The Canadian Journal of Statistics
Vol. 22, No. 2, 1994, Pages 233–45
La Revue Canadienne de Statistique

Bounded-influence rank estimation in the linear model*

Douglas WIENS and Julie ZHOU

University of Alberta

Key words and phrases: Robustness, regression, bounded influence, minimax variance, minimum-volume ellipsoid.
AMS 1985 subject classifications: Primary 62F35; secondary 62G05, 62G20.

ABSTRACT

We introduce and study a class of rank-based estimators for the linear model. The estimate may be roughly described as being calculated in the same manner as a generalized *M*-estimate, but with the residual being replaced by a function of its signed rank. The influence function can thus be bounded, both as a function of the residual and as a function of the carriers. Subject to such a bound, the efficiency at a particular model distribution can be optimized by appropriate choices of rank scores and carrier weights. Such choices are given, with respect to a variety of optimality criteria. We compare our estimates with several others, in a Monte Carlo study and on a real data set from the literature.

RÉSUMÉ

Nous présentons et étudions une classe d'estimateurs fondés sur le rang pour le modèle linéaire. Les valeurs estimées peuvent être décrites sommairement comme étant calculées de la même manière qu'un *M*-estimateur généralisé, mais avec le résidu remplacé par une fonction de son rang avec signe. La fonction d'influence peut ainsi être bornée, tant comme fonction du résidu que comme fonction des porteurs. Avec une telle borne, nous pouvons optimiser l'efficience pour une distribution particulière du modèle, en choisissant de façon appropriée les scores des rangs et les poids des porteurs. De tels choix sont donnés pour un éventail de critères d'optimisation. Nous comparons nos valeurs estimées avec celles de plusieurs autres estimateurs, à l'aide d'une étude de Monte Carlo et d'un ensemble de données réelles de la littérature.

1. INTRODUCTION AND SUMMARY

Jaeckel (1972) proposed a class of rank-based estimates, robust against heavy-tailed error distributions, for the linear model

$$y_i = \mathbf{x}_i^{\mathrm{T}}\boldsymbol{\theta}_* + \epsilon_i, \qquad i = 1, \ldots, n. \tag{1.1}$$

Efficiency at an ideal distribution and robustness against alternative distributions may be balanced against each other by appropriate choices of rank scores—see Wang and Wiens (1992). The estimates, however, are not robust against outlying values of the carriers, as is evidenced by the fact that the influence function is unbounded as a function of $\mathbf{x}$.

Tableman (1990) studied a class of *bounded-influence* rank estimates for the model (1.1). She established asymptotic normality for a one-step version of these estimates, and implemented an algorithm to calculate the estimates, using Wilcoxon scores.

*Research supported by the Natural Sciences and Engineering Research Council of Canada.

In Section 2 of this paper, we define a class of one-step, bounded-influence, rank-based estimates which is somewhat broader than that considered by Tableman (1990). This class includes Tableman's estimates and, as a limiting case, Jaeckel's estimates. The influence of the residuals is bounded by appropriate choices of rank scores, and the influence in the factor space is bounded by placing weights on the carriers.

In Section 3 we choose scores generating functions and weight functions which are optimal according to various criteria. In two cases, we minimize the trace of the asymptotic covariance matrix, subject to a bound on the influence function. We do this first without prior restrictions on the weights, and find that weights of the Schweppe type are optimal. We also optimize within the class of Mallows-type weights, since the corresponding estimator is more robust against asymmetric error distributions. We then specialize to carriers with spherically symmetric distributions, and are thereby able to extend the above optimality, in each case, to minimization of the asymptotic covariance matrix itself, with respect to the ordering by positive definiteness.

In all four of these cases, the distributions of the errors and carriers are held fixed. In a final case, for Mallows weights, we obtain scores which minimize the maximum asymptotic covariance matrix, as the distribution of the errors varies over a particular class of distributions.

In Section 4, using the optimality results of Section 3 as a guide, we implement two modified versions of the estimator. They may be roughly described as the optimal estimates under the assumption of spherically distributed carriers, with the following important difference. Rather than applying weights to the carriers themselves, we apply them to the transformed carriers

$$\tilde{\mathbf{x}} = \mathbf{S}^{-\frac{1}{2}}(\mathbf{x} - \mathbf{m}), \tag{1.2}$$

where $\mathbf{S}$ $[= \mathbf{S}^{\frac{1}{2}}(\mathbf{S}^{\frac{1}{2}})^{\mathrm{T}}]$ and $\mathbf{m}$ are the minimum-volume ellipsoid covariance and location estimates (see Rousseeuw and Leroy 1987). In this respect, our estimators are similar to those of Simpson, Ruppert, and Carroll (1992), who study a class of one-step, bounded-influence M-estimates, starting with an initial estimate with high breakdown point. Weights are applied to $\tilde{\mathbf{x}}$ as at (1.2).

We compare our estimates with several others, in a Monte Carlo study. An application to a particular data set is considered in Section 5.

2. BOUNDED-INFLUENCE *R*-ESTIMATION

Let $(\mathbf{x}_i, y_i) \in \mathbb{R}^{p+1}$, $i = 1, \ldots, n$, with empirical distribution function $F_n(\mathbf{x}, y)$, be i.i.d. observations from the model (1.1). For any $\boldsymbol{\theta}$ define

$$e_i = e_i(\boldsymbol{\theta}) = y_i - \mathbf{x}_i^{\mathrm{T}}\boldsymbol{\theta}.$$

Let $R_i = R_i(\boldsymbol{\theta})$ be the signed rank of e_i:

$$R_i = (\text{rank of } |e_i|) \cdot \operatorname{sign}(e_i).$$

Let $v(\mathbf{x}), w(\mathbf{x})$ be nonnegative functions, and let $\psi(t), \gamma(u)$ be nondecreasing, absolutely continuous functions of $t \in \mathbb{R}$ and $u \in (-1, 1)$. (If the error distribution is symmetric, then the optimal choices of ψ and γ turn out to be odd functions of t and u, respectively.) Define

$$\mathbf{S}_n(\boldsymbol{\theta}) = n^{-1} \sum_{i=1}^{n} v(\mathbf{x}_i) \psi\left(\frac{\gamma\big(R_i/(n+1)\big)}{w(\mathbf{x}_i)}\right) \mathbf{x}_i. \tag{2.1}$$

The estimate $\hat{\boldsymbol{\theta}}_n$ is defined to be a one-step (of a modified Newton's method) approximation to the solution of

$$\mathbf{S}_n(\boldsymbol{\theta}) = \mathbf{0} \tag{2.2}$$

starting with a $\sqrt{n}$-consistent estimate of $\boldsymbol{\theta}_*$. See Section 4 for the computational details.

Special cases:

(1) If $\psi(t) = \psi_r(t)$, where

$$\psi_r(t) := \max\big(-r, \min(t, r)\big), \qquad r > 0, \tag{2.3}$$

is Huber's ψ-function, then (2.1), (2.2) define the estimates of Tableman (1990) for a particular choice of $v(\mathbf{x}) = w(\mathbf{x})$. In her implementations, Tableman takes $\gamma(u) = u$ (Wilcoxon scores).

(2) If $\psi(t) = t$, $w = v = 1$, and the errors are symmetrically distributed, then (2.1) and (2.2) define estimates which are asymptotically equivalent to those of Jaeckel (1972), who uses the ranks of the e_i themselves.

Assume now that the carriers and errors are independent, with d.f.'s $H(\mathbf{x}), G(e)$. Let $F(\mathbf{x}, y)$ be the joint d.f. of $\mathbf{x}$ and $y = \mathbf{x}^{\mathrm{T}}\boldsymbol{\theta} + e$. Define

$$\eta_G(\mathbf{x}, e) = v(\mathbf{x})\psi\left(\frac{\gamma\big(G(e) - G(-e)\big)}{w(\mathbf{x})}\right),$$

$$S(\boldsymbol{\theta};\, F) = \mathcal{E}_F[\mathbf{x}\eta_G(\mathbf{x}, y - \mathbf{x}^{\mathrm{T}}\boldsymbol{\theta})].$$

Define a functional $\boldsymbol{\theta}(F)$ by

$$S\big(\boldsymbol{\theta}(F);\, F\big) = \mathbf{0}.$$

[The uniqueness of $\boldsymbol{\theta}(F) = \boldsymbol{\theta}_*$ is shown below.] Then $\boldsymbol{\theta}(F_n)$ is an M-estimate, defined by

$$\begin{aligned} \mathbf{0} = \mathbf{S}(\boldsymbol{\theta};\, F_n) &= n^{-1}\sum_{i=1}^{n} v(\mathbf{x}_i)\psi\left(\frac{\gamma\big(G_n(e_i) - G_n(-e_i)\big)}{w(\mathbf{x}_i)}\right)\mathbf{x}_i \\ &= n^{-1}\sum_{i=1}^{n} v(\mathbf{x}_i)\psi\left(\frac{\gamma\big((R_i - 1)/n\big)}{w(\mathbf{x}_i)}\right)\mathbf{x}_i. \end{aligned} \tag{2.4}$$

The arguments of γ in (2.1) and (2.4) differ by at most $(n+1)^{-1}$. Under appropriate conditions—see, e.g. Bickel (1975) and Maronna and Yohai (1981)—it follows that $\hat{\boldsymbol{\theta}}_n$ and $\boldsymbol{\theta}(F_n)$ are $\sqrt{n}$-equivalent. Standard theory of M-estimation then yields that $\sqrt{n}(\boldsymbol{\theta}_n - \boldsymbol{\theta}_*)$ is asymptotically normally distributed, with mean zero and covariance matrix

$$\mathbf{C} = \mathbf{M}^{-1}\mathbf{Q}\mathbf{M}^{-1}, \tag{2.5}$$

where

$$\mathbf{Q} = \mathcal{E}_F[\mathbf{x}\mathbf{x}^{\mathrm{T}}\eta_G^2(\mathbf{x}, \epsilon)],$$

$$\mathbf{M} = \mathcal{E}_F[\mathbf{x}\mathbf{x}^{\mathrm{T}}\frac{\partial}{\partial\epsilon}\eta_G(\mathbf{x}, \epsilon)].$$

For a direct proof of asymptotic normality, see Tableman (1990). See also Coakley and Hettmansperger (1993).

The influence function of $\boldsymbol{\theta}(F)$ is

$$\mathrm{IF}(\mathbf{x}, y) = \mathbf{M}^{-1}\mathbf{x}\eta_G(\mathbf{x}, \epsilon), \tag{2.6}$$

with $\epsilon = y - \mathbf{x}^{\mathrm{T}}\boldsymbol{\theta}_*$.

The uniqueness of $\boldsymbol{\theta}(F)$ is proven in Zhou (1992) under the following conditions:

(1) G is symmetric;

(2) H is absolutely continuous, and the density of H is symmetric in each of its arguments;

(3) $v(\mathbf{x}), w(\mathbf{x})$ are symmetric in each of their arguments.

If the model contains a constant term, then even if G is not symmetric, the bias may be transferred to the intercept estimate. That is, we have

$$\boldsymbol{\theta}(F) = \boldsymbol{\theta}_* + a(1, 0, \ldots, 0)^{\mathrm{T}}$$

if a satisfies

$$\mathcal{E}_F[\mathbf{x}\eta_G(\mathbf{x}, \epsilon + a)] = \mathbf{0}. \tag{2.7}$$

For *Mallows* weights $[w(\mathbf{x}) = 1]$, (2.7) is clearly possible. *Schweppe* weights $[w(\mathbf{x}) = v(\mathbf{x})]$ generally preclude (2.7)—see the discussion in Carroll and Welsh (1988).

3. OPTIMALITY THEORY

In this section, we assume

(A1) The d.f. $G(e)$ is symmetric [so that $G(e) - G(-e) = 2G(e) - 1$] and strictly increasing, with finite Fisher information $I(G)$.

By (A1) and Theorem 4.2 of Huber (1981), G has an absolutely continuous density g such that

$$I(G) = \int_{-\infty}^{\infty} \left(\frac{-g'}{g}(e)\right)^2 g(e)\, de < \infty,$$

and $g(x) \longrightarrow 0$ as $x \longrightarrow \pm\infty$. Define

$$\xi_g(e) = \frac{-g'}{g}(e),$$

and assume

(A2) $\xi_g(e)$ is absolutely continuous and nondecreasing.

We will exhibit choices of the functions v, w, γ, and ψ which are optimal according to various criteria.

3.1. General Weights.

Recall the form of the asymptotic covariance matrix $\mathbf{C}$ at (2.5). We first consider the problem

(P1) For fixed distributions G and H, choose functions v, w, γ, ψ so as to minimize the trace of $\mathbf{C}$, subject to

$$\sup_{\mathbf{x},y} \|\mathrm{IF}(\mathbf{x}, y)\| \le c. \tag{3.1}$$

It follows from Proposition 2, Section 6.3 of Hampel *et al.* (1986) that a solution to the problem will be attained if η_G can be put into the form [recall (2.3)]

$$\eta_G(\mathbf{x}, e) = \|\mathbf{Ax}\|^{-1}\psi_c\left(\|\mathbf{Ax}\|\xi_g(e)\right), \tag{3.2}$$

where the matrix $\mathbf{A}$ satisfies $\mathbf{A} = \mathbf{M}^{-1}$.

To complete the solution to (P1) we must put η_G into the form (3.2). This is done by choosing

$$v(\mathbf{x}) = w(\mathbf{x}) = \|\mathbf{Ax}\|^{-1}, \qquad \psi = \psi_c, \qquad \gamma(u) = \xi_g\left(G^{-1}\left(\frac{u+1}{2}\right)\right). \tag{3.3}$$

Thus Wilcoxon scores, as used by Tableman (1990), are optimal in this sense if $G =$ logistic; normal scores

$$\gamma(u) = \Phi^{-1}\left(\frac{u+1}{2}\right)$$

are optimal if $G = \Phi$, the normal d.f. Tableman uses weights $v(\mathbf{x}) = w(\mathbf{x})$ different from those given here, as hers were derived with respect to a bound on *self-standardized sensitivity* (Hampel *et al.* 1986, p. 228).

Under additional assumptions, we can attain a much stronger optimality property and a precise existence result. In addition to (A1) and (A2), assume

(A3) The distribution of $\mathbf{x}$ is spherically symmetric.

A consequence of (A3) is that $\mathbf{x}/\|\mathbf{x}\|$ is distributed independently of $\|\mathbf{x}\| =: Z$. Assume also

(A4) The functions v, w depend on $\mathbf{x}$ only through z.

We shall write $v = v(z)$, $w = w(z)$. Define, for $0 < u < 1$,

$$\delta(z, u) = zv(z)\psi\left(\frac{\gamma(2u-1)}{w(z)}\right).$$

Let the r.v. U be uniformly distributed on $(0, 1)$, indepenently of Z. Then from (2.5) and (2.6),

$$\begin{aligned} \mathbf{C} &= \frac{p\mathcal{E}[\delta^2(Z, U)]}{\{\mathcal{E}[Z\delta(Z, U)\xi_g(G^{-1}(U))]\}^2}\mathbf{I}_p, \\ \|\mathrm{IF}(x, y)\| &= \frac{p|\delta(z, G(\epsilon))|}{\mathcal{E}[Z\delta(Z, U)\xi_g(G^{-1}(U))]}. \end{aligned} \tag{3.4}$$

We consider the problem

(P2) Minimize $\mathbf{C}$, *with respect to the ordering by positive definiteness*, subject to (3.1).

Equivalently, we minimize the scalar functional in (3.4). For

$$c \geq \frac{p}{2g(0)\mathcal{E}[Z]} \tag{3.5}$$

there is a solution given by

$$\delta_G(z, u) = \psi_r\left(z\xi_g(G^{-1}(u)\right), \tag{3.6}$$

with $r > 0$ chosen to satisfy (3.1), i.e.

$$\begin{aligned}\frac{p}{c} &= \mathcal{E}[|Z\xi_g(G^{-1}(U))| I\{|Z\xi_g(G^{-1}(U))| > r\}] \\ &\quad + \mathcal{E}[r^{-1}Z^2\xi_g^2(G^{-1}(U)) I\{|Z\xi_g(G^{-1}(U))| \le r\}].\end{aligned} \tag{3.7}$$

If $G = \Phi$, the right-hand side of (3.7) becomes

$$\mathcal{E}[r^{-1}Z^2\{2\Phi(r/Z) - 1\}]. \tag{3.8}$$

Corresponding to (3.6),

$$w(z) = v(z) = z^{-1}, \qquad \psi = \psi_r, \qquad \gamma(u) = \xi_g\left(G^{-1}\left(\frac{u+1}{2}\right)\right). \tag{3.9}$$

To see that (3.6) and (3.7) give the solution, note that by (3.7), δ_G satisfies

$$\mathcal{E}[Z\delta(Z, U)\xi_g(G^{-1}(U))] = pr/c. \tag{3.10}$$

Since any δ may be divided by an arbitrary positive constant without altering the problem, it suffices to consider only those δ's satisfying (3.10) and $|\delta| \le r$. Note that then

$$\mathcal{E}[\{\delta(Z, U) - Z\xi_g(G^{-1}(U))\}] = \mathcal{E}[\delta^2(Z, U)] - \frac{2pr}{c} + I(G)\mathcal{E}[Z^2], \tag{3.11}$$

and so we may instead minimize (3.11). But on each set $z\xi_g(G^{-1}(u)) \in (-\infty, -r)$, $[-r, r]$, (r, ∞) the integrand in the left-hand side of (3.11) is minimized pointwise by δ_G.

The existence of r satisfying (3.7) follows from (3.5) and the fact that the right-hand side of (3.7) varies from 0 to $\mathcal{E}[|Z\xi_g(G^{-1}(U))] = 2g(0)\mathcal{E}[Z]$ as r varies from ∞ to 0.

3.2. Mallows Weights.

As at the end of Section 2, our estimator has more favourable robustness properties, in the presence of asymmetric errors, if Mallows weights are used. In this case (2.1) depends on ψ and γ only through their composition. Define

$$\begin{gathered} J(u) = \psi(\gamma(2u-1)), \qquad 0 < u < 1; \\ \mathbf{M}_v = \mathcal{E}[\mathbf{x}\mathbf{x}^{\mathrm{T}}v(\mathbf{x})], \qquad \mathbf{Q}_v = \mathcal{E}[\mathbf{x}\mathbf{x}^{\mathrm{T}}v^2(\mathbf{x})], \\ \mathbf{C}_v = \mathbf{M}_v^{-1}\mathbf{Q}_v\mathbf{M}_v^{-1}. \end{gathered}$$

The asymptotic covariance matrix then factors as

$$\mathbf{C} = V(J, G)\mathbf{C}_v,$$

where

$$V(J, G) = \frac{\mathcal{E}[J^2(G(\epsilon))]}{\left(\mathcal{E}\left[\frac{d}{d\epsilon}J(G(\epsilon))\right]\right)^2} = \frac{\int_0^1 J^2(u)\,du}{\left[\int_0^1 J(u)\xi_g(G^{-1}(u))\,du\right]^2}.$$

Similarly,

$$\mathrm{IF}(\mathbf{x}, y) = \frac{\mathbf{M}_v^{-1}\mathbf{x}v(\mathbf{x})J\left(G(\epsilon)\right)}{\int_0^1 J(u)\xi_g\left(G^{-1}(u)\right)du}.$$

We consider the problem

(P3) Determine J and v to minimize the trace of $\mathbf{C}$, for fixed G and H, subject to (3.1).

By the above, we can choose J and v separately. We first find J_G minimizing $V(J, G)$, subject to

$$\sup_e \frac{|J(G(e))|}{\int_0^1 J(u)\xi_g\left(G^{-1}(u)\right)du} \le b. \tag{3.12}$$

We then find v minimizing tr $\mathbf{C}_v$ subject to

$$\sup_{\mathbf{x}} \|\mathbf{M}_v^{-1}\mathbf{x}v(\mathbf{x})\| \le c/b. \tag{3.13}$$

We can find an optimal J for

$$b \ge \{2g(0)\}^{-1}.$$

It is given by

$$J_G(u) = \psi_r(\xi_g\left(G^{-1}(u)\right)), \tag{3.14}$$

with $r > 0$ chosen to satisfy (3.12), i.e.

$$\int_0^1 \psi_r(\xi_g\left(G^{-1}(u)\right))\xi_g\left(G^{-1}(u)\right)du = \frac{r}{b}. \tag{3.15}$$

If $G = \Phi$, (3.14) and (3.15) become

$$J_\Phi(u) = \psi_r\left(\Phi^{-1}(u)\right), \qquad b = \frac{r}{2\Phi(r) - 1}. \tag{3.16}$$

See Wang and Wiens (1992) for details.

Although ψ and γ are not determined uniquely by (3.14), a natural choice is

$$\psi = \psi_r, \qquad \gamma(u) = \xi_g\left(G^{-1}\left(\frac{u+1}{2}\right)\right). \tag{3.17}$$

Optimal weights $v(\mathbf{x})$, subject to (3.13), may be obtained as in Section 6.3 of Hampel *et al.* (1986). They are given, for sufficiently large c, by

$$v(\mathbf{x}) = \frac{\psi_{c/b}(\|\mathbf{B}\mathbf{x}\|)}{\|\mathbf{B}\mathbf{x}\|}, \tag{3.18}$$

where $\mathbf{B}$ satisfies $\mathbf{B} = \mathbf{M}_v^{-1}$.

(a)

Under assumption (A3), we can solve

(P4) Minimize $\mathbf{C}_v$ *with respect to the ordering by positive definiteness*, subject to (3.13).

The solution is given by (3.18) and the condition that $\mathbf{B} = k\mathbf{I}_p$ for some $k > 0$. Equivalently,

$$v(\mathbf{x}) = \|\mathbf{x}\|^{-1}\psi_{c/kb}(\|\mathbf{x}\|), \tag{3.19}$$

$$p = k\mathcal{E}[\|\mathbf{x}\|\psi_{c/kb}(\|\mathbf{x}\|)]. \tag{3.20}$$

Equation (3.19) has a solution $k > 0$ as long as

$$\frac{c}{b} > \frac{p}{\mathcal{E}[\|\mathbf{x}\|]}. \tag{3.21}$$

Equality in (3.21) corresponds to $k = \infty$.

(b)

The function J_G at (3.14) is optimal, subject to (3.1), when only infinitesimal deviations from the model are allowed. An alternative approach is to seek a *minimax* solution, i.e. a solution to the problem

(P5) Determine J to minimize the maximum of $V(J, G)$ as G varies over a given class of distributions.

Since $V(J, G)$ is identical to the asymptotic variance functional of an R-estimator of location, the results of Wiens (1990) apply to (P5). In most cases, such applications result in estimators whose influence functions are bounded as functions of the residuals. They may then be bounded as functions of the carriers as in (3.18) or (3.19).

In particular, for the class

$$G(e) = (1-\epsilon)\Phi(e) + \epsilon K(e)$$

of ϵ-contaminated normal distributions, the minimax solution is

$$J_\Phi(u) = \psi_r\left(\Phi^{-1}\left(\frac{1}{2} + \frac{u-\frac{1}{2}}{1-\epsilon}\right)\right), \tag{3.22}$$

with ϵ and r related by

$$\frac{\epsilon}{2(1-\epsilon)} = \frac{\phi(r)}{r} - \Phi(-r).$$

[In (3.22) we take $\Phi^{-1}(t) = \pm\infty$ if $|t| \geq 1$.] See Wang and Wiens (1992) for details of the solution to (P5) in this and other examples.

Corresponding to (3.22) is

$$\psi = \psi_r, \qquad \gamma(u) = \Phi^{-1}\left(\frac{1}{2} + \frac{u}{2(1-\epsilon)}\right).$$

4. MONTE CARLO

We adopt the view that the optimality results of Section 3 should serve as guides in the choice and implementation of estimators, rather than as a set of exact prescriptions.

In particular, the weight functions should be modified so as to make the corresponding estimators affine equivariant. We thus propose to apply the weight functions to $\tilde{\mathbf{x}}$, as in (1.2), rather than to $\mathbf{x}$ itself. If the model has a constant term, then it is to be understood that the first element of $\mathbf{x}$ is removed before (1.2) is applied. Here we follow Simpson, Ruppert, and Carroll (1992), who find in their study of high-breakdown generalized M -estimates that "the $\mathbf{x}$-dependent weights associated with the GM iteration need to be based on high breakdown point location and scatter estimates rather than the customary multivariate M-estimates".

After the transformation (1.2), we anticipate that the vectors $\tilde{\mathbf{x}}$ may resemble a sample from a spherical distribution. We will thus study estimators similar to the solutions to the optimality problems (P2) and (P4) of Section 3.

RS3 *("Schweppe weights")*: This is the three-step estimate $\boldsymbol{\theta}_3$, starting with the least-median-of-squares (LMS) estimate $\boldsymbol{\theta}_0$ [see Rousseeuw (1984) and Rousseeuw and Leroy (1987)]. A step-by-step description of the ensuing algorithm is as follows. Suppose for definiteness that the regression response contains a constant term θ_0. After the application of (1.2), the model becomes $y = \tilde{\mathbf{x}}^{\mathrm{T}}\tilde{\boldsymbol{\theta}} + \epsilon$, where

$$\tilde{\boldsymbol{\theta}} = \mathbf{U}\boldsymbol{\theta} \qquad \text{with} \quad \mathbf{U} = \begin{pmatrix} 1 & \mathbf{m}^{\mathrm{T}} \\ 0 & (S^{\frac{1}{2}})^{\mathrm{T}} \end{pmatrix}.$$

For $k = 0, 1, 2, 3$ define $\mathbf{M}_k = n^{-1} \sum d_{i,k}\mathbf{x}_i\mathbf{x}_i^{\mathrm{T}}$, where

$$d_{i,k} = \frac{v(\mathbf{x}_i)}{w(\mathbf{x}_i)} n^{-1} \sum_{j=1}^{n} \left\{ \psi' \left(\frac{\gamma\left(R_{j,k}/(n+1)\right)}{w(\mathbf{x}_i)} \right) \frac{\hat{g}(y_j - \mathbf{x}_j^{\mathrm{T}}\boldsymbol{\theta}_k)}{\phi(\gamma\left(R_{j,k}/(n+1)\right))} \right\}$$

and $\hat{g}(\cdot)$ is a kernel density estimate [see Silverman (1986) and Wegman (1972)] computed on S-Plus. Let $\tilde{\mathbf{M}}_k$ and $\tilde{d}_{i,k}$ result from replacing $\mathbf{x}_i$ by $\tilde{\mathbf{x}}_i$ in these definitions. Note that the ranks $R_{j,k} = R_j(\boldsymbol{\theta}_k)$ are invariant under the transformations $(\mathbf{x}, \boldsymbol{\theta}) \longrightarrow (\tilde{\mathbf{x}}, \tilde{\boldsymbol{\theta}})$. The iterates are now defined by $\tilde{\boldsymbol{\theta}}_{k+1} = \tilde{\boldsymbol{\theta}}_k + \tilde{\mathbf{M}}_k^{-1}\tilde{\mathbf{S}}_n(\tilde{\boldsymbol{\theta}}_k)$, where $\tilde{\mathbf{S}}_n(\boldsymbol{\theta})$ is given by (2.1) with $\psi = \psi_{1.5}$, $\gamma(u) = \Phi^{-1}((u+1)/2)$, $v(\mathbf{x}) = w(\mathbf{x}) = \|\mathbf{x}\|^{-1}$, and with $\mathbf{x}$ replaced by $\tilde{\mathbf{x}}$ throughout. The asymptotic covariance matrix $\tilde{\mathbf{C}}$ of $\tilde{\boldsymbol{\theta}}_3$ is estimated by $\tilde{\mathbf{C}}_3 = \tilde{\mathbf{M}}_3^{-1}\tilde{\mathbf{Q}}_3\tilde{\mathbf{M}}_3^{-1}$, where

$$\tilde{\mathbf{Q}}_3 = n^{-1} \sum_{i=1}^{n} \left\{ v^2(\tilde{\mathbf{x}}_i)\psi^2 \left(\frac{\gamma\left(R_{i,3}/(n+1)\right)}{w(\tilde{\mathbf{x}}_i)} \right) \right\} \tilde{\mathbf{x}}_i\tilde{\mathbf{x}}_i^{\mathrm{T}}.$$

The covariance matrix $\mathbf{C}$ of $\boldsymbol{\theta}_3$ is then estimated by $\mathbf{C}_3 = \mathbf{U}^{-1}\tilde{\mathbf{C}}_3\mathbf{U}^{-\mathrm{T}}$.

RM3 *("Mallows weights")*: This estimate is computed in the same manner as RS3, but with

$$w(\tilde{\mathbf{x}}) = 1, \qquad v(\tilde{\mathbf{x}}) = \frac{\psi_{1.6}(\|\tilde{\mathbf{x}}\|)}{\|\tilde{\mathbf{x}}\|}.$$

The rank scores (3.14) are given by

$$J(u) = \psi_{1.5} \left(\Phi^{-1} \left(\frac{1+u}{2} \right) \right).$$

Following a suggestion in Tableman (1990), the choices $r = 1.5$, and $c/b = 1.6$ are made in order that the average weight

$$\text{aver} \left\{ v(\tilde{\mathbf{x}}_i)\psi_r \left(\frac{\gamma\left(R_i/(n+1)\right)}{w(\tilde{\mathbf{x}}_i)} \right) \bigg/ \gamma\left(\frac{R_i}{n+1} \right) \right\}$$

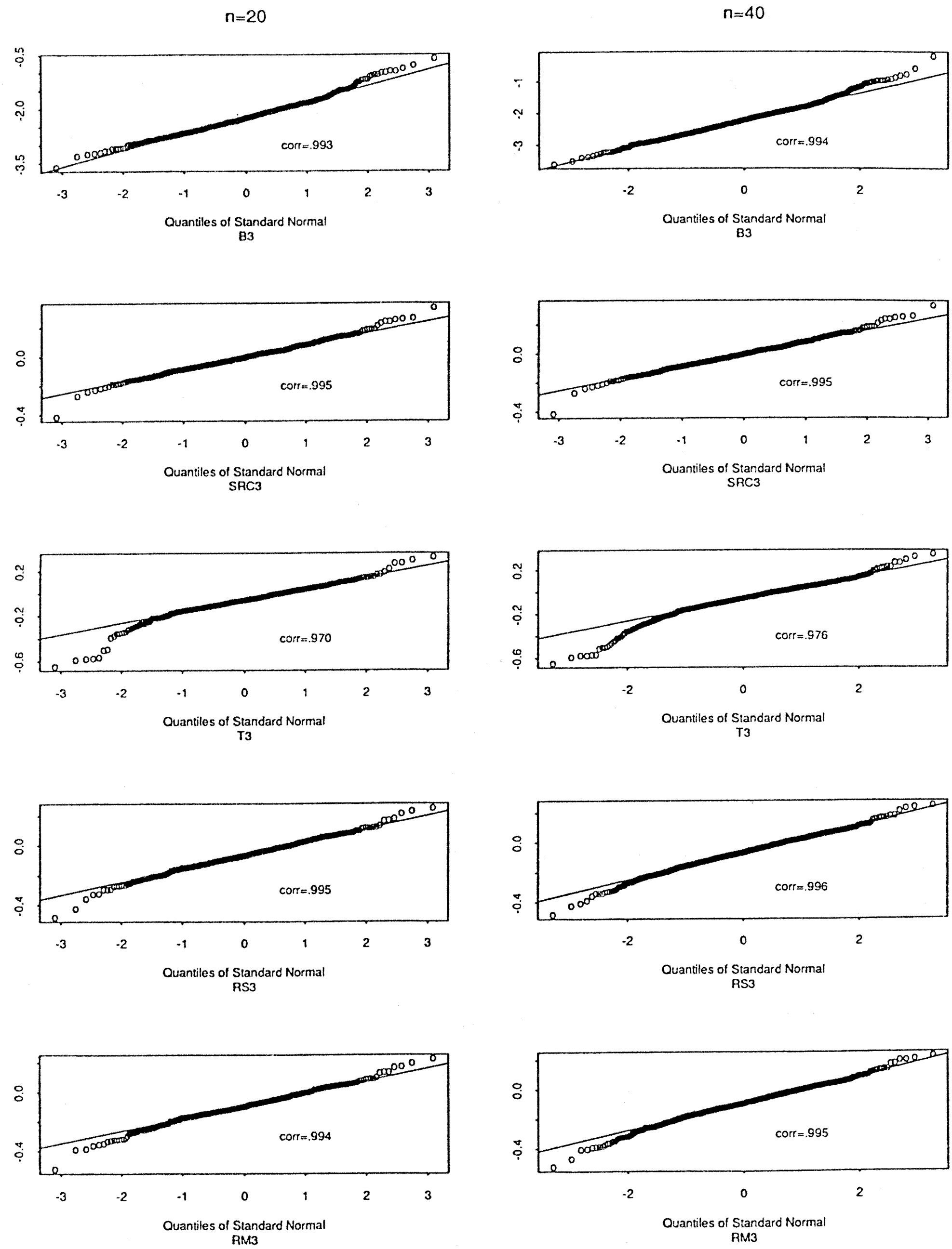

FIGURE 1: $(\hat{\theta}_0 - \theta_0) + (\hat{\theta}_1 - \theta_1)$ vs. normal quantiles: 1000 simulated samples.

be about 0.95 in each case.

We compare our estimates RS3 and RM3 with:

T3: The three-step estimate of Tableman (1990), with Wilcoxon scores and tuning constants as described there.

SRC3: The "Newton-Raphson", "nonexchangeable" version of the Simpson-Ruppert-

TABLE 1: Biases and standard deviations of $\hat{\theta}_0, \hat{\theta}_1$.

	Bias				Standard deviation			
	θ_0		θ_1		θ_0		θ_1	
Estimate	$n=20$	$n=40$	$n=20$	$n=40$	$n=20$	$n=40$	$n=20$	$n=40$
B3	0.188	0.021	2.03	0.306	0.518	0.073	2.037	0.384
LMS	−0.0015	0.032	−0.0008	−0.0025	0.095	0.141	0.108	0.063
SRC3	−0.0027	−0.00052	−0.0002	0.00031	0.061	0.038	0.120	0.040
T3	0.015	−0.027	0.050	0.0048	0.097	2.35	0.088	1.21
RS3	0.0024	0.00079	0.065	0.028	0.068	0.042	0.094	0.050
RM3	0.0056	0.0020	0.089	0.040	0.069	0.042	0.112	0.058

TABLE 2: Coverage properties of 90% confidence intervals on $\hat{\theta}_0 + \hat{\theta}_1$.

	Coverage		Mean width		Lower[a]	
Estimate	$n=20$	$n=40$	$n=20$	$n=40$	$n=20$	$n=40$
B3	0	2.1	1.50	0.230	100	97.9
T3	75.2	32.6	0.291	0.282	21.6	35.7
SRC3	84.0	86.4	0.264	0.171	8.4	6.4
RS3	86.0	91.2	0.336	0.225	13.4	8.0
RM3	78.2	85.7	0.322	0.214	21.2	13.7

	Upper[b]		Estimated s.d.[c]		Sample s.d.[d]	
Estimate	$n=20$	$n=40$	$n=20$	$n=40$	$n=20$	$n=40$
B3	0	0	0.457	0.069	0.451	0.255
T3	3.2	31.7	0.088	0.085	0.120	3.27
SRC3	7.6	7.2	0.080	0.052	0.140	0.056
RS3	0.6	0.8	0.102	0.068	0.096	0.059
RM3	0.6	0.6	0.097	0.065	0.096	0.059

[a]Percentage of times that the entire interval lay below $\theta_0 + \theta_1$.
[b]Percentage of times that the entire interval lay above $\theta_0 + \theta_1$.
[c]Standard deviation of $\hat{\theta}_0 + \hat{\theta}_1$, as computed from the asymptotic covariance matrix.
[d]Sample standard deviation of the 1000 estimates of $\theta_0 + \theta_1$.

TABLE 3: Parameter estimates and standard errors for the example of Section 5.

	Using the whole data set		Excluding observation 5	
Variable	RS3	RM3	RS3	RM3
AC	0.0184(0.0020)	0.0182(0.0023)	0.0176(0.0011)	0.0178(0.0012)
FR	0.0032(0.0012)	0.0031(0.0011)	0.0026(0.00073)	0.0027(0.00061)
UR	0.148 (0.042)	0.153 (0.0104)	0.168 (0.0228)	0.162 (0.0129)
Other	0.0218(0.0078)	0.0220(0.0070)	0.0244(0.0042)	0.0240(0.0036)

Carroll estimator, with tuning constants as described in Section 5 of Simpson, Ruppert, and Carroll (1992).

B3: Bickel's (1975) three-step Huber estimate, using $\psi = \psi_{1.345}$.

LMS: Least median of squares.

For T3, as with RS3 and RM3, we calculate three-step estimates starting with LMS. The iteration procedure for T3 is given in Tableman (1990). For SRC3 and B3 three steps are used, starting with LMS and the minimum-L_1-norm estimate, respectively.

We simulated 1000 samples of size $n = 20$, and 1000 samples of size $n = 40$. Outlying observations and high leverages were modelled by sampling from two data structures. In structure 1, $Y = 2 + 2X + \epsilon$, with $X \sim \mathrm{N}(0, 1)$ and $\epsilon \sim 0.9\mathrm{N}(0, \sigma = 0.2) + 0.1\mathrm{N}(0, 1)$. In structure 2, $X \sim \mathrm{N}(10, 0.2)$, $Y \sim \mathrm{N}(0, 0.2)$. For $n = 20$, 90% of the observations were drawn from structure 1, with the remaining 10% drawn from structure 2. For $n = 40$ these were instead 95% and 5% respectively. The results are shown in Figure 1.

Table 1 gives the biases, and sample standard deviations, of the various estimates. In Table 2 the coverage properties of nominal 90% confidence intervals on $\theta_0 + \theta_1$ are assessed. The intervals are based on the asymptotic normality of the estimates.

From Tables 1 and 2 it can be seen that the proposed estimator RS3 generally outperforms RM3 and, together with SRC3, is a candidate for the best overall performer. In turn, RM3 seems preferable to T3 by most measures. The estimate B3, as expected, fared poorly in the presence of high leverage points. Equally predictably, LMS fared very well in this situation. Note that LMS converges only at the rate $n^{-\frac{1}{3}}$, to a nonnormal limit, and so confidence intervals were not constructed from it.

5. AN EXAMPLE

We analyze a data set collected by Haith (1976) relating land use to water quality, and previously analyzed in Simpson, Ruppert, and Carroll (1992). The data have 20 observations on five variables: the response *nitrogen concentration* (N), and four independent variables representing land use given as a percentage of total land use: *active agriculture* (AC), *forest, brushland, or plantation* (FR), *residential* (RS), and *commercial or industrial* (CI). As in Simpson, Ruppert, and Carroll (1992), we put UR = RS + CI and fit the model $\mathrm{N} = \theta_1 \cdot \mathrm{Other} + \theta_2 \cdot \mathrm{AC} + \theta_3 \cdot \mathrm{FR} + \theta_4 \cdot \mathrm{UR} + \mathrm{error}$, where $\mathrm{Other} = 100 - \mathrm{AC} - \mathrm{FR} - \mathrm{UR}$. Observation 5 (the Hackensack River) is a severe design outlier; analyses are presented both with and without this observation.

The estimates RS3 and RM3 of Section 4 were computed exactly as described there. For RS3 and RM3 we took $r = 3.8$ and $c/b = 5.5$. The parameter estimates and their estimated standard errors are reported in Table 3, and should be compared with those for SRC3, B3, LMS, and OLS (ordinary least squares) as reported by Simpson, Ruppert, and Carroll. It is seen that the parameter estimates using RS3 and RM3 are not greatly influenced by the presence or absence of observation 5. The numbers indicate that the new rank-based estimates should be routinely considered in robust analyses of possibly contaminated data, either as competitors or as supplements to the usual M and GM estimators.

REFERENCES

Bickel, P.J. (1975). One-step Huber estimates in the linear model. *J. Amer. Statist. Assoc.*, 20, 428–434.

Carroll, R.J., and Welsh, A.H. (1988). A note on asymmetry and robustness in linear regression. *Amer. Statist.*, 42, 285–287.

Coakley, C.W., and Hettmansperger T.P. (1993). A bounded influence, high breakdown, efficient regression estimator. *J. Amer. Statist. Assoc.*, 88, 872–880.

Haith, D.A. (1976). Land use and water quality in New York rivers. *J. Environmental Engrg. Div. Proc. Amer. Soc. Civil Engrs.*, 102, 1–15.

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.

Huber, P.J. (1981). *Robust Statistics*. Wiley, New York.

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

Krasker, W.S. (1980). Estimation in linear regression models with disparate data points. *Econometrica*, 48, 1333–1346.

Maronna, R.A., and Yohai, V.J. (1981). Asymptotic behaviour of general *M*-estimates for regression and scale with random carriers. *Z. Wahrsch. Verw. Geb.*, 58, 7–20.

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

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

Silverman, B.W. (1986). *Density Estimation for Statistics and Data Analysis.* Chapman and Hall, London.

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

Tableman, M. (1990). Bounded-influence rank regression: A one-step estimator based on Wilcoxon scores. *J. Amer. Statist. Assoc.*, 85, 508–513.

Wang, Y., and Wiens, D. (1992). Optimal, robust *R*-estimators and test statistics in the linear model. *Statist. Probab. Lett.*, 14, 179–188.

Wegman, E.G. (1972). Nonparametric probability density estimation. *Technometrics*, 14, 533–546.

Wiens, D.P. (1990). Minimax-variance *L*- and *R*-estimators of location. *Canad. J. Statist.*, 18, 47–58.

Zhou, J. (1992). Bounded influence *R*-estimators in the linear model. M.Sc. Thesis, University of Alberta.

Received 5 January 1993
Revised 16 July 1993
Accepted 7 January 1994

Department of Statistics and Applied Probability
University of Alberta
Edmonton, Alberta
Canada T2G 2G1