UNIVERSITY
OF
WISCONSIN-
MADISON

SSRI
Workshop Series

TWO-STEP QUANTILE ESTIMATION OF THE CENSORED REGRESSION MODEL

James L. Powell

8615

SOCIAL SYSTEMS RESEARCH INSTITUTE

Social Systems Research Institute

University of Wisconsin

TWO-STEP QUANTILE ESTIMATION OF THE CENSORED REGRESSION MODEL

James L. Powell

8615

June, 1986

TWO-STEP QUANTILE ESTIMATION OF THE CENSORED REGRESSION MODEL

James L. Powell
Department of Economics
University of Wisconsin
Madison, WI 53706

ABSTRACT

It has previously been shown that consistent estimation of the unknown coefficients of the censored regression (or censored "Tobit") model can be obtained using a quantile estimation approach. Because the functional form of the conditional quantiles of the censored dependent variable does not depend on the parametric form of the distribution of the error terms, the quantile estimators are consistent (and asymptotically normal) under much more general conditions than likelihood-based estimation methods. However, simulation studies have shown that quantile estimates for this model have a small-sample bias in the opposite direction from the well-known least-squares bias for censored data. This bias results from the particular form of the conditional quantile function; quantile estimation in this problem uses only those data points with a positive regression function to estimate the regression coefficients, and the simultaneous estimation of the sign of the regression function and the magnitude of the coefficients induces an asymmetry in the small-sample distribution of the coefficient estimates.

The present paper proposes a two-step quantile estimator for this model which is designed to overcome this finite-sample bias. Specifically, the first stage applies quantile estimation to the sign of the observed dependent variable (i.e., the quantile generalization of the "maximum score" estimator for a binary response model) to estimate the sign of the underlying regression function, while the second stage uses the data points with positive (estimated) regression functions to estimate the regression coefficient magnitudes using standard linear quantile estimation. By separating the estimation of the sign of the regression function from estimation of the coefficents themselves, the finite-sample bias of the quantile approach for this model is attenuated.

The paper gives conditions under which the two-step estimator is asymptotically equivalent to the one-step version, so that no efficiency loss is incurred from the two-step approach. Also, a small scale efficiency study will document the small-sample benefits of the proposed method.

TWO-STEP QUANTILE ESTIMATION FOR THE CENSORED REGRESSION MODEL

by

James L. Powell*

1. Introduction

For the censored regression (or censored "Tobit") model, estimation of the unknown regression parameters using a least absolute deviations (LAD) criterion has recently been investigated (Powell [1984]) and extended to a more general class of regression quantile estimators (Powell [1985]). This class of estimators was shown to be consistent and asymptotically normally distributed for a wide range of distributions of the error terms (unlike maximum likelihood or conditional expectation/nonlinear least squares estimators for the censored regression model, which are inconsistent in general if the parametric form of the error distribution is misspecified). A closely-related estimator for this model, based upon "symmetric trimming" of the dependent variable, was proposed by Powell [1983]; this method, which required symmetry of the error distribution, was shown to be a least squares (LS) analogue of the censored LAD estimator, and was computationally simpler. Consistency and asymptotic normality of this "symmetrically censored LS" estimator was demonstrated under conditions similar to those for the censored regression quantile estimators; it, too, is "semiparametric," in the sense that its consistency does not depend on a specific parametric form of the error distribution .

Given the favorable large-sample properties of the censored quantile and

---

* This research was supported by National Science Foundation Grant No. SES-8510364 at M.I.T. I am grateful to Chuck Manski and Whitney Newey for their helpful comments.

symmetrically trimmed least squares estimators, the question of their finite-sample performance naturally arises. This question was addressed in two simulation studies: Paarsch [1984], which compared censored LAD estimation to normal maximum likelihood and "two-step" least squares estimation, and Powell [1983], which compared symmetrically censored least squares to normal maximum likelihood estimation. Aside from results which would be obvious a priori (e.g., that the behavior of the estimators improves as the sample size increases or as the censoring proportion or error variability decreases), three "stylized facts" emerged from these studies. First, the censored quantile and symmetrically censored estimators were much more efficient (in a mean-squared error sense) than a parametric "two-step" estimator, which was based upon the assumption of Gaussian error terms and which used only those observations with a positive dependent variable in the second stage. Second, unless the sample size is fairly large or the error distribution very nonnormal, the inconsistency of (misspecified) Gaussian maximum likelihood can be small compared to its efficiency advantage (in terms of estimator variance) over the censored LAD or symmetrically censored estimators, at least for the sample designs considered. Finally, the censored LAD and symmetrically censored LS estimators have finite-sample distributions which are (mean) biased in the opposite direction from the well-known classical least squares bias.

This last result -- the finite-sample bias of the semiparametric estimators -- is due to an asymmetry in the sampling distributions of the coefficient estimators rather than an actual "recentering" of those distributions away from the true parameter values. That is, the estimators are nearly median unbiased, but the distribution of the estimator of a slope coefficient is positively skewed (and the intercept estimator has a negatively skewed distribution), so the mean of the sampling distribution of the slope estimator exceeds its median. This asymmetry is less pronounced for designs with less censoring and for error distributions which are heavier-tailed (more kurtotic) distributions than the Gaussian distribution, but appears to be present to some extent in all of the sampling experiments yet conducted.

The present paper describes an attempt to address this problem using a "two-step" modification of the censored regression quantile estimation strategy. In the following section, a numerical example is given to illustrate the reason for the asymmetry in the distribution of "one-step" quantile estimators, and the motivation for the class of two-step estimators as a means of reducing this asymmetry is outlined. The asymptotic properties of the two-step estimators is investigated in section 3; specifically, it is shown that the one- and two-step estimators are asymptotically equivalent (to first order), though under much stronger conditions on the behavior of the regressors than previously imposed for one-step estimation. In section 4, results of a small-scale simulation study for the two-step estimators is presented; unfortunately, this study shows the proposed estimators are also biased in finite samples for the designs considered, though in the opposite direction from the bias for the corresponding one-step procedures. The practical implications of this finding are discussed in the conclusion.

2. Rationale for the Two-Step Approach

The censored regression model can be written in the form

$$y_i = \max\{0, x_i'\beta_0 + u_i\}, \qquad i = 1, \ldots, n, \tag{2.1}$$

where the dependent variable y_i and the regression vector x_i are observed for each i, while the parameter vector β_0 and error term u_i are not observed. For this model, the censored LAD estimator $\hat{\beta}_n$ was defined in Powell [1984] to be that value of β minimizing

$$S_n(\beta) = \frac{1}{n} \sum_{i=1}^{n} |y_i - \max\{0, x_i'\beta\}|, \tag{2.2}$$

over all β in some parameter space B. This estimation method, based upon the conditional median of y_i, was extended in Powell [1985] to arbitrary quantiles estimation of the slope coefficients of β_0 through minimization of

$$Q_n(\beta; \theta) \equiv \frac{1}{n} \sum_{i=1}^{n} \rho_\theta(y_i - \max\{0, x_i'\beta\}) , \tag{2.3}$$

where ρ_θ is the "check function" (Koenker and Bassett [1978, 1982]),

$$\rho_\theta(\lambda) \equiv [\theta - 1(\lambda < 0)]\cdot\lambda , \tag{2.4}$$

for 1(A) denoting the indicator function of the event A (i.e., it takes the value one if A is true, and is zero otherwise). The censored LAD estimator $\hat{\beta}_n$ is easily seen to be a special case of this more general quantile estimator, denoted $\hat{\beta}_n(\theta)$; $Q_n(\beta; \frac{1}{2}) = S_n(\beta)/2$, so $\hat{\beta}_n \equiv \hat{\beta}_n(\frac{1}{2})$, as defined above. Provided the first component of the regressors is an intercept term ($x_{i1} \equiv 1$) and the error terms u_i are i.i.d., the estimand for the censored regression quantile estimator is $\beta_0(\theta) \equiv \beta$sion quantile estimator is $\beta_0(\theta) \equiv \beta_0 + F_u^{-1}(\theta)\cdot e^1$, with $F_u(\lambda)$ the c.d.f. of u_i and e^1 the first unit basis vector, i.e., $e^1 \equiv (1, 0, \ldots, 0)'$.

A closely related estimator, the symmetrically censored least squares estimator (Powell [1983]), was based on the assumption of symmetrically distributed error terms (so that the intercept term of β_0 could be interpreted as either the median or mean -- if it exists -- of the error distribution). The estimator was defined as the minimizer of

$$(2.5) \qquad R_n(\beta) \equiv \frac{1}{n}\sum_{i=1}^{n}(y_i - \max(\frac{1}{2}y_i,\ x_i'\beta))^2 + \frac{1}{n}\sum_{i=1}^{n} 1(y_i > 2x_i'\beta)\cdot[(\frac{1}{2}y_i)^2 - (\max(0,\ x_i'\beta))^2] \ ;$$

the estimator minimizing this objective function (which reduces to $S_n(\beta)$ when absolute replaces squared error loss in (2.5)) was shown to be consistent and asymptotically normally distributed under similar conditions to those for censored LAD estimation.

As discussed in the previous section, simulation studies of the censored LAD and symmetrically censored LS estimators for two-parameter designs (i.e., estimation of an intercept and a single slope coefficient) have revealed a finite-sample bias in the means of their distributions. The reason for this bias concerns the interaction of the estimation of β_0 with the "selection rule" $x_i'\hat{\beta}_n > 0$, which determines the number of observations entering into the calculation of $\hat{\beta}_n$. A simple numerical example will illustrate the cause of the asymmetry in the distribution of $\hat{\beta}_n$. Suppose $n = 4$, $\beta_0 = (0,\ 1)'$, and the regression vector x_i takes the four values $(1,\ -2)'$, $(1,\ -1)'$, $(1,\ 1)'$, and $(1,\ 2)'$; further, suppose the error terms $\{u_i\}$ have a two-point distribution, taking the values $\frac{1}{2}$ and $-\frac{1}{2}$ with equal probability. Then, by equation (2.1), the vector $y \equiv (y_1,\ y_2,\ y_3,\ y_4)'$ of observed dependent variables will take the values $(0,\ 0,\ .5,\ 2.5)'$, $(0,\ 0,\ .5,\ 1.5)'$, $(0,\ 0,\ 1.5,\ 2.5)'$, and $(0,\ 0,\ 1.5,\ 1.5)'$ with equal probability (as illustrated in Figure 1 below). When the vector y assumes one of the first

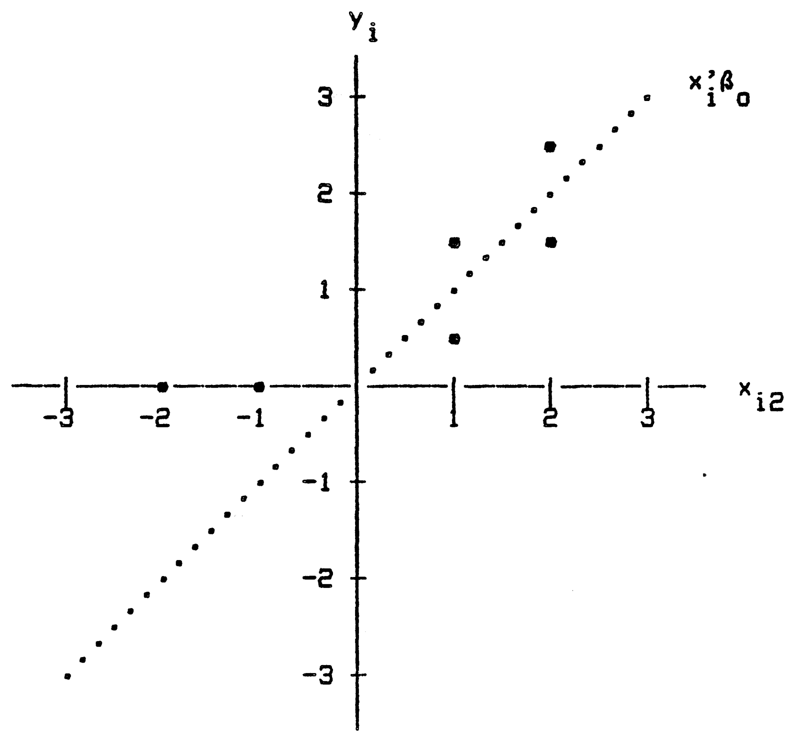

FIGURE 1

Distribution of Dependent Variable in Numerical Example

three possible values, the censored LAD or symmetrically censored regression function will pass through the two data points in the positive quadrant. However, when the last value of y is observed (with $y_3 = y_4 = 1.5$), the fitted regression line will not pass through the points in the positive quadrant (as required for unbiasedness of the estimated coefficients), since this would imply large negative residuals for observations 1 and 2 (i.e., for the data points at $(x_{i2}, y_i) = (-2, 0)$ and $(-1, 0)$). In other words, the observations on the $y_i = 0$ axis are "ignored" (i.e., $\max(0, x_i'\hat{\beta}_n) = 0$) unless the fitted regression line is unusually flat, in which case those observations cause the fitted line to be steeper than if they were ignored. As a result, the expected value of the censored LAD estimator is (−.25, 1.125) for this design, and the expected value of the symmetrically censored least squares estimator is (−.21, 1.13).

Of course, this bias vanishes in large samples: the probability limit of the symmetrically censored least squares estimator is β_0 = (0, 1) when this design is infinitely replicated, and, while the censored LAD is not consistent under these conditions (because of the discrete error distribution), it would be consistent if any continuously-distributed "noise" with zero median and positive density at zero were added to the original error terms $\{u_i\}$. Moreover, the Gaussian maximum likelihood estimator β_n^* is also biased ($E(\beta_n^*)$ = (−.25, 1.16) for this design) and is inconsistent as well ($\operatorname{plim} \beta_n^*$ = (−.10, 1.06)). Nevertheless, for smoother distributions of the regressors, the bias of the semiparametric estimators, which is of probability order smaller than $n^{-1/2}$ by the asymptotic theory, is still evident in moderately-sized samples.

The "two-step" estimator investigated here is meant to attenuate this finite-sample bias while retaining the large-sample properties of the censored regression quantile estimators. This approach is somewhat similar to a parametric "two-step" estimator (as in Heckman [1979]) which estimates the parameter vector (up to a scale term) using a bivariate probit estimator in the first stage. The idea here is to separate the classification of observations into the "$x_i'\beta_0(\theta) > 0$" and "$x_i'\beta_0(\theta) \leq 0$" groups from the estimation of the relative magnitudes of the coefficients in $\beta_0(\theta)$. Thus, the estimation proceeds in two steps: a preliminary semiparametric binary choice estimator is used to estimate which observations have $x_i'\beta_0(\theta) > 0$, and then standard regression quantile estimation is then applied to observations in this group.

For concreteness, consider the special case of LAD estimation (that is, quantile estimation based upon the $.5^{th}$ quantile, with $\beta_o(1/2) \equiv \beta_o$). The censored LAD estimator satisfies the "asymptotic moment restriction"

$$(2.6) \qquad \frac{1}{n} \sum_{i=1}^{n} 1(x_i'\hat{\beta}_n > 0) \cdot \mathrm{sgn}(y_i - x_i'\hat{\beta}_n) \cdot x_i = o_p(n^{-1/2}) \ ;$$

the left-hand side of (2.5) is the subgradient of the function $S_n(\beta)$ of (2.2) evaluated at its optimizing value. As the numerical example above indicates, if the true indicator functions were known -- so that the term "$1(x_i'\hat{\beta}_n > 0)$" could be replaced by "$1(x_i'\beta_o > 0)$" in (2.5) -- the resulting estimator would be (at least approximately) unbiased. Instead, the two-step version of (2.5) would first estimate the indicator variables $1(x_i'\beta_o > 0)$ by estimating a binary choice model for $d_i \equiv 1(y_i > 0) = 1(x_i'\beta_o + u_i > 0)$. The maximum score estimator (Manski [1975, 1984]) is designed for just this problem (i.e., the binary response model with a "zero median" restriction on the error terms); the corresponding estimator $\tilde{\beta}_n$ of β_o (more precisely, of $\beta_o/\|\beta_o\|$, since the scale of β_o is not identified using only the $\{d_i\}$) would minimize

$$(2.7) \qquad M_n(\beta) \equiv \frac{1}{n} \sum_{i=1}^{n} |d_i - 1(x_i'\beta > 0)|$$

over β (as Manski [1985] notes, this minimization problem yields the same estimator as that which maximizes the number of correct predictions of d_i by $1(x_i'\beta > 0)$). From this procedure, the "fitted values" $\tilde{d}_i \equiv 1(x_i'\tilde{\beta}_n > 0)$ would be used to determine the observations to be included in an LAD regression;

since $\max\{0, x_i'\beta\} = x_i'\beta$ when $x_i'\beta > 0$, the appropriate estimator using these fitted values minimizes

$$(2.8) \qquad \tilde{S}_n(\beta) \equiv \frac{1}{n}\sum_{i=1}^{n} \tilde{d}_i \cdot |y_i - x_i'\beta|$$

over β. (This estimator, based on the LAD criterion, is easily extended to general quantile estimation, by replacing absolute values with the $\rho_\theta(\lambda)$ function of (2.4) in each step.)

Applied to the previous numerical example, it is clear that this two-step method would yield an unbiased estimator of β_o. The first step would always yield $\tilde{d}_i = 0$ for observations with $x_{i2} \leq 0$ and $\tilde{d}_i = 1$ otherwise, so minimization of $\tilde{S}_n(\beta)$ here would always amount to LAD estimation over observations in the positive quadrant. In a more realistic case, with a denser distribution of regressors, it is reasonable to expect that the first-step estimates $\{\tilde{d}_i\}$ would be less sensitive to regressors with high leverage, since the magnitudes of the residuals $d_i - \tilde{d}_i$ are necessarily standardized across observations.

Though consistency of the first-step estimator $\tilde{\beta}_n$ of $\beta_o/\|\beta_o\|$ requires stronger restrictions than needed for consistency of the censored LAD estimator, the object of the first step is not consistent estimation of $\beta_o/\|\beta_o\|$ but rather of $\{1(x_i'\beta_o > 0)\}$; for this latter object, weaker regularity conditions suffice. The second-step estimator of β_o from (2.7) would solve the asymptotic moment restriction

$$(2.9) \qquad o_p(n^{-1/2}) = \frac{1}{n}\sum_{i=1}^{n} \tilde{d}_i \cdot \mathrm{sgn}(y_i - x_i'\beta)\cdot x_i$$

$$= \frac{1}{n}\sum_{i=1}^{n} 1(x_i'\beta_o > 0)\cdot \mathrm{sgn}(y_i - x_i'\beta)\cdot x_i$$

$$+ \frac{1}{n}\sum_{i=1}^{n} [1(x_i'\tilde{\beta}_n > 0) - 1(x_i'\beta_o > 0)]\cdot \mathrm{sgn}(y_i - x_i'\beta)\cdot x_i ,$$

so the second-stage estimator will have the same asymptotic distribution as the censored LAD estimator if the last term in this expression is also of $o_p(n^{-1/2})$. It is evident that consistency of the first-stage estimator $\tilde{\beta}_n$ is not necessary for this to be the case, as the numerical example -- for which $1(x_i'\tilde{\beta}_n > 0) \equiv 1(x_i'\beta_o > 0)$ even though $\tilde{\beta}_n$ is clearly not uniquely determined in large samples -- illustrates. In fact, the large-sample theory in the following section will impose the restriction that the regressors $\{x_i\}$ assume one of a finite set of values with probability one, implying that $\tilde{\beta}_n$ does not consistently estimate β_o by Lemma 1 of Manski [1985]; still, under this restriction, asymptotic equivalence of the censored LAD estimator and its two-step analogue will be shown below.

3. Large-Sample Theory

Restating the discussion of the previous section in terms of estimation based on a general conditional quantile of y_i, the two-step regression quantile estimator $\hat{\beta}_n(\theta)$ of $\beta_o(\theta) \equiv \beta_o + \eta_\theta \cdot e^1$ for model (2.1) above is defined to be any minimizing value of

(3.1) $$\tilde{Q}_n(\beta) \equiv \frac{1}{n}\sum_{i=1}^{n} \tilde{d}_i \cdot \rho_\theta(y_i - x_i'\beta) \equiv \tilde{Q}_n(\beta;\ \theta)$$

over all β in some parameter space $B(\theta)$, where $\tilde{d}_i \equiv 1(x_i'\tilde{\beta}_n(\theta) > 0)$ for $\tilde{\beta}_n$ any value of β in $B(\theta)$ which minimizes

(3.2) $$M_n(\beta) \equiv \frac{1}{n}\sum_{i=1}^{n} \rho_\theta(1(y_i > 0) - 1(x_i'\beta > 0)) \equiv M_n(\beta;\ \theta)$$

for $\rho_\theta(\cdot)$ defined in (2.4) above. For this estimator, the same conditions on the parameter space and the error distribution are imposed here that were used to prove consistency and asymptotic normality of the (one-step) censored regression quantile estimator (Powell [1985]).

Assumption P: The parameter vector $\beta_o(\theta)$ is an interior point of a compact parameter space $B(\theta)$.

Assumption E: The error terms u_i are independent and identically distributed random variables with distribution function $F(\lambda)$ which is continuously differentiable with density $f(\lambda)$ that is Lipschitz continuous $(|f(\lambda_1) - f(\lambda_2)| < L\cdot|\lambda_1 - \lambda_2|$ for some L) and positive at $\lambda = \eta_\theta \equiv F^{-1}(\theta)$.

As mentioned in the previous section, the regressors here will be assumed to have finite support, a restriction which is much stronger than that imposed to obtain the large sample properties of one-step censored quantile estimation. The necessity for this stronger restriction comes from the fact that, under more general conditions, the asymptotic distribution of the two-step estimator may depend upon the asymptotic distribution of the maximum score estimator $\tilde{\beta}_n$, which is as yet unknown. For example, if $x_i'\beta_0(\theta)$ is continuously distributed with positive density near zero (as imposed by Manski [1985] to demonstrate consistency of maximum score estimation), the last term in expression (2.9) above can be shown to be $O_p\left\{\sqrt{n}(\tilde{\beta}_n(\theta) - \beta_0(\theta)/\|\beta_0(\theta)\|\right\}$; since Chamberlain [1984] has shown that $\tilde{\beta}_n$ must converge at a rate slower than $\sqrt{n}$ under these conditions, continuously-distributed regressors would pose serious problems in deriving the asymptotic distribution of the two-step estimator $\hat{\beta}_n(\theta)$.

Thus, the conditions imposed on the behavior of the regressors $\{x_i\}$ are given in the following assumption.

<u>Assumption R</u>: (i) The regressors $\{x_i\}$ are elements of a finite set $\{z_j,\ j = 1, \ldots, J\}$ of possible values with probability one, where $z_j'\beta_0(\theta) \neq 0$ for all j, and are distributed independently across i and independently of $\{u_i\}$;

(ii) For each $j = 1, \ldots, J$,

$$\left[\frac{1}{n}\sum_{i=1}^{n} 1(x_i = z_j)\right] > p_0 > 0$$

for $n > n_0$, some n_0 and p_0, with probability approaching one; and

(iii) the minimum characteristic root ν_n of

$$D_n \equiv \left[\frac{1}{n}\sum_{i=1}^{n} 1(x_i'\beta_0(\theta) > 0)\cdot x_i x_i'\right]$$

has $\nu_n > \nu_0 > 0$ for $n > n_0$, some n_0 and ν_0, with probability approaching one.

Condition (i) of the preceeding assumption, in addition to specifying the finite support of the regressors, also requires $x_i'\beta_o(\theta) \neq 0$ (with probability one), which was also imposed for censored reqression quantile estimation. The second condition ensures that the fraction of observations on each possible value of the regressors is strictly positive in large samples, which is needed for the proof of the lemma that follows, while condition (iii) is the usual "identification" condition concerning variability of the regressors previously imposed for one-step quantile estimation. While these conditions imply those imposed in Powell [1985] for the censored regression model, they are neither stronger nor weaker than those imposed by Manski [1985] for the binary choice model (they are only "weaker" in the sense that they preclude consistency of the maximum score estimator).

The following lemma is essential in establishing the first-order properties of the two-step estimator $\hat{\beta}_n(\theta)$:

<u>Lemma 1</u>: Under Assumptions E and R, if $\tilde{d}_i \equiv 1(x_i'\tilde{\beta}_n(\theta) > 0)$ for $\tilde{\beta}_n$ defined to minimize $\tilde{M}_n$ of (3.2), then

$$\sum_{i=1}^{n} |\tilde{d}_i - 1(x_i'\beta_o(\theta) > 0)| = o(1)$$

almost surely.

<u>Proof</u>: For any β in $B(\theta)$, let

$\Gamma(\beta) \equiv \{b \text{ in } B(\theta): 1(z_j'\beta > 0) = 1(z_j'b > 0),\ j = 1, \ldots, J\}$;

that is, $\Gamma(\beta)$ denotes the equivalence class of coefficient vectors which yield the same sign of the regression function for all possible values of the regressors. As noted by Manski [1985], there are at most $J^K + 1$ distinct equivalence classes, where K is the dimension of β. Letting $\Gamma_0 \equiv \Gamma(\beta_0)$, the remaining distinct equivalence classes can be labelled $\Gamma_1, \ldots, \Gamma_G$, where $G \leq J^K$ is finite. Choosing a single element β_ℓ, $\ell \equiv 0, \ldots, G$, from each of these classes, it is clear that $M_n(\beta_\ell) - E[M_n(\beta_\ell)] = o(1)$ almost surely for each β_ℓ by Kolmogorov's strong law (since each term in M_n is uniformly bounded with probability one). Because the set $\{\beta_\ell, \ell = 0, \ldots, G\}$ is finite, this convergence is uniform in the $\{\beta_\ell\}$, and because $M_n(\beta) - E[M_n(\beta)$ is constant for all β in Γ_ℓ, $M_n(\beta) - E[M_n\beta)] = o(1)$ uniformly over β in $B(\theta)$, almost surely.

The uniform convergence of M_n to its expected value implies that, for any $\varepsilon > 0$, $|M_n(\beta) - E[M_n(\beta)]| < \varepsilon/2$ for all n suitably large with arbitrarily high probability. This in turn implies

$$M_n(\beta) - M_n(\beta_0(\theta)) > E[M_n(\beta)] - E[M_n(\beta_0(\theta)] - \varepsilon$$

for any $\varepsilon > 0$ and all n suitably large, with arbitrarily high probability. But

$$E[M_n(\beta) - M_n(\beta_0(\theta))] = \frac{1}{n}\sum_{i=1}^{n} 1[\mathrm{sgn}(x_i'\beta) \neq \mathrm{sgn}(x_i'\beta_0(\theta))]\cdot|\theta - \Pr(y_i = 0)|,$$

where $\mathrm{sgn}(\lambda) \equiv 1 - 2\cdot 1(\lambda \leq 0)$, by calculations analogous to those in Manski [1985], Corollary 2. Since $x_i'\beta_0(\theta)$ is bounded away from zero by Assumption R(i), and since the density function of u_i is strictly positive at $\eta_\theta \equiv F^{-1}(\theta)$,

$$|\theta - \Pr(y_i > 0)| = |\Pr(u_i < \eta_\theta) - \Pr(u_i < -x_i'\beta_o)|$$

$$= \left| \int_0^{-x_i'\beta_o(\theta)} f(\lambda + \eta_\theta)d\lambda \right| > k_o$$

for some positive k_o. Thus, for β not in Γ_o, the equivalence class containing $\beta_o(\theta)$, $1(z_j'\beta > 0) \neq 1(z_j'\beta_o(\theta) > 0)$ for at least one j in $\{1, \ldots, J\}$, so

$$E[M_n(\beta) - M_n(\beta_o(\theta))] > \sum_{j=1}^{J} k_o \cdot 1[\mathrm{sgn}(z_j'\beta) \neq \mathrm{sgn}(z_j'\beta_o(\theta))] \cdot \left[\frac{1}{n} \sum_{i=1}^{n} 1(x_i = z_j) \right]$$

$$> k_o \cdot p_o > 0$$

for all β not in Γ_o and all n suitably large. Finally, since $\tilde{\beta}_n(\theta)$ minimizes $M^n(\beta)$ for each n, $M_n(\tilde{\beta}_n(\theta)) - M_n(\beta_o(\theta)) \leq 0$, but by choosing $\varepsilon < k_o \cdot p_o$, this implies that $\tilde{\beta}_n(\theta)$ must be in Γ^o for all n suitably large, with arbitrarily high probability, by the foregoing inequalities. Since $1(x_i'\tilde{\beta}_n > 0) = 1(x_i'\beta_o(\theta) > 0)$ for all i if $\tilde{\beta}_n$ is in Γ_o, the conclusion of the lemma follows. ∎

With this preliminary result, derivation of the large-sample properties of $\hat{\beta}_n(\theta)$ under the preceeding conditions is straightforward.

<u>Theorem 1</u>: Under Assumptions E, R, and P, the two-step censored regression quantile estimator $\hat{\beta}_n(\theta)$ satisfies the asymptotic linearity relationship

$$\sqrt{n}\,(\hat{\beta}_n(\theta) - \beta_o(\theta)) = [f(\eta_\theta)]^{-1} D_n^{-1} \cdot \frac{1}{\sqrt{n}} \sum_{i=1}^{n} 1(x_i'\beta(\theta) > 0) \cdot [\theta - 1(u_i < \eta_\theta)] \cdot x_i$$

$$+ o_p(1)$$

where D_n and $f(\cdot)$ are defined in Assumptions R and E above.

Proof: First, note that $\hat{\beta}_n(\theta)$ is strongly consistent; this is because $\hat{\beta}_n(\theta)$ can be defined as the minimizer of

$$\tilde{Q}_n(\beta) - \tilde{Q}_n(\beta_o(\theta)) = \frac{1}{n} \sum_{i=1}^{n} \tilde{d}_i \cdot [\rho_\theta(y_i - x_i'\beta) - \rho_\theta(\max\{-x_i'\beta_o(\theta),\ u_i - \eta_\theta\})]$$

$$= \frac{1}{n} \sum_{i=1}^{n} 1(x_i'\beta_o(\theta) > 0) \cdot [\rho_\theta(y_i - x_i'\beta) - \rho_\theta(\max\{-x_i'\beta_o(\theta),\ u_i - \eta_\theta\})]$$

$$+ o(1) \quad \text{a. s.},$$

where the second equality follows from Lemma 1 and the boundedness of $|\rho_\theta(y_i - x_i'\beta) - \rho_\theta(\max\{-x_i'\beta_o(\theta),\ u_i - \eta_\theta\})|$ (which in turn follows from the compactness of $B(\theta)$ and the form of $\rho_\theta(\cdot)$). Since this last expression no longer involves the "first-step" estimation of $1(x_i'\beta_o(\theta) > 0)$ by $\tilde{d}_i$, the same argument given for the strong consistency of the "one-step" censored regression quantile estimators (Powell [1985], Theorem 1) is directly applicable to the present case.

To establish the asymptotic linearity relationship, define $\psi_\theta(\lambda) \equiv \theta - 1(\lambda < 0)$; ψ_θ is thus the left derivative of ρ_θ. Since $\hat{\beta}_n(\theta)$ minimizes $\tilde{Q}_n(\beta)$, it can be shown that

$$\left| \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \tilde{d}_i \cdot \psi_\theta(y_i - x_i'\hat{\beta}_n(\theta)) \cdot x_i \right|$$

$$\leq \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \tilde{d}_i \cdot 1(y_i = 0,\ x_i'\hat{\beta}_n(\theta) = 0) \cdot x_i$$

$$+ \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \tilde{d}_i \cdot 1(y_i > 0,\ y_i - x_i'\hat{\beta}_n(\theta) = 0) \cdot x_i \ ;$$

but both of these latter sums converge to zero in probability, the first because of the consistency of $\hat{\beta}_n(\theta)$ and the assumption that $x_i'\beta_o(\theta) \neq 0$ for all i, and the second because of the continuous distribution of y_i when $y_i > 0$. So

$$o_p(1) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \tilde{d}_i \cdot \psi_\theta(y_i - x_i'\hat{\beta}_n(\theta)) \cdot x_i$$

$$= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} 1(x_i'\beta_o(\theta) > 0) \cdot \psi_\theta(y_i - x_i'\hat{\beta}_n(\theta)) \cdot x_i + o_p(1),$$

where the latter equality follows from Lemma 1. Now the conditional θ-quantile of y_i given x_i is $\max\{0,\ x_i'\beta_o(\theta)\} = x_i'\beta_o(\theta)$ when $x_i'\beta_o(\theta) > 0$, and in this case the conditional distribution of y_i given x_i is continuous at $x_i'\beta_o(\theta)$, with positive density $f(\eta_\theta)$. Therefore, $\hat{\beta}_n(\theta)$ has the same asymptotic behavior as a regression quantile estimator for an uncensored linear regression model, where all regressors are chosen so that the condition $x_i'\beta_o(\theta) > 0$ is satisfied. Apart from the normalizing constant (which would be $(\Sigma_i\ 1(x_i'\beta_o(\theta) > 0))^{-1/2}$ rather than $n^{-1/2}$), the behavior of this estimator is a special case of the analysis of Koenker and Bassett [1982], and the asymptotic linearity result is proved by the same steps used to prove Theorem

3.1 of that paper. ∎

<u>Corollary 1</u>: Under Assumptions P, E, and R, the one- and two-step censored regression quantile estimators are asymptotically equivalent, i.e., $\sqrt{n}$ times their difference converges to zero in probability.

<u>Proof</u>: The same asymptotic linearity relationship given in Theorem 1 was shown for the one-step censored reqression quantile estimators in Powell [1985], under weaker conditions than those given here. ∎

4. Finite Sample Behavior: Preliminary Simulation Results

The preceeding results characterize the large-sample relationship between the one- and two-step censored regression quantile estimators under the stated conditions; however, investigation of the two-step approach was motivated not by a desire for favorable large-sample performance, but rather as a means of attenuating the finite-sample bias observed in the one-step estimators. To determine whether two-step estimation is in fact sucessful in reducing this bias, a sampling experiment conducted by Paarsch [1984] for one-step LAD estimation was replicated for a two-step analogue. As the results below indicate, the two-step estimator is much too "successful" in this respect; not only is the upward (downward) bias in the slope (intercept) estimator eliminated, but an even more substantial bias in the opposite direction is introduced by the two-step method.

Paarsch [1984] conducted several sampling experiments for the censored LAD estimator, with a variety of sample sizes and error distributions. The design replicated here took the sample size n = 200 and a single non-constant regressor which took evenly-spaced values in the interval [0, 20]. The regression function $x_i'\beta_o$ had intercept −10 and slope coefficient 1, and the error distribution was assumed to be Gaussian with variance equal to 100 (the symmetry of the regression function and the error distribution about zero implies that, on average, 50% of the observations will be censored). Even for this relatively-large sample size (relative to the number of parameters), the bias of the one-step estimators was still fairly large (as shown in Table 1 below), so the alternative behavior of two-step estimation should be well represented for this design. Note that the conditions of Assumption R above are not well approximated here -- the non-constant regressors do not repeat, contrary to Assumption R(ii) -- and this may account for the poor performance of the two-step estimator in this example.

Table 1 summarizes the behavior of the two-step estimator for 201 replications of this design; to ensure comparability of the results of the present simulation to Paarsch's, the Gaussian maximum likelihood estimator was calculated for each replication, as it was by Paarsch. As the table shows, the sampling results for Tobit maximum likelihood estimation in this study and Paarsch's are nearly identical, suggesting that comparison of results across these two studies would not be unduly influenced by differences in computation, random number generation, and the like. Because of the simple structure of the bivariate regression model in this study, the weights $\tilde{d}_i$ for the first stage estimator were calculated by direct minimization of $M_n(\beta)$ calculated for each of its $2\cdot(200 + 1) = 402$ possible values of $\tilde{d}_i$. The second-stage LAD estimates were calculated using the "projected gradient" algorithm described by Bloomfield and Stieger [1983], though for this design some numerical stability problems were encountered.

The results for the two-step LAD estimator tabulated in the table clearly show a lack of the upward bias in the estimation of the slope coefficient;

indeed, the downward bias in this coefficient is substantially larger in magnitude than the bias in the one-step procedure. It appears that the interaction of the "$1(x_i'\beta > 0)$" selection rule and the estimation of the slope coefficients works in exactly the opposite direction for two-step as for one-step estimation -- the first step selects observations for which y_i is very likely to be positive, inducing a positive bias in the intercept term (and a corresponding downward bias in the slope) in the second stage.The variability of the number of observations with $\tilde{d}_i = 1$ is also quite substantial, as the last entry of the table shows. This indicates the numerical example of section 2 above was misleading to the extent that it suggested these indicator variables could be very precisely estimated in the first step. Even if the indicators $1(x_i'\beta_o(\theta) > 0)$ were known exactly, though, the second step of the two-step LAD procedure has a downward bias in the slope coefficient estimate, as the results of LAD estimation using the exact indicator variables suggest. Finally, it should be noted that, when multiple minima of M_n occurred, the procedure adopted here chose that set of $\tilde{d}_i$ for which $\Sigma_i \tilde{d}_i$ was largest (i.e., for which the largest possible number of observations were included in the second stage). The bias and variability of the second-stage estimator increased if that set with the smallest $\Sigma_i \tilde{d}_i$ were used instead, so that the results reported here represent the more favorable of two possible implementations of the two-step estimator.

5. Conclusions

Given the poor performance of the two-step procedure in the sampling experiment above, it would not be recommended for practical applications, at least for designs where it was suspected that the regression function was near zero (that is, the probability of censoring is close to θ, if estimation of the conditional θ-quantile is the object), as it is in the experiment above. While the intuition regarding the direction of mean bias for the two-step procedure appears to have been validated, the approach has not been shown to be an improvement over the previous one-step quantile estimation strategy. Further sampling experiments will be conducted the determine whether the behavior of the two-step estimator substantially improves for designs which more closely approximate Assumption R of section 3; however, for such designs the behavior of the one-step estimators will also need to be studied, since their finite sample performance may improve for such designs as well.

TABLE 1

Sampling Behavior of the One- and Two-Step Censored LAD Estimators

Censored LAD Estimator [Paarsch]

	True	Mean	S.D.	L.Qt.	Median	U.Qt.
Intercept	-10.00	-12.48	7.58	-15.55	-10.72	-7.04
Slope	1.00	1.15	0.47	0.78	1.04	1.37

Gaussian Maximum Likelihood [Paarsch]

	True	Mean	S.D.	L.Qt.	Median	U.Qt.
Intercept	-10.00	-10.26	2.03	-11.87	-10.32	-8.95
Slope	1.00	1.02	0.15	0.91	1.00	1.11

Gaussian Maximum Likelihood

	True	Mean	S.D.	L.Qt.	Median	U.Qt.
Intercept	-10.00	-10.22	2.09	-11.87	-9.84	-8.68
Slope	1.00	1.02	0.16	0.92	1.02	1.13

Two-Step Censored LAD

	True	Mean	S.D.	L.Qt.	Median	U.Qt.
Intercept	-10.00	-4.36	7.58	-7.60	-4.76	-1.34
Slope	1.00	0.66	0.49	0.45	0.66	0.91

LAD with True $d_i \equiv 1(x_i'\beta_o(\theta) > 0)$

	True	Mean	S.D.	L.Qt.	Median	U.Qt.
Intercept	-10.00	-8.06	4.71	-11.36	-9.02	-5.57
Slope	1.00	0.89	0.34	0.70	0.94	1.12

Sum of $\tilde{d}_i \equiv 1(x_i'\tilde{\beta}_n(\theta) > 0)$

	True	Mean	S.D.	L.Qt.	Median	U.Qt.
Intercept	100.00	102.00	20.29	89.00	102.00	116.00

REFERENCES

Bloomfield, P. and W.L. Steiger [1983], Least Absolute Deviations: Theory, Applications, and Algorithms. Boston: Birkhauser.

Chamberlain, G. [1984], "Asymptotic efficiency in semiparametric models with censoring," manuscript, Department of Economics, University of Wisconsin-Madison, May.

Cosslett, S.R. [1983], "Distribution-free maximum likelihood estimator of the binary choice model," Econometrica, 51: 765-782.

Heckman, J. [1979], " Sample selection bias as a specification error," Econometrica, 47: 153-161.

Koenker, R. and G.S. Bassett Jr. [1978], "Regression quantiles," Econometrica, 46: 33-50.

Koenker, R. and G.S. Bassett Jr. [1982], "Robust tests for heteroscedasticity based on regression quantiles," Econometrica, 50: 43-61.

Manski, C.F. [1975], "Maximum score estimation of the stochastic utility model of choice," Journal of Econometrics, 3: 205-228.

Manski, C.F. [1984], "Semiparametric analysis of discrete response: asymptotic properties of the maximum score estimator," University of Wisconsin-Madison, SSRI Workshop Paper No. 8413.

Obenhofer, W. [1982], "The consistency of nonlinear regression minimizing the ℓ_1 norm," Annals of Statistics, 10: 316-319.

Paarsch, H.J. [1984], "A Monte Carlo comparison of estimators for censored regression models," Journal of Econometrics, 24: 197-213.

Powell, J.L. [1983a], "Symmetrically trimmed least squares estimation of tobit models," manuscript, Department of Economics, MIT, August.

Powell, J.L. [1984], "Least absolute deviations estimation for the censored regression model," Journal of Econometrics, 25: 303-325.

Powell, J.L. [1985], "Censored regression quantiles," manuscript, Department of Economics, MIT, January.

Womersley, R.S. [1983], "Censored discrete linear ℓ_1 approximation", Mathematical Sciences Research Centre, Austrailian National University, Research Report No. CMA-R42-83.