# BEBR

FACULTY WORKING
PAPER NO. 1263

# L-Estimation for Linear Models

*Roger Koenker*
*Stephen Portnoy*



College of Commerce and Business Administration
Bureau of Economic and Business Research
University of Illinois, Urbana-Champaign

# BEBR

FACULTY WORKING PAPER NO. 1263

College of Commerce and Business Administration

University of Illinois at Urbana-Champaign

June 1986

L-Estimation for Linear Models

Roger Koenker, Professor
Department of Economics

Stephen Portnoy, Professor
Department of Statistics

# L-Estimation for Linear Models

Roger Koenker and Stephen Portnoy

## Abstract

Analogues of linear-combinations-of-order-statistics, or L-estimators, are suggested for estimating the parameters of the linear regression model. The methods are based on linear combinations of the p-dimensional "regression quantiles" proposed by Koenker and Bassett. A uniform Bahadur-type representation of regression quantiles is established, and this permits a general theory of L-estimators based on regression quantiles including those with smooth weight functions. A leading example of the proposed class of estimators is an analogue of the trimmed mean which seems to exhibit certain advantages over earlier proposals by Koenker and Bassett and Ruppert and Carroll. A brief investigation of two proposals for estimating the covariance matrix of this estimator is also reported.

Roger Koenker is Professor of Economics, and Stephen Portnoy is Professor of Statistics, University of Illinois at Urbana- Champaign, Champaign, IL, 61820. This research was supported in part by NSF grants SES-8408567 and MCS-8301834.

KEYWORDS: Regression, quantiles, L-estimation, robustness, Bahadur representation.

## 1. Introduction

Analogues of a broad class of L-estimators for the parameters of the linear regression model are proposed and investigated. The methods are based on the "regression quantile" statistics of Koenker and Bassett (1978).

Consider the linear model

$$y_i = x_i\beta + u_i \qquad i = 1, ..., n \tag{1.1}$$

where $x_i = (1, x_{i2}, ..., x_{ip})$ denotes the $i^{th}$ row of an $n \times p$ design matrix, and $\beta \in \mathbf{R}^p$ is an unknown regression parameter. We will assume throughout that $(u_1, ..., u_n)$ are independent with common distribution function F. Explicit further assumptions on the design and F will be introduced below.

The p-dimensional analogues of the sample quantiles, introduced in KB (1978) solve the problem

$$\min_{b \in \mathbf{R}^p} \sum_{i=1}^{n} \rho_\theta (y_i - x_i b) \tag{1.2}$$

where $\rho_\theta(u)$ denotes the "check" function $\rho_\theta(u) = \theta u^+ + (1-\theta)u^-$ and $u^+$, $u^-$ denote respectively the positive and negative parts of $u$. The set of such solutions will be denoted by $\hat{B}_\theta$. Note that in the location model, i.e., when $x_i \equiv 1$, $\hat{B}_\theta$ is simply the usual $\theta^{th}$ sample quantiles from the (now i.i.d.) sample $(y_1, \cdots, y_n)$ from $F(y-\beta)$. The $l_1$ regression problem, (1.2) with $\theta = 1/2$, is also a familiar special case.

Problem (1.2) may be formulated as a linear program and it is easily shown that $\hat{B}_\theta$ is the convex hull of one or more "basic" solutions of the form $b_h = X_h^{-1} y_h$, where $h$ indexes p-element subsets of $\{1, 2, ..., n\}$ and $X_h$ denotes the sub-design matrix with rows $x_i$: $i \in h$, and $y_h$ is the sub-response vector with coordinates $y_i$: $i \in h$. Thus the "regression quantiles" may be viewed as *order-statistics corresponding to groups of p-observations.* And problem (1.2) serves to identify a small number of "interesting" basic solutions, roughly O(n) in our empirical experience, out of the $\binom{n}{p}$ number of possible basic solutions. Wu (1986) has recently emphasized the fundamental role played by these p-observation subsets in the theory of least-squares estimation.

Computation of regression quantiles is treated in Koenker and d'Orey (1985). There, an algorithm based on Barrodale and Roberts (1974) $l_1$-regression algorithm is provided to efficiently compute solutions to problem (1.2) for all $\theta \in [0,1]$. This may at first appear onerous, but fortunately it is a straightforward exercise in parametric linear programming, or sensitivity analysis. Once one solution has been identified the remaining, O(n), solutions may be found easily and each involves essentially one simplex pivot.

An asymptotic theory of finite linear combinations of regression quantiles was developed in KB (1978), and led to simple analogues of the "systematic statistics" of Mosteller (1946), Tukey (1970), Gastwirth (1966) and others. Ruppert and Carroll (1980) showed that a simple analogue of the trimmed mean could be constructed as,

$$\bar{\beta}_\alpha = (X'WX)^{-1}X'Wy \tag{1.3}$$

where $W$ is a diagonal matrix with typical element $w_i = I(x_i\hat{\beta}_\alpha < y_i < x_i\hat{\beta}_{1-\alpha})$ where $\hat{\beta}_\theta$ denotes some selection from $\hat{B}_\theta$. This estimator trims observations on-or-below the $\alpha^{th}$ and on-or-above the $1-\alpha^{th}$ regression quantile plane, and computes a least squares estimate based on the remaining observations. Ruppert and Carroll established, under mild conditions, that $\sqrt{n}(\bar{\beta}_\alpha - \beta)$ was asymptotically Gaussian with covariance matrix $\sigma^2(\alpha,F)Q^{-1}$ where $Q = \lim n^{-1}X'X$ and $\sigma^2(\alpha,F)$ is the asymptotic variance of the alpha-trimmed mean from a random sample on F.

In simulation experiments, reported briefly in Koenker (1986), it was found that this trimmed least-squares estimator was rather sensitive to influential design points, and exhibited substantial departures from the behavior predicted by its asymptotic theory, especially when $p$ was large relative to $n$. This finding motivated the present investigation into a considerably broader class of L-estimators based on regression quantiles.

Following Serfling (1980), it is natural to consider estimators of the form,

$$\hat{\beta} = \int_0^1 J(\theta)\hat{\beta}(\theta)d\theta + \sum_{i=1}^{n}\omega_i\hat{\beta}(\theta_i) \tag{1.4}$$

where, as above, if necessary, we have adopted a rule for choosing an element $\hat{\beta}(\theta)$ from $\hat{B}_\theta$. Estimators of this general form are scale and reparameterization-of-design equivariant, see KB (1978, Thm. 3.2). This is an important advantage of L-statistics over competing M-estimates. Bickel (1973) proposed analogues of L-estimators for the linear model based on a preliminary estimate, but they are computationally complex and are not equivariant to reparameterization of the design. Recently, Welsh (1985, 1986) has proposed a class of one-step L-estimators which *are* equivariant and reasonably easy to compute.

We will focus here on the first term of (1.4) with J chosen to be reasonably smooth. A leading example of the type we wish to consider is the analogue of the trimmed mean,

$$\tilde{\beta}[J_\alpha] = (1 - 2\alpha)^{-1}\int_\alpha^{1-\alpha} \hat{\beta}(\theta)\, d\theta$$

In the simulations reported in Koenker (1986) this estimator performed extremely well, showing considerably less sensitivity to influential design points than the (asymptotically equivalent) trimmed least squares estimator.

In the next section we establish a uniform $O(n^{-1/4}\log n)$ Bahadur-type representation for the regression quantile process approximating $\sqrt{n}\,(\hat{\beta}(\theta) - \beta(\theta))$ as $1/\sqrt{n}$ times a sum of independent random variables with error negligible to $O(n^{-1/4}\log n)$ *uniformly* in $\theta$. Applications of this result to the asymptotic theory of L-statistics like (1.5) are treated in Section 3, where we also discuss the problem of estimating the covariance matrix of such estimators.

## 2. A Uniform Bahadur Representation for Regression Quantiles

We will assume throughout this section that $\beta = 0$ and $n^{-1}\sum x_i = (1, 0, ..., 0)$; this involves no loss of generality due to equivariance considerations.

The following design conditions are employed:

X1: $\frac{1}{n}X'X = Q + Q_n$ where $Q$ is positive definite and the maximum eigenvalue of $Q_n$ satisfies $\lambda_{\max}(Q_n) = O(n^{-1/4})$

X2: $\sum_{i=1}^{n} \|x_i\|^3 = O(n)$

X3: $\max_i \|x_i\| = O(n^{1/4})$

X4: condition 2.10 of Portnoy (1985):

Partition $\beta = (\alpha, \gamma)$ so $\gamma \in \mathbf{R}^{p-1}$. Then for any constant $a$ (sufficiently large) there is $\eta > 0$ such that for all $\alpha \in [-a,a]$ and all $\gamma \in \mathbf{R}^{p-1}$, $\gamma' M(\alpha,\gamma) \geq \eta\, s(\gamma)$ where $M(\beta) = \sum_{i=1}^{n} x_i F(x_i\beta)$ and $s(\gamma) = \min\{\|\gamma\|^2, \|\gamma\|\}$.

It is not difficult to see that these conditions will hold in typical ANOVA designs, and will hold in probability when the rows of the design $\{x_i : i = 1, 2, \cdots\}$ form a random sample from a very wide class of distributions in $\mathbf{R}^p$. Results along these lines are given in Portnoy (1985), in particular for condition X4.

Our condition on F is the following:

F: F has a density, f, and for some $\epsilon > 0$: $\phi(u) = f(F^{-1}(u)) > 0$ and $\phi'(u)$ is uniformly bounded for $u \in [\epsilon, 1-\epsilon]$.

*Lemma 2.1* Under conditions X1-4, and F, for any $\epsilon > 0$ there is a $K > 0$ such that

$$\sup_{\epsilon < \theta < 1-\epsilon} \|\hat{\beta}(\theta) - \beta(\theta)\| \leq K(\log n / n)^{1/2} \tag{2.1}$$

with probability tending to one.

*Proof.* Following Portnoy (1985) partition $\beta = (\alpha,\gamma)$, where $\gamma \in \mathbf{R}^{p-1}$ and $\hat{\beta}(\theta) = (\hat{\alpha}(\theta), \hat{\gamma}(\theta))$. Let

$$\hat{\theta}(\alpha) = \sup \{\theta \in [0,1]: \hat{\alpha}(\theta) \leq \alpha\} \tag{2.2}$$

See Bassett and Koenker (1982) for further details on this estimate of F. From Lemma 2.1 of Portnoy (1985),

$$\sup_{\epsilon \leq \theta \leq 1-\epsilon} \|\hat{\gamma}(\theta)\| = O_p(\log n / n)^{1/2} \tag{2.3}$$

and, from Proposition 2.1 there, for some $c > 0$,

$$|\hat{\theta}(\alpha) - F(\alpha)| \leq c/\sqrt{n} \tag{2.4}$$

uniformly in $|\alpha| \leq b = \max \{|F^{-1}(\epsilon)|, |F^{-1}(1-\epsilon)|\}$ with probability tending to one. Since $f(x) > 0$ and is continuous, there is a $d > 0$ such that for $\theta \in [\epsilon, 1-\epsilon]$, with probability tending to one

$$\begin{aligned} \hat{\theta}(\alpha + d/\sqrt{n}) &\geq F(\alpha + d/\sqrt{n}) - c/\sqrt{n} \\ &\geq F(\alpha) + Kd/\sqrt{n} - c/\sqrt{n} \\ &\geq F(\alpha) \end{aligned} \tag{2.5}$$

and similarly

$$\hat{\theta}(\alpha - d/\sqrt{n}) \leq F(\alpha) \tag{2.6}$$

where $K = \inf\{ f(u): |u| \leq b + d/\sqrt{n} \}$. Thus from the definition of $\hat{\theta}$, we have for $\theta = F(\alpha)$

$$|\hat{\alpha} - F^{-1}(\theta)| \leq d/\sqrt{n} \tag{2.7}$$

and the lemma follows from (2.3) and (2.7). □

The main result of this section is the following uniform Bahadur representation for $\sqrt{n}\,(\hat{\beta}(\theta) - \beta(\theta))$, extending a result of Jureckovà and Sen (1984).

*Theorem 2.1* Under X1-4 and F, with probability tending to one, for any $\epsilon>0$,

$$\sqrt{n}\ (\hat{\beta}(\theta) - \beta(\theta)) = \frac{1}{\sqrt{n} f(F^{-1}(\theta))} Q^{-1} \sum_{i=1}^{n} x_i[\theta - I(u_i \le F^{-1}(\theta))] + O(n^{-1/4} \log n) \tag{2.8}$$

*uniformly* for $\theta \in [\epsilon, 1-\epsilon]$.

*Proof.* From KB (1978), $\hat{\beta}(\theta) = b_h = X_h^{-1} y_h$ if and only if for j=1, ..., p.

$$\sum_{i=1}^{n} [I(y_i < x_i\hat{\beta}) - \theta I(y_i \ne x_i\hat{\beta})]\, x_{ij} \in [\theta-1, \theta] \tag{2.9}$$

Thus there is a vector $v \in \mathbf{R}^p$ with $\max_j |v_j| \le 1$, for which $\hat{\beta}(\theta) = b_h$ and

$$\| \sum_{i=1}^{n}[I(y_i \le x_i\hat{\beta}) - \theta]x_i - (1-\theta) \sum_{i\in h} x_i \| = O(\|X_h v\|) \tag{2.10}$$

Since $(1-\theta)\sum_{i\in h} \le p \max \|x_i\| = O(n^{1/4})$ and $\|X_h v\| \le [\mathrm{tr}\,(X_h' X_h)]^{1/2} = O(n^{1/4})$ by X3, and $y_i = x_i\beta(\theta) - F^{-1}(\theta) + u_i$ we have

$$\|\sum[I(u_i \le F^{-1}(\theta) + x_i(\hat{\beta} - \beta)) - \theta]x_i\| = O(n^{1/4}), \tag{2.11}$$

with probability tending to one, uniformly in $\theta \in [\epsilon, 1-\epsilon]$. Let

$$g(\delta,\theta) = \sum[I(u_i \le F^{-1}(\theta) + x_i\delta) - \theta]x_i \tag{2.12}$$

and set

$$T(\delta,\theta) = g(\delta,\theta) - g(0,\theta) \tag{2.13}$$

and

$$\tilde{T}(\delta,\theta) = T(\delta,\theta) - ET(\delta,\theta) \tag{2.14}$$

Now, for $\delta \in \{\ \delta \in \mathbf{R}^p \mid \|\delta\| \le K\sqrt{\log n/n}\ \}$

$$\begin{aligned} Eg(\delta,\theta) &= \sum x_i(x_i\delta) f(F^{-1}(\theta)) + \sum x_i(x_i\delta)^2 f'(F^{-1}(\theta^*)) \\ &= n(Q + Q_n)\delta f(F^{-1}(\theta)) + \sum\|x_i\|^3 O(\log n/n) \\ &= nQ\delta f(F^{-1}(\theta)) + O(\log n) \end{aligned} \tag{2.15}$$

by X2 and condition F. And the result follows by Lemma 2.1 and Lemma 2.2 (below) □ .

*Proposition.* For any $\lambda > 0$, $K$ fixed as in Lemma 2.1, and $\delta \in \Delta = \{ \delta \in \mathbf{R}^p \mid \|\delta\| \le K\sqrt{\log n / n} \}$,

$$P\{|\tilde{T}_j| \ge \lambda n^{1/4}\log n\} \le 2\exp\{-\lambda \log n(1 + o(1))\} \tag{2.16}$$

*Proof.* By the Markov inequality, for $t > 0$, and any $\lambda_n > 0$

$$P\{|\tilde{T}_j| \ge \lambda_n\} \le e^{-t\lambda_n}[M_j(t) + M_j(-t)] \tag{2.17}$$

where $M_j(t)$ is the mgf of $\tilde{T}_j$. By independence of the $u$'s, $M_j(t) = \Pi\ M_{ij}(t)$ where

$$M_{ij}(t) = E\ \exp\{tx_{ij}[J_i(\delta,\theta) - EJ_i(\delta,\theta)]\} \tag{2.18}$$

and

$$J_i(\delta,\theta) = I(u_i \le F^{-1} + x_i\delta) - I(u_i \le F^{-1}(\theta)) \tag{2.19}$$

Note that $EJ_i = \text{sgn}(x_i\delta)p_i$ for $p_i = P\{u_i$ between $F^{-1}(\theta)$ and $F^{-1}(\theta) + x_i\delta\}$ thus,

$$M_{ij}(t) = p_i\ \exp\{tx_{ij}(1 - p_i)\ \text{sgn}(x_i\delta)\} + (1 - p_i)\exp\{-t\ x_{ij}\ p_i\ \text{sgn}(x_i\delta)\} \tag{2.20}$$

If $t = O(n^{-1/4})$, $|x_{ij}t|$ is bounded by X3, and since for $\delta \in \Delta$,

$$p_i = |x_i\delta| f(F^{-1}(\theta^*)) \le c_0|x_i\delta| \to 0 \tag{2.21}$$

since $f$ is bounded. Thus,

$$\begin{aligned} \log M_{ij}(t) &\le \log(1 + 2p_i x_{ij}^2 t^2 e^{|x_{ij}t|}) \\ &\le 2p_i(x_{ij}t)^2 e^{|x_{ij}t|} \\ &\le c\ |x_i\delta|(x_{ij}t)^2\exp\{Btn^{1/4}\} \end{aligned} \tag{2.22}$$

for some constant c, by (2.21). Therefore by condition X2, for $t > 0$, and $t = O(n^{-1/4})$,

$$\begin{aligned} \log M_j(t) &\le \sum_{i=1}^{n} c'\ \|\delta\|\ \|x_i\|^3\ t^2\exp\{Btn^{1/4}\} \\ &\le c''\sqrt{n\ \log n}\ \ t^2\exp\{Btn^{1/4}\} \end{aligned} \tag{2.23}$$

Finally, by (2.17), with $t = n^{-1/4}$

$$P\{|T_j| \geq \lambda\, n^{1/4} \log n\} \leq 2 \exp\{-\lambda \log n + c''\sqrt{\log n}\, e^{B}\} \\ = 2 \exp\{-\lambda \log n(1 + o(1))\} \tag{2.24}$$

*Lemma 2.2* Under X1-4 and condition F, with probability tending to one,

$$\sup_{\theta \in [\epsilon, 1-\epsilon]} \|\tilde{T}(\hat{\beta}(\theta) - \beta(\theta), \theta)\| = O(n^{1/4} \log n) \tag{2.25}$$

*Proof.* First let $\theta_i = \epsilon + i/n^3$, $i = 0, 1, 2, \ldots, [(1-2\epsilon)\, n^3]$ and let $\delta_j \in \Delta$ be the centers of spheres of radius $n^{-3}$ covering $\Delta$. Let

$$\mathbf{B} = \{(\theta, \delta) \mid \theta = \theta_i,\ \delta = \delta_j \text{ for some } i \text{ and } j\} \tag{2.26}$$

Then $\# \mathbf{B} \leq a n^3 n^{3p}$, and, hence, from the proposition

$$P\{\sup_{\mathbf{B}} |\tilde{T}_j(\theta, \delta)| \geq (3p + 5) n^{1/4} \log n\} \leq a n^{3p+3}\, e^{-(3p+4)\log n} \to 0 \tag{2.27}$$

Consequently, as $n \to \infty$

$$P\{\sup_{\mathbf{B}} \|\tilde{T}(\theta, \delta)\| \geq p(3p + 5)\, n^{1/4} \log n\} \to 0 \tag{2.28}$$

Now for $\{\delta_1, \delta_2\} \subset \Delta$, with $\|\delta_1 - \delta_2\| \leq n^{-3}$ and $\{\theta_1, \theta_2\} \subset [\epsilon, 1-\epsilon]$ with $|\theta_1 - \theta_2| \leq n^{-3}$ consider

$$\|T(\theta_1, \delta_1) - T(\theta_2, \delta_2)\| = \|\sum x_i\, [I(u_i \leq F^{-1}(\theta_1) + x_i\delta_1) - I(u_i \leq F^{-1}(\theta_2) + x_i\delta_2)]\| \\ \leq \sum \|x_i\|\, [I(u_i \leq F^{-1}(\theta_1) + x_i\delta_1) - I(u_i \leq F^{-1}(\theta_2) + x_i\delta_2)] \tag{2.29}$$

Note that since $f$ is continuous and strictly positive $|F^{-1}(\theta_1) - F^{-1}(\theta_2)| < c_1 n^{-3}$ and hence,

$$|(F^{-1}(\theta) + x_i\delta_1) - (F^{-1}(\theta) + x_i\delta_2)| \leq c_1/n^3 + c_2 n^{1/4} n^{-3} \\ \leq c_3 n^{-2.5} \tag{2.30}$$

Now for $i \neq j$

$$P\{|u_i - u_j| \leq c_3 n^{-2.5}\} = P\{u_i \in [u_j - c_3 n^{-2.5}, u_j + c_3 n^{-2.5}]\} \\ \leq c_4 n^{-2.5} \tag{2.31}$$

since $f$ is bounded. Hence

$$P\{\min_{i \neq j} |u_i - u_j| \leq c_3 n^{-2.5}\} \leq n(n-1) c_3 n^{-2.5} \to 0 \tag{2.32}$$

It follows that with probability tending to one, the term in square brackets in (2.29) is nonzero for at most two values of $i$; and, hence,

$$\|T(\theta_1, \delta_1) - T(\theta_2, \delta_2)\| \le 2 \max \|x_i\| = O(n^{1/4}). \tag{2.33}$$

Also, from (2.29) and (2.21)

$$\begin{aligned} \|ET(\theta_1, \delta_1) - ET(\theta_2, \delta_2)\| &\le \sum \|x_i\| \, |p_i(\theta_1, u_1) - p_i(\theta_2, u_2)| \\ &\le \sum \|x_i\| \, |x_i\delta_i| \, |f(F^{-1}(\bar{\theta}_1) - f(F^{-1}(\bar{\theta}_2)| \\ &+ \sum \|x_i\| \, f(F^{-1}(\bar{\theta}_2)) \Big| |x_i\delta_1| - |x_i\delta_2| \Big| \end{aligned} \tag{2.34}$$

Since $f'(x)$ is bounded the first term above is $O(n(\log n)/n^{1/4})n^{-3} = O(1)$. And since

$$\Big| |x_i\delta_1| - |x_i\delta_2| \Big| \le |x_i(\delta_1 - \delta_2)| \le n^{1/4} n^{-3} \tag{2.35}$$

the second term is also bounded. Hence the left side of (2.34) is bounded, and thus, $\|\tilde{T}(\theta_1, \delta_1) - \tilde{T}(\theta_2, \delta_2)\| = O(n^{1/4})$ uniformly for $\{\theta_1, \theta_2\} \subset [\epsilon, 1-\epsilon]$ with $|\theta_1 - \theta_2| \le n^{-3}$ and $\{\delta_1, \delta_2\} \subset \Delta$ with $\|\delta_1 - \delta_2\| \le n^{-3}$ (with probability tending to one). Thus, using (2.27)

$$P\{\sup\{\|\tilde{T}(\theta,\delta)\|: \theta \in [\epsilon, 1-\epsilon], \delta \in \Delta\} \ge b \; n^{1/4} \log n\} \to 0 \tag{2.36}$$

Therefore, by Lemma 2.1, (2.25) follows □ . 3. L-estimators for the Linear Model

Smooth L-estimators for regression may be expressed as

$$\tilde{\beta} = \int_0^1 J(\theta)\hat{\beta}(\theta)d\theta \tag{3.1}$$

and the results of the previous section immediately yield

*Theorem 3.1* Under conditions of Section 2, let $J(\theta)$ denote a bounded, measurable function on [0,1], and suppose there is an $\epsilon$, satisfying condition $F$, such that $J(\,)$ vanishes outside $[\epsilon, 1-\epsilon]$ then

$$L(\sqrt{n}\,(\tilde{\beta} - \beta(J, F))) \to N(0, \sigma^2(J, F)Q^{-1}) \tag{3.2}$$

where $\beta(J, F) = \int_0^1 \beta(\theta)J(\theta)d\theta$ and

$$\sigma^2(J,F) = \int_0^1\int_0^1 (s \wedge t - ts)[f(F^{-1}(t))f(F^{-1}(s))]^{-1}J(t)J(s)\,ds\,dt \tag{3.3}$$

*Proof.* Theorem 2.1 implies that

$$\sqrt{n}\,(\tilde{\beta} - \beta(J,F)) = \frac{1}{\sqrt{n}}Q^{-1}\sum_{i=1}^{n} x_i\omega_i + O_p(n^{-1/4}\log n) \tag{3.4}$$

where

$$\omega_i = \int_0^1 J(\theta)[f(F^{-1}(\theta))]^{-1}[I(u_i \le F^{-1}(\theta) - \theta]\,d\theta \tag{3.5}$$

[or $\omega_i = \int_{-\infty}^{\infty} J(F(v))[I(u_i \le v) - F(v)]\,dv$]. The $\omega_i$ are iid random variables with mean zero and variance $\sigma^2(J,F)$. Conditions X1 and X3 and the Lindeberg-Feller CLT immediately yield 3.2. □

*Remark.* An intriguing special case, *not* covered by this result is the "untrimmed mean,"

$$\tilde{\beta}_0 = \int_0^1 \hat{\beta}(\theta)\,d\theta \tag{3.7}$$

Under further conditions on the tail behavior of $F$, it is natural to conjecture that $\tilde{\beta}_0$ would have the same limiting behavior as the least squares estimator. The least squares estimator may be written as

$$\hat{\beta} = \sum \omega_h b_h \tag{3.8}$$

where $b_h = X_h^{-1}y_h$ as in Section 1 and $\omega_h = |X_h|^2/\sum|X_h|^2$, and the sums are over *all* $\binom{n}{p}$ possible $h$'s. (See Wu (1986) for further detail on this result.) Thus while *every* subset of $p$ observations gets positive weight in (3.8), the asymptotically equivalent form (3.7) places positive weight on the much smaller subset of $b_h$'s which solve problem (1.2). Thus it may be advantageous to resample from $\hat{\beta}(\theta)$ along the lines recently discussed by Wu (1986) to implement bootstrap methods for regression.

Natural estimates of the asymptotic covariance matrix $\sigma^2(J, F)$ may be constructed in several ways. One approach is to substitute the empirical distribution of the residuals in the expression (3.3) or the equivalent form,

$$\sigma^2(J, F) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} [F(x \wedge y) - F(x)F(y)]J(F(x))J(F(y))\, dx\, dy. \tag{3.9}$$

Welsh (1986) derives a convenient form of this general expression by integrating by parts. An alternative approach to estimating $\sigma^2(J, F)$ is to employ the empirical quantile function

$$\hat{Q}_Y(\theta) = \inf\{\bar{x}\, b \mid b \in \hat{B}_\theta\} \tag{3.10}$$

which arises naturally from problem (1.2). Here $\bar{x}$ is the mean design row, i.e., $n^{-1}\sum x_i$. See Bassett and Koenker (1982, 1986) and Portnoy (1985) for further details on $\hat{Q}_Y(\theta)$. It suffices here to note that under considerably milder conditions than those of Section 2, $Q_Y(\theta)$ is strongly consistent for $Q(\theta) = \bar{x}\beta + F^{-1}(\theta)$, which may be interpreted as the conditional quantile function of the response variable evaluated at the mean design point.

We have investigated both approaches in the important special case of trimming. The asymptotic variance of the trimmed regression quantile estimator given in (1.5) is, when $F$ is symmetric, the Winsorized variance,

$$\sigma^2(\alpha, \xi) = (1 - 2\alpha)^{-2}[\int_{\alpha}^{1-\alpha} \xi^2(\theta)d\theta + \alpha\xi^2(\alpha) + (1 - \alpha)\xi^2(1 - \alpha)] \tag{3.11}$$

where $\xi(\theta) = F^{-1}(\theta)$. The simplest approach to estimating $\sigma^2(\alpha, \xi)$ is simply to replace $\xi$ in (3.11) by the recentered estimate,

$$\hat{\xi}(\theta) = \hat{Q}_Y(\theta) - \bar{x}\tilde{\beta}_\alpha. \tag{3.12}$$

We will denote this estimator as

$$s_0^2(\alpha) = \sigma^2(\alpha, \hat{\xi}). \tag{3.13}$$

De Jongh and de Wet (1986) have investigated several estimators of (3.11) based on residuals from the trimmed least squares estimator. A slight variant of their most successful method is,

$$s_1^2(\alpha) = (1-2\alpha)^{-2}[(n-p)^{-1}\sum r_1^2\, I(\hat{\xi}(\alpha) < r_i < \hat{\xi}(1-\alpha)) + \alpha\hat{\xi}^2(\alpha) + (1-\alpha)\hat{\xi}^2(1-\alpha) \tag{3.14}$$

where $r_i = y_i - x_i\tilde{\beta}_\alpha$, and $\hat{\xi}(\theta)$ is given in 3.12.

To compare the performance of the two estimators we have conducted a small monte-carlo experiment along the lines developed by Gross (1977). Since $\tilde{\beta}_\alpha$ is translation equivariant, and $s_0^2(\alpha), s_1^2(\alpha)$ are scale equivariant, we can exploit Gross's monte-carlo swindle for error distributions from the normal/independent family. Given a design matrix $X$, we draw $y_i = u_i = z_i/v_i$, $i = 1, 2, ..., n$, where the $z_i$ are independent standard normal and the $v_i$ are independent root chi-squared random variables divided by degrees of freedom. Thus the $u_i$ are i.i.d. Student random variables and we may compute the optimal weighted least squares estimate $\hat{\beta} = (X'WX)^{-1}X'Wy$, with $W = \text{diag}(v_i^2)$. Then, as in Gross (1977), for any linear contrast $\tilde{a} = c'\tilde{\beta}$,

$$\text{P}\,(\tilde{a} > ks_i) = 1 - \Phi((ks_i - \tilde{a} + \hat{a})/\sigma_c)$$

and by symmetry considerations,

$$\text{P}\,(\tilde{a} > ks_i) = \Phi((-ks_i - \tilde{a} + \hat{a})/\sigma_c)$$

where $\Phi$ is the standard normal distribution function, $\hat{a} = c'\hat{\beta}$, and $\sigma_c = c'(X'WX)^{-1}c$. We average these two probabilities over a number of replications of the experiment for several values of $k$, yielding estimates $\hat{p}(k_i)$, $i = 1, ..., k$. Logit $(\hat{p})$ is then regressed on $k$ and we interpolate in logit $(p)$ to find $k^*$ such that $\hat{p}(k^*) \simeq .025$.

Expected confidence interval lengths (ECILs) may be estimated by averaging $s/((y - X\hat{\beta})'W(y - X\hat{\beta}))^{\frac{1}{2}}$ over monte-carlo replications and finally multiplying by $2k^*$ times the factor

$$\begin{aligned} E\hat{\sigma} &= E((y - X\hat{\beta})'W(y - X\hat{\beta}))^{\frac{1}{2}} \\ &= \sqrt{2}\,\Gamma((n - p + 1)/2)/\Gamma((n - p)/2). \end{aligned}$$

There are 27 different experimental configurations. The factors are

Design: The $X$ matrix is drawn at random once for each configuration and fixed over experimental replications. The first column of $X$ consists of ones, the remaining columns consist of i.i.d. draws from a Student's t distribution with 1, 3, and $\infty$ degrees of freedom. The design matrix $X$ is then orthogonalized for each configuration.

Errors: The error distribution is also chosen to be Student's t with 1, 3, and $\infty$ degrees of freedom.

Sample Size: The sample size is chosen to be 25, 50, and 100.

All other factors are fixed over experimental replications: $p = 3$ parameters are estimated in every case, ten percent trimming is applied, and the linear contrast employed was $c = (\sqrt{3}, \sqrt{3}, \sqrt{3})'$. The experiment was conducted entirely with the 'S' system of Becker and Chambers (1984). 1000 replications were preformed for each configuration. An 'S' macro to compute results for a given configuration is available on request. The random number generator used is the 'S' portable implementation of the Marsaglia uniform generator and thus, recalling the seeds used in the experiment, results should be reproducible on any machine supporting this generator.

In Table 3.1 we report estimated 5% critical values for a two-tailed test on the specified linear contrast. Results are reported for both $s_0$ and $s_1$ and the former yields consistently slightly smaller critical values. The estimated critical values for the $n = 25$ cases are somewhat larger than one would be led to expect from a naive t-table inspection, however, it is only in the extreme case of Cauchy response *and* Cauchy design where the discrepancy is substantial. This point is reinforced by examining the results for larger sample sizes. Standard errors for the elements of Table 3.1 are approximately .01, but particularly for the Cauchy design cases it should be emphasized that the results are conditioned on the initial draw of the design.

In Table 3.2 we report ECIL's for each of the experimental configurations. Consistently the scale estimate, $s_0$, based on the regression quantile function yields slightly shorter intervals than $s_1$, the estimate employing residuals.

In the case of both tables the results compare favorably with those of Gross for the bisquare m-estimator. They suggest that reliable hypothesis testing and confidence interval estimation is possible for the trimmed regression quantile estimator with modest sample sizes. Further investigation is clearly needed to suggest methods for improving on the simple methods studied here. The bootstrapping suggestions of deJongh and deWet(1986) provide a natural alternative approach.

Table 3.1

**ESTIMATED CRITICAL VALUES**

| Response Distribution | | Design Distribution: Normal | Student (3) | Cauchy |
|---|---|---|---|---|
| sample size = 25 | | | | |
| Normal | $s_0$ | 2.21 | 2.22 | 2.33 |
| | $s_1$ | 2.31 | 2.32 | 2.53 |
| Student (3) | $s_0$ | 2.10 | 2.38 | 2.54 |
| | $s_1$ | 2.22 | 2.53 | 2.73 |
| Cauchy | $s_0$ | 1.83 | 2.30 | 2.28 |
| | $s_1$ | 2.00 | 2.60 | 2.48 |
| sample size = 50 | | | | |
| Normal | $s_0$ | 2.09 | 2.09 | 2.14 |
| | $s_1$ | 2.14 | 2.14 | 2.18 |
| Student (3) | $s_0$ | 2.02 | 2.08 | 2.33 |
| | $s_1$ | 2.06 | 2.12 | 2.38 |
| Cauchy | $s_0$ | 1.80 | 1.96 | 3.07 |
| | $s_1$ | 1.88 | 1.96 | 3.07 |
| sample size = 100 | | | | |
| Normal | $s_0$ | 2.01 | 2.02 | 2.06 |
| | $s_1$ | 2.03 | 2.04 | 2.08 |
| Student (3) | $s_0$ | 1.99 | 2.01 | 2.26 |
| | $s_1$ | 2.01 | 2.02 | 2.29 |
| Cauchy | $s_0$ | 1.85 | 1.91 | 2.99 |
| | $s_1$ | 1.88 | 1.95 | 3.02 |

Table 3.2

**EXPECTED CONFIDENCE INTERVAL LENGTHS**

| Response Distribution | | Design Distribution: Normal | Student (3) | Cauchy |
|---|---|---|---|---|
| sample size = 25 | | | | |
| Normal | $s_0$ | 4.36 | 4.36 | 4.36 |
| | $s_1$ | 4.44 | 4.44 | 4.44 |
| Student (3) | $s_0$ | 6.20 | 6.20 | 6.20 |
| | $s_1$ | 6.30 | 6.30 | 6.30 |
| Cauchy | $s_0$ | 17.84 | 17.84 | 17.40 |
| | $s_1$ | 18.10 | 18.10 | 18.10 |
| sample size = 50 | | | | |
| Normal | $s_0$ | 4.19 | 4.19 | 4.24 |
| | $s_1$ | 4.22 | 4.22 | 4.26 |
| Student (3) | $s_0$ | 5.19 | 5.36 | 5.85 |
| | $s_1$ | 5.23 | 5.41 | 5.88 |
| Cauchy | $s_0$ | 10.10 | 10.62 | 15.92 |
| | $s_1$ | 10.24 | 10.73 | 15.80 |
| sample size = 100 | | | | |
| Normal | $s_0$ | 4.08 | 4.11 | 4.16 |
| | $s_1$ | 4.09 | 4.12 | 4.16 |
| Student (3) | $s_0$ | 5.12 | 5.14 | 5.70 |
| | $s_1$ | 5.13 | 5.15 | 5.72 |
| Cauchy | $s_0$ | 9.21 | 9.41 | 14.10 |
| | $s_1$ | 9.28 | 9.47 | 14.08 |

References

Barrodale,I. and Roberts,F.D.K. (1974). Solution of an overdetermined system of equations in the l1 norm. *Communications of the ACM.* 17, 319-320 .

Bassett, G. and Koenker R. (1986). Strong Consistency of Regression Quantiles and Related Processes. *Econometric Theory.* forthcoming.

Bassett, G. and Koenker R. (1982). An empirical quantile function for linear models with iid errors. *Journal of the American Statistical Association.* 77, 407-415 .

Becker, R.A. and Chambers J.M (1984). *S: An Interactive Environment for Data Analysis and Graphics.* Wadsworth. 1

Bickel,P.J. (1973). On some analogues to linear combinations of order statistics in the linear model. *Annals of Statistics.* 1, 597-616 .

deJongh P.J. and deWet T (1986). Confidence Intervals for Regression Parameters based on Trimmed Means. *South African Journal of Statistics.* forthcoming.

Gastwirth, J.L. (1966). On robust procedures. *Journal of the American Statistical Association.* 61, 929-948 .

Gross, A.M. (1977). Confidence intervals for bisquare regression estimates. *Annals of Statistics.* 72, 341-354 .

Jureckova J. and Sen P.K. (1984). On adaptive scale equivariant m-estimators in linear models. *Statistics and Decisions.* 1, 31-46.

Koenker,R.W. and Bassett,G.W. (1978). Regression quantiles. *Econometrica.* 46 , 33-50 .

Koenker, R. and d'Orey V. (1986) Computing Regression Quantiles *Applied Statistics.* forthcoming.

Mosteller,F. (1946). On some useful "inefficient" statistics. *Annals of Mathematical Statistics.* 17 , 377-408 .

Portnoy S. (1985). Tightness of the Sequence of empiric CDF Processes Defined from Regression Quantiles. J. Franke, W. Hardle, and D. Martin(ed.) *Robust and Nonlinear Time Series Analysis.* Springer-Verlag.

Ruppert, D. and Carroll, R.J. (1980). Trimmed least squares estimation in the linear model. *Journal of the American Statistical Association.* 75, 828-838 .

Serfling, R. (1980). *Asymptotic Theory of Statistics.* Wiley. New York .

Tukey J.W. (1970). *Exploratory Data Analysis.* Addison-Wesley.

Welsh A.H. (1985). The Trimmed Mean in the Linear Model. *Annals of Statistics.* forthcoming.

Welsh A.H. (1986). One-Step L-estimators for the Linear Model. University of Chicago Statistics Department Technical Report.

Wu C.F.J. (1986). Jackknife, bootstrap and Other Resampling Methods in Regression analysis. *Annals of Statistics.* forthcoming.