Faculty Working Papers

GENERALIZED SAMPLE QUANTILE ESTIMATORS FOR THE LINEAR MODEL

Gilbert Bassett and Roger Koenker

#225

College of Commerce and Business Administration
University of Illinois at Urbana-Champaign

FACULTY WORKING PAPERS

College of Commerce and Business Administration

University of Illinois at Urbana-Champaign

December 12, 1974

GENERALIZED SAMPLE QUANTILE ESTIMATORS FOR THE LINEAR MODEL

Gilbert Bassett and Roger Koenker

#225

GENERALIZED SAMPLE QUANTILE ESTIMATORS

FOR THE LINEAR MODEL

Gilbert Bassett

and

Roger Koenker*

To be presented at

the Meetings of the Econometric Society

San Francisco

December, 1974

*University of Illinois at Chicago-Circle and Urbana-Champaign respectively. The second author was on the staff of the Office of Research, U.S. Bureau of Labor Statistics, while the first draft of the paper was being written. We wish to thank Robert Knapp for his help with the argument of Appendix II although he should bear no responsibility for errors or malapropisms there or in the remainder of the paper.

I. Introduction

The precision of the sample mean, compared to certain other location estimators, is notoriously sensitive to the shape of the underlying population distribution function. Furthermore, some estimators which are considerably less sensitive than the mean to the "outliers" of heavy tailed distributions have been found to be comparable in precision to the mean for near-normal samples.[1] This robustness to distributional assumptions has led to an intensive search for similarly robust alternatives to least squares for the general linear model.[2] Since many robust estimators of location are linear combinations of order statistics, some attention has naturally focused on the problem of generalizing the notions of order statistics and sample quantiles to the linear model.[3]

This paper proposes such a generalized sample quantile (GSQ) estimator. We contend that it generalizes the sample quantile in much the same way that least-squares regression generalizes the sample mean. The θth GSQ hyperplane is estimated by minimizing a weighted sum of absolute deviations from the hyperplane. Positive deviations are assigned weight θ; negative deviations--weight $(1-\theta)$.

The GSQ minimization problem may be solved by standard linear programming techniques. In the location sub-model the GSQ estimator reduces to the conventional sample quantile. Important invariance properties of sample quantiles readily generalize to the GSQ. The asymptotic distribution of the coefficients of the GSQ hyperplane are, with very weak restrictions on the sample design and the underlying error distribution of the model, asymptotically normal with variance

covariance matrix $\phi(\Theta, F)(X'X)^{-1}$, where X is the design matrix of the model and $\phi(\Theta, F)$ is the asymptotic variance of the Θth sample quantile from a population with CDF F. These features of the GSQ lead us to expect they may prove useful in the search for robust alternatives to least aquares.

The plan of the paper is as follows. Section 2 defines notation and states assumptions. Section 3 sets forth the defining minimization problem for the GSQ and characterizes its solution. Invariance properties of the GSQ are established in Section 4. Section 5 derives the finite sample density of the GSQ estimator. The asymptotic distribution of the GSQ coefficients is deduced in Section 6. A final section contains a brief summary of results and some remarks on lines of future inquiry.

II. Notation and Assumptions

Let $\{y_t : t = 1,\ldots,T\}$ be independent random variables such that

$$\Pr\{y_t \leq Y\} = F(Y - x_t\beta) = F(u)$$

where β is a K-vector of unknown parameters, F is a probability distribution function, and x_t is the t^{th} row of a known TxK design matrix X of rank K. Equivalently, we can regard the T-vector y as being generated by the linear model,

$$y = X\beta + u$$

where u is a T-vector of independent and identically distributed (i.i.d) random variables with distribution function F.

In order to include the location sub-model as a special case, we assume X contains a constant vector so that $x_t = (1, z_t)$ or in matrix notation $X = (i\ Z)$ where i is a T-vector of ones and Z is a Tx(K-1) matrix of non-constant explanatory variables. In considering the limiting distribution of the GSQ's we allow T to increase with K fixed. The elements of the design matrix are assumed bounded and $\lim T^{-1}X'X = Q$, a K dimensional positive definite matrix.

Further we assume that the density function of u, f(u), is: strictly positive, bounded for all u, and twice continuously differentiable. These derivitives will be denoted f' and f". It may be noted that under these assumptions a realization of more than K observations (y_t, z_t) on the same hyperplane in R^K is an event of probability zero. Thus the solution to the linear programming problem of the next section will be, almost surely, nondegenerate.[4]

I_N will denote an N- dimensional identity matrix. The subscript is dropped when the dimensional is clear from the context. The transpose of a matrix will be denoted by prime '. Finally, vector inequalities are to be read componentwise.

III. Generalized Sample Quantiles

Given the data {y, X} the θth generalized sample quantile is defined as the solution to minimization problem:

Problem 1

$$\min \Psi(\beta,u) = \sum_{t\in\{t:u_t\geq 0\}} \theta|u_t| + \sum_{t\in\{t:u_t<0\}} (1-\theta)|u_t| \tag{3.1}$$

subject to,

$$(3.2)\qquad \begin{aligned} &y - X\beta = u \\ &\beta \epsilon R^K \\ &u \epsilon R^T \end{aligned}$$

A solution to this problem is denoted by the ordered pair (β^*, u^*) or often by merely β^* since $u^* = y - X\beta^*$. Both elements of the solution pair are sometimes expressed as functions of the data and the weighting parameter $\{y,X,\theta\}$.[5]

For computational purposes it is convenient to consider the equivalent linear programming problem:

<u>Problem 2</u>

$$(3.3)\qquad \min \psi(\beta,r^+,r^-) = \sum_{t=1}^{T} \theta r_t^+ + (1-\theta)r_t^-$$

subject to:

$$(3.4)\qquad \begin{aligned} &y-X\beta = r^+-r^- \\ &\beta \epsilon\ R^K \\ &r^+ \geq 0 \\ &r^- \geq 0 \end{aligned}$$

A solution to Problem 2 will be denoted (β^*,r^{+*},r^{-*}). To establish the equivalence of the two problems we first prove the following lemma.[6]

<u>Lemma</u> If (β^*,r^{+*},r^{-*}) is a solution to Problem 2 then $\min\{r_t^{+*},r_t^{-*}\}=0$ $t=1,\ldots T$.

<u>Proof</u>: Suppose (β^*,r^{+*},r^{-*}) is a solution but that,

$$\min\{r_1^{+*},r_1^{-*}\} = \alpha>0$$

Consider the vector,

$$[\beta^*,(r_1^{+*} - \alpha),r_2^{+*},\ldots,r_T^{+*};(r_1^{-*} - \alpha),r_2^{-*},\ldots,r_T^{-*}],$$

which still satisfies the constraints (34) of Problem 2. Substituting it into (33), the value of the objective function is reduced by α, so the lemma is established by contradiction.

Theorem If (β_1^*, u^*) is a solution to Problem 1 and $(\beta_2^*, r^{+*}, r^{-*})$ is a solution to Problem 2 for identical (y, X, θ) then,

$$(\beta_1^*, u^*) = (\beta_2^*, r^{+*} - r^{-*})$$

Proof Since $\min\{r_t^{+*}, r_t^{-*}\} = 0$ for all t,

$$\theta r_t^+ + (1-\theta) r_t^- = \begin{cases} \theta |r_t^+ - r_t^-| & \text{for } r_t^+ - r_t^- \geq 0 \\ (1-\theta) |r_t^+ - r_t^-| & \text{for } r_t^+ - r_t^- \leq 0 \end{cases}$$

Hence the objective functions and constraints of the two problems are identical.

To solve Problem 2 and obtain a characterization of the GSQ estimator we introduce the following notation. Let τ be the set of integers $\{1, 2, \ldots, T\}$, and H be the set of K-element subsets of τ. A typical element $h \in H$ has the relative complement $\bar{h} = \tau - h$. The notation $y(h)$ and $X(h)$ will be used to define subsets, or partitions, of the observations on the data, e.g. $y(h)$ denotes a K-vector of observations on y corresponding to $t \in h$.[7]

We now invoke a fundamental linear programming theorem to establish,

Proposition 1 A solution to Problem 2 has the form,

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

for some $h^* \in H$, where rank $X(h) = K$.

Proof: Since the rank of the constraint matrix $[X, I, -I]$ is T, a basic solution to Problem 2 has K+T zero variables and T non-zero variables. Of the T nonzero variables K must correspond to β since β receives no weight in the objective function. The lemma implies therefore that $u_t^*=0$ for K distinct elements of τ. This index set of "solution observations" is denoted by h^*, and the result follows.

We now proceed to characterize the GSQ estimator by establishing feasibility and optimality conditions for a particular $h \epsilon H$ given y, X, and Θ. The linear programming methods utilized below are discussed in Dantzig [7] and Spivey and Thrall [12].

For some $h \epsilon H$ partition the equality constraints of Problem 2 into 2 blocks. We then have the tableau,

$$(3.5) \quad \begin{bmatrix} \Psi \\ y(h) \\ y(\bar{h}) \end{bmatrix} = \begin{bmatrix} 0 & \Theta i_K' & \Theta i_{T-K}' & (1-\Theta) i_K' & (1-\Theta) i_{T-K}' \\ X(h) & I_K & 0 & -I_K & 0 \\ X(\bar{h}) & 0 & I_{T-K} & 0 & -I_{T-K} \end{bmatrix} \begin{bmatrix} \beta \\ r+(h) \\ r+(\bar{h}) \\ r-(h) \\ r-(\bar{h}) \end{bmatrix}$$

The basic columns in this tableau which yield $\beta^* = X(h)^{-1} y(h)$ as a proposed solution are given by

$$\begin{bmatrix} X(h) & 0 \\ X(\bar{h}) & D \end{bmatrix}$$

where D is some T-K dimensional diagonal matrix with 1 or -1 diagonal elements. The set of all such matrices will be denoted $\mathcal{D}$. Let M be the bordered matrix,

$$M = \begin{bmatrix} 1 & 0 & i'\Phi \\ 0 & X(h) & 0 \\ 0 & X(\bar{h}) & D \end{bmatrix}$$

where $\Phi = (\Theta-\frac{1}{2}) D + \frac{1}{2}I$ a $(T-K)^2$ diagonal matrix with Θ or $1-\Theta$ diagonal elements. The proposed basic columns may be introduced by pre-multiplying (35) by M^{-1}. Using partitioned inverse rules we have,

$$M^{-1} = \begin{bmatrix} 1 & i'\Phi DX(\bar{h})\, X(h)^{-1} & -i'\Phi D \\ 0 & X(h)^{-1} & 0 \\ 0 & -DX(\bar{h})X(h)^{-1} & D \end{bmatrix}$$

The revised tableau is,

$$\text{(3.6)} \qquad \begin{bmatrix} \Psi - i'\Phi D[y(\bar{h}) - X(h)\, X(\bar{h})^{-1}\, y(h)] \\ X(h)^{-1}\, y(h) \\ D[y(\bar{h}) - X(\bar{h})\, X(h)^{-1} y(h)] \end{bmatrix} =$$

$$\left[\begin{array}{c|c|c|c|c} 0 & \Theta i' + i'\Phi DX(\bar{h})\, X(h)^{-1} & i'[\Theta I + \Phi D] & (1-\Theta)i' - i'\Phi DX(h)\, X(h)^{-1} & i'[(I-\Theta)I - \Phi D] \\ I_k & X(h)^{-1} & 0 & -X(h)^{-1} & 0 \\ 0 & -D\, X(\bar{h})\, X(h)^{-1} & D & DX(\bar{h})\, X(h)^{-1} & -D \end{array}\right] \begin{bmatrix} \beta \\ r+(h) \\ r+(h) \\ r+(\bar{h}) \\ r-(\bar{h}) \\ r-(\bar{h}) \end{bmatrix}$$

The proposed solution for Probelm 2 is given by

$$\beta = X(h)^{-1}y(h)$$
$$r+(h) = 0$$
$$r+(\bar{h}) = 1/2(D+I)[y(\bar{h})-X(\bar{h})X(h)^{-1}y(h)]$$
$$r-(h) = 0$$
$$r-(\bar{h}) = 1/2(D-I)[y(\bar{h})-X(\bar{h})X(h)^{-1}y(h)]$$

On substituting these values into the revised tableau it is readily verified that the proposed solution satisfies the equality constraints. In linear programming terminology, the T basic variables of the proposed solution are the K elements of β and the T-K nonzero elements in $\{r+(\bar{h}), r-(\bar{h})\}$. The latter elements correspond to the diagonal elements of D which are +1 and -1 respectively.

The feasibility and optimality of the proposed solution can now be checked by referring to the revised tableau. Feasibility requires the non-negativity of (r+,r-). This is satisfied for (r+(h),r-(h)); for $(r+(\bar{h}), r-(\bar{h}))$ we require,

$$1/2(D+I)[y(\bar{h})-X(\bar{h})X(h)^{-1}y(h)] \geq 0$$
$$1/2(D-I)[y(\bar{h})-X(\bar{h})X(h)^{-1}y(h)] \geq 0$$

or combining these inequalities we have,

I. $$D[y(\bar{h})-X(\bar{h})X(h)^{-1}y(h)] \geq 0.$$

This will be called Condition I.

Optimality requires that the weights of the non-basic variables in the revised tableau be non-negative. This is readily verified for (r+(h), r-(h)) for any D. For the remaining non basic variables optimality requires,

$$\theta i' + i'\Phi\ D\ X(\bar{h})X(h)^{-1} \geq 0$$
$$(1-\theta)i' - i'\Phi\ D\ X(\bar{h})X(h)^{-1} \geq 0$$

or combining these inequalities we have,

II. $$-\theta i' \leq i'\Phi\ D\ X(\bar{h})X(h)^{-1} \leq (1-\theta)i'$$

This will be called Condition II.

A unique optimality condition can also be obtained with a slight strengthening of Condition II. Uniqueness requires strictly positive weights in the revised tableau. The unique optimality condition,

II.' $$-\theta i' < i'\Phi\ DX(\bar{h})X(h)^{-1} < (1-\theta)i'$$

will be called Condition II'.

These conditions now yield the following characterization of the GSQ:

<u>Theorem</u> $\beta^*(y, X, \theta) = X(h^*)^{-1}\ y(h^*)$ is a solution to Problem 2 if there exists a D satisfying Conditions I and II. The solution is unique if Condition II' is also satisfied.

By reversing the above argument the converse may also be established. That is, if $\beta^*(y, X, \theta) = X(h^*)^{-1}y(h^*)$ then there exists a D such that Conditions I and II are satisfied.[8]

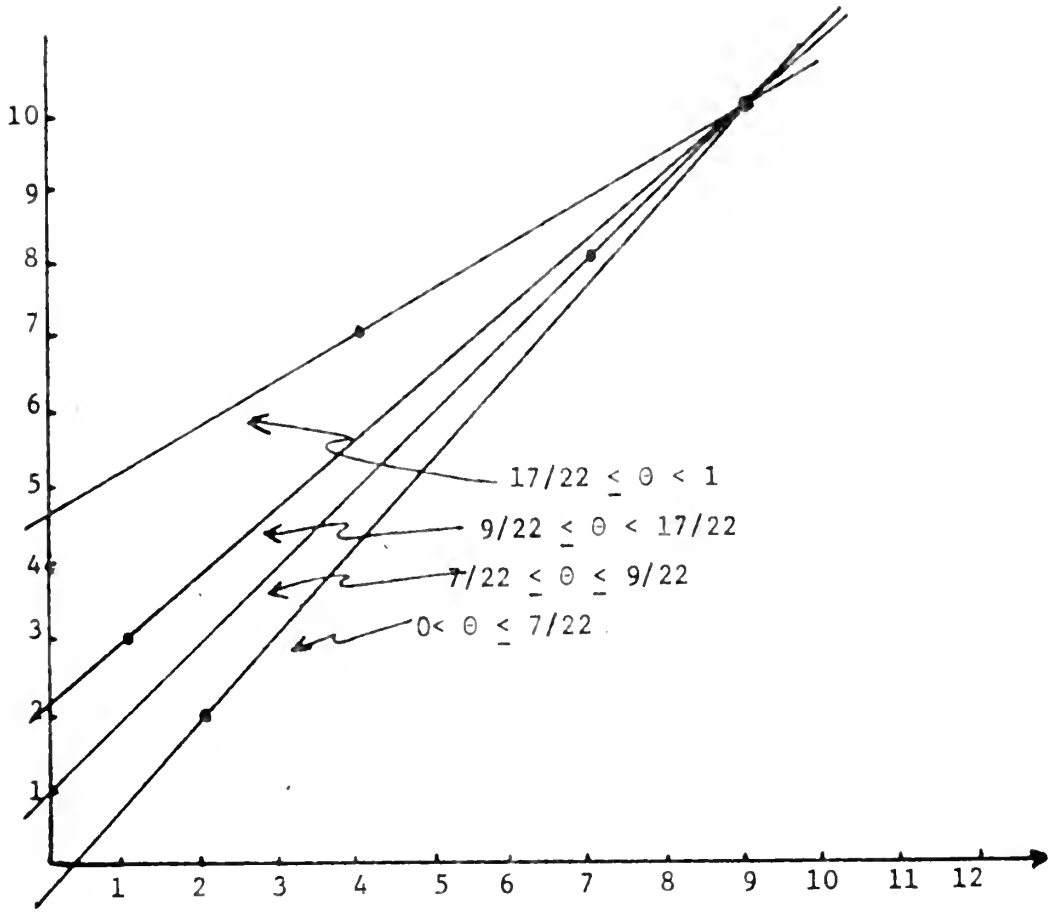
In the following sections it will prove useful to have a notation for the set of $D\varepsilon\mathcal{D}$ which satisfy Condition II for a particular index set h. Hence, let

$$\mathcal{D}(h) = \{D\varepsilon\mathcal{D} \mid -\theta i' \leq i'\Phi\ DX(\bar{h})\quad X(h)^{-1} \leq (1-\theta)i'\}.$$

It should be emphasized that $\mathcal{D}(h)$ is nonstochastic since it does not depend on y. The solution element of $\mathcal{D}(h^*)$ is denoted D* below.

Two sets of GSQ "regression" lines are illustrated in Figures 1 and 2. Data are presented in Tables 1 and 2. It will be noted that the GSQ estimator is not uniquely defined for certain values of Θ and that these points of non-uniqueness depend upon the realization of y. (The data for the two examples differ only in the value taken by y_5.) This situation is in contrast to the location submodel in which non-uniqueness arises only when $T\Theta$ is an integer and therefore does not depend upon sample information. The figures also suggest a conjecture which we do not pursue here. In the location submodel it is well known that the order-statistics are sufficient statistics for any sample regardless of the underlying CDF. Does an analogous result hold for the set $\{\beta^*(\Theta): 0<\Theta<1\}$? Can we reconstruct the sample vector y given the design X and the coefficients of all GSQ hyperplanes? The plausibility of this conjecture is increased by the observation that a GSQ hyperplane passes through each of the sample points as Θ varies between 0 and 1.

Figure 1



Model

$y = \beta_1 + \beta_2 x + u$

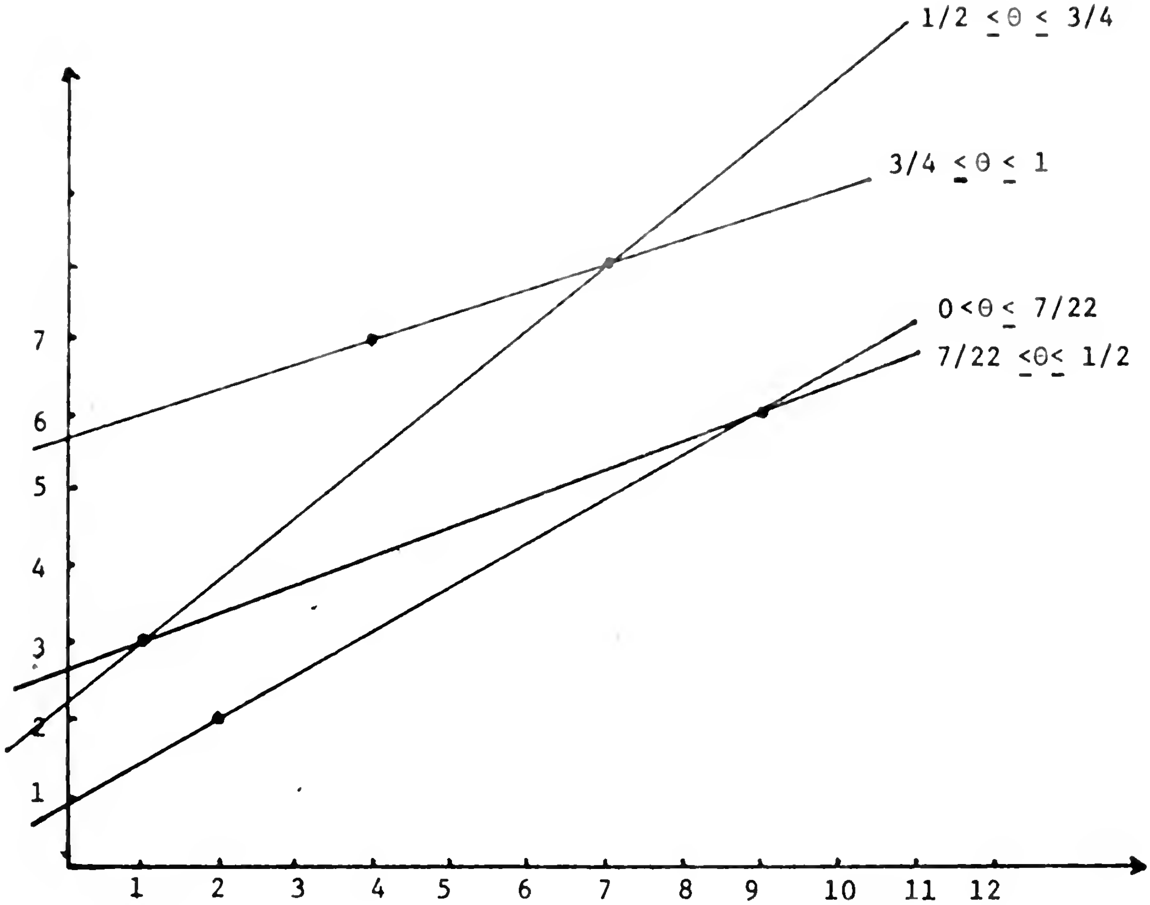
Data

$$y = \begin{bmatrix} 3 \\ 2 \\ 7 \\ 8 \\ 10 \end{bmatrix} \qquad X = \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 4 \\ 1 & 7 \\ 1 & 9 \end{bmatrix}$$

Estimates*

	$0<\theta\leq 7/22$	$7/22\leq\theta\leq 9/22$	$9/22\leq\theta\leq 17/22$	$17/22\leq\theta<1$
$\beta_1^*(\theta)$	-2/7	1	17/8	23/5
$\beta_2^*(\theta)$	8/7	1	7/8	3/5

Figure 2



Model

$$y = \beta_1 + \beta_2 x + u$$

Data

$$y = \begin{bmatrix} 3 \\ 2 \\ 7 \\ 8 \\ 6 \end{bmatrix} \qquad X = \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 4 \\ 1 & 7 \\ 1 & 9 \end{bmatrix}$$

Estimates

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

IV. Properties of Generalized Sample Quantiles

We now establish a number of properties of $\beta^*(y,X,\theta)$. The proofs generally verify that h^* and D^* are invariant with respect to some transformation of the data by checking Conditions I and II.

Property 1 $\beta^*(\lambda y,X,\theta) = \lambda\beta^*(y,X,\theta)$ for $\lambda\geq 0$.

Proof: Condition II does not depend upon y hence it is undisturbed by the transformation. Substituting in Condition I we have,

$$D^*[\lambda y(\bar{h}^*) - X(\bar{h}^*) X(h^*)^{-1}\lambda y(h^*)] \geq 0$$

which is undisturbed if $\lambda\geq 0$, hence

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

Property 2 $\beta^*(\lambda y,X,1-\theta) = \lambda\beta^*(y,X,\theta)$, for $\lambda<0$.

Proof: We show that if h^*, D^* satisfy Conditions I and II for $\{y,X,\theta\}$ then h^*, $-D^*$ satisfy conditions I and II for $\{y,X,1-\theta\}$, and conversely. Substituting the transformed data in Condition I gives,

$$D^*[\lambda y(\bar{h}^*) - X(\bar{h}^*) X(h^*)^{-1} \lambda y(h^*)] \geq 0$$

and this is satisfied for $\lambda<0$ with $-D^*$. From Condition II we know

$$-\theta\imath' \leq \imath'[(\theta-\tfrac{1}{2})I + \tfrac{1}{2}D^*] X(\bar{h}^*) X(h^*)^{-1} \leq (1-\theta)\imath' .$$

So multiplying by -1,

$$-(1-\theta)\imath' \leq \imath'[(1-\theta-\tfrac{1}{2})I -\tfrac{1}{2}D^*] X(\bar{h}^*) X(h^*)^{-1} \leq (1-(1-\theta))\imath'$$

so that $-D^*$ satisfies Condition II for $1-\theta$. Hence, for $\lambda<0$

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

Property 3 $\beta^*(y + X\gamma, X,\theta) = \beta^*(y,X,\theta) + \gamma$; $\gamma\epsilon R^K$

Proof: Condition II is undisturbed by the transformation of y. Condition I on the transformed data requires,

$$D^*[y(\bar{h}^*) + X(\bar{h}^*)\gamma - X(\bar{h}^*)X(h^*)^{-1}[y(h^*) + X(h^*)\gamma]] \geq 0$$

which simplifies to the original condition. Hence

$$\beta^*(y + X\gamma, X, \theta) = X(h^*)^{-1}[y(h^*) + X(h^*)\gamma] = \beta^*(y,X,\theta) + \gamma \ .$$

In the special case $\gamma = -\beta$, Property 3 implies

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

hence,

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

and Condition I may be written as,

$$D^*[u(\bar{h}^*) - X(\bar{h}^*)\ X(h^*)^{-1}\ u(h^*)] \geq 0$$

where $D^* \ \varepsilon \mathscr{D}(h^*)$.

<u>Property 4</u> $\beta^*(y,XA,\theta) = A^{-1}\ \beta^*(y,X,\theta)$, for any KxK matrix A of rank K.

Proof: Note that,

$$X(\bar{h}^*)\ A[X(h^*)A]^{-1} = X(\bar{h}^*)\ X(h^*)^{-1}$$

hence Conditions I and II are undisturbed by the reparameterization of the space spanned by the X's and,

$$\beta^*(y,XA,\theta) = [X(h^*)A]^{-1}\ y(h^*) = A^{-1}\ \beta^*(y,X,\theta).$$

<u>Property 5</u> $\beta^*([X\beta^*(y,X,\theta) + Mu^*(y,X,\theta)], X,\theta) = \beta^*(y,X,\theta)$; where M is any TxT diagonal matrix with non-negative elements.

Proof: Let

$$\tilde{y} = \begin{bmatrix} X(h^*) \\ X(\bar{h}^*) \end{bmatrix} \beta^*(y,X,\theta) + M\left[\begin{pmatrix} y(h^*) \\ y(\bar{h}^*) \end{pmatrix} - \begin{pmatrix} X(h^*) \\ X(h^*) \end{pmatrix} \beta^*(y,X,\theta)\right]$$

so,

$$\tilde{y} = \begin{bmatrix} y(h^*) \\ X(\bar{h}^*)\,X(h^*)^{-1}y(h^*) \end{bmatrix} + M\begin{bmatrix} 0 \\ y(\bar{h}^*) - X(h^*)\,X(h^*)^{-1}\,y(h^*) \end{bmatrix}$$

Let $\bar{M}$ be the lower $(T-K)^2$ submatrix of M, then we must verify,

$$D^*[\bar{M}\,y(h^*) - (\bar{M}-I)\,X(h^*)\,X(h^*)^{-1}\,y(h^*) - X(\bar{h}^*)\,X(h^*)^{-1}\,y(h^*)] \geq 0.$$

But this is clearly satisfied for M with non-negative elements if it was satisfied for h^*; D^* before the transformation.

<u>Property 6</u> Let P and N denote the respective number of positive and negative elements in $u^*(y,X,\theta)$. Then

$$N \leq T\theta \leq K+N = T-P .$$

Moreover these inequalies are strict when β^* is unique.

Proof: We first note that P and N are the number of positive and negative elements in D^*. Hence,

$$\iota' D^* \iota = P - N .$$

Next, Condition II implies that,

$$[(\theta-\tfrac{1}{2})\,\iota + \tfrac{1}{2}\iota'\,D^*]\,X(\bar{h}^*)\,X(h^*)^{-1} \equiv s'$$

where elements of the K-vector s satisfy $-\theta \leq s_k \leq 1-\theta$, $k=1,\ldots,K$. If β^* is unique the inequalities are strict. Postmultiplying by $X(h^*)$ and explicitly writing $X = [\iota\ Z]$ yields

$$(\theta-\tfrac{1}{2})\,\iota'\iota + \tfrac{1}{2}\iota'\,D^*\iota = s'\iota$$

$$(\theta-\tfrac{1}{2})\,\iota' Z(h^*) + \tfrac{1}{2}\iota'\,D^* Z(\bar{h}^*) = s' Z(h^*) .$$

From the first of these inequalities we have

$$(\theta-\tfrac{1}{2})(T-K) + \tfrac{1}{2}(P-N) = \sum_{k=1}^{K} s_k$$

But,

$$-K\theta \leq \sum_{k=1}^{K} s_k \leq K(1-\theta)$$

so that,

$$N \leq T\theta \leq N + K = T - P$$

with strict inequalities holding when β^* is unique.

Property 7 $\beta^*(y,\imath,\theta)$ is the θth sample quantile of y.

Proof: This follows directly from Property 6 with $X=\imath$. However a direct proof is instructive since it shows how Conditions I and II simplify to the sample quantile case. With $X=\imath$ we have,

$X(h) = 1$

$X(\bar{h}) = \imath'$: a (T-1)-vector of ones

$y(h) =$ the h^{th} element of y

$y(\bar{h}) =$ a (T-1)-vector of y's excluding $y(h)$.

Conditions I and II become

Iq. $D^*[y(\bar{h}^*) - \imath' \, y(h^*)] \geq 0$

IIq. $\tfrac{1}{2}(T-1) - \tfrac{1}{2}\imath'D^*\imath \leq \theta T \leq \tfrac{1}{2}(T+1) - \tfrac{1}{2}\imath'D^*\imath$

From (Iq) we see that the diagonal elements of D^* are +1 and -1 as the sign of the corresponding element of $[y(\bar{h}^*) - \imath \, y(h^*)]$ is positive or negative. Hence substituting $\imath'D^*\imath = P - N$ into (IIq) we have

$$N \leq T\theta \leq N + 1$$

which yields the conventional quantile. When $T\theta$ is not an integer the inequalities of (IIq) will be strictly satisfied and the quantile will be uniquely defined. When $T\theta$ is integral, the sample quantile will be non-unique. In this case the solution set to Problems 1 and 2 will consist of convex combinations of two observations on y. In this case, as with the GSQ, some convention may be adopted to resolve the ambiguity.

Property 7 may be extended to a comparison of quantiles design defined by,

$$X = \begin{bmatrix} i_{T_1} & \cdot & \cdot\ \cdot & 0 \\ 0 & i_{T_2} & & \cdot \\ \vdots & & \ddots & \cdot \\ 0 & & & i_{T_K} \end{bmatrix}.$$

<u>Property 8</u> If X has the comparison of quantiles design then $\beta^*(\theta)$ is a vector of θ^{th} sample quantiles for the K subsamples of y defined by X.

Proof: Conditions I and II reduce t K independent pairs of restrictions and the result follows.

From Properties 1 and 2 it will be noted that $\beta(\frac{1}{2})$, the minimization of absolute deviations, is the only scale invariant estimator within the class of generalized sample quantiles. However, we may also note in passing (defering a detailed discussion) that any estimator which is a linear combination of GSQ's with symmetric weights on $\beta^*(\theta)$ and $\beta^*(1-\theta)$ terms is also scale invariant.

Property 3 establishes that GSQ's are "shift" or "regression" invariant, Property 4 that they are invariant to a reparameterization of the design matrix.

Property 5 is a distinctive feature of GSQ's: $\beta^*(\theta)$ is invariant

to transformations of y as long as the signs of all of the residuals are left undisturbed. (The median of the two samples {1,2,3} and {-10,2,2.1} are identical.)

Property 6 places narrow bounds on the number of residuals which will be on either side of the GSQ hyperplane. It might be underlined that these bounds come from the intercept term in the X matrix, and tighter bounds could be calculated if all of the information in a particular X matrix were utilized.

Properties 7 and 8 specialize GSQ's to simple forms of the design matrix, establishing their equivalence with conventional definitions of the sample quantiles.

V. The Distribution of Generalized Sample Quantiles

We now consider the probability density function of the estimation error $\beta^*(\theta) - \beta$. We write the probability element of $\beta^*(\theta) - \beta$ as,

$$(5.1) \quad g(\delta)d\delta_1 \cdots d\delta_K = \Pr[\delta < \beta^*(\theta) - \beta < \delta + d\delta]$$

where $d\delta = (d\delta_1, \ldots, d\delta_K)'$ and $g(\delta)$ is the probability density function of $\beta^*(\theta) - \beta$.

For a given $(h, D) \in H \times \mathcal{D}(h)$ the joint probability of the event,

$$(5.2) \quad \delta < \beta^*(\theta) - \beta = X(h)^{-1} u(h) < \delta + d\delta$$

and the event

$$(5.3) \quad D[u(\bar{h}) - X(\bar{h}) X(h)^{-1} u(h)] \geq 0$$

can be written as the probability of (5.2) times the probability of (5.3) conditional on (5.2), or

$$(5.4) \quad \Pr[\delta < X(h)^{-1} u(h) < \delta + d\delta] \Pr[Du(h) > DX(\bar{h})\delta].$$

For distinct pairs (h,D) the events in the brackets of (5.4) are mutually exclusive. Hence the probability element of $(\beta^*(\theta) - \beta)$ is,

$$(5.5) \quad g(\delta)d\delta_1 \cdots d\delta_K = \sum_{h \in H} \sum_{D \in \mathcal{D}(h)} \Pr[\delta < X(h)^{-1} u(h) < \delta + d\delta] \cdot \Pr[Du(\bar{h}) > D X(\bar{h})\delta]$$

We may now invoke our assumptions on u in order to obtain an expression for $g(\delta)$ in terms of F.

From the i.i.d assumption we have,

$$(5.6) \quad \Pr[\delta < X(h)^{-1} u(h) < \delta + d\delta] = | X(h)' X(h)|^{\frac{1}{2}} \prod_{t \in h} f(x_t \cdot \delta) d\delta_1 \cdots d\delta_K$$

where $|X(h)'X(h)|^{\frac{1}{2}}$ is the Jacobian of $X(h)^{-1}$. Let $\Delta_t^D \varepsilon\{-1,1\}$ denote the t^{th} diagonal element of $D\varepsilon\mathcal{D}(h)$ where the index t runs over the integers in $\bar{h}$. Then, again by the i.i.d. assumption, we have,

$$(5.7) \qquad Pr[Du(h) \geq D\ X(h)\ \delta] = \prod_{t\varepsilon \bar{h}} [\tfrac{1}{2}(1 + \Delta_t^D) - \Delta_t^D\ F(x_t.\ \delta)]$$

A factor in this product is $F(x_t.\ \delta)$ if $\Delta_t^D = -1$ and $(1-F(x_t.\delta))$ if $\Delta_t^D = +1$. That is, the factor equals the probability that $u_t \leq x_t.\delta$ if $\Delta_t^D = -1$ and the probability that $u_t \geq x_t.\delta$ if $\Delta_t^D = +1$. Now substituting (5.6) and (5.7) into (5.5) we obtain,

$$(5.8) \quad g(\delta) = \sum_{h\varepsilon H} \sum_{D\varepsilon\mathcal{D}(h)} |X(h)'X(h)|^{\frac{1}{2}} \prod_{t\varepsilon h} f(x_t.\delta) \prod_{t\varepsilon \bar{h}} [\tfrac{1}{2}(1 + \Delta_t^D) - \Delta_t^D\ F(x_t.\delta)]$$

To illustrate (5.8) we consider the special case of the median $(\theta=\frac{1}{2})$ $X = 1$, $T = 2N + 1$.

Then,

$$g(\delta) = \sum_{h\varepsilon H} \sum_{D\varepsilon\mathcal{D}(h)} f(\delta) \prod_{t\varepsilon \bar{h}} [\tfrac{1}{2}(1 + \Delta_t^D) - \Delta_t^D\ F(\delta)].$$

and since every $D\varepsilon\mathcal{D}(h)$ contains N positive and N negative elements by Property 7 we have,

$$g(\delta) = \sum_{h\varepsilon H} \sum_{D\varepsilon\mathcal{D}(h)} f(\delta)\ [F(\delta)]^N\ [1 - F(\delta)]^N.$$

Finally, $\mathcal{D}(h)$ contains $\binom{T-1}{N}$ elements since this is the number of ways that exactly N plus ones can be assigned to $T-1$ places. H has T elements so that

$$g(\delta) = \frac{T!}{(n!)^2}\ [F(\delta)]^N\ [1-F(\delta)]^N\ f(\delta).$$

VI. The Asymptotic Distribution of the GSQ

To investigate the asymptotic behavior of $\beta^*(\theta)-\beta$ we consider the random variable $\sqrt{T/\theta(1-\theta)}\, f(\lambda)[\beta^*(\theta)-\beta-\bar{\delta}]$ where $\bar{\delta}$ is the K-vector $(\lambda,0,0,\ldots,0)$ and $F(\lambda)=\theta$. This variable has the density function,

$$\left[\frac{\theta(1-\theta)}{T}\right]^{K/2} f(\lambda)^{-K}\, g\left(\bar{\delta}+\frac{1}{f(\lambda)}\sqrt{\frac{\theta(1-\theta)}{T}}\;\delta\right),$$

or fully written out,

$$(6.1)\quad \left[\frac{\theta(1-\theta)}{T}\right]^{K/2} f(\lambda)^{-K} \sum_{h\in H}\sum_{D\in\mathcal{D}(h)} |X(h)'\,X(h)|^{\frac{1}{2}}$$

$$\prod_{t\in h} f(x_t\zeta)\cdot\prod_{t\in\bar{h}}\left[\tfrac{1}{2}(\Delta_t^D+1)-\Delta_t^D F(x_t\zeta)\right]$$

where $\zeta=\bar{\delta}+\sqrt{\theta(1-\theta)/T}\;f(\lambda)^{-1}\delta$.

Expanding the final product in Taylor series around $\zeta=\bar{\delta}$ we have (see Appendix I),

$$\prod_{t\in\bar{h}}\left[\tfrac{1}{2}(\Delta_t^D+1)-\Delta_t^D F(x_t\zeta)\right]=$$

$$|I-\Phi|\prod_{t\in\bar{h}}\frac{\frac{1}{2}(\Delta_t^D+1)-\Delta_t^D F(x_t\zeta)}{\frac{1}{2}(\Delta_t^D+1)-\Delta_t^D\theta}$$

$$(6.2)\quad = |I-\Phi|\,\exp\Big\{-[\theta(1-\theta)T]^{-\frac{1}{2}}i'\quad DX(\bar{h})\delta$$

$$-\tfrac{1}{2}T^{-1}\delta'\,X(h)'\,X(h)\delta$$

$$+\tfrac{1}{2}[\theta(1-\theta)T]^{-1}\delta'\,X(\bar{h})'\quad X(\bar{h})\delta$$

$$-\tfrac{1}{2}f'(\lambda)\,f(\lambda)^{-2}T^{-1}\,\delta'\,X(\bar{h})'\quad D\,X(\bar{h})\delta$$

$$+R(\tilde{\zeta})\Big\}$$

where $\tilde{\zeta}=(1-\tau)\bar{\delta}+\tau\zeta;\quad 0<\tau<1.$

The uniform boundedness of (6.1) is essential if we are to be able to investigate its asymptotic behavior term by term. In Appendix II a uniform bound is constructed for the case in which the design matrix, X, expands by replication. Persuasive intuitive evidence suggests that a much weaker condition (like, $\lim T^{-1}(X'X) = Q$, positive definite) would suffice to obtain a uniform bound, but this has not yet been formally established. In the remainder of this section we proceed under the hypothesis that (6.1) is uniformly bounded and therefore its limit may be investigated term by term.[10]

We now state four convergence results. (Proofs will be found in Appendix I.

1. $\lim\limits_{T\to\infty} \quad T^{-\frac{1}{2}} 1' \Phi \quad D\, X(\bar{h}) = 0$

2. $\lim\limits_{T\to\infty} \quad T^{-1} X(\bar{h})' \Phi \quad D\, X(\bar{h}) = \lim\limits_{T\to\infty} \quad T^{-1} X(\bar{h})' \Phi \quad X(\bar{h}) = 0$

3. $\lim\limits_{T\to\infty} \quad T^{-1} X(\bar{h})'\, X(\bar{h}) = Q$ for any $h \in H$.

4. $\lim\limits_{T\to\infty} \quad R(\zeta) = 0$

Noting that,

$$\lim_{T\to\infty} \prod_{t\varepsilon h} f(x_t;\zeta) = f(\lambda)^K,$$

we conclude that, asymptotically, the density (6.1) is proportional to,

$$\exp\{-\tfrac{1}{2}\delta' Q\delta\} ,$$

the constant of proportionality being,

$$\lim_{T\to\infty} \sum_{h\varepsilon H} \sum_{D\varepsilon \mathcal{D}(h)} |X(h)'X(h)|^{\frac{1}{2}} \, |I - \Phi| \left[\frac{\theta(1-\theta)}{T}\right]^{K/2} .$$

The constant must converge to $(2\pi)^{K/2}\ |Q|^{\frac{1}{2}}$ by the integrability of the limiting form of the density over all of R^K.

Thus $\beta^*(\theta)$, the GSQ estimator, is asymptotically multivariate normal

$$N(\beta + \bar{\delta},\ \phi(X'X)^{-1})$$

where,

$$\bar{\delta} = (\lambda,\ 0,\ 0,\ \ldots,\ 0)$$

$$\phi = \phi(\theta,F) = \frac{\theta(1-\theta)T}{f(\lambda)^2}$$

$$F(\lambda) = \theta\ .$$

The scale parameter of the asymptotic variance-covariance matrix will be recognized as the asymptotic variance of the θ^{th} sample quantile from a population with cumulative distribution function F.

For a normal distribution with mean 0 and variance σ^2, $f(0) = 1/\sigma\sqrt{2\pi}$; $\beta^*(\frac{1}{2})$, the least absolute error (LAE) estimator, is asymptotically normal with mean vector β and variance-covariance matrix,

$$\frac{\sigma^2\pi}{2T}\ (X'X)^{-1}\ .$$

This estimate has less precision than the least squares estimate by a factor of $2/\pi$. However, the median is known to have greater precision than the mean for a large number of "heavy-tailed" distributions and the results of this section generalize this fact to the linear model.

VII. Conclusion

We have shown that by placing asymmetric weights on positive and negative residuals in the conventional least absolute error (LAE) estimation problem, a new class of estimators for the linear model may be derived which generalize, in a natural way, the notion of sample quantiles. Since order statistics and conventional sample quantiles have proven so fruitful for robust estimation of location in the univariate model, it is hoped that the GSQ estimator can play an analogous role in constructing robust alternatives to least squares for models with "heavy-tailed" error distributions.

In the simple location submodel the GSQ estimator simplifies to the conventional sample quantile. The important invariance properties of sample quantiles generalize nicely to the GSQ. And we have shown that the coefficients of the θ^{th} GSQ hyperplane are asymptotically multivariate normal with variance-covariance matrix $\phi(\theta,F)(X'X)^{-1}$, where $\phi(\theta,F)$ is the asymptotic variance of the θ^{th} sample quantile from a population with distribution function F.

The joint (asymptotic) distribution of the vector of GSQ parameter estimates must now be investigated, with the ultimate aim of studying the distribution of linear combinations of GSQ statistics. Natural generalizations to the linear model of trimmed means, inter-quantile midranges and other location estimators are suggested by the GSQ estimator; it remains to investigate their statistical properties. At some point it will become essential to study small-sample properties of these estimators via Monte-Carlo techniques, but at present the vast literature on sample quantiles and order statistics provides a rich source of theoretical analogues to be considered.

Appendix I

We first develop the expansion (5.2). The product over $t\epsilon\bar{h}$ is multiplied and divided by the determinant of the matrix $I - \Phi$, which from the definition of Φ (see page 7 above) is diagonal, (T-K) x (T-K), with typical element $[\frac{1}{2}(\Delta_t^D + 1) - \Delta_t^D\theta]$. Taking logs we consider the sum,

$$\ln S(\zeta) = \sum_{t\epsilon\bar{h}} \ln\left[\frac{\frac{1}{2}(\Delta_t^D+1) - \Delta_t^D F(x_t\zeta)}{\frac{1}{2}(\Delta_t^D+1) - \Delta_t^D \theta}\right] .$$

Expanding in Taylor series around $\zeta = \bar{\delta}$,

$$\ln (\zeta) = \ln (\bar{\delta}) + [(\zeta - \bar{\delta}) \cdot \nabla] \ln S(\bar{\delta})$$

$$+ 1/2![(\zeta - \bar{\delta})\cdot\nabla]^2 \ln S(\bar{\delta})$$

$$+ 1/3![(\zeta - \bar{\delta})\cdot\nabla]^3 \ln S(\tilde{\zeta})$$

where ∇ is the partial differentiation operator and the third order term is evaluated at $\tilde{\zeta} = (1-\tau)\bar{\delta} + \tau\zeta$.

Evaluating $\ln S$ and the higher order terms yields

$$\ln S(\zeta) = \sum_{t\epsilon\bar{h}} \ln\left[\frac{\frac{1}{2}(\Delta_t^D+1) - \Delta_t^D\theta}{\frac{1}{2}(\Delta_t^D+1) - \Delta_t^D\theta}\right] = 0$$

$$[(\zeta-\bar{\delta})\cdot\nabla] \ln S(\bar{\delta}) = \sum_{t\epsilon\bar{h}} [\sqrt{\frac{\theta(1-\theta)}{T}}\, f(\lambda)^{-1}\delta]\cdot \frac{-\Delta_t^D f(\lambda)}{\frac{1}{2}(1+\Delta_t^D) - \Delta_t^D\theta}\, x_t$$

$$= -[\theta(1-\theta)T]^{-\frac{1}{2}} \sum_{t\epsilon\bar{h}} [\tfrac{1}{2}(1+\Delta_t^D) - \Delta_t^D]\, \Delta_t^D(x_t\delta)$$

$$[(\zeta-\bar{\delta})\cdot\nabla]^2 \ln S(\bar{\delta}) = \sum_{t\epsilon\bar{h}} \frac{\theta(1-\theta)}{T}\cdot f(\lambda)^{-2}\cdot f(\lambda)^2\cdot[\tfrac{1}{2}(1+\Delta_t^D) - \Delta_t^D\theta]^{-2}(x_t\delta)^2$$

$$- \sum_{t\epsilon\bar{h}} \frac{\theta(1-\theta)}{T} f(\lambda)^{-2} f'(\lambda)\, \Delta_t^D[\tfrac{1}{2}(1+\Delta_t^D) - \Delta_t^D\theta]^{-1} (x_t\delta)^2$$

$$- \sum_{t\epsilon\bar{h}} [\theta(1-\theta)T]^{-1} [(\theta-\tfrac{1}{2}) + \tfrac{1}{2}\Delta_t^D]^2 (x_t \cdot \delta)^2$$

$$- \sum_{t\epsilon\bar{h}} f'(\lambda) f(\lambda)^{-2} [(\theta-\tfrac{1}{2}) + \tfrac{1}{2}\Delta_t^D] (x_t \cdot \delta)^2$$

$$= - \sum_{t\epsilon\bar{h}} \{T^{-1} - [\theta(1-\theta)T]^{-1} [(\theta-\tfrac{1}{2}) + \tfrac{1}{2}\Delta_t^D]\} (x_t \cdot \delta)^2$$

$$- \sum_{t\epsilon\bar{h}} f'(\lambda) f(\lambda)^{-2} [(\theta-\tfrac{1}{2}) + \tfrac{1}{2}\Delta_t^D] (x_t \cdot \delta)^2$$

$$[(\zeta-\bar{\delta})\cdot\nabla]^3 \ln S(\tilde{\zeta}) = \sum_{t\epsilon\bar{h}} [\theta(1-\theta)T]^{-3/2} f(\lambda)^{-3} W_t(\tilde{\zeta})(x_t \cdot \delta)^3$$

where,

$$W_t(\tilde{\zeta}) = \left[\frac{2\Delta_t^D f(x_t \cdot \tilde{\zeta})}{1+\Delta_t^D - 2\Delta_t^D F(x_t \cdot \tilde{\zeta})}\right]^3 + \frac{12 f(x_t \cdot \tilde{\zeta}) f'(x_t \cdot \tilde{\zeta})}{(1+\Delta_t^D - 2\Delta_t^D F(x_t \cdot \tilde{\zeta}))^2}$$

$$+ \frac{2\Delta_t^D f''(x_t \cdot \tilde{\zeta})}{1+\Delta_t^D - 2\Delta_t^D F(x_t \cdot \tilde{\zeta})}$$

In the matrix notation of earlier sections the expansion may be written as,

$$\begin{aligned} \ln S(\zeta) = & -[(\theta(1-\theta)T]^{-\frac{1}{2}} 1' \Phi D X(\bar{h})\delta \\ & -\tfrac{1}{2}T^{-1}\delta' X(\bar{h})' X(\bar{h})\delta \\ & + \tfrac{1}{2}[\theta(1-\theta)T]^{-1} \delta' X(\bar{h})' \Phi X(\bar{h})\delta \\ & - \tfrac{1}{2}f'(\lambda) f(\lambda)^{-2} T^{-1} \delta' X(\bar{h})' \Phi D X(\bar{h})\delta \\ & + R(\tilde{\zeta}). \end{aligned}$$

We now establish the four convergence results of Section 6.

(1) $\lim_{T\to\infty} T^{-\frac{1}{2}} 1' \Phi D X(\bar{h}) = 0$

Proof: From Condition II we have that,

$$-\theta i' \leq i' \, \Phi \, D \, X(\bar{h}) \, X(\bar{h})' \leq (1-\theta) i'$$

Denote the middle term by s', then

$$i' \, \Phi \, I^{*} X(\bar{h}) = s'X(h).$$

$X(h)$ is a fixed matrix and the elements of s are bounded so the result follows.

(2) $$\lim_{T\to\infty} T^{-1} X(\bar{h}) \, \Phi \, D \, X(\bar{h}) = \lim_{T\to\infty} T^{-1} X(\bar{h})' \, \Phi \, X(\bar{h}) = 0$$

Proof: We show that the matrix

$$X(\bar{h})' \, \Phi \, D \, X(\bar{h}) = \sum_{t\epsilon\bar{h}} [(\theta-\tfrac{1}{2}) + \tfrac{1}{2}\Delta_t^{D}] \, x_t x_t'$$

is bounded for all T. The KxK matrices $x_t x_t'$ are bounded, therefore there exists a matrix C such that

$$\Sigma[\theta-\tfrac{1}{2} + \tfrac{1}{2}\Delta_t^{D}] \, x_t x_t' \leq C \, \Sigma \, [\theta-\tfrac{1}{2} + \tfrac{1}{2}\Delta_t^{D}].$$

From result 1, $i' \, \Phi \, D \, X(\bar{h}) = s'X(h)$, and explicitly partitioning $X = [i \vdots Z]$ we have from the intercept term,

$$i' \, \Phi \, i = s'i \ .$$

But the elements of s are all bounded in the interval $[-\theta, 1-\theta]$, hence their sum is less than K, so,

$$-K\theta \leq \Sigma[\theta-\tfrac{1}{2} + \tfrac{1}{2}\Delta_t^{D}] \leq K(1-\theta)$$

and,

$$-K\theta C \leq X(\bar{h})' \, \Phi \, D \, X(\bar{h}) \leq K(1-\theta) \, C \ .$$

Similarly, $i' \, \Phi \, i = -s'i$ and therefore the corresponding sum is also bounded and the result follows.

(3) $\lim_{T\to\infty} T^{-1}X(\bar{h})'X(\bar{h}) = Q$ for any $h\epsilon H$

Proof: $X'X = X(h)'X(h) + X(\bar{h})'X(\bar{h})$ but $X(h)$ is fixed so,

$$\lim_{T\to\infty} T^{-1}X(\bar{h})'X(\bar{h}) = \lim_{T\to\infty} T^{-1}X'X = Q$$

(4) $\lim_{T\to\infty} R(\tilde{\zeta}) = 0$

Proof: Since

$$0 < 1 + \Delta_t^D - 2\Delta_t^D F(x_t\tilde{\zeta}) < 2$$

and f, f', f" are bounded, it follows that $W_t(\tilde{\zeta})$ is bounded. The vector x_t is bounded by hypothesis, hence there exists a number C such that

$$|W_t(\tilde{\zeta})\ (x_t\tilde{\zeta})^3| < C < \infty$$

and

$$|R(\tilde{\zeta})| < \sum_{t\epsilon\bar{h}} 1/6(\theta(1-\theta))^{3/2} f(\lambda)^{-3}T^{-3/2}C$$

$$< 1/6(\theta(1-\theta))^{-3/2} f(\lambda)^{-3}T^{-\frac{1}{2}}C$$

Therefore,

$$\lim_{T\to\infty} R(\tilde{\zeta}) = 0.$$

Appendix II

Consider the density function,

$$\text{(II.1)} \qquad \left[\frac{\theta(1-\theta)}{T}\right]^{K/2} f(\lambda)^{-K} \sum_{h\varepsilon H} \sum_{D\varepsilon\mathcal{D}(h)} |X(h)'X(h)|^{\frac{1}{2}} \, |I-\Phi|$$

$$\cdot \prod_{t\varepsilon h} f(x_t;\zeta) \prod_{t\varepsilon \bar{h}} \frac{\frac{1}{2}(\Delta_t^D+1)-\Delta_t^D F(x_t;\zeta)}{\frac{1}{2}(\Delta_t^D+1)-\Delta_t^D \theta}$$

The matrices X(h) are bounded, as is f(·) and the bracketed fraction is uniformly bounded so there exists a constant C_1 such that (II.1) is less than,

$$C_1 \; \frac{\theta(1-\theta)}{T}^{K/2} \; |I-\Phi| \; \#(H) \; \#[\mathcal{D}(h)]$$

uniformly for all T and $\zeta \in R^K$. (#(A) denotes the number of elements in the set A.) Clearly,

$$\#(H) \leq \binom{T}{K} < T^K$$

$$|I-\Phi| \leq 2^{-(T-K)}.$$

So (II.1) is bounded from above by

$$C_1[\theta(1-\theta)]^{K/2} \, T^{K/2} \, 2^{-(T-K)} \#[\mathcal{D}(h)]$$

$$\leq C_2 \, T^{K/2} \, 2^{-T} \, \#[\mathcal{D}(h)]$$

when $C_2 = C_1 2^{K/2}$. It remains to show that the number of elements in $\mathcal{D}(h)$ increases no faster than $2^T T^{-K/2}$.

Let A be a KxT matrix and ε a T-vector whose elements are ±1. We wish to estimate the number of solutions, ε, to the symmetric system,

$$-i \leq A\varepsilon \leq i \; .$$

We begin by considering the case in which A expands by replication.

Let $A_{(1)} = (a_{ij})$ be a KxM matrix whose first K columns are assumed to be invertible. Denote by $A_{(L)}$ the L^{th} replication of $A_{(1)}$, and by $\alpha_{ij}^{(L)}$ the entries of $A_{(L)}$. If $s<M$, we have that $\alpha_{irM+s}^{(r)} = a_{is}$ for $r=0,1,\ldots,L-1$. Let $B = (b_{ij})$ be the inverse of the KxK submatrix of $A_{(1)}$; i.e. $\sum_{k=1}^{K} b_{ik}a_{kj} = \delta_{ij}$ when $j \leq K$. Finally, set $T = LM$.

We now consider the system of inequalities

$$-1 \leq \sum_{rM+s=1}^{T} \alpha_{irM+s}^{(r)} \quad \varepsilon_{rM+s} \leq 1 \qquad (i = 1, \ldots, K)$$

We have,

$$\sum_{rM+s=1}^{T} \alpha_{irM+s}^{(r)} \varepsilon_{rM+s} = \sum_{s=1}^{M} \sum_{r=0}^{L-1} a_{is}\varepsilon_{rM+s}$$

For any T-triple $(\varepsilon_1, \ldots, \varepsilon_T)$ let p_s denote the number of the ε_{rM+s} which are positive, $r = 0,1,\ldots,L-1$. Thus the system may be rewritten as,

$$-1 \leq \sum_{s=1}^{M} a_{is}(p_s-(L - p_s)) \leq 1 \qquad (i = 1, \ldots, K)$$

$$-\tfrac{1}{2} \leq \sum_{s=1}^{M} a_{ks}p_s - A(k)\ L \leq \tfrac{1}{2} \qquad (k = 1, \ldots, K)$$

where $A(i) = \frac{1}{2} \sum_{s=1}^{M} a_{is}$. Multiplying by b_{ik} we have,

$$-\frac{|b_{ik}|}{2} \leq \sum_{s=1}^{M} b_{ik}a_{ks}p_s - b_{ik}A(k)\ L \leq \frac{|b_{ik}|}{2}$$

Summing over k and setting,

$$B(i) = \sum_{k=1}^{K} \frac{|b_{ik}|}{2}; \qquad C(i) = \sum_{k=1}^{K} b_{ik}A(k)$$

we obtain,

$$-B(i) \le \sum_{s=1}^{M} \sum_{k=1}^{K} b_{ik} a_{ks} p_s - C(i)\, L \le B(i)$$

$$-B(i) \le p_s + \sum_{s=k+1}^{M} b_{ik} i_{ks} p_s - C(i)\, L \le B(i) \qquad (i = 1,\dots,K)$$

Denote the L-tuple $(\varepsilon_s, \varepsilon_{M+s}, \dots, \varepsilon_{(L-1)M+s})$ by E_s. Then for each $s > K$ there are 2^L L-tuples E_s, so there are $2^{L(M-K)} = 2^{T-LK}$ simultaneous choices for the E_s: $s = K+1, \dots, M$. Let C be a fixed integer larger than $2B(i)+1$ for all i. Then for each choice of the E_s, $s > K$, there are less than C values for p_i, $i \le K$ and for each of these p_i there are at most $\binom{L}{L/2}$ choices of E_i with p_i positive elements. By Stirling's formula there is a uniform constant $\tilde{C}$ (independent of L) such that

$$\binom{L}{L/2} \le \tilde{C} \frac{2^L}{L^{1/2}}$$

Setting $\hat{C} = C\tilde{C}$ we obtain,

$$\left(\hat{C} \frac{2^L}{\sqrt{L}}\right)^K 2^{T-LK} = \hat{C}^K \frac{2^T}{T^{K/2}} = \hat{C}^K M^{K/2} \frac{2^T}{T^{K/2}} = C' \frac{2^T}{T^{K/2}}$$

as a bound for the number of solutions. The constant C' is, of course, independent of L. Had we chosen an asymmetric system ($\theta \ne 1/2$), a smaller bound could have been constructed, since $\binom{L}{L/2} \ge \binom{L}{L\theta}$ $0 < \theta < 1$.

The generalization of this result to weaker conditions on the design matrix remains elusive. A geometric analogy is suggestive however. The T-vectors, ε, may be thought of as vertices of a T dimensional hypercube in a sphere of radius $\sqrt{T}$. The surface area of the sphere is proportional to $T^{T/2}$ and the fraction of that area satisfying the K independent inequalities

$$-1 \leq A\,c \leq 1$$

is proportional to $T^{-K/2}$. Unfortunately the discrete vertices of the hypercube are not distributed uniformly on the sphere necessitating a more delicate argument.

Footnotes

1. See e.g. Andrews, et.al. [1] and Huber [9].

2. See the review articles by Huber [8] and Bickel [3], and the recent work of Yohai [13] and Relles [10].

3. The primary exponent of this approach has been Bickel [3,4].

4. For a discussion of degeneracy in linear programming see Spivey and Thrall [12; pp. 90–99].

5. It will be noted that the case $\theta = \frac{1}{2}$ specializes the problem to the minimization of absolute deviations which has been extensively studied in recent years. For a review of this literature see Taylor [11] and Bassett [2].

6. The equivalence of Problems 1 and 2 for $\theta = \frac{1}{2}$ was first pointed out by Charnes, Cooper, and Ferguson [5].

7. This approach to the LAE ($\theta=\frac{1}{2}$) problem was developed by Bassett and is discussed in Taylor [11] and Bassett [2].

8. See Spivey and Thrall [12; chapter 3].

9. In the remainder of the paper Conditions I and II will be used to characterize the GSQ estimator. Hence a unique solution to Problems 1 and 2 need not always exist. The properties of the GSQ estimator discussed in the next section are valid for any element $\beta^*(\theta)$ in the solution set to Problems 1 and 2. For practical purposes, it may be convenient to adopt some arbitrary rule to select a single solution vector in the case of multiple optimal solutions.

10. The general argument in this section parallels closely Cramér's proof of the asymptotic normality of ordinary sample quantiles; see [6; pp. 367-70]. The restrictive hypothesis of a replicated experi-

mental design is employed in Appendix II to calculate a uniform bound for the expression (6.1). A revised version of the paper will generalize these results to weaker design conditions.

References

[1] Andrews, D.F., P.J. Bickel, F.R. Hampel, P.J. Huber, W.H. Rogers, and J.W. Tukey, Robust Estimates of Location: Survey and Advances, Princeton: Princeton University Press, 1972.

[2] Bassett, G.W., Some Properties of the Least Absolute Error Estimator, Unpublished Dissertation, University of Michigan, December, 1973.

[3] Bickel, P.J., "Analogues of Linear Combinations of Order Statistics in the Linear Model,"

[4] Bickel, P.J., "On Some Analogues to Linear Combinations of Order Statistics in the Linear Model," The Annals of Statistics. (1), 1973, 597-616.

[5] Charnes, A., W.W. Cooper, and R.O. Ferguson, "Optimal Estimation of Executive Compensation by Linear Programming," Management Science, January, 1955.

[6] Cramér, H., Mathematical Methods of Statistics, Princeton: Princeton University Press, 1946.

[7] Dantzig, G.B., Linear Programming and Extensions, Princeton: Princeton University Press, 1963.

[8] Huber, P.J., "Robust Regression: Asymptotics, Conjectures and Monte-Carlo," The Annals of Mathematical Statistics, 1973, 799-821.

[9] Huber, P.J., "Robust Statistics: A Review," The Annals of Mathematical Statistics, 1972, 1041-67.

[10] Relles, D. Robust Regression by Modified Least Squares, Unpublished Ph.D. Dissertation, Yale University, 1968.

[11] Spivey, W.A.,and R.M. Thrall, Linear Optimization, New York: Holt, Rinehart, Winston, 1970.

[12] Taylor, L.D., "Estimation by Minimizing the Sum of Absolute Errors," in P. Zarembka (ed.), Frontiers in Econometrics, New York: Academic Press, 1973.

[13] Yohai, V.J., "Robust Estimation in the Linear Model," Annals of Statistics, 1974, 562-67.