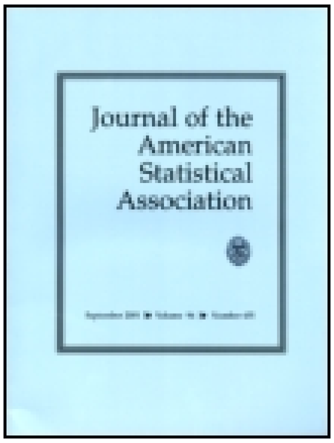

Journal of the American Statistical Association
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/uasa20

An Empirical Quantile Function for Linear Models with iid Errors

Gilbert Bassett Jr. [a] & Roger Koenker [b]
[a] University of Illinois at Chicago Circle , Chicago , IL , 60680 , USA
[b] Bell Laboratories , Murray Hill , NJ , 07974 , USA
Published online: 12 Mar 2012.

To cite this article: Gilbert Bassett Jr. & Roger Koenker (1982) An Empirical Quantile Function for Linear Models with iid Errors, Journal of the American Statistical Association, 77:378, 407-415, DOI: 10.1080/01621459.1982.10477826

To link to this article: http://dx.doi.org/10.1080/01621459.1982.10477826

An Empirical Quantile Function for Linear Models with iid Errors

GILBERT BASSETT, JR. and ROGER KOENKER*

The regression quantile statistics of Koenker and Bassett (1978) are employed to construct an estimate of the error quantile function in linear models with iid errors. Some finite sample properties and the asymptotic behavior of the proposed estimator are derived. Comparisons with procedures based on residuals are made. The stackloss data of Brownlee (1965) is reanalyzed to illustrate the technique.

1. INTRODUCTION

In the simplest location model where we observe a random sample on a variable Y having symmetric distribution $F(y - \mu)$, a valuable approach to robust estimation of μ has been to *adapt* the choice of estimator, $\hat{\mu}$, to the shape of the empirical distribution function. This adaptation may involve a simple pretesting step, or relatively sophisticated density estimation techniques (see, e.g., Hogg 1972; Stone 1975). The success of such methods in the location model raises several questions concerning analogous methods for general linear statistical models. Are there reliable methods of estimating the *shape* of the error distribution for linear models? Can efficient estimates of the error density be constructed for adaptive estimation of linear models? Do analogs of tests of distributional hypotheses based on the empirical distribution function in location models exist for linear models? In this paper we take some tentative steps toward affirmative answers to some of these questions, employing analogs of the sample quantiles for linear models introduced in Koenker and Bassett (1978).

The classical response to the questions raised above is based on the examination of residuals from preliminary estimation of the model, which is usually (under Gaussian assumptions) by least squares. The fundamental articles are Anscombe (1961), Anscombe and Tukey (1963), and Anscombe (1967). See also the recent article by White and MacDonald (1980) and the comment of Weisberg (1980). The appeal of adaptive estimators for linear models based on residuals from a preliminary fit is somewhat diminished by their dependence through the likelihood function of the preliminary estimate on an a priori notion of distributional shape. In contrast, our proposals do not rely on residuals from any preliminary estimate. Instead we suggest a natural generalization to the linear model of a little known one-parameter family of minimization problems that gives rise to the empirical quantile function in the location model.[1] Thus we believe that our approach opens a new and potentially quite fruitful line of inquiry to these and other questions.

The plan of the remainder of the article is as follows. In Section 2 we define and state some fundamental properties of the regression quantile function. Section 3 is devoted to the asymptotic properties of the proposed estimator; consistency and finite-dimensional asymptotic Gaussianity are established. Section 4 compares procedures based on residuals from a preliminary fit of the model. Section 5 is devoted to an example involving the analysis of the well-known stackloss data of Brownlee. An appendix is devoted to computational considerations.

2. THE REGRESSION QUANTILE FUNCTION

Let $\mathbf{Y} = (Y_1, \ldots, Y_n)$ be a vector of independent random variables from the linear model

$$\Pr[Y_i \le y \mid \mathbf{x}_i] = F\left(y - \sum_{j=1}^{p} x_{ij}\beta_j\right) \qquad i = 1, \ldots, n, \quad (2.1)$$

where $\mathbf{x}_i$ denotes a row of the $n \times p$ design matrix $\mathbf{X}$. We will assume throughout that the design contains an intercept, more explicitly that $x_{i1} = 1$, for $i = 1, \ldots, n$. The *conditional quantile function* of Y given $\mathbf{x}$ may now be defined as

$$Q_Y(\theta \mid \mathbf{x}) = \mathbf{x}\boldsymbol{\beta} + F^{-1}(\theta),$$

which due to the intercept we may express as

$$Q(\theta \mid \mathbf{x}) = \mathbf{x}\boldsymbol{\beta}(\theta),$$

where $\boldsymbol{\beta}(\theta) = \boldsymbol{\beta} + F^{-1}(\theta)\mathbf{e}_1$ and $\mathbf{e}_1' = (1, 0, \ldots, 0) \in \mathbf{R}^p$.

We are interested in finding a definition of the ordinary sample quantiles in the location model that will extend naturally to the general model (2.1). By the location model we mean the important special case of (2.1) in which $\boldsymbol{\beta}$ is a scalar and $\mathbf{X} = 1_n$, a column of ones. The

[1] Parzen (1979) has recently stressed the importance of the empirical quantile function in the analysis of one-sample data.

* Gilbert Bassett, Jr. is Associate Professor of Economics, University of Illinois at Chicago Circle, Chicago, IL 60680. Roger Koenker is a member of the technical staff, Bell Laboratories, Murray Hill, NJ 07974.

© Journal of the American Statistical Association
June 1982, Volume 77, Number 378
Theory and Methods Section

usual notions involving an ordering of the sample observations are obviously inadequate when $p \geq 2$, since such orderings depend upon $\boldsymbol{\beta}$. However, a simple minimization problem provides the device we seek. Consider the "check" function

$$\rho_\theta(u) = \begin{cases} \theta u & \text{for } u \geq 0 \\ (\theta - 1)u & \text{for } u < 0 \end{cases} \quad \theta \in [0, 1]. \tag{2.2}$$

Any θth sample quantile from a realization $\mathbf{y} = (y_1, \ldots, y_n)$ of $\mathbf{Y}$ in the location model is a solution to the problem,

$$\min_{b \in \mathbf{R}} \left[\sum_{i=1}^{n} \rho_\theta(y_i - b) \right]. \tag{2.3}$$

The defining minimization problem (2.3) generalizes naturally to the linear model (2.1) as

$$\min_{\mathbf{b} \in \mathbf{R}^p} \left[\sum_{i=1}^{n} \rho_\theta(y_i - \mathbf{x}_i \mathbf{b}) \right], \tag{2.4}$$

whose p-dimensional solutions we call regression quantiles and have studied in Koenker and Bassett (1978). Least absolute deviation regression is an obviously important (median) special case (see Bassett and Koenker 1978). We denote the potentially set-valued solutions to (2.4) by $\hat{\mathbf{B}}(\theta)$, with generic element $\hat{\boldsymbol{\beta}}$. These p-dimensional analogs of the sample quantiles appear to successfully extend certain well-known L-estimates of location to the general linear model. In Koenker and Bassett (1978) we study finite linear combinations of regression quantiles of the form

$$\hat{\boldsymbol{\beta}}[\omega] = \sum_{i=1}^{m} \omega(\theta_i)\hat{\boldsymbol{\beta}}(\theta_i)$$

and it is shown that such estimators have essentially the same asymptotic theory as linear combinations of sample quantiles in the location model. See Bickel (1973) and Hogg (1975) for alternative approaches to L-estimators for linear models. Recently, Ruppert and Carroll (1980) have confirmed a conjecture appearing in Koenker and Bassett (1978) that trimmed least squares based on $\hat{\boldsymbol{\beta}}(\theta)$ has asymptotic behavior analogous to ordinary trimmed means in the location model. Perhaps more surprisingly, Ruppert and Carroll also show that trimmed least squares based upon residuals from a preliminary estimate has a different and considerably less appealing asymptotic behavior.

In Figures 1 and 2 we illustrate the regression quantiles for a very simple bivariate example with only 5 observations consisting of (x, y) pairs (1, 3), (2, 2), (4, 7), (7, 8), and (9, 6). For θ in the open intervals (0, 7/22), (7/22, 1/2), (1/2, 3/4), and (3/4, 1) there are unique solutions to (2.4) at, respectively, $\hat{\boldsymbol{\beta}}$ = (6/7, 4/7), (21/8, 3/8), (13/6, 5/6), and (17/3, 1/3). These solutions are illustrated by the solid lines in Figure 1. For $\theta \in \{0, 7/22, 1/2, 3/4, 1\}$ the solution set to (2.4) is not single valued. In Figure 2 we illustrate in parameter space the regression quantile estimates for this example. The four labeled points are the

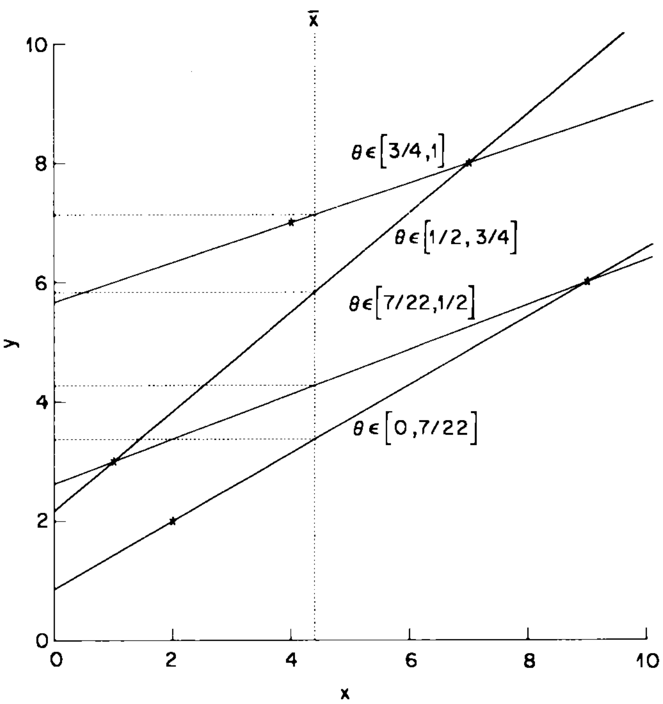

Figure 1. REGRESSION QUANTILES IN SAMPLE SPACE. This is a bivariate example with five observations. The four solid lines are the regression quantile estimates for the intervals (0, 7/22), (7/22, 1/2), (1/2, 3/4), and (3/4, 1) in ascending order, based on intersections with the vertical dashed line located at $\bar{x}$ = 4.4. The ordinates at the intersections determine the empirical quantile function for this example, which is illustrated in Figure 3.

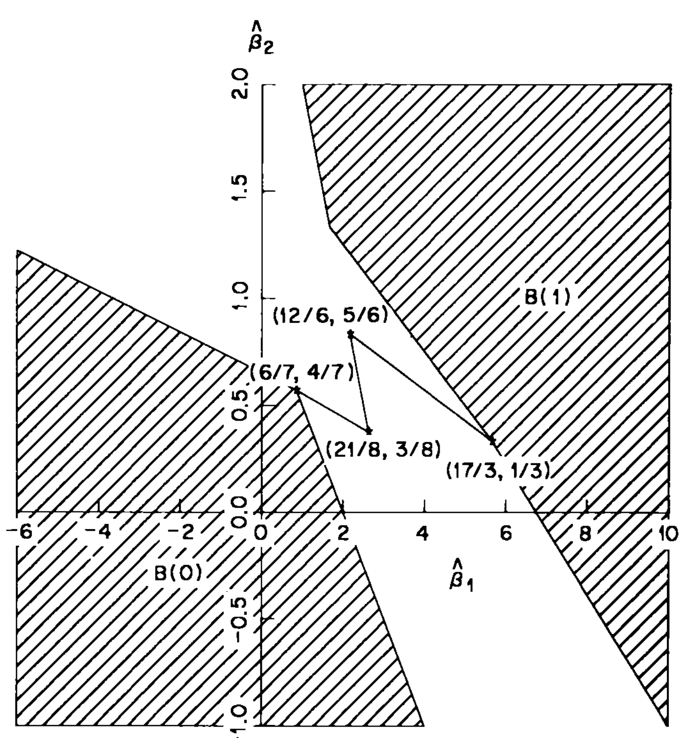

Figure 2. REGRESSION QUANTILES IN PARAMETER SPACE. The hatched regions are the sets of regression quantile estimates for θ = 0 and θ = 1. The four labeled points correspond to the four solid lines of the previous figure. The line segments connecting these points define the solutions to (2.4) at 7/22, 1/2, and 3/4.

unique solutions described above. The three closed line segments connecting these points illustrate the solutions at $\theta = 7/22$, $1/2$, $3/4$. At $\theta = 0$ ($\theta = 1$), any $b \in \mathbf{R}^p$ that makes all the residuals positive (negative) solves (2.4). The sets $B(0)$ and $B(1)$ are the hatched regions in Figure 2. Readers interested in computational details may wish to refer to the Appendix at this point.

As θ varies from zero to one the regression quantile estimates describe an "ascending" sequence of distinct $(p - 1)$-dimensional hyperplanes each of which pass through (at least) p sample points. The adjective "ascending" is necessarily somewhat vague at this juncture; nevertheless, it may be partially justified by the following general result from Koenker and Bassett (1978 Theorem 3.4): *there are at most $n\theta$ sample observations below and at least $(n - p)\theta$ observations above any θth regression quantile hyperplane*. Note that in the location model ($p = 1$) this property completely characterizes the quantiles.

For any design point $\mathbf{x}$ we define the empirical conditional quantile function

$$\hat{Q}(\theta \mid \mathbf{x}) = \inf\{\mathbf{x}\boldsymbol{\beta} \mid \hat{\boldsymbol{\beta}} \in \hat{\mathbf{B}}(\theta)\} \qquad (2.5)$$

In the location model, $\hat{Q}$ is simply a left-continuous version of the ordinary empirical quantile function as discussed recently, for example, by Parzen (1979). The geometry of $\hat{Q}$ may be easily seen in Figure 1. Choose a design point, say $\bar{x} = 4.4$, draw in the vertical line through the point, and note the intersections of the regression quantile lines with this vertical line. The ordinates at these intersections define the jump function $\hat{Q}$ depicted in Figure 3. The jumps occur at 7/22, 1/2, and 3/4. Superficially, this function looks like a garden-variety empirical quantile function. However, unlike the (one-sample) location model the jumps are not equally spaced on the θ-axis, but depend upon the design configuration as well as the realization of $\mathbf{Y}$. Furthermore, note that if we had chosen a design point like $x_2 = 8$, $\hat{Q}$ would not even have been monotone! However, at the centroid of the design, $\bar{\mathbf{x}} = n^{-1} \sum \mathbf{x}_i$, $\hat{Q}$ is assured to be monotone by the following result.

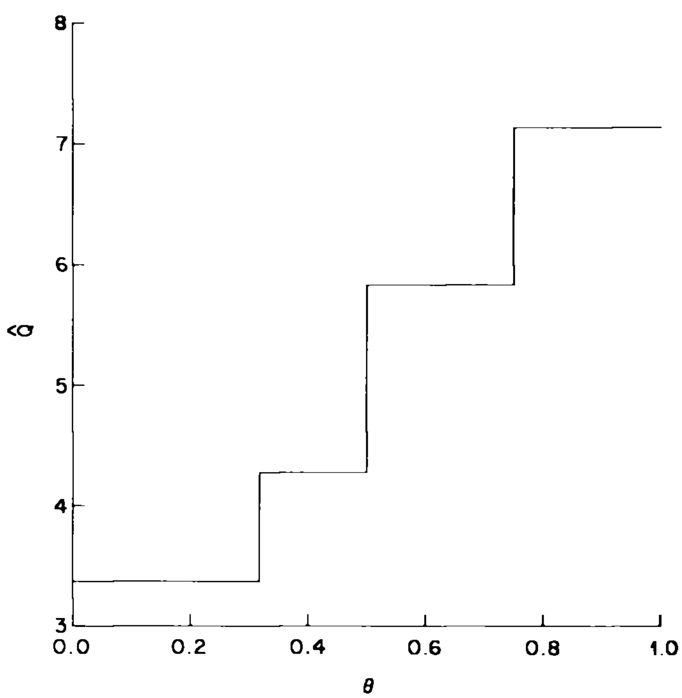

Figure 3. AN EMPIRICAL QUANTILE FUNCTION (EQF). This is $\hat{Q}(\theta \mid \bar{x} = 4.4)$ based on the data in Figure 1.

Theorem 2.1. The sample paths of $\hat{Q}(\theta \mid \bar{\mathbf{x}})$ are nondecreasing, left-continuous, jump functions on $(0, 1)$.

Proof. Let $r(\theta)$ denote the function implicitly defined in (2.4). We show that $\hat{Q}(\theta \mid \bar{\mathbf{x}})$ is a translate of the left derivative of $r(\theta)$, and that $r(\theta)$ is piecewise linear and concave. The theorem then follows from elementary monotonicity and continuity properties of derivatives of convex functions. Given a realization y from the model (2.1), define the polyhedral convex set

$$V = \{\mathbf{v} \in \mathbf{R}^p \times \mathbf{R}_+^{2n} \mid y = Z\mathbf{v}\},$$

where $Z = [X \; I \; -I]$. For $\mathbf{c} \in \mathbf{R}^{p+2n}$ let

$$\phi(\mathbf{c}) \equiv \sup\{\mathbf{c}'\mathbf{v} \mid \mathbf{v} \in V\} \qquad (2.6)$$

denote the support function of V. Let $\hat{V}(\mathbf{c}) \subseteq V$ denote the set on which $\mathbf{c}'\mathbf{v}$ achieves its maximum on V. Since V is polyhedral, it follows from Rockafellar (1970), Theorems 23.5.3 and 23.10, that the directional derivative function,

$$\begin{aligned} \phi'(\mathbf{c}; \mathbf{d}) &\equiv \lim_{\lambda \downarrow 0} [(\phi(\mathbf{c} + \lambda\mathbf{d}) - \phi(\mathbf{c}))/\lambda] \\ &= \sup\{\mathbf{d}'\hat{\mathbf{v}} \mid \hat{\mathbf{v}} \in \hat{V}(\mathbf{c})\}, \end{aligned} \qquad (2.7)$$

is a proper polyhedral convex function of $\mathbf{d}$.

Set $\mathbf{c}(\theta) = (\mathbf{0}_p, \theta\mathbf{1}_n, (1 - \theta)\mathbf{1}_n)'$. Then standard linear programming results imply (see, for example, the discussion in the Appendix)

$$r(\theta) = -\phi(-\mathbf{c}(\theta)) \qquad (2.8)$$

and $\hat{V}(-\mathbf{c}(\theta)) = \hat{\mathbf{B}}(\theta) \times \hat{\mathbf{U}}(\theta)$ having elements $(\hat{\boldsymbol{\beta}}, \hat{\mathbf{u}}^+, \hat{\mathbf{u}}^-) \in \hat{V}$ with $\hat{u}_i^+ = \max(0, y_i - \mathbf{x}_i\hat{\boldsymbol{\beta}})$ and $\hat{u}_i^- = -\min(0, y_i - \mathbf{x}_i\hat{\boldsymbol{\beta}})$ for $i = 1, \ldots, n$. The function $\phi(\cdot)$ is polyhedral convex, so $r(\cdot)$ is piecewise linear and concave. The left derivative of r is given by

$$\begin{aligned} r_-'(\theta) &= -\lim_{\lambda \downarrow 0} [r(\theta - \lambda) - r(\theta)]/\lambda \\ &= \lim_{\lambda \downarrow 0} [\phi(-\mathbf{c}(\theta - \lambda)) - \phi(-\mathbf{c}(\theta))]/\lambda. \end{aligned}$$

Letting $\mathbf{c}_\theta = \partial\mathbf{c}(\theta)/\partial\theta = (\mathbf{0}_p, \mathbf{1}_n, -\mathbf{1}_n)'$, then

$$\begin{aligned} r_-'(\theta) &= \lim_{\lambda \downarrow 0} [\phi(-\mathbf{c}(\theta) + \lambda\mathbf{c}_\theta) - \phi(-\mathbf{c}(\theta))]/\lambda \\ &= \phi'(-\mathbf{c}(\theta); \mathbf{c}_\theta) = \sup\{\mathbf{c}'_\theta\hat{\mathbf{v}} \mid \hat{\mathbf{v}} \in \hat{V}(-\mathbf{c}(\theta))\} \\ &= \sup\{\mathbf{1}'_n(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}(\theta) \mid \hat{\boldsymbol{\beta}}(\theta) \in \hat{\mathbf{B}}(\theta)\} \\ &= \mathbf{1}'_n\mathbf{y} - n\hat{Q}(\theta, \bar{\mathbf{x}}) \end{aligned} \qquad (2.9)$$

The stated properties of $\hat{Q}$ now follow immediately from the properties of r.

This result establishes that $\hat{Q}(\theta \mid \bar{\mathbf{x}})$, henceforth denoted simply by $\hat{Q}(\theta)$ is a proper quantile function. In the location model the jumps in the empirical quantile function occur at equally spaced points $\{i/n: i = 1, \ldots, n\}$ on $(0, 1)$. However, in the linear model the jumps in $\hat{Q}(\cdot)$ occur at random points on $(0, 1)$ depending in a rather complicated way upon the design as well as the realization of $\mathbf{Y}$. The following result establishes that on intervals of fixed length the number of jumps is at least of order n.

Theorem 2.2. If F is continuous then for any t and s in $(0, 1)$ satisfying $t - s > p/n$, $\hat{Q}(t) = \hat{Q}(s)$ with probability zero.

Proof. Suppose that $\hat{Q}(t) = \hat{Q}(s)$ for some $t - s > p/n$. Let $\hat{\boldsymbol{\beta}}_s \in \hat{\mathbf{B}}(s)$ satisfy $\hat{Q}(s) = \bar{\mathbf{x}}\hat{\boldsymbol{\beta}}_s$. Then

$$r(\theta) = r(s) + (\theta - s)r_-'(s) \tag{2.10}$$

for $\theta \in [s, t]$ since by hypothesis $r_-'(s)$ is constant on this interval. Since $r_-'(s) = \mathbf{1}'_n(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}_s)$, the left side evaluated at $\theta = t$ equals $\sum_i \rho_t(y_i - \mathbf{x}_i\hat{\boldsymbol{\beta}}_s)$, hence $\hat{\boldsymbol{\beta}}_s \in \hat{\mathbf{B}}(t)$. However, by Theorem 3.4 of Koenker and Bassett (1978) we have for all $\hat{\boldsymbol{\beta}} \in \hat{\mathbf{B}}(\theta)$ the inequalities,

$$\#\{i \mid u(\hat{\boldsymbol{\beta}}) < 0\} \leq n\theta \leq \#\{i \mid u(\hat{\boldsymbol{\beta}}) \leq 0\}, \tag{2.11}$$

where $u(\mathbf{b}) = \mathbf{y} - \mathbf{X}\mathbf{b}$ and $\#S$ denotes the cardinality of S. Since F is continuous $P[\#\{i \mid u(b) = 0\} > p] = 0$ for all $\mathbf{b} \in \mathbf{R}^p$ so with probability one (2.11) becomes

$$n\theta - p \leq \#\{i \mid u(\hat{\boldsymbol{\beta}}) < 0\} \leq n\theta \tag{2.12}$$

for all $\hat{\boldsymbol{\beta}} \in \hat{\mathbf{B}}(\theta)$ and $\theta \in (0, 1)$. However, $ns \leq nt - p$ by hypothesis so $\hat{\mathbf{B}}(s) \cap \hat{\mathbf{B}}(t)$ is empty with probability one, contradicting $\hat{\boldsymbol{\beta}}_s \in \hat{\mathbf{B}}(t)$.

Remark. It is tempting to conjecture that the subset of $(0, 1)$ on which $\hat{\mathbf{B}}(\theta)$ is nonunique must have measure zero. However, as the proof of the preceding result implies, this need not be the case. This is illustrated in the following example. Consider the (x, y) sample points $\{(-1, -3), (-1, 3), (0, 0), (1, -1), \text{and } (1, 1)\}$. Then, $\hat{\mathbf{B}}(\theta) = \{\mathbf{b} \in \mathbf{R}^2 \mid \lambda(0, 1) + (1 - \lambda)(0, -1)\}$ for $\theta \in (2/5, 3/5)$. This example also illustrates that $\hat{\mathbf{B}}(\theta)$ may have dimension p since $\hat{\mathbf{B}}(2/5)$ is the convex hull of the parameter points $\{(0, 1), (0, -1), (-2, 1), (-1, 2)\}$. The salient feature of this rather perverse example is that the sequence of basic solutions to the parametric programming problem posed in the Appendix pivots on a design point at $\bar{\mathbf{x}}$, in this case $\bar{\mathbf{x}} = 0$. This example also illustrates a case in which the upper bound of p/n on the length of an interval without a jump is achieved, $\hat{\mathbf{B}}(\theta) = (-2, 1)$ for $\theta \in (0, 2/5)$.

A number of equivariance properties of $\hat{Q}(\theta \mid \mathbf{x})$ follow immediately from Theorem 3.2 of Koenker and Bassett (1978). Expressing $\hat{Q}(\theta \mid \mathbf{x}; \mathbf{y}, \mathbf{X})$ as an explicit function of the response vector, $\mathbf{y}$, and the design matrix $\mathbf{X}$, we have

Theorem 2.3. For $\lambda \geq 0$, $\boldsymbol{\delta} \in \mathbf{R}^p$, and any nonsingular $p \times p$ matrix $\mathbf{A}$:

(i) $\hat{Q}(\theta \mid \mathbf{x}; \lambda\mathbf{y}, \mathbf{X}) = \lambda\hat{Q}(\theta \mid \mathbf{x}; \mathbf{y}, \mathbf{X})$
(ii) $\hat{Q}(\theta \mid \mathbf{x}; -\lambda\mathbf{y}, \mathbf{X}) = \lambda\hat{Q}((1 - \theta) \mid \mathbf{x}; \mathbf{y}, \mathbf{X})$
(iii) $\hat{Q}(\theta \mid \mathbf{x}; \mathbf{y} + \mathbf{X}\boldsymbol{\delta}, \mathbf{X}) = \hat{Q}(\theta \mid \mathbf{x}; \mathbf{y}, \mathbf{X}) + \mathbf{x}\boldsymbol{\delta}$
(iv) $\hat{Q}(\theta \mid \mathbf{x}; \mathbf{y}, \mathbf{X}\mathbf{A}) = \hat{Q}(\theta) \mid \mathbf{x}; \mathbf{y}, \mathbf{X})$.

$\hat{Q}$ is scale and location equivariant by (i)–(iii) and invariant to reparameterization of design by (iv).

3. ASYMPTOTIC BEHAVIOR

To study the large-sample behavior of $\hat{Q}$ we make the following additional assumptions:

A1. (Density) The error distribution, F, has continuous and strictly positive density, f, for all z such that $0 < F(z) < 1$.

A2. (Design) $\lim_{n\to\infty} n^{-1}\mathbf{X}'\mathbf{X} = \mathbf{D}$, a positive definite matrix.

Under these assumptions we have the following result from Koenker and Bassett (1978).

Theorem 3.1. Let $\{\hat{\boldsymbol{\beta}}_n(\theta_1), \ldots, \hat{\boldsymbol{\beta}}_n(\theta_m)\}$ with $0 < \theta_1 < \cdots < \theta_m < 1$ denote a sequence $(n \to \infty)$ of regression quantiles. Then

$$\sqrt{n}\,(\hat{\boldsymbol{\beta}}_n(\theta_1) - \boldsymbol{\beta}(\theta_1), \ldots, \hat{\boldsymbol{\beta}}_n(\theta_m) - \boldsymbol{\beta}(\theta_m))$$

converges in law to an mp-variate Gaussian random vector with zero mean and covariance matrix $\boldsymbol{\Omega}(\theta_1, \ldots, \theta_m; F) \otimes \mathbf{D}^{-1}$, where $\boldsymbol{\Omega}$ has typical element

$$\omega_{ij} = \frac{\min(\theta_i, \theta_j) - \theta_i\theta_j}{f(F^{-1}(\theta_i))f(F^{-1}(\theta_j))}. \tag{3.1}$$

Thus in linear models with iid errors the asymptotic behavior of linear combinations of regression quantiles is remarkably similar to the large-sample theory of ordinary sample quantiles in the one-sample, location model. Note in the latter case $\mathbf{D} = 1$.

We may now state the main result of this section.

Theorem 3.2. The finite dimensional distributions of the random function

$$Z_n(\theta \mid \mathbf{x}) = \sqrt{n}\,(\hat{Q}(\theta \mid \mathbf{x}) - Q(\theta \mid \mathbf{x})) \tag{3.2}$$

are asymptotically Gaussian with zero mean and covariance matrix $\mathbf{x}\mathbf{D}^{-1}\mathbf{x}'\boldsymbol{\Omega}$. Specializing to $\mathbf{x} = \bar{\mathbf{x}}$ we have $Z_n(\theta \mid \bar{\mathbf{x}}) \to G(0, \boldsymbol{\Omega})$.

Proof. This is immediate from the definition of $\hat{Q}$, Theorem 3.1, and the identity $\bar{\mathbf{x}}\mathbf{D}^{-1}\bar{\mathbf{x}}' = 1$ which follows from the fact that $\mathbf{1}$ is in the column space of $\mathbf{X}$.

Thus, $\hat{Q}(\theta \mid \mathbf{x})$ is (weakly) consistent for $Q(\theta \mid \mathbf{x})$. In Bassett and Koenker (1982) we study the strong consistency of $\hat{Q}$ under somewhat different regularity conditions on F. The asymptotic Gaussianity of the finite-dimensional distributions of Z_n extends classical results of Mosteller (1946) and Cramér (1946). Weak convergence of the function $Z_n(\cdot)$ is an open problem.

Extending our analogy to the estimation of the conditional distribution function of Y, we may define the estimator,

$$\hat{F}(y \mid \mathbf{x}) = \sup\{\theta \mid \hat{Q}(\theta \mid \mathbf{x}) \le y\}. \tag{3.3}$$

We have from Theorem 2.1 that $\hat{F}(y) \equiv \hat{F}(y \mid \bar{\mathbf{x}})$ is a nondecreasing, right-continuous, jump function on $\mathbf{R}^1$. Versions of $\hat{Q}(\cdot)$ and $\hat{F}(\cdot)$ with continuous sample paths are easily constructed. Let $\{(q_i, p_i): i = 1, \ldots, n\}$ denote the points at which the function $\hat{Q}(\cdot)$ jumps. Set $q_0 = q_1 - 1/n$, and $p_0 = 0$. Now define the piecewise linear function

$$\tilde{Q}(\theta) = \left(\frac{\theta - p_{i-1}}{p_i - p_{i-1}}\right) q_{i-1} + \left(\frac{p_i - \theta}{p_i - p_{i-1}}\right) q_i, \tag{3.4}$$
$$\theta \in [p_{i-1}, p_i), \quad i = 1, \ldots, n.$$

Inverting $\tilde{Q}$, we define $\tilde{F}(y) \equiv \tilde{Q}^{-1}(y)$. Suppose $\tilde{Q}$ has the same finite-dimensional asymptotic distributions as $\hat{Q}$; then we would have

$$\begin{aligned} P[\sqrt{n}\,(\tilde{Q}(\theta) - Q(\theta)) \le y] &= P[\tilde{Q}(\theta) \ge Q(\theta) + y/\sqrt{n}] \\ &= P[\theta \le \tilde{F}(Q(\theta) + y/\sqrt{n})] \\ &= P[\sqrt{n}(\hat{F}(z_n) - F(z_n)) \ge \sqrt{n}\,(\theta - F(z_n)) \\ &\quad + \sqrt{n}\,(\hat{F}(z_n) - \tilde{F}(z_n))], \end{aligned} \tag{3.5}$$

where $z_n = Q(\theta) + y/\sqrt{n}$. However, $\sqrt{n}\,(\hat{F}(z_n) - \tilde{F}(z_n)) = 0(1/\sqrt{n})$ by Theorem 2.2, and expanding $F(z_n)$ around $Q(\theta)$ we have the right side of (3.5) converging to $-yf(Q(\theta))$. Hence $\hat{F}(\cdot)$ would have the asymptotic covariance function

$$\sigma(z_1, z_2) = F(z_1)(1 - F(z_2)) \quad z_1 \le z_2. \tag{3.6}$$

Finally, under the null hypothesis H_0: $F = F_0$, $\sqrt{n}\,(\hat{F}(F_0^{-1}(\theta)) - \theta)$ would have the covariance function,

$$\sigma_0(\theta_1, \theta_2) = \theta_1(1 - \theta_2) \quad \theta_1 \le \theta_2. \tag{3.7}$$

Tightness of the random function $\sqrt{n}\,(\hat{F}(F_0^{-1}(\theta)) - \theta)$ and hence its convergence to tied-down Brownian motion also remains an open and rather tantalizing question. Affirmative answers would open the way to Kolmogorov-Smirnov tests of distributional hypotheses, and other tests based on the empirical distribution function.

4. COMPARISON WITH PROCEDURES BASED ON RESIDUALS

It is interesting at this point to compare the asymptotic behavior of $\hat{Q}$ with that of alternative estimators of the conditional quantile function based on residuals from a preliminary estimate, say $\boldsymbol{\beta}^*$, of $\boldsymbol{\beta}$. Let $\mathbf{u}^* = \mathbf{y} - \mathbf{X}\boldsymbol{\beta}^*$ and let $\hat{\mathbf{u}}^*(\theta)$ denote a θth sample quantile from the residual vector $\mathbf{u}^* \in \mathbf{R}^n$. For convenience assume $\bar{\mathbf{x}} = \mathbf{e}_1$. Let $\boldsymbol{\delta}_n = \sqrt{n}\,(\boldsymbol{\beta}_n^* - \boldsymbol{\beta})$ and let $\psi_\theta = \theta - I(x < 0)$ denote the (θth-quantile) influence function, which is simply the (right) derivative of $\rho_\theta(\cdot)$ defined in (2.2). Ruppert and Carroll (1980) obtain the following useful asymptotic linearity result. If $\boldsymbol{\delta}_n = 0_p(1)$, then

$$\begin{aligned} \sqrt{n}\,(\hat{\mathbf{u}}^*(\theta) - F^{-1}(\theta)) &= f(F^{-1}(\theta))^{-1}\left[\frac{1}{\sqrt{n}}\sum_{i=1}^{n} \psi_\theta(z_i - F^{-1}(\theta))\right] \\ &\quad + \mathbf{e}'_1\boldsymbol{\delta}_n + o_p(1). \end{aligned} \tag{4.1}$$

When $\boldsymbol{\beta}^*$ is an M-estimator with influence function proportional to ψ satisfying the typical asymptotic expansion

$$\sqrt{n}\,(\boldsymbol{\beta}_n^* - \boldsymbol{\beta}) = \mathbf{D}^{-1}\sum_{i=1}^{n} [\psi(z_i)/E\psi']\mathbf{x}'_i + o_p(1), \tag{4.2}$$

then $\mathbf{e}'_1\boldsymbol{\delta}_n = \sum \psi(z_i)/E\psi' + o_p(1)$, and $\sqrt{n}\,(\hat{\mathbf{u}}_n^*(\theta) - F^{-1}(\theta))$ has asymptotically Gaussian finite dimensional distributions with zero mean and covariance function

$$\sigma(\theta_i, \theta_j) = \omega_{ij} + \frac{1}{E\psi'}\left[\frac{\boldsymbol{\Psi}(\theta_i)}{f(F^{-1}(\theta_i))} + \frac{\boldsymbol{\Psi}(\theta_j)}{f(F^{-1}(\theta_j))}\right] + \frac{V\psi}{(E\psi')^2}, \tag{4.3}$$

where $\boldsymbol{\Psi}(\theta) = \int_{-\infty}^{F^{-1}(\theta)} \psi(z)dF(z)$.

When $\boldsymbol{\beta}^*$ is a maximum likelihood estimator so $\psi = f'/f$, the $E\psi' = V\psi = I(F)$, the Fisher information of F. Also $\boldsymbol{\Psi}(\theta) = -f(F^{-1}(\theta))$, so the covariance in (4.3) becomes simply,

$$\sigma(\theta_i, \theta_j) = \omega_{ij} - I(F)^{-1}. \tag{4.4}$$

Corresponding formulas for the covariance function of the empirical distribution function from M-estimator residuals appear in Bloomfield (1974). Pierce and Kopecky (1979) treat the maximum likelihood case. Durbin (1973) provides very general treatment of the asymptotic behavior of empirical processes when parameters are estimated, but does not treat the regression case explicitly.

Note that adding $\mathbf{e}'_1\boldsymbol{\delta}_n$ to both sides of (4.1) and denoting $\hat{Q}_n^*(\theta) = \hat{\mathbf{u}}_n^*(\theta) + \mathbf{e}'_1\boldsymbol{\beta}^*$ gives

$$\sqrt{n}\,(\hat{Q}_n^*(\theta) - Q(\theta)) = f(F^{-1}(\theta))^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^{n} \psi_\theta(z_i - F^{-1}(\theta)) + o_p(1). \tag{4.5}$$

This random function has finite dimensional distributions that are asymptotically the same as the $Z_n(\theta)$ defined above based on the regression quantile function under the crucial proviso that $\sqrt{n}\,(\boldsymbol{\beta}^* - \boldsymbol{\beta})$ is bounded in probability. This fails, for example, in the case of least squares estimates with Cauchy errors; in such a case $\hat{Q}_n^*(\theta)$ need not even be consistent.

In finite samples there is a commonly observed tendency for least squares residuals to look more Gaussian than the true errors. Each residual is the sum of two components: one is the true error, the other is a linear function of the entire n vector of errors. The latter component is (typically) asymptotically Gaussian irrespective of the true F. Hence, the distribution of residuals is a convolution of F with an approximately Gaussian distribution. However, since the Gaussian component is $O(T^{-1})$ the bias vanishes asymptotically. Gnanadesikan (1977) has called this tendency of least squares to appear Gaussian "supernormality." Bloomfield (1974) provides asymptotic expansions that give an explicit justification of this view.

The example discussed in Section 5 illustrates this effect and helps to contrast the behavior of $\hat{Q}(\theta)$ with $\hat{Q}^*(\theta)$ based on a least squares preliminary estimate.

5. AN EXAMPLE

We now present a brief analysis of the well-known "stackloss" data of Brownlee (1965). The data consist of 21 observations taken on successive days from a plant for the oxidation of ammonia to nitric acid. The dependent variable is the proportion of ammonia escaping unabsorbed. The independent variables are x_1, the rate of air flow; x_2, the inlet temperature of the cooling water; and x_3, the concentration of nitric acid in the absorbing liquid. For a detailed analysis of the data see Daniel and Wood (1971). Subsequent analysis of the data has been made by Andrews (1974), Cook (1979), Denby and Mallows (1977), Hettmansperger and McKean (1977), and many others. Despite the intensive analysis the data have received, our methods reveal some new and rather surprising features.

In Table 1 we report the complete regression quantile statistics for the stackloss data. Each row of the table gives an estimated coefficient vector $\hat{\beta}(\theta)$ that minimizes function (2.4) for the interval of θ's given in columns one and two. The last column lists the observations that have zero residuals for each quantile estimate. We note as expected that the outlying observations 3, 4, and 21 appear only in the very extreme quantiles, while observations 1 and 2, which are frequently also identified as outliers in robust analyses, appear in a wide range of quantiles. Both observations 1 and 2 arise from outlying design points with substantial *leverage*; for a detailed analysis of their influence see Cook (1979).

Perhaps the most remarkable feature of the analysis is the fact that *eight* of the 21 sample points lie on the plane represented by

$$y = -36 + \tfrac{1}{2}x_1 + x_2,$$

which is the first quartile estimate. This is a linear model equivalent of "tied" observations in a one-sample problem. It is interesting that the tied observations {6, 7, 13, 14, 16, 17, 18, 19} are nearly consecutive, possibly suggesting a distinct "regime" for plant operation.

With a continuous error distribution ties occur with probability zero; in this example we would "almost surely" have regression quantile estimates with four nonzero coefficients and four zero residuals. Ties would indicate a mass point in the error distribution. However, with the data rounded to two significant figures as in this example we would obviously not infer a mass point in F from the first quantile fit, but as we shall see there is some evidence of a mode in the distribution near the first quartile.

In Figure 4 we plot the function $\hat{Q}(\theta)$ for the stackloss example centered at zero. Several features of this figure stand out. $\hat{Q}(\theta)$ is quite flat at all three quartiles, and particularly so at the first quartile, where we have already identified the surprising "tied" observations. The derivative of a smooth quantile function is the reciprocal of a density function, aptly called the sparsity function by Tukey (1965). Thus flat segments in $\hat{Q}$ indicate low sparsity, hence high density. This impression is strengthened by examining density estimates based on smoothing $\hat{Q}$.

Table 1. Regression Quantile Estimates for the Stackloss Data

Quantiles						
Upper	Lower	Intercept	Airflow	Temperature	Acid	Basic Observations
.0	.124	−29.01	.31	1.22	−.03	9, 17, 19, 21
.124	.130	−36.08	.35	1.75	−.09	7, 17, 19, 21
.130	.275	−36.00	.50	1.00	.0	6, 7, 13, 14, 16, 17, 18, 19
.275	.331	−37.90	.76	.79	−.10	5, 7, 14, 17
.331	.375	−38.53	.84	.73	−.13	2, 7, 14, 17
.375	.392	−32.64	.83	.74	−.19	2, 7, 14, 18
.392	.409	−33.25	.85	.63	−.17	2, 7, 10, 18
.409	.490	−39.65	.83	.58	−.06	2, 8, 10, 18
.490	.565	−39.69	.83	.57	−.06	2, 8, 16, 18
.565	.592	−39.73	.83	.56	−.06	2, 8, 12, 18
.592	.604	−39.68	.84	.56	−.06	8, 12, 16, 18
.604	.620	−54.03	.87	.98	−.003	1, 12, 16, 18
.620	.651	−54.06	.87	.98	−.002	1, 10, 12, 18
.651	.690	−54.07	.87	.98	.002	1, 10, 12, 19
.690	.762	−54.19	.87	.98	.0	1, 10, 11, 19
.762	.768	−54.09	.86	1.00	.0	1, 10, 11, 12
.768	.774	−54.34	.78	1.18	.02	1, 10, 12, 15
.774	.777	−56.68	.78	1.25	.03	10, 12, 15, 20
.777	.814	−58.55	.80	1.27	.03	3, 10, 12, 20
.814	.834	−59.38	.80	1.25	.05	3, 10, 15, 20
.834	.913	−58.54	.79	1.30	.04	3, 10, 12, 15
.913	1.000	−58.46	.52	1.85	.11	3, 4, 10, 12

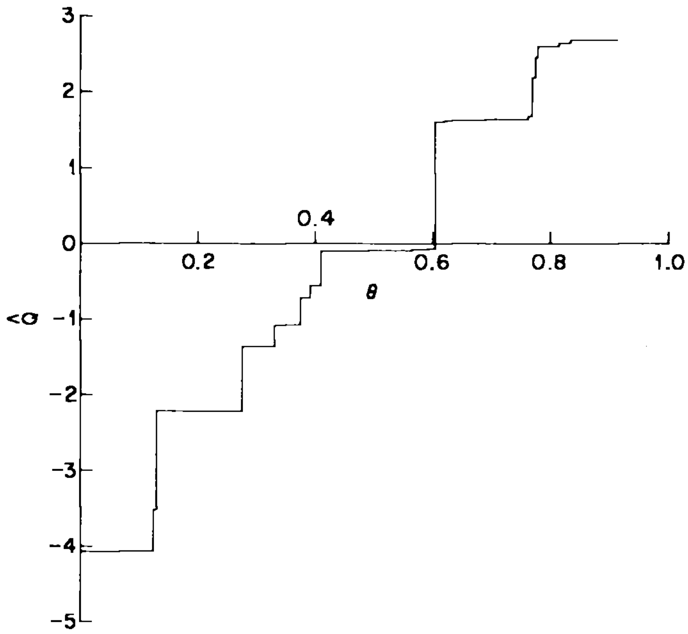

Figure 4. STACKLOSS EQF BASED ON REGRESSION QUANTILES. This is $\hat{Q}(\theta \mid \bar{x}) - \bar{y}$ based on the stackloss data.

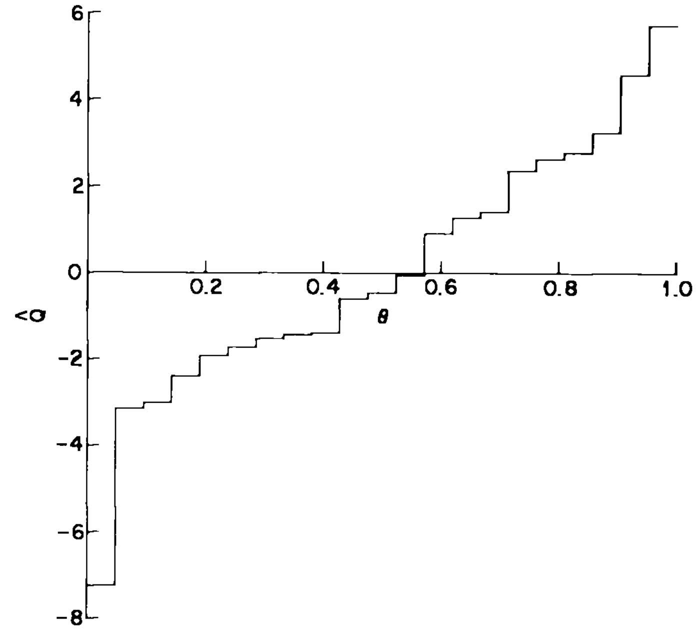

Figure 5. STACKLOSS EQF BASED ON LEAST-SQUARES RESIDUALS. This is the sample quantile function of the residuals from a least-squares fit of the stackloss data.

It is interesting to compare our estimates of $\hat{Q}$ based on regression quantiles with estimates based on least squares and least absolute residuals. These alternative estimates are depicted in Figures 5 and 6, respectively. The estimates are quite different. The least squares estimate is quite uniform, that is, linear, except in the tails where the influence of the outlying observations 3, 4, and 21 is felt. The l_1 estimate has a large flat segment at the median which as we have already noted implies a large mode. The effect is attributable to the four residuals that are exactly zero and "determine" the l_1 fit. Four observations 1, 3, 4, and 21 are extreme outliers in this plot. Figures 5 and 6 correspond closely to Figures 1 and 3 of Andrews (1974), which give probability plots of residuals from least squares and robust fits of the stackloss model. A striking difference among the three estimates is the range of $\hat{Q}$. As we would expect, the l_1 range is somewhat larger than the l_2 range since the latter estimate is extremely sensitive to outlying observations. However, the quantile regression estimate has considerably smaller range than either the l_1 or l_2 alternative; Why is this?

An answer is most easily provided by considering a simple artificial example. Figure 7 illustrates a sample of 100 observations generated from the (heteroscedastic) model

$$y_i = x_i + x_i\epsilon_i,$$

where x_i's are independently drawn from a uniform distribution on [0, 1] and the ϵ_i's are independently drawn from a standard Gaussian distribution. Superimposed on the scatter are the regression quantile estimates for $\theta \in \{.05, .25, .50, .75, .95\}$. A vertical line is drawn through $\bar{x}$, and ordinates at intersections of the quantile lines with this vertical line are points on $\hat{Q}(\cdot)$. Clearly the range of $\hat{Q}(\cdot)$ is considerably smaller than the range of any empirical quantile function based on residuals, least squares or otherwise.

Confirmation of the hypothesis that the differences in observed ranges in Figures 4 through 6 can be attributed to heteroscedasticity in the stackloss data may be found in Table 1. In a strictly iid case we would expect to find that the slope coefficients were essentially invariant to the quantile being estimated. In Koenker and Bassett (1982) we propose tests for heteroscedasticity based on the large-sample behavior of the regression quartiles

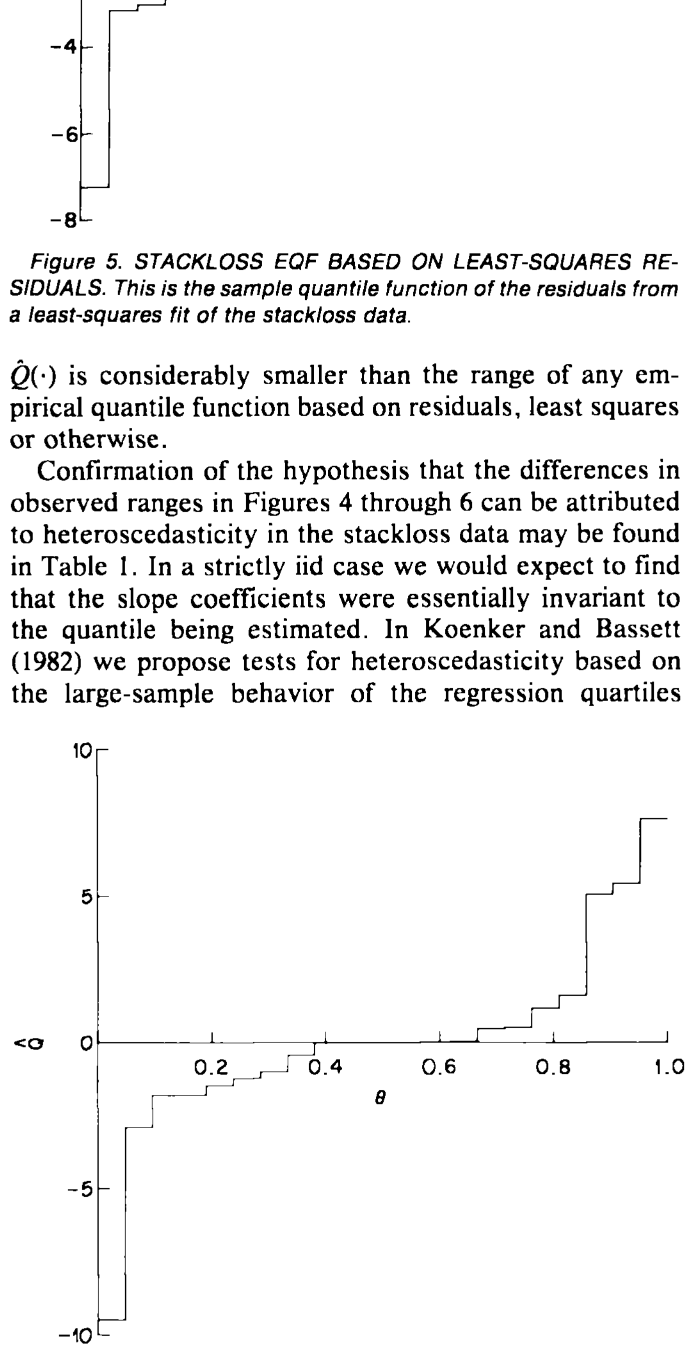

Figure 6. STACKLOSS EQF BASED ON LEAST ABSOLUTE RESIDUALS. This is the sample quantile function of the residuals from a least absolute residual fit of the stackloss data.

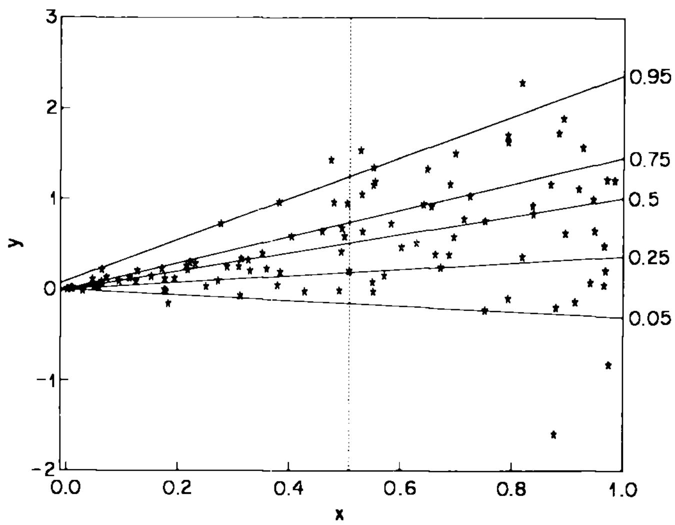

Figure 7. REGRESSION QUANTILES IN AN HETEROSCEDASTIC CASE. The scatter is an artificial sample of 100 points from a heteroscedastic model with Gaussian noise. Superimposed on the scatter are the regression quantile estimates for $\theta \in \{.05, .25, .50, .75, .95\}$. The vertical dashed line is drawn through $\bar{x}$.

under the null hypothesis of iid errors. For the stackloss data our test that the slope coefficients are identical at the three quartiles yields a χ^2 of 25.2 on six degrees of freedom, which is significant at 1 percent. The predominant component of this χ^2 is the difference of .33 in the airflow coefficient between the first and second quartiles. With this significant degree of heteroscedasticity it is not surprising, in light of the preceding examples, that $\hat{Q}$ exhibits a smaller range than estimates based on residuals.

We have identified two important features of the stackloss data that have not been apparent in previous investigations. First, we have seen that the first quartile estimate fits a substantial fraction (more than a third) of the data "exactly" and may provide a reasonable alternative to robust fits proposed to date. Second, a significant degree of heteroscedasticity has been found to be associated with the rate of air flow.

APPENDIX

The problem posed in (2.4) is equivalent to the linear program:

$$\min[\theta \mathbf{1}'_n \mathbf{u}^+ + (1 - \theta)\mathbf{1}'_n \mathbf{u}^-] \qquad (A.1)$$

subject to:

$$\mathbf{y} = \mathbf{X}\mathbf{b} + \mathbf{u}^+ - \mathbf{u}^- \quad (\mathbf{b}, \mathbf{u}^+, \mathbf{u}^-) \in \mathbf{R}^p \times \mathbf{R}_+^{2n}.$$

Clearly, $\min\{\mathbf{u}_j^+, \mathbf{u}_j^-\} = 0$ at any solution. Barrodale and Roberts (1974) and Bartels and Conn (1980) have suggested computationally efficient methods of solving (A.1) for the important case $\theta = \frac{1}{2}$. Both algorithms are easily adapted to the general case. Two forms of the dual problem to (A.1) are given in Koenker and Bassett (1978).

Given a solution to (A.1) for a fixed $\theta_0 \in [0, 1]$, how do we find solutions for the rest of the interval? In particular, how do we compute the function $r(\theta)$, the value of the optimized objective function, on the interval $[0, 1]$? Since $r(\theta)$ is piecewise linear this question reduces to finding the mesh on $[0, 1]$ on which kinks occur in r. These kink points are points at which $\hat{Q}$ jumps; elsewhere it is flat. Fortunately, this mesh may be easily found by standard parametric linear programming techniques. A useful general reference to this subject is the recent book by Gal (1979).

We will briefly describe the iterative process involved in identifying the mesh. Let $\mathbf{N} = \{1, 2, \ldots, n\}$, and $\mathbf{H}$ be the set of p-element subsets of $\mathbf{N}$. Elements $h \in \mathbf{H}$ and $\bar{h} \equiv \mathbf{N} - h$ will be used to partition the sample observations. Thus, y_h will denote a p vector with elements $\{y_i: i \in h\}$ and $\mathbf{X}_h$ denotes a p by p matrix with rows $\{\mathbf{x}_i: i \in h\}$. Now, $\boldsymbol{\beta}_h = \mathbf{X}_h^{-1}\mathbf{y}_h$ is a unique, nondegenerate solution to (A.1) for $\theta = \theta_0$ if and only if (see Koenker and Bassett 1978, Theorem 3.3)

$$(\theta_0 - 1)\mathbf{1}'_p < \sum_{i \in \bar{h}} [\tfrac{1}{2} - \tfrac{1}{2}\operatorname{sgn}(y_i - \mathbf{x}_i\boldsymbol{\beta}_h) - \theta_0]\mathbf{x}_i\mathbf{X}_h^{-1} < \theta_0 \mathbf{1}'_p. \qquad (A.2)$$

For $\theta \neq \theta_0$, $\boldsymbol{\beta}_h$ remains optimal until these conditions are violated. Thus, starting from θ_0 we have a system of p double inequalities in θ,

$$\theta - 1 < a_j + b_j\theta < \theta \quad j = 1, \ldots, p \qquad (A.3)$$

Let $\boldsymbol{\Theta}_0 = \{a_j/(1 - b_j), (a_j + 1)/(1 - b_j): j = 1, \ldots, p\}$ and define

$$\theta_1 = \min_{\theta > \theta_0}\{\theta \in \boldsymbol{\Theta}_0\},$$

and

$$\theta_{-1} = \max_{\theta < \theta_0}\{\theta \in \boldsymbol{\Theta}_0\}.$$

$\boldsymbol{\beta}_h$ solves (A.1) for all $\theta \in [\theta_{-1}, \theta_1]$. At each of the endpoints there will be two distinct solutions with index sets of "basic" observations, h and h', differing in one element. At θ_1 we make one simplex pivot from $\boldsymbol{\beta}_h$ to a new basic solution, $\boldsymbol{\beta}_{h'}$, recompute the a's and b's of the system (A.3) and continue the iteration to find θ_2, etc. Thus once an initial solution is determined the entire function may be found with $O(n)$ pivots.

[*Received March 1979. Revised December 1981.*]

REFERENCES

ANDREWS, D.F. (1974), "A Robust Method for Multiple Linear Regression," *Technometrics*, 16, 523–531.

ANSCOMBE, F.J. (1961), "Examination of Residuals," *Proceedings of the 4th Berkeley Symposium on Mathematical Statistics and Probability*, 1, 1–36.

——— (1967), "Topics in the Investigation of Linear Relations Fitted by Least Squares," *Journal of the Royal Statistical Society*, Series B, 29, 1–52.

ANSCOMBE, F.J., and TUKEY, J.W. (1963), "The Examination and Analysis of Residuals," *Technometrics*, 5, 141–162.

BARRODALE, I., and ROBERTS, F.D.K. (1974), "Solution of an Overdetermined System of Equations in the l1 Norm," *Communications of the Association for Computing Machinery*, 17, 319–320.

BARTELS, R.H., and CONN, A.R. (1980), "Linearly Constrained Discrete l1 Problems," *Association for Computing Machinery Transactions on Mathematical Software*, 6, 594–608.

BASSETT, G.W., and KOENKER, R.W. (1978), "The Asymptotic Distribution of the Least Absolute Error Estimator," *Journal of the American Statistical Association*, 73, 618–622.

BASSETT, G.W., and KOENKER, R.W. (1982), "Strong Consistency of Regression Quantiles," *Journal of Econometrics*, forthcoming.

BICKEL, P.J. (1973), "On Some Analogues to Linear Combinations of Order Statistics in the Linear Model," *Annals of Statistics*, 1, 597–616.

BLOOMFIELD, P. (1974), "On the Distribution of the Residuals From a Fitted Linear Model," Technical Report 56, Department of Statistics, Princeton University.

BROWNLEE, K.A. (1965), *Statistical Theory and Methodology in Science and Engineering*, New York: John Wiley.

COOK, R.D. (1979), "Influential Observations in Regression," *Journal of the American Statistical Association*, 74, 169–174.

CRAMER, H. (1946), *Mathematical Methods of Statistics*, Princeton: Princeton University Press.

DANIEL, C., and WOOD, F.S. (1971), *Fitting Equations to Data*, New York: John Wiley.

DENBY, L., and MALLOWS, C.L. (1977), "Two Diagnostic Displays for Robust Regression Analysis," *Technometrics*, 19, 1–13.

DURBIN, J. (1973), *Distribution Theory for Tests Based on the Sample Distribution Function*, Philadelphia: SIAM.

GAL, T. (1979), *Postoptimal Analyses, Parametric Programming and Related Topics*, New York: McGraw-Hill.

GNANADESIKAN, R. (1977), *Methods for Statistical Data Analysis of Multivariate Observations*, New York: John Wiley.

HETTMANSPERGER, T.P., and McKEAN, J.W. (1977), "A Robust Alternative Based on Ranks to Least Squares in Analyzing Linear Models," *Technometrics*, 19, 274–284.

HOGG, R.V. (1972), "Adaptive Robust Procedures," *Journal of the American Statistical Association*, 43, 1041–1067.

——— (1975), "Estimates of Percentile Regression Lines Using Salary Data," *Journal of the American Statistical Association*, 70, 56–59.

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

——— (1982), "Robust Tests For Heteroscedasticity Based on Regression Quantiles," *Econometrica*, 50, 43–61.

MOSTELLER, F. (1946), "On Some Useful 'Inefficient' Statistics," *Annals of Mathematical Statistics*, 17, 377–408.

PARZEN, E. (1979), "Nonparametric Statistical Data Modeling," *Journal of the American Statistical Association*, 74, 105–131.

PIERCE, D.A., and KOPECKY, K.J. (1979), "Testing Goodness of Fit for the Distribution of Errors in Regression Models," *Biometrika*, 66, 1–5.

RUPPERT, D., and CARROLL, R.J. (1980), "Trimmed Least Squares Estimation in the Linear Model," *Journal of the American Statistical Association*, 75, 828–838.

ROCKAFELLAR, R.T. (1970), *Convex Analysis*, Princeton: Princeton University Press.

STONE, C.J. (1975), "Adaptive Maximum Likleihood Estimators of a Location Parameter," *Annals of Statistics*, 3, 267–284.

TUKEY, J.W. (1965), "Which Part of the Sample Contains the Information," *Proceedings of the National Academy of Sciences*, 53, 127–134.

WEISBERG, S. "Comment," *Journal of the American Statistical Association*, 75, 28–31.

WHITE, H., and MacDONALD, G.M. (1980), "Some Large Sample Tests for Nonnormality in the Linear Regression Model," *Journal of the American Statistical Association*, 75, 16–28.