Econometrica, Vol. 46, No. 1 (January, 1978)

REGRESSION QUANTILES[1]

BY ROGER KOENKER AND GILBERT BASSETT, JR.

A simple minimization problem yielding the ordinary sample quantiles in the location model is shown to generalize naturally to the linear model generating a new class of statistics we term "regression quantiles." The estimator which minimizes the sum of absolute residuals is an important special case. Some equivariance properties and the joint asymptotic distribution of regression quantiles are established. These results permit a natural generalization to the linear model of certain well-known robust estimators of location.

Estimators are suggested, which have comparable efficiency to least squares for Gaussian linear models while substantially out-performing the least-squares estimator over a wide class of non-Gaussian error distributions.

1. INTRODUCTION

IN STATISTICAL PARLANCE the term robustness has come to connote a certain resilience of statistical procedures to deviations from the assumptions of hypothetical models. The paradigm may be briefly stated as follows.[2] The process generating observed data is thought to be approximately described by an element of some parametric class of models. The investigator seeks statistics, i.e., a mapping from the sample space to a parameter space, whose distribution will be as concentrated as possible near the true parameters—if the hypothesized model is correct. If however, as seems almost certain, the parametric model is not quite true, one would like to use estimators whose distributions were altered only slightly if the distribution of the observations were close, in some reasonable sense, to that of some member of the parametric class. In important special cases this modest robustness requirement is not met by estimators in common use.[3]

We consider the familiar problem of estimating a vector of unknown (regression) parameters, β, from a sample of independent observations on random variables $Y_1, Y_2, \ldots, Y_T$, distributed according to

$$P(Y_t < y) = F(y - x_t\beta) \qquad (t = 1, \ldots, T), \tag{1.1}$$

where $x_t: t = 1, \ldots, T$, denote rows of a known $(T \times K)$ design matrix *and the shape of F is not precisely known.*[4] If *F is* known precisely then it is frequently

[1] An early version of this paper [**5**] was presented at the Winter, 1974 meeting of the Econometric Society in San Francisco.

[2] A rigorous statement of this point of view on the robustness problem may be found in the work of Hampel [**16**, **17**, **18**].

[3] We need only mention the sample mean, an estimator *nonpareil* of location if the sample observations are generated by an independent and identically distributed Gaussian process. However, in any open neighborhood of a Gaussian distribution there exists distribution functions which would take the distribution of the sample mean arbitrarily far away from its distribution in the Gaussian case; cf., Hampel [**16**].

[4] This formulation restricts attention to what may be called the "robustness to distributional assumptions" problem, and leaves aside problems involving possible dependence among observations, non-linearities of functional form, etc.

possible to show that the maximum likelihood estimator or some one-step (M) approximant to it is efficient in the Cramèr-Rao sense. In particular, when F is known to be Gaussian (normal), Rao has shown that the least squares estimator, $\hat{\beta}$, is minimum variance in the class of unbiased estimators. Unfortunately the extreme sensitivity of the least squares estimator to modest amounts of outlier contamination makes it a very poor estimator in many non-Gaussian, especially long-tailed, situations. This paper introduces new classes of robust alternatives to the least squares estimator for the linear model. Estimators are suggested which have comparable efficiency to $\hat{\beta}$ for Gaussian models while substantially outperforming the least squares estimator over a wide class of non-Gaussian error distributions. The proposed estimators are analogues to linear combinations of sample quantiles in the location model.[5]

2. BACKGROUND AND MOTIVATION

The aphorism made famous by Poincaré and quoted by Cramér [**12**] that, "everyone believes in the [Gaussian] law of errors, the experimenters because they think it is a mathematical theorem, the mathematicians because they think it is an experimental fact," is still all too apt. This "dogma of normality" as Huber has called it, seems largely attributable to a kind of wishful thinking. As Box and Tiao put it,

> Classical statistical arguments lead us to treat assumptions as if they were in some way axiomatic and yet consideration will show that, in fact, they are conjectures which in practice may be expected to be more or less true .... If we assume normality, we can proceed with an 'objective' classical analysis ... however ... as seems to be inevitably the case in other problems as well as this one, our 'objectivity' is gained by pretending to knowledge we do not have ... [**11**, p. 419].

Following Haavelmo's [**15**] classic paper it is sometimes argued that the errors encountered in econometric models are known to be the sum of a large number of small and independent elementary errors, and therefore will be approximately Gaussian due to central limit theorem considerations. However, it is rather puzzling that investigators, who are generally loathe to adopt informative priors about the systematic structure of their models about which theoretical considerations and past empirical experience should provide substantive evidence, should feel themselves so well informed about the unobservable constituents of their model's unobservable errors to argue that they satisfy a Lindeberg condition! A few gross errors occurring with low probability can cause serious deviations from normality: to dismiss the possibility of these occurrences almost invariably requires a leap of Gaussian faith into the realm of pure speculation.

The need for robust alternatives to the sample mean (the least squares estimator in the location model) has been apparent since the eighteenth century. The median, other trimmed means, and more complicated linear combinations of order statistics were in common use especially in astronomical calculations, in the

[5] For obvious reasons the term "location model" is used to describe (1.1) when $x_t = 1$: $t = 1, \ldots, T$.

nineteenth century.[6] By 1821 Gauss had shown that the sample mean provided the "most probable" estimate of the location parameter from a random sample with probability density proportional to $e^{-x^2/2\sigma^2}$, but this result was *explicitly an ex post rationalization* for the use of the sample mean rather than a claim for the empirical validity of this particular error law.[7] In fact, it was noted by a number of authors that error distributions with longer tails than that of the Gaussian distribution were commonly observed. In such cases it appeared desirable to choose estimators which modified the sample mean by putting reduced weight on extreme observations.

There was a parallel early recognition of the need for robust alternatives to the least squares estimator for the linear model. Wild observations, or "outliers" as they came to be called, were more difficult to identify in such models and the fruitful notion from the location model of an ordering of sample observations had no simple analogue in the more complicated models. Many illustrious figures (Gauss, Laplace and Legendre, to name only three) suggested that the minimization of absolute deviations might be preferable to least squares when some sample observations are of dubious reliability.[8] In 1818 Laplace proved that in the simple model of bivariate regression through the origin, this least absolute error (LAE) estimator had smaller asymptotic variance than the least squares estimator if the error law of the model had variance, σ^2, and density at the median, $f(0)$, satisfying $[2f(0)]^{-1} < \sigma$. This result paved the way for investigations of the large sample theory of statistics based on sample quantiles in the location model.[9]

The elementary point that there may exist nonlinear, or for that matter—biased, estimators superior to least squares for the non-Gaussian linear model is a well kept secret in most of the econometrics literature. Statements of the Gauss-Markov theorem too often seem to imply that linearity in y and unbiasedness are added virtues of the least squares estimator instead of restrictions on the class of its potential competitors. Indeed one sometimes encounters the view that infinite variance of the errors constitutes the only possible rationale for seeking robust alternatives to least squares in the linear model. This is, emphatically, false. While least squares is obviously abysmal for distributions having infinite variance (having zero efficiency for the Cauchy for example) its gross inferiority to a variety of nonlinear estimators is by no means confined to distributions with infinite variance.

The wave of current interest[10] in the problem of robust estimation has focused primarily on the location model. While we cannot hope to do justice to the vast recent literature on this subject we briefly sketch the main lines of the developments which are most relevant to our work on the linear model.

[6] See the excellent survey of Harter [**19**] and the fascinating paper by Stigler [**38**].

[7] This point is made emphatically by Huber [**25**] in his survey paper on robust estimation.

[8] Boscovitch is generally credited with first proposing estimators which minimize the sum of absolute deviations. See Harter [**19**].

[9] Our paper [**6**] generalizes this result to the general linear model.

[10] Tukey has written of robust estimation as the third wave of statistical theory (see Hampel [**18**]); after parametric and non-parametric theory, a theory of "almost parametric models" is slowly emerging. See also the fundamental survey papers of Huber [**25**] and Hogg [**21**].

Mosteller [**32**], in 1946, proposed the use of a variety of so-called "inefficient statistics" based on a few sample quantiles as "quick and dirty" substitutes for more conventional estimators. It was found that estimators of this type could be constructed which were almost as efficient as the maximum likelihood estimators for most conventional parametric models. This approach was further developed by Bennett [**7**] for strictly parametric models, while Gastwirth [**14**] and others established that some estimators of this type had good efficiency properties for a wide variety of distributions. For example, the weighted average of the 1/3, 1/2, and 2/3 quantiles with weights .3, .4, .3 has asymptotic efficiency of nearly eighty per cent for the Gaussian, Laplace, logistic, *and* Cauchy distributions.[11] In contrast, the sample mean has asymptotic efficiency of one in the Gaussian case, but is less than half as efficient as the median for the Laplace distribution and has zero efficiency for the Cauchy distribution. Thus, although these "quick and dirty" estimators may be "inefficient statistics" for any particular parametric model, in practice they may actually be preferable to putatively "optimal" estimators, like the sample mean, if there is some uncertainty about the shape of the distribution generating the sample. Much recent work has been devoted to extending results of this type beyond a fixed number of quantiles to estimators which are linear combinations of order statistics. The asymptotic theory of such (L) estimators has achieved almost classical standing through the efforts of Bickel [**8**], Stigler [**39**] and others. The most common (L) estimator of location is the α-trimmed mean which is simply the mean of the sample after the proportion α of largest and smallest observation have been removed. This venerable[12] estimator was revived by Tukey in the late forties and has played an important role in recent work on robust estimation of location. Huber's now classic paper [**24**] on robust estimation of location solves for the least favorable (minimal Fisher information) distribution in the class of Gaussian contaminants—distributions of the form $F = (1-\varepsilon)\Phi + \varepsilon H$, where Φ denotes the standard Gaussian cumulative, H ranges over all symmetric cumulative distribution functions, and $0 \leq \varepsilon < 1$ is a fixed proportion of contamination: the least favorable distribution has exponential tails, and Gaussian center so the minimax estimator is quadratic in the center and linear in the tails. Asymptotically Huber's minimax estimator behaves like an ε-trimmed mean.

In order to provide some quantitative evidence on the performance of some alternative estimates of location we have abstracted a small subset of estimators and a small subset of distributions from those considered in the Princeton Robustness Study [**3**]. Table I gives Monte-Carlo variances for six selected estimators and five selected distributions. We have purposely chosen simple estimators which are nonadaptive. The table clearly illustrates that estimators of location are available which, while making a small sacrifice of efficiency to the mean at the Gaussian distribution, are greatly superior to the mean for non-Gaussian distributions. Only an unshakable *a priori* faith in the Gaussian "law of

[11] This estimator is due to Gastwirth [**14**] and bears his name in the Princeton Robustness Study [**3**].
[12] Stigler [**38**] and Harter [**19**] both discuss the historical background of the trimmed mean.

TABLE I

EMPIRICAL VARIANCES OF SOME ALTERNATIVE LOCATION ESTIMATORS[a]

(Sample Size 20)

Estimators	Distributions				
	Normal	10% 3σ[b]	10% 10σ[c]	Laplace	Cauchy
Mean	1.00	1.88	11.54	2.10	12,548.0
10% trimmed mean	1.06	1.31	1.46	1.60	7.3
25% trimmed mean	1.20	1.41	1.47	1.33	3.1
Median	1.50	1.70	1.80	1.37	2.9
Gastwirth[d]	1.23	1.45	1.51	1.35	3.1
Trimean[e]	1.15	1.37	1.48	1.43	3.9

[a] Abstracted from Exhibit 5 in Andrews, *et al.* [3].
[b] Gaussian Mixture: $.9\Phi(1)+.1\Phi(3)$.
[c] Gaussian Mixture: $.9\Phi(1)+.1\Phi(10)$.
[d] $\tilde{\beta}=.3\beta^*(1/3)+.4\beta^*(1/2)+.3\beta^*(2/3)$, where $\beta^*(\theta)$ is the θth sample quantile.
[e] $\tilde{\beta}=1/4\beta^*(1/4)+1/2\beta^*(1/2)+1/4\beta^*(3/4)$.

errors" would seem to justify selecting the sample mean.[13] In practice, of course, it is common to discard certain observations which seem to be deviant on the basis of a preliminary inspection of the data. This procedure obviously amounts to a rough and ready trimmed mean, but more formal procedures seem desirable. Such ad hoc empirical accommodations are even more problematic within the context of the linear model since deviant observations are more difficult to identify there. In Sections 4 through 6 below we propose new classes of robust alternatives to the least squares estimator for the linear model and show that members of these classes have, unlike the least squares estimator, high efficiency over a wide range of error distributions.

Huber's work on the location model has been extended to other estimators of the (M) maximum likelihood type by Relles [**35**], Huber [**26**], Andrews [**2**], Bickel [**10**], and others for the linear model. Another line of inquiry, based on analogues of rank procedures in the location model, has been extended to the linear model by Jurečková [**29**], Jaeckel [**28**], and others. Bickel [**9**] has suggested a third line of attack based on analogues of linear combinations of order statistics, (L) estimates, from the linear model. His estimates are one-step iterations from an ordering of observations based on some preliminary robust estimate of location like, for example, the LAE estimate.[14]

Our approach, although substantially different from that taken by Bickel, may also be viewed as an attempt to extend to the linear model the notions of systematic statistics and linear combinations of order statistics which have proven so fruitful in dealing with the robust estimation problem in the location model. We begin by introducing a natural generalization to the linear model of the concept of

[13] Based on the full study of 65 different estimators and ten distributions, Hampel concludes that the sample mean is the "clear candidate for . . . worst estimator of the study." Andrews, et al., [**3**, p. 239].

[14] A number of more sophisticated robust (M) estimators use this (LAE) estimator as an initial robust estimate and make a one-step Newton iteration. See, e.g., Hill and Holland [**20**].

sample quantiles from the location model.[15] The LAE estimator will be an important special case.[16]

3. REGRESSION QUANTILES

Our point of departure is an elementary definition of the sample quantiles which, by circumventing the usual reliance on an ordered set of sample observations, is readily extended to the linear model. As above, let $\{y_t: t=1,\ldots,T\}$ be a random sample on a random variable Y having distribution function F. Then the θth sample quantile, $0<\theta<1$, may be defined as any solution to the minimization problem:

$$\min_{b\in\mathbb{R}}\left[\sum_{t\in\{t:y_t\geq b\}}\theta|y_t-b|+\sum_{t\in\{t:y_t<b\}}(1-\theta)|y_t-b|\right].$$

The case of the median ($\theta=1/2$) is, of course, well known, but the general result has languished in the status of curiosum—appearing for example as an exercise in Ferguson [**13**].

Huber's [**26**] observation that outliers are difficult to identify in the regression context underlines the essential ambiguity involved in extending to the linear model the ordinary notions of sample quantiles based on an ordering of sample observations. A direct generalization of the minimization problem posed above resolves this ambiguity. Letting $\{x_t: t=1,\ldots,T\}$ denote a sequence of (row) K-vectors of a known design matrix, suppose $\{y_t: t=1,\ldots,T\}$ is a random sample on the regression process $u_t=y_t-x_t\beta$ having distribution function F. The *θth regression quantile*, $0<\theta<1$, is defined as any solution to the minimization problem:

$$\min_{b\in\mathbb{R}^K}\left[\sum_{t\in\{t:y_t\geq x_tb\}}\theta|y_t-x_tb|+\sum_{t\in\{t:y_t<x_tb\}}(1-\theta)|y_t-x_tb|\right].$$

In the location model ($K=1$, $x_t=1$, for all t) the two minimization problems coincide. The least absolute error estimator is the regression median, i.e., the regression quantile for $\theta=1/2$.

We now introduce some crucial notation and state some fundamental properties of elements, $\beta^*(\theta)$, of the solution sets $B^*(\theta)$ of the regression quantile minimization problem.

Let $\mathcal{T}=\{1,2,\ldots,T\}$ and $\mathcal{H}$ denote the set of K-element subsets of $\mathcal{T}$. Elements $h\in\mathcal{H}$ have relative complement, $\bar{h}=\mathcal{T}-h$, and both serve to partition

[15] Professor Hogg [**22, 23**] has recently proposed an alternative method of estimating "percentile hyperplanes" for the linear model which revives and generalizes the median regression methods of Mood and Brown. While the details of his method differ greatly from ours, the approaches are quite similar in spirit.

[16] The large sample theory of LAE is discussed in [**6**]; see also Taylor [**41**] for a recent survey, and Bassett [**4**]. Examples of empirical applications where LAE has performed extremely well in comparison with least squares in forecasting tests may be found in Meyer and Glauber [**31**] and Overson [**33**].

y and X. Thus, for example $y(h)$ denotes the (K)-vector with elements $\{y_t\colon t\in h\}$ while $X(\bar{h})$ denotes a $(T-K)\times K$ matrix with rows $\{x_t\colon t\in\bar{h}\}$. The (K)-dimensional vector of ones will be denoted by ι_K. Finally, let,

$$H=\{h\in\mathcal{H}|\text{rank } X(h)=K\}.$$

THEOREM 3.1: *If X has rank K then the set of regression quantiles, $B^*(\theta)$, has at least one element of the form,*

$$\beta^*(\theta)=X(h)^{-1}y(h)$$

for some $h\in H$. Moreover, $B^*(\theta)$, is the convex hull of all solutions having this form.

PROOF: This result follows immediately from the linear programming formulation of the defining minimization problem; see Appendix 1 and the recent paper by Abdelmalek [**1**] for details.

REMARK: Sample quantiles in the location model are either identified with a single order statistic from the observed sample, or, for example in the case of the median from an even sample, they are identified with a closed interval between two adjacent order statistics. Theorem 1 generalizes this feature to regression quantiles where normals to hyperplanes defined by subsets of K observations play the role of order statistics.

THEOREM 3.2: *If $\beta^*(\theta, y, X)\in B^*(\theta, y, X)$ then the following are elements of the solution of the specified transformed problems*:

(i) $\beta^*(\theta, \lambda y, X)=\lambda\beta^*(\theta, y, X), \qquad \lambda\in[0,\infty),$

(ii) $\beta^*(1-\theta, \lambda y, X)=\lambda\beta^*(\theta, y, X), \qquad \lambda\in(-\infty, 0],$

(iii) $\beta^*(\theta, y+X\gamma, X)=\beta^*(\theta, y, X)+\gamma, \qquad \gamma\in\mathbb{R}^K,$

(iv) $\beta^*(\theta, y, XA)=A^{-1}\beta^*(\theta, y, X), \qquad A_{K\times K}$ *nonsingular.*

PROOF: Let

$$\psi(b;\theta,y,X)=\sum_{\{t:y_t>x_tb\}}\theta|y_t-x_tb|+\sum_{\{t:y_t<x_tb\}}(1-\theta)|y_t-x_tb|$$

$$=\sum_{t=1}^{T}[\theta-1/2+1/2\,\text{sgn}\,(y_t-x_tb)][y_t-x_tb]$$

where sgn (u) takes values 1, 0, -1 as $u\gtreqless 0$. Now, note that

(i) $\lambda\psi(b;\theta,y,X)=\psi(\lambda b;\theta,\lambda y,X), \qquad \lambda\in[0,\infty),$

(ii) $-\lambda\psi(b;\theta,y,X)=\psi(\lambda b;1-\theta,\lambda y,X), \qquad \lambda\in(-\infty,0],$

(iii) $\psi(b;\theta,y,X)=\psi(b+\gamma;y+X\gamma,X), \qquad \gamma\in\mathbb{R}^K,$

(iv) $\psi(b;y,X)=\psi(A^{-1}b;y,XA), \qquad |A_{K\times K}|\neq 0.$

REMARK: Theorem 3.2 collects a number of equivariance properties of regression quantiles. Note that (i) and (ii) imply $\beta^*(1/2)$ is scale equivariant, (iii) states $\beta^*(\theta)$ is location (or "shift" or "regression") equivariant, and (iv) states that $\beta^*(\theta)$ is equivariant to reparameterization of design.

THEOREM 3.3: *If F is continuous then with probability one: $\beta^* = X(h)^{-1}y(h)$ is a unique solution to Problem (P) if and only if,*

$$(3.1)\qquad (\theta-1)\iota_K' < \sum_{t\in \bar{h}} [1/2 - 1/2\ sgn\ (y_t - x_t\beta^*) - \theta]x_t X(h)^{-1} < \theta\iota_K'.$$

PROOF: Consider the directional derivative of $\psi(b)$ in direction w,

$$(3.2)\qquad \psi'(b; w) = \sum_{t=1}^{T} [1/2 - 1/2\ \mathrm{sgn}^*\ (y_t - x_t b; -x_t w) - \theta]x_t w$$

where

$$\mathrm{sgn}^*\ (u; z) = \begin{cases} \mathrm{sgn}\ u & \text{if} \quad u \neq 0, \\ \mathrm{sgn}\ z & \text{if} \quad z = 0. \end{cases}$$

Since $\psi(b)$ is convex, it attains a unique minimum at β^* if and only if $\psi'(\beta^*; w) > 0$ for all $w \neq 0$. At $\beta^* = X(h)^{-1}y(h)$,

$$(3.3)\qquad \psi'(\beta^*; w) = \sum_{t\in h} [1/2 + \mathrm{sgn}\ (x_t w) - \theta]x_t w + \sum_{t\in \bar{h}} [1/2 - 1/2\ \mathrm{sgn}^*\ (y_t - x_t\beta^*; -x_t w) - \theta]x_t w.$$

Letting $v = X(h)w$, we have that $\psi'(\beta^*; w) > 0$ for all $w \neq 0$, if and only if,

$$(3.4)\qquad 0 < \sum_{k=1}^{K} [(1/2 - \theta)v_k + |v_k|] + \sum_{t\in \bar{h}} [1/2 - 1/2\ \mathrm{sgn}^*\ (y_t - x_t\beta^*; x_t X(h)^{-1}v) - \theta]x_t X(h)^{-1}v$$

for all $v \neq 0$. But this is equivalent to

$$(3.5)\qquad (\theta-1)\iota_K' < \sum_{t\in \bar{h}} [1/2 - 1/2\ \mathrm{sgn}^*\ (y_t - x_t\beta^*; x_t X(h)^{-1}v) - \theta]x_t X(h)^{-1} < \theta\iota_K',$$

for all $v \neq 0$. Finally F continuous implies that the events $[y_t - x_t X(h)^{-1}y(h) = 0, t \in h]$ have probability measure zero since they require $M > K$ observations to lie exactly on hyperplanes of dimension $K-1$. Thus (3.5) simplifies to (3.1).

REMARK: It is perhaps instructive to pause to consider the ordinary sample quantiles in the light of Theorem 3.3. If $x_t = 1$, $t = 1, \ldots, T$, so $H = \mathcal{T}$ and F is continuous, then Theorem 3.3 asserts that $\beta^*(\theta) = y(h)$ is a unique θth sample quantile if and only if,

$$(3.6)\qquad \theta - 1 < \sum_{t\in \bar{h}} [1/2 - 1/2\ \mathrm{sgn}\ (y_t - y(h)) - \theta] < \theta.$$

The expression in brackets takes the values of $-\theta$ if $y_t > y(h)$ and $1-\theta$ if $y_t < y(h)$, so (3.6) reduces to the requirement that the number of y_t's less that $y(h)$ be strictly between $T\theta - 1$ and $T\theta$. This in turn demands that $T\theta$ be nonintegral. The continuity of F removes the tiresome problem of "ties" in the location model and accomplishes the same task in the general linear model. It may be noted that in the absence of this degeneracy phenomena the condition for uniqueness is purely a design condition, reducing in the location model to the requirement that $T\theta$ be non-integral. This suggests that for any sequence $\{X_T\}$ of designs one should be able to extract a subsequence, or at worst some "perturbed" subsequence whose elements have unique solutions. An alternative approach which is frequently employed in the location model is to adopt some arbitrary rule to choose a single element from sets of quantiles when they occur. Either approach suffices to obtain the sequence of unique solutions considered in the next section dealing with the large sample distribution of regression quantiles.

THEOREM 3.4: *Let* $P(u^*(\theta))$, $N(u^*(\theta))$, *and* $Z(u^*(\theta))$ *denote the number of positive, negative, and zero elements in the vector* $u^*(\theta) = y - X\beta^*(\theta)$. *Then if* X *contains a column of ones,*

$$N(u^*) \leq T\theta \leq T - P(u^*) = N(u^*) + Z(u^*) \tag{3.7}$$

for every $\beta^*(\theta) \in B^*(\theta)$. *If* β^* *is unique, i.e.,* $\beta^* = B^*$, *then the inequalities hold strictly.*

PROOF: Partition the design so that $X = [\iota_. : \tilde{X}]$. By the argument used to obtain Theorem 3.3, $\beta^* \in B^*$ if and only if

$$\psi'(\beta^*; w) = \sum_{t=1}^{T} [1/2 - 1/2 \operatorname{sgn}^*(y_t - x_t\beta^*; -x_t w) - \theta] x_t w \geq 0, \tag{3.8}$$

for all $w \neq 0$. So in particular (3.8) must hold for $w^+ = (1, 0, \ldots, 0)$ and $w^- = (-1, 0, \ldots, 0)$ in $\mathbb{R}^K$. Since $x_t w^+ = 1$ and $x_t w^- = -1$ for all x_t, (3.8) implies

$$\sum_{t=1}^{T} \pm [1/2 - 1/2 \operatorname{sgn}^*(y_t - x_t\beta^*; \mp 1) - \theta] > 0 \tag{3.9}$$

which is equivalent to the two conditions

$$-\theta P + (1-\theta)N + (1-\theta)Z > 0,$$

$$-\theta P + (1-\theta)N - Z < 0,$$

which in turn are equivalent to (3.7). If $\beta^*(\theta)$ is unique then all inequalities are strict.

REMARK: Note that if F is continuous then $Z(u^*) = K$ with probability one, so there are at least $T\theta$ observations below the θth regression quantile hyperplane and at most $T\theta + K$ observations above it.

THEOREM 3.5: *If* $\beta^*(\theta) \in B^*(\theta, y, X)$, *then* $\beta^*(\theta) \in B^*(\theta, X\beta^* + Du^*, X)$ *where* $u^* = y - X\beta^*$ *and* D *is any* $T \times T$ *diagonal matrix with nonnegative elements.*

PROOF: $\beta^* \in B^*(\theta, y, X)$ implies

$$\sum_{t=1}^{T} [1/2 - 1/2 \operatorname{sgn}^*(y_t - x_t\beta^*; -x_t w) - \theta] x_t w \geq 0$$

for all $w \neq 0$. Note that

$$\begin{aligned}
&[1/2 - 1/2 \operatorname{sgn}^*(x_t\beta^* + d_t(y_t - x_t\beta^*) - x_t\beta^*; -x_t w) - \theta] x_t w \\
&\quad = (1/2 - \theta) x_t w - 1/2 \operatorname{sgn}^*(d_t(y_t - x_t\beta^*); -x_t w) x_t w \\
&\quad \geq (1/2 - \theta) x_t w - 1/2 \operatorname{sgn}^*(y_t - x_t\beta^*; -x_t w) x_t w
\end{aligned}$$

for $d_t \geq 0$ and the result follows.

REMARK: This result has a simple geometric interpretation. Imagine a scatter of sample observations in $\mathbb{R}$ with the θth regression quantile line slicing through the scatter. Now consider the effect (on the position of the θth RQ line) of moving observations up or down in the scatter. The result states that as long as these movements leave observations on the same side of the original line its position is unaffected. This property is obvious in the location model, sample quantile context, but its generalization is perhaps somewhat less obvious.

We conclude this section with an illustration in Table II of the regression quantiles for all $\theta \in (0, 1)$ in a simple bivariate model with five observations. Note that for $\theta \in \{7/22, 1/2, 3/4\}$ the problem admits multiple solutions.

TABLE II

	$0 < \theta \leq 7/22$	$7/22 \leq \theta \leq 1/2$	$1/2 \leq \theta \leq 3/4$	$3/4 \leq \theta < 1$
$\beta_1^*(\theta)$	6/7	21/8	13/6	17/3
$\beta_2^*(\theta)$	4/7	3/8	5/6	1/3

4. THE ASYMPTOTIC DISTRIBUTION THEORY OF REGRESSION QUANTILES

The following well known result concerning sample quantiles in the location model is due to Mosteller [**32**].

THEOREM 4.1: *Let* $\{\xi_T^*(\theta_1), \ldots, \xi_T^*(\theta_M)\}$ *with* $0 < \theta_1 < \theta_2 < \ldots < \theta_M < 1$, *denote a sequence of unique sample quantiles from random samples of size* T *from a population with inverse distribution function* $\xi(\theta) = F^{-1}(\theta)$. *If* F *is continuous and has continuous and positive density,* f, *at* $\xi(\theta_i)$, $i = 1, \ldots, M$, *then,*

$$\sqrt{T}[\xi_T^*(\theta_1) - \xi(\theta_1), \ldots, \xi_T^*(\theta_M) - \xi(\theta_M)]$$

converges in distribution to an (M)*-variate Gaussian random vector with mean,* 0, *and covariance matrix* $\Omega(\theta_1, \ldots, \theta_M; F)$ *with typical element,*

$$\omega_{ij} = \frac{\theta_i(1 - \theta_j)}{f(\xi(\theta_i)) f(\xi(\theta_j))}, \qquad i \leq j.$$

This theorem provides the foundation for a large-sample theory of so-called "systematic statistics"—estimators which are linear combinations of a few sample quantiles. The median is the most important special case having asymptotic variance $[2f(\xi(1/2))]^{-2}$. As we have noted above, this value will be less than the variance of the mean for a large class of long-tailed distributions. The asymptotic variance of the Gastwirth or trimean estimator of Table I can be easily calculated in a similar manner from Theorem 4.1 for arbitrary distributions. We merely require the evaluation of the density function of the distribution F at specified quantiles.

The analogy between sample quantiles in the location model and regression quantiles in the linear model is considerably strengthened by the striking resemblance in their asymptotic behavior. This is made explicit in the following theorem which plays a central role in the theory developed in the remainder of the paper.

THEOREM 4.2: *Let* $\{\beta_T^*(\theta_1), \beta_T^*(\theta_2), \ldots, \beta_T^*(\theta_M)\}$ *with* $0<\theta_1<\theta_2<\ldots<\theta_M<1$ *denote a sequence of unique regression quantiles from model* (1.1). *Let* $\xi(\theta)=F^{-1}(\theta)$, $\boldsymbol{\xi}(\theta)=(\xi(\theta), 0, \ldots, 0)\in\mathbb{R}^K$ *and* $\boldsymbol{\xi}_T^*(\theta)=\beta_T^*(\theta)-\beta$. *Assume*:

(i) *F is continuous and has continuous and positive density, f, at $\xi(\theta_i)$, $i=1,\ldots,M$, and*

(ii) $x_{1t}=1$: $t=1,2,\ldots$ *and* $\lim_{T\to\infty} T^{-1}X'X=Q$, *a positive definite matrix.*

Then,

$$\sqrt{T}[\boldsymbol{\xi}_T^*(\theta_1)-\boldsymbol{\xi}(\theta_1), \ldots, \boldsymbol{\xi}_T^*(\theta_M)-\boldsymbol{\xi}(\theta_M)]$$

converges in distribution to an (MK)-variate Gaussian random vector with mean 0 *and covariance matrix* $\Omega(\theta_1,\ldots,\theta_M; F)\otimes Q^{-1}$, *where Ω is the covariance matrix of the corresponding M ordinary sample quantiles from random samples from distribution F.*

PROOF: The proof for $M=1$ is a trivial modification of the proof given in [**6**] for the special case $\beta^*(1/2)$. The case $M=2$ is treated below explicitly, but the generalization to arbitrary M is obvious, albeit somewhat tedious. Consider the probability element,

$$(4.1)\qquad g_T(\delta_1,\delta_2)\,d\delta_1\cdot d\delta_2=\text{pr}\,[\delta_i<\sqrt{T}(\boldsymbol{\xi}_T^*(\theta_i)-\boldsymbol{\xi}(\theta_i))<\delta_i+d\delta_i,\ i=1,2].$$

We demonstrate that the joint density function $g_T(\delta_1,\delta_2)$ converges to a specified Gaussian density and Scheffé's theorem on convergence of densities completes the proof.

By Theorem 3.2,

$$(4.2)\qquad \beta^*(\theta,u,X)=\beta^*(\theta,y,X)-\beta$$

where $u=y-X\beta$ is a vector of T independent realizations from F. So, by

Theorem 3.1 and the uniqueness of $\{\beta_T^*(\theta_1), \beta_T^*(\theta_2)\}$ there must exist index sets h_1, h_2 such that

$$\delta_i < \sqrt{T}\left(X(h_i)^{-1}u(h_i) - \xi(\theta_i)\right) < \delta_i + d\delta_i, \qquad i = 1, 2, \tag{4.3}$$

and

$$\begin{aligned}(\theta_i - 1)\iota_K' &< \sum_{t \in \bar{h}_i} \left[1/2 - 1/2 \operatorname{sgn}\left(u_t - x_t X(h_i)^{-1}u(h_i)\right) - \theta_i\right] x_t X(h_i)^{-1} \\ &< \theta_i \iota_K', \qquad i = 1, 2.\end{aligned} \tag{4.4}$$

Since the first column of X is the unit vector, the events of (4.3) may be written as

$$u_t \in \left(\xi(\theta_i) + T^{-1/2}x_t\delta_i,\ \xi(\theta_i) + T^{-1/2}x_t(\delta_i + d\delta_i)\right) \tag{4.5}$$

for $t \in h_i$, $i = 1, 2$. By hypothesis $\xi(\theta_1) < \xi(\theta_2)$ so (4.5) implies that for sufficiently large T, the solution index sets must be disjoint. Thus proceeding as in [**6**], we may write,

$$\begin{aligned}g_T(\delta_1, \delta_2)\, d\delta_1 \cdot d\delta_2 = \sum_{\substack{h_1 \in H \\ h_1 \cap h_2 = \phi}} \sum_{h_2 \in H} &\operatorname{pr}\left[\delta_i < \sqrt{T}\left(X(h_i)^{-1}u(h_i)\right.\right. \\ &\left.\left. - \xi(\theta_i)\right) < \delta_i + d\delta_i,\ i = 1, 2\right] \\ &\cdot \operatorname{pr}\left[(\theta_i - 1)\iota_K' < \sum_{t \in \bar{h}_i} z_{ti} < \theta_i \iota_K',\ i = 1, 2\right]\end{aligned} \tag{4.6}$$

where

$$z_{ti} = \left[1/2 - 1/2 \operatorname{sgn}\left(u_t - \xi(\theta_i) - T^{-1/2}x_t\delta_i\right) - \theta_i\right] x_t X(h_i)^{-1}.$$

But since $h_1 \cap h_2 = \varnothing$ and the u's are i.i.d.,

$$\begin{aligned}&\operatorname{pr}\left[\delta_i < \sqrt{T}\left(X(h_i)^{-1}u(h_i) - \xi(\theta_i)\right) < \delta_i + d\delta_i,\ i = 1, 2\right] \\ &\quad = T^{-K}|X(h_1)||X(h_2)| \prod_{t \in h_1} f\left(\xi(\theta_1) + T^{-1/2}x_t\delta_1\right) \prod_{t \in h_2} f\left(\xi(\theta_2)\right. \\ &\qquad \left. + T^{-1/2}x_t\delta_2\right) d\delta_1 \cdot d\delta_2.\end{aligned} \tag{4.7}$$

Now note that the random variables,

$$z_{ti} = \begin{cases} -\theta_i x_t X(h_i)^{-1} \\ (1 - \theta_i)x_t X(h_i)^{-1} \end{cases} \quad \text{with probability} \quad \begin{matrix} 1 - F\left(\xi(\theta_i) + T^{-1/2}x_t\delta_i\right), \\ F\left(\xi(\theta_i) + T^{-1/2}x_t\delta_i\right), \end{matrix} \tag{4.8}$$

for $t \in \bar{h}_1 \cap \bar{h}_2$. And by (4.5),

$$\begin{aligned}z_{t1} &= (1 - \theta_1)x_t X(h_1)^{-1}, \qquad t \in h_2, \\ z_{t2} &= -\theta_2 x_t X(h_2)^{-1}, \qquad t \in h_1,\end{aligned}$$

for sufficiently large T.

A bit of calculation reveals that the stabilized sums

$$Z_{Ti} = T^{-1/2} \sum_{t \in \bar{h}_i \cap \bar{h}_2} z_{ti}, \qquad i = 1, 2,$$

converge to a $(2K)$-variate Gaussian random vector with mean $(f(\xi(\theta_1))\delta_1' QX'(h_1)^{-1}, f(\xi(\theta_2))\delta_2' QX'(h_2)^{-1})$ and covariance matrix

$$\begin{bmatrix} \theta_1(1-\theta_1)X'(h_1)^{-1}QX(h_1)^{-1} & \theta_1(1-\theta_2)X'(h_1)^{-1}QX(h_2)^{-1} \\ \theta_1(1-\theta_2)X'(h_2)^{-1}QX(h_1)^{-1} & \theta_2(1-\theta_2)X'(h_2)^{-1}QX(h_2)^{-1} \end{bmatrix}.$$

This can be verified by expanding F around the population quantiles $\xi(\theta_i)$: $i = 1, 2$, noting that $T^{-1} \Sigma_{t \in \bar{h}_1 \cap \bar{h}_2} x_t' x_t \to Q$, and

$$T^{-1/2} \max_{k,t<T} |x_{kt}| = o(1);$$

see Malinvaud [**30**, 226–27]. It then follows that

$$\begin{aligned} (4.9) \qquad & T^K \operatorname{pr}[-T^{-1/2}(\theta_i - 1)\iota_K' < Z_{Ti} < T^{-1/2}\theta_i \iota_K', \ i = 1, 2] \\ & = (2\pi)^{-K} |\Theta|^{-K/2} |X'(h_1)^{-1}QX(h_1)^{-1}|^{-1/2} |X'(h_2)^{-1}QX(h_2)^{-1}|^{-1/2} \\ & \quad \cdot \exp\{-\tfrac{1}{2}\delta'[\Omega^{-1} \otimes Q]\delta\} + o(1), \end{aligned}$$

where $\Omega = \Omega(\theta_1, \theta_2; F)$, $\delta = (\delta_1, \delta_2)'$, and

$$\Theta = \begin{bmatrix} \theta_1(1-\theta_1) & \theta_1(1-\theta_2) \\ \theta_1(1-\theta_2) & \theta_2(1-\theta_2) \end{bmatrix}.$$

The continuity of the density at $\xi(\theta_1)$ and $\xi(\theta_2)$ implies

$$(4.10) \qquad \prod_{t \in h_i} f(\xi(\theta_i) + T^{-1/2} x_t \delta_i) = f(\xi(\theta_i))^K + o(1), \qquad i = 1, 2.$$

Substituting (4.10) and (4.9) into (4.6), we have

$$\begin{aligned} (4.11) \qquad g_T(\delta) = & \sum_{\substack{h_1 \in H \\ h_1 \cap h_2 = \varnothing}} \sum_{h_2 \in H} T^{-2K} |X(h_1)|^2 |X(h_2)|^2 f(\xi(\theta_1))^K f(\xi(\theta_2))^K |\Theta|^{-K/2} |Q|^{-1} \\ & \cdot \exp\{-\tfrac{1}{2}\delta'[\Omega^{-1} \otimes Q]\delta\} + o(1) \sum_{h_1} \sum_{h_2} T^{-2K} |X(h_1)|^2 |X(h_2)|^2. \end{aligned}$$

But,

$$(4.12) \qquad T^{-2K} \sum \sum |X(h_1)|^2 |X(h_2)|^2 = T^{-2K} |X'X|^2 = |T^{-1}X'X|^2$$

converges to $|Q|^2$; see Rao [**34**, p. 32]. So simplifying (4.11), we have

$$g_T(\delta) \to (2\pi)^{-K} |\Omega^{-1} \otimes Q|^{1/2} \exp\{-\tfrac{1}{2}\delta'[\Omega^{-1} \otimes Q]\delta\}.$$

And finally, Scheffe's [**37**] theorem on convergence of densities yields the desired conclusion.

REMARK: An extremely important special case is the regression median, or least absolute error estimator, $\beta^*(1/2)$. Without loss of generality β may be located so that $F(0)=1/2$, so $\xi(1/2)=0$. Then the asymptotic distribution of the random variable $\sqrt{T}(\beta^*(1/2)-\beta)$ is K-variate Gaussian with mean zero and covariance matrix $[2f(0)]^{-2}Q^{-1}$. See our detailed treatment of this special case in [**6**]. Thus, *the regression median is seen to be more efficient than the least squares estimator in the linear model for any distribution for which the median is more efficient than the mean in the location model.* While such a result has been the subject of conjecture among a number of authors we believe ours to be its first formal statement.[17]

The remarkable parallelism between the asymptotic behavior of ordinary sample quantiles in the location model and regression quantiles in the linear model suggests a straightforward extension of the large sample theory of so-called systematic statistics from the location model to the linear model. This is made explicit in the following result.

THEOREM 4.3: *Let* $\pi(\theta)=(\pi(\theta_1),\ldots,\pi(\theta_M))'$ *be a discrete, symmetric probability measure on* $(0,1)$ *concentrating mass on* $\{\theta_i: i=1,\ldots,M | 0<\theta_1<\theta_2<\ldots<\theta_M<1\}$. *Suppose* $F(0)=1/2$, *so* $\xi(1/2)\equiv F^{-1}(1/2)=0$, *and conditions* (*i*) *and* (*ii*) *of Theorem* 4.2 *hold. Then*

$$\tilde{\beta}_T(\pi)=\sum \pi(\theta)\beta_T^*(\theta)$$

is invariant to location, scale and reparameterization of design in the sense of Theorem 3.2, *and* $\sqrt{T}(\tilde{\beta}_T(\pi)-\beta)$ *converges in distribution to a* (K)*-variate Gaussian random vector with mean zero and covariance matrix* $\pi'\Omega\pi\cdot Q^{-1}$.

PROOF: Immediate from Theorems 3.2 and 4.2.

REMARK: Obvious examples of weight functions $\pi(\theta)$, in addition to the regression median example $(\pi(1/2)=1)$ already discussed, are the Gastwirth: $(\pi(1/3),\pi(1/2),\pi(2/3))=(.3,.4,.3)$ and the trimean: $(\pi(1/4),\pi(1/2),\pi(3/4))=(1/4,1/2,1/4)$. When F is non-Gaussian, weight functions $\pi(\theta)$ are readily found which make $\pi'\Omega\pi$ smaller than the variance of F, i.e., yielding estimators $\tilde{\beta}(\pi)$ having superior asymptotic efficiency to the least squares estimator. Furthermore, as we have noted above, many estimators of the form $\tilde{\beta}(\pi)$ like Gastwirth's and the trimean, have high efficiency over a large class of distributions. The efficiency of the mean and its least squares analogue, $\hat{\beta}$, in the

[17] In finite samples the form of the design matrix obviously has some impact on the sampling distribution of the estimates, and is likely therefore to affect the sample size required to achieve a given degree of convergence to the asymptotic distribution. Thus there is no contradiction between our general result and the finding of Rosenberg and Carlson [**36**] that different designs have somewhat different sampling distributions in small sample Monte-Carlo experiments. Rosenberg and Carlson's results do seem to suggest the kurtosis of the explanatory variables produces a slower convergence. These problems of robustness of design are only now beginning to receive some much deserved attention. See, e.g., the pioneering work of Huber [**27**] on minimax designs.

linear model, on the contrary, deteriorates rapidly as the error distribution tends away from the Gaussian toward longer tailed distributions.

The contrast between the ordinary least squares estimator and linear combination of regression quantiles, $\tilde{\beta}(\pi)$-type, estimators for the linear model is highlighted by writing the former in the partitioning notation of Section 3 as

$$\hat{\beta}=\sum w(h)\hat{\beta}(h),$$

where $w(h)=|X(h)|^2/\Sigma\,|X(h)|^2$, summations are over $h \in \mathscr{H}$, and $\hat{\beta}(h)$ satisfies the normal equations,

$$X(h)\hat{\beta}(h)=y(h).$$

This remarkable formulation of the least squares estimate as a weighted average of all possible coefficient vectors defined by subsamples of K observations was apparently first noted by M. Subrahmanyan [**40**]. Least squares places positive weight on *all* $\hat{\beta}(h)$ with nonzero design determinant; the $\tilde{\beta}(\pi)$ estimators place positive weight on only a few select $\hat{\beta}(h)$ vectors. The simple connection between the sample mean and the sample quantiles in the location model thus generalizes in a somewhat more sophisticated form to the connection between the least squares estimator $\hat{\beta}$ and the class of $\tilde{\beta}(\pi)$ estimators for the linear model.

In a very stimulating critique of least squares estimation procedures, Tukey [**42**] suggests that efforts should be made to find estimators which modify the least squares method, reducing its notorious sensitivity to outlying observations, but preserving its essentially good qualities. In this spirit we discuss a simple modification of least squares which employs regression quantiles.

An obvious analogue to the trimmed mean of the location model is suggested by regression quantiles. Let α be the desired trimming proportion. Calculate $\beta^*(\alpha)$ and $\beta^*(1-\alpha)$ for the sample. By Theorem 3.4 there will be approximately $T\alpha$ observations below the former and above the latter. These observations are set aside, and the remaining observations are subjected to a least squares fit. Since the β^* vectors are consistent estimators of the α and $(1-\alpha)$ population quantile hyperplanes, least squares on the "censored" sample is (asymptotically) like sampling from the distribution F truncated at $F^{-1}(\alpha)$ and $F^{-1}(1-\alpha)$. This estimator, say $\hat{\beta}_\alpha$, is location and scale invariant, and will be asymptotically K-variate Gaussian with covariance matrix $\sigma^2(\alpha, F)Q^{-1}$ where $\sigma^2(\alpha, F)$ denotes the asymptotic variance of the corresponding α-trimmed mean from a population with distribution F. This class offers promising robustness properties and should behave similarly to Huber's (M) estimates. Computation of these trimmed least squares estimates merely requires the solutions of two simple linear programming problems in addition to the usual least squares computation. The choice of the trimming proportion α will depend, obviously, on one's confidence in the quality of available data. With data of very dubious reliability one may want to use the ultimate trimmed least squares estimator with $\alpha = 1/2$, the regression median or LAE estimator. For somewhat better data, a trimming proportion of five to ten per cent is probably preferable.

5. CONCLUSION

We argue that the conventional least squares estimator may be seriously deficient in linear models with non-Gaussian errors. In the absence of a well-specified prior on the set of plausible distribution functions it is useful, following Huber [**25**] and others, to view the problem of estimation in terms of insurance. It seems reasonable to pay a small premium in the form of sacrificed efficiency "at the Gaussian distribution" (if that is the hypothesized parametric model), in order to achieve a substantial improvement over least squares in the event of a non-Gaussian situation.

We introduce a new class of statistics for the linear model which we have called "regression quantiles" since they appear to have analogous properties to the ordinary sample quantiles of the location model. Natural generalizations based on regression quantiles of linear combinations of sample quantiles and trimmed means which appear to have promising robustness properties are then proposed for the general linear model.

Bell Laboratories
and
University of Illinois—Chicago Circle

Manuscript received February, 1975; revision received January, 1977.

APPENDIX

COMPUTATIONAL ASPECTS OF REGRESSION QUANTILES[18]

The regression quantile minimization problem posed above in Section 3 is equivalent to the linear program

$$\text{(P)} \qquad \min\,[\theta\iota' r^+ + (1-\theta)\iota' r^-]$$

subject to

$$y = Xb + r^+ - r^-,$$
$$(b, r^+, r^-) \in \mathbb{R}^K \times \mathbb{R}_+^{2T},$$

where $\iota' = (1, 1, \ldots, 1)$, a T vector of ones. The dual to (P) is the bounded variables problem,

$$\text{(D)} \qquad \max\,[y'd]$$

subject to

$$X'd = 0,$$
$$d \in [\theta - 1, \theta]^T,$$

where $[\theta-1, \theta]^T$ denotes the T-fold Cartesian product of the closed interval $[\theta-1, \theta]$. It proves convenient to make one more minor adjustment. Translate the dual variables setting $\Delta = d + 1 - \theta$, so we have

$$\text{(D')} \qquad \max\,[y'\Delta]$$

[18] For further details on the important special case, $\theta = 1/2$, see [**1**, **43**].

subject to

$$X'\Delta = (1-\theta)X'\iota,$$

$$\Delta \in [0, 1]^T.$$

This translated dual formulation proves to be extremely convenient for computational purposes. Standard linear programming algorithms provide quite efficient algorithms for such bounded variables problems. So-called "parametric variation of the right-hand-side" of (D') permits economical solution for many, or indeed all, $\theta \in [0, 1]$.

REFERENCES

[1] ADELMALEK, N. N.: "On the Discrete Linear L_1 Approximation and L_1 Solutions of Overdetermined Linear Equations," *Journal of Approximation Theory*, 11 (1974), 38–53.

[2] ANDREWS, D. F.: "A Robust Method for Multiple Linear Regression," *Technometrics*, 16 (1974), 523–31.

[3] ANDREWS, D. F. ET AL.: *Robust Estimates of Location*. Princeton: Princeton University Press, 1972.

[4] BASSETT, G. W.: "Some Properties of the Least Absolute Error Estimator," unpublished Ph.D. Dissertation, Department of Economics, University of Michigan, 1972.

[5] BASSETT, G., AND R. KOENKER: "Generalized Sample Quantile Estimators for the Linear Model," presented at the Winter, 1974, meeting of the Econometric Society, San Francisco.

[6] ———: "The Asymptotic Distribution of the Least Absolute Error Estimator," manuscript, 1976.

[7] BENNETT, C. A.: "Asymptotic Properties of Ideal Linear Estimators," unpublished Ph.D. Dissertation, Department of Mathematics, University of Michigan, 1952.

[8] BICKEL, P. J.: "Some Contributions to the Theory of Order Statistics," *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics*, Vol. 1, 575–91.

[9] ———: "On Some Analogues to Linear Combinations of Order Statistics in the Linear Model," *Annals of Statistics*, 1 (1973), 597–616.

[10] ———: "One-Step-Huber Estimates in the Linear Model," *Journal of the American Statistical Association*, 70 (1975), 428–34.

[11] BOX, G. E. P., AND G. C. TIAO: "A Further Look at Robustness via Bayes' Theorem," *Biometrika*, 49 (1962), 419–32.

[12] CRAMÉR, *Mathematical Methods of Statistics*. Princeton: Princeton University Press, 1946.

[13] FERGUSON, T. S.: *Mathematical Statistics: A Decision Theoretic Approach*. New York: Academic Press, 1967.

[14] GASTWIRTH, J. L.: "On Robust Procedures," *Journal of the American Statistical Association*, 65 (1966), 946–73.

[15] HAAVELMO, T.: "The Probability Approach in Econometrics," *Econometrica*, supplement 12 (1944).

[16] HAMPEL, F. R.: "Contributions to Theory of Robust Estimation," unpublished Ph.D. Dissertation, Department of Statistics, University of California, Berkeley, 1968.

[17] ———: "A General Qualitative Definition of Robustness," *Annals of Mathematical Statistics*, 42 (1971), 1887–96.

[18] ———: "Robust Estimation: A Condensed Partial Survey," *Zeitschrift fur Wahrscheinlichkeitstheorie und verwandte Gebiete*, 27 (1973), 87–104.

[19] HARTER, H. L.: "The Method of Least Squares and Some Alternatives," (in five parts), *International Statistical Review*, 42 (1974), 147–74, 235–64, 43 (1975), 1–44, 125–90, 269–78.

[20] HILL, R. W., AND P. W. HOLLAND: "A Monte-Carlo Study of Two Robust Alternatives to Least Squares Regression Estimation," NBER Working Paper, No. 58, Cambridge, Mass., 1974.

[21] HOGG, R. V.: "Adaptive Robust Procedures: A Partial Review and Some Suggestions for Future Applications and Theory," *Journal of the American Statistical Association*, 43 (1972), 1041–67.

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

[23] HOGG, R. V., AND R. H. RANDLES: "Adaptive Distribution Free Regression Methods and their Applications," *Technometrics*, 17 (1975), 399–407.

[24] HUBER, P. J.: "Robust Estimation of a Location Parameter," *Annals of Mathematical Statistics*, 35 (1964), 73–101.
[25] ———: "Robust Statistics: A Review," *Annals of Mathematical Statistics*, 43 (1972), 1041–67.
[26] ———: "Robust Regression, Asymptotics, Conjectures and Monte-Carlo," *Annals of Statistics*, 1 (1973), 799–821.
[27] ———: "Robustness and Designs," in *A Survey of Statistical Design and Linear Models*, ed. by J. Srivastava. Amsterdam: North-Holland, 1975.
[28] JAECKEL, L. A.: "Estimating Regression Coefficients by Minimizing the Dispersion of the Residuals," *Annals of Mathematical Statistics*, 43 (1972), 1449–58.
[29] JUREČKOVÁ, J.: "Nonparametric Estimate of Regression Coefficients," *Annals of Mathematical Statistics*, 42 (1971), 1328–38.
[30] MALINVAUD, E.: *Statistical Methods of Econometrics*, 2nd ed. Amsterdam: North-Holland, 1970.
[31] MEYER, J. R., AND R. R. GLAUBER: *Investment Decision, Economic Forecasting and Public Policy*. Cambridge: Harvard Business School Press, 1964.
[32] MOSTELLER, F.: "On Some Useful "Inefficient" Statistics," *Annals of Mathematical Statistics*, 17 (1946), 377–408.
[33] OVERSON, R. M.: "Regression Parameter Estimation by Minimizing the Sum of Absolute Errors," unpublished Ph.D. Dissertation, Department of Economics, Harvard University, 1968.
[34] RAO, C. R.: *Linear Statistical Inference and Its Applications*, 2nd ed. New York: Wiley, 1973.
[35] RELLES, D. A.: "Robust Regression by Modified Least Squares," unpublished Ph.D. Dissertation, Yale University, 1968.
[36] ROSENBERG, B., AND D. CARLSON: "The Sampling Distribution of Least Absolute Residuals Regression Estimates," Working Paper No. IP-164, Institute of Business and Economic Research, University of California, Berkeley, 1971.
[37] SCHEFFÉ, H.: "A Useful Convergence Theorem for Probability Distributions," *Annals of Mathematical Statistics*, 18 (1947), 434–58.
[38] STIGLER, S. M.: "Simon Newcomb, Percy Daniell, and the History of Robust Estimation 1885–1920," *Journal of the American Statistical Association*, 68 (1973), 872–79.
[39] ———: "Linear Functions of Order Statistics with Smooth Weight Functions," *Annals of Statistics*, 2 (1974), 676–693.
[40] SUBRAHMANYAN, M.: "A Property of Simple Least Squares Estimates," *Sankhyā (B)*, 34 (1972), 355–56.
[41] TAYLOR, L. D.: "Estimation by Minimizing the Sum of Absolute Errors," in *Frontiers in Econometrics*, ed. by P. Zarembka. New York: Academic Press, 1973.
[42] TUKEY, J. W.: "Instead of Gauss-Markov Least Squares, What?" in *Applied Statistics*, ed. by R. P. Gupta. Amsterdam: North-Holland, 1975.
[43] WAGNER, H.: "Linear Programming Techniques for Regression Analysis," *Journal of the American Statistical Association*, 56 (1959), 206–12.