Sensitivity Functions and Numerical Analysis of the Repeated Median Slope

Peter J. Rousseeuw

Department of Mathematics and Computer Science, UIA, Universiteitsplein 1, B-2610 Antwerpen, Belgium

Christophe Croux

Department of Mathematics and Computer Science, UIA, Universiteitsplein 1, B-2610 Antwerpen, Belgium

Ola Hössjer

Department of Mathematical Statistics, Lund Institute of Technology, Box 118, S-22100 Lund, Sweden

Summary

Recently it has been shown that the repeated median slope estimator converges at the $n^{-1/2}$ rate, and is asymptotically normal with a gaussian efficiency of 40.5%. On the other hand, simulations indicate that its finite-sample efficiency converges extremely slowly to the limiting value. Our goal is to explain this unusual phenomenon by constructing finite-sample sensitivity functions, and by investigating key quantities with numerical analysis methods.

Keywords: Asymptotic normality; Breakdown point; Influence function; Robust regression; Simulation.

1. Introduction

Consider the simple linear regression model

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

where $\mathbf{z}_i = (x_i, y_i)$ are the observations and e_i represents noise. We assume that the (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

$$\begin{aligned} h(\mathbf{z}_i, \mathbf{z}_j) &= \frac{y_j - y_i}{x_j - x_i} \qquad \text{where } x_i \neq x_j \\ &= 0 \qquad\qquad\quad \text{where } x_i = x_j. \end{aligned} \tag{1.2}$$

For instance, the least squares estimator $\hat{\beta}_{LS}$ may be written as a weighted average,

$$\hat{\beta}_{LS} = \sum_{i<j} w_{ij} h(\mathbf{z}_i, \mathbf{z}_j) \Big/ \sum_{i<j} w_{ij}$$

with weights $w_{ij} = (x_i - x_j)^2$. Other estimators based on weighted medians and trimmed means of pairwise slopes have been proposed by Theil (1950), Adichie (1967), Sen (1968), Jaeckel (1972), and Scholz (1978). Frees (1991) gives a survey of these estimators.

Another estimator, the repeated median slope (or "RM slope")

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

was proposed by Siegel (1982). He showed that when all x_i are distinct (an event with probability one if G is continuous), $\hat{\beta}_n$ has a finite-sample breakdown point $\varepsilon_n^* = [n/2]/n$, i.e. 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), yielding an asymptotic breakdown point of 0.5. (For a slope estimator, regression equivariance comes down to

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

for any real constants c and d.) Siegel (1982) also showed that $\hat{\beta}_n$ is a Fisher-consistent estimate of β.

To keep the notation simple, we will assume from now on that $\alpha = \beta = 0 = F^{-1}(\frac{1}{2}) = G^{-1}(\frac{1}{2})$, which is no restriction because of the regression equivariance. In this case $y_i = e_i$ and we can denote the distribution of $\mathbf{z}_i = (x_i, y_i)$ by $K = G \times F$.

Recently it was proved (Hössjer et al, 1994) that the RM slope satisfies

$$\hat{\beta}_n - \beta = \frac{1}{n} \sum_{i=1}^{n} IF(\mathbf{z}_i) + o_p(n^{-1/2}) \tag{1.4}$$

where the influence function (IF) is given by

$$IF(\mathbf{z}) = IF(x, y) = \frac{\operatorname{sgn}(xy)}{2f(0)E_G|X|}. \tag{1.5}$$

Here f is the density of F. Applying the Central Limit Theorem to (1.4) yields the asymptotic normality

$$\sqrt{n}(\hat{\beta}_n - \beta) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} IF(\mathbf{z}_i) + o_p(1) \xrightarrow{d} N\left(0, \int IF(\mathbf{z})^2 dK(\mathbf{z})\right) \tag{1.6}$$

and the asymptotic variance is therefore

$$\int IF(x,y)^2 dK(x,y) = \frac{1}{4f^2(0)(E_G|X|)^2}. \tag{1.7}$$

Note that the influence function in (1.4) is defined through a first order von Mises expansion (Hampel et al 1986, page 85). We see from (1.5) that this influence function is bounded, again illustrating the robustness of the RM slope.

The proof of the expansion (1.4) is far from trivial, due to the "nested" operations in (1.3). Moreover, it turns out that the limit in (1.6) is reached very slowly because of some unusual remainder terms in (1.4). In the present paper we will study this phenomenon by means of numerical methods, and provide some graphical displays to facilitate their interpretation. In Section 2 we see that Monte Carlo efficiencies of the RM slope approach their limit very slowly, the difference still being large at $n = 40,000$. Section 3 recapitulates the derivation of the IF, on which the numerical work will be based. In Section 4 we compute finite-sample sensitivity functions, including averaged and stylized versions which were constructed especially for this bivariate situation. Section 5 presents an explanation of this strange behavior based on a numerical investigation of certain quantities which play a crucial role in the derivation of the influence function.

2. Monte Carlo results

In order to study the behavior of the RM slope estimator empirically, we will have to compute it very often. As indicated by Siegel (1982, page 242) the repeated median slope can be computed in $O(n^2)$ time. This is achieved by a brute force algorithm for (1.3) which computes the inner median for each i, and then computes the outer median once. Very recently, faster algorithms for the RM slope have been obtained. In our computations we used the algorithm described in (Rousseeuw, Netanyahu and Mount 1993), which runs in expected time $O(n \log^2 n)$ and requires only $O(n)$ storage. This algorithm makes use of dualization and mergesort, and yields the exact result. Its speed allowed us to perform simulations for n up to $40,000$ whereas with the brute force algorithm we could not go beyond $n = 3,000$.

From (1.6) it follows 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 $\hat{\beta}_n^{(k)}$ for

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%

$k = 1, \ldots, m$. Table 1 lists the bias given by

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

which is very small, as well as the n-fold variance

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

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 and the congruential generator of Cheney and Kincaid (1985, page 335). 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 $\{\hat{\beta}_n^{(k)}, k = 1, \ldots, m\}$ of estimated slopes. From these it does appear that the sampling distribution of the estimator $\hat{\beta}_n$ 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%.

In conclusion, the RM slope estimator performs better at finite samples than would be expected from its asymptotic behavior. In this respect it outperforms the Brown-Mood estimator, which has the same IF as the RM and hence also the same asymptotic efficiency of 40.5%, but the finite-sample efficiency of which remains around 40% (see Table 3 of Johnstone and Velleman 1985).

3. Influence functions

In this section we will sketch the derivation of (1.4) given in Hössjer et al (1994), where this expansion was used for obtaining the influence function of the RM slope and proving its asymptotic normality. In the next sections we will then investigate the main steps of this construction by means of numerical methods.

As in Section 1, we assume that (1.1) holds with $\alpha = \beta = 0 = F^{-1}(\frac{1}{2}) = G^{-1}(\frac{1}{2})$. Therefore, the observations $\mathbf{z}_1, \ldots, \mathbf{z}_n$ are i.i.d. with common distribution $K = G \times F$. Denote the "inner" median in (1.3) by T_1 and the "outer" median by T_2. For each $\mathbf{z}$ we put $H(\mathbf{z}) = T_1(L_{\mathbf{z}})$ where

$$L_{\mathbf{z}}(t) = P_K(h(\mathbf{z}, \mathbf{Z}) \leq t) \tag{3.1}$$

Then $T(K) = T_2(L)$, where

$$L(t) = P_K(H(\mathbf{Z}) \leq t) \tag{3.2}$$

is the functional corresponding to the RM slope.

We obtained

$$IF(\mathbf{z}) = IF_2\big(H(\mathbf{z})\big) + E_K\big[IF_2'\big(H(\mathbf{Z})\big) IF_1(\mathbf{Z}, \mathbf{z})\big] \tag{3.3}$$

with

$$IF_1(\mathbf{z}_1, \mathbf{z}_2) = IF(h(\mathbf{z}_1, \mathbf{z}_2), T_1, L_{\mathbf{z}_1}) = \frac{\operatorname{sgn}\big(h(\mathbf{z}_1, \mathbf{z}_2) - H(\mathbf{z}_1)\big)}{2l_{\mathbf{z}_1}\big(H(\mathbf{z}_1)\big)}$$

and

$$IF_2(x) = IF(x, T_2, L) = \frac{\operatorname{sgn}(x - T(K))}{2l(T(K))},$$

where $l_{\mathbf{z}} = L_{\mathbf{z}}'$ and $l = L'$. Unfortunately, this IF_2 is not differentiable at $T(K)$, making it difficult to deal with (3.3). Therefore, we temporarily

replace T_2 by an M-estimator T_2^ε of Huber type, based on the function

$$\psi_\varepsilon(x) = \begin{cases} \operatorname{sgn}(x),\ |x| > \varepsilon \\ x/\varepsilon,\ |x| \le \varepsilon, \end{cases} \tag{3.4}$$

and then we let $\varepsilon \to 0+$. Putting $T_\varepsilon(K) = T_2^\varepsilon(L)$, Formula (3.3) for the influence function becomes

$$IF^\varepsilon(\mathbf{z}) = \frac{\varepsilon\psi_\varepsilon\big(H(\mathbf{z}) - T_\varepsilon(K)\big)}{L\{[T_\varepsilon(K) - \varepsilon, T_\varepsilon(K) + \varepsilon]\}} + E_{K_\varepsilon}\frac{\operatorname{sgn}\big(h(\mathbf{Z},\mathbf{z}) - H(\mathbf{Z})\big)}{2l_{\mathbf{Z}}\big(H(\mathbf{Z})\big)}, \tag{3.5}$$

where K_ε denotes the conditional distribution of $\mathbf{Z} \sim K$, given that $H(\mathbf{Z}) \in [T_\varepsilon(K) - \varepsilon, T_\varepsilon(K) + \varepsilon]$. If now $\varepsilon \to 0+$ implies that $T_\varepsilon(K) \to T(K)$ as well as $(L\{[T_\varepsilon(K) - \varepsilon, T_\varepsilon(K) + \varepsilon]\})/\varepsilon \to 2l(T(K))$ and $K_\varepsilon \xrightarrow{d} K_0$ for some distribution K_0, it follows that

$$IF^\varepsilon(\mathbf{z}) \to IF(\mathbf{z}) = \frac{\operatorname{sgn}\big(H(\mathbf{z}) - T(K)\big)}{2l(T(K))} + E_{K_0}\frac{\operatorname{sgn}\big(h(\mathbf{Z},\mathbf{z}) - H(\mathbf{Z})\big)}{2l_{\mathbf{Z}}\big(H(\mathbf{Z})\big)}. \tag{3.6}$$

This expression can be simplified because in Hössjer et al (1994) it is found that $T(K) = 0$, and also that

$$\operatorname{sgn}(H(\mathbf{z})) = \operatorname{sgn}(xy) \quad \text{and} \quad l(0) = \infty. \tag{3.7}$$

Therefore, the first term in (3.6) vanishes. It has also been proved that K_0 equals the Dirac measure at $(0,0)$, because the set $A_\varepsilon = \{z;\, H(\mathbf{z}) \le \varepsilon\}$ looks roughly like

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

and in particular, around the origin, like

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

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

$$P\big(\Omega_d \cap \{\mathbf{z};\ |H(\mathbf{z})| \le \varepsilon\}\big) \asymp \varepsilon\log(\frac{1}{\varepsilon}),\ \text{as } \varepsilon \to 0+, \tag{3.8}$$

while

$$P\big(\Omega_d^c \cap \{\mathbf{z};\ |H(\mathbf{z})| \le \varepsilon\}\big) = O(\varepsilon),\ \text{as } \varepsilon \to 0+. \tag{3.9}$$

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 $T_\varepsilon(K) = 0$. Hence, in this case K_ε is the conditional distribution of $\mathbf{Z} \sim K$ on the set $A_\varepsilon = \{\mathbf{z}; |H(\mathbf{z})| \le \varepsilon\}$, and so by (3.8)–(3.9) it converges weakly to δ_0 as $\varepsilon \to 0+$. Summarizing, the influence function in (3.6) becomes

$$IF(\mathbf{z}) = \frac{\operatorname{sgn}\big(h(\mathbf{0},\mathbf{z})\big)}{2l_{\mathbf{0}}(\mathbf{0})} = \frac{\operatorname{sgn}(xy)}{2f(0)E_G|X|} \tag{3.10}$$

confirming (1.5). Note that the fact that K_0 is a one point distribution has simplified the expression for the influence function a lot.

4. Sensitivity functions

In Section 3 the influence function of the RM slope was obtained in the case where the data are i.i.d. according to $K = G \times F$, which amounts to assuming that the true parameters α and β in (1.1) are zero. If both G and F equal the standard gaussian distribution Φ, (3.10) becomes

$$IF(x, y) = \frac{\pi}{2}\text{sgn}(xy). \tag{4.1}$$

This IF has a positive constant value in quadrants 1 and 3 and takes on the opposite value in quadrants 2 and 4, as shown in Figure 1a. Therefore $IF(x, y)$ is bounded in both x and y, unlike the least absolute deviations method.

Note that the influence function in (4.1) is an asymptotic concept, whereas we saw in Section 2 that the finite-sample variability of the RM slope its quite different from its asymptotic limit. In order this examine this discrepancy, we would like to have an empirical version of the IF. For location estimators based on univariate data, Tukey (1970) introduced the *sensitivity curve* as a finite-sample approximation to the influence function. However, the influence function of a slope estimator depends on two arguments x and y, and hence becomes a surface in three-dimensional space. We therefore have to extend Tukey's definition accordingly. Let us start with a sample $\{\mathbf{z}_1, \ldots, \mathbf{z}_n\}$ of the bivariate distribution $K = \Phi \times \Phi$. Then the sensitivity function is defined in each $\mathbf{z} = (x, y)$ by

$$SF_n(\mathbf{z}) = n(\hat{\beta}_{n+1}(\mathbf{z}_1, \ldots, \mathbf{z}_n, \mathbf{z}) - \hat{\beta}_n(\mathbf{z}_1, \ldots, \mathbf{z}_n)). \tag{4.2}$$

Note that $SF_n(\mathbf{z})$ also depends on which slope estimator we are using, as well as on the original sample $\mathbf{Z} = \{\mathbf{z}_1, \ldots, \mathbf{z}_n\}$. Therefore, a more complete notation would be $SF_n(\mathbf{z}; T, \mathbf{Z})$. Also note that $\mathbf{z}$ is not restricted to the observations, but that it can be any point in $I\!R^2$. In fact, when making a plot of $SF_n(\mathbf{z})$ we will let $\mathbf{z}$ range over a rectangular grid.

Figure 1b shows such a sensitivity function SF_n with $n = 20$. Note that the constant parts (both positive and negative) are clearly visible, but that the transitions between them do not occur exactly on the x-axis and the y-axis, and that in the intermediate region SF_n looks somewhat "jumpy". This is due to the fact that SF_n is based on a finite sample $\{\mathbf{z}_1, \ldots, \mathbf{z}_n\}$ from $K = \Phi \times \Phi$, which may provide 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 different samples, as was proposed in the univariate situation by Rousseeuw and Leroy (1987, page 193). This yields the *averaged sensitivity function* given by

$$ASF_n(x, y) = \underset{j=1,\ldots,m}{\text{average}} SF_n(x, y; T, \mathbf{Z}^{(j)}) \tag{4.3}$$

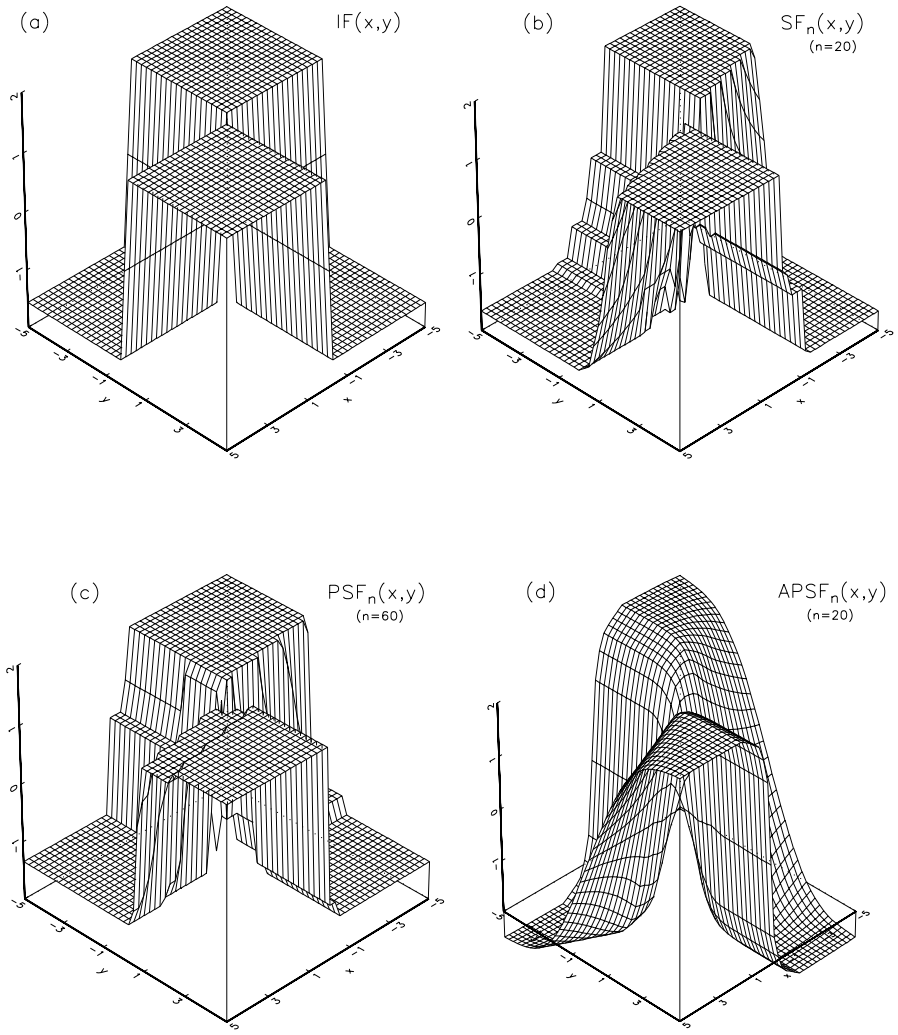

Figure 1: (a) Influence function of RM slope estimator; (b) Sensitivity function for $n = 20$; (c) Permutation-stylized SF for $n = 60$; (d) Averaged permutation-stylized SF for $n = 20$.

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 (the number of replications) is increased. However, this approach needs rather large amounts of computation time compared to the univariate situation. (One way to improve the resulting ASF is to prestandardize each $\mathbf{Z}^{(j)}$ by replacing all x_i by $x_i - \text{med}_k x_k$ and all y_i by $y_i - \text{med}_k y_k$.)

In the case of univariate data, another approach is to work with a "stylized sample"

$$x_i^s = \Phi^{-1}(\frac{i}{n+1}) \quad \text{for } i = 1, \ldots, n$$

as proposed by Andrews et al (1972). Compared to random samples, a stylized sample gives a better approximation to the population distribution (for instance, it inherits the symmetry and the zero median of Φ). It also offers the advantage of being unique, so no averaging is needed. Unfortunately, there is no straightforward way to define stylized samples in the bivariate case. (It seems hard to preserve some form of spherical symmetry in combination with the right marginal distributions.) One way out would be to use "semi-stylized" data sets in which the x_i are stylized and the y_i are randomly generated (or vice versa). Instead, we propose to construct "permutation-stylized" data sets which are defined as

$$\mathbf{Z}(\pi) = \{(x_i^s, x_{\pi(i)}^s); i = 1, \ldots, n\} \tag{4.4}$$

where π is a random permutation on $\{1, \ldots, n\}$. This has the advantage that the marginal distribution of y is also stylized (and hence symmetric with zero median). Definition (4.4) treats both variables in the same way, because interchanging x and y merely corresponds to using the inverse permutation π^{-1}. Figure 1c shows a permutation-stylized sensitivity function (PSF), obtained by inserting a data set of the form $\mathbf{Z}(\pi)$ with $n = 60$ in (4.2). We clearly recognize the horizontal parts of this surface, but in the boundary regions there are some irregularities which depend on the particular choice of π.

As a final step, the effect of the permutation can be averaged out by computing

$$APSF_n(x, y) = \underset{\pi}{\text{average}}\, SF_n(x, y; T, \mathbf{Z}(\pi)) \tag{4.5}$$

where π ranges over a collection of random permutations. This approach seems to give somewhat better results (for the same amount of computation time) than the ASF of (4.3). Figure 1d shows this averaged permutation-stylized SF for $n = 20$ and $m = 1,000$. It provides a fairly close approximation to the theoretical influence function given in Figure 1a. Note that there is a rather smooth transition between the constant parts of the $APSF$ in Figure 1d, which is caused by averaging many rough jumps occurring in the same region.

A way to explain the high finite-sample efficiencies in Section 2 is by looking at the sensitivity functions above. Whereas the (asymptotic) influence function is discontinuous on the x-axis and the y-axis (causing the low asymptotic efficiency of 40.5%), the finite-sample SF's make a more gradual transition between the positive and negative horizontal parts. (It could be argued that a less pronounced version of this effect exists for the univariate median, whose small-sample SF's look like those of a trimmed mean.) This effect seems to wear off slowly when n increases substantially, in which case the SF's of the RM estimator begin to look more and more like its IF.

As the asymptotic variance equals $E_K[IF^2]$, it is natural to investigate whether the expected value of SF_n^2 provides a good approximation to the n-fold Monte Carlo variance. In the univariate case, Rousseeuw and Leroy (1987, page 189) approximated $E_K[SF_n^2]$ by means of $V_n = \text{ave}_k SF_n(\mathbf{z}^{(k)})^2$, where the points $\mathbf{z}^{(k)}$ (for $k = 1, \ldots, m$) were generated according to K. In the bivariate situation this does not appear to be sufficiently accurate, so we used numerical integration (the composed Simpson rule) instead. As was to be expected, the result based on SF_n was quite variable due to the dependence on $\mathbf{z}_1, \ldots, \mathbf{z}_n$ in (4.2). Replacing SF_n by ASF_n (for $m = 1000$) was already somewhat better, but we could only consider small n due to the large amount of computation time needed. By far the closest approximation was obtained by computing $E_K[APSF_n^2]$ for the same value of m. Therefore, the smooth shape of $APSF_n$ does help to explain the finite-sample efficiency of the RM slope estimator.

5. The function H and related quantities

In the proof of (1.4) given in (Hössjer et al 1994) and summarized in Section 3 above, certain remainder terms tend to zero at a very slow rate. As a consequence, unusually large samples are needed before the finite-sample efficiency comes close to its asymptotic limit of 40.5%. In this section, we will investigate the underlying cause (the slow convergence of K_ε to K_0) in a numerical way.

Throughout, an important role is played by the function H defined in the beginning of Section 3. For each point $\mathbf{z} = (x, y)$ in $\mathbb{R}^2$, the value $H(\mathbf{z})$ is given by

$$H(\mathbf{z}) = L_{\mathbf{z}}^{-1}(0.5) = \operatorname*{med}_{\mathbf{Z} \sim K} h(\mathbf{z}, \mathbf{Z}). \tag{5.1}$$

The function H arises naturally when we write down the RM functional at the model distribution $K = \Phi \times \Phi$, because

$$T(K) = \operatorname*{med}_{\mathbf{Z} \sim K} \operatorname*{med}_{\mathbf{Z}' \sim K} h(\mathbf{Z}, \mathbf{Z}') = \operatorname*{med}_{\mathbf{Z} \sim K} H(\mathbf{Z}) \tag{5.2}$$

(note that $T(K) = \beta = 0$ by assumption).

Although H is a deterministic function, no simple expression is available.

Table 2: Values of $H(x,y)$ for x and y ranging between 0.0 and 5.0

	x										
y	**0.0**	**0.5**	**1.0**	**1.5**	**2.0**	**2.5**	**3.0**	**3.5**	**4.0**	**4.5**	**5.0**
0.0	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.5	0.00	0.23	0.30	0.28	0.24	0.20	0.17	0.14	0.12	0.11	0.10
1.0	0.00	0.51	0.62	0.57	0.48	0.40	0.33	0.29	0.25	0.22	0.20
1.5	0.00	0.86	0.97	0.86	0.72	0.59	0.50	0.43	0.37	0.33	0.30
2.0	0.00	1.25	1.33	1.16	0.96	0.79	0.67	0.57	0.50	0.44	0.40
2.5	0.00	1.65	1.70	1.47	1.21	0.99	0.83	0.71	0.62	0.56	0.50
3.0	0.00	2.04	2.06	1.77	1.45	1.19	1.00	0.86	0.75	0.67	0.60
3.5	0.00	2.43	2.43	2.07	1.69	1.39	1.16	1.00	0.87	0.78	0.70
4.0	0.00	2.81	2.79	2.38	1.94	1.59	1.33	1.14	1.00	0.89	0.80
4.5	0.00	3.18	3.15	2.68	2.18	1.79	1.50	1.29	1.12	1.00	0.90
5.0	0.00	3.56	3.51	2.98	2.43	1.99	1.66	1.43	1.25	1.11	1.00

We can estimate $H(\mathbf{z})$ by

$$H_m(\mathbf{z}) = \underset{k=1,\ldots,m}{\text{med}} h(\mathbf{z}, \mathbf{z}^{(k)}) \tag{5.3}$$

where the points $\mathbf{z}^{(k)}$ (for $k = 1, \ldots, m$) are generated according to K. We also have implemented an iterative algorithm which needs considerable computation time, but can calculate $H(\mathbf{z})$ with arbitrarily high precision. The basic idea is to solve $L_{\mathbf{z}}(t) = 1/2$ by means of Newton-Raphson, starting from $t_0 = H_m(\mathbf{z})$, and making use of numerical integration in the expressions

$$L_{\mathbf{z}}(t) = \int_{-\infty}^{x} (1 - F(y + t(x' - x)))dG(x') + \int_{x}^{\infty} F(y + t(x' - x))dG(x') \tag{5.4}$$

and

$$\partial L_{\mathbf{z}}(t)/\partial t = l_{\mathbf{z}}(t) = \int_{-\infty}^{\infty} |x' - x| f(y + t(x' - x))dG(x'). \tag{5.5}$$

By comparing the final solution to the starting value, we found that $H_m(\mathbf{z})$ is a reasonably good approximation to $H(\mathbf{z})$ provided m is fairly large (we used $m = 3000$).

Table 2 gives some values of H, obtained with the numerical algorithm, for x and y ranging between 0 and 5. Only values for the first quadrant are listed, because $H(x,y) = H(-x,-y) = -H(-x,y) = -H(x,-y)$. Note that H is an unbounded function, which attains large values when x is small (but not zero) and y is large. (Indeed, from such a point (x,y) the median slope to the bivariate gaussian data can be very steep! Observations in that region would be very harmful, if the outer estimator in (5.2) were not robust.) We can also verify that $H(x,y)$ is increasing in y when $x > 0$ is held fixed. (Note

that $H(x,y)$ equals the median of the doubly noncentral Cauchy distribution with noncentrality parameters $\delta = -y$ and $\lambda = -x$.)

Figure 2a gives an idea of the shape of H around the origin. Figure 2b displays some contours (obtained numerically), given by $H(x,y) = \varepsilon$ for various values of ε. The plot resembles a spider web. We see that each contour of H has the y-axis as an asymptote, but not the x-axis. Indeed, letting x tend to infinity yields a skewed asymptote with slope ε. (This slope follows from the interpretation of H.) When ε tends to zero, the skewed asymptote tends to the x-axis. For small ε, the minimum of the contour curves occurs at $x' = 1.109$.

It is interesting to note the similarity between the function H in Figure 2a and sensitivity functions of the RM estimator. Indeed, if we consider truncated (and standardized) versions of H given by

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

the resulting surface looks like H near the x-axis and y-axis, but is flattened beyond the contours of levels ε and $-\varepsilon$ in Figure 2b. For moderately large ε this resembles the sensitivity function for $n = 20$ shown in Figure 2c. In particular, note that the change of sign is gradual at the x-axis and steep at the y-axis. The PSF for $n = 40$ (Figure 2d) looks like H_ε for a smaller ε, whereas $n = 60$ (Figure 1c) corresponds to a still smaller ε. Letting n tend to infinity corresponds to letting ε tend to zero, which leads to the discontinuous function $\mathrm{sign}(xy)$ in the expression of $IF(x,y)$.

The above reasoning illustrates the importance of the sets

$$A_\varepsilon = \{(x,y);\ |H(x,y)| \le \varepsilon\}$$

which play a crucial role in the derivation of the influence function in Section 3. Combining (3.8) and (3.9), we obtain

$$K\{A_\varepsilon\} \asymp \varepsilon \log\frac{1}{\varepsilon} \quad \text{for } \varepsilon \to 0+. \tag{5.6}$$

This implies that $l(0) = \infty$ which makes the first term of the influence function (3.6) vanish. In order to illustrate (5.6), we generated 2500 points $\mathbf{z}_j$ according to $K = \Phi \times \Phi$ and computed $H(\mathbf{z}_j)$ for each of them, using the abovementioned numerical algorithm for H. The construction of this data base

$$\{(\mathbf{z}_j, H(\mathbf{z}_j));\ j = 1, \ldots, 2500\} \tag{5.7}$$

required substantial computation time. We then considered 20 values of ε, equispaced between 0.005 and 0.100, and approximated $K\{A_\varepsilon\}$ in each case by the fraction of points in (5.7) satisfying $|H(\mathbf{z}_j)| \le \varepsilon$. Figure 3 plots these estimated probabilities versus $-\varepsilon \log \varepsilon$. We see that the points are close to a straight line through the origin (with slope ≈ 1), confirming (5.6).

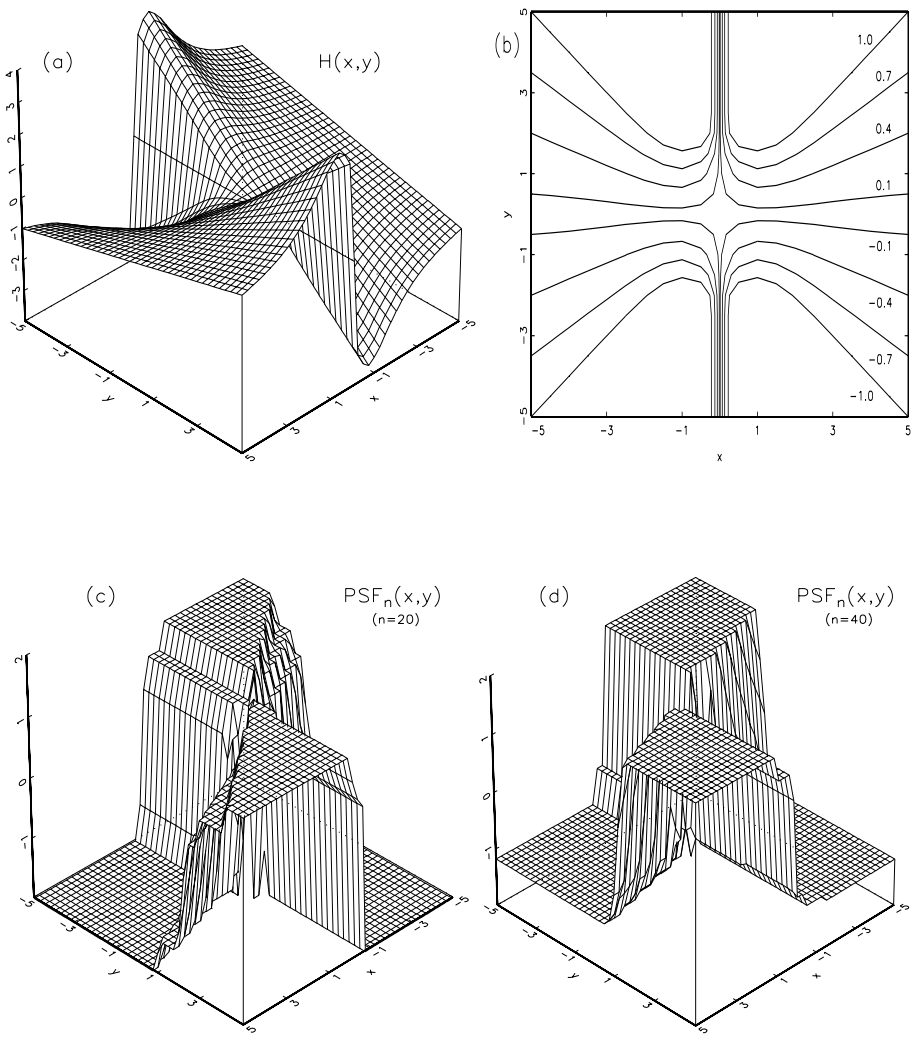

Figure 2: (a) Plot of $H(x, y)$; (b) Contour curves of $H(x, y)$; (c) Permutation-stylized sensitivity function for $n = 20$; (d) for $n = 40$.

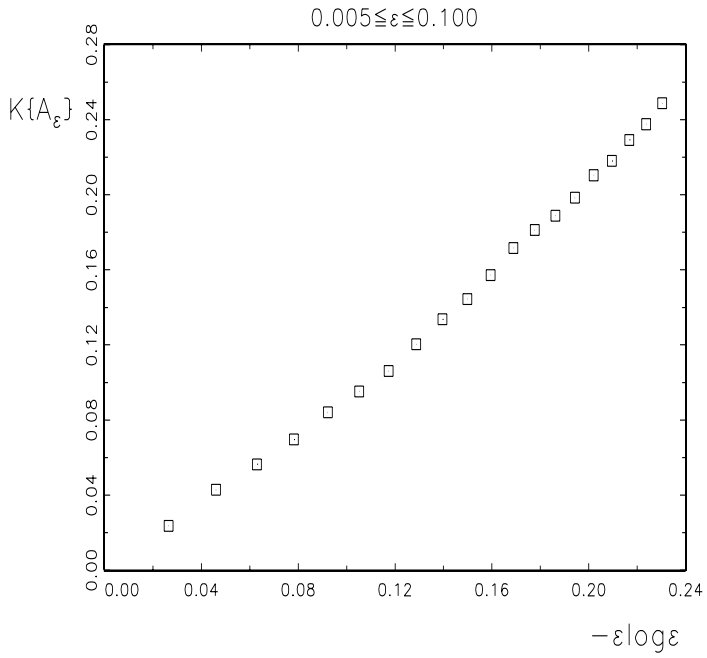

Figure 3: Plot of (estimated) $K\{A_\varepsilon\}$ versus $-\varepsilon \log \varepsilon$.

Note that the same computational effort allows us to obtain an idea of the distribution of $H(\mathbf{Z})$, the cdf of which was denoted as L in Section 3. Figure 4 gives the empirical cdf of H, computed from the 2500 values in (5.7). The graph of L_{2500} is very smooth, and its tangent at zero does become vertical, in accordance with the fact that the density of L satisfies $l(0) = \infty$.

The second key property of A_ε, also following from (3.8) and (3.9), is that

$$K\{A_\varepsilon \setminus \Omega_d\}/K\{A_\varepsilon\} \asymp -1/\log \varepsilon \quad \text{for } \varepsilon \to 0+ \tag{5.8}$$

where Ω_d is the square $[-d, d] \times [-d, d]$ for $d > 0$. This implies that all the mass of A_ε moves into Ω_d when ε tends to zero, which greatly simplifies the second term of the influence function (3.6). In order to illustrate (5.8), we chose $d = 0.5$ (hence, $K\{\Omega_d\} \approx 0.14$). For ε we took values such that $-1/\log \varepsilon$ is roughly equispaced between 0.2 and 0.6. Then we estimated the left member of (5.8) by counting the number of points $\mathbf{z}_j$ of (5.7) falling into the appropriate regions. The result is plotted in Figure 5, which offers some support for (5.8) but looks rather incomplete because we would prefer to include some smaller values of $-1/\log \varepsilon$. However, for smaller values of ε the estimate of (5.8) becomes highly variable, because very few points $\mathbf{z}_j$ belong to A_ε. For instance, for $-1/\log \varepsilon = 0.1$ we find $\varepsilon \approx 0.000045$ hence $K\{A_\varepsilon\} \approx -\varepsilon \log \varepsilon \approx 0.00045$, which means that by generating 2500 points

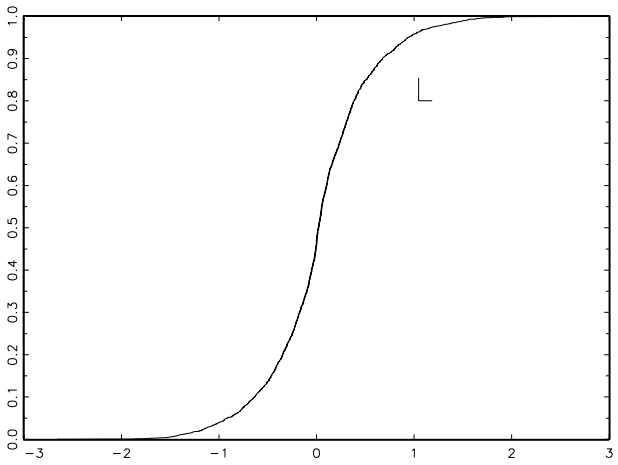

Figure 4: Empirical distribution function of H, based on 2500 points.

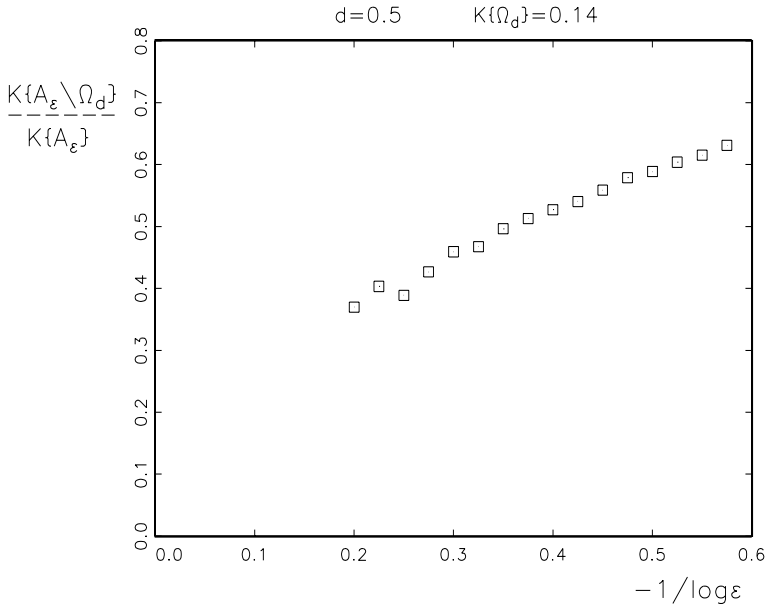

Figure 5: Plot of $K\{A_\varepsilon \setminus \Omega_d\}/K\{A_\varepsilon\}$ versus $-1/\log\varepsilon$.

from K we obtain (on average) only one point in A_ε. By using more refined techniques (e.g., importance sampling) we could come closer to $\varepsilon = 0$, but not very much.

This slow rate of $-1/\log\varepsilon$ in (5.8) also has practical consequences. For instance, if we assume that the proportionality factor at $d = 0.5$ is above 1 (which is not unreasonable when looking at Figure 5), we find that $K\{A_\varepsilon \setminus \Omega_d\}/K\{A_\varepsilon\} < 0.05$ only when $-1/\log\varepsilon < 0.05$, hence $\varepsilon < 2.06 * 10^{-9}$. In the proof of the asymptotic normality, certain remainder terms become negligible for $n > 1/\varepsilon^2$ which yields $n > 2.3 * 10^{17}$. Therefore, Figure 5 gives some insight in the excruciatingly slow convergence of the estimator's finite-sample efficiency to its asymptotic limit.

In Section 3, the conditional distribution of $\mathbf{Z} \sim K$ on the set A_ε was denoted by K_ε. The expression of the influence function hinges on the fact that K_ε converges weakly to the point mass in (0,0), which is proved from (5.8). Therefore, we should be able to see that when ε becomes smaller, the distribution K_ε becomes more and more concentrated around (0,0). Let us look at the marginal distributions of K_ε. Figure 6a shows the x-marginal density, which has been obtained by numerical methods, for $\varepsilon = 0.2$ and $\varepsilon = 0.1$. We see that the second density has a smaller dispersion, but also that the relative factor is less than 2 which reflects the slow convergence. Similarly, Figure 6b shows the y-marginal density for the same ε.

Another way to investigate this convergence is to estimate the scale of the marginal distributions. For values of ε between 0.01 and 0.6, we selected the points of (5.7) belonging to A_ε and then computed the interquartile range of their x-coordinates. Plotting these scale estimates versus ε yielded Figure 7a. As in Figure 5, the "wiggles" at small ε were due to the small number of points in A_ε. The scale estimates in Figure 7a decrease at a slow rate, whereas the scale estimates of the y-marginal distribution (shown in Figure 7b) converge more quickly.

Acknowledgements: The work of the second author was supported by the Belgian Science Foundation, and that of the third author by the Swedish Board for Technical Development, contract 88-01350 P. We would like to thank Dr. Nathan Netanyahu and Bert van Zomeren for running part of the simulations in Table 1, and the referee for helpful suggestions.

References

Adichie, J.N. (1967) Estimation of regression coefficients based on rank tests. *Ann. Math. Statist.*, 38, 894–904.

Andrews, D.F., Bickel P.J., Hampel F.R., Huber P.J., Tukey J.W., and Rogers W.H.(1972) *Robust Estimates of Location: Survey and Advances* (Princeton University Press, Princeton NJ).

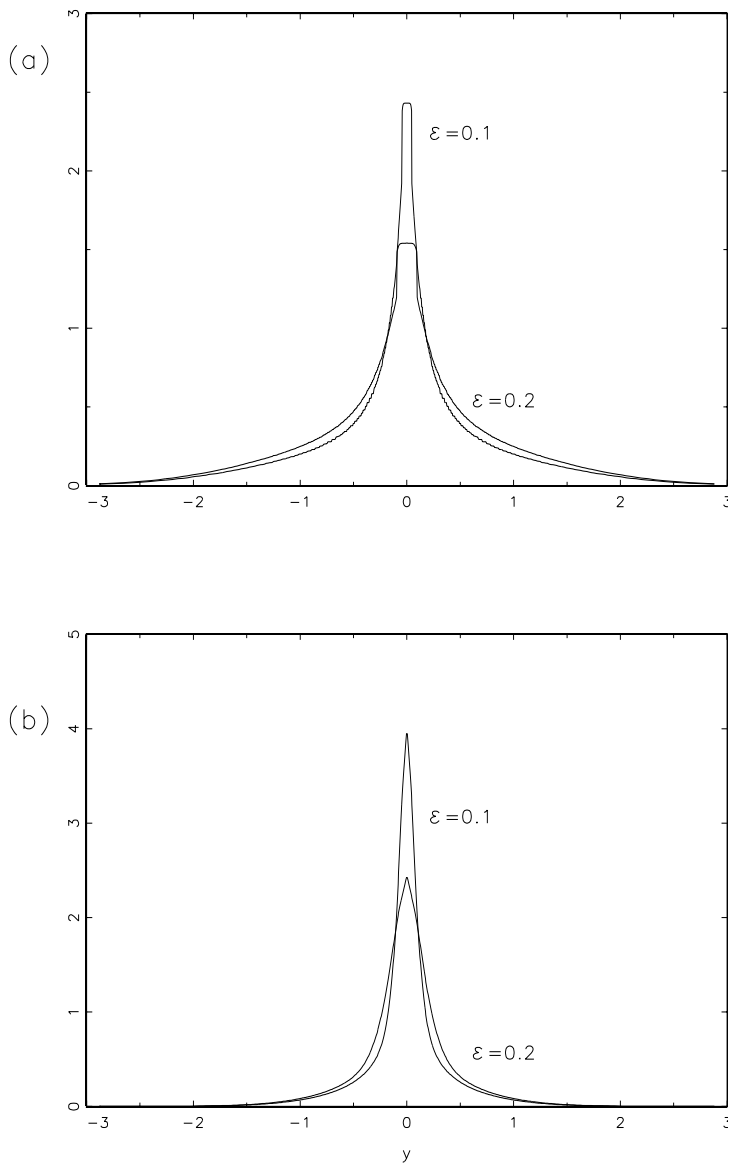

Figure 6: (a) Density of x-marginal distribution of K_ε for $\varepsilon = 0.1$ and $\varepsilon = 0.2$; (b) same for y-marginal distribution.

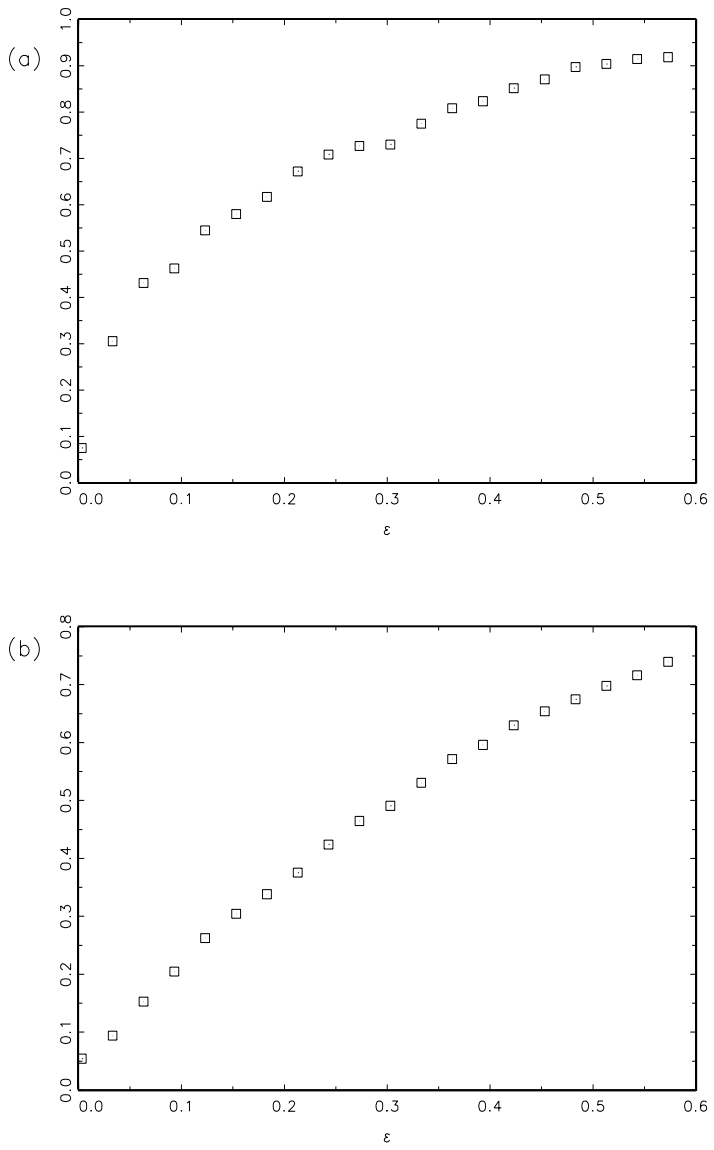

Figure 7: (a) Interquartile range of x-marginal distribution of K_ε plotted versus ε; (b) same for y-marginal distribution.

Cheney, W., and Kincaid, D. (1985) *Numerical Methods and Computing,* Pacific Grove: California.

Frees, E.W. (1991) Trimmed slope estimates for simple linear regression. *J. of Stat. Planning and Inference*, 27, 203–221.

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

Hössjer, O., Rousseeuw, P.J., and Croux, C. (1994) Asymptotics of the repeated median slope estimator. *The Annals of Statistics,* Vol. 22, September issue.

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

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

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

Rousseeuw, P.J., and Leroy, A. (1987) *Robust Regression and Outlier Detection*, New York: John Wiley.

Rousseeuw, P.J., Netanyahu, N.S., and Mount, D.M. (1993). New statistical and algorithmic results on the repeated median line. In *New Directions in Statistical Data Analysis and Robustness,* edited by S. Morgenthaler, E. Ronchetti and W.A. Stahel, Birkhäuser Verlag, Basel, pp. 177-194.

Scholz, F.-W. (1978) Weighted median regression estimates. *Ann. Statist.*, 6, 603–609.

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

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

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.

Tukey, J.W.(1970) *Exploratory Data Analysis* (Limited Preliminary Edition), Reading, Massachusetts: Addison-Wesley.