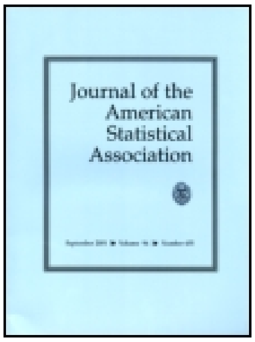

Journal of the American Statistical Association

ISSN: 0162-1459 (Print) 1537-274X (Online) Journal homepage: http://www.tandfonline.com/loi/uasa20

The Least Trimmed Differences Regression Estimator and Alternatives

Arnold J. Stromberg , Ola Hössjer & Douglas M. Hawkins

The Least Trimmed Differences Regression Estimator and Alternatives

Arnold J. STROMBERG, Ola HÖSSJER, and Douglas M. HAWKINS

This article proposes and studies the performance in theory and practice of the least trimmed differences (LTD) linear regression estimator. The estimator minimizes the sum of the smallest quartile of the squares of the differences in each pair of residuals. We obtain the breakdown point, maxbias curve, and large-sample properties of a class of estimators including the LTD as special case. The LTD estimator has a 50% breakdown point and Gaussian efficiency of 66%—substantially higher than other common high-breakdown estimators such as least median of squares and least trimmed squares. The LTD estimator is difficult to compute, but can be performed using a "feasible solution" algorithm. Half-sample jackknifing is effective in producing standard errors. In simulations we find the LTD to be more stable than other high-breakdown estimators. In an example, the LTD still shows instability like other high-breakdown estimators when there are small changes in the data.

KEY WORDS: Robust high breakdown efficient estimation.

1. INTRODUCTION

We consider the linear regression model given by

$$y_i = \alpha_0 + \beta_0^T x_i + \varepsilon_i \qquad 1 \le i \le n,$$

where α_0 is the intercept parameter and β_0 is the $p-1$-dimensional slope parameter. The x_i are random with distribution G and independent of the ε_i, which are independent and identically distributed with distribution F that is not necessarily symmetric. The residuals are denoted by $r_i(\alpha, \beta) = y_i - \alpha - \beta^T x_i$. The classical least squares (LS) estimator minimizes the sum of the squared residuals. It is not robust in the sense that it is heavily influenced by outliers. In fact, it has a breakdown point (Donoho and Huber 1983) of $1/n$, meaning that altering one point can drive the estimator to infinity. One way to robustify the LS estimate is to use an M estimate (Huber 1964), which minimizes

$$\sum_{i=1}^{n} \rho(r_i(\alpha, \beta)),$$

where ρ is a symmetric continuously differentiable function for which $\rho(0) = 0$. With a suitable choice of ρ, M estimators and related estimators can be effective in reducing or eliminating the influence of outliers. One drawback of M estimators is that ρ must be chosen by the practitioner. As an alternative without that drawback, Rousseeuw (1983, 1984) introduced the least median of squares (LMS) estimator, which minimizes the median of the squared residuals, and the least trimmed squares (LTS) estimator, which minimizes the sum of the smallest half of the squared residuals. These estimators have a breakdown point of almost 50% in most situations. Thus they handle outliers well but have drawbacks; in particular, poor efficiencies for Gaussian data. LMS has asymptotic efficiency of 0%, and LTS has asymptotic efficiency of only 8%. Croux, Rousseeuw, and Hössjer (1994) overcame this difficulty by proposing the least quartile difference (LQD) estimator, which minimizes the lower quartile of the ordered absolute differences in residual pairs; that is,

$$\hat{\beta}_{\mathrm{LQD}} = \underset{\beta}{\operatorname{argmin}}\{|r_i - r_j|; i < j\}_{\binom{h_p}{2}:\binom{n}{2}}, \tag{1}$$

where $h_p = [(n+p+1)/2]$, p is the number of regression parameters, and the notation $\binom{h_p}{2} : \binom{n}{2}$ means minimize the the $\binom{h_p}{2}$th order statistic among the $\binom{n}{2}$ elements of the set $\{|r_i - r_j|; i < j\}$.

The LQD is a 50% breakdown estimator with Gaussian efficiency of 67%. Note that the LQD does not involve the intercept parameter α, and so it must be estimated separately. We propose instead the least trimmed difference (LTD) estimator, which minimizes the sum of the smallest quartile of the squared differences of residual pairs; that is,

$$\hat{\beta}_{\mathrm{LTD}} = \underset{\beta}{\operatorname{argmin}} \sum_{i=1}^{\binom{h_p}{2}:\binom{n}{2}} \{(r_i - r_j)^2; i < j\}. \tag{2}$$

Section 2 studies the theoretical properties of the LTD estimator. We prove that the LTD estimator has a 50% breakdown point and is asymptotically normal with Gaussian efficiency of 66%. Section 3 discusses methods for computing the LTD estimator as well as the LQD, LMS, and LTS estimators and some simulation results. Section 4 presents an example and provides some closing remarks.

2. THEORETICAL PROPERTIES OF THE ESTIMATOR

2.1 Breakdown Point

We use the finite-sample version of the breakdown point introduced by Donoho and Huber (1983). Let $\mathbf{Z}$ denote the original sample $\{z_i = (x_i, y_i)\}$ of data points and let $|\cdot|$ denote the Euclidean norm.

Arnold J. Stromberg is Associate Professor, Department of Statistics, University of Kentucky, Lexington, KY 40506 (E-mail: *astro@ms.uky.edu*). Ola Hössjer is Professor, Center of Mathematics, Lund University 5-221 00 Lund, Sweden (E-mail: *ola@maths.lth.se*). Douglas M. Hawkins is Professor, School of Statistics, University of Minnesota, Minneapolis, MN 55455 (E-mail: *doug@stat.umn.edu*).

Journal of the American Statistical Association
September 2000, Vol. 95, No. 451, Theory and Methods

Definition. The breakdown point of the estimator T at sample $\mathbf{Z}$ is defined as

$$\varepsilon_n^*(T,\mathbf{Z}) = \min\left\{\frac{m}{n}; \sup_{\mathbf{Z}'} |T(\mathbf{Z}') - T(\mathbf{Z})| = \infty\right\},$$

where the supremum is taken over all samples $\mathbf{Z}'$, defined by replacing m observations in $\mathbf{Z}$.

Notice that $r_i - r_j = (y_i - y_j) - \boldsymbol{\beta}^T(x_i - x_j)$ can be interpreted as a residual when the pairwise y differences are regressed against the corresponding x differences. The breakdown point of $\hat{\boldsymbol{\beta}}_{\text{LTD}}$ can be affected if many $x_i - x_j$ are linearly dependent. To this end, we impose the following.

Definition. We say that the differences of the x_i are in *general position* if no $\binom{p-1}{2} + 1$ of them belong to the same hyperplane in $\mathbf{R}^{p-1}$ through the origin.

This definition is slightly stronger than the one given by Croux et al. (1994), that only required that no set of $\binom{p}{2}$ differences belong to the same hyperplane. Both definitions imply that no set of p x_i lies in the same affine hyperplane.

Consider now the objective function

$$D_n(\boldsymbol{\beta}) = \binom{n}{2}^{-1} \sum_{l=1}^{k} \{(r_i - r_j)^2; i < j\}_{l:\binom{n}{2}} \tag{3}$$

with trimming point $k = k_n$, and let

$$q(j) = \min\left\{i; \binom{i}{2} \geq j\right\}$$

be the inverse function of the binomial coefficients. The following theorem states how ε_n^* depends on k.

Theorem 1. The breakdown point of the LTD estimator is given by

$$\varepsilon_n^*(\hat{\boldsymbol{\beta}}_{\text{LTD}}, \mathbf{Z}) = \min(q(k) - p + 1, n + 1 - q(k))/n,$$

if the differences $x_i - x_j$ are in general position. In particular, if $k_n/\binom{n}{2} \to \gamma \in [0,1]$ as $n \to \infty$, then

$$\varepsilon^* = \lim_{n\to\infty} \varepsilon_n^* = \min(\sqrt{\gamma}, 1 - \sqrt{\gamma}).$$

Remark 1. Let $r_{ij} = r_i - r_j$ and put $|r_{\cdot\cdot}|_{k:\binom{n}{2}} = \{|r_{ij}|; i < j\}_{k:\binom{n}{2}}$. Theorem 1 applies to any objective function such that $g_1(|r_{\cdot\cdot}|_{k:\binom{n}{2}}) \leq D_n \leq g_2(|r_{\cdot\cdot}|_{k:\binom{n}{2}})$, where g_1 and g_2 are strictly increasing with $g_1(0) = g_2(0) = 0$ and $g_1(\infty) = g_2(\infty) = \infty$. This includes the LQD estimator $D_n = |r_{\cdot\cdot}|_{k:\binom{n}{2}}$, the LS estimator $D_n = \sum_{i<j}(r_i - r_j)^2, k = \binom{n}{2}$, and the rank-based estimators considered by Croux et al. (1994), $D_n = \sum_i a(i)|r_{\cdot\cdot}|_{i:\binom{n}{2}}$, with a a nonnegative function and $k = \max\{i; a(i) > 0\}$.

Remark 2. For $q(k) = [(n+p)/2]$ or $[(n+p+1)/2]\varepsilon_n^*$ attains the maximal value $([(n-p)/2]+1)/n$, which is also the maximal breakdown point among all regression equivariant estimators (Rousseeuw and Leroy 1987). The maximum is attained for

$$\binom{[\frac{n+p}{2} - 1] - 1}{2} < k \leq \binom{[\frac{n+p+1}{2}]}{2} = \binom{h_p}{2}.$$

This is a slightly larger range of k values than that given by Croux et al. (1994), due to the fact that our definition of general position is stronger.

Remark 3. Theorem 1 is an analog of a result of Hössjer (1994), who proved that $\varepsilon_n^* = \min(k - p + 1, n + 1 - k)/n$ for objective functions with $g_1(|r_\cdot|_{k:n}) \leq D_n \leq g_2(|r_\cdot|_{k:n})$ and $|r_\cdot|_{k:n} = \{|r_i|, 1 \leq i \leq n\}_{k:n}$.

2.2 Maximum Bias Curves

We first develop a functional version of a class of estimators including the LTD estimator. We refer to work of Martin, Yohai, and Zamar (1989) and Croux et al. (1994), where the maxbias functions of S and generalized S (GS) estimators are computed. Define the kernel function $h(z_1, z_2; \boldsymbol{\beta}) = (r_1(\alpha,\boldsymbol{\beta}) - r_2(\alpha,\boldsymbol{\beta}))^2 = (\varepsilon_1 - \varepsilon_2 - (\boldsymbol{\beta} - \boldsymbol{\beta}_0)^T(x_1 - x_2))^2$. Let $K = \mathcal{L}((X, \boldsymbol{\beta}_0^T X + \varepsilon))$ be the distribution of each z_i and define the distribution function $H_K(t;\boldsymbol{\beta}) = P_{K\times K}(h(Z_1, Z_2; \boldsymbol{\beta}) \leq t)$. Consider then (a functional version of) the objective function

$$D(\boldsymbol{\beta}; K) = \int_0^1 J(u) H_K^{-1}(u;\boldsymbol{\beta})\,du, \tag{4}$$

with $H_K^{-1}(u;\boldsymbol{\beta}) = \inf\{t; H_K(t;\boldsymbol{\beta}) \geq u\}$. Here $J(u)$ is the density (which may contain, e.g., delta functions) of a positive measure μ on [0, 1], so that $J(u)du = d\mu(u)$. Let $\gamma = \inf\{u \in (0,1]; \mu(u,1] = 0\}$. For instance, the LTD estimator corresponds to $J(u) = I(0 \leq u \leq \gamma)$, and the LQD estimator corresponds to $J(u) = \delta_\gamma(u)$, a delta function located at γ. Minimizing D yields (a functional version of) our estimator,

$$T(K) = \operatorname*{argmin}_{\boldsymbol{\beta}} D(\boldsymbol{\beta}; K).$$

Let K_0 be the nominal distribution of z, and let G_0 and F_0 be the corresponding nominal distributions of x and ε. Consider the ε-contamination neighborhood $V_\varepsilon = \{K; K = (1-\varepsilon)K_0 + \varepsilon K^*\}$, where K^* is arbitrary. The maxbias function is defined as

$$B_\varepsilon(T) = \sup\{|T(K) - \boldsymbol{\beta}_0|; K \in V_\varepsilon\}$$

with asymptotic breakdown point $\varepsilon^* = \inf\{\varepsilon; B_\varepsilon(T) = \infty\}$. Suppose that the following conditions hold:

A. F_0 has a continuous and symmetric density, which is strictly decreasing on the positive real line.

B. G_0 is spherical, $P_{G_0}(\boldsymbol{\beta}^T X = 0) = 0$ for all $\boldsymbol{\beta} \neq 0$ in $\mathbf{R}^{p-1}$ and for all $\boldsymbol{\beta}$ the distribution of $\boldsymbol{\beta}^T X$ is unimodal.

Put

$$g_1(\varepsilon) := \sup_{K\in V_\varepsilon} D(\boldsymbol{\beta}_0; K) = \int J(u) H_1^{-1}(u;\varepsilon)\,du$$

and

$$g_2(\varepsilon,\boldsymbol{\beta}) := \inf_{K\in V_\varepsilon} D(\boldsymbol{\beta}; K) = \int J(u) H_2^{-1}(u;\varepsilon,\boldsymbol{\beta})\,du$$

with $H_1(t;\varepsilon) = (1-\varepsilon)^2 H_{K_0}(t;\boldsymbol{\beta}_0), H_2(t;\varepsilon,\boldsymbol{\beta}) = (1-\varepsilon)^2 H_{K_0}(t;\boldsymbol{\beta}) + 2\varepsilon(1-\varepsilon)\tilde{H}_K(t;\boldsymbol{\beta}) + \varepsilon^2$, and $\tilde{H}_K(t;\boldsymbol{\beta}) = P_K((\varepsilon_1 - \boldsymbol{\beta}^T X)^2 \le t)$. Loosely speaking, H_1 describes the (stochastically) largest possible choice of $H_K(\cdot;\boldsymbol{\beta}_0)$ within V_ε, whereas H_2 is the (stochastically) smallest choice of $H_K(\cdot;\boldsymbol{\beta})$. Note that H_1 is substochastic, whereas H_2 is a proper distribution function. Because G_0 is spherically symmetric, g_2 is only a function of $|\boldsymbol{\beta}-\boldsymbol{\beta}_0|$, and this function is continuous and strictly increasing (provided that $\varepsilon^2 < \gamma$). This follows because $\mathcal{L}(\varepsilon_1 - \varepsilon_2)$ satisfies A and $\mathcal{L}(X_1 - X_2)$ satisfies B. Thus we may define $b = g_2^{-1}(\varepsilon, a)$ as the inverse of g_2, in the sense that $g_2(\varepsilon,\boldsymbol{\beta}) = a$ when $|\boldsymbol{\beta}-\boldsymbol{\beta}_0| = b$.

Theorem 2. Suppose that conditions A and B hold. Then the maxbias function is given by

$$B_\varepsilon(T) = \begin{cases} g_2^{-1}(\varepsilon, g_1(\varepsilon)), & 0 \le \varepsilon \le \min(\sqrt{\gamma}, 1-\sqrt{\gamma}) \\ \infty, & \min(\sqrt{\gamma}, 1-\sqrt{\gamma}) < \varepsilon \le 1. \end{cases} \quad (5)$$

In particular, $\varepsilon^* = \min(\sqrt{\gamma}, 1-\sqrt{\gamma})$.

Consider the multivariate Gaussian case $(X,\varepsilon) \sim \mathrm{N}_p(0,I)$. Put $H_0 \sim \chi^2(1)$ and $g_0(t) = \int_0^{H_0(t)} J(u) H_0^{-1}(u)\, du$. Then

$$g_1(\varepsilon) = 2(1-\varepsilon)^2 g_0\left(H_0^{-1}\left(\frac{\gamma}{(1-\varepsilon)^2}\right)\right),$$

$$g_2(\varepsilon,\boldsymbol{\beta}_0) = 2(1-\varepsilon)^2 g_0\left(\frac{H_2^{-1}(\gamma;\varepsilon,0)}{2}\right) + 2\varepsilon(1-\varepsilon) g_0(H_2^{-1}(\gamma;\varepsilon,0)),$$

and

$$g_2(\varepsilon,\boldsymbol{\beta}) = (1+|\boldsymbol{\beta}-\boldsymbol{\beta}_0|^2) g_2(\varepsilon,\boldsymbol{\beta}_0),$$

so that

$$B_\varepsilon(T) = \sqrt{\frac{g_1(\varepsilon)}{g_2(\varepsilon,\boldsymbol{\beta}_0)} - 1}. \quad (6)$$

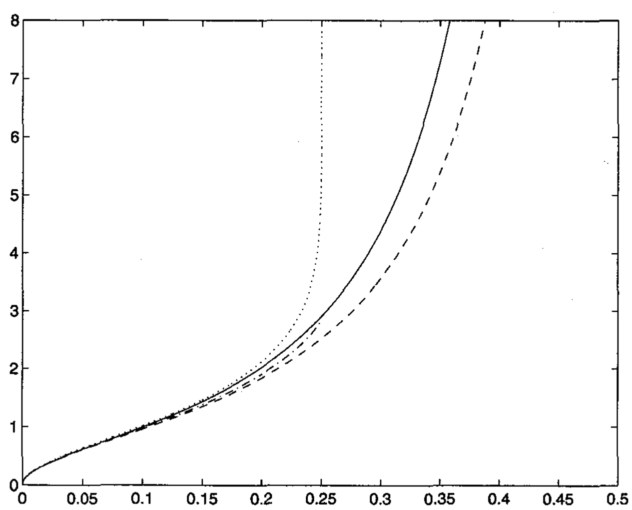

Figure 1. Bias Curves, for Multivariate Gaussian Data, of the LTD Estimator With $\varepsilon^* = .5$ (Solid Line) and $\varepsilon^* = .25$ (Dashed-Dotted Line), and the LQD Estimator With $\varepsilon^* = .5$ (Dashed Line) and $\varepsilon^* = .25$ (Dotted Line). The x-axis is ε; the y-axis is $B_\varepsilon(T)$.

Figure 1 plots the maxbias curve of the LTD and LQD estimators for multivariate Gaussian data and $\varepsilon^* = .25, .5$. Note that the LQD bias explodes when $\varepsilon \to \varepsilon^* = .25$, but this is not the case for the LTD estimator.

The modified gross error sensitivity, (Yohai and Zamar 1992), is given by

$$\gamma^{**} = \lim_{\varepsilon\to 0} \frac{B_\varepsilon(T)}{\sqrt{\varepsilon}} = \sqrt{\frac{g_1'(0) - g_2'(0,\boldsymbol{\beta}_0)}{g_1(0)}},$$

with g_2' the derivative of g_2 with respect to its first argument. For the last equality, we used $g_1(0) = g_2(0,\boldsymbol{\beta}_0)$. Specializing to LTD estimators, we get

$$\gamma^{**} = \sqrt{\frac{(c^2-1)(2\Phi(c)-1) + 2c\phi(c)}{\gamma - \sqrt{2}c\phi(c/\sqrt{2})}}$$

with Φ the cdf of a standard normal distribution and $c = \sqrt{2}\Phi^{-1}((1+\gamma)/2)$. In particular, with $\varepsilon^* = .5$, we have $\gamma^{**} = 2.3907$. This is almost the same as the value of 2.3992 found by Croux et al. (1994) for the LQD.

Remark 4. Instead of A and B, we may simply assume that $P_{G_0}(\boldsymbol{\beta}^T X = 0) = 0$ when $\boldsymbol{\beta} \ne 0$. The definitions of g_1 and g_2 are the same, but the infimum of $D(\boldsymbol{\beta};K)$ over V_ε may not be attained at $H_K = H_2$. Moreover, $g_2(\varepsilon,\cdot)$ is generally not an (increasing) function of $|\boldsymbol{\beta}-\boldsymbol{\beta}_0|$. However, Theorem 2 still holds with $g_2^{-1}(\varepsilon,a) = \sup\{b;$ there exists $\boldsymbol{\beta}$ such that $g_2(\varepsilon,\boldsymbol{\beta}) \le a$ and $|\boldsymbol{\beta}-\boldsymbol{\beta}_0| = b\}$.

Remark 5. A GS estimator has an implicit objective function $D(\boldsymbol{\beta};K) = S(H_K(\cdot;\boldsymbol{\beta}))$. Here $S(H)$ is the solution of $\int \tilde{\rho}(t/S)\, dH(t) = k_0$, with $\tilde{\rho}$ an even and bounded function that is increasing on the positive real line. The constant k_0 is usually chosen so that $S(H_K(\cdot;\boldsymbol{\beta}_0)) = \sigma$ for Gaussian errors $\varepsilon_i \sim \mathrm{N}(0,\sigma^2)$. Using our notation, the maxbias function of GS estimators is given by (5), with $g_1(\varepsilon) = S(H_1(\cdot;\varepsilon))$ and $g_2(\varepsilon,\boldsymbol{\beta}) = S(H_2(\cdot;\varepsilon,\boldsymbol{\beta}))$ (cf. thm. 4 in Croux et al. 1994). In fact, (5) holds for other scale functionals S as well, provided that they satisfy some mild regularity conditions.

Remark 6. Note that $B_{\varepsilon^*}(T) = \lim_{\varepsilon \nearrow \varepsilon^*} B_\varepsilon(T)$ is finite for the LTD estimator and multivariate Gaussian data, as soon as $\gamma > .25$ (Fig. 1). This is not the case for the LQD estimator (and GS estimators in general). Depending on the error distribution and how much mass μ puts around γ, the integral $g_1(\varepsilon^*) = \int_0^1 H_K^{-1}(u;\boldsymbol{\beta}_0) J(u/\gamma)\, du$ may converge or diverge. If $\gamma \le .25$, then $g_2(\varepsilon^*,\cdot) \equiv 0$, so that $B_{\varepsilon^*}(T) = \infty$. If $\gamma > .25$, then $g_2^{-1}(\varepsilon^*, g_1(\varepsilon^*))$ exists finitely as soon as $g_1(\varepsilon^*)$ is finite.

Remark 7. Consider for a moment a regression model $y_i = \boldsymbol{\beta}^T x_i + \varepsilon_i$ with no intercept (or with an intercept included in the $\boldsymbol{\beta}$ parameter). The LTS estimator has an objective function (4), with $J(u) = I(0 \le u \le \gamma)$ and $H_K(t;\boldsymbol{\beta}) = P_K((Y - \boldsymbol{\beta}^T X)^2 \le t)$. Theorem 2 also holds for the LTS estimator (and its generalizations with other choices of μ), if we put $H_1(t;\varepsilon) = (1-\varepsilon)H_{K_0}(t;0)$ and $H_2(t;\varepsilon,\boldsymbol{\beta}) = (1-\varepsilon)H_{K_0}(t;\boldsymbol{\beta}) + \varepsilon$. In particular, $\varepsilon^* =$

$\min(\gamma, 1-\gamma)$, as is well known. The phenomenon described in Remark 6 occurs here as well; $B_{\varepsilon^*}(T) < \infty$ for $\gamma > .5$ as soon as μ puts little mass around γ.

2.3 Asymptotic Normality and Efficiency

Define the distribution function $H_n(t;\boldsymbol{\beta}) = \sum_{i<j} I(h(z_i, z_j;\boldsymbol{\beta}) \le t)/\binom{n}{2}$, which is an empirical version of H_K. For simplicity, we first consider the LTD case $J(u) = I(0 \le u \le \gamma)$. The (empirical) objective function (3) then takes the form

$$D_n(\boldsymbol{\beta}) = \int_0^\gamma H_n^{-1}(u;\boldsymbol{\beta})\, du, \tag{7}$$

with $k_n = [\gamma\binom{n}{2}]$. We impose the following regularity conditions on F and G:

C. The carrier distribution G satisfies $E_G|X|^{2+\delta} < \infty$ for some $\delta > 0$, and $X_1 - X_2$ is not concentrated on hyperplanes in the sense that $\sup_{\boldsymbol{\beta}\neq 0} P(\boldsymbol{\beta}^T(X_1 - X_2) = 0) < \gamma$. Thus $\Sigma = \operatorname{cov}(X)$ is nonsingular, because $\operatorname{cov}(X_1 - X_2) = 2\Sigma$.

D. The error distribution F has a unimodal and strictly positive density f, and f' is continuous and bounded.

As D_n is a linear combination of the ordered $h(z_i, z_j)$, it is a generalized L (GL) statistic, introduced by Serfling (1984). As $n \to \infty$,

$$D_n(\boldsymbol{\beta}) \xrightarrow{p} D(\boldsymbol{\beta}; K) = D(\boldsymbol{\beta}). \tag{8}$$

The linear expansion

$$D_n(\boldsymbol{\beta}) = \binom{n}{2}^{-1} \times \sum_{i<j} \rho(\varepsilon_i - \varepsilon_j - (\boldsymbol{\beta}-\boldsymbol{\beta}_0)^T(x_i - x_j);\boldsymbol{\beta}) + R_n(\boldsymbol{\beta}) \tag{9}$$

will be useful, with

$$\rho(t;\boldsymbol{\beta}) = \int_0^\gamma \frac{I(t^2 > H_K^{-1}(u;\boldsymbol{\beta})) - (1-u)}{h_K(H_K^{-1}(u;\boldsymbol{\beta});\boldsymbol{\beta})}\, du = H_K^{-1}(\gamma;\boldsymbol{\beta}) \wedge t^2 - (1-\gamma) H_K^{-1}(\gamma;\boldsymbol{\beta}),$$

and $h_K(t;\boldsymbol{\beta}) = \partial H_K(t;\boldsymbol{\beta})/\partial t$. Equation (9) expresses D_n as a sum of a U statistic of order 2 (e.g., Lee 1990) and a remainder term R_n. After some manipulations, one finds

$$E\rho(\varepsilon_i - \varepsilon_j - (\boldsymbol{\beta}-\boldsymbol{\beta}_0)^T(X_i - X_j);\boldsymbol{\beta}) = D(\boldsymbol{\beta}) \tag{10}$$

and

$$R_n(\boldsymbol{\beta}) = \int_0^\gamma \left(H_n^{-1}(u) - H_K^{-1}(u) - \frac{u - H_n(H_K^{-1}(u))}{h_K(H_K^{-1}(u))} \right) du, \tag{11}$$

where we have dropped $\boldsymbol{\beta}$ in the notation for convenience. The integrand in the last integral is a remainder term of Bahadur type for each u. Now

$$\hat{\boldsymbol{\beta}}_{\mathrm{LTD}} = \operatorname*{argmin}_{\boldsymbol{\beta}} D_n(\boldsymbol{\beta}) = \operatorname*{argmin}_{\boldsymbol{\beta}} (D_n(\boldsymbol{\beta}) - D_n(\boldsymbol{\beta}_0)),$$

so the asymptotic behavior of the process $\{D_n(\boldsymbol{\beta}) - D_n(\boldsymbol{\beta}_0); \boldsymbol{\beta} \in \mathbf{R}^{p-1}\}$ determines the large-sample properties of $\hat{\boldsymbol{\beta}}_{\mathrm{LTD}}$. We call this process a GL process. By (9), it can be decomposed as a sum of a U process and a remainder process $\{R_n(\boldsymbol{\beta}) - R_n(\boldsymbol{\beta}_0)\}$. Ignore for a moment the remainder, and approximate the U process locally around $\boldsymbol{\beta}_0$ by a quadratic function,

$$\begin{aligned} D_n(\boldsymbol{\beta}) - D_n(\boldsymbol{\beta}_0) &\approx \binom{n}{2}^{-1} \sum_{i<j} \rho(\varepsilon_i - \varepsilon_j - (\boldsymbol{\beta}-\boldsymbol{\beta}_0)^T \\ &\quad \times (x_i - x_j);\boldsymbol{\beta}) \\ &\quad - \binom{n}{2}^{-1} \sum_{i<j} \rho(\varepsilon_i - \varepsilon_j; 0) \\ &\approx -(\boldsymbol{\beta}-\boldsymbol{\beta}_0)^T V_n + \frac{1}{2}(\boldsymbol{\beta}-\boldsymbol{\beta}_0)^T \\ &\quad \times D''(\boldsymbol{\beta}_0)(\boldsymbol{\beta}-\boldsymbol{\beta}_0), \end{aligned} \tag{12}$$

with $V_n = \sum_{i<j}(x_i - x_j)\psi(\varepsilon_i - \varepsilon_j)/\binom{n}{2}$ and $\psi(t) = \partial\rho(t;\boldsymbol{\beta}_0)/\partial t = 2tI(|t| \le \sqrt{H_K^{-1}(\gamma;\boldsymbol{\beta}_0)})$. The first term in (12) approximates the stochastic variation of $D_n(\cdot) - D_n(\boldsymbol{\beta}_0)$, and the second term approximates the mean value function $D(\cdot) - D(\boldsymbol{\beta}_0)$ (because $D'(\boldsymbol{\beta}_0) = 0$). Introduce $\bar{\psi}(t) = E\psi(t - \varepsilon)$. By the asymptotic theory of nondegenerate U statistics (Lee 1990, p. 76),

$$\sqrt{n}V_n = \frac{2}{\sqrt{n}} \sum_{i=1}^n \bar{\psi}(\varepsilon_i)(x_i - E(X)) + o_p(1) \xrightarrow{\mathcal{L}} \mathrm{N}_{p-1}(0, 4E\bar{\psi}(\varepsilon)^2\Sigma). \tag{13}$$

Differentiating twice with respect to $\boldsymbol{\beta}$ in (10) gives $D''(\boldsymbol{\beta}_0) = 2\Sigma E\bar{\psi}'(\varepsilon)$ (see also Lem. C.1 in Appendix C). Hence (12) leads us to the following.

Theorem 3. Assume that C and D hold. Then

$$\sqrt{n}(\hat{\boldsymbol{\beta}}_{\mathrm{LTD}} - \boldsymbol{\beta}_0) = \sqrt{n}D''(\boldsymbol{\beta}_0)^{-1}V_n + o_p(1) \xrightarrow{\mathcal{L}} \mathrm{N}_{p-1}\left(0, \frac{E\bar{\psi}(\varepsilon)^2}{(E\bar{\psi}'(\varepsilon))^2}\Sigma^{-1}\right) \tag{14}$$

as $n \to \infty$.

The argument outlined earlier is made strict in Appendix C. Technically, we first need to establish consistency of $\hat{\boldsymbol{\beta}}_{\mathrm{LTD}}$ (with a certain rate of convergence). Then we use empirical process theory (Pollard 1984) and U process theory (Nolan and Pollard 1987) to prove that two processes are uniformly small over local neighborhoods of $\boldsymbol{\beta}_0$. These processes, of which $R_n(\cdot) - R_n(\boldsymbol{\beta}_0)$ is one, correspond to the two approximations made in (12).

The Gaussian efficiency of $\hat{\boldsymbol{\beta}}_{\mathrm{LTD}}$ is given by

$$e = \frac{(\int \bar{\psi}'(t)\phi(t)\,dt)^2}{\int \bar{\psi}(t)^2\phi(t)\,dt},$$

and the cutoff point for ψ is $\sqrt{H_K^{-1}(\gamma;\boldsymbol{\beta}_0)} = \sqrt{2}\Phi^{-1}((1+\gamma)/2)$. Table 1 lists efficiencies, asymptotic variances $V = 1/e$, and breakdown points.

Table 1. Breakdown Points and Gaussian Efficiencies and Asymptotic Variances for the LTD Estimator

ε^*	γ	e	V
.50	.2500	.6626	1.5092
.45	.3025	.6688	1.4952
.40	.3600	.6773	1.4765
.35	.4225	.6885	1.4524
.30	.4900	.7033	1.4219
.25	.5625	.7228	1.3835
.20	.6400	.7484	1.3362
.15	.7225	.7826	1.2778
.10	.8100	.8291	1.2061
.05	.9025	.8956	1.1166

According to (13)–(14), the influence function (IF) of $\hat{\beta}_{\mathrm{LTD}}(z_1, \ldots, z_n)$ is

$$\mathrm{IF}((x,y),K,\hat{\beta}_{\mathrm{LTD}}) = \frac{\bar{\psi}(y-\alpha_0-\beta_0^T x)}{E(\bar{\psi}'(\varepsilon))}\Sigma^{-1}(x-E(X)),$$

in the sense that $\hat{\beta}_{\mathrm{LTD}} = \beta_0 + \sum_{i=1}^n \mathrm{IF}(z_i, K, \hat{\beta}_{\mathrm{LTD}})/n + o_p(n^{-1/2})$. Figure 2 depicts the IF for standard normal errors for various values of ε^*. Note that the IF is unbounded in the x-direction and bounded in the residual direction for each fixed x. This fact is more clear from the three-dimensional plot of the influence function given in Figure 3. As pointed out by the associate editor, the fact that the IF is unbounded in x may suggest instability in the LTD estimator (see Davies 1993; Sheather, McKean, and Hettmansperger 1997). This is further investigated in the example in Section 5.

Remark 5 (continued). The (empirical) objective function of the GS estimator is defined as the solution of $\sum_{i<j}\rho((\varepsilon_i-\varepsilon_j-(\beta-\beta_0)^T(x_i-x_j))/D_n)/\binom{n}{2} = k_0$. According to Hössjer, Croux, and Rousseeuw (1994, thm. 3.1) $\hat{\beta}_{\mathrm{LTD}}$ is asymptotically equivalent to a GS estimator with $\rho(\cdot;\beta_0) = \tilde{\rho}(\cdot/S(H_K(\cdot;\beta_0)))$.

Remark 7 (continued). The LTS estimator, $\hat{\beta}_{\mathrm{LTS}}$, has an objective function of the form (7), if we let $H_n(\cdot;\beta)$ denote the edf computed from $\{(\varepsilon_i-(\beta-\beta_0)^T x_i)^2\}$. Heuristically, the argument to establish asymptotic normality goes through with $V_n = \sum_i x_i\psi(\varepsilon_i)/n, D''(\beta_0) = \Sigma E(\psi'(\varepsilon)), \Sigma = E(X^TX)$ and asymptotic covariance matrix $E(\psi^2(\varepsilon))\Sigma^{-1}/(E(\psi'(\varepsilon)))^2$ (see also Rousseeuw 1985; Yohai and Maronna 1976). However, it is much more difficult to prove that the remainder processes are asymptotically negligible, because they oscillate more for the LTS. This is because empirical processes oscillate more than U processes. In this case, it is probably easier to use the method of proof of Hössjer (1994) and establish asymptotic linearity of $D_n'(\cdot)$ locally around β_0.

Remark 8. The objective function of LTD could be generalized to $D_n = \int J(u)H_n^{-1}(u)\,du$, with $J(u)du$ a finite measure on [0, 1]. Then $\rho(t;\beta) = \int_0^{t^2} J(H_K(s))\,ds + \int(H_K(s)-1+sh_K(s))J(H_K(s))\,ds$ and $\psi(t) = 2tJ(H_K(t^2;\beta_0))$. The functions $\bar{\psi}$ and $\bar{\psi}'$ are defined as earlier in terms of ψ. If J is a Dirac function, then we obtain the LQD estimator. The proof in Appendix C could be extended to cover smooth J functions.

Remark 9. Rank-type estimators with objective function $D_n = \int J(u)\bar{H}_n^{-1}(u)\,du$ were considered by Croux et al. (1994), where the influence function of these estimators was derived. Here $\bar{H}_n^{-1} = \sqrt{H_n^{-1}}$; that is, $\bar{H}_n(\cdot,\beta)$ is the edf of $\{|\varepsilon_i-\varepsilon_j-(\beta-\beta_0)^T(x_i-x_j)|; i<j\}$. (This is the same estimator as in Remark 1, if we put $a(i) = \int_{(i-1)/\binom{n}{2}}^{i/\binom{n}{2}} J(u)\,du$.) Among other things, it was found that $\gamma = .25$ and $J(u) = I(0 \le u \le \gamma)$ result in a Gaussian efficiency of .6604. Asymptotic normality can be proved using the methods in Appendix C.

3. COMPUTATION AND SIMULATIONS

3.1 Computing the Estimator

Computing high-breakdown estimators is notoriously difficult, as the objective functions usually have many local minima. The case of the LTD and LQD estimators is even worse than most, as we need to minimize a function of $\binom{n}{2}$ residuals. But both criteria can be fitted using more familiar estimators as the LTD estimate turns out to be the LTS estimate (and the LQD estimate is the LMS estimate) of a modified dataset.

From Section 1,

$$\begin{aligned}\hat{\beta}_{\mathrm{LTD}} &= \operatorname*{argmin}_{\beta} \sum_{i=1}^{\binom{h_p}{2}:\binom{n}{2}} \{(r_i-r_j)^2; i<j\}\\ &= \operatorname*{argmin}_{\beta} \sum_{i=1}^{\binom{h_p}{2}:\binom{n}{2}}\\ &\quad \times \{((y_i-\alpha-\beta^T x_i)-(y_j-\alpha-\beta^T x_j))^2; i<j\}\end{aligned}$$

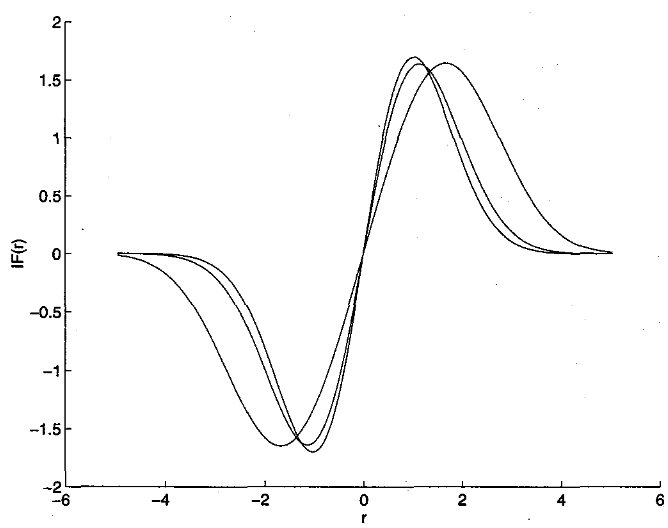

Figure 2. Influence Function (Residual Part) of the LTD Estimator for Gaussian Errors, Corresponding to ε^* = .05, .25, and .5. A higher breakdown point is indicated by a larger derivative of the influence function at the origin.

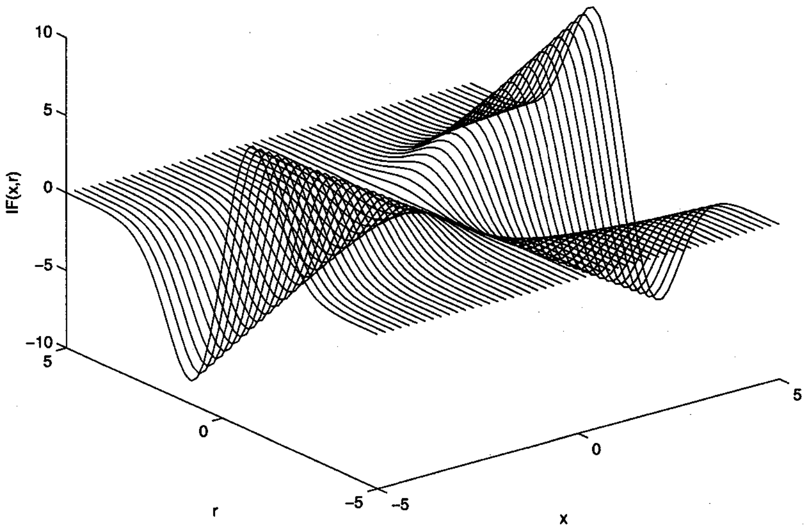

Figure 3. Influence Function for Standard Normal Errors When $\varepsilon^* = .5$, $p = 2$, $E(X) = 0$, and $\Sigma = 1$.

$$= \underset{\beta}{\operatorname{argmin}} \sum_{i=1}^{\binom{h_p}{2}:\binom{n}{2}} \{((y_i - y_j) - \beta^T(x_i - x_j))^2; i < j\}. \tag{15}$$

Thus the LTD fit for any dataset is the LTS fit for the data points

$$\{(y_i - y_j), (x_i - x_j); i < j\}, \tag{16}$$

while (as shown in Croux et al. 1994) the LQD fit is the LMS fit for the same pairs of points.

The computational approach chosen by Croux et al. (1994) for the LQD estimator is based on the Hawkins and Simonoff (1993) refinement of the original "elemental set" PROGRESS (e.g., Rousseeuw and Leroy 1987) LMS algorithm. In it, the LQD is estimated by computing the exact fit to all p point subsets of the data and then choosing the fit with the smallest value of the LQD objective function as the estimate of the LQD fit. This algorithm cannot find the exact LQD fit except in very special circumstances, although in most cases it can find a close approximation.

Hawkins's (1994) feasible solution algorithm (FSA) for LTS uses a different approach. It selects a random half of the data and makes pairwise swaps between the selected and unselected cases until it finds the half whose residual sum of squares cannot be further improved with pairwise swaps. This is done repeatedly using different random starting points. The method is guaranteed to find the global optimum if it is repeated with enough random starts, and has generally been found to produce much better approximations in a given amount of computing time than does the elemental set approach. The LTD can then be fitted using the FSA applied to the dataset of differences. A parallel FSA for LMS (Hawkins 1993) can be used for LQD.

Initial simulations for this article used an improved version of the FSA. It used a preliminary necessary condition to reduce the number of pairwise swaps studied, and typically ran one or more orders of magnitude faster than the original FSA on large datasets. Unfortunately, computation times were still excessive. Using an HP 715-75 workstation, it took about 2 hours to compute the LTD estimator for one dataset when $n = 30$ and $p = 5$.

Currently, we compute the LTD (and LQD) using a version of the SURREAL algorithm (Ruppert 1992). An initial search is done in the parameter space for the LTD estimate. The second step calls the improved version of the FSA discussed earlier to locate a local minimum of the objective function. The resulting algorithm reduced the computation time for the LTD estimator for one dataset when $n = 10$ and $p = 2$ to less than 2 seconds. For $n = 30$ and $p = 5$, the LTD estimator took about 10 minutes to compute. Using a newer HP C-110 workstation, the $n = 30$ and $p = 5$ LTD estimate takes about 5 minutes to compute. These computation times are clearly greater than for the LMS or LTS estimates, but a substantial gain in efficiency is achieved.

3.2 Simulation Study

3.2.1 Gaussian Data. LMS and LTS are popular with practitioners but they suffer from low asymptotic efficiency for Gaussian data. Further, as Hettmansperger and Sheather (1992) and Sheather et al. (1997) showed, the LMS (and LTS) fits may shift substantially when there are minor shifts in the data. LQD and LTD have much higher asymptotic efficiency, and thus they overcome, at least asymptotically, that problem with the LMS and LTS estimators.

Table 2 shows results of a simulation study comparing the LMS, LTS, LQD, and LTD estimators for various sample sizes (n) and numbers of explanatory variables (p). Datasets for $n = 10, 15, 20, 25, 30$ and $p = 2, 3, 4, 5, 6$ using standard normal data were randomly generated and then all five estimators were computed for each dataset. n was limited to at most 30 because of the computation time. All the Monte Carlo means were close to 0, as would be expected for unbiased estimators. The LMS and LTS estimators are quite a bit less efficient than either the LQD or LTD estimators. Neither LMS nor LTS has a clear efficiency advantage over the other. For normal data, it appears that the LTD performs

Table 2. Means of Monte Carlo Standard Errors and Efficiencies for the LS, LMS, LTS, LQD, and LTD Estimators and Standard Normal Data

		p				
n	Estimator	2	3	4	5	6
	LS	.3909	.4373	.4497	.4910	.5711
	LMS	.7082 (30%)	.7850 (31%)	1.314 (12%)	.9626 (26%)	1.448 (16%)
10	LTS	.6907 (32%)	.8261 (28%)	1.089 (17%)	.9197 (28%)	1.428 (16%)
	LQD	.5616 (48%)	1.012 (19%)	1.019 (19%)	1.269 (15%)	1.437 (16%)
	LTD	.5909 (44%)	.7588 (33%)	.7798 (33%)	.8141 (36%)	.9518 (36%)
	LS	.2597	.2699	.3275	.3039	.3491
	LMS	.8519 (9%)	.5581 (23%)	.6951 (22%)	.7029 (19%)	.7692 (21%)
15	LTS	.8237 (10%)	.5792 (22%)	.6828 (23%)	.7010 (19%)	.7637 (21%)
	LQD	.4994 (27%)	.4822 (31%)	.5739 (33%)	.7506 (16%)	.7390 (22%)
	LTD	.4766 (30%)	.4385 (38%)	.5349 (37%)	.4804 (40%)	.5317 (43%)
	LS	.2556	.2501	.2429	.2579	.2887
	LMS	.4625 (31%)	.4621 (29%)	.5549 (19%)	.5707 (20%)	.6276 (21%)
20	LTS	.5174 (24%)	.4313 (34%)	.5317 (21%)	.5722 (20%)	.6581 (19%)
	LQD	.3794 (45%)	.3628 (48%)	.4576 (28%)	.5092 (26%)	.5381 (29%)
	LTD	.3744 (47%)	.3410 (54%)	.4046 (36%)	.4296 (36%)	.4755 (37%)
	LS	.2181	.2108	.2341	.2358	.2387
	LMS	.5046 (19%)	.4967 (18%)	.5244 (20%)	.5540 (18%)	.5046 (22%)
25	LTS	.5369 (16%)	.5024 (18%)	.5149 (21%)	.5651 (17%)	.5034 (22%)
	LQD	.4187 (27%)	.4009 (28%)	.4108 (32%)	.4305 (30%)	.4020 (35%)
	LTD	.3597 (37%)	.3495 (36%)	.3470 (46%)	.3472 (46%)	.3720 (41%)
	LS	.2069	.1821	.1907	.1931	.2131
	LMS	.4106 (25%)	.4224 (19%)	.3987 (23%)	.4346 (20%)	.4525 (22%)
30	LTS	.4440 (22%)	.4391 (17%)	.3932 (24%)	.4302 (20%)	.4822 (20%)
	LQD	.3470 (36%)	.3239 (32%)	.3006 (40%)	.3528 (30%)	.3809 (31%)
	LTD	.2972 (48%)	.2989 (37%)	.2975 (41%)	.3134 (38%)	.3208 (44%)

somewhat better than the LQD. The average percent reduction in the LTD Monte Carlo standard errors over the LQD Monte Carlo standard errors was 20%. Both the LTD and LQD finite-sample efficiencies were substantially less than their asymptotic efficiencies.

3.2.2 Heavy-Tailed Distributions. In Table 3, both the explanatory and response variables are randomly generated from a t-distribution with one df. This results in both influential points and outliers. For each estimator, n and p, 100 datasets were randomly generated. The mean of the means of the 100 parameter estimates is reported along with the mean of the Monte Carlo standard error estimates in parentheses. Because of the superiority of LTD and LQD for Gaussian data, LMS and LTS results are not included in the tables here. In Table 4 the explanatory variables are again generated from a t distribution with 1 df to create influential points, then the response variable is from a standard normal with 20% of the responses increased by 50 to generate outliers. As would be expected, the LS estimator does poorly in these situations. Neither LTD nor LQD has a clear advantage over the other in these situations. It is interesting that the LQD estimates for $n = 20$ and $p = 5$ appear somewhat unstable. This is investigated further in Section 4.

4. STANDARD ERRORS

Standard errors are another vital aspect of estimation. Asymptotic standard errors were not reliable for the moderate sample sizes considered here, because the weak convergence of Theorem 3 is very slow. First, we tried approximating the asymptotic variance in (14), both for normal errors and for a general unknown residual distribution. Because these approximations were poor for moderate n, we then tried a second-order correction of the numerator $E\bar{\psi}(\varepsilon)^2$ based on an exact expression of $n \operatorname{var}(V_n)$ for normal errors. Even this method gave poor estimates of the standard error.

We then considered standard errors based on resampling methods. Stromberg (1997) argued that jackknife, rather

Table 3. Monte Carlo Means of Means and MC Standard Errors for the LS, LQD, and LTD Estimators and t Distribution Data

		p		
n	Estimator	2	4	6
	LS	.6822 (4.175)	−2.396 (40.47)	−.2474 (5.322)
10	LQD	−.0013 (.4519)	−.0510 (.7692)	−.2690 (2.760)
	LTD	.0174 (.4349)	−.0632 (.7170)	−.2947 (2.960)
	LS	−.1279 (.5846)	−.5108 (5.019)	−.1057 (2.890)
20	LQD	−.0093 (.1542)	.0192 (.2861)	−.4045 (4.254)
	LTD	−.0049 (.1688)	.0032 (.2802)	.0279 (.3200)
	LS	−.1632 (1.4137)	.1925 (2.1715)	.0825 (1.2181)
30	LQD	−.0065 (.1116)	−.0054 (.1340)	−.0070 (.1554)
	LTD	.0077 (.1403)	.0002 (.1626)	.0087 (.1980)

Table 4. Monte Carlo Means of Means and MC Standard Errors for the LS, LQD, and LTD Estimators and 20% Outlier Data

		p		
n	Estimator	2	4	6
	LS	.0845 (2.697)	−.0989 (3.935)	.2082 (5.275)
10	LQD	.0365 (.3165)	−.0228 (.5180)	.0780 (1.205)
	LTD	.0204 (.3860)	−.0235 (.5177)	.0750 (1.266)
	LS	−.0645 (1.218)	−.0886 (1.470)	−.0128 (1.545)
20	LQD	.0079 (.1484)	.0041 (.1773)	.6296 (9.883)
	LTD	−.0200 (.1958)	.0045 (.2021)	−.0014 (.2361)
	LS	.0754 (.7187)	.0314 (.8893)	.0227 (.9914)
30	LQD	.0107 (.0719)	−.0019 (.1038)	.0021 (.0908)
	LTD	.0042 (.0548)	−.0020 (.1191)	.0060 (.1097)

Table 5. Standard Error Estimates for LTD and LQD Estimates

Type of data	SE estimate	10/2	20/2	20/4	30/2	30/4	30/6
x – N(0, 1)	LTD MC	.6079	.3929	.4557	.3590	.2943	.3607
	LTD dn/2	.4132	.2130	.3248	.1809	.2156	.2972
	LQD MC	.5616	.3794	.4576	.3470	.3158	.3809
y – N(0, 1)	LQD dn/2	1.041	.6599	.8625	.6676	.6630	.7081
	LTD MC	.4349	.1688	.2802	.1403	.1626	.1980
x – t (1 df)	LTD dn/2	.5104	.1784	.2962	.1239	.1516	.1408
y – t (1 df)	LQD MC	.4519	.1542	.2861	.1116	.1340	.1554
	LQD dn/2	1.922	.9758	1.077	.8881	.7628	1.051
x – t (1 df)	LTD MC	.3860	.1958	.2021	.0548	.1191	.1097
y – N(0, 1)	LTD dn/2	.5515	.1418	.1938	.0772	.1094	.1553
20% outliers	LQD MC	.3165	.1484	.1773	.0719	.1038	.0908
	LQD dn/2	7.446	.6051	1.499	.4433	.8810	2.322

than bootstrap, standard errors are more appropriate for robust estimators because jackknifed standard errors have a higher breakdown point. In the case of LQD and LTD, bootstrap standard errors have another disadvantage in that they are based on differences in pairs of residuals, and so if a bootstrap-resampled dataset contained replicates, then the LQD or LTD objective function for these pairs of points would be 0 regardless of the estimate. The result would be lower efficiency for bootstrap standard errors. In all simulations the jackknife standard errors performed much more reasonably than the bootstrap standard errors; thus we present results only for the jackknife. Clearly, just as the delete-one jackknife fails for the median, we must expect the delete-one jackknife to underestimate standard errors for LQD and LTD. Breakdown point considerations suggest the delete-$n/3$ jackknife would perform well (breakdown point $1/3$). Although it has a breakdown point of only $1/4$, Wu suggested the delete-$1/2$ jackknife. Extensive simulations showed that the delete-$1/3$ jackknife underestimated standard errors for LTD, so we present results for only the delete-$1/2$ jackknife. Because of the computational load, the smallest number of resampled datasets is desirable. For the simulations considered here, n resampled datasets appeared to be adequate.

We consider the cases of the previous sections: standard normal data for both the explanatory and response variables, t distribution with 1 df for both the explanatory and response variables, and t distribution with 1 df for the explanatory variables and a standard normal response with 50 added to 20% of the responses to generate outliers.

Table 5 presents means of Monte Carlo and delete-$n/2$ jackknife standard error estimates for the LQD and LTD estimators. The Monte Carlo estimates are ideal but are not available in practice. Good performance can be measured by how close the standard errors are to the Monte Carlo results. The table shows that the delete-$1/2$ jackknife standard errors for LTD, but not those for LQD, are close to the Monte Carlo standard errors. The LTD standard errors appear to be quite a bit more stable than the LQD standard errors.

5. EXAMPLE

5.1 Engine Knock Data

An interesting data set that has received considerable attention in the robustness literature is the engine knock data initially studied by Mason, Gunst, and Hess (1989,

Table 6. LMS Residuals for Several 15-Point Subsets of the Engine Knock Data

Deleted point 1	3	5	10
.16	–.23	.19	–.23
–4.02	–.23	–3.42	–.23
–.34	.22	–.19	.22
–.45	4.18	–.07	4.18
6.66	5.02	6.05	5.02
.45	.10	.10	.10
–.45	–4.19	–1.24	–4.19
.45	.22	–.01	.22
1.72	.22	1.60	.22
–.02	–.14	.19	–.14
.26	–.85	.17	–.85
–.07	–2.42	–.19	–2.42
3.48	–.12	3.14	–.12
–.45	5.69	–.19	5.69
3.98	–.17	3.55	–.17
.45	–.23	.19	–.23

Table 7. LTS Residuals for Several 15-Point Subsets of the Engine Knock Data

Deleted point 1	2	4	5
–.24	.13	–.23	.13
–.16	–3.31	–.11	–3.31
.23	–.17	.34	–.17
4.44	–.03	4.03	–.03
4.82	5.95	4.87	5.95
–.10	.05	0	.05
–4.45	–1.39	–4.28	–1.39
.11	–.08	.06	–.08
.49	1.50	.24	1.50
.12	.17	–.02	.17
–.57	.08	–.83	.08
–2.13	–.29	–2.35	–.29
–.04	3.03	0	3.03
6.11	–.03	5.63	–.03
–.11	3.41	–.09	3.41
.02	.17	–.18	.17

Table 8. LQD Residuals for Several 15-Point Subsets of the Engine Knock Data

Deleted point			
1	2	4	15
−.68	.04	−.40	−4.55
−4.32	−3.55	−4.65	−37.76
−1.12	−.37	−1.19	−25.86
−.25	−.04	0	9.54
4.45	5.90	5.58	7.57
−1.57	−.06	−.70	−9.84
−2.96	−1.46	−1.83	−2.92
−1.42	−.12	−.39	2.25
1.81	1.48	1.91	17.45
.31	.06	0	−2.85
.40	.06	.47	14.46
.17	−.33	.08	12.58
2.71	2.92	2.83	−.55
−.12	−.13	−.09	−6.81
3.13	3.33	3.37	6.11
.12	.08	.25	.55

p. 529). Hettmansperger and Sheather (1992) discussed the instability of the LMS estimate using the PROGRESS (e.g., Rousseeuw and Leroy 1987) computational algorithm due to a minor change shift in one of the data points. Further discussion of this issue was given by Sheather et al. (1997). Stromberg (1993) presented an exact algorithm for computing the LMS estimate, as well as two methods for detecting instability in the LMS estimate and showed that this instability remains when the exact algorithm is used and so is inherent to the estimator, not to the computational approximation.

It is of interest to check whether LQD and LTD are also susceptible to this instability. Tables 6–9 present the residuals for selected 15-point (1-point-deleted) subsets of the data from LMS, LTS, LQD, and LTD using 500 random starts of the FSA algorithm. All four estimates yield vastly different residuals depending on what point is deleted. At least in this case, it appears that despite the higher efficiency of LTS, LQD, and LTD, they are no more stable than LMS.

Table 9. LTD Residuals for Several 15-Point Subsets of the Engine Knock Data

Deleted point			
1	3	8	15
.12	−.26	−.01	.06
−3.28	.02	−.30	−3.22
−.18	.47	.05	−.17
−.06	3.81	5.08	−.06
5.87	4.82	5.02	5.88
0	.01	−.01	.05
−1.48	−4.26	−4.41	−1.49
−.16	−.01	.37	−.12
1.42	.10	.85	1.27
.13	−.04	.25	.02
0	−.97	−.23	−.13
−.37	−2.44	−1.89	−.54
2.96	.05	−.13	2.84
−.14	5.55	6.23	−.01
3.33	−.08	−.10	3.20
0.05	−.19	.01	.01

As discussed in Section 2.3, this instability could be due to the fact that LTD influence function is unbounded in the x space. As pointed out by the associate editor, it also suggests that the residuals from the LTD fit may perform poorly in model identification as discussed by Cook, Hawkins, and Wiesburg (1992) and McKean, Sheather, and Hettmansperger (1993).

5.2 Conclusions

The LTD (and LQD) estimators have strong intuitive appeal based on their desirable breakdown and asymptotic properties. Like other high-breakdown estimators, they can be unstable under minor shifts in the data, but their higher Gaussian efficiency is likely to make them less sensitive than LMS and LTS to such shifts. The simulation results for both Gaussian and non-Gaussian data support using the LTD over the LQD.

APPENDIX A: PROOF OF THEOREM 1

Throughout the proof, we assume that m data points of $\mathbf{Z}$ are replaced. Denote the new sample by $\mathbf{Z}' = \{z_i' = (x_i', y_i')\}$. Let G be set of indices of the unchanged, "good," data points, and let B be the indices of the changed, "bad," points. Put also $z_{ij} = (x_{ij}, y_{ij}) = (x_i - x_j, y_i - y_j) = z_i - z_j$, $z_{ij}' = z_i - z_j$, $r_{ij}(\boldsymbol{\beta}) = y_{ij} - \boldsymbol{\beta}^T x_{ij}$, $r_{ij}'(\boldsymbol{\beta}) = y_{ij}' - \boldsymbol{\beta}^T x_{ij}'$, and $H_{\boldsymbol{\beta}} = \{z = (x, y); y = \boldsymbol{\beta}^T x\}$. The proof comprises three steps.

Step 1. $m \geq q(k) - p + 1 \Rightarrow$ breakdown. It suffices to prove that $\binom{m+p-1}{2} \geq k$ implies breakdown. Choose an arbitrary set $G_0 \subset G$ with $|G_0| = p - 1$. Let $\mathcal{B}$ denote the set of $\boldsymbol{\beta}$-vectors for which there exists and $\alpha = \alpha(\boldsymbol{\beta})$ such that $y_i = \boldsymbol{\beta}^T x_i + \alpha$, for all $i \in G_0$. Fix $\boldsymbol{\beta} \in \mathcal{B}$, and let α be the corresponding intercept. For all $i \in B$, let $z_i' \in H = \{z = (x, y); y = \boldsymbol{\beta}^T x + \alpha\}$. But then $z_{ij}' \in H_{\boldsymbol{\beta}}$ for all $i < j$ with $i, j \in G_0 \cup B$. Hence we have at least $\binom{|G_0|+|B|}{2} = \binom{m+p-1}{2} \geq k$ pairs z_{ij}' that belong to $H_{\boldsymbol{\beta}}$; that is, $D_n(\boldsymbol{\beta}) = 0$. Because $\{x_i\}$ are in general position, $\mathcal{B}$ is a one-dimensional affine hyperplane, and thus $\boldsymbol{\beta}$ can be chosen arbitrarily large.

Step 2. $m \geq n+1-q(k) \Rightarrow$ breakdown. It suffices to prove that $\binom{n-m}{2} \leq k - 1$ implies breakdown. Place all new points $z_i', i \in B$ in $H_{\boldsymbol{\beta}'}$, where $\boldsymbol{\beta}'$ is chosen later. Then $(i \neq j)$,

$$r_{ij}'(\boldsymbol{\beta}') = \begin{cases} 0; & i, j \in B \\ y_i - (\boldsymbol{\beta}')^T x_i; & i \in G, \quad j \in B \\ y_{ij} - (\boldsymbol{\beta}')^T x_{ij}; & i, j \in G. \end{cases}$$

Hence

$$|r_{\cdot\cdot}'(\boldsymbol{\beta}')|_{k:\binom{n}{2}} \leq 2(\max_i |y_i| + |\boldsymbol{\beta}'| \max_i |x_i|), \tag{A.1}$$

using the same notation as in Remark 1, with r' referring to the new sample. On the other hand, for an arbitrary $\boldsymbol{\beta}$, we have

$$r_{ij}'(\boldsymbol{\beta}) = \begin{cases} (\boldsymbol{\beta}' - \boldsymbol{\beta})^T x_{ij}'; & i, j \in B \\ (\boldsymbol{\beta}' - \boldsymbol{\beta})^T x_j' + y_i - \boldsymbol{\beta}^T x_i; & i \in G, \quad j \in B \\ y_{ij} - \boldsymbol{\beta}^T x_{ij}; & i, j \in G. \end{cases}$$

Let $\hat{\boldsymbol{\beta}}$ be the old minimum of $D_n(\cdot)$ and fix $b > 0$ as a large number. Because $\binom{|G|}{2} \leq k - 1$, it follows that

$$\begin{aligned} \inf_{\boldsymbol{\beta}} |r_{\cdot\cdot}'(\boldsymbol{\beta})|_{k:\binom{n}{2}} &\geq \inf_{x', \boldsymbol{\beta}} |(\boldsymbol{\beta}' - \boldsymbol{\beta})^T x'| \\ &\quad - (\max_i |y_i| + (|\hat{\boldsymbol{\beta}}| + b) \max_i |x_i|) \\ &= \tau - (\max_i |y_i| + (|\hat{\boldsymbol{\beta}}| + b) \max_i |x_i|), \end{aligned} \tag{A.2}$$

where the infima are taken over $B(\hat{\boldsymbol\beta}, b) = \{\boldsymbol\beta; |\boldsymbol\beta - \hat{\boldsymbol\beta}| \le b\}$ and $x' \in \{x_i'; i \in B\} \cup \{x_{ij}', i < j, i, j \in B\}$. Because x' ranges over a finite set, we can always choose $\boldsymbol\beta'$ outside $B(\hat{\boldsymbol\beta}, b)$ in such a way that $\tau > 0$. If now $z_i' \to Mz_i'$ for all $i \in B$, then the upper bound in (A.1) is unchanged, whereas τ may be replaced by $M\tau$ in (A.2). Hence, for M large enough, we have $\inf_{\boldsymbol\beta \in B(\hat{\boldsymbol\beta}, b)} D_n(\boldsymbol\beta) > D_n(\boldsymbol\beta')$ for the objective function based on the new data points. Because b was arbitrarily chosen, we have a breakdown.

Step 3. $m \le \min(n + 1 - q(k), q(k) - p + 1) - 1 \Rightarrow$ no breakdown. It suffices to prove that $\binom{m+p-1}{2} \le k - 1$ and $\binom{n-m}{2} \ge k$ imply no breakdown. Because $\binom{|G|}{2} \ge k$, it follows that

$$\begin{aligned} D_n(0) &\le k|r'_{\cdot\cdot}(\boldsymbol\beta')|^2_{k:\binom{n}{2}} \Big/ \binom{n}{2} \le k(2 \max_i |y_i|)^2 \Big/ \binom{n}{2} \\ &:= k(2M)^2 \Big/ \binom{n}{2} < \infty. \end{aligned}$$

Because the pairs $\{x_{ij}\}$ are in general position, $\inf_{|\boldsymbol\beta|=1} \{|\boldsymbol\beta^T x_{ij}|; i < j\}_{\binom{p-1}{2}+1:\binom{n}{2}} = \tau > 0$. Let $\rho = \tau/n$ and fix $\boldsymbol\beta$. As in the proof of theorem 2 of Croux et al. (1994), divide G into classes $G_0, \dots, G_s$ (depending on $\boldsymbol\beta$) such that

$$i, j \in G_l, l \ge 1 \Rightarrow |\boldsymbol\beta^T x_{ij}| \le \tau|\boldsymbol\beta|,$$

$$i, j \in G_0 \Rightarrow |\boldsymbol\beta^T x_{ij}| > \rho|\boldsymbol\beta|,$$

and

$$i \in G_l, j \in G_{l'}, l \ne l' \Rightarrow |\boldsymbol\beta^T x_{ij}| > \rho|\boldsymbol\beta|.$$

In the latter two cases, $|r_{ij}'(\boldsymbol\beta)| = |r_{ij}(\boldsymbol\beta)| = |y_{ij} - \boldsymbol\beta^T x_{ij}| \ge \rho|\boldsymbol\beta| - 2M$. Next, subdivide B into disjoint sets $B_0, \dots, B_{s+1}$, with $B_j = \{i;$ there exists $l \in G_j$ such that $|r_{il}'(\boldsymbol\beta)| \le (\rho|\boldsymbol\beta| - 2M)/4\}$ for $j = 0, \dots, s$ and $B_{s+1} = B \setminus (B_0 \cup \cdots \cup B_s)$. Consequently, $|r_{ij}'(\boldsymbol\beta)| \le (\rho|\boldsymbol\beta| - 2M)/4$ for at most

$$\begin{aligned} &\sum_{l=1}^{s} \binom{|G_l|}{2} + \sum_{l=1}^{s} |G_l||B_l| + |B_0| + \binom{|B|}{2} \\ &\le \binom{p-1}{2} + (p-1)m + \binom{m}{2} \\ &= \binom{m+p-1}{2} \le k - 1 \end{aligned}$$

pairs. We have used $\sum_{l\ge1}\binom{|G_l|}{2} \le \binom{p-1}{2}$, and hence $|G_l| \le p - 1$ for each $l \ge 1$. This follows because $\{x_{ij}\}$ are in general position. Hence $|r'_{\cdot\cdot}(\boldsymbol\beta)|_{k:\binom{n}{2}} > (\rho|\boldsymbol\beta| - 2M)/4$, which implies $D_n(\boldsymbol\beta) \ge k|r'_{\cdot\cdot}(\boldsymbol\beta)|^2_{k:\binom{n}{2}}/\binom{n}{2} > D_n(0)$ for all sufficiently large $|\boldsymbol\beta|$.

APPENDIX B: PROOF OF THEOREM 2

The basic ideas of the proof are similar to the proof of theorem 4 of Croux et al. (1994), so we are rather brief here. Put $c = g_2^{-1}(\varepsilon, g_1(\varepsilon))$. We divide the proof into three steps. We assume $\varepsilon < \min(\sqrt{\gamma}, 1 - \sqrt{\gamma})$ throughout the first two steps.

Step 1. $B_\varepsilon(T) \le c$. It suffices to prove

$$|\boldsymbol\beta - \boldsymbol\beta_0| > c \Rightarrow D(\boldsymbol\beta; K) > g_1(\varepsilon) \ge D(0; K) \qquad \text{(B.1)}$$

for any $K \in V_\varepsilon$. Given two nonnegative distribution functions H'' and H', let $H'' \overset{s}{>} H'$ mean "H'' is strictly stochastically larger than H'," in the sense that $H''(t) < H'(t)$ for all $t > 0$. Note that

$$\begin{aligned} H_K(\cdot; \boldsymbol\beta) &= (1 - \varepsilon)^2 H_{K_0}(\cdot; \boldsymbol\beta) + 2\varepsilon(1 - \varepsilon)\mathcal{L}_{K_0 \times K^*} \\ &\quad \times (\varepsilon_1 - \varepsilon_2 - (\boldsymbol\beta - \boldsymbol\beta_0)^T(X_1 - X_2)) \\ &\quad + \varepsilon^2 H_{K^*}(\cdot; \boldsymbol\beta) \overset{s}{>} H_2(\cdot; \boldsymbol\beta) \overset{s}{>} H_2(\cdot; \boldsymbol\beta') \end{aligned}$$

with $|\boldsymbol\beta' - \boldsymbol\beta_0| = c$. The second-last relation follows from conditions A and B (cf. lemma 1 in Croux et al. 1994). By the definition of c, $\int J(u) H_2^{-1}(u; \boldsymbol\beta')\, du = g_1(\varepsilon)$, so the first inequality in (B.1) follows. The second inequality is immediate.

Step 2. $B_\varepsilon(T) \ge c$. Take any $0 < c_1 < c$ and pick $\boldsymbol\beta'$ so that $|\boldsymbol\beta' - \boldsymbol\beta_0| = c_1$. Define a sequence $K_n = (1 - \varepsilon)K_0 + \varepsilon K_n^* \in V_\varepsilon$. The contaminating distribution K_n^*, corresponds to (x_n^*, y_n^*), with $y_n^* = (\boldsymbol\beta')^T x_n^*$ and x_n^* uniformly distributed on the line segment $[\lambda_n \boldsymbol\beta', 2\lambda_n \boldsymbol\beta']$. Because c_1 is arbitrary, it suffices to prove

$$\sup_n |T(K_n) - \boldsymbol\beta_0| \ge c_1. \qquad \text{(B.2)}$$

If (B.2) fails, then we may construct a subsequence of K_n (which we still call K_n) such that $T(K_n) = \boldsymbol\beta_n$ and $\boldsymbol\beta_n \to \tilde{\boldsymbol\beta}$, with $|\tilde{\boldsymbol\beta} - \boldsymbol\beta_0| < c_1$. As done by Croux et al. (1994, lem. 2), one proves $\liminf_{n\to\infty} D(\boldsymbol\beta_n; K_n) \ge g_1(\varepsilon)$, with $\lambda_n \to \infty$ as a key ingredient in the proof. Moreover, because $y_n^* = (\boldsymbol\beta')^T x_n^*$ holds exactly under K_n^*, it is clear that $D(\boldsymbol\beta'; K_n) = g_2(\varepsilon; \boldsymbol\beta') < g_1(\varepsilon)$. Thus $T(K_n) = \boldsymbol\beta_n$ leads to a contradiction, and (B.2) must hold.

Step 3. Breakdown. Note that

$$g_1(\varepsilon) = \begin{cases} \int J(u/(1-\varepsilon)^2) H_{K_0}^{-1}(u; \boldsymbol\beta_0)\, du < \infty, & \varepsilon < 1 - \sqrt{\gamma} \\ \infty, & \varepsilon > 1 - \sqrt{\gamma} \end{cases}$$

Further, for any $0 < a < \infty, g_2^{-1}(\varepsilon, a) < \infty$ if $\varepsilon < \sqrt{\gamma}$ and $g_2^{-1}(\varepsilon, a) = \infty$ if $\varepsilon \ge \sqrt{\gamma}$.

APPENDIX C: PROOF OF THEOREM 3

For brevity, put $\hat{\boldsymbol\beta}_{\text{LTD}} = \hat{\boldsymbol\beta}_n$ and assume without loss of generality that $\boldsymbol\beta_0 = \alpha_0 = 0$, so that $z_i = (x_i, y_i) = (x_i, \varepsilon_i)$. Put also $z_{ij} = z_i - z_j, x_{ij} = x_i - x_j, \varepsilon_{ij} = \varepsilon_i - \varepsilon_j$, and let C denote a positive constant whose value may change from line to line. To prove Theorem 3, we first need a series of preliminary lemmas.

Lemma C.1. The limiting objective function $D(\cdot)$ satisfies

$$D(t\boldsymbol\beta) \ge D(\boldsymbol\beta) \quad \text{if } t > 1, \qquad \text{(C.1)}$$

$$D'(0) = 0, \qquad \text{(C.2)}$$

$D''(\cdot)$ is continuous with

$$\begin{aligned} D''(0) &= 8\boldsymbol\Sigma \int_0^{\bar{H}_K^{-1}(\gamma)} (f^*(t) - f^*(\bar{H}_K^{-1}(\gamma)))\, dt \\ &= 2\boldsymbol\Sigma E(\bar{\psi}'(\varepsilon)), \end{aligned} \qquad \text{(C.3)}$$

which is positive definite, where $\bar{H}_K^{-1}(u) = \sqrt{H_K^{-1}(u; 0)}$ and f^* is the density of ε_{ij}.

Proof. Formula (C.1) is derived as was done by Hössjer et al. (1994, lem. B.1), using the fact that unimodality of f implies unimodality of f^*. The rest of the proof follows after some manipulations; for example, differentiation twice with respect to $\boldsymbol\beta$ in (8) (see Stromberg, Hawkins, and Hössjer 1995 for details.

Lemma C.2. For some $r > 0$, we have $P(\hat{\boldsymbol\beta}_n \in K) \to 1$, with $K = \{\boldsymbol\beta; |\boldsymbol\beta| \le r\}$. For any $\tau > 0, \hat{\boldsymbol\beta}_n = o_p(n^{-1/4+\tau})$.

Proof. For the proof of the first part, we refer to work of Stromberg et al. (1995). The second part has two main ingredients. First,

$$\sup_{\boldsymbol\beta \in K} |D_n(\boldsymbol\beta) - D(\boldsymbol\beta)| = o_p(n^{-1/2+\tau}), \quad \forall \tau > 0$$

is proved using empirical process theory (see Stromberg et al. 1995 for details). Second, by (C.1)–(C.3) there exists a positive constant λ (take, e.g., one-quarter of the smallest eigenvalue of $D''(0)$) such that

$$\inf_{\boldsymbol{\beta};|\boldsymbol{\beta}|\geq\zeta} D(\boldsymbol{\beta}) \geq D(0) + \lambda\zeta^2$$

for all ζ small enough. The lemma now follows by combining the last two displays with $P(\hat{\boldsymbol{\beta}}_n \in K) \to 1$.

For the next two lemmas, introduce the neighborhoods $U_n = \{\boldsymbol{\beta}; |\boldsymbol{\beta}| \leq \delta_n\}$, with $\delta_n \to 0$ a given sequence of positive numbers.

Lemma C.3. Define the U process

$$S_n(\boldsymbol{\beta}) = \binom{n}{2}^{-1} \sum_{i<j} \eta(z_i, z_j; \boldsymbol{\beta}), \tag{C.4}$$

with $\eta(z_i, z_j; \boldsymbol{\beta}) = \rho(\varepsilon_{ij} - \boldsymbol{\beta}^T x_{ij}; \boldsymbol{\beta}) - D(\boldsymbol{\beta}) - \rho(\varepsilon_{ij}; 0) + D(0) + \boldsymbol{\beta}^T x_{ij}\psi(\varepsilon_{ij})$. Then

$$\sup_{\boldsymbol{\beta}\in U_n} |S_n(\boldsymbol{\beta})| = o_p(\delta_n^{1+\delta/2} n^{-1/2+\tau} + n^{-1})$$

for any $\tau > 0$, with δ as defined in Condition C.

Proof. By construction, $E\eta(Z_1, Z_2; \boldsymbol{\beta}) = 0$. We start by making a Hoeffding decomposition of S_n,

$$\begin{aligned} S_n(\boldsymbol{\beta}) &= \frac{2}{n} \sum_{i=1}^{n} \eta_1(z_i; \boldsymbol{\beta}) + \binom{n}{2}^{-1} \sum_{i<j} \eta_2(z_1, z_2; \boldsymbol{\beta}) \\ &:= S_{n1}(\boldsymbol{\beta}) + S_{n2}(\boldsymbol{\beta}), \end{aligned}$$

with $\eta_1(z; \boldsymbol{\beta}) = E(\eta(z, Z; \boldsymbol{\beta}))$ and $\eta_2(z_1, z_2; \boldsymbol{\beta}) = \eta(z_1, z_2; \boldsymbol{\beta}) - \eta_1(z_1; \boldsymbol{\beta}) - \eta_1(z_2; \boldsymbol{\beta})$. Thus S_n is written as a sum of an empirical process S_{n1} and a U process S_{n2} with a degenerate kernel η_2. Stromberg et al. (1995) proved that $\sup_{\boldsymbol{\beta}\in U_n} |S_{n1}(\boldsymbol{\beta})| = \delta_n^{1+\delta/2} o_p(n^{-1/2+\tau})$ for any $\tau > 0$, using empirical process theory, and $\sup_{\boldsymbol{\beta}\in U_n} |S_{n2}(\boldsymbol{\beta})| = \delta_n O_p(n^{-1})$, using U process theory.

Lemma C.4. Let R_n be as defined in (11). Then

$$\sup_{\boldsymbol{\beta}\in U_n} |R_n(\boldsymbol{\beta}) - R_n(0)| = o_p((\delta_n^{1/2} n^{-1} + \delta_n^2 n^{-1/2} + n^{-5/4}) n^{\tau}),$$

for any $\tau > 0$.

Proof. Put $\bar{R}_n(u; \boldsymbol{\beta}) = H_n^{-1}(u; \boldsymbol{\beta}) - H_K^{-1}(u; \boldsymbol{\beta}) - (u - H_n(H_K^{-1}(u; \boldsymbol{\beta}); \boldsymbol{\beta}))/h_K(H_K^{-1}(u; \boldsymbol{\beta}); \boldsymbol{\beta})$, which is a Bahadur representation remainder term. Then, according to (11),

$$R_n(\boldsymbol{\beta}) - R_n(0) = \int_0^{\gamma} (\bar{R}_n(u; \boldsymbol{\beta}) - \bar{R}_n(u; 0))\, du.$$

Stromberg et al. (1995) proved that $\bar{R}_n(u; \boldsymbol{\beta})$ is uniformly small for $(u, \boldsymbol{\beta}) \in (0, \gamma) \times K$, with a rate that implies the lemma.

Proof of Theorem 3

We specify the expansion (12). Using (9), (C.4), and the continuity of $D''(\cdot)$, we obtain

$$\begin{aligned} D_n(\boldsymbol{\beta}) = D_n(0) + \left(\frac{1}{2}\boldsymbol{\beta}^T D''(0)\boldsymbol{\beta} + o(|\boldsymbol{\beta}|^2)\right) \\ - \boldsymbol{\beta}^T V_n + S_n(\boldsymbol{\beta}) + (R_n(\boldsymbol{\beta}) - R_n(0)). \end{aligned} \tag{C.5}$$

Suppose we know that $\hat{\boldsymbol{\beta}}_n = O_p(n^{-\xi})$ for some $\xi \in (0, 1/2]$. Recall that $|V_n| = O_p(n^{-1/2})$ according to (13) and that $D''(0)$ is positive definite by (C.3). These facts and Lemmas C.3 and C.4 imply that

$$\begin{aligned} D_n(\hat{\boldsymbol{\beta}}_n) \geq\ & D_n(0) + \lambda|\hat{\boldsymbol{\beta}}_n|^2(1 + o_p(1)) + |\hat{\boldsymbol{\beta}}_n| O_p(n^{-1/2}) \\ & + o_p(n^{-1} + n^{-\xi(1+\delta/2)-1/2+\tau}) \\ & + o_p((n^{-\xi/2-1} + n^{-2\xi-1/2} + n^{-5/4}) n^{\tau}) \end{aligned}$$

for some positive λ and any $\tau > 0$. The quantity $\delta > 0$ is the same number as in condition C. Because $D_n(\hat{\boldsymbol{\beta}}_n) \leq D_n(0)$, we obtain $\hat{\boldsymbol{\beta}}_n = O_p(n^{-\xi'})$, with

$$\xi' = \min\left(\frac{1}{2}, \frac{\xi}{2}\left(1 + \frac{\delta}{2}\right) + 1/4 - \tau, \xi + \frac{1}{4} - \tau\right)$$

for any $\tau > 0$. Starting with Lemma C.2, we may now iterate and obtain $\hat{\boldsymbol{\beta}}_n = O_p(n^{-\xi_k})$, with $1/4 - \tau = \xi_0 < \xi_1 < \cdots < \xi_N = 1/2$, so that after a finite number of steps, we reach 1/2. Hence $\hat{\boldsymbol{\beta}}_n = O_p(n^{-1/2})$ and

$$D_n(\hat{\boldsymbol{\beta}}_n) = D_n(0) + \frac{1}{2}\hat{\boldsymbol{\beta}}_n^T D''(0)\hat{\boldsymbol{\beta}}_n + \hat{\boldsymbol{\beta}}_n^T V_n + o_p(n^{-1}),$$

which implies

$$\hat{\boldsymbol{\beta}}_n = -D''(0)^{-1} V_n + o_p(n^{-1/2}).$$

The theorem now follows from (13) and (C.3).

[Received August 1995. Revised September 1999.]

REFERENCES

Cook, D., Hawkins, D., and Wiesberg, S. (1992), "Comparison of Model Misspecification Diagnostics Using Residuals From Least Mean of Squares and Least Median of Squares Fits," *Journal of the American Statistical Association*, 87, 419–424.

Croux, C., Rousseeuw, P., and Hössjer, O. (1994), "Generalized S-Estimators," *Journal of the American Statistical Association*, 89, 1271–1281.

Davies, P. L. (1993), "Aspects of Robust Linear Regression," *The Annals of Statistics*, 21, 1843–1899.

Donoho, D. L., and Huber, P. J. (1983), "The Notion of Breakdown Point," in *A Festschrift for Erich L. Lehmann*, eds. P. Bickel, K. Doksum, and J. Hodges, Belmont, CA: Wadsworth.

Hawkins, D. M. (1993), "The Feasible Set Algorithm for Least Median of Squares Regression," *Computational Statistics and Data Analysis*, 16, 81–101.

——— (1994), "The Feasible Solution Algorithm for Least Trimmed Squares Regression," *Computational Statistics and Data Analysis*, 17, 185–196.

Hawkins, D. M., and Simonoff, J. S. (1993), "High-Breakdown Regression and Multivariate Estimation," *Applied Statistics*, 42, 423–432.

Hettmansperger, T. P., and Sheather, S. J. (1992), "A Cautionary Note on the Method of Least Median Squares," *The American Statistician*, 46, 79–83.

Hössjer, O. (1994), "Rank-Based Estimates in the Linear Model With High Breakdown Point," *Journal of the American Statistical Association*, 89, 149–158.

Hössjer, O., Croux, C., and Rousseeuw, P. (1994), "Asymptotics of Generalized S Estimators," *Journal of Multivariate Analysis*, 51, 148–177.

Huber, P. J. (1964), "Robust Estimation of a Location Parameter," *Annals of Mathematical Statistics*, 35, 73–101.

Lee, A. J. (1990), *U Statistics: Theory and Practice*, New York: Marcel Dekker.

Martin, D., Yohai, V., and Zamar, R. (1989), "Min-Max Robust Regression," *The Annals of Statistics*, 17, 1608–1630.

Mason, R. L., Gunst, R. F., and Hess, J. L. (1989), *Statistical Design and Analysis of Experiments*, New York: Wiley.

McKean, W. M., Sheather, S. J., and Hettmansperger, T. P. (19??), The Use and Interpretation of Residuals Based on Robust Estimation," *Journal of the American Statistical Association*, 88, 1254–1263.

Nolan, D., and Pollard, D. (1987), "U Processes: Rates of Convergence," *The Annals of Statistics*, 15, 780–799.

Pollard, D. (1984), *Convergence of Stochastic Processes*, New York: Springer-Verlag.

Rousseeuw, P. J. (1983), "Multivariate Estimation With High Breakdown Point," *The IMS Bulletin*, 12, 234.

——— (1984), "Least Median of Squares Regression," *Journal of the American Statistical Association*, 79, 871–880.

Rousseeuw, P. (1985), "Multivariate Estimation With High Breakdown Point," in *Mathematical Statistics and Applications, Vol. B*, eds. W. Grossmann, G. Pflug, I. Vincze, and W. Wertz, Dordrecht, The Netherlands: Reidel, pp. 283–297.

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

Ruppert, D. (1992), "Computing S Estimators for Regression and Multivariate Location/Dispersion," *Journal of Computational and Graphical Statistics*, 1, 253–270.

Serfling, R. J. (1984), "Generalized L, M, and R Statistics," *The Annals of Statistics*, 12, 76–86.

Sheather, S. J., McKean, J. W., and Hettmansperger, T. P. (1997), "Finite-Sample Stability Properties of the Least Median of Squares Estimator," *Journal of Statistical Computing and Simulation*, 58, 371–383.

Stromberg, A. J. (1993), "Computing the Exact Value of the Least Median of Squares Estimate and Stability Diagnostics in Multiple Linear Regression," *SIAM Journal on Scientific and Statistical Computing*, 14, 1289–1299.

——— (1997), "Robust Covariance Estimates Based on Resampling," *Journal of Statistical Planning and Inference*, 57, 321–324.

Stromberg, A. J., Hawkins, D. M., and Hössjer, O. (1995), "The Least Trimmed Differences Regression Estimator and Alternatives," University of Lund, Dept. of Mathematical Statistics.

Yohai, V., and Maronna, R. (1976), "Location Estimators Based on Linear Combinations of Order Statistics," *Communications in Statistics, Part A—Theory and Methods*, 5, 481–486.

Yohai, V., and Zamar, R. (1992), "Optimally Bounding a Generalized Gross-Error Sensitivity," unpublished manuscript.