working paper
department
of economics

SYMMETRICALLY TRIMMED LEAST SQUARES ESTIMATION
FOR TOBIT MODELS

M.I.T. Working Paper #365
by
James L. Powell
August, 1983 revised January, 1985

massachusetts
institute of
technology

50 memorial drive
cambridge, mass. 02139

SYMMETRICALLY TRIMMED LEAST SQUARES ESTIMATION FOR TOBIT MODELS

M.I.T. Working Paper #365

by

James L. Powell

August, 1983 revised January, 1985

SYMMETRICALLY TRIMMED LEAST SQUARES ESTIMATION FOR TOBIT MODELS

by

James L. Powell

Department of Economics

Massachusetts Institute of Technology

August, 1983

Revised January, 1985

ABSTRACT

This paper proposes alternatives to maximum likelihood estimation of the censored and truncated regression models (known to economists as "Tobit" models). The proposed estimators are based on symmetric censoring or truncation of the upper tail of the distribution of the dependent variable. Unlike methods based on the assumption of identically distributed Gaussian errors, the estimators are consistent and asymptotically normal for a wide class of error distributions and for heteroscedasticity of unknown form. The paper gives the regularity conditions and proofs of these large sample results, demonstrates how to construct consistent estimators of the asymptotic covariance matrices, and presents the results of a simulation study for the censored case. Extensions and limitations of the approach are also considered.

SYMMETRICALLY TRIMMED LEAST SQUARES ESTIMATION

FOR TOBIT MODELS[1/]

by

James L. Powell

1. Introduction

Linear regression models with nonnegativity of the dependent variable--known to economists as "Tobit" models, due to Tobin's [1958] early investigation--are commonly used in empirical economics, where the nonnegativity constraint is often binding for prices and quantities studied. When the underlying error terms for these models are known to be normally distributed (or, more generally, have distribution functions with a known parametric form), maximum likelihood estimation and other likelihood-based procedures yield estimators which are consistent and asymptotically normally distributed (as shown by Amemiya [1973] and Heckman [1979]). However, such results are quite sensitive to the specification of the error distribution; as demonstrated by Goldberger [1980] and Arabmazar and Schmidt [1982], likelihood-based estimators are in general inconsistent when the assumed parametric form of the likelihood function is incorrect. Furthermore, heteroscedasticity of the error terms can also cause inconsistency of the parameter estimates even when the shape of the error density is correctly specified, as Hurd [1979], Maddala and Nelson [1975], and Arabmazar and Schmidt [1981] have shown.

Since economic reasoning generally yields no restrictions concerning the form of the error distribution or homoscedasticity of the residuals, the sensititivy of likelihood-based procedures to such assumptions is a serious concern; thus, several strategies for consistent estimation without these assumptions have been investigated. One class of estimation methods combines the likelihood-based approaches with nonparametric estimation of the distribution function of the residuals; among these are the proposals of Miller [1976], Buckley and James [1979], Duncan [1983], and Fernandez [1983]. At present, the large-sample distributions of these estimators are not established, and the methods appear sensitive to the assumption of identically distributed error terms. Another approach was adopted in Powell [1984], which used a least absolute deviations criterion to obtain a consistent and asymptotically normal estimator of the regression coefficient vector. As with the methods cited above, this approach applies only to the "censored" version of the Tobit model (defined below), and, like some of the previous methods, involves estimation of the density function of the error terms (which introduces some ambiguity concerning the appropriate amount of "smoothing"). [2/]

In this study, estimators for the "censored" and "truncated" versions of the Tobit model are proposed, and are shown to be consistent and asymptotically normally distributed under suitable conditions. The approach is not completely general, since it is based upon the assumption of symmetrically (and independently) distributed error terms; however, this assumption is often reasonable if the data have been appropriately transformed, and is far more general than the assumption of normality usually imposed. The estimators will also be consistent if the residuals are not identically distributed; hence, they are robust to (bounded) heteroscedasticity of unknown form.

In the following section, the censored and truncated regression models are defined, and the corresponding "symmetrically trimmed" estimators are proposed and discussed. Section 3 gives sufficient conditions for strong consistency and asymptotic normality of the estimators, and proposes consistent estimators of the asymptotic covariance matrices for purposes of inference, while section 4 presents the results of a small scale simulation study of the censored "symmetrically trimmed" estimator. The final section discusses some generalizations and limitations of the approach adopted here. Proofs of the main results are given in an appendix.

2. Definitions and Motivation

In order to make the distinction between a censored and a truncated regression model, it is convenient to formulate both of these models as appropriately restricted version of a "true" underlying regression equation

$$(2.1)\quad y_t^* = x_t'\beta_o + u_t, \qquad t = 1, \ldots, T,$$

where x_t is a vector of predetermined regressors, β_o is a conformable vector of unknown parameters, and u_t is a scalar error term. In this context a censored regression model will apply to a sample for which only the values of x_t and $y_t = \max\{0, y_t^*\}$ are observed, i.e.,

$$(2.2)\quad y_t = \max\{0, x_t'\beta_o + u_t\}, \qquad t = 1, \ldots, T.$$

The censored regression model can alternatively be viewed as a linear regression model for which certain values of the dependent variable are "missing"--namely, those values for which $y_t^* < 0$. To these "incomplete" data points, the value zero is assigned to the dependent variable. If the error term u_t is continuously distributed, then, the censored dependent variable y_t will be continuously distributed for some fraction of the sample, and will assume the value zero for the remaining observations.

A truncated regression model corresponds to a sample from (2.1) for which the data point (y_t^*, x_t') is observed only when $y_t^* > 0$; that is, no information about data points with $y_t^* < 0$ is available. The truncated regression model thus applies to a sample in which the observed dependent variable is generated from

$$(2.3) \quad y_t = x_t'\beta_o + v_t, \qquad t = 1, \ldots, T,$$

where x_t and β_o are defined above and v_t has the conditional distribution of u_t given $u_t > -x_t'\beta_o$. That is, if the conditional density of u_t given x_t is $f(\lambda | x_t, t)$, the error term v_t will have density

$$(2.4) \quad g(\lambda | x_t, t) \equiv 1(\lambda < -x_t'\beta_o) f(\lambda | x_t, t) \left[\int_{-x_t'\beta_o}^{\infty} f(\lambda | x_t, t) d\lambda\right]^{-1},$$

where "1(A)" denotes the indicator function of the event "A" (which takes the value one if "A" is true and is zero otherwise).

If the underlying error terms $\{u_t\}$ were symmetrically distributed about zero, and if the latent dependent variables $\{y_t^*\}$ were observable, classical least squares estimation (under the usual regularity conditions) would yield consistent estimates of the parameter vector β_o. Censoring or truncation of the dependent variable, however, introduces an asymmetry in its distribution which is systematically related to the regressors. The approach taken here is to symmetrically censor or truncate the dependent variable in such a way that symmetry of its distribution about $x_t'\beta_o$ is restored, so that least squares will again yield consistent estimators.

To make this approach more explicit, consider first the case in which the dependent variable is truncated at zero. In such a truncated sample, data points for which $u_t < -x_t'\beta_o$ are omitted; but suppose data points with $u_t > x_t'\beta_o$ were also excluded from the sample. Then any observations for which $x_t'\beta_o < 0$ would automatically be deleted, and any remaining observations would have error terms lying within the interval $(-x_t'\beta_o, x_t'\beta_o)$. Because of the symmetry of the distribution of the original error terms, the residuals for the "symmetrically truncated" sample will also be symmetrically distributed about zero; the corresponding dependent variable would take values between zero and $2x_t'\beta_o$, and would be symmetrically distributed about $x_t'\beta_o$, as illustrated in Figure 1.

Thus, if β_o were known, a consistent estimator for β_o could be obtained as a solution (for β) to the set of "normal equations"

$$(2.5)\quad 0 = \sum_{t=1}^{T} 1(v_t < x_t'\beta_o)(y_t - x_t'\beta)x_t$$

$$= \sum_{t=1}^{T} 1(y_t < 2x_t'\beta_o)(y_t - x_t'\beta)x_t.$$

Of course, β_o is not known, since it is the object of estimtion; nonetheless, using a "self consistency" (Efron [1967]) argument which uses the desired estimator to perform the "symmetric truncation," an estimator $\tilde{\beta}_T$ for the truncated regression model can be defined to satisfy the asymptotic "first-order condition"

$$(2.6)\quad o_p(T^{1/2}) = \sum_{t=1}^{T} 1(y_t < 2x_t'\beta)(y_t - x_t'\beta)x_t,$$

which sets the sample covariance of the regressors $\{x_t\}$ and the "symmetrically trimmed" residuals $\{1(-x_t'\beta < y_t - x_t'\beta < x_t'\beta)\cdot(y_t - x_t'\beta)\}$ equal to zero (asymptotically).[3/] Corresponding to this implicit equation is a minimization problem; letting $(-\frac{1}{2})\partial R_T(\beta)/\partial\beta$ denote the (discontinuous) right-hand side of (2.6), the form of the function $R_T(\beta)$ is easily found by integration:

$$(2.7)\quad R_T(\beta) \equiv \sum_{t=1}^{T} (y_t - \max\{\tfrac{1}{2}y_t, x_t'\beta\})^2.$$

Thus the symmetrically truncated least squares estimator $\tilde{\beta}_T$ of β_o is defined as any value of β which minimizes $R_T(\beta)$ over the parameter space B (assumed compact).

Definition of the symmetrically trimmed estimator for a censored sample is similarly motivated. The error terms of the censored regression model are of the form $u_t^* = \max\{u_t, -x_t'\beta_o\}$, so "symmetric censoring" would replace u_t^* with $\min\{u_t^*, x_t'\beta_o\}$ whenever $x_t'\beta_o > 0$, and would delete the observation otherwise. Equivalently, the dependent variable y_t would be replaced with $\min\{y_t, 2x_t'\beta_o\}$, as shown in Figure 2; the resulting "normal equations"

$$(2.8)\quad 0 = \sum_{t=1}^{T} 1(x_t'\beta_o > 0)(\min\{y_t,\ 2x_t'\beta_o\} - x_t'\beta)x_t$$

for known β_o would be modified to

$$(2.9)\quad 0 = \sum_{t=1}^{T} 1(x_t'\beta > 0)(\min\{y_t,\ 2x_t'\beta\} - x_t'\beta)x_t,$$

since β_o is unknown. An objective function which yields this implicit equation is

$$(2.10)\quad S_T(\beta) \equiv \sum_{t=1}^{T} (y_t - \max\{\tfrac{1}{2} y_t,\ x_t'\beta\})^2 + \sum_{t=1}^{T} 1(y_t > 2x_t'\beta)[(\tfrac{1}{2} y_t)^2 - (\max\{0,\ x_t'\beta\})^2].$$

The symmetrically censored least squares estimator $\ddot{\beta}_T$ of β_o is thus defined to be any value of β minimizing $S_T(\beta)$ over the parameter space B.

The reason $\hat{\beta}_T$ and $\tilde{\beta}_T$ were defined to be minimizers of $S_T(\beta)$ and $R_T(\beta)$, rather than solutions to (2.6) and (2.9, respectively, is the multiplicity of (inconsistent) solutions to these latter equations, even as $T \to \infty$. The "symmetric trimming" approach yields nonconvex minimization problems, (due to the "$1(x_t'\beta > 0)$" term) so that multiple solutions to the "first-order conditions" will exist (for example, $\beta = 0$ always satisfies (2.6) and (2.9)). Nevertheless, under the regularity conditions to be imposed below, it will be shown that the global minimizers of (2.7) and (2.10) above are unique and close to the true value βo with high probability as $T \to \infty$.

It is interesting to compare the "symmetric trimming" technique for the censored regression model to the "EM algorithm" for computation of maximum likelihood estimates when the parametric form of the error distribution is assumed known.[4/] This method computes the maximum likelihood estimate β^*_T of β_0 as an interative solution to the equation

$$(2.11)\quad \beta^*_T = [\sum_t x_t x_t']^{-1} \sum_t y^*_t(\beta^*_T) x_t$$

where

$$(2.12)\quad y^*_t(\beta) \equiv \begin{cases} y_t & \text{if} \quad y_t > 0 \\ E(y^*_t = x_t'\beta + u_t \mid u_t < -x_t'\beta) & \text{otherwise.} \end{cases}$$

That is, the censored observations, here treated as missing values of y^*_t, are replaced by the conditional expectations of y^*_t evaluated at the estimated value β^*_T (as well as other estimates of unknown nuisance parameters). It is the calculation of the functional form of this conditional expectation that requires knowledge of the shape of the error density. On the other hand, the symmetrically censored least squares estimator $\hat{\beta}_T$ solves

$$(2.13)\quad \hat{\beta}_T = [\sum_t 1(x_t'\hat{\beta}_T > 0) x_t x_t']^{-1} \sum_t 1(x_t'\hat{\beta}_T > 0) \cdot \min\{y_t, 2x_t'\hat{\beta}_T\} x_t;$$

instead of replacing censored values of the dependent variable in the lower tail of its distribution, the "symmetric trimming" method replaces uncensored observations in the upper tail by their estimated "symmetrically censored" values.

The rationale behind the symmetric trimming approach makes it clear that consistency of the estimators will require neither homoscedasticity nor known distributional form of the error terms. However, symmetry of the error distribution is clearly essential; furthermore, the true regression function $x_t'\beta_o$ must satisfy stronger regularity conditions than are usually imposed for the estimators to be consistent. Since observations with $x_t'\beta_o < 0$ are deleted from the sample (because trimming the "upper tail" in this case amounts to completely eliminating the observations), consistency will require that $x_t'\beta_o > 0$ for a positive fraction of the observations, and that the corresponding regressors x_t be sufficiently variable to identify β_o uniquely. Also, for the truncated regression model the underlying error terms $\{u_t\}$ must be restricted to have a distribution which is unimodal as well as symmetric. Such a condition is needed to ensure the uniqueness of the value $\tilde{\beta}_T$ which minimizes $R_T(\beta)$ as $T \to \infty$. Figure 3 illustrates the problems that may arise if this assumption is not satisfied; for the bimodal case pictured in the upper panel, omitting data points for which $y_t < 2x_t'\beta^*$ will yield a symmetric distribution about $x_t'\beta^*$, just as trimming by the rule $y_t < 2x_t'\beta_o$ yields a symmetric distribution about $x_t'\beta_o$ (in fact, for the case in which $x_t \equiv 1$, the minimum of the limiting function $R_T(\beta)$ as $T \to \infty$ will occur at β^*, not at the true value β_o). When the error terms are uniformly distributed, as pictured in the lower panel of Figure 3, there is clearly a continuum of possible values of $x_t'\beta$ for which truncation of data points with $y_t > 2x_t'\beta$ would yield a symmetric distribution about $x_t'\beta$; thus such cases will be ruled out in the regularity conditions given below.

3. Large Sample Properties of the Estimators

The asymptotic theory for the symmetrically trimmed estimators will first be developed for the symmetrically censored least squares estimator $\hat{\beta}_T$. Thus, for the censored regression model (2.2), the following conditions on β_o, x_t, and u_t

are imposed:

Assumption P: The true parameter vector β_0 is an interior point of a compact parameter space B.

Assumption R: The regressors $\{x_t\}$ are independently distributed random vectors with $E\,\|x_t\|^{4+\eta} < K_0$ for some positive K_0 and η, and ν_T, the minimum characteristic root of the matrix

$$N_T \equiv \frac{1}{T}\sum_t E[1(x_t'\beta_0 > \varepsilon_0)x_t x_t'],$$

has $\nu_T > \nu_0$ whenever $T > T_0$, some positive ε_0, ν_0, and T_0.

Assumption E1: The error terms $\{u_t\}$ are mutually independently distributed, and, conditionally on x_t, are continuously and symmetrically distributed about zero, with densities which are bounded above and positive at zero, uniformly in t. That is, if $F(\lambda|x_t, t) \equiv F_t(\lambda)$ is the conditional c.d.f. of u_t given x_t, then $dF_t(\lambda) = f_t(\lambda)d\lambda$, where $f_t(\lambda) = f_t(-\lambda)$, $f_t(\lambda) < L_0$, and $f_t(\lambda) > \zeta_0$ whenever $|\lambda| < \zeta_0$, some positive L_0 and ζ_0.

Assumption P is standard for the asymptotic distribution theory of extremum estimators; only the compactness of B is used in the proof of consistency, but the argument for asymptotic normality requires β_0 to be in the interior of B. The assumptions on the regressors are more unusual; however, because the $\{x_t\}$ are not restricted to be identically distributed, the assumption of their mutual independence is less restrictive than might first appear. In particular, the condition will be satisfied if the regressors take arbitrarily fixed values (with probability one), provided they are uniformly bounded. The condition on the minimum characteristic root of N_T is the essential "identification condition" discussed in the previous section.

Continuity of the error distributions is assumed for convenience; it suffices that the distributions be continuous only in a neighborhood of zero (uniformly in t). The bounds on the density function $f_t(\lambda)$ essentially require the heteroscedasticity of the errors to be bounded away from zero and infinity (taking the inverse of the conditional density at zero to be the relevant scale parameter), but these restrictions can also be weakened; with some extra conditions on the regressors, uniform positivity of the error densities at zero can be replaced by the condition

$$\frac{1}{T} \sum_{t=1}^{T} \Pr(|u_t| \leqq \varepsilon) > K > 0$$

for some K and ε in $(0, \varepsilon_0)$, with ε_0 defined in Assumption R. Finally, note that no assumption concerning the existence of moments of u_t is needed; since the range of the "symmetrically trimmed" dependent variable is bounded by linear functions of x_t, it is enough that the regressors have sufficiently bounded moments.

With these conditions, the following results can be established:

<u>Theorem 1</u>: For the censored regression model (2.2) under Assumptions P, R, and E1, the symmetrically censored least squares estimator $\hat{\beta}_T$ is strongly consistent and asymptotically normal; that is, if $\hat{\beta}_T$ minimizes $S_T(\beta)$, defined in (2.10), then

(i) $\lim_{T\to\infty} \hat{\beta}_T = \beta_o$ almost surely, and

(ii) $D_T^{-1/2} C_T \cdot \sqrt{T}(\hat{\beta}_T - \beta_o) \overset{d}{\to} N(0, I)$,

where

$$C_T \equiv \frac{1}{T} \sum_t E[1(-x_t'\beta_o < u_t < x_t'\beta_o)x_t x_t']$$

and $D_T^{-1/2}$ is any square root of the inverse of

$$D_T \equiv \frac{1}{T} \sum_t E[1(x_t'\beta_o > 0)\cdot \min\{u_t^2, (x_t'\beta_o)^2\}x_t x_t'].^{5/}$$

The proof of this and the following theorems are given in the mathematical appendix below.

Turning now to the truncated regression model (2.3)~(2.4), additional restrictions must be imposed on the error distributions to show consistency and asymptotic normality of the symmetrically trimmed estimator $\tilde{\beta}_T$:

Assumption E2: In addition to the conditions of Assumption E1, the error densities $\{f_t(\lambda)\}$ are unimodal, with strict unimodality, uniformly in t, in some neighborhood of zero. That is, $f_t(\lambda_2) \geq f_t(\lambda_1)$ if $|\lambda_1| \geq |\lambda_2|$, and there exists some $\alpha_o > 0$ and some function $h(\lambda)$, strictly decreasing in $[0, \alpha_o]$, such that $f_t(\lambda_2) - f_t(\lambda_1) \geq h(|\lambda_2|) - h(|\lambda_1|)$ when $|\lambda_2| < |\lambda_1| < \alpha_o$.

With these additional conditions, results similar to those in Theorem 1 can be established for the truncated case.

Theorem 2: For the truncated regression model (2.3)~(2.4) under Assumptions P, R, and E2, the symmetrically truncated least squares estimator $\tilde{\beta}_T$ is strongly consistent and asymptotically normal; that is, if $\tilde{\beta}_T$ minimizes $R_T(\beta)$, defined in (2.7), then

(i) $\lim_{T\to\infty} \tilde{\beta}_T = \beta_o$ almost surely, and

(ii) $Z_T^{-1/2}(W_T - V_T)\cdot\sqrt{T}(\tilde{\beta}_T - \beta_o) \xrightarrow{d} N(0, I)$, where

$$W_T \equiv \frac{1}{T}\sum_t E[1(-x_t'\beta_o < v_t < x_t'\beta_o)x_t x_t'],$$

$$V_T \equiv \frac{1}{T}\sum_t E[1(x_t'\beta_o > 0)\left[\frac{2(x_t'\beta_o)f_t(x_t'\beta_o)}{F_t(x_t'\beta_o)}\right] x_t x_t'],$$

and $Z_T^{-1/2}$ is any square root of the inverse of

$$Z_T \equiv \frac{1}{T}\sum_t E[1(-x_t'\beta_o < v_t < x_t'\beta_o)v_t^2\, x_t x_t'].$$

In order to use the asymptotic normality of $\hat{\beta}_T$ and $\tilde{\beta}_T$ to construct large sample hypothesis tests for the parameter vector β_o, consistent estimators of the asymptotic covariance matrices must be derived. For the censored regression model, "natural" estimators of relevant matrices C_T and D_T exist; they are

(3.1) $\hat{C}_T \equiv \frac{1}{T}\sum_t 1(0 < y_t < 2x_t'\hat{\beta}_T)x_t x_t'$

$$\equiv \frac{1}{T} \sum_t 1(-x_t'\hat{\beta}_T < \hat{u}_t < x_t'\hat{\beta}_T)x_t x_t'$$

and

$$(3.2) \quad \hat{D}_T \equiv \frac{1}{T} \sum_t 1(x_t'\hat{\beta}_T > 0)\cdot\min\{\hat{u}_t^2,\ (x_t'\hat{\beta}_T)^2\}x_t x_t',$$

where as usual, $\hat{u}_t \equiv y_t - x_t'\hat{\beta}_T$.

These estimators are consistent with no further conditions on the censored regression model.

<u>Theorem 3</u>: Under the conditions of Theorem 1, the estimators $\hat{C}_T$ and $\hat{D}_T$, defined in (3.1) and (3.2) are strongly consistent, i.e., $\hat{C}_T - C_T = o(1)$ and $\hat{D}_T - D_T = o(1)$ almost surely.

For the truncated regression model, estimation of the matrices W_T and Z_T is also straightforward:

$$(3.3) \quad \tilde{W}_T \equiv \frac{1}{T} \sum_t 1(y_t < 2x_t'\tilde{\beta}_T)x_t x_t'$$

$$\equiv \frac{1}{T} \sum_t 1(-x_t'\tilde{\beta}_T < \tilde{v}_t < x_t'\tilde{\beta}_T)x_t x_t'$$

and

$$(3.4) \quad \tilde{Z}_T \equiv \frac{1}{T} \sum_t 1(-x_t'\tilde{\beta}_T < \tilde{v}_T < x_t'\tilde{\beta}_T)\tilde{v}_t^2\ x_t x_t',$$

for $\tilde{v}_t \equiv y_t - x_t'\tilde{\beta}_T$. It is only the estimation of the matrix V_T that is problematic, since this involves estimation of the density functions of the error terms. Since density function estimation necessarily involves some ambiguity about the amount of "smoothing" to be applied to the empirical distribution function of the residuals, a single "natural" estimator of V_T cannot be defined.

Instead, a class of consistent estimators can be constructed by combining kernel density estimation with the heteroscedasticity-consistent covariance matrix estimators of Eicker [1967] and White [1980]. Specifically, let

$$(3.5)\quad \tilde{V}_T \equiv \frac{1}{T}\sum_t 1(x_t'\tilde{\beta}_T > 0)(x_t'\tilde{\beta}_T) \cdot \tilde{c}_T^{-1}[1(y_t < \tilde{c}_T) + 1(2x_t'\tilde{\beta}_T < y_t < 2x_t'\tilde{\beta}_T+\tilde{c}_T)]x_t x_t'$$

$$\equiv \frac{1}{T}\sum_t 1(x_t'\tilde{\beta}_T > 0)(x_t'\tilde{\beta}_T) \cdot \tilde{c}_T^{-1}[1(-x_t'\tilde{\beta}_T < \tilde{v}_t < -x_t'\tilde{\beta}_T+\tilde{c}_T) + 1(x_t'\tilde{\beta}_T < \tilde{v}_t < x_t'\tilde{\beta}_T+\tilde{c}_T)]x_t x_t',$$

there $\tilde{v}_t \equiv y_t - x_t'\tilde{\beta}_T$ as before and $\tilde{c}_T$ is a (possibly stochastic)[6/] sequence to be specified more fully below. Heuristically, the density functions of the $\{v_t\}$ at $\{x_t'\beta_o\}$—that is, $\{f_t(x_t'\beta_o)/F_t(x_t'\beta_o)\}$--are estimated by the fraction of residuals $\tilde{v}_t$ falling into the intervals $(-x_t'\tilde{\beta}_T, -x_t'\tilde{\beta}_T + \tilde{c}_T)$ and $(x_t'\tilde{\beta}_T, x_t'\tilde{\beta}_T + \tilde{c}_T)$, divided by the width of these intervals, $\tilde{c}_T$.

Of course, different sequences $\{\tilde{c}_T\}$ of smoothing constants will yield different estimators of V_T; thus, some conditions must be imposed on $\{\tilde{c}_T\}$ for the corresponding estimator to be consistent. The following condition will suffice:

<u>Assumption W</u>: Corresponding to the sequence of interval widths $\{\tilde{c}_T\}$ is some monstochastic sequence $\{c_T\}$ such that $(1 - \tilde{c}_T/c_T) = o_p(1)$ and $c_T = o(1)$, $c_T^{-1} = o(\sqrt{T})$.

That is, the sequence $\{\tilde{c}_T\}$ tends to zero (in probability) at a rate slower than $T^{-1/2}$. With this condition, as well as a stronger condition on the moments of the regressors, the estimator $\tilde{V}_T$ can be shown to be (weakly) consistent.

<u>Theorem 4</u>: Under the conditions of Theorem 2, the estimators $\tilde{W}_T$ and $\tilde{Z}_T$, defined in (3.3) and (3.4), are strongly consistent, i.e., $\tilde{W}_T - W_T = o(1)$ and $\tilde{Z}_T - Z_T = o(1)$ almost surely. In addition, if Assumption R holds with $\eta = 3$ and if Assumption W holds, the estimator $\tilde{V}_T$ is weakly consistent, i.e., $\tilde{V}_T - V_T = o_p(1)$.

Of course, the $\{\tilde{c}_T\}$ sequence must still be specified for the estimator $\tilde{V}_T$ to be operational. Even restricting the sequence to be of the form

$$(3.6) \quad \tilde{c}_T = (c_0 T^{-\gamma})\tilde{\sigma}_T, \qquad c_0 > 0, \qquad 0 < \gamma < .5,$$

for $\tilde{\sigma}_T$ some estimator of the scale of $\{v_t\}$, the constants c_0 and γ must be chosen in some manner. Also, there is no guarantee that the matrix $(\tilde{W}_T - \tilde{V}_T)$ will be positive definite for finite T, though the associated covariance matrix estimator, $(\tilde{W}_T - \tilde{V}_T)^{-1}\tilde{Z}_T(\tilde{W}_T - \tilde{V}_T)^{-1}$, should be positive definite in general.

Thus, further research on the finite sample properties of $\tilde{V}_T$ is needed before it can be used with confidence in applications.

Given the conditions imposed in Theorems 1 through 4 above, normal sampling theory can be used to construct hypothesis tests concerning β_0 without prior knowledge of the likelihood function of the data. In addition, the estimators $\hat{\beta}_T$ and $\tilde{\beta}_T$ can be used to test the null hypothesis of homoscedastic, normally-distributed error terms by comparing these estimators with the corresponding maximum likelihood estimators, as suggested by Hausman [1978]. Of course, rejection of this null hypothesis may occur for reasons which cause the symmetrically-trimmed estimators to be inconsistent as well--e.g., asymmetry of the error distribution, or misspecification of the regression function.[7/]

4. Finite Sample Behavior of the Symmetrically Censored Estimator

Having established the limiting behavior of the symmetrically trimmed estimators under the regularity conditions imposed above, it is useful next to consider the degree to which these large sample results apply to sample sizes likely to be encountered in practice. To this end, a small-scale simulation study of a bivariate regression model was conducted, using a variety of assumptions concerning sample size, degree of censoring, and error distribution. It is, of course, impossible to completely characterize the finite-sample distribution of $\hat{\beta}_T$ under the weak conditions imposed above; nonetheless, the behavior of $\hat{\beta}_T$ in a few special cases may serve as a guide to its finite sample behavior, and perhaps as well to the sampling behavior of the symmetrically truncated estimator $\tilde{\beta}_T$.

The "base case" simulations took the dependent variable y_t to be generated from equation (2.2), with the parameter vector β_0 having two components (an intercept term of zero and a slope of one), the sample size T=200, and the error terms being genereated as i.i.d. standard Gaussian variates.[8/] In this base case, the regression vectors were of the form $x_t'=(1,z_t)$, where z_t assumed evenly-spaced values in the interval $[-b,b]$, where b was chosen to set the sample variance of z_t equal to one (i.e., $b\approx1.7$). While no claim is made that this design is "typical" in applications of the censored regression model, certain aspects do correspond to data configurations encountered in practice -- specifically, the relatively low "signal-to-noise ratio" (for uncensored data, the overall R^2 is .5, and this falls to .2 for the subsample with $x_t'\beta_0 > 0$) and the moderately large number of observations per parameter.

For this design, and its variants, the symmetrically censored least squares (SCLS) estimator $\hat{\beta}_T$ and the Tobit maximum likelihood estimator (TOBIT MLE) for an assumed Gaussian error distribution were calculated for 201 replications of the experiment. The results are summarized under the heading "Design 1" in Table 1 below. The other entries of Table 1 summarize the results for other designs which maintain the homoscedastic Gaussian error distribution. For each design, the "true" parameter values are reported, along with the sample means, standard deviations, root mean-squared errors (RMSE), lower, middle, and upper quartiles (LQ, Median, and UQ), respectively, and median absolute errors (MAE) for the 201 replications.

One feature of the tabulated results which is not surprising is the relative efficiency of the Gaussian MLE (which is the "true" MLE for these designs) to the SCLS estimator for all entries in the table. Comparison of the SCLS estimator to the MLE using the RMSE criterion is considerably less favorable to SCLS than the same comparison using the MAE criterion, suggesting that the tail behavior of the SCLS estimator is an important determinant of this efficiency. The magnitude of the relative efficiency is about what would be expected if the data were uncensored and classical least squares estimates for the entire sample were compared to least squares estimates based only upon those observations with $x_t'\beta_0 > 0$; that is, the approximate three-to-one efficiency advantage of the MLE can largely be attributed to the exclusion (on average) of half of the observations having $x_t'\hat{\beta}_T < 0$.

A more surprising result is the finite sample bias of the SCLS estimator, with bias in the opposite direction (i.e., upward bias in slope, downward bias in intercept) from the classical least squares bias for censored data. While this bias represents a small conponent in the RMSE of the estimator, it is interesting in that it reflects an asymmetry in the sampling distribution of $\ddot{\beta}_T$ (as evidenced by the tabulated quartiles) rather than a "recentering" of this distribution away from the true parameter values. In fact, this asymmetry is caused by the interaction of the $x_t'\hat{\beta}_T < 0$ "exclusion" rule with the magnitude of the estimated $\ddot{\beta}_T$ components. Specifically, SCLS estimates with high slope and low intercept components rely on fewer observations for their numerical values than those with low slope and high intercept terms (since the condition "$x_t'\hat{\beta}_T > 0$" occurs more frequently in the latter case); as a result, high slope estimates have greater dispersion than low slope estimates, and the corresponding distribution is skewed upward. Though this is a second-order effect in terms of MSE efficiency, it does mean that conclusions concerning the "center" of the sampling distributions will depend upon the particular notion of center (i.e., mean or median) used.

The remaining designs in Table 1 indicate the variation in the sampling distribution of $\hat{\beta}_T$ when the sample size, signal-to-noise ratio, censoring proportion, and number of regressors changes. In the second design, a 50% reduction of the sample size results in an increase of about $\sqrt{2}$ in the standard deviations, and of more than $\sqrt{2}$ in the mean biases, in accord with the asymptotic theory (which specifies the standard deviations of the estimators to be $O(T^{-1/2})$ and the bias to be $o(T^{-1/2})$). Doubling the scale of the errors (which reduces the population R^2 of the uncensored regression to .2, and the R^2 for the observations with $x_t'\beta_o > 0$ to .06) has much more drastic effects on the bias and variability of the SCLS estimator, as illustrated by design 3; on the other hand, reduction of the censoring proportion to 25% (by increasing the intercept of the regression to one in design 4) substantially reduces the bias of the SCLS estimator and greatly improves its efficiency relative to the MLE (of course, as the censoring proportion tends to zero, this efficiency will tend to one, as both estimators reduce to classical least squares estimators).

The last two entries of Table 1 illustrate the interaction of number of parameters with sample size in the performance of the SCLS estimator. In these designs, an additional regressor w_t, taking on the repeating sequence

$\{-1,1,1,-1,\ldots\}$ of values; w_t thus had zero mean, unit variance, and was uncorrelated with z_t (both for the entire sample and the subsample with $x_t'\beta_o > 0$), and its coefficient in the regression function was taken to be zero, leaving the explanatory power of the regression unaffected. Comparison of designs 5 and 6 with designs 1 and 2 suggest that the convergence of the sampling distribution is slower than $O(p/n)$; that is, the sample size must increase more than proportionately to the number of regressors to achieve the same precision of the sampling distribution.

In Table 2, the design parameters of the base case -- T=200, censoring proportion = 50%, and overall error variance = 1 (except for design 8, which used the standard Cauchy distribution) -- are held fixed, and the effects of departures of the error distribution from normality and homoscedasticity are investigated. For these experiments, the RMSE or MAE efficiency of the SCLS estimator to the Gaussian MLE depends upon whether the inconsistency of the latter is numerically large relative to its dispersion. In the Laplace example (design 6), the 5% bias in the MLE slope coefficient (which reflects an actual recentering of the sampling distribution rather than asymmetry) is a small component of its RMSE, which is less than that for the SCLS estimator; with Cauchy errors, the performance of the Gaussian MLE deteriorates drastically relative to symmetrically censored least squares. For the 10% normal mixtures of designs 9 and 10, the MLE is more efficient for a relative scale (ratio of the standard deviations of the contaminating and contaminated distributions) of 5, but less efficient for a relative scale of 9. It is interesting to note that the performance of the SCLS estimator improves for these heavy-tailed distributions; holding the error variance constant, an increase in kurtosis of the error distribution reduces the variance of the "symmetrically trimmed" errors $\text{sgn}(u_t)\cdot\min\{|u_t|, |x_t'\beta_o|\}$, and so improves the SCLS performance.

Designs 11 and 12 make the scale of the error terms a linear function of $x_t'\beta_o$, while maintaining an overall variance $T^{-1}\sum E(u_t)^2$ of unity. For these designs, the "relative scale" is the ratio of the standard deviation of the 200th observation to the first (e.g., in design 11, $E(u_{200})^2=3E(u_1)^2$). With increasing heteroscedasticity (design 11), most of the error variability is associated with positive $x_t'\beta_o$ values, and the SCLS estimator performs relatively poorly; for decreasing heteroscedasticity (design 12), this conclusion is reversed. In both of these designs, the bias of the Gaussian MLE is much more pronounced, suggesting that failure of the homoscedasticity assumption may have more serious consequences than failure of Gaussianity in censored regression models.

While the results in Table 2 are not unambiguous with respect to the relative merits of the SCLS estimator to the misspecified MLE, there is reason to beleive that, for practical purposes, the foregoing results present an overly-favorable picture of the performance of the Gaussian MLE. As Ruud [1984] notes, more substantial biases of misspecified maximum likelihood estimation result when the joint distribution of the regressors is asymmetric, with low and high values of $x_t'\beta_o$ being associated with different components of the x_t vector. The designs in Table 2, which were based upon a single symmetrically distributed covariate, do not allow for this source of bias, which is likely to be important in empirical applications.

In any case, the performance of the symmetrically censored least squares estimator is much better than that of a two-step estimator (based upon a normality assumption) which uses only those observations with $y_t > 0$ in the second step. A simulation study by Paarsch [1984], which used designs very similar to those used here, found a substantial efficiency loss of this two-step estimator relative to Gaussian maximum likelihood. In terms of both relative efficiency and comutational ease, then, the symmetrically censored least squares estimator is preferable to this two-step procedure.

5. Extensions and Limitations

The "symmetric trimming" approach to estimation of Tobit models can be generalized in a number of respects. An obvious extension would be to permit more general loss functions than squared error loss. That is, suppose $\rho(\lambda)$ is a symmetric, convex loss function which is strictly convex at zero; then for the truncated regression model, the analogue to $R_T(\beta)$ of (2.7) would be

$$(5.1)\quad R_T(\beta;\ \rho) \equiv \sum_{t=1}^{T} \rho(y_t - \max\{\tfrac{1}{2} y_t,\ x_t'\beta\}),$$

which reduces to $R_T(\beta)$ when $\rho(\lambda) = \lambda^2$. A similar analogue to $S_T(\beta)$ of (2.10) is

$$(5.2)\quad S_T(\beta;\ \rho) \equiv \sum_{t=1}^{T} \rho(y_t - \min\{\tfrac{1}{2} y_t,\ x_t'\beta\})$$
$$+ \sum_{t=1}^{T} 1(y_t > 2x_t'\beta)\lfloor \rho(\tfrac{1}{2} y_t) - \rho(\max\{0,\ x_t'\beta\})\rfloor.$$

When $\rho(\lambda) = |\lambda|$, the minimand in (5.2) reduces to the objective function which defines the censored LAD estimator,

$$(5.3)\quad S_T(\beta;\ LAD) \equiv \sum_{t=1}^{T} |y_t - \max\{0,\ x_t'\beta\}|;$$

under somewhat different conditions (which did not require symmetry of the error distribution), the corresponding estimator was shown in Powell [1984] to be consistent and asymptotically normal. On the other hand, if the loss function $\rho(\lambda)$ is twice continuously differentiable, then the results of Theorems 1 and 2 can be proved as well for the estimators minimizing (5.2) and (5.1), with appropriate modification of the asymptotic covariance matrices.[9/]

To a certain extent, the "symmetric trimming" approach can be extended to other limited dependent variable models as well. The censored and truncated regression models considered above presume that the dependent variable is known to be censored (or truncated) to the left of zero. Such a sample selection mechanism may be relevant in some econometric contexts (e.g., individual labor supply or demand for a specific commodity), but there are other situations in which the censoring (or truncation) point would vary with each individual. A prominent example is the analysis of survival data; such data are often left-truncated at the individual's age at the start of the study, and right censored at the sum of the duration of the study and the individual's age when it begins.

The symmetrically trimmed estimators are easily adapted to limited dependent variable models with known limit values. In the survival model described above, let L_t and U_t be the lower truncation and upper censoring point for the t^{th} observation, so that the dependent variable y_t is generated from the model

$$(5.4)\quad y_t = \min\{x_t'\beta_o + v_t,\ U_t\},\ t = 1, \ldots, T,$$

where v_t has density function

$$(5.5)\quad g_t(\lambda) \equiv 1(\lambda > L_t - x_t'\beta_o)f_t(\lambda)[F_t(L_t - x_t'\beta_o)]^{-1},$$

for $F_t(\cdot)$ and $f_t(\cdot)$ satisfying Assumption E2. The corresponding symmetrically trimmed least squares estimator would minimize

$$(5.6)\quad R_T^*(\beta) \equiv \sum_t 1(x_t'\beta < \tfrac{1}{2}(U_t + L_t))(y_t - \max\{\tfrac{1}{2}(y_t + L_t),\ x_t'\beta\})^2$$

$$+ \sum_t 1(x_t'\beta > \tfrac{1}{2}(U_t + L_t))(y_t - \min\{\tfrac{1}{2}(y_t + U_t),\ x_t'\beta\})^2$$

$$+ \sum_t 1(x_t'\beta > \tfrac{1}{2}(U_t + L_t))\cdot 1(\tfrac{1}{2}(y_t + U_t) < x_t'\beta)$$

$$\cdot[(\tfrac{1}{2}(y_t - U_t))^2 - (\min\{0,\ x_t'\beta - U_t\})^2],$$

and consistency and asymptotic normality for this estimator could be proved using appropriate modifications of the corresponding assumptions for $\hat{\beta}_T$ and $\tilde{\beta}_T$.

In the model considered above, the upper and lower "cutoff" points U_t and L_t are assumed to be observable for each t; however, many limited dependent variable models in use do not share this feature. In survival analysis, for example, it is often assumed that the censoring point is observed only for those data points which are in fact censored, on the grounds that, once an uncensored observation occurs, the value of the "potential" cutoff point is no longer relevant.[10/] As another example, a model commonly used in economics has a censoring (truncation) mechanism which is not identical to the regression equation of interest; in such a model (more properly termed a "sample selection" model, since the range of the observed dependent variable is not necessarily restricted), the complete data point (y_t^*, x_t') from the linear model $y_t^* = x_t'\beta_0 + u_t$ is observed only if some (unobserved) index function I_t exceeds zero, and the dependent variable is censored (or the data point is truncated) otherwise. The index function I_t is typically written in linear form $I_t = z_t'\alpha_o + e_t$, depending upon observed regressors z_t, unobserved parameters α_o, and an error term e_t which is independent across t, but is neither independent of, nor perfectly correlated with, the residual u_t in the equation of interest.

Unfortunately, the "symmetric trimming" approach outlined in Section 2 does not appear to extend to models with unobserved cutoff points. In the sample selection model described above, only the sign of I_t and not its value is observed, so "symmetric trimming" of the sample is not feasible, since it is not possible to determine whether $I_t < 2x_t'\alpha_o$, which is the trimming rule corresponding to $y_t < 2x_t'\beta_o$ in the simpler Tobit models considered previously. This example illustrates the difficulty of estimation of (α_o', β_o') in the model with $I_t \neq y_t^*$ when the joint distribution of the error terms is unknown.[11/]

Finally, the approach can naturally be extended to the nonlinear regression model, i.e., one in which the linear regression function is $h_t(x_t, \beta_o)$. With some additional regularity conditions on the form of $h_t(\cdot)$(such as differentiability and uniform boundedness of moments of $h_t(\cdot)$ and $\|\partial h_t/\partial\beta\|$ over t and B), the results of Theorems 1 through 4 should hold if "$x_t'\beta$" and "x_t" are replaced by "$h_t(x_t, \beta)$" and "$\partial h_t/\partial\beta$" throughout. Of course, just as in the standard regression model, the symmetric trimming approach will not be robust to misspecification of the regression function $h_t(\cdot)$; nonetheless, theoretical considerations are more likely to suggest the paper form of the regression function than the functional form of the error distribution in empirical applications, so the approach proposed here addresses the more difficult of these two specification problems.

APPENDIX

Proofs of Theorems in Text

Proof of Theorem 1

(i) Strong Consistency: The proof of consistency of $\hat{\beta}_T$ follows the approach of Amemiya [1973] and White [1980] by showing that, in large samples, the objective function $S_T(\beta)$ is minimized by a sequence of values which converge a.s. to β_o. First, note that minimization of $S_T(\beta)$ is equivalent to minimiztion of the normalized function

$$(A.1)\quad Q_T(\beta) \equiv \frac{1}{T}\left[S_T(\beta) - S_T(\beta_o)\right],$$

which can be written more explicitly as

$$\begin{aligned}(A.2)\quad Q_T(\beta) \equiv \frac{1}{T}\sum_{t=1}^{T} \{ & 1(\tfrac{1}{2}\, y_t < \min\{x_t'\beta,\ x_t'\beta_o\})\left[(y_t - x_t'\beta)^2 - (y_t - x_t'\beta_o)^2\right] \\ & + 1(x_t'\beta < \tfrac{1}{2}\, y_t < x_t'\beta_o)\left[(\tfrac{1}{2}\, y_t)^2 - (\max\{0,\ x_t'\beta\})^2 - (y_t - x_t'\beta_o)^2\right] \\ & + 1(x_t'\beta_o < \tfrac{1}{2}\, y_t < x_t'\beta)\left[(y_t - x_t'\beta)^2 - (\tfrac{1}{2}\, y_t)^2 + (\max\{0,\ x_t'\beta_o\})^2\right] \\ & + 1(\tfrac{1}{2}\, y_t > \max\{x_t'\beta, x_t'\beta_o\})\left[(\max\{0,\ x_t'\beta\})^2 - (\max\{0,\ x_t'\beta_o\})^2\right]\}.\end{aligned}$$

By inspection, each term in the sum is bounded by $\|x_t\|^2(\|\beta\| + \|\beta_o\|)$, so by Lemma

2.3 of White [1980],

$$\text{(A.3)} \quad \lim_{T\to\infty} |Q_T(\beta) - E[Q_T(\beta)]| = 0$$

uniformly in β B by the compactness of B. Since $E[Q_T(\beta_o)] = 0$, $\hat{\beta}_T$ will be strongly consistent if, for any $\varepsilon > 0$, $E[Q_T(\beta)]$ is strictly greater than zero uniformly in β for $\|\beta - \beta_o\| > \varepsilon$ and all T sufficiently large, by Lemma 2.2 of White [1980].

After some manipulations (which exploit the symmetry of the conditional distribution of the errors), the conditional expectation of $Q_T(\beta)$ given the regressors can be written as

$$\text{(A.4)} \quad E[Q_T(\beta)|x_1,\ldots,x_T]$$

$$= \frac{1}{T}\sum_{t=1}^{T} \{1(x_t'\beta_o < 0 < x_t'\beta)[\int_{x_t'\beta_o}^{-x_t'\beta_o}(x_t'\beta)^2 dF_t(\lambda)$$

$$+ \int_{-x_t'\beta_o}^{-x_t'\beta_o+2x_t'\beta} \tfrac{1}{2}(\lambda+x_t'\beta_o-2x_t'\beta)^2 dF_t(\lambda)]$$

$$+ 1(x_t'\beta < 0 < x_t'\beta_o)[\int_{-x_t'\beta_o}^{x_t'\beta_o} \tfrac{1}{2}[(x_t'\beta_o)^2 - \lambda^2]dF_t(\lambda)]$$

$$+ 1(0 < x_t'\beta_o < x_t'\beta)[\int_{-x_t'\beta_o}^{x_t'\beta_o} (x_t'\delta)^2 dF_t(\lambda)$$

$$+ \int_{x_t'\beta_o}^{-x_t'\beta_o+2x_t'\beta} \tfrac{1}{2}(\lambda+x_t'\beta_o-2x_t'\beta)^2 dF_t(\lambda)]$$

$$+ 1(0 < x_t'\beta < x_t'\beta_o)[\int_{-x_t'\beta_o}^{x_t'\beta_o} \tfrac{1}{2}(x_t'\delta)^2[1-(\lambda/x_t'\beta_o)^2]dF_t(\lambda)$$

$$+ \int_{-x_t'\beta_o}^{-x_t'\beta_o+2x_t'\beta} \tfrac{1}{2}(x_t'\delta/x_t'\beta_o)^2(\lambda+x_t'\beta_o)^2 dF_t(\lambda)$$

$$+ \int_{-x_t'\beta_o+2x_t'\beta}^{x_t'\beta_o} \tfrac{1}{2}[(x_t'\delta/x_t'\beta_o)^2(\lambda+x_t'\beta_o)^2 - (\lambda+x_t'\beta_o-2x_t'\beta)^2]dF_t(\lambda)]\},$$

where

(A.5) $\delta \equiv \beta - \beta_o$

here and throughout the appendix, and where $F_t(\lambda)$ is the conditional c.d.f. of u_t given x_t. All of the integrals in this expression are nonnegative, so

(A.6) $E[Q_T(\beta)|x_1,\ldots,x_T]$

$$\geq \frac{1}{T}\sum_t \{1(x_t'\beta_o \geq \varepsilon_o,\ x_t'\beta \leq 0)\int_{-x_t'\beta_o}^{x_t'\beta_o} \tfrac{1}{2}[(x_t'\beta_o) - \lambda^2]dF_t(\lambda)$$

$$+ 1(x_t'\beta_o \geq \varepsilon_o,\ x_t'\beta_o > x_t'\beta > 0)\int_{-x_t'\beta_o}^{x_t'\beta_o} \tfrac{1}{2}(x_t'\delta)^2[1-(\lambda/x_t'\beta_o)^2]dF_t(\lambda)$$

$$+ 1(x_t'\beta_o \geq \varepsilon_o,\ x_t'\beta > x_t'\beta_o)\int_{-x_t'\beta_o}^{x_t'\beta_o} (x_t'\delta)^2 dF_t(\lambda)\},$$

where ε_o is chosen as in Assumptions R and E1. Without loss of generality, let

$\varepsilon_o = \min\{\varepsilon_o, \zeta_o\}$; then, for any τ in $(0, \varepsilon_o]$,

$$\text{(A.7)} \quad E[Q_T(\beta)|x_1,\ldots,x_T]$$

$$\geq \frac{1}{T}\sum_t \{1(x_t'\beta_o \geq \varepsilon_o,\ x_t'\beta < 0)\int_0^{\varepsilon_o} [\varepsilon_o - \lambda^2]\varepsilon_o d\lambda$$

$$+ 1(x_t'\beta_o \geq \varepsilon_o,\ x_t'\beta_o > x_t'\beta > 0)1(|x_t'\delta| > \tau)\int_0^{\varepsilon_o} \tau^2[1-(\lambda/\varepsilon_o)^2]\varepsilon_o d\lambda$$

$$+ 1(x'_t\beta_o \geq \varepsilon_o,\ x_t'\beta > x_t'\beta_o)1(|x_t'\delta| > \tau)\int_0^{\varepsilon_o} 2\tau^2\varepsilon_o d\lambda$$

$$\geq \frac{2}{3}\varepsilon_o^2\tau^2\frac{1}{T}\sum_t 1(x_t'\beta_o \geq \varepsilon_o)1(|x_t'\delta| > \tau)$$

Finally, let $K \equiv \frac{2}{3}\varepsilon_o^2(K_o)^{3/2(4+\eta)}$, for K_o defined in Assumption R. Then taking expectations with respect to the distribution of the regressors and applying Holder's and Jensen's inequalities yields

$$\text{(A.8)} \quad E[Q_T(\beta)]$$

$$\geq K\tau^2\frac{1}{T}\sum_t E[1(x_t'\beta_o \geq \varepsilon_o)\cdot 1(|x_t'\delta| > \tau)]^3[E\|x_t\|^3]^2$$

$$\geq K\tau^2\frac{1}{T}\sum_t E[1(x_t'\beta_o \geq \varepsilon_o)\cdot 1(|x_t'\delta| > \tau)\|x_t\|^2]^3$$

$$> K\tau^2\{\frac{1}{T}\sum_t E[1(x_t'\beta_o > \varepsilon_o)\cdot 1(|x_t'\delta| > \tau)(x_t'\delta)^2\|\delta\|^{-2}]\}^3$$

$$> K\tau^2[(\nu_T\|\delta\|^2 - \tau^2)\|\delta\|^{-2}]^3$$

where ν_T, the smallest characteristic root of the matrix N_T, is bounded below by ν_o for large T by Assumption R. Thus, whenever $\|\delta\|\|\beta - \beta_o\| > \varepsilon$, choosing τ so that $\tau^2 < \nu_o\varepsilon^2$ shows that $E[Q_T(\beta)]$ is uniformly bounded away from zero for all T $> T_o$, as desired.

(ii) Asymptotic Normality: The proofs of asymptotic normality here and below are based on the approach of Huber [1967], suitably modified to apply to the non-identically distributed case. (For complete details of the modifications required, the reader is referred to Appendix A of Powell [1984]). Define

$$\text{(A.9)}\quad \psi(u_t, x_t, \beta) \equiv 1(x_t'\beta > 0)x_t(\min\{y_t, 2x_t'\beta\} - x_t'\beta)$$
$$= 1(x_t'\beta > 0)x_t(\min\{\max\{u_t - x_t'\delta, -x_t'\beta\}, x_t'\beta\});$$

then Huber's argument requires that

$$\text{(A.10)}\quad \frac{1}{\sqrt{T}}\sum_t \psi(u_t, x_t, \hat{\beta}_T) = o_p(1).$$

But the left-hand side of (A.10), multiplied by $-2\sqrt{T}$, is the gradient of the objective function $S_T(\beta)$ defining $\ddot{\beta}_T$, so it is equal to the zero vector for large

T by Assumption P and the strong consistency of $\hat{\beta}_T$.

Other conditions for application of Huber's argument involve the quantities

$$\mu_t(\beta, d) \equiv \sup_{\|\beta-\gamma\|<d} \|\psi(u_t, x_t, \beta) - \psi(u_t, x_t, \gamma)\|$$

and

$$\lambda_T(\beta) \equiv \frac{1}{T} \sum_t E[\psi(u_t, x_t, \beta)].$$

Aside from some technical conditions which are easily verified in this situation (such as the measurability of the $\{\mu_t\}$), Huber's argument is valid if the first two moments of μ_t are $O(d)$, and $\lambda_T(\beta) \geq a \|\beta - \beta_o\|$ (some $a > 0$), for all β near β_o, d near zero, and $T > T_o$.

Simple algebraic manipulations show that

$$\text{(A.11)} \quad \mu_t(\beta, d) \equiv \sup_{\|\beta-\gamma\|d} \|1(x_t'\beta > 0)x_t(\min\{y_t - x_t'\beta, x_t'\beta\}) \\ -1(x_t'\gamma > 0)x_t(\min\{y_t - x_t'\gamma, x_t'\gamma\})\| \\ \leq \|x_t\|^2 d,$$

so that

$$\text{(A.12)} \quad E[\mu_t(\beta, d)] \leq K_o^{2/(4+\eta)} d,$$

$$E[\mu_t(\beta, d)]^2 \leq K_o^{4/(4+\eta)} d^2,$$

so both are $O(d)$ for d near zero. Furthermore,

$$\text{(A.13)} \quad \lambda_T(\beta) = -C_T(\beta - \beta_o) + O(\|\beta - \beta_o\|^2),$$

where C_T is the matrix defined in the statement of the theorem. This holds because

(A.14) $\|\lambda_T(\beta) + C_T(\beta - \beta_o)\|$

$$= E\,\|\frac{1}{T}\sum_t 1(x_t'\beta_o > -x_t'\delta)x_t\{\int_{-x_t'\beta_o}^{-x_t'\beta_o+2x_t'\beta}(-x_t'\delta)dF_t(\lambda)+\int_{-x_t'\beta_o+2x'\beta}^{x_t'\beta_o}(x_t'\beta)dF_t(\lambda)\}$$

$$-\frac{1}{T}\sum_t 1(x_t'\beta_o > 0)x_t\int_{-x_t'\beta_o}^{x_t'\beta_o}(-x_t'\delta)dF_t(\lambda)\|$$

$$\leq E\,\|\frac{1}{T}\sum_t 1(|x_t'\beta_o| < |x_t'\delta|)x_t\{\int_{-x_t'\beta_o}^{-x_t'\beta_o+2x_t'\beta}(-x_t'\delta)dF_t(\lambda)$$

$$+\int_{-x_t'\beta_o+2x_t'\beta}^{x_t'\beta_o}(x_t'\beta)dF_t(\lambda)\}$$

$$+\frac{1}{T}\sum_t 1(x_t'\beta_o > |x_t'\delta|)x_t\int_{-x_t'\beta_o+2x_t'\beta}^{x_t'\beta_o}(-x_t'\beta_o+2x_t'\beta-\lambda)dF_t(\lambda)\|$$

$$\leq 4L_oE\,\|\frac{1}{T}\sum_t \lfloor x_t(x_t'\delta)^2\rfloor\|$$

$$\leq 4L_o(K_o)^{3/(4+\eta)}\,\|\delta\|^2.$$

Thus $\lambda_T(\beta)\cdot\|\beta - \beta_o\|^{-1} > a > 0$ for all β in a suitably small neighborhood of β_o, since C_T is positive definite for large T by Assumption R and E1. The argument leading to Huber's [1967] Lemma 3 thus yields the asymptotic relation

$$\text{(A.15)} \quad C_T \cdot \sqrt{T}(\hat{\beta}_T - \beta_o) = \frac{1}{\sqrt{T}} \sum_t \psi(u_t, x_t, \beta_o) + o_p(1).$$

Application of the Liapunov form of the central limit theorem yields the result, where the covariance matrix is calculated in the usual way.

Proof of Theorem 2:

(i) Strong Consistency: The argument is the same as for Theorem 1 above; strong consistency of $\tilde{\beta}_T$ holds if β_o uniquely minimizes $T^{-1}E[R_T(\beta) - R_T(\beta_o)] \equiv E[M_T(\beta)]$ uniformly in β for T sufficiently large. By some tedious algebra, it can be shown that the following inequality holds:

$$\text{(A.16)} \quad E[M_T(\beta)|x_1,\ldots,x_T]$$

$$\geq \frac{1}{T}\sum_t [4F_t(x_t'\beta_o)]^{-1}\{1(x_t'\beta_o > 0,\ x_t'\beta < 0)\int_0^{x_t'\beta_o}[(x_t'\beta_o-\lambda)^2 - \lambda^2]dF_t(\lambda)$$

$$+ 1(x_t'\beta_o < 0,\ x_t'\beta > 0)\int_0^{x_t'\beta}[(x_t'\beta-\lambda)^2 - \lambda^2]dF_t(\lambda)$$

$$+ 1(x_t'\beta_o > 0,\ x_t'\beta > 0)\int_0^{|x_t'\delta|}[(|x_t'\delta|-\lambda)^2 - \lambda^2]dF_t(\lambda)\}.$$

Using the conditions in E2, it can be shown that, by choosing τ sufficiently close to zero,

$$\text{(A.17)} \quad E[M_T(\beta)] > \tau^3[h(\tfrac{\tau}{2}) - h(\tfrac{3\tau}{2})] \frac{1}{T}\sum_t E[1(x_t'\beta_o > \varepsilon_o)\cdot 1(|x_t'\delta| > \tau)],$$

which corresponds to inequality (A.7) above. The same reasoning as in (A.8) thus shows that $E[M_T(\beta)] > 0$ uniformly in β for $\|\beta - \beta_o\| > \varepsilon$ and $T > T_o$, as desired.

(ii) Asymptotic Normality: The proof here again follows that in Theorem 1; in this case, defining

$$\text{(A.18)} \quad \psi(v_t, x_t, \beta) \equiv 1(y_t < 2x_t'\beta)(y_t - x_t'\beta)x_t$$

$$= 1(v_t - x_t'\delta < x_t'\beta)(v_t - x_t'\delta)x_t,$$

the condition

$$\text{(A.19)} \quad \frac{1}{\sqrt{T}}\sum_t \psi(v_t, x_t, \beta_T) = o_p(1)$$

must again be established; but the left-hand side of (A.19), multiplied by $-2\sqrt{T}$, is the (left) derivative of the objective function $R_T(\beta)$ evaluated at $\tilde{\beta}_T$, so (A.19) can be established using the facts that $\tilde{\beta}_T$ minimizes $R_T(\beta)$ and that the distribution of v_t continuous (a similar argument is given for Lemma A.2 of Ruppert and Carroll [1980]).

For this problem, the appropriate quantities $\mu_t(\beta, d)$ and $\lambda_T(\beta)$ can be shown to satisfy

$$(A.20) \quad \mu_t(\beta, d) < \frac{1}{T} \sum_t \|x_t\|^2 d$$

and

$$(A.21) \quad \|\lambda_T(\beta) + (W_T - V_T)(\beta - \beta_o)\| = O(\|\beta - \beta_o\|^2)$$

for β near β_o, so

$$(A.22) \quad (W_T - V_T)\sqrt{T}(\beta - \beta_o) = \frac{1}{\sqrt{T}} \sum_t \psi(v_t , x_t , \beta_o) + o_p(1)$$

as before, and application of Liapunov's central limit theorem yields the result.

Proof of Theorem 3:

Only the consistency of $\hat{D}_T$ will be demonstrated here; consistency of $\hat{C}_T$ can be shown in an analogous way. First, note that $\hat{D}_T$ is asymptotically equivalent to the matrix

$$(A.23) \quad \Delta_T \equiv \frac{1}{T} \sum_t 1(x_t'\beta_o > 0)\cdot\min\{u_t^2, (x_t'\beta_o)^2\}x_t x_t';$$

that is, each element of the difference matrix $\hat{D}_T - \Delta_T$ converges to zero almost surely. This holds since, defining $\hat{\delta}_T \equiv \hat{\beta}_T - \beta_o$, the $(i,j)^{th}$ element of the difference satisfies

(A.24) $|[\hat{D}_T - \Delta_T]_{ij}|$

$$= |\frac{1}{T}\sum_t 1(x_t'\beta_o > -x_t'\hat{\delta}_T)\min\{(u_t - x_t'\hat{\delta}_T)^2, (x_t'\beta_o + x_t'\hat{\delta}_T)^2\}x_{it}x_{jt}$$

$$- \frac{1}{T}\sum_t 1(x_t'\beta_o > 0)\min\{u_t^2, (x_t'\beta_o)^2\}x_{it}x_{jt}|$$

$$< \frac{1}{T}\sum_t 1(|x_t'\beta_o| < |x_t'\hat{\delta}_T|)[(x_t'\beta_o)^2 + (x_t'\hat{\beta}_T)^2]|x_{it}x_{jt}|$$

$$+ \frac{1}{T}\sum_t 1(x_t'\beta_o > |x_t'\hat{\delta}_T|)\cdot 1(|u_t| < x_t'\beta_o)[-2u_t(x_t'\hat{\delta}_T) + (x_t'\hat{\delta}_T)^2]|x_{it}x_{jt}|$$

$$+ \frac{1}{T}\sum_t 1(x_t'\beta_o > |x_t'\hat{\delta}_T|)\cdot 1(|u_t| > x_t'\beta_o)[-2(x_t'\beta_o)(x_t'\hat{\delta}_T) + (x_t'\hat{\delta}_T)^2]|x_{it}x_{jt}|$$

$$< \frac{2}{T}\sum_t \|x_t\|^4\|\hat{\delta}_T\|(\|\beta_o\| + \|\hat{\delta}_T\|).$$

But since $E\|x_t\|^{4+\eta}$ is uniformly bounded, these terms converge to zero almost surely by the strong consistency of $\hat{\beta}_T$. Consistency of $\hat{D}_T$ then follows from the almost sure convergence of $\Delta_T - E[\Delta_T] \equiv \Delta_T - D_T$ to zero by Markov's strong law of large numbers.

<u>Proof of Theorem 4</u>:

Again, strong consistency of $\tilde{W}_T$ and $\tilde{Z}_T$ can be demonstrated in the same way it was shown for $\hat{D}_T$ in the preceeding proof. To prove consistency of $\tilde{V}_T$, it will first be shown that $\tilde{V}_T$ is asymptotically equivalent to

$$\text{(A.25)} \quad \Gamma_T \equiv \frac{1}{T}\sum_t 1(x_t'\beta_o > 0)(x_t'\beta_o)x_t x_t' \cdot c_T^{-1}[1(y_t < c_T) + 1(0 < y_t - 2x_t'\beta_o < c_T)].$$

First, note that $(\tilde{c}_T/c_T)\tilde{V}_T$ is asymptotically equivalent to $\tilde{V}_T$, since $(\tilde{c}_T/c_T) \overset{p}{\rightarrow} 1$; next, note that each element of $(\tilde{c}_T/c_T)\tilde{V}_T - \Gamma_T$ satisfies

$$\text{(A.26)} \quad |[(\tilde{c}_T/c_T)\tilde{V}_T - \Gamma_T]_{ij}|$$

$$< \frac{1}{T}\sum_t 1(|x_t'\beta_o| < |x_t'\tilde{\delta}_T|)\|x_t\|^2 c_T^{-1}|x_t'\beta_o|$$

$$+ \frac{1}{T}\sum_t 1(x_t'\beta_o > |x_t'\tilde{\delta}_T|)\|x_t\|^2 c_T^{-1}|x_t'\tilde{\delta}_T|$$

$$+ \frac{1}{T}\sum_t (x_t'\beta_o > |x_t'\tilde{\delta}_T|)\|x_t\|^2|x_t'\beta_o|c_T^{-1}\cdot 1(|y_t - c_t| < |\tilde{c}_t - c_T|)$$

$$+ \frac{1}{T}\sum_t 1(x_t'\beta_o > |x_t'\tilde{\delta}_T|)\|x_t\|^2|x_t'\beta_o|c_T^{-1}\cdot 1(|y_t - 2x_t'\beta_o - c_T| < |\tilde{c}_T - c_T| + |x_t'\tilde{\delta}_T|)$$

$$< \frac{1}{T}\sum_t \|x_t\|^3\, \tilde{c}_T^{-1}\|\tilde{\delta}_T\| + \frac{1}{T}\sum_t \|x_t\|^3\|\beta_o\|c_T^{-1}1(|v_t + x_t'\beta_o - c_T| < c_T\cdot o_p(1))$$

$$+ \frac{1}{T}\sum_t \|x_t\|^3\|\beta_o\|c_T^{-1}\cdot 1(|v_t - x_t'\beta_o - c_T| < c_T[o_p(1) + c_T^{-1}\|\tilde{\delta}_T\|\ \|x_t\|]),$$

where $\tilde{\delta}_T \equiv \tilde{\beta}_T - \beta_o$. Chebyshev's inequality and the fact that $\sqrt{T}(\tilde{\beta}_T - \beta_o) = o_p(1)$ can thus be used to show that the difference between $\tilde{V}_T$ and Γ_T converges to zero in probability.

The expected value of Γ_T is

$$(A.27)\quad E[\Gamma_T] = E\{E[\Gamma_T | x_1, \ldots, x_T]\}$$

$$= E[\frac{1}{T}\sum_t 1(x_t'\beta_o > 0)(x_t'\beta_o)x_t x_t'[F_t(x_t'\beta_o)]^{-1}\int_{-c_T}^{c_T} c_T^{-1} f_t(\lambda - x_t'\beta_o)d\lambda]$$

$$= E[\frac{1}{T}\sum_t 1(x_t'\beta_o > 0)(x_t'\beta_o)x_t x_t'[F_t(x_t'\beta_o)]^{-1}\int_{-1}^{1} f_t(c_T\lambda - x_t'\beta_o)d\lambda]$$

$$= V_T + o(1)$$

by dominated convergence, while

$$(A.28)\quad Var([\Gamma_T]_{ij})$$

$$\leq (\frac{1}{T})^2\sum_t E\{1(x_t'\beta_o > 0)\|x_t\|^3\|\beta_o\|\cdot c_T^{-2}\{[1(0 < v_t - x_t'\beta_o < c_T) + 1(0 < v_t + x_t'\beta_o < c_T)]\}$$

$$= o(1/c_T T) = o(T^{-1/2}),$$

so each element of $\Gamma_T - V_T$ converges to zero in quadratic mean.

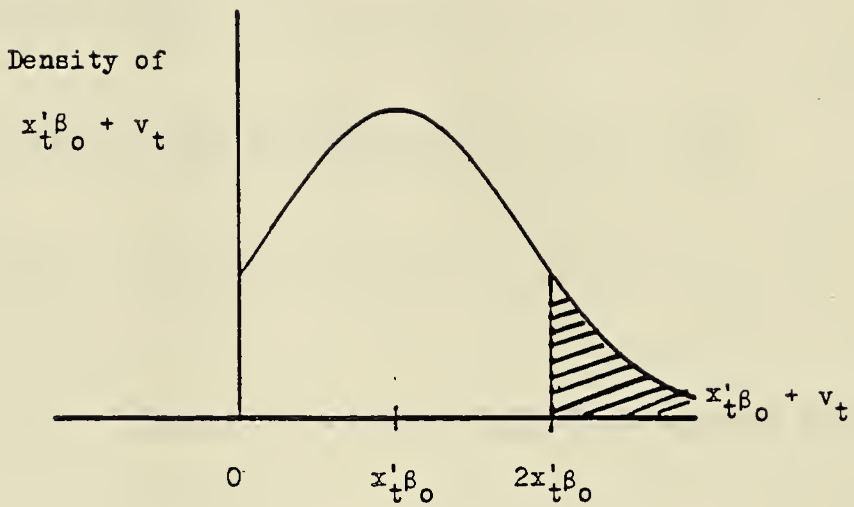

FIGURE 1

Density Function of y_t in Truncated

and "Symmetrically Truncated" Sample

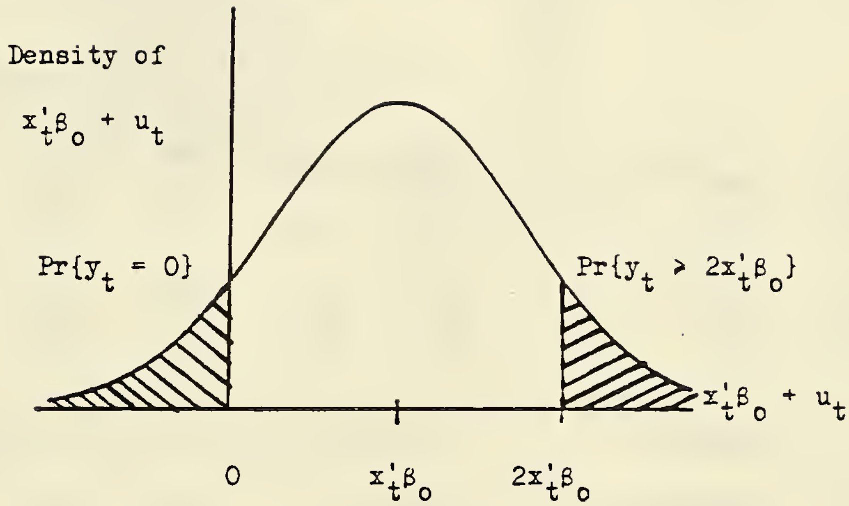

FIGURE 2

Distribution of y_t in Censored and

"Symmetrically Censored" Sample

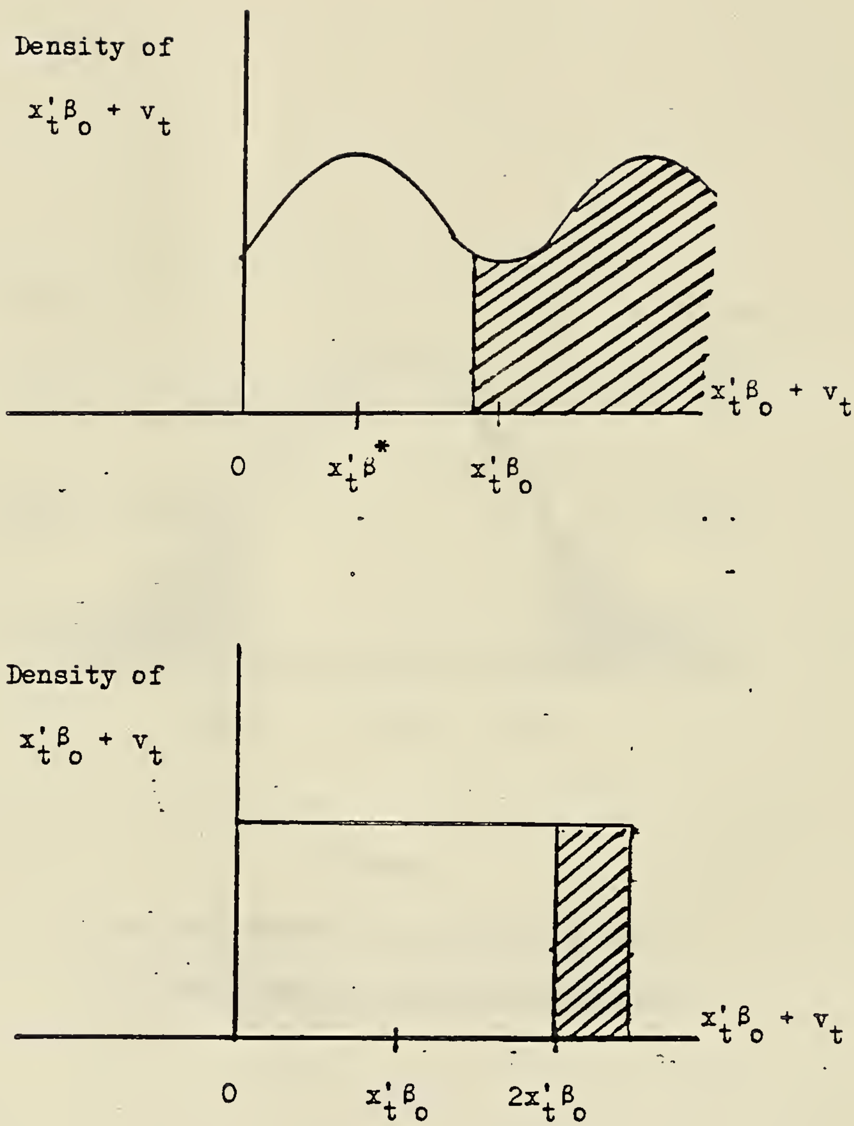

FIGURE 3

Density of y_t in Bimodal

and Uniform Cases

TABLE 1

Results of Simulations for Designs With Homocedastic Gaussian Errors

Design 1 - Scale=1, T=200, Censoring=50%

Tobit MLE	True	Mean	SD	RMSE	LQ	Median	UQ	MAE
Intercept	.000	-.005	.093	.093	-.071	.008	.065	.069
Slope	1.000	1.003	.097	.097	.934	1.000	1.053	.063

SCLS	True	Mean	SD	RMSE	LQ	Median	UQ	MAE
Intercept	.000	-.080	.411	.418	-.135	.028	.145	.167
Slope	1.000	1.060	.325	.331	.844	.988	1.182	.171

Design 2 - Scale=1, T=100, Censoring=50%

Tobit MLE	True	Mean	SD	RMSE	LQ	Median	UQ	MAE
Intercept	.000	-.001	.142	.142	-.088	.008	.109	.102
Slope	1.000	1.009	.141	.142	.903	.994	1.108	.102

SCLS	True	Mean	SD	RMSE	LQ	Median	UQ	MAE
Intercept	.000	-.129	.566	.580	-.300	.074	.233	.263
Slope	1.000	1.101	.463	.474	.780	.994	1.252	.273

Design 3 - Scale=2, T=200, Censoring=50%

Tobit MLE	True	Mean	SD	RMSE	LQ	Median	UQ	MAE
Intercept	.000	-.011	.194	.195	-.135	.012	.145	.141
Slope	1.000	.987	.161	.162	.875	.980	1.100	.117

SCLS	True	Mean	SD	RMSE	LQ	Median	UQ	MAE
Intercept	.000	-.903	2.974	3.108	-.821	-.079	.208	.311
Slope	1.000	1.622	1.887	1.987	.847	1.132	1.665	.342

Design 4 - Scale=1, T=200, Censoring=25%

Tobit MLE	True	Mean	SD	RMSE	LQ	Median	UQ	MAE
Intercept	1.000	.999	.080	.080	.949	1.000	1.046	.050
Slope	1.000	1.002	.077	.077	.946	1.003	1.055	.054

SCLS	True	Mean	SD	RMSE	LQ	Median	UQ	MAE
Intercept	1.000	.981	.123	.124	.919	.997	1.063	.072
Slope	1.000	1.019	.126	.128	.928	1.016	1.096	.085

TABLE 1 (continued)

Design 5 - Scale=1, T=200, Censoring=50%, Two Slope Coefficients

Tobit MLE	True	Mean	SD	RMSE	LQ	Median	UQ	MAE
Intercept	.000	-.001	.102	.102	-.067	.004	.077	.069
Slope 1	1.000	.996	.095	.095	.942	.995	1.056	.057
Slope 2	.000	-.011	.083	.084	-.071	-.011	.045	.062

SCLS	True	Mean	SD	RMSE	LQ	Median	UQ	MAE
Intercept	.000	-.174	.497	.527	-.322	-.023	.129	.200
Slope 1	1.000	1.124	.380	.399	.878	1.042	1.256	.193
Slope 2	.000	-.015	.128	.129	-.098	-.015	.064	.080

Design 6 - Scale=1, T=300, Censoring=50%, Two Slope Coefficients

Tobit MLE	True	Mean	SD	RMSE	LQ	Median	UQ	MAE
Intercept	.000	-.002	.082	.082	-.056	.008	.055	.056
Slope 1	1.000	.998	.083	.083	.940	.996	1.055	.059
Slope 2	.000	.003	.066	.066	-.037	.003	.041	.040

SCLS	True	Mean	SD	RMSE	LQ	Median	UQ	MAE
Intercept	.000	-.120	.407	.425	-.252	-.011	.117	.160
Slope 1	1.000	1.086	.315	.326	.897	1.020	1.222	.146
Slope 2	.000	.002	.094	.094	-.058	.004	.061	.058

TABLE 2

Results of Simulations for Designs Without Homocedastic Gaussian Errors

Design 7 - Laplace, T=200, Censoring=50%

Tobit MLE	True	Mean	SD	RMSE	LQ	Median	UQ	MAE
Intercept	.000	-.037	.101	.108	-.102	-.032	.032	.071
Slope	1.000	1.052	.109	.121	.974	1.047	1.125	.081

SCLS	True	Mean	SD	RMSE	LQ	Median	UQ	MAE
Intercept	.000	-.033	.203	.205	-.148	.018	.110	.122
Slope	1.000	1.034	.197	.200	.894	1.005	1.168	.129

Design 8 - Std. Cauchy, T=200, Censoring=50%

Tobit MLE	True	Mean	SD	RMSE	LQ	Median	UQ	MAE
Intercept	.000	-29.506	183.108	185.470	-10.522	-4.333	-2.045	4.422
Slope	1.000	25.826	166.976	168.812	2.559	4.241	7.768	3.241

SCLS	True	Mean	SD	RMSE	LQ	Median	UQ	MAE
Intercept	.000	-.352	1.588	1.626	-.404	-.002	.185	.248
Slope	1.000	1.439	2.537	2.575	.834	1.052	1.353	.240

Design 9 - Normal Mixture, Relative Scale=4, T=200, Censoring=50%

Tobit MLE	True	Mean	SD	RMSE	LQ	Median	UQ	MAE
Intercept	.000	-.065	.098	.118	-.126	-.052	.005	.072
Slope	1.000	1.087	.124	.152	.999	1.074	1.161	.095

SCLS	True	Mean	SD	RMSE	LQ	Median	UQ	MAE
Intercept	.000	-.013	.215	.215	-.122	.024	.130	.130
Slope	1.000	1.019	.191	.192	.896	1.001	1.123	.110

Design 10 - Normal Mixture, Relative Scale=9, T=200, Censoring=50%

Tobit MLE	True	Mean	SD	RMSE	LQ	Median	UQ	MAE
Intercept	.000	-.180	.125	.219	-.237	-.159	-.086	.160
Slope	1.000	1.194	.166	.255	1.078	1.174	1.268	.174

SCLS	True	Mean	SD	RMSE	LQ	Median	UQ	MAE
Intercept	.000	-.020	.108	.110	-.082	-.007	.054	.066
Slope	1.000	1.019	.118	.120	.936	1.009	1.092	.069

TABLE 2 (continued)

Design 11 - Heteroscedastic Normal, $3E(u_1)^2=E(u_{200})^2$, T=200, Censoring=50%

Tobit MLE	True	Mean	SD	RMSE	LQ	Median	UQ	MAE
Intercept	.000	-.274	.135	.306	-.361	-.256	.171	.261
Slope	1.000	1.283	.139	.315	1.187	1.285	1.378	.285

SCLS	True	Mean	SD	RMSE	LQ	Median	UQ	MAE
Intercept	.000	-.228	.613	.654	-.334	.021	.125	.208
Slope	1.000	1.166	.476	.504	.833	1.109	1.306	.234

Design 12 - Heteroscedastic Normal, $E(u_1)^2=3E(u_{200})^2$, T=200, Censoring=50%

Tobit MLE	True	Mean	SD	RMSE	LQ	Median	UQ	MAE
Intercept	.000	.193	.075	.207	.144	.197	.242	.197
Slope	1.000	.807	.071	.206	.754	.806	.856	.195

SCLS	True	Mean	SD	RMSE	LQ	Median	UQ	MAE
Intercept	.000	-.010	.211	.211	-.113	-.003	.154	.136
Slope	1.000	1.014	.161	.162	.887	1.004	1.117	.117

FOOTNOTES

1/ This research was supported by National Science Foundation Grants No. SES79-12965 and SES79-13976 at Stanford University and SES83-09292 at the Massachusetts Institute of Technology. I would like to thank T. Amemiya, T. W. Anderson, T. Bresnahan, C. Cameron, J. Hausman, T. MaCurdy, D. McFadden, W. Newey, J. Rotemberg, P. Ruud, T. Stoker, R. Turner, and a referee for their helpful comments.

2/ These references are by no means exhaustive; alternative estimation methods are given by Kalbfleisch and Prentice [1980], Koul, Suslara, and Van Ryzin [1981], and Ruud [1983], among others.

3/ Since the right-hand side of (2.6) is discontinuous, it may not be possible to find a value of β which sets it to zero for a particular sample. In the asymptotic theory to follow, it is sufficient for the summation in (2.6), when normalized by $\frac{1}{T}$ and evaluated at the proposed solution $\tilde{\beta}_T$, to converge to the zero vector in probability at a faster rate than $T^{-1/2}$.

4/ A more detailed discussion of the EM algorithm is given by Dempster, Laird, and Rubin [1977].

5/ If C_T and D_T have limits C_∞ and D_∞, then result (ii) can be written in the more familiar form $T(\hat{\beta}_T - \beta_o) \xrightarrow{d} N(0, C_\infty^{-1} D_\infty C_\infty^{-1})$.

6/ The scaling sequence $\hat{c}_T$ may depend, for example, upon some concomittant estimate of the scale of the residuals.

7/ Turner [1983] has applied symmetrically censored least squares estimation to data on the share of fringe benefits in labor compensation. The successive approximation formula (2.13) was used iteratively to compute the regression coefficients, and standard errors were obtained using $\hat{C}_T$ and $\hat{D}_T$. above. The resulting estimates, while generally of the same sign as the maximum likelihood estimates assuming Gaussian errors, differed substantially in their relative magnitudes, indicating failure of the assumptions of normality or homoscedasticity.

8/ Gaussian deviates were generated by the sine-cosine transform, as applied to uniform deviates generated by a linear congruential scheme; deviates for the other error distributions below were generated by the inverse-transform method. The symmetrically censored least squares estimator was calculated throughout using (2.13) as a recursion formula, with classical least squares estimates as starting values. The Tobit maximum likelihood estimates used the Newton-Raphson iteration, again with least squares as starting values.

9/ Of course, if the loss function $\rho(\cdot)$ is not scale equivariant, some concommitant estimate of scale may be needed, complicating the asymptotic theory.

10/ This argument is somewhat more compelling in the case of an i.i.d. dependent variable than in the regression context, where conditioning on the censoring points is similar to conditioning on the observed values of the regressors. In practice, even if the data do not include censoring values for the uncensored observations, such values can usually be imputed as the potential survival time assuming the individual had lived to the completion of the study.

11/ Cosslett [1984a] has recently proposed a consistent, nonparametric two-step estimator of the general sample selection model, but its large-sample distribution is not yet known. Chamberlain [1984] and Cosslett [1984b] have investigated the maximum attainable efficiency for selection models when the error distribution is unknown.

REFERENCES

Amemiya, T. [1973], "Regression Analysis When the Dependent Variable is Truncated Normal," Econometrica, 41: 997-1016.

Arabmazar, A. and P. Schmidt [1981], "Further Evidence on the Robustness of the Tobit Estimator to Heteroskedasticity," Journal of Econometrics, 17: 253-258.

Arabmazar, A. and P. Schmidt [1982], "An Investigation of the Robustness of the Tobit Estimator to Non-Normality," Econometrica, 50: 1055-1063.

Buckley, J. and I. James [1979], "Linear Regression with Censored Data," Biometrika, 66: 429-436.

Chamberlain, G. [1984], "Asymptotic Efficiency in Semiparametric Models with Censoring," Department of Economics, University of Wisconsin-Madison, manuscript, May.

Cosslett, S.R. [1984a], "Distribution-Free Estimator of a Regression Model with Sample Selectivity," Center for Econometrics and Decision Sciences, University of Florida, manuscript, June.

Cosslett, S.R. [1984b], "Efficiency Bounds for Distribution-Free Estimators of the Binary Choice and Censored Regression Models," Center for Econometrics and Decision Sciences, University of Florida, manuscript, June.

Dempster, A.P., N.M. Laird, and D.B. Rubin [1977], "Maximum Likelihood from Incomplete Data via the EM Algorithm," Journal of the Royal Statistical Society, Series B, 39: 1-22.

Duncan, G.M. [1983], "A Robust Censored Regression Estimator", Department of Economics, Washington State University, Working Paper No. 581-3.

Efron, B. [1967], "The Two-Sample Problem with Censored Data," Proceedings of the Fifth Berkeley Symposium, 4: 831-853.

Eicker, F. [1967], "Limit Theorems for Regressions with Unequal and Dependent Errors," Proceedings of the Fifth Berkeley Symposium, 1: 59-82.

Fernandez, L. [1983], "Non-parametric Maximum Likelihood Estimation of Censored Regression Models," manuscript, Department of Economics, Oberlin College.

Goldberger, A.S. [1980], "Abnormal Selection Bias," Social Systems Research Institute, University of Wisconsin-Madison, Workshop Paper No. 8006.

Hausman, J.A. [1978], "Specification Tests in Econometrics," Econometrica, 46: 1251-1271.

Heckman, J. [1979], "Sample Bias as a Specification Error," Econometrica, 47: 153-162.

Huber, P.J. [1965], "The Behaviour of Maximum Likelihood Estimates Under Nonstandard Conditions," Proceedings of the Fifth Berkeley Symposium, 1: 221-233.

Hurd, M. [1979], "Estimation in Truncated Samples When There is Heteroskedasticity," Journal of Econometrics, 11: 247-258.

Kalbfleisch, J.O. and R.L. Prentice [1980], The Statistical Analysis of Failure Time Data, New York: John Wiley and Sons.

Koul, H., Suslarla, V., and J. Van Ryzin [1981], "Regression Analysis With Randomly Right Censored Data," Annals of Statistics, 9: 1276-1288.

Maddala, G.S. and F.D. Nelson [1975], "Specification Errors in Limited Dependent Variable Models," National Bureau for Economic Research, Working Paper No. 96.

Miller, R. G. [1976], "Least Squares Regression With Censored Data," Biometrika, 63: 449-464.

Paarsch, H.J. [1984], "A Monte Carlo Comparison of Estimators for Censored Regression," Annals of Econometrics, 24: 197-213.

Powell, J.L. [1984], "Least Absolute Deviations Estimation for the Censored Regression Model," Journal of Econometrics, 25: 303-325.

Ruppert, D. and R.J. Carroll [1980], "Trimmed Least Squares Estimation in the Linear Model," Journal of the American Statistical Association, 75: 828-838.

Ruud, P.A. [1983], "Consistent Estimation of Limited Dependent Variable Models Despite Misspecification of Distribution," manuscript, Department of Economics, University of California, Berkeley.

Tobin, J. [1958], "Estimation of Relationships for Limited Dependent Variables," Econometrica, 26: 24-36.

Turner, R. [1983], The Effect of Taxes on the Form of Compensation, Ph.D. dissertation, Department of Economics, Massachusetts Institute of Technology.

White, H. [1980], "A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity," Econometrica, 48: 817-838.

bar code is on back cover