots,m} \hat{\alpha}_n^{(k)} \tag{4.4}$$

TABLE I

Simulation Results of Two Intercept Estimators, Applied to Bivariate Gaussian Data

	Repeated median intercept		Hierarchical intercept	
n	Bias	n-Fold variance	Bias	n-Fold variance
10	−0.003	2.108	−0.008	1.788
20	0.005	1.783	0.003	1.704
40	−0.006	1.564	−0.006	1.523
60	0.003	1.590	−0.001	1.607
80	0.002	1.542	0.001	1.539
100	0.002	1.616	0.002	1.636
200	−0.001	1.537	−0.001	1.591
300	0.001	1.601	0.001	1.625
500	0.000	1.515	0.000	1.584
1000	0.000	1.544	−0.001	1.610
∞	0.000	1.571	0.000	1.571

as well as the n-fold variance

$$n \underset{k=1,\dots,m}{\text{variance}} \hat{\alpha}_n^{(k)} \tag{4.5}$$

which should converge (as n tends to ∞) to 1.571. (The second part of the table concerns another estimator, which will be discussed in Section 4.4.)

The Gaussian variables in the simulation were obtained by means of the Box–Muller transform and the congruential generator of Cheney and Kincaid (1985, p. 335). The calculations were carried out on a workstation running under Unix. The n-fold variances in the table have a standard error of approximately 0.025. We also made Gaussian Q–Q plots of the set $\{\hat{\alpha}_n^{(k)},\ k=1, \dots, m\}$ of estimated intercepts, to confirm that the sampling distribution of $\hat{\alpha}_n^{(k)}$ is approximately Gaussian.

In Table I we see that the RM intercept is indeed unbiased. Also, the asymptotics provide a good approximation to the Monte Carlo variances for $n \geqslant 40$. (For $n=10$ and $n=20$ the RM intercept is somewhat less efficient.) At first glance, the fast convergence of the n-fold variance to the asymptotic variance appears to be at odds with the slow convergence of the ASF_n to the IF, because the asymptotic variance equals $E_K[\mathrm{IF}^2]$. However, the difference between $\mathrm{ASF}_n(x, y)$ and $\mathrm{IF}(x, y)$ is at its smallest for (x, y) close to $(0, 0)$, which corresponds to the region where $K = \Phi \times \Phi$ has most of its mass. This is exactly the opposite of the situation for the RM *slope*, whose ASF_n (see Rousseeuw *et al.*, 1995) most resembles its IF at points (x, y) far away from $(0, 0)$, in regions where K has little mass. Therefore the n-fold variance of the RM slope converges much more slowly to its asymptotic variance.

4.3. *The Function H*

The function H defined in Section 2 plays an essential role throughout the paper. For each point $\mathbf{z} = (x, y)$ the real value $H(\mathbf{z})$ is given by (2.2). Although H is a deterministic function, no simple expression is available. We can approximate $H(\mathbf{z})$ by

$$H_m(\mathbf{z}) = \underset{k=1,\dots,m}{\text{med}}\ h(\mathbf{z}, \mathbf{Z}^{(k)}) \tag{4.6}$$

where the points $\mathbf{Z}^{(k)}$ (for $k=1, \dots, m$) are generated according to K. We used $m=6000$ to obtain good accuracy.

Fig. 3 gives an idea of the shape of H around the origin. By definition, H satisfies the symmetry properties $H(x, y) = H(-x, y) = -H(x, -y) = -H(-x, -y)$ and it takes on positive values for $y > 0$ and negative values for $y < 0$. Note that H is an unbounded function, which attains large values when $|x|$ is small and $|y|$ is large. From the definition of H it immediately follows that $H(x, 0) = 0$ for all x, and that $H(0, y) = y$ for all y. Figure 3b displays several contours, given by $H(x, y) = t$ for some positive and some negative values of t. When t tends to zero, the contour becomes the x-axis.

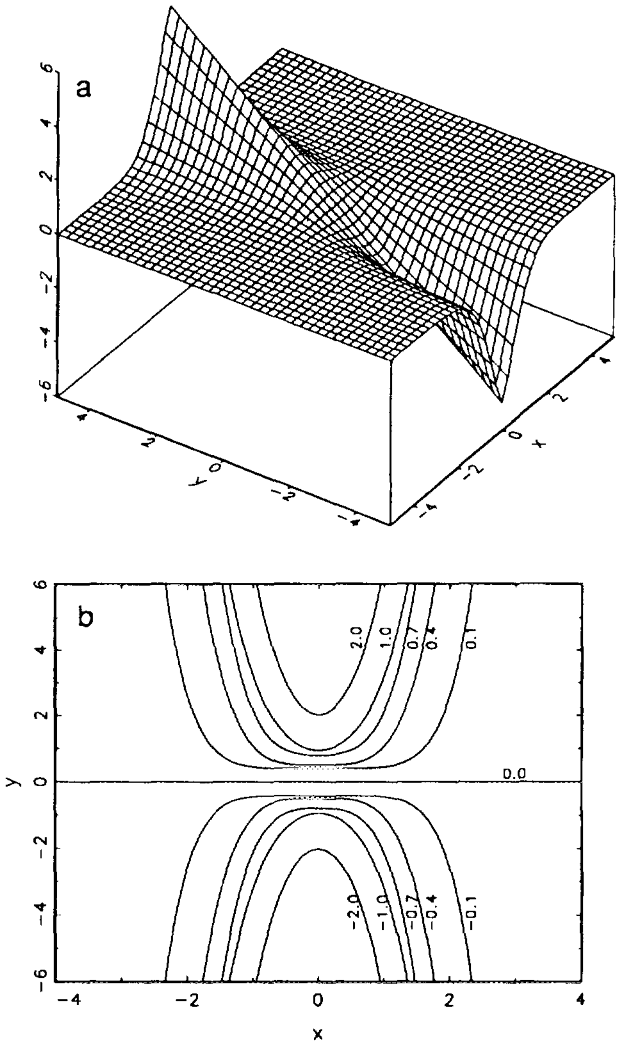

FIG. 3. (a) Plot of $H(x, y)$; (b) contour curves of $H(x, y)$; (c) function H_ε for $\varepsilon = 1$; (d) empirical distribution function of H, based on 2500 points.

It is interesting to note the similarity between the function H in Fig. 3a and sensitivity functions of the intercept estimator. If we consider truncated (and standardized) versions of H given by

$$H_\varepsilon(x, y) = \frac{1}{\varepsilon} \min\{\varepsilon, \max\{H(x, y), -\varepsilon\}\}$$

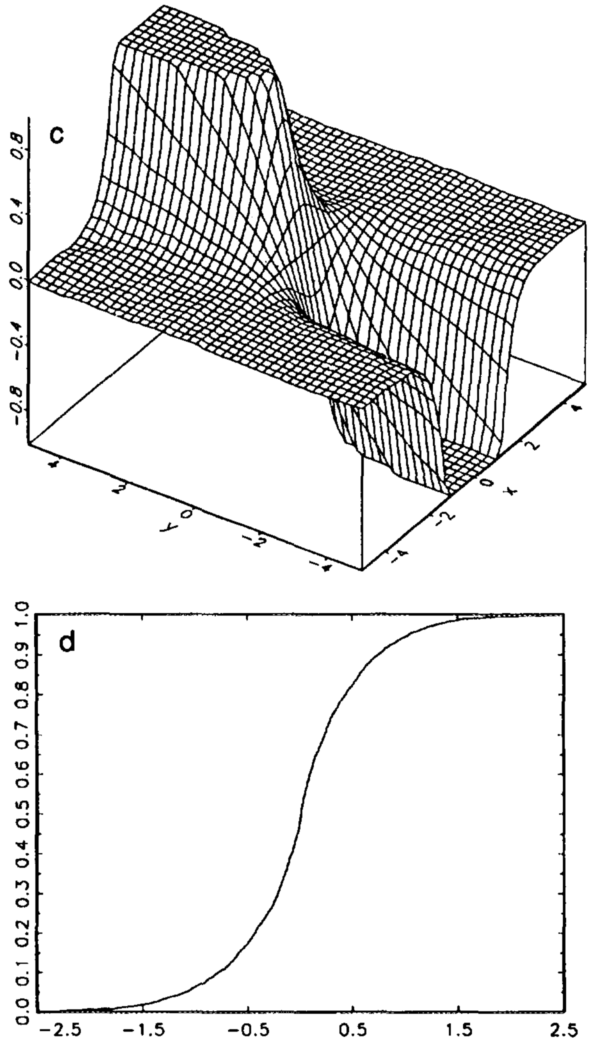

FIG. 3. *Continued.*

for any $\varepsilon > 0$, the resulting surface looks like H near the x-axis, but is cut at the contours of levels ε and $-\varepsilon$. Figure 3c shows the function H_ε with $\varepsilon = 1$, which looks quite similar to the averaged sensitivity functions in Fig. 2. In particular, note the steep part which occurs at the origin along the y-axis. The ASF_n with large n correspond to functions H_ε with small ε, which converge (for ε tending to zero) to the discontinuous function $\operatorname{sign}(y)$ in the expression of $\mathrm{IF}(x, y)$. For more details, see Appendix B.

We can also study the distribution of $H(\mathbf{Z})$ for $\mathbf{Z} \sim K$. For this we generated 2500 observations $\mathbf{Z}_j$ according to $K = \Phi \times \Phi$ and computed $H(\mathbf{Z}_j)$ for each of them, yielding $\{(\mathbf{Z}_j, H(\mathbf{Z}_j)); j = 1, ..., 2500\}$. From these values we can construct the empirical cdf of H shown in Fig. 3d. It provides an approximation to the theoretical cdf (2.3) denoted by L. From Lemma B.2 it follows that L has a vertical tangent in zero.

4.4. *The Hierarchical Intercept*

Apart from the repeated median intercept (1.2) we have studied until now, Siegel (1982) also proposed another estimator of the intercept. For this we start by computing the RM slope

$$\hat{\beta}_n = \operatorname*{med}_i \operatorname*{med}_{j, j \neq i} \frac{Y_j - Y_i}{X_j - X_i} \tag{4.7}$$

and then we estimate α by

$$\tilde{\alpha}_n = \operatorname*{med}_i (Y_i - \hat{\beta}_n X_i) \tag{4.8}$$

which ensures that the median of the final residuals will be zero. Because this is a two-stage procedure, we call $\tilde{\alpha}_n$ the *hierarchical intercept*. An obvious advantage of $\tilde{\alpha}_n$ is its computational economy: assuming that we already have $\hat{\beta}_n$, it can be computed in just $O(n)$ time. If we use the $O(n \log n)$-time algorithm for $\hat{\beta}_n$ proposed by Mount and Netanyahu (1991), then the additional computation of $\tilde{\alpha}_n$ does not increase the order of complexity.

The hierarchical intercept has the same influence function (4.1) as the RM intercept, and also the same asymptotic variance. This follows from results of Jurečková (1971), where it is proved that it is asymptotically equivalent to compute the sign test statistic either from $Y_i - \hat{\beta}_n X_i$ or from $Y_i - \beta X_i$ when $\hat{\beta}_n$ is a consistent estimator of the true parameter β. Therefore, the asymptotic behavior of $\tilde{\alpha}_n$ is that of the univariate median applied to observations drawn from $\tilde{F}(v) = F(v - \alpha)$.

Figure 4 shows averaged sensitivity functions of $\tilde{\alpha}_n$ for two values of n. We see immediately that they are much closer to $\mathrm{IF}(x, y)$ than the ASF_n of $\hat{\alpha}_n$ are. The main difference is that the ASF_n of $\tilde{\alpha}_n$ do not go to zero for increasing $|x|$, provided $|y|/|x|$ is large enough. This is because contamination at (x, y) can only have a bounded effect on the RM slope, which has a bounded IF as well. Outside a wedge-shaped region $|y| \leqslant c_n |x|$, the contamination (x, y) therefore has its maximal influence on $\tilde{\alpha}_n$, which is a positive or negative constant because of the median in (4.8). Inside the region $|y| \leqslant c_n |x|$ the effect of (x, y) on the initial slope $\hat{\beta}_n$ causes the gradual transition between the positive and the negative constant in

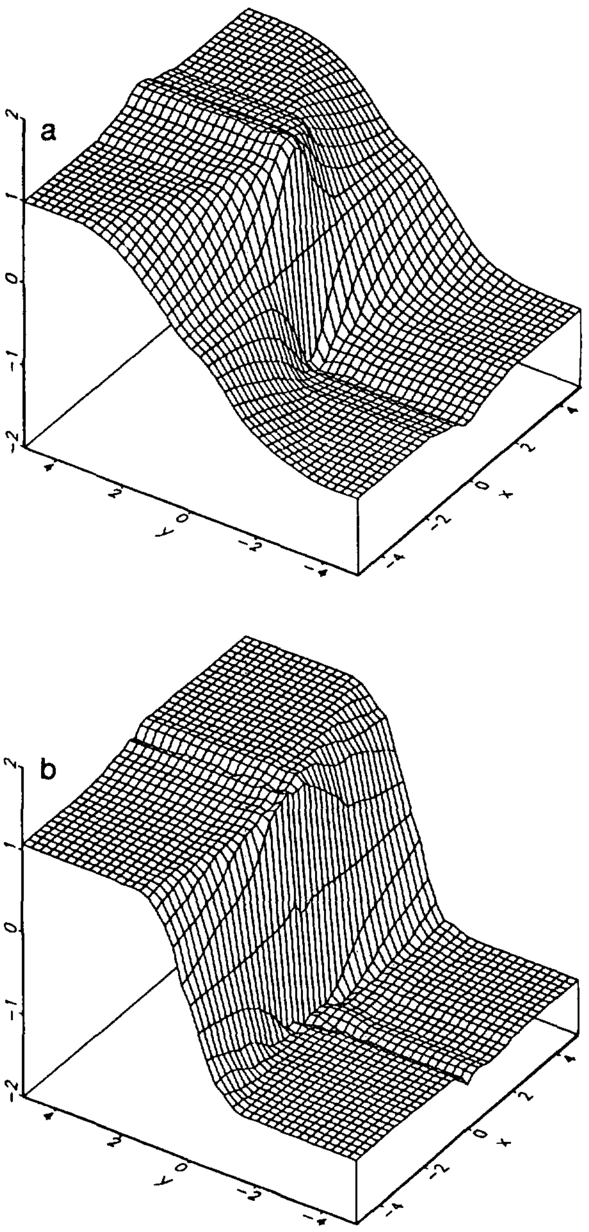

FIG. 4. (a) Averaged sensitivity function of hierarchical intercept for $n = 20$ and $m = 1000$; (b) $\mathrm{ASF}_n(x, y)$ for $n = 100$ and $m = 200$.

the ASF_n of $\tilde{\alpha}_n$. When n increases, the wedge becomes more narrow (that is, $c_n \to 0$) and the ASF_n converges to the IF.

In Table I we see that the n-fold Monte Carlo variances of $\tilde{\alpha}_n$ are quite stable, and that they are rather close to the asymptotic variance 1.571. This is in accordance with the relatively fast convergence of the ASF_n to the IF, at least in the central region around $(0, 0)$ where $K = \Phi \times \Phi$ has nearly all its mass, which explains why the Monte Carlo variances converge almost as fast as in the case of a truly univariate median.

In the same simulation, we also constructed Q–Q plots of $\tilde{\alpha}_n$ to confirm that its distribution is approximately Gaussian. We also computed empirical correlations between the three estimators $\hat{\alpha}_n$, $\hat{\beta}_n$, and $\tilde{\alpha}_n$. As was to be expected, the correlation between the intercept estimators $\hat{\alpha}_n$ and $\tilde{\alpha}_n$ becomes very large, whereas the correlations between each of the intercept estimators and $\hat{\beta}_n$ tend to zero.

Appendix A

Lemma A.1. *The function $H(\mathbf{z})$ defined in* (2.2) (*see Fig.* 3a) *satisfies*

$$H(x, y) = H(-x, y) = -H(x, -y), \tag{A.1}$$

and the distribution corresponding to $L(\cdot)$ defined in (2.3) *is symmetric.*

Proof. Let $A_x(x, y) = (x, -y)$ denote the operator that mirrors points through the x-axis and $A_y(x, y) = (-x, y)$ the corresponding operator for the y-axis. Then, by the symmetry of G and the relation $h(A_y(\mathbf{z}_1), A_y(\mathbf{z}_2)) = h(\mathbf{z}_1, \mathbf{z}_2)$, it follows that

$$H(A_y(\mathbf{z})) = H(\mathbf{z}). \tag{A.2}$$

Similarly, the symmetry of F and $h(A_x(\mathbf{z}_1), A_x(\mathbf{z}_2)) = -h(\mathbf{z}_1, \mathbf{z}_2)$ imply that

$$H(A_x(\mathbf{z})) = -H(\mathbf{z}). \tag{A.3}$$

This proves (A.1). The symmetry of L follows from (A.3) and the symmetry of F. ∎

Lemma A.2. *Define the interval $I_\varepsilon = [0.5 - \varepsilon, 0.5 + \varepsilon]$. Then, for some $0 < \varepsilon < 0.5$, there exists a positive integer n_0, and numbers $\underline{l}$ and $\bar{L}$, such that*

$$\inf_{u \in I_\varepsilon} l_{\mathbf{z}}(L_{\mathbf{z}}^{-1}(u)) \geq \underline{l} > 0 \tag{A.4}$$

and

$$\sup_{t_1 \neq t_2} \frac{|l_{\mathbf{z}}(t_1) - l_{\mathbf{z}}(t_2)|}{|t_1 - t_2|^{\eta}} \leqslant \bar{L} < \infty \tag{A.5}$$

hold for all $\mathbf{z} \in A_1$ *whenever* $n \geqslant n_0$. *Here* A_1 *is defined in* (3.5), $L_{\mathbf{z}}$ *in* (2.1), *and* $l_{\mathbf{z}} = L'_{\mathbf{z}}$.

Proof. Since the set A_1 decreases with n, we only have to establish the lemma for $n = n_0$ (which will be chosen below). For any $\mathbf{z} = (x, y)$, $x \neq 0$, let

$$q_{t,\mathbf{z}}(x') = t + \frac{y-t}{x} x' = y + \frac{y-t}{x}(x' - x)$$

describe the line through $(0, t)$ and $\mathbf{z}$ as x' varies. Then

$$L_{\mathbf{z}}(t) = P_K(Y \leqslant q_{t,\mathbf{z}}(X), X < x) + P_K(Y > q_{t,\mathbf{z}}(X), X > x). \tag{A.6}$$

Differentiating (A.6) w.r.t. t yields

$$l_{\mathbf{z}}(t) = \int_{-\infty}^{\infty} f(q_{t,\mathbf{z}}(x')) \frac{|x' - x|}{|x|} \, dG(x'). \tag{A.7}$$

Since G is continuous, we can find an $a > 0$ and $0 < \zeta < 1$ such that $G(a) < 1$ and $G(\zeta a) > 0.5$. Now choose n_0 so large that $A_1 \subseteq \{\mathbf{z};\, |x| \geqslant a, |y| \leqslant 1\}$. Consider a fixed $\mathbf{z} \in A_1$. By symmetry (cf. Lemma A.1) we may also assume that $\mathbf{z}$ lies in the first quadrant, so that $x \geqslant a$ and $0 \leqslant y \leqslant 1$. We first show that for any $u \in I_{\varepsilon}$ and $t = L_{\mathbf{z}}^{-1}(u)$,

$$|q_{t,\mathbf{z}}(x')| \leqslant 2 \qquad \text{whenever} \quad 0 \leqslant x' \leqslant a. \tag{A.8}$$

Let

$$\underline{f} = \inf_{y;\, |y| \leqslant 2} f(y) > 0. \tag{A.9}$$

Together with (A.7), (A.8) will imply that

$$\begin{aligned} l_{\mathbf{z}}(t) &\geqslant \underline{f} \int_0^a \frac{|x' - x|}{|x|} \, dG(x') \geqslant \underline{f} \int_0^{\zeta a} \frac{|(1-\zeta)x|}{|x|} \, dG(x') \\ &= (1-\zeta)\underline{f}(G(\zeta a) - 0.5) \triangleq \underline{l} > 0, \end{aligned} \tag{A.10}$$

which proves (A.4). Actually, (A.8) will follow if we show that

$$L_{\mathbf{z}}^{-1}(0.5 + \varepsilon) \leqslant 2 \tag{A.11}$$

and

$$L_{\mathbf{z}}^{-1}(0.5-\varepsilon) \geqslant -2. \tag{A.12}$$

But from (A.7),

$$\begin{aligned}
L_{\mathbf{z}}(2) &\geqslant (L_{\mathbf{z}}(2) - L_{\mathbf{z}}(1)) + L_{\mathbf{z}}(y) \\
&\geqslant \underline{f} \int_0^{\zeta a} (q_{2,\mathbf{z}}(x') - q_{1,\mathbf{z}}(x'))\, dG(x') \\
&\quad + F(x)\, G(y) + (1 - F(x))(1 - G(y)) \geqslant \underline{f}(1-\zeta)(G(\zeta a) - 0.5) + 0.5.
\end{aligned}$$

Hence, with $\varepsilon < \underline{f}(1-\zeta)(G(\zeta a) - 0.5)$, (A.11) follows. Inequality (A.12) is proved in a way similar to that for (A.11), and this completes the proof of (A.4).

The relation (A.5) is a consequence of (A.7) and the Lipschitz continuity of f,

$$\begin{aligned}
|l_{\mathbf{z}}(t_2) - l_{\mathbf{z}}(t_1)| &\leqslant \frac{1}{|x|} \int |f(q_{t_2,\mathbf{z}}(x')) - f(q_{t_1,\mathbf{z}}(x'))|\, |x' - x|\, dG(x') \\
&\leqslant \frac{\|f\|_\eta}{|x|} \int |q_{t_2,\mathbf{z}}(x') - q_{t_1,\mathbf{z}}(x')|^\eta\, |x' - x|\, dG(x') \\
&= \frac{\|f\|_\eta\, E_G\, |X - x|^{1+\eta}}{|x|^{1+\eta}}\, |t_2 - t_1|^\eta \leqslant \bar{L}\, |t_2 - t_1|^\eta,
\end{aligned}$$

where the last inequality holds uniformly in $\mathbf{z}$ whenever $x \geqslant a$, for some finite $\bar{L}$ which depends on η. ▌

Appendix B

In this appendix we will consider the region

$$A_\varepsilon = \{\mathbf{z};\, |H(\mathbf{z})| \leqslant \varepsilon\} \tag{B.1}$$

shown in Fig. 3b. We start with the following crucial lemma.

Lemma B.1. *There exist positive constants $C_1 < C_2$ such that for small enough ε,*

$$\begin{aligned}
\left[-\frac{C_1 \varepsilon}{1 - G(|x|)}, \frac{C_1 \varepsilon}{1 - G(|x|)} \right] &\subseteq \{y;\, |y| \leqslant 1,\, (x, y) \in A_\varepsilon\} \\
&\subseteq \left[-\frac{C_2 \varepsilon}{1 - G(|x|)}, \frac{C_2 \varepsilon}{1 - G(|x|)} \right].
\end{aligned} \tag{B.2}$$

Proof. In view of Lemma A.1 we only consider the first quadrant, that is, $x \geqslant 0$, $0 \leqslant y \leqslant 1$. Let δ be a number such that $0.5 < G(\delta) < 0.75$. We will treat the cases $0 < x < \delta$ and $x \geqslant \delta$ separately. In the former case, it suffices to show that

$$[0, C_1'\varepsilon] \subseteq \{y; 0 \leqslant y \leqslant 1, (x, y) \in A_\varepsilon\} \subseteq [0, C_2'\varepsilon] \tag{B.3}$$

for some positive constants C_1' and C_2', since the variation of $\log(1/(1-G(|x|)))$ on $[0, \delta]$ can be absorbed into the constants C_1 and C_2 in (B.3). Since f is strictly positive on $[-1, 1]$, it follows that for any $t > y$

$$\begin{aligned} L_\mathbf{z}(t) > L_\mathbf{z}(y) &= G(x)\, F(y) + (1 - G(x))(1 - F(y)) \\ &= 0.5 + 2(G(x) - 0.5)(F(y) - 0.5) \geqslant 0.5, \end{aligned}$$

and this implies that $H(x, y) \leqslant y$. Hence, we may choose $C_1' = 1$ in (B.3). As for C_2', we intend to show that for large enough $D > 0$, and small enough $\varepsilon > 0$,

$$P_K(h((x, D\varepsilon), \mathbf{Z}) \leqslant \varepsilon) < 0.5 \qquad \text{when} \quad 0 < x < \delta, \tag{B.4}$$

so that we can put $C_2' = D$ in (B.3). Let B_i, $i = 1, ..., 4$, be the intersection of the ith quadrant and $\{\mathbf{z}'; h(\mathbf{z}, \mathbf{z}') \leqslant \varepsilon\}$. Then (B.4) follows for large ε and D, since $P_K(B_1) \leqslant 0.25$, $P_K(B_2) \leqslant F(\varepsilon) - 0.5$, and

$$P_K(B_3 \cup B_4) \rightarrow G(x) - 0.5 \leqslant G(\delta) - 0.5 < 0.25 \qquad \text{as} \quad D \rightarrow \infty.$$

We now treat the case $x \geqslant \delta$. According to (A.6) and the symmetry of F and G,

$$\begin{aligned} L_\mathbf{z}(0) &= P_K\left(\frac{Y}{X} \leqslant \frac{y}{x}, X < x\right) + P_K\left(\frac{Y}{X} \geqslant \frac{y}{x}, X > x\right) \\ &= P_K\left(\frac{Y}{X} < \frac{y}{x}\right) - P_K\left(-\frac{y}{x} \leqslant \frac{Y}{X} \leqslant \frac{y}{x}, X > x\right) \\ &= 0.5 - 2P_K(\mathbf{Z} \in \Omega_\mathbf{z}), \end{aligned} \tag{B.5}$$

where

$$\Omega_\mathbf{z} = \left\{\mathbf{z}' = (x', y'); 0 \leqslant \frac{y'}{x'} \leqslant \frac{y}{x}, x' > x\right\}.$$

Consequently, for any $\mathbf{z}$ in the first quadrant,

$$\mathbf{z} \in A_\varepsilon \Leftrightarrow L_\mathbf{z}(\varepsilon) \geqslant 0.5 \Leftrightarrow L_\mathbf{z}(\varepsilon) - L_\mathbf{z}(0) \geqslant 2P_K(\mathbf{Z} \in \Omega_\mathbf{z}). \tag{B.6}$$

In order to apply (B.6) we need upper and lower bounds of $L_{\mathbf{z}}(\varepsilon)-L_{\mathbf{z}}(0)$ and $P_K(\mathbf{Z}\in\Omega_{\mathbf{z}})$, respectively. We obtain from (A.6) that

$$\begin{aligned} L_{\mathbf{z}}(\varepsilon)-L_{\mathbf{z}}(0) &= P_K(q_{0,\mathbf{z}}(X)\leqslant Y\leqslant q_{\varepsilon,\mathbf{z}}(X),\, X<x) \\ &\quad + P_K(q_{\varepsilon,\mathbf{z}}(X)\leqslant Y\leqslant q_{0,\mathbf{z}}(X),\, X>x) \\ &\leqslant \frac{\varepsilon\,\|f\|_\infty\, E_G\,|X-x|}{x}, \end{aligned} \tag{B.7}$$

and moreover (remember $0\leqslant y\leqslant 1$),

$$\begin{aligned} \underline{f}(1-G(x))y &\leqslant P_K(\mathbf{Z}\in\Omega_{\mathbf{z}}) \\ &\leqslant \|f\|_\infty(1-G(x))y+\frac{\|f\|_\infty\, E_G(X-x)_+}{x}\,y, \end{aligned} \tag{B.8}$$

with $\underline{f}$ defined in (A.9). It remains to give a lower bound for $L_{\mathbf{z}}(\varepsilon)-L_{\mathbf{z}}(0)$. From (A.7) we obtain

$$\begin{aligned} L_{\mathbf{z}}(\varepsilon)-L_{\mathbf{z}}(0) \geqslant \underline{f}\int_{-x}^{x} & |q_{\varepsilon,\mathbf{z}}(x')-q_{0,\mathbf{z}}(x')|\; I(|q_{\varepsilon,\mathbf{z}}(x')|, \\ & |q_{0,\mathbf{z}}(x')|\leqslant 2)\; dG(x'). \end{aligned} \tag{B.9}$$

Since $0\leqslant y\leqslant 1$, it follows that for any $|x'|\leqslant|x|$,

$$|q_{0,\mathbf{z}}(x')|=\frac{y\,|x'|}{x}\leqslant 1<2 \tag{B.10}$$

and, for any $\varepsilon<0.5$,

$$|q_{\varepsilon,\mathbf{z}}(x')|=\left|\varepsilon+\frac{(y-\varepsilon)\,x'}{x}\right|\leqslant\varepsilon+|y-\varepsilon|<2. \tag{B.11}$$

Hence, from (B.9)–(B.11),

$$\begin{aligned} L_{\mathbf{z}}(\varepsilon)-L_{\mathbf{z}}(0) &\geqslant \frac{\underline{f}\varepsilon}{x}\int_{-x}^{x}|x'-x|\; dG(x') \\ &= \frac{\underline{f}\varepsilon}{x}\,E_G(|X-x|\; I(|X|\leqslant x)). \end{aligned} \tag{B.12}$$

We now obtain from (B.6)–(B.8)

$$\{y; 0 \leqslant y \leqslant 1, (x, y) \in A_\varepsilon\} \subseteq \left[0, \frac{\|f\|_\infty E_G |X-x| \, \varepsilon/x}{2\underline{f}(1-G(x))}\right]$$
$$\subseteq \left[0, \frac{\|f\|_\infty}{2\underline{f}} \left(\frac{E_G |X|}{\delta} + 1\right) \frac{\varepsilon}{1-G(x)}\right]. \tag{B.13}$$

Similarly, (B.6), (B.8), and (B.12) imply that

$$\{y; 0 \leqslant y \leqslant 1, (x, y) \in A_\varepsilon\}$$
$$\supseteq \left[0, \frac{\underline{f}\varepsilon E_G(|X-x| \, I(|X| \leqslant x))/x}{2\,\|f\|_\infty (1-G(x)) + 2\,\|f\|_\infty E_G(X-x)_+/x}\right]$$
$$\supseteq \left[0, \frac{C\varepsilon}{1-G(x)}\right], \tag{B.14}$$

for some constant $C > 0$. The last step in (B.14) follows since $E_G(|X-x| \, I(|X| \leqslant x))/x$ can be lower bounded by the same positive constant for all $x \geqslant \delta$ and, secondly, $E_G(X-x)_+/x \leqslant C'(1-G(x))$ for some $C' > 0$ because of (G). The case $x \geqslant \delta$ is now proved because of (B.13)–(B.14). ∎

LEMMA B.2. *Let A_ε be the set defined in* (B.1). *Then there exist positive constants $C_3 < C_4$ such that for small enough $\varepsilon > 0$,*

$$C_3 \varepsilon \log \frac{1}{\varepsilon} \leqslant K\{A_\varepsilon\} \leqslant C_4 \varepsilon \log \frac{1}{\varepsilon}. \tag{B.15}$$

Proof. By symmetry we have (cf. Lemma A.1) that $K\{A_\varepsilon\} = 4K\{A_{\varepsilon 1}\}$, where $A_{\varepsilon 1}$ is the intersection of A_ε and the first quadrant. It therefore suffices to consider $A_{\varepsilon 1}$ instead of A_ε. With C_2 the constant in (B.2), define $x_\varepsilon(b)$ as any solution of

$$1 - G(x_\varepsilon(b)) = \frac{C_2 \varepsilon}{b}. \tag{B.16}$$

Then, by Lemma B.1, for small enough $\varepsilon > 0$,

$$K\{A_{\varepsilon 1}\} \leqslant \int_0^{x_\varepsilon(1)} \frac{C_2 \varepsilon}{1-G(x')} \, dG(x') + 0.5 P_G(X > x_\varepsilon(1))$$
$$+ P_K(\mathbf{Z} \in A_{\varepsilon 1}, Y > 1)$$
$$= C_2 \varepsilon \log\left(\frac{1}{2C_2\varepsilon}\right) + 0.5 C_2 \varepsilon + P_K(\mathbf{Z} \in A_{\varepsilon 1}, Y > 1) \tag{B.17}$$

and

$$K\{A_{\varepsilon 1}\} \geqslant \int_0^{x_\varepsilon(1)} \frac{C_1 \varepsilon}{1 - G(x')}\, dG(x') = C_1 \varepsilon \log\left(\frac{1}{2C_2 \varepsilon}\right). \qquad \text{(B.18)}$$

It remains to bound the last term in (B.17). We use (B.6) for this. When $x > 0$ and $y > 1$, a lower bound for $K\{\Omega_{\mathbf{z}}\}$ is

$$K\{\Omega_{\mathbf{z}}\} \geqslant (1 - G(x))(F(1) - 0.5). \qquad \text{(B.19)}$$

Let δ be defined as in the proof of Lemma B.1. Then (B.6), (B.7), and (B.19) imply that

$$\begin{aligned} P_K(\mathbf{Z} \in A_{\varepsilon 1},\, Y > 1,\, X > \delta) &\leqslant P_G\Bigg(X > \delta;\, 2(1 - G(X))(F(1) - 0.5) \\ &\leqslant \frac{\|f\|_\infty\, E_G\, |X - x|\, \varepsilon}{x}\Bigg) \leqslant P_G((1 - G(X)) \\ &\leqslant \frac{\|f\|_\infty}{2(F(1) - 0.5)}\left(1 + \frac{E_G\,|X|}{\delta}\right) \varepsilon\Bigg) = O(\varepsilon). \qquad \text{(B.20)} \end{aligned}$$

Finally, from (B.4),

$$P_K(\mathbf{Z} \in A_{\varepsilon 1},\, Y > 1,\, 0 < X \leqslant \delta) \leqslant P_K(0 \leqslant Y \leqslant D\varepsilon,\, 0 < X \leqslant \delta) = O(\varepsilon). \qquad \text{(B.21)}$$

The lemma now follows from (B.17)–(B.18) and (B.20)–(B.21). ▮

LEMMA B.3. *The set A_2 defined in* (3.6) *satisfies*

$$K\{A_2\} = O((\log n)^{\gamma + \tau}\, n^{-1/2}) = o((\log n)^{3/4}\, n^{-1/2}). \qquad \text{(B.22)}$$

Proof. As in the proof of Lemma B.2, it suffices to consider A_{21}, the intersection between A_2 and the first quadrant. We decompose A_{21} into three components,

$$\begin{aligned} A_{21} = A_{211} \cup A_{212} \cup A_{213} &= \{\mathbf{z};\, 0 \leqslant x < a_n,\, y > 0,\, |H(\mathbf{z})| \leqslant \varepsilon_n\} \\ &\cup \{\mathbf{z};\, x \geqslant 0,\, b_n < y \leqslant 1,\, |H(\mathbf{z})| \leqslant \varepsilon_n\} \\ &\cup \{\mathbf{z};\, x \geqslant 0,\, y > 1,\, |H(\mathbf{z})| \leqslant \varepsilon_n\}, \qquad \text{(B.23)} \end{aligned}$$

and we will estimate the probability of each set separately. Suppose n is so large that $a_n \leqslant x_{\varepsilon_n}(1)$ (cf. (B.16)), that is, $1 - G(a_n) \geqslant C_2 \varepsilon_n$, where C_2 is the same constant as in (B.2). This is possible in view of (3.2) and (3.4). It then follows from Lemma B.1 that

$$K\{A_{211}\} \leqslant \int_0^{a_n} \frac{C_2 \varepsilon}{1-G(x')}\, dG(x') = C_2 \varepsilon_n \log \left(\frac{1}{2(1-G(a_n))} \right)$$
$$= O((\log n)^{\gamma+\tau}\, n^{-1/2}). \qquad \text{(B.24)}$$

We now turn to A_{212}. Then from Lemma B.1 and (B.16),

$$K\{A_{212}\} \leqslant \int_{x_{\varepsilon_n}(b_n)}^{x_{\varepsilon_n}(1)} \frac{C_2 \varepsilon}{1-G(x')}\, dG(x')$$
$$= C_2 \varepsilon_n \log \left(\frac{1-G(x_{\varepsilon_n}(b_n))}{1-G(x_{\varepsilon_n}(1))} \right)$$
$$= C_2 \varepsilon_n \log \frac{1}{b_n} = O((\log n)^{\gamma+\tau}\, n^{-1/2}). \qquad \text{(B.25)}$$

Finally,

$$K\{A_{213}\} = O(\varepsilon_n) \qquad \text{(B.26)}$$

is a consequence of (B.20), (B.21). Formula (B.22) now follows from (B.24)-(B.26). ∎

Lemma B.4. *Let A_3 be as defined in* (3.7). *Then*

$$K\{A_3\} = O(\varepsilon_n). \qquad \text{(B.27)}$$

Proof. By symmetry it suffices to consider A_{31}, the intersection of A_3 and the first quadrant. Then

$$A_{31} \subseteq \{\mathbf{z};\, L_{\mathbf{z}}(\varepsilon) > 0.5 - \rho'\varepsilon_n,\, 0 \leqslant x \leqslant \delta,\, y \geqslant 0\}$$
$$\cup\, \{\mathbf{z};\, L_{\mathbf{z}}(\varepsilon_n) > 0.5 - \rho'\varepsilon_n,\, x > \delta,\, y > 1\} \triangleq A_{311} \cup A_{312}.$$

Using a completely analogous argument to that in connection with (B.4) (which corresponds to $\rho' = 0$), we may choose a $D > 0$ such that for large enough n, $A_{311} \subseteq [0, \delta] \times [0, D\varepsilon_n]$, and hence $K\{A_{311}\} = O(\varepsilon_n)$. The fact that $K\{A_{312}\} = O(\varepsilon_n)$ follows in the same way as (B.20) (which corresponds to $\rho' = 0$). ∎

Acknowledgments

The research of the first author was supported by the Swedish Board for Technical Development under Contract 88-01359P. The authors thank Christophe Croux for sharing useful software.

References

[1] Cheney, W., and Kincaid, D. (1985). *Numerical Methods and Computing.* Brooks–Cole, Pacific Grove, CA.

[2] Cox, D. R., and Hinkley, D. V. (1974). *Theoretical Statistics.* Chapman and Hall, London.

[3] 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.

[4] Hössjer, O., Rousseeuw, P. J., and Croux, C. (1994). Asymptotics of the repeated median slope estimator. *Ann. Statist.*

[5] Hössjer, O., Rousseeuw, P. J., and Ruts, I. (1992). The repeated median intercept estimator: Influence function and asymptotic normality, Tech. Rep. 1992:13. Uppsala University Dept. of Mathematics.

[6] Jurečková, J. (1971). Asymptotic Independence of the Rank Test Statistic for Testing Symmetry in Regression. *Sankhya Ser. A* 10–18.

[7] Mount, D. M., and Netanyahu, N. S. (1991). Computationally efficient algorithms for a highly robust slope estimator, Tech. Rep. Center for Automation Research, University of Maryland.

[8] Rousseeuw, P. J., Croux, C., and Hössjer, O. (1995). Sensitivity functions and numerical analysis of the repeated median slope, *Computational Statistics*, to appear.

[9] Rousseeuw, P. J., and Leroy, A. (1987). *Robust Regression and Outlier Detection.* Wiley, New York.

[10] Serfling, R. A. (1980). *Approximation Theorems of Mathematical Statistics.* Wiley, New York.

[11] Siegel, A. F. (1982). Robust regression using repeated medians. *Biometrika* **69** 242–244.

[12] Tukey, J. W. (1970). *Exploratory Data Analysis*, limited preliminary ed. Addison–Wesley, Reading, MA.