The Annals of Statistics
1989, Vol. 17, No. 1, 362–381

ADAPTIVE *L*-ESTIMATION FOR LINEAR MODELS[1]

BY STEPHEN PORTNOY AND ROGER KOENKER

University of Illinois at Urbana-Champaign

Asymptotically efficient (adaptive) estimators for the slope parameters of the linear regression model are constructed based upon the "regression quantile" statistics suggested by Koenker and Bassett. The estimators are natural analogues of the adaptive L-estimators of location of Sacks, but employ kernel-density type estimators of the optimal L-estimator weight function.

1. Introduction. The existence of asymptotically efficient estimators of a Euclidean parameter, β, in the presence of an infinite-dimensional nuisance parameter, F, has attracted considerable recent attention. The problem, formulated by Stein (1956) for asymptotically estimating β when F is unknown, as well as when it is known, has been treated in increasing generality. In a remarkable confluence of papers Beran (1974), Sacks (1975) and Stone (1975) independently proposed *adaptive* R-, L- and M-estimators, respectively, of the center of symmetry of an unknown (symmetric) distribution. In his 1980 Wald lectures, Bickel (1982), developing the approach of Stein, extended adaptation to a broad array of problems. In particular, he proposed an adaptive M-estimator for the parameters of the linear model

$$y_i = x_i'\beta_0 + u_i \tag{1.1}$$

with $\{x_j' = (x_{ji}, \dots, x_{jp})\}$ a sequence of known p-vectors, $\beta_0 \in \mathbf{R}^p$ an unknown regression parameter to be estimated and $\{u_i\}$ a sequence of independent random variables with common distribution function F. When F is symmetric, Bickel constructed an adaptive estimator of the entire vector β_0. Dropping the symmetry condition, he further showed that if the design contains an intercept, that is, $x_i' = (1, \dot{x}_i')$ so that

$$y_i = x_i'\beta_0 + u_i = \alpha + \dot{x}_i'\gamma + u_i, \tag{1.2}$$

then the $(p - 1)$-vector of "slope" parameters can be adaptively estimated. Manski (1984) reviewed these results and offered some extensions to nonlinear regression models. Manski and Hsieh (1987) have studied several variants of Bickel's adaptive M-estimator via Monte Carlo methods. Newey (1987) has recently proposed adaptive method-of-moment type estimators for the linear model which are asymptotically efficient under rather weak regularity conditions. Hogg (1981) has also proposed various partially adaptive methods based on

Received July 1987; revised March 1988.
[1]Research supported by NSF Grants SES-87-07169 and DMS-85-03785.
AMS 1980 *subject classifications*. 62J05, 62G35, 62F20.
Key words and phrases. Regression quantiles, kernel-density estimation, adaptive estimation, linear models.

M-estimators and de Jongh and de Wet (1986) have recently suggested an adaptive choice of the trimming proportion for trimmed least squares estimators.

In this paper we propose fully adaptive L-estimators for the slope parameters of the linear model, under the least restrictive assumptions possible on F (needed only to make the asymptotic efficiency well defined). These results extend results of Sacks (1975) to the case of linear regression and Koenker and Portnoy (1987) to the adaptive case. In the remainder of this section, we introduce notation and state our main results. Section 2 gives a detailed treatment of our construction of the adaptive estimator. Section 3 treats the problem of constructing a satisfactory estimate of the score function. Section 4 constructs a practical version of an adaptive L-estimator and describes a small Monte Carlo experiment designed to evaluate the performance of the estimator in moderate-sized samples. We conclude that a *practical* adaptive L-estimator can be constructed for the slope parameters of the linear model. The estimator achieves high finite-sample efficiency in a wide variety of error situations and outperforms standard robust methods in all situations we investigated. Substantial gains in efficiency are achieved relative to simpler robust procedures in asymmetric error situations.

Let X_n denote the $n \times p$ matrix with ith row x_i'. We will assume throughout that $n^{-1}X_n'X_n \to Q$, a positive definite matrix. The Euclidean norm of x will be denoted $\|x\|$ and $\lambda_1(M)$ will denote the largest eigenvalue for the matrix M. We will focus attention on Bickel's (1982) Example 3: the linear model (1.2) *with* an explicit intercept and *without* any symmetry condition on F. We also assume that the means have been subtracted in X_n so that $\sum \dot{x}_i = 0$, where $\dot{x}_i$ is the last $p - 1$ coordinates of x_i. Thus, if Q is partitioned so that $\dot{Q}$ is the lower $(p-1) \times (p-1)$ corner, $\dot{Q}^{-1}$ is the corresponding corner of Q^{-1}. The following regularity condition on the sequence of designs $\{X_n\}$ will be maintained.

CONDITION X. There exist positive constants b, $\bar{b}$, $\underline{b}$ and c, such that

(X1) $$\lambda_1\big(Q - n^{-1}X_n'X_n\big) \le bn^{-1/4},$$

(X2) $$\sum_{i=1}^{n} \|x_i\|^3 \le bn,$$

(X3) $$\max_i \|x_i\| \le bn^{1/4},$$

(X4) $$\inf_{\|\delta\|=1} \#\big\{i: \underline{b} \le x_i'\delta \le \bar{b}\big\} \ge cn.$$

In Portnoy (1984) it is shown that such conditions are satisfied for a broad class of random designs, as well as for ANOVA designs when the number of observations per cell tends to infinity. On F we require only:

CONDITION F. F is absolutely continuous with finite, nonzero Fisher information $I(F)$.

Let $\mathbf{F}$ be the set of F satisfying this condition. Our methods are based on the regression quantiles of Koenker and Bassett (1978) which solve for $t \in [0,1]$,

$$\min_{\beta \in \mathbf{R}^p} \sum_{t=1}^{n} \rho_t(y_i - x_i'\beta), \tag{1.3}$$

where $\rho_t(u) = u(t - I(u < 0))$. Let $\{\hat{\beta}_n(t) = (\hat{\alpha}_n(t), \hat{\gamma}_n(t)\}$ denote the sequence of regression quantile processes so defined. In the Appendix, a uniform Bahadur representation with explicit remainder is established for $\hat{\beta}_n(t)$. This result strengthens somewhat similar results of Jurečková and Sen (1984) and Koenker and Portnoy (1987).

Our adaptive estimator, T_n, of γ is a linear function of $\hat{\beta}_n(t)$, that is, we consider

$$T_n = \int_0^1 \hat{\gamma}_n(t)\hat{J}_n(t)\,dt, \tag{1.4}$$

where $\hat{J}_n(t)$ is an estimate of the *optimal* score function

$$J_0(t) = \psi'(F^{-1}(t)),$$

where $\psi(x) = -L'(x)$ and $L(x) = \ln f(x)$. Theorem 2.1 provides conditions on $\hat{J}_n(t)$ which make T_n adaptive for any F satisfying Condition F. A kernel estimator $\hat{J}_n(t)$ is constructed in Section 3 which satisfies the condition of Theorem 2.1, verifying our claim. Some further remarks on practical aspects of estimating $J_0(t)$ are contained in Section 4.

Our estimator of the optimal score function is based on the estimators of the conditional quantile and conditional distribution functions introduced in Bassett and Koenker (1982). Denoting the set of solutions to (1.3) by $\hat{B}_n(t)$, we may define a natural estimator of the tth conditional quantile of Y given x, as

$$\hat{Q}_n(t|x) = \inf\{x'b | b \in \hat{B}_n(t)\}. \tag{1.5}$$

Correspondingly,

$$\hat{F}_n(y|x) = \sup\{t \in [0,1] | \hat{Q}(t|x) \le y\}, \tag{1.6}$$

affords a natural estimator of the conditional distribution function. At the mean of the design, $\bar{x} = n^{-1}\Sigma_i x_i$, $\hat{Q}(u|\bar{x})$ is a proper quantile function (a nondecreasing, left-continuous, step function on $u \in [0,1]$ [see Bassett and Koenker (1982), Theorem 2.1]), so $\hat{F}_n(y) \equiv \hat{F}_n(y|\bar{x})$ is a proper distribution function (a nondecreasing, right-continuous step-function on $y \in \mathbf{R}$). $\hat{F}_n$ behaves asymptotically exactly like a sample distribution function [see Portnoy (1984)]. The results of Section 3 give methods of estimating $J_0(t)$, based on $\hat{F}_n(y)$ which satisfy the conditions for adaptation of T_n given in Section 2.

2. The adaptive estimators. In order to treat asymptotics for L-estimators it is necessary to have smooth, positive densities. Following Stone (1975) this may be accomplished in great generality by convolving the original error distribution with a vanishingly small smooth contaminant. In particular, define

$$\tilde{u}_i = u_i + \frac{W_i}{s} + \frac{W_i'}{t}, \qquad \tilde{Y}_i = x_i'\beta + \tilde{u}_i, \tag{2.1}$$

where $\{W_i\}$ and $\{W_i'\}$ are independent i.i.d. sequences (independent of u_i) with density

$$g(w) = \frac{c}{(1+\rho(w))^2}, \qquad -\infty < w < \infty. \tag{2.2}$$

Here $\rho(w)$ is an even continuously three times differentiable, positive function, increasing on $[0,1]$, with $\rho(w) = |w|$ for $|w| \geq 1$. Let $G(w)$ denote the c.d.f. corresponding to g and define (for $F \in \mathbf{F}$)

$$\begin{aligned} f_t(x) &= t\int g(t(x-y))\,dF(y), \\ f_s(x) &= s\int g(s(x-y))\,dF_t(y), \\ F_t(x) &= \int G(t(x-y))\,dF(y), \\ F_s(x) &= \int G(s(x-y))\,dF_t(y). \end{aligned} \tag{2.3}$$

That is, f_s and F_s are the density and c.d.f. for $\tilde{u}_i$, and f_t and F_t are the density and c.d.f. for $u_i + W_i/t$. Lastly define for fixed $\eta \leq \frac{1}{2}$ and arbitrary $b > 0$,

$$s_n = (\log n)^{\eta}, \qquad t_n = (\log n)^{b}. \tag{2.4}$$

Note that the subscript n on s_n and t_n will often be suppressed.

Furthermore, since the uniform Bahadur representation (Theorem A.1) holds only on a compact subinterval of $[0,1]$, the interval of integration must also be restricted to the subinterval. Thus, for fixed $\delta \geq 0$, $0 \leq \varepsilon \leq \delta + \eta < \frac{1}{2}$, and $a < \frac{1}{2}$, define (for $F \in \mathbf{F}$)

$$\begin{aligned} \alpha_n &= (\log n)^{-\varepsilon} + F_t\left(-\tfrac{1}{2}(\log n)^{\delta}\right) + 1 - F_t\left(+\tfrac{1}{2}(\log n)^{\delta}\right), \\ \hat{\alpha}_n &= n^{-a} + (\log n)^{-\varepsilon} + \hat{F}_n^*\left(-\tfrac{1}{2}(\log n)^{\delta}\right) + 1 - \hat{F}_n^*\left(+\tfrac{1}{2}(\log n)^{\delta}\right), \end{aligned} \tag{2.5}$$

where $\hat{F}_n^*$ is the Koenker–Bassett c.d.f. estimator [see (1.6)] based on observations $Y_i + W_i'/t_n$. Also let $\hat{F}_n$ denote the Koenker–Bassett c.d.f. estimator based on observations $\tilde{Y}_i = Y_i + W_i/s_n + W_i'/t_n$ (that is, $\hat{F}_n$ estimates F_s). Now define the adaptive (slope parameter) estimator

$$T_n = \frac{\int_{\hat{\alpha}_n}^{1-\hat{\alpha}_n} \hat{\gamma}_n(t)\hat{J}_n(t)\,dt}{\int_{\hat{\alpha}_n}^{1-\hat{\alpha}_n} \hat{J}_n(t)\,dt}, \tag{2.6}$$

where $\hat{J}_n(t)$ is any appropriately consistent estimator of the score function $J_{s_n}(t) \equiv -L_{s_n}''(F_{s_n}^{-1}(t))$. An appropriate example [satisfying (2.7)] generated by kernel estimation based on $\hat{F}_n$ is given in Section 3.

THEOREM 2.1. *Let $\hat{J}_n(t)$ be an estimator of $J_s(t)$ satisfying*

$$\int_{\alpha_n}^{1-\alpha_n} |\hat{J}_n(t) - J_s(t)|\, dt = o_p\big((\log n)^{-(2\delta+\eta)}\big). \tag{2.7}$$

Then for any $F \in \mathbf{F}$, and $\{X_n\}$ satisfying Condition X,

$$\sqrt{n}\,(T_n - \gamma) \to_D \mathbf{N}_{p-1}(0, Q^{-1}/\mathbf{I}(F)),$$

where $\mathbf{I}$ *is the Fisher information for* F.

This theorem will be proved after some preliminary properties of f_s are developed. The following lemmas each assume the hypotheses of Theorem 2.1 and that $F \in \mathbf{F}$.

LEMMA 2.1. *Given α_n defined by* (2.5), *define*

$$x_n = \max\{-F_s^{-1}(\alpha_n), F_s^{-1}(1-\alpha_n)\}. \tag{2.8}$$

Then there is a constant c^* such that for $B_n \le (\log n)^\delta$,

$$\inf_{\alpha_n} f_s(x) \equiv \inf\{f_s(x): -x_n - B_n \le x \le x_n + B_n\} \ge c^*(\log n)^{-(2\delta+\eta)}.$$

PROOF. Note that [by (2.2)]

$$\begin{aligned} F_s(-x_n) &= \int G(s(x_n - y))\, dF_t(y) \\ &\le G\big(-\tfrac{1}{2}x_n\big) + P\{|u + W/t| \ge \tfrac{1}{2}x_n\} \\ &\le c^*/\big(1 + \tfrac{1}{2}sx_n\big) + F_t\big(-\tfrac{1}{2}x_n\big) + 1 - F_t\big(\tfrac{1}{2}x_n\big) \end{aligned} \tag{2.9}$$

and a similar inequality holds for $1 - F_s(x_n)$. Hence, from (2.8), $\alpha_n = F_s(-x_n)$ or $\alpha_n = 1 - F_s(x_n)$, and if x_n were larger than $(\log n)^\delta$, (2.9) would be contradicted by (2.5) (for n large enough). Thus, it follows that (for n large enough)

$$0 \le x_n \le (\log n)^\delta \quad \text{and} \quad x_n \to +\infty \tag{2.10}$$

[since $\alpha_n \to 0$ by (2.5)]. Now (for $x > 0$)

$$f_s(x) = \int \frac{cs}{(1 + \rho(s(x-y)))^2}\, dF_t(y) \ge \frac{cs}{(1 + \rho(sx))^2} P\left\{u + \frac{W}{t} \le 0\right\}$$

and, hence, for $|x| \le 2(\log n)^\delta$, with $c^* = P\{u + W/t \le 0\}$,

$$f_s(x) \ge \frac{c^* s}{\big(1 + \rho\big(2s(\log n)^\delta\big)\big)^2} = \frac{c^* s}{\big(1 + 2s(\log n)^\delta\big)^2}$$

and the result follows from (2.4). □

LEMMA 2.2. *For constants c_ν ($\nu = 0, 1, 2, 3$) with $c_0 = c$ in* (2.2),

$$|f_s^{(\nu)}(x)| \le c_\nu s^{\nu+1} \quad \textit{and} \quad |f_t^{(\nu)}(x)| \le c_\nu t^{\nu+1}$$

uniformly in x.

PROOF. Differentiate $f_s(x)$ or $f_t(x)$ [see (2.3)] under the integral and use the fact that derivatives of ρ are uniformly bounded. □

LEMMA 2.3. *$f_{s_n}(x_n) \to 0$ and $f_{s_n}'(x_n) \to 0$ as $n \to \infty$.*

PROOF. As in (2.9) [using (2.4) and (2.10)],

$$|f_s(x_n)| \le cs/\left(1 + \tfrac{1}{2}sx_n\right)^2 + \tilde{c}\left\{F_t\left(-\tfrac{1}{2}x_n\right) + 1 - F_t\left(\tfrac{1}{2}x_n\right)\right\} \to 0,$$

$$|f_s'(x_n)| \le c^* s^2/\left(1 + \tfrac{1}{2}sx_n\right)^3 + \tilde{c}^*\left\{F_t\left(-\tfrac{1}{2}x_n\right) + 1 - F_t\left(\tfrac{1}{2}x_n\right)\right\} \to 0. \qquad \square$$

LEMMA 2.4. *$\int_{\alpha_n}^{1-\alpha_n} J_s(t)\,dt \to \mathbf{I}(F)$ as $n \to \infty$.*

PROOF. A slight modification of the proof of Theorem 4.1 in Stone (1975) provides the result here. □

LEMMA 2.5. *As $n \to \infty$,*

$$\int_{\alpha_n}^{1-\alpha_n} |\hat{J}_n(t)|\,dt = O_p\left(s_n^2\right) = O_p\left((\log n)^{2\eta}\right).$$

PROOF. By condition (2.7), we need only consider

$$\begin{aligned} \int_{\alpha_n}^{1-\alpha_n} |J_s(t)|\,dt &\le \int_{-x_n}^{x_n} |L_s''(x)|\, f_s(x)\,dx \\ &\le \int_{-x_n}^{x_n} |f_s''(x)|\,dx + \int_{-x_n}^{x_n} \left(L_s'(x)\right)^2 f_s(x)\,dx. \end{aligned} \tag{2.11}$$

Now differentiating f_s in (2.3) twice (under the integral) and using the fact that derivatives of ρ are bounded,

$$\begin{aligned} |f_s''(x)| &\le \int \frac{2cs^3|\rho''(s(x-y))|}{\left(1 + \rho(s(x-y))\right)^3}\,dF_t(y) + \int \frac{3cs^3\left(\rho'(s(x-y))\right)^2}{\left(1+\rho(s(x-y))\right)^4}\,dF_t(y) \\ &\le c_1 s^2 \int \frac{cs}{\left(1+\rho(s(x-y))\right)^2}\,dF_t(y) = c_1 s^2 f_s(x). \end{aligned}$$

Hence, the first term in (2.11) is $O(s^2)$. The last term in (2.11) converges to $\mathbf{I}(F)$ by Lemma 2.4 and, thus, the desired result follows. □

LEMMA 2.6. *Let α_n, $\hat{\alpha}_n$ and a be given by (2.5) and assume $F \in \mathbf{F}$. Then with probability tending to 1,*

$$\alpha_n \le \hat{\alpha}_n \le \alpha_n + 2n^{-a}.$$

PROOF. By Proposition 3.1, $|\hat{F}_n^*(x) - F_t(x)| = O_p(n^{-1/2})$ uniformly for $F_s^{-1}(\alpha_n) \le x \le F_s^{-1}(1-\alpha_n)$. By (2.10) and (2.8), $\pm\,\frac{1}{2}(\log n)^{\delta}$ lies in this interval, and the result follows immediately. □

LEMMA 2.7. *Let* $S_n = (\alpha_n, \hat{\alpha}_n) \cup (1 - \hat{\alpha}_n, 1 - \alpha_n)$. *Then* $\int_{S_n} |J_s(t)|\, dt = O(n^{-a/2})$ *as* $n \to \infty$.

PROOF. Following the argument of Lemma 2.5 and using Lemma 2.6, with probability tending to 1,

$$\int_{\alpha_n}^{\hat{\alpha}_n} |J_s(t)|\, dt \le c_1 s^2 \big(F_s^{-1}(\alpha_n + 2n^{-a}) - F_s^{-1}(\alpha_n)\big) + \int_{-x_n}^{F_s^{-1}(\alpha_n + 2n^{-a})} \big(L_s'(x)\big)^2 f_s(x)\, dx.$$

From Lemma 2.1 (and the mean value theorem) the first term is of order $O(\log n)^b n^{-a} = O(n^{-a/2})$. A similar argument shows that the second term has this same order. The same argument applies to the integral from $1 - \hat{\alpha}_n$ to $1 - \alpha_n$. □

PROOF OF THEOREM 2.1. From (2.6),

$$\sqrt{n}\,(T_n - \gamma) = \frac{\int_{\hat{\alpha}_n}^{1-\hat{\alpha}_n} \sqrt{n}\,(\hat{\gamma}_n(t) - \gamma)\hat{J}_n(t)\, dt}{\int_{\hat{\alpha}_n}^{1-\hat{\alpha}_n} \hat{J}_n(t)\, dt} \equiv \frac{A_n}{B_n}. \tag{2.12}$$

By Lemmas 2.4 and 2.7 and condition (2.7), the denominator, B_n, tends to $\mathbf{I}(F)$ in probability; so it remains to consider the numerator. Define

$$U_n = \frac{1}{\sqrt{n}} \dot{Q}^{-1} \sum_{i=1}^{n} \dot{x}_i K_{in}(t), \qquad K_{in}(t) = t - I\big(\tilde{u}_i \le F_s^{-1}(t)\big). \tag{2.13}$$

Then, by Theorem A.1,

$$\begin{aligned} &\sqrt{n}\left|\hat{\gamma}_n(t) - \gamma - (1/\sqrt{n})U_n(t)\big(f_s(F_s^{-1}(t))\big)^{-1}\right| \\ &\quad \le \big(n^{-1/4}(\log n)B(X, F_s) + n^{-1/2} b_1(X)\big)/f_s\big(F_s^{-1}(t)\big) \end{aligned} \tag{2.14}$$

on $(\alpha_n, 1 - \alpha_n)$ except with probability bounded by $q(X, F)$ (see Lemma A.3). By Lemmas 2.1 and 2.2, uniformly on $(\alpha_n, 1 - \alpha_n)$,

$$\frac{B(X, F_s)}{f_s(F_s^{-1}(t))} = O\big((\log n)^{\eta + 3(2\delta + \eta)}\big),$$

$$q(X, F) = O\Big(n^{-1/2} \exp\big(c^2 b_4 (\log n)^{2\delta + \eta}\big)\Big) \to 0.$$

Therefore (using Theorem A.1), with probability tending to 1, the numerator in (2.12) satisfies (since $\alpha_n \le \hat{\alpha}_n$ in probability)

$$\begin{aligned} \left| A_n - \int_{\alpha_n}^{1-\alpha_n} \frac{U_n(t) J_s(t)}{f_s(F_2^{-1}(t))}\, dt \right| &\le \left| \int_{\hat{\alpha}_n}^{1-\hat{\alpha}_n} \frac{U_n(t)\hat{J}_n(t)}{f_s(F_s^{-1}(t))}\, dt - \int_{\alpha_n}^{1-\alpha_n} \frac{U_n(t) J_s(t)}{f_s(F_s^{-1}(t))}\, dt \right| \\ &\quad + O_p\big(n^{-1/4}(\log n)\big) \int_{\alpha_n}^{1-\alpha_n} |\hat{J}_n(t)|\, dt \\ &\quad + O_p\big(n^{-1/2}\big) \int_{\alpha_n}^{1-\alpha_n} \frac{|\hat{J}_n(t)|}{f_s(F_s^{-1}(t))}\, dt, \end{aligned}$$

where the last two terms arise from the error terms in (2.14). Bounding the integrands yields

$$\left| A_n - \int_{\alpha_n}^{1-\alpha_n} \frac{U_n(t) J_s(t)}{f_s(F_s^{-1}(t))} dt \right| \le O_p\left(n^{-1/4}(\log n)\right) \frac{\int_{\alpha_n}^{1-\alpha_n} |\hat{J}_n(t)|\, dt}{\inf_{\alpha_n} f_s(x)}$$

$$\text{(2.15)} \qquad + \frac{\sup_t |U_n(t)|}{\inf_{\alpha_n} f_s(x)} \int_{\alpha_n}^{1-\alpha_n} |\hat{J}_n(t) - J_s(t)|\, dt$$

$$+ \frac{\sup_t |U_n(t)|}{\inf_{\alpha_n} f_s(x)} \int_{S_n} |J_s(t)|\, dt,$$

where $S_n + (\alpha_n, \hat{\alpha}_n) \cup (1 - \hat{\alpha}_n, 1 - \alpha_n)$. By condition (2.7) and Lemma 2.5, $\int_{\alpha_n}^{1-\alpha_n} |\hat{J}_n(t)|\, dt = O_p(s_n^2)$ and, hence, by (2.4) and Lemma 2.1, the first term in (2.15) tends to 0 in probability. Using an invariance principle for $U_n(t)$ [e.g., see Koul (1969), Theorem A.3], $\sup_t |U_n(t)| = O_p(1)$. Thus, combining Lemma 2.1 and condition (2.7), the second term in (2.15) also tends to 0 in probability. Last, the third term converges to 0 by Lemmas 2.1 and 2.7. Therefore, the right side of (2.15) tends to 0 in probability; it remains to consider

$$V_n \equiv \int_{\alpha_n}^{1-\alpha_n} U_n(t) J_s(t) / f_s(F_s^{-1}(t))\, dt.$$

Fix $t \in R^p$ and consider $t'V_n$. Define $a_{in} = t'\dot{Q}^{-1}\dot{x}_i / \sqrt{n}$. Then

$$\text{(2.16)} \qquad \sum_{i=1}^{n} a_{in}^2 \to t'\dot{Q}^{-1}t \quad \text{as } n \to \infty$$

and $t'V_n$ is a weighted sum of n i.i.d. random variables [see (2.13)]:

$$t'V_n = \sum_{i=1}^{n} a_{in} \int_{\alpha_n}^{1-\alpha_n} K_{in}(t) J_s(t) / f_s(F_s^{-1}(t))\, dt.$$

To apply the Liapounov central limit theorem, compute third moments: Since $|K_{in}(t)| \le 2$,

$$E|t'V_n|^3 \le 8 \sum_{i=1}^{n} |a_{in}|^3 \left\{ \int_{\alpha_n}^{1-\alpha_n} |J_s(t)| / f_s(F_s^{-1}(t))\, dt \right\}^3 \le \sum_{i=1}^{n} |a_{in}|^3 O\left((\log n)^{2\delta + 3\eta}\right),$$

where Lemmas 2.1 and 2.5 are applied. Last, from Condition (X3), the definition of a_{in} and (2.16),

$$E|t'V_n|^3 \le \sum a_{in}^2 O\left(n^{-1/4} (\log n)^b\right) \to 0 \quad \text{as } n \to \infty.$$

So by Liapounov's theorem [e.g., see Breiman (1968), page 275], it remains to check that the variance converges to $t'\dot{Q}^{-1} t \mathbf{I}(F)$. Direct calculation gives

$$\sigma_n \equiv \frac{\operatorname{Var} t'V_n}{t'\dot{Q}^{-1}t} = \int_{\alpha_n}^{1-\alpha_n} \int_{\alpha_n}^{1-\alpha_n} \frac{\min(t, t') - tt'}{f_s(F_s^{-1}(t)) f_s(F_s^{-1}(t'))} J_s(t) J_s(t')\, dt\, dt'$$

$$= \int_{x_n}^{y_n} \int_{x_n}^{y_n} \{\min(F_s(x), F_s(y)) - F_s(x) F_s(y)\} L_s''(x) L_s''(y)\, dx\, dy,$$

where $x_n = F_s^{-1}(\alpha_n)$, $y_n = F_s^{-1}(1-\alpha_n)$. Let σ_s denote the above double integral with $x_n = -\infty$ and $y_n = \infty$. Then $\sigma_s - \sigma_n$ can be expressed as the sum of integrals over rectangles disjoint from $(x_n, y_n) \times (x_n, y_n)$. Consider one such integral, the integral over $(-\infty, x_n) \times (x_n, \infty)$. Integrating by parts,

$$\left|\int_{-\infty}^{x_n}\int_{x_n}^{\infty} F_s(y)(1-F_s(x))L_s''(x)L_s''(y)\,dx\,dy\right|$$
$$= \left|\{-(1-F_s(x_n))L_s'(x_n) - f_s(x_n)\}\{F_s(x_n)L_s'(x_n) - f_s(x_n)\}\right|$$
$$\le (L_s'(x_n))^2F_s(x_n) + |L_s'(x_n)|f_s(x_n) + f_s^2(x_n).$$

By Lemma 2.3, $f_s^2(x_n) \to 0$ and $|L_s'(x_n)|f_s(x_n) = |f_s'(x_n)| \to 0$ as $n \to \infty$. Also by L'Hospital's rule,

$$\begin{aligned}\lim F_s(x_n)(L_s'(x_n))^2 &= \lim F_s(x_n)L_s'(x_n)\lim\frac{f_s'(x_n)}{f_s(x_n)}\\ &= \lim F_s(x_n)L_s'(x_n)\lim\frac{f_s(x_n)}{F_s(x_n)}\\ &= \lim f_s'(x_n) = 0\end{aligned}$$

by Lemma 2.3. Treating other contributions to $|\sigma_n - \sigma_s|$ similarly, we see that $|\sigma_n - \sigma_s| \to 0$ as $n \to \infty$. But integrating by parts, $\sigma_s = \int_{-\infty}^{\infty} L_s''(x)f_s(x)\,dx \to \mathbf{I}(F)$ as $n \to \infty$ (by Lemma 2.4). Therefore, $\sigma_n \to \mathbf{I}(F)$ and, hence, $V_n \to_D \mathbf{N}_{p-1}(0, \mathbf{I}(F)\dot{Q}^{-1})$. As noted above, this implies A_n [in (2.12)] has the same limiting distribution. Therefore $A_n/B_n \to_D \mathbf{N}_{P-1}(0, \dot{Q}/\mathbf{I}(F))$ and the proof is complete. □

3. An appropriate estimator of the score function. Here, as in Section 2, we assume that the errors are distributed according to F_s defined in (2.3) with $F_s \in \mathbf{F}$. For such smooth F_s, it is relatively easy to construct an estimator $\hat{J}_n(t)$ of $J_s(t)$ satisfying (2.7) by using appropriate density estimators based on $\hat{F}_n$. Since (2.7) requires only logarithmic convergence, the following conditions on the density estimators will be seen to be sufficient. Let $s_n = (\log n)^{\eta}$ [as in (2.4)],

$$U_n = \{x: F_s^{-1}(\alpha_n) - B \le x \le F_s^{-1}(1-\alpha_n) + B\} \tag{3.1}$$

for any constant B and define $K_n = (\log n)^{-(2\delta+\eta)}$ [so that, by Lemma 2.1, $\{\inf_{U_n} f_s(x)\}^{-1} = O(1/K_n)$]. Suppose there are density estimators, $\hat{f}_n(x)$ [with derivatives $\hat{f}_n^{(\nu)}(x)$] and (smooth) c.d.f. estimators, $\tilde{F}_n(x)$ (generally the integral of $\hat{f}_n$) such that for $\nu = 0, 1, 2, 3$,

$$\sup_{U_n}|\hat{f}_n^{(\nu)}(k) - f_s^{(\nu)}(x)| = o_p\big((s_n/K_n)^{-5}\big) = o_p\big((\log n)^{-10(\delta+\eta)}\big), \tag{3.2}$$

$$\sup_{U_n}|\tilde{F}_s^{-1}(F_s(x)) - x| = o_p\big((s_n/K_n)^{-4}\big) = o_p\big((\log n)^{-8(\delta+\eta)}\big), \tag{3.3}$$

where f_s and F_s are given by (2.3). As in Section 2, define $L_s(x) = \log f_s(x)$, $\hat{L}_n(x) = \log \hat{f}_n(x)$, $J_s(t) = -L_s''(F_s^{-1}(t))$ and $\hat{J}_n(t) = -\hat{L}_n''(\tilde{F}_n^{-1}(t))$.

Lemma 3.1. *If* (3.2) *holds, then*

$$\sup_{U_n}|\hat{L}_n''(x) - L_s''(x)| = O_p\big((\log n)^{-(2\delta+\eta)}\big).$$

Proof. First note that by (3.2) and Lemma 2.1,

$$\inf_{U_n} \hat{f}_n(x) \ge c^* K_n - o_p(1). \tag{3.4}$$

Hence, $\{\inf_{U_n} \hat{f}_n(x)\}^{-1} = O_p(1/K_n)$. Similarly, by (3.2) and Lemma 2.2, we also have

$$|\hat{f}_n^{(\nu)}(x)| \le c_\nu s^{\nu+1} \quad \text{for } x \in U_n. \tag{3.5}$$

Therefore, letting Δf denote absolute differences between $\hat{f}_n$ and f_s (and their derivatives) and (with n suppressed) writing $L''(x) = -[f''(x)/f(x) - (f'(x)/f(x))^2]$,

$$\begin{aligned}
&\sup_{U_n}|\hat{L}_n''(x) - L_s''(x)| \\
&\quad \le \left(\frac{\Delta f''}{K} + \frac{\sup f'' \Delta f}{K^2} + \frac{\sup(\hat{f}' + f')\Delta f'}{K^2} + \frac{\sup f'^2 \sup(\hat{f} + f)\Delta F}{K^4} \right) \\
&\quad = O_p\left(\frac{s_n^5}{K_n^4}\right) o_p\left(\left(\frac{s_n}{K_n}\right)^{-5}\right) \\
&\quad = o_p(K_n). \qquad \square
\end{aligned}$$

Theorem 3.1. *If* (3.2) *and* (3.3) *hold, then*

$$\int_{\alpha_n}^{1-\alpha_n} |\hat{J}_n(t) - J_s(t)|\, dt = o_p\big(\log n^{-(2\delta+\eta)}\big).$$

Proof. Changing variables using $t = F_s^{-1}(x)$ and letting $x_n = F_s^{-1}(\alpha_n)$, $y_n = F_s^{-1}(1 - \alpha_n)$,

$$\begin{aligned}
\int_{\alpha_n}^{1-\alpha_n} |\hat{J}_n(t) - J_s(t)|\, dt &= \int_{x_n}^{y_n} |\hat{L}_n''(\tilde{F}_n^{-1}(F_s(x))) - L_s''(x)| f_s(x)\, dx \\
&\le \int_{x_n}^{y_n} |\hat{L}_n''(x) - L_s''(x)| f_s(x)\, dx \\
&\quad + \int_{x_n}^{y_n} \left| \int_x^{\tilde{F}^{-1}(F_s(x))} \hat{L}'''(u)\, du \right| f_s(x)\, dx.
\end{aligned}$$

The first term has the desired order by Lemma 3.1. By Lemmas 2.1 and 2.2 and (3.4) and (3.5),

$$\sup_{U_n}|\hat{L}'''(x)| = O_p\big(s_n^4/K_n^3\big). \tag{3.6}$$

Hence, by conditions (3.3) and (3.6), the inner integral in the second term above is $O_p(s_n^4/K_n^3) o_p((s_n/K_n)^{-4}) = o_p(K_n)$, and the result follows. $\square$

Last, estimates $\hat{f}_n$ satisfying (3.2) and (3.3) need to be constructed. In fact, it is generally easy to construct estimates where the error terms are even smaller than those required in conditions (3.2) and (3.3). For example, if there is a c.d.f. estimator, $\hat{F}_n(x)$, satisfying

$$\sup_{U_n}|\hat{F}_n(x) - F_s(x)| = O_p(n^{-a}) \quad \text{for some } a > 0, \tag{3.7}$$

then kernel estimators satisfying (3.2) and (3.3) can be constructed (and similarly for estimating F_t). We first show that (3.7) holds for $a = \frac{1}{2}$ for the Koenker–Bassett c.d.f. estimator, $\hat{F}_n$, given by (1.6) based on observations $\tilde{Y}_i$. However, it is no harder to show that the empirical distribution of residuals from any estimator, $\hat{\beta}$ (with $\hat{\beta}$ consistent at rate n^{-a}), will also satisfy (3.7).

Proposition 3.1. *Assume that the result of Theorem* A.1 *holds. Then condition* (3.7) *holds for* $F \in \mathbf{F}$ *with* $a = \frac{1}{2}$.

Proof. By Theorem A.1 and Lemmas 2.1 and 2.2,

$$\sup_{U_n}|\hat{F}_n(x) - F_s(x)| \le \sup_{U_n}\left|\frac{1}{n}\sum_{i=1}^{n} I(\tilde{u}_c \le x) - F_s(x)\right| + O_p\big(n^{-3/4}(\log n)^b\big)$$

for some $b > 0$. By Kolmogorov's result [e.g., see Breiman (1968), page 287] the sup on the right is $O_p(n^{-1/2})$, and, hence, (3.7) holds. The same argument works for $|\hat{F}_n^*(x) - F_t(x)|$, where $\hat{F}_n^*$ is based on $Y_i + W_i/t$. □

Now, let $k(x)$ be a kernel which is a (symmetric) density with support in $[-1, 1]$ such that $|k^{(\nu)}(x)| \le b$ (for some $b > 0$) uniformly for all x and $\nu =$ 0, 1, 2, 3, 4. Given $\hat{F}_n(x)$ satisfying (3.7) define

$$\begin{aligned} \hat{f}_n(x) &= r_n \int_{-\infty}^{\infty} k(r_n(x-y))\, d\hat{F}_n(y), \\ \tilde{F}_n(x) &= \int_{-\infty}^{x} \hat{f}_n(x)\, dx, \end{aligned} \tag{3.8}$$

where

$$r_n = n^{a_0} \quad \text{with } a_0 < a/4. \tag{3.9}$$

Lemma 3.2. *If* (3.7) *holds, then* (3.2) *holds for estimates given by* (3.8).

Proof. Integrating by parts, for $\nu = 0, 1, 2, 3$,

$$\hat{f}_n^{(\nu)}(x) = r^{\nu+2} \int_{-\infty}^{\infty} k^{(\nu+1)}(r(x-y))\hat{F}_n(y)\, dy.$$

Therefore,

$$\begin{aligned} &|\hat{f}_n^{(\nu)}(x) - f_s^{(\nu)}(x)| \\ &\quad \le r^{\nu+2} \int_{-\infty}^{\infty} k^{(\nu+1)}(r(x-y))\, |\hat{F}_n(y) - F_s(y)|\, dy \\ &\qquad + \left| \int_{-\infty}^{\infty} r^{\nu+2} k^{(\nu+1)}(r(x-y)) F_s(y)\, dy - f_s^{(\nu)}(x) \right|. \end{aligned} \tag{3.10}$$

By (3.7) and the conditions on k, the supremum of the first term above is of order $r^{\nu+1}O_p(n^{-a})$, which decreases as a power of n by (3.9). For the second term, integrating by parts yields

$$\begin{aligned}\int_{-\infty}^{\infty} r^{\nu+2}k^{(\nu+1)}(r(x-y))F_s(y)\,dy &= \int_{-\infty}^{\infty} rk(r(x-y))f_s^{(\nu)}(y)\,dy \\ &= \int_{-\infty}^{\infty} k(u)f_s^{(\nu)}\left(x-\frac{u}{r}\right)du \\ &= f_s^{(\nu)}(x) - \frac{1}{r}\int_{-\infty}^{\infty} uk(u)f_s^{(\nu+1)}(\tilde{x}(u))\,du.\end{aligned}$$

Thus, by Lemma 2.2 and the conditions on k,

$$\sup_{U_n}\left|\int_{-\infty}^{\infty} r^{\nu+2}k^{(\nu+1)}(r(x-y))F_s(y)\,dy - f_s^{(\nu)}(y)\right| = O\left(\frac{s^{\nu+2}}{r}\right)$$

and, hence, the supremum of the second term in (3.10) also decreases as a power of n. Thus, (3.2) holds, in fact, with an error of order n^{-a^*} with $a^* < \min(a_0, a - 4a_0)$, where a_0 is defined in (3.9). □

LEMMA 3.3. *If* (3.7) *holds with* B *replaced by* $3B$ *in the definition of* U_n [(3.1)], *then* (3.3) *holds for estimates given by* (3.8).

PROOF. Let $\mathrm{U}_n(B)$ denote the set U_n in (3.1) with dependence on B explicit and define

$$D_n = \sup_{U_n(3B)}|\hat{F}_n(x) - F_s(x)| = O_p(n^{-a}).$$

Let $\varepsilon > 0$ be given and choose n large enough so that by Lemmas 2.1, 3.7 and 3.9,

$$1/r + D_n/\inf_{U_n(3B)} f_s(x) \le cn^{-a_1} \le B$$

for some $a_1 < a_0$ and constant c, with probability at least $1 - \varepsilon$. Then since the support of k is contained in $[-1, 1]$, (3.8) implies that for $y \in U_n(2B)$, with probability at least $1 - \varepsilon$ (for n large enough),

$$\begin{aligned}\tilde{F}_n(y) \le \hat{F}_n(y + 1/r) &\le F_s(y + 1/r) + D_n \\ &\le F_s\big(y + 1/r + D_n/\inf_{U_n(3B)} f_s(x)\big).\end{aligned}$$

Now let $x = y + 1/r + D_n/\inf_{U_n(3B)} f_s(x)$. Then for $x \in U_n(B)$,

$$x - 1/r - D_n/\inf_{U_n(3B)} f_s(x) \le \tilde{F}_n^{-1}(F_s(x))$$

or

$$x \le \tilde{F}_n^{-1}F_s(x) + cn^{-a_1}$$

with probability at least $1 - \varepsilon$ for n large enough. The reverse inequality follows similarly and, hence, the result holds. □

4. Practical experience. To assess the performance of adaptive L-estimation in practical applications, a small scale Monte Carlo experiment was conducted. Before describing the experiment in detail, we should explicitly describe the version of the adaptive estimator (1.4) as it is employed in the experiment.

In Section 3, it is shown that the estimator $\hat{F}_n(y) \equiv \hat{F}_n(y|\bar{x})$ defined in (1.6) and described in detail in Bassett and Koenker (1982) and Portnoy (1984) satisfies the condition

$$\sup_{y \in U_n} |\hat{F}_n(y) - F_s(y)| = O_p(n^{-1/2}) \tag{4.1}$$

and $F_s(y)$ defined in Section 2, for U_n given in (3.1), and further, that kernel density estimators of f_s and its derivatives based on $\hat{F}_n(y)$ can be used to achieve the sufficient condition (2.7) for an adaptive $\hat{J}_n(t)$ required by the estimator defined in (1.4).

Rather than randomly perturbing the observed y's as suggested by the theory of Sections 2 and 3, we have chosen instead to smooth $\hat{F}_n(y)$ directly by kernel methods. For an appropriate choice of the kernel, this may be viewed as taking expectations with respect to the randomized estimator treated in Section 2 [cf. Stone (1975)]. Obviously, any sufficiently small amount of initial "dithering" would have no appreciable effect on the reported results. In all other respects the implementation of the estimators reported here is identical to the construction in Sections 2 and 3 above. $\hat{F}_n(y)$ takes the form

$$\hat{F}_n(y) = \sum_{i=1}^{n} p_i I(y \geq \xi_i) \tag{4.2}$$

for numbers $0 < p_1 < p_1 + p_2 < \cdots < \sum_{i=1}^{m-1} p_i < 1$ and $\xi_1 < \xi_2 < \cdots < \xi_m$. So, we may write kernel estimates of the density and its derivatives as

$$\hat{f}_n^{(\nu)}(x) = \sum_{i=1}^{m} p_i r_{in}^{\nu+1} k^{(\nu)}(r_{in}(x - \xi_i)), \tag{4.3}$$

where $k(\cdot)$ denotes a proper kernel and r_{in}^{-1}: $i = 1, \ldots, m$ are local bandwidth numbers which control the degree of smoothness of the estimate. The latter are chosen by the procedure outlined in Silverman (1986), pages 101–102. A pilot estimate, $\tilde{f}(x)$, of the density is constructed based on a fixed bandwidth, say h. Then the local bandwidth factors

$$\lambda_i = \left[\tilde{f}(\xi_i)/g\right]^{-\sigma}$$

are computed with $\log g = \sum p_i \log \tilde{f}(\xi_i)$. The sensitivity parameter, σ, controls the responsiveness of the local bandwidths

$$r_{in} = (h\lambda_i)^{-1}$$

to the pilot density. We have adopted the (standard) choice $\sigma = \frac{1}{2}$ after some brief experimentation with other values.

The choice of the kernel $k(\cdot)$ is critical to the success of the method. Guided by the theory of Section 2 we have chosen the Cauchy kernel,

$$k(x) = \left(\pi(1 + x^2)\right)^{-1},$$

which has the salient characteristic that it tends to control the tail behavior of our estimated $J(\cdot)$ much more successfully than more conventional, thinner-tailed kernels.

Given the estimates (4.3), it is natural to define

$$\tilde{J}_n(t_i) = \left(\frac{\hat{f}_n^{(1)}(\xi_i)}{\hat{f}_n(\xi_i)}\right)^2 - \left(\frac{\hat{f}_n^{(2)}(\xi_i)}{\hat{f}_n(\xi_i)}\right), \qquad i = 1, 2, \dots, m,$$

where $t_i = \sum_{j=1}^{i} p_j$ is the cumulative mass associated with the quantile ξ_i. In theory *and practice* it is essential to trim the tails of the weight function so for a sequence $\alpha_n \to 0$, as $n \to \infty$, we compute

$$\hat{J}_n(t_i) = \tilde{p}_i \tilde{J}_n(t_i) \Big/ \sum_{j=1}^{m} \tilde{p}_j \tilde{J}_n(t_j) \tag{4.4}$$

with $\tilde{p}_i = \max(\min(t_i, 1 - \alpha_n) - \max(t_{i-1}, \alpha_n), 0)$.

It remains only to describe the choice of the initial window width h and the trimming proportion α. The latter is straightforward; we simply report results for *both* of the traditional trimming proportions $\alpha = 0.05$ and $\alpha = 0.1$. The theory of Section 3 suggests that $\alpha_n \to 0$ as a negative power of $\log n$; thus these traditional values should be reasonable for a wide range of sample sizes (say $n < 1000$). The choice of h is a delicate issue and warrants considerable further investigation. We began with a conventional rule for density estimation [see Silverman (1986), Section 3.4],

$$h = \kappa \min(s_1, s_2)/n^{1/5},$$

where s_1 and s_2 are alternative estimates of the dispersion of $\hat{F}_n(y)$: standard deviation and (interquartile range)/1.34, respectively, and κ is a constant to be determined. The choice $\kappa = 0.9$ tuned to minimizing integrated mean-squared error of the normal density is clearly inappropriate in the present instance. Virtually imperceptible bulges in $\hat{f}$ give rise to violent oscillations in $\hat{J}$. We have adopted $\kappa = 2.5$ provisionally, although this tends to oversmooth to a significant degree in some cases. In Figures 1 and 2 we illustrate several estimated J functions for the Gaussian and Cauchy cases, respectively, for a bivariate linear model with 100 observations. The smooth curves in each case depict the "true" J.

We should emphasize at this point that many of the choices described above may be easily criticized. Indeed the choice of kernel estimation of J is itself questionable. Cox (1985) has proposed an elegant smoothing spline approach to the estimation of $-f'(x)/f(x)$ which may prove attractive in the present instance as well, if a satisfactory approach can be found for controlling the tail behavior of the estimator. In some preliminary experiments we found this to be difficult. Clearly, many alternatives exist to the particular choice of initial and

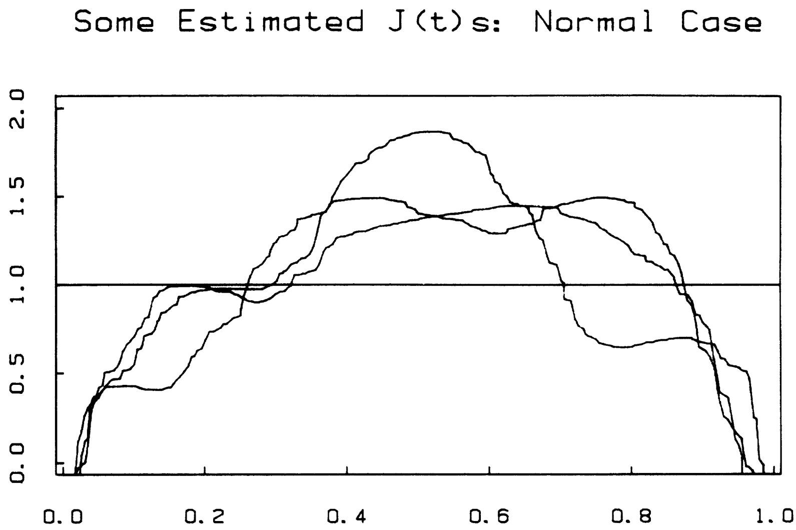

FIG. 1. *Three $\hat{J}$'s with Gaussian errors.*

local bandwidths described above. We regard the current methods as simply illustrative of one approach which yields quite promising results.

The experiment is limited to the bivariate linear model,

$$y_i = \alpha + \beta x_i + u_i,$$

with the x_i drawn as i.i.d. Gaussian and u_i also i.i.d. from one of the distribu-

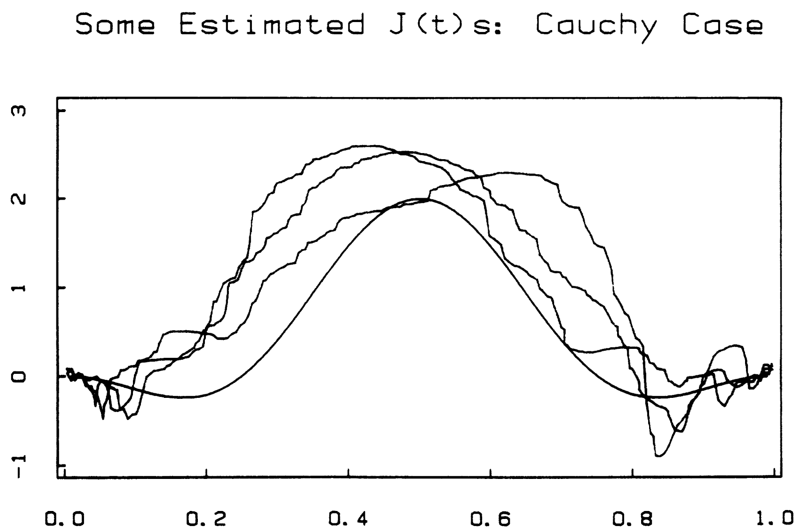

FIG. 2. *Three $\hat{J}$'s with Cauchy errors.*

TABLE 1
Distributions, densities and their optimal J's

Name	Density[1]	Optimal J[2]
Gaussian	$\phi(x)$	1
Cauchy	$(\pi(1+x^2))^{-1}$	$\cos(2\pi u)(\cos(2\pi u)-1)$
Uniform	$I_{[0,1]}(x)$	$0.5\delta_0(u)+0.5\delta_1(u)$
Laplace	$\frac{1}{2}e^{-\lvert x\rvert}$	$\delta_{1/2}(u)$
Exponential	$e^{-x}, x \geq 0$	$\delta_0(u)$
Lognormal	$x^{-1}\phi(\log x)$	$-\log(\Phi^{-1}(u)(u))/\Phi^{-1}(u)^2(u))$
Bimodal	$0.5\phi(x-3)+0.5\phi(x+3)$	$1+9\left[\frac{\left(\phi(\Phi^{-1}(u)+3)-\phi(\Phi^{-1}(u)-3)\right)^2}{\left(\phi(\Phi^{-1}(u)+3)+\phi(\Phi^{-1}(u)-3)\right)^2}-1\right]$

[1] $\phi(x) = (2\pi)^{-1/2}e^{-x^2/2}$.
[2] $\delta_x(u)$ denotes the Dirac density with point mass 1 at x, $\Phi(z) = \int_{-\infty}^{z}\phi(x)\,dx$.

tions appearing in Table 1. Since asymmetric distributions are of substantial interest we restrict attention to the relative performance of several estimators of the *slope* parameter, β. To control computing costs we restrict attention to only a few competing estimators. Once the regression quantile process, $\hat{\beta}(t)$, implicitly defined in (1.3) has been computed, it is easy to compute a variety of L-estimators. See Koenker and D'Orey (1987) for a detailed description of the algorithm used to compute $\hat{\beta}(t)$. For example, the analogues of the trimmed means

$$\hat{\beta}_\alpha = (1-2\alpha)^{-1}\int_\alpha^{1-\alpha}\hat{\beta}(t)\,dt,$$

termed "trimmed regression quantiles" (TRQ) are readily calculated as in (4.4) setting $\tilde{J}_n(t) \equiv 1$ on $(\alpha, 1-\alpha)$ and 0 otherwise. These estimators are, asymptotically, closely related to the Huber M-estimators. We consider three members of this family: TRQ(0.5), the l_1-estimator; TRQ(0.25), a regression midmean; and TRQ(0.1), the 10% trimmed regression quantile. Finally, we compute the ordinary least squares estimator (l_2) and the maximum likelihood estimator (MLE). In the Laplace and exponential cases the MLE's are the regression quantiles $\hat{\beta}(\frac{1}{2})$ and $\hat{\beta}(0)$, respectively. In the uniform case the MLE is the l_∞-estimator which minimizes the maximum residual and can be computed by linear programming methods. The Cauchy, lognormal and bimodal MLE's were computed by generic numerical optimization methods using the PORT3 routine MNF by Gay (1983).

In Table 2 we report Monte Carlo relative efficiencies for each of these estimators based on 10,000 trials. The reported efficiencies are all relative to the maximum likelihood estimator. The random number generator was the portable version of the Marsaglia generator as implemented in the PORT3 library [Fox (1984)], so results should be reproducible (up to differences in machine precision) across machines given the seeds used here. The computations were carried out on the Cray XMP-48 at the National Center for Supercomputing Applications at the University of Illinois.

TABLE 2
Monte Carlo efficiencies of various estimators of the slope parameter of a bivariate linear model[1]

Distribution	ARQ(0.05)	ARQ(0.10)	TRQ(0.10)	TRQ(0.25)	TRQ(0.5)	l_2
	Estimator					
Normal	0.91	0.89	0.93	0.82	0.62	1.00
Cauchy	0.72	0.77	0.45	0.77	0.79	0.00
Uniform	0.16	0.14	0.15	0.10	0.07	0.19
Laplace	1.00	1.00	0.89	1.03	1.00	0.67
Exponential	0.19	0.16	0.07	0.07	0.05	0.05
Lognormal	0.27	0.23	0.08	0.10	0.09	0.03
Bimodal	0.47	0.44	0.13	0.07	0.02	0.11

[1]Reported entries are efficiencies relative to the maximum likelihood estimator for each error distribution, e.g., $\mathrm{mse}(\hat{\beta}_{\mathrm{MLE}})/\mathrm{mse}(\hat{\beta}_{\mathrm{ARQ}})$. In each case the efficiencies are based on 10,000 replications of the bivariate linear model with 100 observations.

As the theory predicts, the adaptive L-estimators offer good performance over the entire range of distributions investigated. To our delight, they are particularly successful in the asymmetric and bimodal cases. But they offer high efficiency in the more familiar symmetric unimodal cases as well. Finally, we must emphasize that these results are based solely on the bivariate model and we did very little experimentation with the smoothing methods employed to estimate the J functions. In future work we hope to report more extensive experimental results.

APPENDIX

The uniform Bahadur representation for regression quantiles with explicit bounds. Basically, the proof of Theorem 2.1 of Koenker and Portnoy (1987) will be followed exactly with bounds expressed explicitly as functions of the distribution and interval $(\alpha, 1-\alpha)$. However, this requires the result of Lemma 2.1 of Portnoy (1984) showing that $\|\hat{\gamma}\|\, O_p(\log n/n)^{1/2}$. To obtain explicit bounds, condition (2.10) of Portnoy (1984) must be replaced by Condition (X4) as described in Proposition 3.2 of Portnoy (1984) (with some modification of the argument). The conditions required here are:

CONDITION F1. Conditions (X1)–(X4) hold and the density f is continuous, bounded and strictly positive.

CONDITION F2. In addition to (F1), the derivative f' exists and is uniformly bounded.

Note that for $F \in \mathbf{F}$, F_s satisfies Conditions (F1) and (F2) and, hence, Theorem A.1 holds for $\hat{F}_n$ and $\hat{\gamma}$ defined in Section 1 (based on observations $\tilde{Y}_i$) and F_s given by (2.3) for any $F \in \mathbf{F}$. Following the proofs of Lemma 2.1 and

Proposition 2.2 in Portnoy (1984) and keeping careful track of explicit bounds yields the following results:

LEMMA A.1. *Assume Condition* (F.1). *Then there exist n_0 and constants $b_i(X)$ depending only on the constants in Conditions* (X1)–(X4) *such that for $n \geq n_0$,*

$$\|\hat{\gamma}\| \leq K(X, f)R(\log n/n)^{1/2}, \tag{A.1}$$

where

$$\begin{aligned} P\{|R| \geq w\} &\leq \exp -b_1(X)(w-1)^2 \log n \quad \text{for } w \geq 2, \\ K(X, f) &= b_2(X)/\left\{\inf_{\alpha,\, b_3(X)} f(t)\right\}. \end{aligned} \tag{A.2}$$

Here, we define

$$\inf_{\alpha, b} f(t) = \inf\left\{ f(t) \colon F^{-1}(\alpha) - b \leq t \leq F^{-1}(1-\alpha) + b \right\}.$$

The results of Koenker and Portnoy (1987) can also be extended by providing firm bounds in terms of the density, f, and the constants in (X1)–(X4). Again with $b_i(x)$ denoting constants [depending only on (X1)–(X4)], careful consideration of the proofs in Koenker and Portnoy (1987) yields the following results.

LEMMA A.2. *Assume Condition* (F2) *and define for $\delta \in R^p$ and $0 \leq \theta \leq 1$,*

$$\begin{aligned} T(\delta, \theta) &= \sum_{i=1}^{n} x_i \left\{ I\left(u_i \leq F^{-1}(\theta) + x_i' \delta\right) - I\left(u_i \leq F^{-1}(\theta)\right) \right\}, \\ \tilde{T}(\delta, \theta) &= T(\delta, \theta) - ET(\delta, \theta). \end{aligned} \tag{A.3}$$

Then, for $\delta \in \Delta \equiv \{\delta \colon \|\delta\| \leq K(\log n/n)^{1/2}\}$ and $\alpha \leq \theta \leq 1 - \alpha$,

$$\begin{aligned} &|ET(\delta, \theta) - nQ\delta f\left(F^{-1}(\theta)\right)| \\ &\qquad \leq K(K+1) b_1(X) \left\{ \sup_x f(x) + \sup_x |f'(x)| \right\} n^{1/4} (\log n)^{1/2} \end{aligned}$$

and

$$\begin{aligned} P&\left\{ \sup_{\delta \in \Delta,\, \alpha \leq \theta \leq 1-\alpha} \|\tilde{T}(\delta, \theta)\| \geq \left(n^{1/4} \log n\right) K^2 b_2(X) \left\{ \sup_x f(x) + \sup_x |f'(x)| \right\} \right\} \\ &\leq K \exp\left\{ b_3(X) \sup_x f(x) - (\log n) \right\} \\ &\quad + 1/\sqrt{n}\, \left\{ 2 \sup_x f(x) / \inf_\alpha f(t) + b_4(X) \right\}. \end{aligned}$$

Combining Lemmas A.1 and A.2 yields

LEMMA A.3. *Under* (F2),

$$P\left\{ \sup_{\alpha \leq \theta \leq 1-\alpha} \|T((0, \hat{\gamma}), \theta)\| \geq n^{1/4} (\log n) B(X, F) \right\} \leq q(X, F),$$

where

$$B(X, F) = \frac{b_1(X)\{\sup_x f(x) + \sup_x |f'(x)|\}}{\{\inf_{\alpha,\, b_2(X)} f(t)\}^2},$$

$$q(X, F) = \frac{b_3(X)}{\sqrt{n}} \frac{\exp\{b_4(X)\sup f(X)\}}{\{\inf_{\alpha,\, b_5(X)} f(t)\}}.$$

Lastly, as a consequence we have

THEOREM A.1. *Under Condition* (F2), *using* $B(X, F)$ *and* $q(X, F)$ *defined above*,

$$\sup_{\alpha \le \theta \le 1-\alpha} \left| \hat{F}_n(F^{-1}(\theta)) - 1/n \sum_{i=1}^{n} I(u_i \le F^{-1}(\theta)) \right| \le n^{-3/4}(\log n) B(X, F)$$

and

$$\sup_{\alpha \le \theta \le 1-\alpha} \left\| \dot{Q}(\hat{\gamma}(\theta) - \gamma) f(F^{-1}(\theta)) - 1/n \sum_{i=1}^{n} \dot{x}_i(\theta - I(u_i \le F^{-1}(\theta))) \right\|$$
$$\le n^{-3/4}(\log n) B(X, F) + b_1(X)/n,$$

except on a set with probability bounded above by $q(X, F)$.

REFERENCES

BASSETT, G. and KOENKER, R. (1982). An empirical quantile function for the iid linear model. *J. Amer. Statist. Assoc.* **77** 407–415.

BERAN, R. (1974). Asymptotically efficient adaptive rank estimates in location models. *Ann. Statist.* **2** 63–74.

BICKEL, P. J. (1982). On adaptive estimation. *Ann. Statist.* **10** 647–671.

BREIMAN, L. (1968). *Probability*. Addison-Wesley, Reading, Mass.

COX, D. (1985). A penalty method for nonparametric estimation of the logarithmic derivative of a density function. *Ann. Inst. Statist. Math.* **37** 271–288.

DE JONGH, P. and DE WET, T. (1986). Adaptive trimmed least squares. Unpublished.

FOX, P. A. (1984). *The Port Mathematical Subroutine Library*. AT & T Bell Laboratories, Murray Hill, N.J.

GAY, D. M. (1983). Algorithm 611. Subroutines for unconstrained minimization using a model/trust-region approach. *ACM Trans. Math. Software* **9** 503–524.

HOGG, R. (1981). On adaptive robust inferences. In *Statistics and Related Topics*. (M. Csörgő, D. A. Dawson, J. N. K. Rao and A. K. Md. Ehsanes Saleh, eds.) 253–264. North-Holland, Amsterdam.

JUREČKOVÁ, J. and SEN, P. K. (1984). On adaptive scale-equivalent M-estimators in linear models. *Statist. Decisions* **1** 31–46.

KOENKER, R. and BASSETT, G. (1978). Regression quantiles. *Econometrica* **46** 33–50.

KOENKER, R. and D'OREY, V. (1987). Computing regression quantiles. *Appl. Statist.* **36** 383–393.

KOENKER, R. and PORTNOY, S. (1987). L-estimators for linear models. *J. Amer. Statist. Assoc.* **82** 851–857.

KOUL, H. L. (1969). Asymptotic behavior of Wilcoxon type confidence regions in multiple linear regression. *Ann. Math. Statist.* **40** 1950–1979.

MANSKI, C. F. (1984). Adaptive estimation of non-linear regression models. *Econom. Rev.* **3** 145–194.

MANSKI, C. F. and HSIEH, D. A. (1987). Monte Carlo evidence on adaptive maximum likelihood estimation of a regression. *Ann. Statist.* **15** 541–551.

NEWEY, W. (1987). Adaptive estimation of regression models via moment restrictions. Princeton Research Memorandum 330, Princeton Univ.

PORTNOY, S. (1984). Tightness of the sequence of empiric c.d.f. processes defined from regression fractiles. *Robust and Nonlinear Time Series Analysis. Lecture Notes in Statist.* **26** 231–246. Springer, New York.

SACKS, J. (1975). An asymptotically efficient sequence of estimators of a location parameter. *Ann. Statist.* **3** 285–298.

SILVERMAN, B. (1986). *Density Estimation for Statistics and Data Analysis*. Chapman and Hall, New York.

STEIN, C. (1956). Efficient nonparametric testing and estimation. *Proc. Third Berkeley Symp. Math. Statist. Probab.* **1** 187–196. Univ. California Press.

STONE, C. (1975). Adaptive maximum likelihood estimators of a location parameter. *Ann. Statist.* **3** 267–284.

DEPARTMENT OF STATISTICS
UNIVERSITY OF ILLINOIS
CHAMPAIGN, ILLINOIS 61820

DEPARTMENT OF ECONOMICS
UNIVERSITY OF ILLINOIS
CHAMPAIGN, ILLINOIS 61820