Staff Papers Series

Staff Paper P90-8 September 1990

ITERATIVE LEAST SQUARES ESTIMATION OF CENSORED REGRESSION MODELS WITH UNKNOWN ERROR DISTRIBUTIONS

Yacov Tsur and Amos Zemel

January 31, 1990
Current Version: September 1990

Department of Agricultural and Applied Economics

University of Minnesota
Institute of Agriculture, Forestry and Home Economics
St. Paul, Minnesota 55108

Staff Paper P90-8 September 1990

ITERATIVE LEAST SQUARES ESTIMATION OF CENSORED REGRESSION MODELS WITH UNKNOWN ERROR DISTRIBUTIONS

Yacov Tsur and Amos Zemel

January 31, 1990
Current Version: September 1990

Current version: September 1990

Iterative Least Squares Estimation of Censored Regression Models With Unknown Error Distributions*

Yacov Tsur[1] and Amos Zemel[2]

A simple and tractable algorithm to estimate censored regression models with unknown error distribution is described. The algorithm is based on a new empirical estimator of the conditional expectation of the errors and is designed to yield solutions to a fixed point equation via an iterative least squares procedure. The resulting estimator is $\sqrt{N}$-consistent and asymptotically normal.

*The authors are grateful to L. Breiman for suggesting the empirical conditional expectations and for his advice. They also wish to thank L.-F. Lee, Y. Ritov, C. Sims, H. Ichimura and S. Thompson for helpful discussions.

[1]Department of Agricultural and Applied Economics, University of Minnesota, 1994 Buford Avenue, St. Paul, MN 55108, USA. Supported by BARD Grant No. US-1161-86.

[2]Center for Energy and Environmental Physics, the J. Blaustein Institute for Desert Research, Ben Gurion University of the Negev, Sede Boqer Campus, 84993, Israel.

Iterative Least Squares Estimation of Censored Regression Models With Unknown Error Distributions

Yacov Tsur and Amos Zemel

1. Introduction

Regression models with dependent variables which are incompletely observed are pervasive. Restrictions on the observations occur, for example, if a measuring device fails to give correct values beyond a given level, or when the dependent variable, by its nature, is limited to a specific range (e.g., can take only positive values). The literature is abundant with further examples. Models describing such situations, known as censored regression models, cannot be readily estimated using standard Least Squares (LS) techniques due to the bias introduced by the censoring. When the error distribution is specified the elimination of the bias is relatively simple; a detailed treatment of this case is given in Breiman, Tsur and Zemel (1989). The problem becomes involved in the more practical case of unknown error distributions.

In this work we develop an estimator of the parameter vector of censored regression models which does not require knowledge of the underlying error distribution. We describe a simple, iterative algorithm to obtain this estimator and show that it is consistent and asymptotically normal. Each iteration consists of two steps: First, one fills in the missing data using predictors based on the observations and on current parameter estimates. Improved parameter estimates are then obtained by applying LS methods as if no data are missing. The predictors used to fill in the missing data are derived from estimators of the corresponding expectations of the errors conditional on all available information. The construction of such predictors in a way which generates consistent estimators, yet requires only few, simple and fast computations is the key feature of our algorithm.

A similar procedure was suggested by Buckley and James (1979) and further investigated by James and Smith (1984) and by Ritov (1990). Our procedure differs in the way the empirical estimators of the error conditional expectations are evaluated. The proposed empirical estimators were also used by Lee (1988) in the context of semiparametric truncated regression models.

Other approaches to estimate censored regression models without specifying the error distribution have been discussed by Powell (1984, 1986), Duncan (1986), Fernandez (1986), Horowitz (1986), Nawata (1990), Tsiatis (1990), and Ritov and Fygenson (1990). The estimators studied in these works were shown to be consistent and asymptotically normal. Some of them, however, are not easy to implement and their application to data may become computationally cumbersome.

When the EM algorithm of Dempster, Laird and Rubin (1977) is applied to a censored regression model with Gaussian errors, one obtains an iterative LS procedure, where in each iteration the missing data are replaced by their expectations conditional on all available information (Tsur (1983)). The resulting estimator was shown to maximize the likelihood function, hence it is consistent and efficient. It is of interest, then, to find out whether such an iterative LS procedure maintains desirable large sample properties when applied to models with non normal (but known) distributions. Breiman, Tsur and Zemel (1989) answered this question in the affirmative and further showed that this iterative LS procedure (referred to as the EP algorithm) possesses excellent convergence properties.

Proceeding along this line of thought, the next quest to pursue concerns the properties of a similar procedure in which distribution-free empirical estimates of the error conditional expectations are employed. Indeed this is the main theme of this work. Analyzing the algorithm based on the new empirical conditional expectations, we find that the governing equations are very similar (asymptotically) to those corresponding to the case of known error distributions, and that the distribution-free EP estimator is consistent and asymptotically normal.

We begin, in Section 2, by describing the EP algorithm and define the EP estimator. In its strict form, the estimator is defined as a solution of a fixed point equation. Due to discontinuities in the empirical estimates, such a solution may not always exist in finite samples. We thus generalize the solution concept from a point to a neighborhood that shrinks (at a rate faster than $1/\sqrt{N}$) with the sample size N and prove, in Section 3, the existence of a $\sqrt{N}$-consistent solution. Consistency of the EP estimator, then, requires the identification of the consistent root if more than one solution exists. The selection of this root is based on a minimization criterion recently proposed by Lee (1988) for the truncated case. Finally, we show that the EP estimator is asymptotically normal and derive its limiting covariance matrix.

2. The EP Algorithm and Estimator

We consider a model in which the data $(y_i, \tilde{x}_i)$ are generated by the mechanism

$$y_i = \text{MAX}(0, \alpha_0 + \tilde{x}_i'\beta_0 + u_i), \quad i=1,2,\ldots,N$$

where y_i are observed scalars, $\tilde{x}_i$ are iid K-dimensional observed vectors, α_0 is an unknown intercept parameter, β_0 is a K-dimensional vector of unknown slope parameters to be estimated, u_i are iid error terms distributed according to some unknown cdf F with $E(u_i) = 0$, and N is the number of measurements. The value $(y_i=0)$ indicates that y_i is missing, otherwise $y_i>0$.

Let $x_i = \tilde{x}_i - \bar{x}$, where $\bar{x} = E_x(\tilde{x}_i)$, be the shifted regressors expressed as deviations from the mean. Let $\epsilon_i = u_i + \alpha_0 + \bar{x}'\beta_0$ be the shifted errors with mean $E(\epsilon_i) = \bar{\epsilon} = \alpha_0 + \bar{x}'\beta_0$ and cdf $F_\epsilon(z) = F(z-\bar{\epsilon})$. Let $\tilde{z}_i^o = -\alpha_0 - \tilde{x}_i'\beta_0$ and $z_i^o = -x_i'\beta_0$, so that $\tilde{z}_i^o = z_i^o - \bar{\epsilon}$. Thus, recalling $\epsilon_i = u_i + \bar{\epsilon}$, one has $F_\epsilon(z_i^o) = F(\tilde{z}_i^o)$ and $E(\epsilon|\epsilon<z_i^o) = E(u|u<\tilde{z}_i^o) + \bar{\epsilon}$. Because F_ϵ is always evaluated at a shifted argument we suppress the subscript ϵ without risking confusion.

Denoting the index sets corresponding to the observed and missing cases by $M^+ = \{i: z_i^o < \epsilon_i\} = \{i: y_i > 0\}$ and $M^- = \{i: z_i^o \geq \epsilon_i\} = \{i: y_i = 0\}$, the model can be equivalently presented as

$$y_i = \begin{cases} x_i'\beta_0 + \epsilon_i & ;\ i \in M^+ \\ 0 & ;\ i \in M^- \end{cases}, \quad i=1,2,\ldots,N. \tag{2.1}$$

Model (2.1), expressed in terms of the shifted regressors and errors, will be referred to as the shifted model. The N by K matrix whose i'th row is x_i' is denoted by X and X^+ is its partition to the observed cases, $i \in M^+$.

The EP algorithm is an iterative procedure to estimate the slope vector β_0. Each iteration consists of two steps: an Expectation (E) step and a Projection (P) step. The idea is to replace the missing values y_i, $i \in M^-$, by their expectations, using all the available information including the parameter estimate obtained in the previous iteration. These values for y_i

are then employed to find an improved estimator for β_o.

E-step: Given the values $\beta^{(r)}$ of the r'th iteration, the next values of y_i are calculated as:

$$y_i(\beta^{(r)}) = \begin{cases} y_i & ; i \in M^+ \\ x_i'\beta^{(r)} + E\{\epsilon \mid \epsilon < -x_i'\beta^{(r)}\} - \bar{\epsilon}/F(-x_i'\beta^{(r)}) & ; i \in M^- \end{cases} \tag{2.2}$$

P-step: In this step $\beta^{(r+1)}$ is found by projecting $Y(\beta^{(r)})$ on the space spanned by the columns of X:

$$\beta^{(r+1)} = (X'X)^{-1}X'Y(\beta^{(r)}), \tag{2.3}$$

where $Y(\beta^{(r)})$ is the N-dimensional vector whose elements are $y_i(\beta^{(r)})$ of eq. (2.2). Eq. (2.3) is the usual LS formula for unlimited observations. With a known error distribution, $Y(\beta^{(r)})$ is readily calculated yielding a procedure which converges geometrically to a *unique* point which is consistent and asymptotically normal (Breiman, Tsur and Zemel 1989). Lacking knowledge on the error distribution, $E\{\epsilon \mid \epsilon < -x_i'\beta^{(r)}\} - \bar{\epsilon}/F(-x_i'\beta^{(r)})$ must be estimated empirically. Buckley and James (1979) used the Kaplan-Meier Product Limit Estimator of the error distribution. We suggest the following estimator. For a real variable z and a given vector β, let

$$H(\beta,z) = \sum_{j \in M^+} (y_j - x_j'\beta)\ I(y_j - x_j'\beta \geq z > -x_j'\beta) \tag{2.4}$$

and

$$M(\beta,z) = \mathrm{MAX}\left(\sum_{j \in M^+} I(y_j - x_j'\beta \leq z) + \sum_{j \in M^-} I(-x_j'\beta < z)\ ,\ 1 \right) \tag{2.5}$$

where $I(\cdot)$ is the indicator function defined as unity when its argument is true and zero otherwise. The empirical estimator of $E\{\epsilon \mid \epsilon < z\} - \bar{\epsilon}/F(z)$, evaluated at β, is defined as

$$E_e(\beta,z) = -H(\beta,z)/M(\beta,z). \tag{2.6}$$

Substituting $E_e(\beta^{(r)}, -x_i'\beta^{(r)})$ for $E\{\epsilon \mid \epsilon < -x_i'\beta^{(r)}\} - \bar{\epsilon}/F(-x_i'\beta^{(r)})$ in eq. (2.2), the E-step is complete and can be followed by the P-step.

The implementation of the algorithm proceeds along the following steps:

1) Form the shifted regressor matrix X using the mean $\sum_{i=1}^{N} \tilde{x}_i/N$ as an estimate of $\bar{x}$.

2) Set $\beta^{(o)}$, an initial value for the parameter vector.

3) Fill in the missing y_i values as given by eq. (2.2) using $E_e(\beta, -x_i'\beta)$ with the current β-estimate (E-step).

4) Update β according to eq. (2.3) (P-step).

5) Return to step 3 unless

$$\|\beta^{(r+1)} - \beta^{(r)}\|^2 = \sum_{k=1}^{K} (\beta_k^{(r+1)} - \beta_k^{(r)})^2$$

decreases below some predetermined convergence requirement.

6) Once the convergence criterion is satisfied, adopt the last value of β as the final estimate.

The limit of this iterative process is the EP estimator and may be considered as the solution to the Fixed Point Equation (FPE)

$$\beta = (X'X)^{-1}X'Y(\beta). \tag{2.7}$$

The discontinuities (with respect to β) in E_e may cause situations in which the FPE does not have a solution (the same problem was noticed by Buckley and James (1979) for their estimator). Practical experience shows that the EP algorithm then settles to oscillations among several fixed values instead of converging to a unique fixed point. Once this situations has been identified, the iterations are terminated; the criterion to select the proper oscillation point is described below. A related ambiguity stems from the theoretical difficulty in proving the asymptotic uniqueness of the solution of the FPE. Although it is shown in the next section that the FPE must have a consistent root, it is not clear that every solution is indeed consistent. One can, however, identify the consistent root among a given set of possible solutions (arrived at, for example, by starting the algorithm at different initial values in a situation where such non uniqueness occurs), according to the following criterion: Let

$$M_c(\beta,z) = \sum_{j\in M^+} I(y_j - x_j'\beta \geq z > -x_j'\beta) \qquad (2.8)$$

and

$$E_{ec}(\beta,z) = H(\beta,z)/M_c(\beta,z) \qquad (2.9)$$

be an estimator of $E(\epsilon|\epsilon>z)$ (cf. Eq. 2.6). Define also

$$Q_N^e(\beta) = \sum_{i\in M^+} \left(y_i - x_i'\beta - E_{ec}(\beta,-x_i'\beta)\right)^2/N \qquad (2.10)$$

and evaluate Q_N^e for every solution of the FPE. The root corresponding to the minimum value of Q_N^e is adopted as the EP estimate.

The FPE is here presented as a result of a specific iterative procedure. In the following section the solutions of this equation are analyzed without reference to the particular algorithm used to obtain them. Thus, the results presented below are valid for any method of solution, yet the EP algorithm proposed here is particularly convenient for numerical implementation.

3. Asymptotic properties

We derive in this section the consistency and asymptotic normality of the EP estimator. The analysis is based on properties of the FPE. As explained above, this equation does not necessarily have a solution for every finite sample. Therefore, the term "solution to the FPE" must be generalized, allowing deviations that diminish as the sample size N increases. We show the existence of a consistent and asymptotically normal "generalized solution" to the FPE, and verify that the EP estimator coincides with this particular solution. Aiming at simplicity, the derivations are based on a set of assumptions which are somewhat restrictive but clarify the proofs. In general, we assume uniform bounds when weaker moment conditions may suffice. Several generalizations are possible, but the investigation of the minimal conditions under which the results hold will be carried out elsewhere.

In addition to the standard condition that the regressors are statistically independent of the errors, we require:

Assumption 1:

(i) x_i are iid with a distribution having a bounded support.

(ii) $X'X/N$ is uniformly positive definite (upd).

(iii) The distribution of $z_i^o = -x_i'\beta_o$, induced by the distribution of x_i, has a bounded density.

(iv) For z restricted to the bounded support of z_i^o, the error distribution is bounded: $1 - \delta > F(z) > \delta$ for some $1 > \delta > 0$, and the density $f(z) = F'(z)$ is bounded.

(This implies that the functions $E(z) = E(\epsilon|\epsilon<z)$ and $E'(z) = dE/dz$ are also bounded.)

(v) $E(\epsilon^4) < \infty$.

The empirical conditional expectation E_e is the key ingredient of the algorithm. We begin by deriving an important consistency property of E_e. Let $z_j = -x_j'\beta$ and $N(\beta,z) = \sum_{j=1}^{N} I(z_j < z)$. For $F(z) > 0$, the following result holds:

Theorem 1: $E_e(\beta_o,z) \xrightarrow{p} E(\epsilon|\epsilon<z) - \bar{\epsilon}/F(z)$ provided $N(\beta_o,z) \to \infty$.

(All proofs are presented in the appendix.) Theorem 1 claims that for the true parameter, β_o, the evaluation of E_e at any z within the support of z_i^o provides a consistent estimator of the quantity required to fill in the missing y-values (cf. Eq. (2.2)). This consistency property has motivated definitions (2.4)-(2.6) of the empirical conditional expectations. Note that the relevant sample size is $N(\beta,z)$ rather than N. This technical difficulty is addressed throughout the derivations.

Using Theorem 1 we next show that β_o is an asymptotic solution of the FPE. The notation $O_{qm}(1)$ is used to denote a variable having a mean and a variance which are o(1) and O(1), respectively. A variable is $o_{qm}(1)$ if both its mean and variance are o(1). Let $W(\beta) = Y(\beta)-X\beta$ and observe that, since $X'X/N$ is upd, $\psi(\beta)=X'W(\beta)/\sqrt{N}$ can serve as a measure of the degree of precision to which the FPE is satisfied. At β_o, ψ assumes a particularly simple form. Let $s_i = \epsilon_i I(\epsilon_i>z_i^o) + (E(z_i^o)-\bar{\epsilon}/F(z_i^o))I(\epsilon_i\le z_i^o)$; $\sigma_i^2=\mathrm{Var}(s_i)$, Σ be the N by N diagonal matrix with elements σ_i^2 and $V = E_x(X'\Sigma X/N)$. Assumption 1 ensures that V exists and is positive definite. Then, we can prove

Theorem 2: Under Assumption 1, $\psi(\beta_o) \xrightarrow{D} N(0,V)$.

Theorem 2 immediately implies that β_o is an asymptotic solution of the FPE:

Corollary: Under Assumption 1, $\sqrt{N}((X'X)^{-1}X'Y(\beta_o) - \beta_o) = O_{qm}(1)$.

Moreover, Theorem 2 plays a key role in the derivation of the asymptotic distribution of the EP estimator (cf. Theorem 6 below).

The observation that β_o solves the FPE asymptotically suggests the existence of solutions that approach β_o as $N \to \infty$. However, as mentioned in

Section 2, the FPE may not have an exact solution for any finite sample. Nonetheless, the solution concept can be slightly generalized to vectors that satisfy the FPE to a *better* approximation than β_o. Then, as shown below, it is possible to verify the existence of a consistent and asymptotically normal "generalized solution", which coincides with the EP estimate.

Definition: A vector β is a solution to the FPE if $\psi(\beta) = o_{qm}(1)$, that is if $\sqrt{N}((X'X)^{-1}X'Y(\beta) - \beta) \to_{qm} 0$.

Since $\psi(\beta_o) = O_{qm}(1)$, β_o does not qualify as a solution and we need some preparations to show that generalized solutions indeed exist.

Let $x_i^{(o)}$, $X^{(o)}$ and $z_i^{(o)}$ represent respectively x_i, X and z_i^o after ordering the regressors x_i according to their projections on β_o:

$$i < j \iff z_i^o < z_j^o.$$

Define the K by K matrix

$$\Omega_N = X^{(o)\prime}\Gamma(I-A)X^{(o)}/N. \tag{3.1}$$

Here Γ is an N by N diagonal matrix with

$$\Gamma_{ii} = \gamma(z_i^{(o)}) = F_iE_i' + 1 - F_i + \bar{\epsilon}f_i/F_i, \tag{3.2}$$

where F_i, E_i' and f_i are evaluated at $z_i^{(o)}$ and A is the N by N matrix

$$A = \begin{pmatrix} 0 & 0 & 0 & 0 & \cdots & 0 & 0 \\ 1 & 0 & 0 & 0 & \cdots & 0 & 0 \\ 1/2 & 1/2 & 0 & 0 & \cdots & 0 & 0 \\ 1/3 & 1/3 & 1/3 & 0 & \cdots & 0 & 0 \\ \vdots & & & & & & \vdots \\ 1/(n-1) & 1/(n-1) & 1/(n-1) & 1/(n-1) & \cdots & 1/(n-1) & 0 \end{pmatrix} \tag{3.3}$$

The ordering has been introduced in order to specify A as a fixed (non random) matrix. Without ordering, the rows of A would have to be permuted according to the (random) order of z_i^o.

We are now ready to establish the following result:

Theorem 3: Under Assumption 1, for any β such that $\Delta\beta = \beta-\beta_o = O(1/\sqrt{N})$,

$$\Delta\psi = \psi(\beta) - \psi(\beta_o) = -\Omega_N\sqrt{N}\Delta\beta + o_{qm}(1).$$

According to Theorem 3, the FPE is essentially linear in a small region around

β_0 and the matrix Ω_N is nothing but the derivative of $-\psi(\beta)/\sqrt{N}$ with respect to β. The condition needed to guarantee the existence of a consistent solution to the FPE is therefore equivalent to the condition required for Ω_N to be uniformly nonsingular. In fact, a solution can (in principle) be constructed explicitly, using the Newton-Raphson value $\hat{\beta}_N = \beta_0 + \Omega_N^{-1}\psi(\beta_0)/\sqrt{N}$. The matrix Ω_N plays here the role played by $X'X/N$ in the uncensored regression model. For censored regression with a known error distribution, the corresponding matrix is $X'\Gamma X/N$ (Breiman, Tsur and Zemel, 1989). The matrix A represents, therefore, the modifications introduced by the use of the empirical conditional expectation.

The explicit form of Ω_N can now be used to investigate its properties, taking into account the random nature of X. (It so happens that the matrix corresponding to Ω_N in the unshifted model becomes singular as $N \to \infty$; this explains the use of the shifted model.)

As mentioned above, the cost of having a non-random A is the need to order x_i, which disturbs the independence among the rows of X. Therefore, it is expedient to introduce a normalized regressor matrix $Z = \begin{pmatrix} \zeta_1' \\ \vdots \\ \zeta_N' \end{pmatrix}$ on which the effect of the ordering is restricted in the following sense: the elements of the first column retain the original ordering of z_i^o whereas the elements of the other columns, while not yet independent, are uncorrelated with zero means.

Let $\Lambda = E_X(X'\Gamma X/N)$, $C_0 = \|\Lambda^{1/2}\beta_0\|$ and $b_1 = \Lambda^{1/2}\beta_0/C_0$. Construct K-1 unit vectors $b_2 \ldots b_K$ such that the K by K matrix $B = (b_1 \ldots b_K)$ is orthogonal. The normalized regressor matrix is given by $Z = -X\Lambda^{-1/2}B$. It is verified that $\zeta_{i1} = -x_i'\beta_0/C_0$. We denote the cdf of these quantities, induced by the distribution of x_i, by $F_z(\cdot)$. The definition of Z is meaningful, and its desirable properties are guaranteed if the following assumptions hold:

Assumption 2:

(i) $X'\Gamma X/N$ is upd;

(ii) $E\left(\zeta_{ik}|\zeta_{i1}=z\right) = 0$ for all $k>1$ and all z.

Condition (i) is standard for censored regression models and its validity is discussed in detail in Breiman, Tsur and Zemel (1989) in the context of known error distributions. In view of Assumption 1, it holds trivially if $\bar{\epsilon}\geq 0$. Condition (ii) implies some symmetry on the distribution of the vectors ζ_i. It holds, for example, for any distribution that depends only on the norm of its argument, i.e. the regressors x_i (after normalization) have no preferred direction in the K-dimensional space. Weaker symmetries are, in fact, sufficient.

We denote by $Z^{(o)} = \left(\zeta_{ik}^{(o)}\right)$ the ordered normalized regressor matrix and observe that $Z'^{(o)}\Gamma(I-A)Z^{(o)}/N = B'\Lambda^{-1/2}\Omega_N\Lambda^{-1/2}B$. Thus it is sufficient to investigate the conditions for the nonsingularity of the former matrix. The condition involves the distributions of both the errors and regressors and takes the form:

Assumption 3: $\theta = \int\zeta\ \gamma(C_o\zeta)\left(\zeta-E_z(\zeta)\right)f_z(\zeta)d\zeta = \int\zeta\ \gamma(C_o\zeta)E'_z(\zeta)F_z(\zeta)d\zeta \neq 0$
where $E_z(\zeta) = E(\zeta_{i1}|\zeta_{i1}<\zeta)$ and $E'_z(\zeta) = dE_z(\zeta)/d\zeta = \left(\zeta-E_z(\zeta)\right)f_z(\zeta)/F_z(\zeta)$.

Assumption 3 implies that the typical increase in $\zeta-E_z(\zeta)$ is not exactly counter balanced by the decreasing function γ. By the definition of Z, $\int\zeta^2\gamma(C_o\zeta)f_z(\zeta)d\zeta = 1$, so that an alternative way of writing the condition is $\int\zeta\ \gamma(C_o\zeta)E_z(\zeta)f_z(\zeta)d\zeta \neq 1$. The next result is based on the observation that $Z'^{(o)}\Gamma(I-A)Z^{(o)}/N$ converges (in quadratic mean) to the unit matrix except for its 1,1 element which equals θ. Thus we arrive at the following theorem, which establishes the existence of a $\sqrt{N}$-consistent solution to the FPE.

Theorem 4: Under Assumptions 1-3:

(i) Ω_N converges in quadratic mean to a nonsingular limit Ω.

(ii) $\hat{\beta} = \beta_0 + \Omega^{-1}\psi(\beta_0)/\sqrt{N}$ is a $\sqrt{N}$-consistent solution to the FPE.

In view of Theorem 4, the EP algorithm (or any alternative method of solving the FPE) seems to provide a promising procedure for generating consistent estimators. Indeed, if one starts at a β-value which is close enough to β_0, the linear nature of $\psi(\beta)$ in that region ensures that the consistent solution will be found. However, the results presented so far discuss local properties only, and do not rule out the existence of solutions which are remote from β_0. In order to establish consistency one needs global results that ensure uniqueness (to $O(1/\sqrt{N})$) of the solution. For this purpose, too, the Ω matrix formalism might prove useful since a generalization of Theorem 3 to arbitrary $\Delta\beta$ entails consistency if the generalized Ω is nonsingular. Indeed, for certain simple distributions explicit expressions for this matrix could be derived. The identification of sufficient conditions for nonsingularity is, however, more involved.

An alternative approach is based on recent results obtained by Lee (1988) for truncated regression models. Using a smooth version of the empirical conditional expectations, Lee (1988) constructed a consistent estimator for the truncated model by minimizing a sum of mean-corrected squared errors. Obviously, Lee's method can be applied also to censored models. However, smoothing procedures tend to complicate the computations and an attractive feature of the EP algorithm is lost.

In the censored case it is preferable, therefore, to employ the EP algorithm to obtain solutions to the FPE. When more than one solution is found, the selection of the consistent root proceeds as described in Section 2. Let $F_C(z) = 1 - F(z)$; $E_{1C}(z) = E(\epsilon|\epsilon>z)$ and $E_{2C}(z) = E(\epsilon^2|\epsilon>z)$. An empirical estimator to $E_{1C}(z)$ is defined in eq. (2.9) and used to construct the objective function $Q_N^e(\beta)$ in eq. (2.10). Taking expectation over ϵ_j, we define the following quantities:

$$h_j(\beta,z) = F_C(z + \Delta_j)I(z > z_j)(E_{1C}(z + \Delta_j) - \Delta_j);$$

$$m_j(\beta,z) = F_c(z + \Delta_j)I(z > z_j); \quad \overline{E_c}(\beta,z) = \sum_{j=1}^{N} h_j(\beta,z) \Big/ \sum_{j=1}^{N} m_j(\beta,z)$$

and

$$Q_N(\beta) = \sum_{i=1}^{N} F_c(z_i^o)\left\{\mathrm{Var}(\epsilon|\epsilon>z_i^o) + \left(\overline{E_c}(\beta,z_i)+\Delta_i-E_{1c}(z_i^o)\right)^2\right\} / N$$

$$= Q_N(\beta_o) + \sum_{i=1}^{N} F_c(z_i^o)\left(\overline{E_c}(\beta,z_i)+\Delta_i-E_{1c}(z_i^o)\right)^2 / N.$$

Obviously, β_o minimizes Q_N (note that the rightmost term above vanishes at β_o). Since it can be shown that $\left(Q_N^e(\beta)-Q_N(\beta)\right) \rightarrow_{qm} 0$, we can expect that of all solutions to the FPE, the one that yields the lowest value of $Q_N^e(\beta)$ is the consistent root. In fact, some additional assumptions are required. Let $E_c(\beta,z_i) = E_{ux}(y+z|y+z>z_i>z)$. Note that E_c is continuous in β, and $E_c(\beta_o,z_i^o) = E_{1c}(z_i^o)$.

Assumption 4:

(i) β_o is an interior point of a compact set S_β.

(ii) For every $\beta \in S_\beta$ and x_i, x_j in the support of X, the error cdf satisfies $F_c(z_i+\Delta_j) > \delta > 0$.

(This condition is a generalization of Assumption 1(iv)).

(iii) For every $\beta \in S_\beta$, $E_X\left\{|E_{1c}(z^o)-z^o-(E_c(\beta,z)-z)|\right\} \neq 0$ if $\beta \neq \beta_o$.

The identification condition (iii) ensures that β_o is the unique minimizer of $Q_N(\beta)$. It was originally proposed by Lee (1988) who gave heuristic arguments for its validity. With the aid of Assumption 4, the following consistency theorem can be derived:

Theorem 5: Let $\{\beta_m\}_{m=1...M}$ be the set of distinct (to $O(1/\sqrt{N})$) solutions of the FPE. Then, under Assumptions 1-4, $\hat{\beta} = \underset{m=1..M}{\mathrm{argmin}}\ Q_N^e(\beta_m)$ is a $\sqrt{N}$-consistent estimator of β_o.

Theorem 5 is similar to Theorem 4.1 of Lee (1988) but differs in two important respects. First, Q_N^e is the nonsmooth objective function. Second, the minimization is carried out over a finite set. In this way we utilize the global properties of Lee's procedure while retaining the computational simplicity of the EP algorithm.

Applying Theorems 3 and 4 to the $\sqrt{N}$-consistent $\hat{\beta}$ gives

$$\psi(\beta_o) = -\Delta\psi + o_{qm}(1) = \Omega_N\sqrt{N}(\hat{\beta} - \beta_o) + o_{qm}(1) = \Omega\sqrt{N}(\hat{\beta} - \beta_o) + o_{qm}(1).$$

According to Theorem 2, $\psi(\beta_o) \xrightarrow{D} N(0,V)$ while Theorem 4 ensures that Ω is nonsingular. Thus we arrive at

Theorem 6: Under Assumptions 1-3, $\sqrt{N}(\hat{\beta} - \beta_o) \xrightarrow{D} N(0,\Omega^{-1}V\Omega'^{-1})$.

4. Concluding remarks

A major problem in the study of estimation procedures of censored or truncated regression models which are robust with respect to the specification of the error distribution is the need to establish asymptotic uniqueness of the solutions. Many of the estimators are defined as the extremum points of some underlying objective function. Estimators of this kind benefit from the well-developed techniques to analyze extremum estimators, and the conditions under which they are consistent (i.e., the true parameter is asymptotically a unique extremum point) are usually identified. However, these estimators tend to be computationally cumbersome, since they entail optimization of objective functions which are either non-differentiable or require smoothing procedures.

A different class of estimators is defined by fixed points of some iterative estimation procedure. These estimators are often more tractable computationally, but it is more difficult to verify that their estimation equations have unique solutions. (See, for example, Ritov (1990) on the properties of the Buckley and James (1979) estimator). Under favorable conditions, the fixed point solutions are also the extremum points of objective functions. The EP estimator, for example, maximizes the likelihood function when the errors have normal distribution (Tsur, 1983) and a convex generalized sum-of-squares function for non-normal, but known, error distributions (Breiman, Tsur and Zemel, 1989). For the general, distribution-free, case the construction of a proper objective function is more difficult.

These observations lead us to the structure of the EP estimator proposed in this work. It is produced by an iterative algorithm which first locates the solutions of the estimation equation, then selects the consistent root if multiple roots are found. The second stage establishes the connection to the extremum estimators and ensures the asymptotic uniqueness of the solution. However, both stages employ the simple, discontinuous empirical conditional expectations, permitting fast and easy numerical implementations.

The insistence on the simple empirical conditional expectation entails a certain complication in the theoretical analysis: the relevant sample size for the empirical estimators is only a fraction of N. Thus, the convergence of these estimators is not uniform. This situation is in contrast to Lee's study (1988), where smoothing and trimming procedures ensure uniform convergence and simplify the subsequent analysis. For the EP case, however, it is found that the relevant correction terms are typically $O(\log N/\sqrt{N})$ and do not affect the asymptotic properties of the estimator.

Appendix: *Proofs of theorems*

The following result is central to the analysis.

Lemma 1: For some vector β, let $z_j = -x_j'\beta$, $\Delta\beta = \beta - \beta_o$, $\Delta_j = x_j'\Delta\beta$,
$N(\beta,z) = \sum_{j=1}^{N} I(z_j<z)$, $\bar{\Delta} = \sum_{j=1}^{N} \Delta_j I(z_j<z)/N(\beta,z)$,

$$E_1(z) = E(z) - \bar{\epsilon}/F(z) + \bar{\Delta}\Big(E'(z)+(1-F(z))/F(z)+\bar{\epsilon}F'(z)/F^2(z)\Big)$$

Then, for z such that $F(z) \neq 0$, $F(z+\bar{\Delta}) \neq 0$ and $N(\beta,z) \neq 0$:

(i) $E\Big\{E_e(\beta,z)\Big\} = E_1(z) + O(N(\beta,z)^{-1}) + O(\bar{\Delta}^2)$;

(ii) $Var\Big\{E_e(\beta,z)\Big\} = O(N(\beta,z)^{-1})$.

Proof: Rewrite Eqs. (2.4) and (2.5) as

$$H(\beta,z) = \sum_{j=1}^{N}(\epsilon_j-\Delta_j)I(\epsilon_j\geq z+\Delta_j)I(z_j<z),$$
$$M(\beta,z) = \sum_{j=1}^{N}I(\epsilon_j<z+\Delta_j)I(z_j<z).$$

(The probability that $M(\beta,z) = 0$ is $O\Big((1-F(z+\bar{\Delta}))^{N(\beta,z)}\Big)$ so the corrections due to the exclusion of this case can be neglected). Taking expectation with respect to ϵ, one finds

$$E\Big\{\epsilon I(\epsilon\geq z+\Delta)\Big\} = \bar{\epsilon} - E(z+\Delta)F(z+\Delta) = \bar{\epsilon} - E(z)F(z) - (E(z)F(z))'\Delta + O(\Delta^2).$$

Thus

$$E_H = E\left\{\frac{-H(\beta,z)}{N(\beta,z)}\right\} = -\bar{\epsilon} + E(z)F(z) + (E(z)F(z))'\bar{\Delta} + (1-F(z))\bar{\Delta} + O(\bar{\Delta}^2).$$

Similarly

$$E_M = E\left\{\frac{M(\beta,z)}{N(\beta,z)}\right\} = F(z) + F'(z)\bar{\Delta} + O(\bar{\Delta}^2),$$
$$Cov\left\{\frac{-H(\beta,z)}{N(\beta,z)}, \frac{M(\beta,z)}{N(\beta,z)}\right\} = -E_H E_M/N(\beta,z) + O(\bar{\Delta}/N(\beta,z))$$

and

$$Var\left\{\frac{-H(\beta,z)}{N(\beta,z)}\right\} = O(N(\beta,z)^{-1}).$$

Note that the central moments of $\frac{M(\beta,z)}{N(\beta,z)}$ are of the same order (in powers of $N(\beta,z)$) as those of $\frac{\bar{M}(\beta,z)}{N(\beta,z)} = \sum_{j=1}^{N} I\Big(\epsilon_j<F^{-1}(E_M)\Big)I(z_j<z) / N(\beta,z)$, which may be viewed as the average of $N(\beta,z)$ iid Bernoulli trials with success probability

E_M. Thus, we apply the well known bounds on the moments of Bernoulli trials and the bound $M(\beta,z)\geq 1$ to verify, using Lemma 1.1, that the mean and variance of $B = \left(\frac{\{E_M - M(\beta,z)/N(\beta,z)\}^2}{M(\beta,z)/N(\beta,z)} - \frac{1-E_M}{N(\beta,z)}\right)$ satisfy:

$$E\{B\} = E\left(\frac{\{E_M - M(\beta,z)/N(\beta,z)\}^2}{E_M}\right) - \frac{1-E_M}{N(\beta,z)} + E\left(\frac{\{E_M - M(\beta,z)/N(\beta,z)\}^3}{E_M\, M(\beta,z)/N(\beta,z)}\right) = O(N(\beta,z)^{-2})$$

Similarly, $\mathrm{Var}(B) = O(N(\beta,z)^{-2})$.

The empirical conditional expectation is written as

$$E_e(\beta,z) = \frac{-H(\beta,z)}{M(\beta,z)} = \frac{-H(\beta,z)/N(\beta,z)}{E_M}\left(1 + \frac{E_M - M(\beta,z)/N(\beta,z)}{E_M} + \frac{(1-E_M)/E_M}{N(\beta,z)} + B/E_M\right)$$

from which we derive $E\{E_e\} = E_H/E_M + O(N(\beta,z)^{-1})$. (The contribution of B is neglected using $\mathrm{Cov}^2\left\{\frac{-H(\beta,z)}{N(\beta,z)}, B\right\} \leq \mathrm{Var}\left\{\frac{-H(\beta,z)}{N(\beta,z)}\right\}\cdot\mathrm{Var}(B)$). Expanding to $O(\bar{\Delta}^2)$, one finds

$$E_H/E_M = E_1(z) + O(\bar{\Delta}^2)$$

yielding the desired value for $E\{E_e\}$.

The bound on $\mathrm{Var}\{E_e\}$ is obtained along the same lines and requires the evaluation of higher moments of $H(\beta,z)$, $M(\beta,z)$ and B. The details are omitted.

Lemma 1.1. Let M be the maximum between 1 and the sum of N i.i.d Bernoulli variates with success probability $F > 0$, then, for all $k \geq 3$:

$$\text{(i)}\quad E\left\{\frac{(F-M/N)^k}{M/N}\right\} = O(1/N^2);\qquad \text{(ii)}\quad E\left\{\frac{(F-M/N)^k}{(M/N)^2}\right\} = O(1/N^2).$$

Proof: Let $\alpha > 0$ and $d_k = E\left\{\frac{|F-M/N|^k}{M/N}\right\}$, then $d_k = O(N^{1-\alpha})$ for all $k \geq 3\alpha$. To see this, define $b = \begin{cases} N^{-\alpha} & \text{if } |F-M/N|^k < N^{-\alpha} \\ 1 & \text{if } |F-M/N|^k \geq N^{-\alpha} \end{cases}$, thus $b \geq |F-M/N|^k$ and therefore $d_k \leq E(b/(M/N)) = E(1/(M/N)|b=1)\Pr(b=1) + N^{-\alpha}E(1/(M/N)|b\neq 1)\Pr(b\neq 1)$. Now, $\Pr(b=1) \leq N^{2\alpha}E(|F-M/N|^{2k}) = O(N^{2\alpha-k})$ and $E(1/(M/N)|b=1) \leq N$, so the first term is $O(N^{2\alpha+1-k}) = O(N^{1-\alpha})$ if $k \geq 3\alpha$. For the second term, $E(1/(M/N)|b\neq 1) \leq N$ and $N^{-\alpha}E(1/(M/N)|b\neq 1)Pr(b\neq 1) = O(N^{1-\alpha})$. Choosing $\alpha = 3$, it follows that $d_k = O(N^{-2})$ for all $k \geq 9$. For $k = 8$, write $E\left\{\frac{|F-M/N|^8}{M/N}\right\} \leq E\left\{\frac{|F-M/N|^8}{F}\right\} + E\left\{\frac{|F-M/N|^9}{M/N}\right\}/F$. The first term is $O(N^{-4})$ and the second term has $k = 9$. In this way we can reduce the exponent to $k=4$. For $k=3$, use $E((F-M/N)^3) = O(N^{-2})$ to get $E\left\{\frac{(F-M/N)^3}{M/N}\right\} = O(N^{-2})$. Part (ii) is derived in the same way.

Theorem 1: $E_e(\beta_0,z) \xrightarrow{p} E(\epsilon|\epsilon<z) - \bar{\epsilon}/F(z)$ provided $N(\beta_0,z) \to \infty$.

Proof: Follows immediately from Lemma 1, setting $\bar{\Delta} = 0$ and letting $N(\beta_0,z) \to \infty$.

Theorem 2: Under Assumption 1, $\psi(\beta_0) \xrightarrow{D} N(0,V)$.

Proof: The derivation is based on U-statistics techniques to resolve the difficulties due to the dependence among w_i. We introduce the following short-hand notation: $I_{ji}=I(z_j^o<z_i^o)$; $N_i=N(\beta_0,z_i^o)=\sum_{j=1}^{N} I_{ji}$; $H_i=H(\beta_0,z_i^o)$; $M_i=M(\beta_0,z_i^o)$; $F_i=F(z_i^o)$; $E_{1i}=E(\epsilon|\epsilon\leq z_i^o)-\bar{\epsilon}/F_i$; $I_i=I(\epsilon_i\leq z_i^o)$; $s_i=\epsilon_i(1-I_i)+E_{1i}I_i$; $r_i=(E_e(\beta_0,z_i^o)-E_{1i})I_i$; $w_i=s_i+r_i$; $(N_{max},s_{max},z^o_{max}) = \begin{cases}(N_i,s_i,z_i^o) & \text{if } z_i^o>z_j^o\\ (N_j,s_j,z_j^o) & \text{if } z_j^o\geq z_i^o\end{cases}$; and HOT denotes high-order-terms involving powers of N_i and N_j such that $\frac{1}{N}\sum_i\sum_j HOT = o(1)$.

Following Lemma 1 we write

$$E_e(\beta_0,z_i^o) = \frac{-H_i}{N_iF_i}\left(1+\frac{F_i-M_i/N_i}{F_i}+\frac{(1-F_i)/F_i}{N_i}+B_i^o/F_i\right).$$

Thus, $r_i = r_{1i} + r_{2i}$ where $r_{1i}=(E_{1i}F_i-H_iB_i^o)I_i/(N_iF_i^2)$ and

$$r_{2i} = \left\{ \frac{-H_i}{N_iF_i} \left(1+ \frac{F_i - M_i/N_i}{F_i} + \frac{(1-F_i)/F_i}{N_i}\right) - E_{1i}\left(1+ \frac{1}{N_iF_i}\right)\right\} I_i .$$

The term involving r_{1i} can be ignored, as $\sum_{i=1}^{N} x_i r_{1i}/\sqrt{N} \to_{qm} 0$ (cf. Lemma 1). The second term has been constructed to ensure that $E(r_{2i})=0$, thus $\sum_{i=1}^{N} x_i r_{2i}/\sqrt{N}$ has the structure of U-statistics, albeit not in the standard symmetric form. Indeed, a straightforward evaluation gives

$$\Psi_{Nm} = E\left(\sum_{i=1}^{N} x_i r_{2i}/\sqrt{N} \mid \epsilon_m\right) = - \sum_i \frac{x_i I_{mi}}{N_i} \left(s_{mi} + \frac{s'_{mi}}{N_iF_i}\right)/\sqrt{N},$$

where $s_{mi} = \epsilon_m I(\epsilon_m > z_i^\circ) + E_{1i} I(\epsilon_m \le z_i^\circ)$ and $s'_{mi} = (\epsilon_m I(\epsilon_m > z_i^\circ) + E_{1i}F_i) + E_{1i}F_i(I(\epsilon_m \le z_i^\circ) - F_i)$. Notice that each Ψ_{Nm} depends only on ϵ_m and hence is independent of $\Psi_{Nm'}$, for all $m' \neq m$. Thus $\Psi_N = \sum_{m=.}^{N} \Psi_{Nm}$ is a convenient approximation to $\sum_{i=1}^{N} x_i r_i/\sqrt{N}$. Obviously, $E(\Psi_N)=0$. Furthermore, $E(s_{mi}s_{mj}) = Var(s_{max})$, $E(s'_{mi}s_{mj}) = O(1)$, $E(s'_{mi}s'_{mj}) = O(1)$ and $Var(\Psi_{Nm}) = \frac{1}{N}\sum_i \sum_j x_i x'_j I_{mi} I_{mj} Var(s_{max})/(N_iN_j) + HOT$. Since $I_{mi}I_{mj} = I\left(z_m^\circ < \min(z_i^\circ, z_j^\circ)\right)$, the summation over m is easily carried out:

$$Var(\Psi_N) = \sum_m Var(\Psi_{Nm}) = \frac{1}{N}\sum_i \sum_j x_i x'_j Var(s_{max})/N_{max} + o(1).$$

Following Lemma 1, we find $E(r_{2i}) = 0$ and $Cov(r_{2i}, r_{2j}) = Var(s_{max})/N_{max} + HOT$, so that $Var\left(\sum_{i=1}^{N} x_i r_{2i}/\sqrt{N}\right) = \frac{1}{N}\sum_i \sum_j x_i x'_j Var(s_{max})/N_{max} + o(1)$. It follows that $E\left(\sum_{i=1}^{N} x_i r_{2i}/\sqrt{N} - \Psi_N\right)^2 = Var\left(\sum_{i=1}^{N} x_i r_{2i}/\sqrt{N}\right) - Var(\Psi_N) = o(1)$ (cf. Lehmann, 1975, pp. 362-363). Thus, $\left(\sum_{i=1}^{N} x_i r_i/\sqrt{N} - \Psi_N\right) \to_{qm} 0$.

Having observed that the term with s'_{mi} has a negligible contribution, we consider the quantities $\omega_{Nm} = \left(x_m s_m - \sum_i \frac{x_i I_{mi}}{N_i} s_{mi}\right)/\sqrt{N}$ and obtain $\left(\psi(\beta_0) - \sum_m \omega_{Nm}\right) \to_{qm} 0$. The moments of ω_{Nm} are evaluated in the same way as those of Ψ_{Nm}: $E(\omega_{Nm}) = 0$ and

$$E(\omega_{Nm}\omega'_{Nm}) = x_m x'_m Var(s_m)/N + \frac{1}{N}\sum_i \sum_j x_i x'_j I_{mi} I_{mj} Var(s_{max})/(N_i N_j)$$

$$- \frac{1}{N}\sum_i x_i x'_m I_{mi} Var(s_i)/N_i - \frac{1}{N}\sum_j x_m x'_j I_{mj} Var(s_j)/N_j .$$

Summing over m, the contributions of the last 3 terms add to

$\frac{1}{N}\sum_m x_m x'_m Var(s_m)/N_m = o(1)$. Thus,

$$Var\left(\sum_m \omega_{Nm} \right) = \sum_m x_m x'_m Var(s_m)/N + o(1) \rightarrow X'\Sigma X/N \rightarrow_{qm} V$$

(the qm-convergence is with respect to the X-distribution). By Assumption 1, V is positive definite. A similar derivation, using $E(|s^3|)<\infty$, yields $E\left[(\sum_m \omega_{Nmk})^3\right] = O(\log^2(N)/\sqrt{N}) = o(1)$ for every component k=1,2,..,K of ω_{Nm}. It follows that the quantities $(X'\Sigma X/N)^{-1/2}\omega_{Nm}$ form a double array satisfying all the required conditions for the CLT to hold (cf. Chung, 1974, Theorem 7.1.2). Thus $(X'\Sigma X/N)^{-1/2}\sum_m \omega_{Nm} \xrightarrow{D} N(0,I)$ and $\psi(\beta_o) \xrightarrow{D} N(0,V)$, as asserted.

Theorem 3: Under Assumption 1, for any β such that $\Delta\beta \equiv \beta-\beta_o = O(1/\sqrt{N})$,

$$\Delta\psi = \psi(\beta) - \psi(\beta_o) = -\Omega_N\sqrt{N}\Delta\beta + o_{qm}(1).$$

Proof: For $\beta = \beta_o + \Delta\beta$ we write

$w_i(\beta) = (\epsilon_i-\Delta_i)I(\epsilon_i>z_i^o) + E_e(\beta,z_i)I(\epsilon_i\le z_i^o)$. Using Lemma 1, we obtain

$$E\left\{w_i(\beta)\right\} = \bar{\epsilon} - E(z_i^o)F(z_i^o) - \Delta_i(1-F(z_i^o)) + F(z_i^o)E\left\{E_e(\beta,z_i)\right\} + O(\Delta^2)$$

$$= \gamma(z_i^o)(\bar{\Delta}_i(\beta)-\Delta_i) + O(N(\beta,z_i)^{-1}) + O(\Delta^2).$$

The function γ is defined by eq. (3.2) and $\bar{\Delta}_i(\beta) = \sum_{j=1}^N \Delta_j I(z_j<z_i)/N(\beta,z_i)$. Ordering x_i according to z_i, it is seen that the vectors with elements Δ_i and $\bar{\Delta}_i$ can be written as $X^{(o)}\Delta\beta$ and $AX^{(o)}\Delta\beta$, respectively where $X^{(o)}$ is formed from the ordered regressors and A is defined by eq. (3.3). Thus, recalling that $E\{\psi(\beta_o)\} = o(1)$,

$$E\{\Delta\psi\} = X^{(o)\prime}\Gamma(A-I)X^{(o)}\Delta\beta/\sqrt{N} + O\left(\sum_{i=1}^N i^{-1}/\sqrt{N}\right) + O\left(\sum_{i=1}^N N^{-3/2}\right)$$

$$= -\Omega_N \sqrt{N}\Delta\beta + o(1).$$

This result is not yet quite what is needed, since the ordering is carried out according to z_i rather than z_i^o, leaving Ω_N dependent on β. To remedy for this we show in Lemma 2 that $\bar{\Delta}_i(\beta) - \bar{\Delta}_i(\beta_o) = O(N(\beta_o, z_i^o)^{-1})$. It follows that the additional term introduced by evaluating Ω_N at β_o is also of o(1). The more tedious evaluation of $\mathrm{Var}\{\Delta\psi\}$ is carried out in the same way as the derivation of $\mathrm{Var}\{\psi(\beta_o)\}$ in the proof of Theorem 2. One notes that the leading terms are proportional to $\Delta\beta$ and therefore $\mathrm{Var}\{\Delta\psi\} = o(1)$ although both $\mathrm{Var}\{\psi(\beta_o)\}$ and $\mathrm{Var}\{\psi(\beta)\}$ are of O(1).

Lemma 2: Under Assumption 1

(i) $E_X\left\{N(\beta, z_i) - N(\beta_o, z_i^o)\right\} = O(N\|\Delta\beta\|)$.

(ii) $E_X\left\{N(\beta_o, z_i^o)\left(\bar{\Delta}_i(\beta) - \bar{\Delta}_i(\beta_o)\right)\right\} = O(N\|\Delta\beta\|^2)$.

In particular, for $\Delta\beta = O(1/\sqrt{N})$, $E_X\left\{N(\beta_o, z_i^o)\left(\bar{\Delta}_i(\beta) - \bar{\Delta}_i(\beta_o)\right)\right\} = O(1)$.

Proof: $N(\beta, z_i) - N(\beta_o, z_i^o) = \sum_{j\neq i} I(z_j<z_i) - I(z_j^o<z_i^o) = \sum_{j\neq i} I(z_j^o-\Delta_j<z_i^o-\Delta_i) - I(z_j^o<z_i^o)$.

Let $\epsilon = 2\cdot\mathrm{Sup}_x\|x\|\,\|\Delta\beta\| = O(\Delta\beta)$, then $|N(\beta, z_i) - N(\beta_o, z_i^o)| \le \sum_{j\neq i} I(|z_j^o - z_i^o|<\epsilon)$.

For a given z_i^o, (i) follows immediately from Assumption 1-(iii). Furthermore,

$$\bar{\Delta}_i(\beta) - \bar{\Delta}_i(\beta_o) = \sum_{j\neq i} \frac{\Delta_i(I(z_j<z_i) - I(z_j^o<z_i^o))}{N(\beta_o, z_i^o)} - \bar{\Delta}_i(\beta)\frac{N(\beta, z_i) - N(\beta_o, z_i^o)}{N(\beta_o, z_i^o)}$$

or $N(\beta_o, z_i^o)|\bar{\Delta}_i(\beta) - \bar{\Delta}_i(\beta_o)| \le 2\epsilon \sum_{j\neq i} I(|z_j^o - z_i^o|<\epsilon)$ and (ii) is derived in the same way as (i). Corresponding bounds can be deduced for the variances.

The proof of Theorem 4 utilizes:

Lemma 3: Under assumption 2, for all ς in the domain of F_z

(i) $E\left(\varsigma_{ik}^{(o)} | \varsigma_{i1}^{(o)} = \varsigma\right) = 0$ for k>1 and all i.

(ii) $E\left(\zeta_{ik}^{(o)}\zeta_{jk'}^{(o)}|\zeta_{i1}^{(o)}=\zeta\right) = 0$ for $k>1$ or $k'>1$ and $i\neq j$.

(iii) $E\left(\zeta_{ik}^{(o)}\zeta_{jk'}^{(o)}\zeta_{1k''}^{(o)}\zeta_{mk'''}^{(o)}|\zeta_{i1}^{(o)}=\zeta\right) = 0$ for $k>1$ and $i\neq j$ $i\neq l$ $i\neq m$, or $k'>1$ and $j\neq i$ $j\neq l$ $j\neq m$, or $k''>1$ and $l\neq i$ $l\neq j$ $l\neq m$, or $k'''>1$ and $m\neq i$ $m\neq j$ $m\neq l$.

Remark: It follows that the corresponding unconditional expectations also vanish, so that except for $k=1$, $\zeta_{ik}^{(o)}$ mimic the properties of the unordered ζ_{ik}.

Proof: Denote by $p(\zeta_{i1})$ the index of ζ_{i1} after ordering, indicating that $p(\zeta_{i1})-1$ elements of the first column of Z are smaller than ζ_{i1} while $N-p(\zeta_{i1})$ elements are larger. Thus $\zeta_{ik}^{(o)} = \sum_{j=1}^{N}\zeta_{jk}\, I\left(p(\zeta_{j1})=i\right)$. Taking expectation, we obtain

$$E\left(\zeta_{ik}^{(o)}|\zeta_{i1}^{(o)}=\zeta\right) = \sum_{j=1}^{N}E\left(\zeta_{jk}|p(\zeta_{j1})=i;\zeta_{j1}=\zeta\right)\text{Prob}\left(p(\zeta_{j1})=i|\zeta_{i1}^{(o)}=\zeta\right)$$
$$= \sum_{j=1}^{N}E\left(\zeta_{jk}|\zeta_{j1}=\zeta\right)\text{Prob}\left(p(\zeta_{j1})=i|\zeta_{i1}^{(o)}=\zeta\right).$$

The last step follows since given $\zeta_{j1}=\zeta$, $p(\zeta_{j1})=i$ entails conditions on ζ_{m1} for $m\neq j$ only and hence is independent of ζ_{jk}. The resulting sum vanishes identically because $E\left(\zeta_{jk}\middle|\zeta_{j1}=\zeta\right)=0$ according to Assumption 2(ii). Thus, (i) is established. Parts (ii) and (iii) are derived following the same reasoning, utilizing the factorization property

$$E\left(\zeta_{jk}\zeta_{mk'}\middle|\zeta_{j1}=\zeta;\zeta_{m1}=\eta\right) = E\left(\zeta_{jk}\middle|\zeta_{j1}=\zeta\right)\; E\left(\zeta_{mk'}\middle|\zeta_{m1}=\eta\right).$$

Theorem 4: Under Assumptions 1-3:

(i) Ω_N converges in quadratic mean to a nonsingular limit Ω;

(ii) $\hat{\beta} = \beta_0 + \Omega^{-1}\psi(\beta_0)/\sqrt{N}$ is a $\sqrt{N}$-consistent solution to the FPE.

Proof: We recall that (i) is equivalent to the proposition that $Z'^{(o)}\Gamma(I-A)Z^{(o)}/N$ has a nonsingular limit. Ordering plays no role in the evaluation of $Z'^{(o)}\Gamma Z^{(o)}/N$, so the definition of Z implies $Z'^{(o)}\Gamma Z^{(o)}/N \rightarrow_{qm} I$, and only $Z'^{(o)}\Gamma A Z^{(o)}/N$ requires further consideration. We begin by showing that $\left(Z'^{(o)}\Gamma A Z^{(o)}/N\right)_{mk} \rightarrow_{qm} 0$ unless m=k=1. For j≥2 let $R_{jk}^{(o)} = \sum_{i=1}^{j-1} \zeta_{ik}^{(o)}/(j-1)$ and $\left(Z'^{(o)}\Gamma A Z^{(o)}/N\right)_{mk} = \sum_{j=2}^{N} \zeta_{jm}^{(o)} \gamma(C_o \zeta_{j1}^{(o)}) R_{jk}^{(o)}/N$. Lemma 3 implies that

$$E\left(\zeta_{jm}^{(o)} \gamma(C_o \zeta_{j1}^{(o)}) R_{jk}^{(o)}\right) = \sum_{i=1}^{j-1} \int E\left(\zeta_{jm}^{(o)} \zeta_{ik}^{(o)} \middle| \zeta_{j1}^{(o)} = \zeta\right) \gamma(C_o \zeta) f_j(\zeta) d\zeta/(j-1) = 0$$

(where f_j is the density function of $\zeta_{j1}^{(o)}$), and

$\mathrm{Var}\left(\sum_{j=2}^{N} \zeta_{jm}^{(o)} \gamma(C_o \zeta_{j1}^{(o)}) R_{jk}^{(o)}/N\right) \rightarrow 0$ if k≠1 or m≠1. Thus, only $\left(Z'^{(o)}\Gamma A Z^{(o)}/N\right)_{11}$ survives. To evaluate this element we write it in terms of the unordered regressors as $\sum_{j=1}^{N} \zeta_{j1} \gamma(C_o \zeta_{j1}) R_{j1}/N$, where $R_{j1} = \begin{cases} \sum_i \zeta_{i1} I(\zeta_{i1} < \zeta_{j1})/N_j & \text{if } N_j > 0 \\ 0 & \text{if } N_j = 0 \end{cases}$ and $N_j = N(\beta_o, \zeta_{j1}) = \sum_i I(\zeta_{i1} < \zeta_{j1})$. Here, we evaluate the expectation by conditioning on the unordered variable ζ_{j1}:

$E\left(\zeta_{j1} \gamma(C_o \zeta_{j1}) R_{j1}\right) = \int \zeta \gamma(C_o \zeta) E(R_{j1} | \zeta_{j1} = \zeta) f_z(\zeta) d\zeta$. It is convenient to treat differently the cases where $F_z(\zeta)$ is large or small, i.e., for some ε>0, let $\zeta_\epsilon = F_z^{-1}(\epsilon)$ and consider first $\zeta > \zeta_\epsilon$. As in the derivation of Lemma 1, we write

$$R_{j1} = \frac{\sum_i \zeta_{i1} I(\zeta_{i1} < \zeta_{j1})/N}{F_z(\zeta_{j1})} \left(1 - \frac{N_j/N - F_z(\zeta_{j1})}{N_j/N}\right)$$

and verify that $E\left(\frac{N_j/N - F_z(\zeta_{j1})}{N_j/N} \middle| \zeta_{j1} = \zeta\right) = O\left(1/(NF_z(\zeta))\right)$ and

$\mathrm{Var}\left(\frac{N_j/N - F_z(\zeta_{j1})}{N_j/N} \middle| \zeta_{j1} = \zeta\right) = O\left(1/(NF_z(\zeta))\right)$. Thus,

$E(R_{j1} | \zeta_{j1} = \zeta) = E_z(\zeta) + O\left(1/(NF_z(\zeta))\right)$. For $\zeta < \zeta_\epsilon$ we use the fact that R_{j1} is bounded to obtain $\int_{\zeta_{min}}^{\zeta_\epsilon} \zeta \gamma(C_o \zeta) E(R_{j1} | \zeta_{j1} = \zeta) f_z(\zeta) d\zeta = O(\epsilon)$. Finally, by choosing ε such that $\epsilon \rightarrow 0$ and $N\epsilon \rightarrow \infty$ we get

$E\left(\zeta_{j1} \gamma(C_o \zeta_{j1}) R_{j1}\right) \rightarrow \int \zeta \gamma(C_o \zeta) E_z(\zeta) f_z(\zeta) d\zeta$. A similar derivation gives

$Var\left(R_{j1}-E_z(\zeta_{j1})\right) \to 0$. It follows that

$\left(Z'^{(o)}\Gamma A Z^{(o)}/N\right)_{11} - \sum_{j=1}^{N}\zeta_{j1}\gamma(C\circ\zeta_{j1})E_z(\zeta_{j1})/N \to_{qm} 0$. The sum on the lhs' consists of independent terms, each with the mean $\int\zeta\gamma(C\circ\zeta)E_z(\zeta)f_z(\zeta)d\zeta$. Thus

$$\left(Z'^{(o)}\Gamma A Z^{(o)}/N\right)_{11} \to_{qm} \int\zeta\gamma(C\circ\zeta)E_z(\zeta)f_z(\zeta)d\zeta .$$

In fact, the same reasoning can be used to show that

$$\left(Z'^{(o)}\Gamma Z^{(o)}/N\right)_{11} \to_{qm} \int\zeta^2\gamma(C\circ\zeta)f_z(\zeta)d\zeta = 1.$$

Summarizing, all the elements of $Z'^{(o)}\Gamma(I-A)Z^{(o)}/N$ converge to the corresponding elements of I except for the 1,1 element, whose limit equals $\theta = \int \zeta\ \gamma(C\circ\zeta)\left(\zeta-E_z(\zeta)\right)f_z(\zeta)d\zeta$, which establishes (i). Let Ω denote the probability limit of Ω_N and define $\hat{\beta} = \beta_o + \Omega^{-1}\psi(\beta_o)/\sqrt{N}$. According to theorem 2, $\psi(\beta_o) = O_{qm}(1)$, and the nonsingularity of Ω implies that $\sqrt{N}(\hat{\beta}-\beta_o)$ is also $O_{qm}(1)$. Thus, we can use Theorem 3 to obtain $\psi(\hat{\beta}) = \psi(\beta_o)-\Omega_N\sqrt{N}(\hat{\beta}-\beta_o) + o_{qm}(1) = \psi(\beta_o) - \Omega\sqrt{N}(\hat{\beta}-\beta_o) + o_{qm}(1) = o_{qm}(1)$, implying that $\hat{\beta}$ is a $\sqrt{N}$-consistent solution to the FPE.

Theorem 5: Let $\{\beta_m\}_{m=1...M}$ be the set of distinct (to $O(1/\sqrt{N})$) solutions of the FPE. Then, under assumptions 1-4, $\hat{\beta} = \operatorname{argmin}_{m=1..M} Q_N^e(\beta_m)$ is a $\sqrt{N}$-consistent estimator of β_o.

Proof: We first show that for every $\beta \in S_\beta$, $Q_N^e(\beta)-Q_N(\beta) \to_{qm} 0$.

$$\begin{aligned} Q_N^e(\beta) - Q_N(\beta) = & \sum_{i=1}^{N}\left((y_i+z_i)^2 I(y_i>0) - (E_{2c}(z_i^o)-2E_{1c}(z_i^o)\Delta_i+\Delta_i^2)F_c(z_i^o)\right) / N \\ & + 2\sum_{i=1}^{N} (y_i+z_i)I(y_i>0)\left(\overline{E}_c(\beta,z_i) - E_{ec}(\beta,z_i)\right) / N \\ & + 2\sum_{i=1}^{N} \overline{E}_c(\beta,z_i)\left((E_{1c}(z_i^o)-\Delta_i)F_c(z_i^o) - (y_i+z_i)I(y_i>0)\right) / N \\ & + \sum_{i=1}^{N} I(y_i>0)\left(E_{ec}^2(\beta,z_i) - \overline{E}_c^2(\beta,z_i)\right) / N \end{aligned}$$

$$+ \sum_{i=1}^{N} \overline{E_C}^2(\beta,z_i) \left[I(y_i>0) - F_C(z_i^o)\right] / N.$$

For given X, we evaluate the expectation and the variance of each term. Since $y_i+z_i = \epsilon_i-\Delta_i$; $E\left[I(y_i>0)\right] = F_C(z_i^o)$; $E\left[\epsilon_i I(y_i>0)\right] = E_{1C}(z_i^o)F_C(z_i^o)$; and $E\left[\epsilon_i^2 I(y_i>0)\right] = E_{2C}(z_i^o)F_C(z_i^o)$, the expectations of the first, third and fifth terms vanish while the corresponding variances are O(1/N). Moreover, $(y_i+z_i)I(y_i>0)$ and $E_{eC}(\beta,z_i)$ are independent and $E\left[E_{eC}(\beta,z_i) - \overline{E_C}(\beta,z_i)\right] = O(1/N(\beta,z_i))$, $Var\left[E_{eC}(\beta,z_i)\right] = O(1/N(\beta,z_i))$ and $Var\left[E_{eC}^2(\beta,z_i)\right] = O(1/N(\beta,z_i))$ (cf. the derivation of Lemma 1). It follows that the expectations of the remaining terms, involving $E_{eC}(\beta,z_i)$, are $O(\log(N)/N)$ and the variances are $O(\log^2(N)/N^2)$.

For the rest of the derivation we consider moments with respect to the distribution of X. First, we fix z_i for some i and verify that, $E\left[\overline{E_C}(\beta,z_i)|z_i\right] = E_C(\beta,z_i) + O(1/(NF_z(z_i)))$ and $Var\left[\overline{E_C}(\beta,z_i)|z_i\right] = O(1/(NF_z(z_i)))$ where $NF_z(z_i) = E\left[N(\beta,z_i)|z_i\right]$. Next, we replace $\overline{E_C}(\beta,z_i)$ with $E_C(\beta,z_i)$ and define the following sum of independent quantities:

$$Q_N^*(\beta) = \sum_{i=1}^{N} F_C(z_i^o)\left\{Var(\epsilon|\epsilon>z_i^o) + \left[E_C(\beta,z_i)+\Delta_i-E_{1C}(z_i^o)\right]^2\right\} / N.$$

Now, when $F_z(z_i)$ is small, the corresponding variance of $\overline{E_C}(\beta,z_i)$ is large. Nevertheless, we can follow the reasoning of the proof of Theorem 4, separate the cases where F_z is large and small and integrate over the distribution of z_i to obtain $E\left[Q_N(\beta) - Q_N^*(\beta)\right]^2 \to 0$ uniformly on S_β. $Q_N(\beta)$ can, therefore, be approximated by $Q_N^*(\beta)$. Furthermore, $E\left[Q_N^*(\beta) - Q^*(\beta)\right]^2 \to 0$ uniformly on S_β, where

$$Q^*(\beta) = E\left\{F_C(z_i^o)\left\{Var(\epsilon|\epsilon>z_i^o) + \left[E_C(\beta,z_i)+\Delta_i-E_{1C}(z_i^o)\right]^2\right\}\right\}$$

is continuous in β.

It has already been noted that β_o minimizes $Q_N(\beta)$ for every sample X.

Thus, it must also minimize $Q^*(\beta)$. In fact, the identification condition 4(iii) ensures that β_o is the unique minimizer. Calculated at the consistent root of the FPE, $Q_N^e(\beta)$ converges (in quadratic mean) to the global minimum $Q^*(\beta_o)$, whereas, by virtue of the identification condition and the continuity of $Q^*(\beta)$, at any other root the corresponding value of Q_N^e is kept well above this minimum. It follows that the choice of the root that minimizes $Q_N^e(\beta)$ provides a consistent estimator.

Theorem 6: Under Assumptions 1-3, $\sqrt{N}(\hat{\beta} - \beta_o) \xrightarrow{D} N(0, \Omega^{-1} V \Omega'^{-1})$.

Proof: Given in the text.

References

Breiman, L., Y. Tsur and A. Zemel (1989): "A Simple Estimator for Censored Regression Models with Known Error Distribution", Technical Report No. 198, Dept. of Statistics, University of California, Berkeley.

Buckley, J. and I. James (1979): "Linear Regression with Censored Data", *Biometrica*, 66, 429-436.

Chung, K. L. (1974): "*A Course in Probability Theory*, 2nd ed. Academic Press, New York.

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. (1986): "A Robust Censored Regression Estimator", *Journal of Econometrics*, 32, 5-34.

Fernandez, L. (1986): "Non-parametric Maximum Likelihood Estimation of Censored Regression Models", *Journal of Econometrics*, 32, 35-57.

Horowitz, J. L. (1986): "A Distribution-Free Least Squares Estimator for Censored Linear Regression Models", *Journal of Econometrics*, 32, 59-84.

James, I. R. and P. J. Smith (1984): "Consistency Results for Linear Regression with Censored Data", *Ann. Statist.*, 12, 590-600.

Lee, L.-F. (1988): "A Semiparametric Nonlinear Least Square Estimator of Truncated Regression Models", Dept. of Economics, University of Minnesota.

Lehmann, E. L. (1975): "*Nonparametrics: Statistical Methods Based on Ranks*", Holden-Day, San Francisco.

Nawata, K. (1990): "Robust Estimation Based on Grouped-Adjusted Data in Censored Regression Models," *Journal of Econometrics*, 43, 337-362.

Powell, J. L. (1986): "Symmetrically Trimmed Least Squares Estimation for Tobit Models", *Econometrica*, 54, 1435-1460.

Powell, J. L. (1984): "Least Absolute Deviations Estimation for Censored Regression Model", *Journal of Econometrics*, 25, 303-325.

Ritov, Y. (1990): "Estimation in a Linear Regression Model with Censored Data", *Ann. Statist.*, 18, 303-328.

Ritov, Y. and M. Fygenson (1990): "A Monotone Estimating Equation for Censored Regression," Dept. of Statistics, The Hebrew University of Jerusalem.

Tsiatis, A. A. (1990): "Estimating Regression Parameters Using Linear Rank Tests for Censored Data", *Ann. Statist.*, 18, 354-372.

Tsur, Y. (1983): "On Efficient Estimation of Some Limited Dependent Variable Models", Dept. of Agricultural & Resource Economics, Giannini Foundation Working Paper 281, University of California, Berkeley.