*The Annals of Statistics*
1993, Vol. 21, No. 4, 1711–1720

# ON A SIMPLE ESTIMATION PROCEDURE FOR CENSORED REGRESSION MODELS WITH KNOWN ERROR DISTRIBUTIONS

By Leo Breiman,[1] Yacov Tsur and Amos Zemel

*University of California, University of Minnesota and Ben-Gurion University*

A simple and tractable iterative least squares estimation procedure for censored regression models with known error distributions is analyzed. It is found to be equivalent to a well-defined Huber type $M$-estimate. Under a regularity condition, the algorithm converges geometrically to a unique solution. The resulting estimate is $\sqrt{N}$-consistent and asymptotically normal.

**1. Introduction.** An important virtue of least squares (LS) estimates is their desirable large sample properties for a wide variety of error distribution functions. For regression models with incomplete data on the dependent variable (censored regression models), LS methods cannot be directly applied without first correcting for the potential bias inherent in the missing data. Bias correction procedures depend on the pattern of the missing data and on the error distribution. The pattern of missing data is usually assumed known, and the various estimation procedures differ mainly with respect to the level of knowledge assumed about the error distribution. In the simplest case, the complete knowledge of the error distribution is assumed. While this is not a very general situation, it is sometimes encountered: In the physical sciences, for example, the errors are often mainly due to inaccuracies of some measuring device and can be studied in the course of a calibration procedure.

In principle, an "optimal" solution to the problem is always available: The maximum likelihood estimator (MLE) is well known to be consistent and efficient. However, depending on the pattern of missing data and on the form of the error distribution, the likelihood function may turn out to be quite complicated and the associated normal equations may possess multiple roots. When this is indeed the case, the task of maximizing the likelihood function may become tedious, requiring an exhaustive search over all possible local maxima until the global maximum is found. In such cases it may be preferable to consider an alternative simple estimation which requires a few fast computations and converges rapidly to a unique solution. In the following, a simple and tractable iterative algorithm to estimate the coefficients of a censored regression model is analyzed.

The algorithm is based on an old and simple idea [Orchard and Woodbury, (1972)]: First, one fills in the missing data using predictors based on the

Received November 1990; revised October 1992

[1]Supported in part by NSF Grant DMS-8718362.

*AMS* 1991 *subject classifications*. 62F10, 62F12.
*Key words and phrases*. Censored data, iterative least squares, geometrical convergence.

observations and current parameter estimates; using these data, improved parameter estimators are obtained by applying LS methods as if no data are missing. These two steps are iterated until the procedure converges. Interestingly, this algorithm yields the MLE when the errors are normally distributed. This is so because, with normal errors, the present algorithm coincides with the EM algorithm of Dempster, Laird and Rubin (1977), which is known to yield the MLE [see Wu (1983) and Louis (1982) for more on the properties of the EM algorithm and Tsur (1983) for an account of the EM in the context of censored regression models]. With nonnormal errors, the present estimator differs from the MLE (and from EM) and the choice between the two entails a trade-off between the desirable (finite sample) computation properties of the proposed estimator versus the desirable (large sample) efficiency of the MLE. However, since we show $\sqrt{N}$-consistency, a single Newton–Raphson step achieves efficiency.

Any limit point of the algorithm satisfies a fixed point equation. From this, we show that the estimate is also the minimizer of a convex function which is a sum of convex error terms. This is used to uniquely define the estimate. We then show, under a regularity condition on the error distribution, that the sequence of iterates converges geometrically to the limit. Finally, using the fixed point equation we give simple proofs of $\sqrt{N}$-consistency and asymptotic normality.

Similar procedures were proposed by Schmee and Hahn (1979) and by Buckley and James (1979): The former employed simulation methods to investigate the case of normal errors with an unknown scale; the latter considered the case of an unknown error distribution. Several estimators of this type were described by Chatterjee and McLeish (1986), who compared their numerical performance using empirical heart transplant data. Extensions and further properties of the Buckley and James procedure were investigated by James and Smith (1984) and by Ritov (1990). In assuming a known distribution, we are tackling a simpler problem. The payoff is that the framework is more transparent, the derivations simple and illuminating and the conclusions much stronger. In particular, the connection to the minimization of a sum of convex error functions leading to an $M$-estimate gives an interesting alternative way of looking at the estimate.

The usefulness of the analysis described here extends, however, beyond the simple case of a known error distribution. In a recent study, Tsur and Zemel (1990) extended the estimation procedure proposed here to the case of unknown error distributions by employing new (distribution-free) estimates of the conditional expectations of the errors. Their analysis has been motivated to a large extent by the present study [in particular, see the discussion following (5.6)]. Obviously, the results derived by Tsur and Zemel (1990) are weaker and rely on restrictive assumptions, but the study here presented proves an important intermediate step from the uncensored regression model to the general censored case.

The paper is organized as follows: Section 2 describes the iterative method for computing the estimate. Section 3 gives the equivalence to the minimization of a sum of convex error terms. The convergence of the algorithm is studied in Section 4. Section 5 establishes consistency and asymptotic normality.

**2. The Iterative procedure.** We observe data $(y_i, x_i)$ where $y_i$ are scalars and the $x_i$ are $K$-dimensional vectors. The value $\{y_i = 0\}$ is taken to indicate that the value of $y_i$ is missing, otherwise $y_i > 0$. Define $M^+ = \{i; y_i > 0\}$ and $M^- = \{i; y_i = 0\}$. The data are generated by the mechanism

$$y_i = \max\{0, x_i'\beta_0 + u_i\}, \qquad i = 1, \dots, N,$$

where the $\{u_i\}$ are i.i.d. zero-mean error terms with known distribution and $\beta_0$ is an unknown $K$-dimensional parameter. Consider the following iterative procedure for estimating $\beta_0$. 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 available information including the estimate for $\beta_0$ obtained in the previous iteration. These filled in values for $y_i$ are then used to find an improved estimator for $\beta_0$.

E-STEP. Given the value $\beta^{(r)}$ of the $r$th iteration, the next values of the dependent variable are calculated as:

$$y_i(\beta^{(r)}) = \begin{cases} y_i, & i \in M^+, \\ x_i'\beta^{(r)} + E(u|u < -x_i'\beta^{(r)}), & i \in M^-. \end{cases} \tag{2.1}$$

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

$$\beta^{(r+1)} = (X'X)^{-1}X'Y(\beta^{(r)}), \tag{2.2}$$

where $Y(\beta^{(r)})$ is the $N$-dimensional vector whose elements are $y_i(\beta^{(r)})$ of (2.1). Equation (2.2) is the usual LS formula for uncensored observations.

The implementation of the algorithm proceeds as follows:

1. Set $\beta^{(0)}$, an initial value for the parameter vector.
2. $E$-step: Fill in the missing $y_i$ values as given by (2.1) using $E(u|u < z)$ with $z = -x'\beta$ calculated at the current $\beta$-estimate.
3. P-step: Calculate a new $\beta$-estimate according to (2.2).
4. Return to 2 unless the norm of the difference vector $\|\beta^{(r+1)} - \beta^{(r)}\|$ decreases below some predetermined convergence requirement.
5. Once the convergence criterion is satisfied, adopt the last value of $\beta$ as the final estimate.

Assuming this process converges, the limit $\hat{\beta}_N$ is the estimator and satisfies the fixed point equation (FPE)

$$\hat{\beta}_N = (X'X)^{-1}X'Y(\hat{\beta}_N). \tag{2.3}$$

The difficulty is that this is not, at this point, a satisfactory definition. The estimate $\hat{\beta}_N$ is defined only if the iterative process converges to the same limit for all starting values. We fix this up in the following section.

**3. A unique definition.** An obvious question is whether the FPE (2.3) has a unique solution. Assume $E(u|u < s)$ is continuously differentiable in $s$ and let

$$G(z) = \int_0^z E(u|u < s)\, ds.$$

As $G''(z) \geq 0$, $G(-x_i'\beta)$ is convex in $\beta$. Define

$$Q(\beta) = \tfrac{1}{2} \sum_{i \in M^+} (y_i - x_i'\beta)^2 + \sum_{i \in M^-} G(-x_i'\beta).$$

Then $Q(\beta)$ is smoothly convex in $\beta$ and has a unique minimum $\beta^*$. [It is easily verified that $Q(\beta) \to \infty$ as $\|\beta\| \to \infty$.] The role of the $G(-x_i'\beta)$ terms becomes clearer when one notes that $G(z)$ is nonincreasing, hence these terms tend to adjust $\beta$ so as to achieve low $x_i'\beta$ values for the missing data. This is in accord with our convention that the data are left-censored at zero.

Let $X'X^+ = \sum_{i \in M^+} x_i x_i'$ and $X'X^- = \sum_{i \in M^-} x_i x_i'$ be the matrices formed by partitioning $X'X$ according to $i \in M^+$ and $i \in M^-$, respectively, and assume that $X'X^+$ is positive definite (pd). We can now show:

PROPOSITION 3.1. *$\beta^*$ is the unique solution of the fixed point equation* (2.3).

The proof is a simple verification that setting the partial derivatives of $Q(\beta)$ equal to zero gives (2.3). (The pd nature of $X'X^+$ ensures that $M^+$ is not empty and that $X'X$ is pd.) A consequence of this proposition is that if the EP algorithm converges, its limit does not depend on the starting value. Other algorithms, such as Newton–Raphson, could be used to directly minimize $Q(\beta)$, but they do not have the attractive simplicity of the EP method.

We can also write $Q(\beta)$ in the form

$$Q(\beta) = \sum_i \omega(y_i, x_i'\beta),$$

$$\omega(y_i, x_i'\beta) = \tfrac{1}{2}(y_i - x_i'\beta)^2 I(y_i > 0) + G(-x_i'\beta) I(y_i = 0),$$

where $I(\cdot)$ denotes the indicator function. The estimate we want minimizes $Q(\beta)$ and therefore is an $M$-estimate of the type proposed and studied by Huber (1981). Thus, we can define $\hat{\beta}_N$ either as the unique minimizer of $Q(\beta)$ or the unique solution of (2.3).

**4. Convergence of the EP algorithm.** The convergence analysis becomes particularly simple when the error distribution $F(\cdot)$ satisfies the condition

(R) $$1\Big/\int_{-\infty}^z F(s)\, ds \quad \text{is convex in } z.$$

We call distributions satisfying (R) *regular cdf* and note that a number of well known cdf, including the normal, fall into this category. For a regular cdf the stronger condition $0 \le dE(u|u < z)/dz \le 2$ replaces the inequality $0 \le dE(u|u < z)/dz$ which holds for any distribution. We begin by proving a convergence theorem for regular distributions.

THEOREM 4.1 (Geometrical convergence). *If $F$ is regular and $X'X^+$ is pd, then the EP algorithm converges geometrically to a unique fixed point $\hat{\beta}_N$.*

Let $B$ be the symmetric positive definite square root of $X'X$. Take $\{\beta^{(r)}\}$, $r = 1, 2, \dots$, to be the sequence of iterates produced by the algorithm from any starting point, then Theorem 4.1 is a direct consequence of the following.

PROPOSITION 4.2. *There is an $L, 0 \le L < 1$, such that for all $r$,*

$$\| B(\beta^{(r+1)} - \beta^{(r)}) \| \le L \| B(\beta^{(r)} - \beta^{(r-1)}) \|.$$

PROOF. Write (2.2) for $\beta^{(r)}$ and $\beta^{(r+1)}$; subtracting gives

$$\beta^{(r+1)} - \beta^{(r)} = (X'X)^{-1} X' [Y(\beta^{(r)}) - Y(\beta^{(r-1)})]. \tag{4.3}$$

Define

$$\gamma_i(\beta) = 1 - I(y_i = 0)[1 - dE(u|u < z)/dz],$$

where the derivative is evaluated at $z = -x_i'\beta$, and let $\Gamma(\beta)$ be the $N$ by $N$ diagonal matrix with elements $\gamma_i$. Using definition (2.1) we can write

$$Y(\beta^{(r)}) - Y(\beta^{(r-1)}) = (I - \Gamma^{(r)}) X (\beta^{(r)} - \beta^{(r-1)}), \tag{4.4}$$

where $\Gamma^{(r)}$ is the matrix $\Gamma$ evaluated at some intermediate point on the line joining $\beta^{(r)}$ and $\beta^{(r-1)}$. Substituting (4.4) into (4.3) gives

$$\beta^{(r+1)} - \beta^{(r)} = A^{(r)} (\beta^{(r)} - \beta^{(r-1)}) \tag{4.5}$$

with $A^{(r)} = (X'X)^{-1} X'(I - \Gamma^{(r)}) X = (X'X)^{-1} C^{(r)}$. Write (4.5) as

$$B(\beta^{(r+1)} - \beta^{(r)}) = B^{-1} C^{(r)} B^{-1} B(\beta^{(r)} - \beta^{(r-1)}).$$

Since $B^{-1} C^{(r)} B^{-1}$ is a symmetric matrix

$$\| B(\beta^{(r+1)} - \beta^{(r)}) \| \le \bar{\lambda}^{(r)} \| B(\beta^{(r)} - \beta^{(r-1)}) \|,$$

where $\bar{\lambda}^{(r)}$ is the maximum of the absolute values of the eigenvalues of $B^{-1} C^{(r)} B^{-1}$. Then

$$\bar{\lambda}^{(r)} = \max_w \left| \frac{w'X'(I - \Gamma^{(r)}) X w}{w'X'Xw} \right|.$$

Using the regularity condition $(R)$, $|1 - \gamma_i| \le 1$, $i \in M^-$ and $|1 - \gamma_i| = 0$, $i \in M^+$, so that

$$|w'X'(I - \Gamma^{(r)}) X w| \le w'X'X^- w.$$

Let $\lambda^+_{\min}$ be the minimum eigenvalue of $X'X^+$ and $\lambda^-_{\max}$ the maximum eigenvalue of $X'X^-$. Then,

$$\bar{\lambda}^{(r)} \le \max_w \left\{ \frac{w'X'X^-w}{w'X'X^+w + w'X'X^-w} \right\} \le \frac{\lambda^-_{\max}}{\lambda^+_{\min} + \lambda^-_{\max}} = L < 1.$$

Now, the proposition implies that there is limit $\beta^{(\infty)}$ with $\|\beta^{(\infty)} - \beta^{(r)}\| \le \alpha L^r$, and proves the theorem. □

Condition (R) is sufficient, but far from necessary, for convergence. If $M$ is the upper bound on $dE(u|u < z)/dz$ of an irregular cdf, then arguments similar to the above show that Theorem 4.1 holds for all samples with $\lambda^+_{\min}/\lambda^-_{\max} > M - 2$, and even this weaker condition is far from necessary.

**5. Asymptotic behavior.** We consider in this section the solutions $\hat{\beta}_N$ of the fixed point equation (2.3) and look at their large $N$ behavior. We make the following assumptions:

A1. $\underline{\lambda}_N \ge \delta > 0$, for all $N$ sufficiently large, where $\underline{\lambda}_N$ is the minimum eigenvalue of $X'X^+/N$.

A2. $\mathrm{Trace}(X'X/N) = O(1)$.

A3. $\max_{i \le N} \|x_i\|^2/N \to 0$.

A4. $E|u_i|^3 < \infty$.

A5. $E(u|u < z)$ has a uniformly continuous bounded derivative.

The $\{x_i\}$ can either be considered nonstochastic, satisfying the above assumptions, or stochastic independent of the $\{u_i\}$. In the latter case, all our results are conditional on fixed $\{x_i\}$.

The following notation will be used: $z_i^0 = -x_i'\beta_0$, $z_i = -x_i'\beta$, $I_i = I(u_i < z_i^0)$, $E(I_i) = F(z_i^0)$, $E_i = E(u|u < z_i^0)$ and

$$s_i = u_i - I_i(u_i - E_i), \qquad S = (s_1, s_2, \dots, s_N)'. \tag{5.1}$$

With the above notation it can be verified that

$$Y(\beta_0) = X\beta_0 + S. \tag{5.2}$$

As (5.2) suggests, the quantities $s_i$ play in the present model the role of the errors $u_i$ in the uncensored case. The variables $s_i$ have zero means and variances given by

$$\tau_i^2 \equiv \mathrm{Var}\{s_i\} = \sigma_0^2 - \mathrm{Var}\{u|u < z_i^0\}F(z_i^0) \qquad i = 1, 2, \dots, N, \tag{5.3}$$

where $\sigma_0^2$ is the variance of $u_i$.

As in Section 4, take

$$\gamma_i(\beta) = 1 - I_i[1 - dE(u|u < z)/dz], \tag{5.4}$$

where the derivative is evaluated at $z = z_i$, and let $\Gamma(\beta)$ be the $N$ by $N$ diagonal matrix with elements $\gamma_i(\beta)$. Consider the two difference vectors $\Delta\beta = \hat{\beta}_N - \beta_0$ and $\Delta Y = Y(\hat{\beta}_N) - Y(\beta_0)$. Using (2.1) we get

$$\Delta Y = (I - \tilde{\Gamma})X\Delta\beta, \tag{5.5}$$

where the matrix $\tilde{\Gamma}$ is $\Gamma(\tilde{\beta})$ with $\tilde{\beta}$ an intermediate value between $\beta_0$ and $\hat{\beta}_N$.

Being a solution to the FPE (2.3),

$$\hat{\beta}_N = (X'X)^{-1}X'Y(\hat{\beta}_N) = (X'X)^{-1}X'(Y(\beta_0) + \Delta Y).$$

Hence, using (5.2) and (5.5), we obtain

$$(X'\tilde{\Gamma}X/N)\sqrt{N}\Delta\beta = (X'S/\sqrt{N}). \tag{5.6}$$

Note the similarity between this equation and the standard result for uncensored regression, with $X'\tilde{\Gamma}X$ replacing $X'X$ and $S$ replacing the error vector $(u_1, u_2, \dots, u_N)'$. A corresponding expression has been derived in the unknown error distribution case [Tsur and Zemel (1990)], with $X'\tilde{\Gamma}X$ and $S$ replaced by closely related, albeit more complicated, quantities.

Partitioning according to the index sets $M^+$ and $M^-$ gives

$$X'\tilde{\Gamma}X = X'X^+ + X'\tilde{\Gamma}X^-.$$

Because $dE(u|u < z)/dz \geq 0$ for any $z$, $X'\tilde{\Gamma}X^-$ is positive semidefinite. This implies that the minimum eigenvalue of $X'\tilde{\Gamma}X/N$ is not smaller than $\underline{\lambda}_N$. Hence, from (5.6),

$$\|\Delta\beta\| \leq \underline{\lambda}_N^{-1}\|X'S\|/N.$$

Because of (5.3),

$$E\|X'S/N\|^2 \leq \frac{\sigma_0^2}{N}\operatorname{Trace}(X'X/N),$$

and this leads to

$$E\|\Delta\beta\|^2 \to 0; \qquad E(N\|\Delta\beta\|^2) = O(1).$$

This result, which relies only on A1 and A2, establishes consistency for the estimate and also shows that $\sqrt{N}\Delta\beta$ is an $L_2$-bounded sequence of random vectors.

It is now possible to derive the limiting distribution of the estimator.

THEOREM 5.1. *Let $V$ and $\Omega$ be the respective probability limits of $X'\Sigma X/N$ and $X'\Gamma^* X/N$, where $\Sigma$ and $\Gamma^*$ are the $N$ by $N$ diagonal matrices with elements $\tau_i^2$ and $\gamma_i^* = E\{\gamma_i(\beta_0)\}$, respectively. Then*

$$\sqrt{N}\Delta\beta \to_D N(0, \Omega^{-1}V\Omega^{-1}).$$

PROOF. Let $f_N$ be the characteristic function of $\sqrt{N}\,\Delta\beta$ and $\tilde{f}$ the characteristic function of a $N(0, \Omega^{-1}V\Omega^{-1})$ variate (it is verified below that $\Omega$ is nonsingular), then it is sufficient to show that $f_N - \tilde{f} \to 0$ uniformly in a neighborhood of zero. Let $\Gamma_0 = \Gamma(\beta_0)$. Then, $\|(X'(\tilde{\Gamma} - \Gamma_0)X/N)\sqrt{N}\,\Delta\beta\| \le \alpha_N\|\sqrt{N}\,\Delta\beta\|$ where $\alpha_N$ is the maximum of the absolute values of the eigenvalues of $X'(\tilde{\Gamma} - \Gamma_0)X/N$:

$$\alpha_N = \max_{\|w\|=1}\left|w'X'\left(\tilde{\Gamma} - \Gamma_0\right)Xw/N\right| \le \max_i\left|\gamma_i(\tilde{\beta}) - \gamma_i(\beta_0)\right|\mathrm{Trace}(X'X/N).$$

Denoting $\theta(z) = dE(u|u < z)/dz$,

$$\alpha_N \le c_N \max_i\left|\theta\left(-x_i'\hat{\beta}_N\right) - \theta(-x_i'\beta_0)\right|.$$

Let $h$ be the modulus of continuity of $\theta$, then

$$\begin{aligned}\alpha_N &\le c_N \max_i h\left(\left|x_i'\left(\hat{\beta}_N - \beta_0\right)\right|\right)\\ &\le c_N \max_i h\left(\|\sqrt{N}\,\Delta\beta\| \cdot \|x_i\|/\sqrt{N}\right)\\ &= c_N h\left(\|\sqrt{N}\,\Delta\beta\| \cdot \max_i\|x_i\|/\sqrt{N}\right).\end{aligned}$$

Now, $c_N = O(1)$ and $\varepsilon_N = \max(\|x_i\|/\sqrt{N}) \to 0$, and we assert that

$$h\left(\varepsilon_N\|\sqrt{N}\,\Delta\beta\|\right) \cdot \|\sqrt{N}\,\Delta\beta\| \to_p 0,$$

because for any random variable $X \ge 0$, $P(h(\varepsilon X)X \ge d) = P(g(\varepsilon X) \ge \varepsilon d)$ where $g(v) = vh(v)$, and

$$P(g(\varepsilon X) \ge \varepsilon d) = P\left(\varepsilon X \ge g^{-1}(\varepsilon d)\right) \le E(X^2)\left[\varepsilon/g^{-1}(\varepsilon d)\right]^2.$$

Finally, for $d$ fixed and $\varepsilon \to 0$, the ratio $\varepsilon/g^{-1}(\varepsilon d) \to 0$. Then we claim that

$$\left\|(X'(\Gamma_0 - \Gamma^*)X/N)\sqrt{N}\,\Delta\beta\right\| \to_p 0. \tag{5.7}$$

To see this, let $\Lambda_{mm'}$ be the $m, m'$ element of $X'(\Gamma_0 - \Gamma^*)X/N$. Then, for every vector $\pi \in R^K$,

$$\left\|(X'(\Gamma_0 - \Gamma^*)X/N)\pi\right\|^2 \le \left(\sum_{m,m'}\Lambda_{mm'}^2\right)\|\pi\|^2. \tag{5.8}$$

Now,

$$\Lambda_{mm'} = -\frac{1}{N}\sum_i (x_{im}x_{im'})q_i\left[I(y_i = 0) - P(y_i = 0)\right],$$

where $q_i = 1 - \theta(z_i)$. By A5, $|q_i| \le \bar{q}$, so

$$\sum_{m,m'} E\left(\Lambda_{mm'}^2\right) \le \frac{\bar{q}^2}{N^2}\sum_i\|x_i\|^4 \le \bar{q}^2 \max_i\left(\|x_i\|^2/N\right)\frac{1}{N}\sum_i\|x_i\|^2.$$

Hence, $E(\Sigma_{m,m'}\Lambda^2_{mm'}) \to 0$. Since $E\|\sqrt{N}\,\Delta\beta\|^2 = O(1)$, (5.8) implies (5.7). Furthermore, if $\underline{\lambda}^*_N$ is the minimum eigenvalue of $X'\Gamma^*X/N$ and $\pi$ is the corresponding eigenvector, we use (5.8) again to show that $\underline{\lambda}^*_N \geq \underline{\lambda}_N - (\Sigma_{m,m'}\Lambda^2_{mm'})^{1/2}$. Hence, for large enough $N$, $\underline{\lambda}^*_N > \delta/2$ and $\Omega$ is nonsingular.

Rewrite (5.6) as

$$(X'\Gamma^*X/N)\sqrt{N}\,\Delta\beta = \left(X'S/\sqrt{N}\right) + \xi_N \tag{5.9}$$

where $\|\xi_N\| \to_p 0$. Therefore

$$\sqrt{N}\,\Delta\beta = \left(X'\Gamma^*X/N\right)^{-1}\left(\left(X'S/\sqrt{N}\right) + \xi_N\right).$$

The second term is dominated by $2\delta^{-1}\|\xi_N\|$ and hence can be ignored. The first term is of the form $(1/\sqrt{N})\Sigma_{i=1}^N a_{ik}s_i$, with $E(s_i) = 0$, $E(s_i^2) \leq \sigma_0^2$ and, by A4, $E|s_i|^3 \leq c < \infty$. By standard characteristic function expansions, the theorem follows if

$$\frac{1}{N}\sum_{i=1}^N a_{ik}^2 = O(1) \quad \text{and} \quad \frac{1}{N^{3/2}}\sum_{i=1}^N |a_{ik}|^3 \to 0.$$

Now, the latter follows from the former if $\max_{i\leq N}(|a_{ik}|/\sqrt{N}) \to 0$. So we need (a) $(1/N)\Sigma_{i=1}^N\|a_i\|^2 = O(1)$ and (b) $\max_{i\leq N}(\|a_i\|^2/N) \to 0$. By (5.9), $a_i = (X'\Gamma^*X/N)^{-1}x_i$, hence $\|a_i\|^2 \leq 4\delta^{-2}\|x_i\|^2$. It therefore suffices if $(1/N)\Sigma_{i=1}^N\|x_i\|^2 = O(1)$ and $\max_{i\leq N}(\|x_i\|^2/N) \to 0$. The latter holds by A3, and $(1/N)\Sigma_{i=1}^N\|x_i\|^2 = \text{Trace}(X'X/N) = O(1)$ by A2. $\square$

Note that the unknowns in computing the asymptotic covariance matrix are $\Sigma$ and $\Gamma^*$; by arguments similar to the above, they can be estimated by the corresponding matrices with diagonal elements $\tau_i^2(\hat{\beta}_N)$ and $\gamma_i(\hat{\beta}_N)$.

## REFERENCES


BUCKLEY, J. and JAMES, I. (1979). Linear regression with censored data. *Biometrika* **66** 429–436.

CHATTERJEE, S. and MCLEISH, D. L. (1986). Fitting linear regression models to censored data by least squares and maximum likelihood methods. *Comm. Statist. Theory Methods* **15** 3227–3243.

DEMPSTER, A. P., LAIRD, N. M. and RUBIN, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *J. Roy. Statist. Soc. Ser. B* **39** 1–22.

HUBER, P. J. (1981). *Robust Statistics*. Wiley, New York.

JAMES, I. R. and SMITH, P. J. (1984). Consistency results for linear regression with censored data. *Ann. Statist.* **12** 590–600.

LOUIS, T. (1982). Finding the observed information matrix when using the EM algorithm. *J. Roy. Statist. Soc. Ser. B* **44** 226–233.

ORCHARD, T. and WOODBURY, M. A. (1972). A missing information principle: Theory and applications. *Proc. Sixth Berkeley Symp. Math. Statist. Probab.* 697–715. Univ. California Press, Berkeley.

RITOV, Y. (1990). Estimation in a linear regression model with censored data. *Ann. Statist.* **18** 303–328.

SCHMEE, J. and HAHN, G. J. (1979). A simple method for regression analysis with censored data. *Technometrics* **21** 417–432.

TSUR, Y. (1983). On efficient estimation of some limited dependent variables models. Dept. Agricultural and Resource Economics, Univ. California, Berkeley, Giannini Foundation Working Paper 281.

TSUR, Y. and ZEMEL, A. (1990). Iterative least squares estimation of censored regression models with unknown error distribution. Dept. Agricultural and Applied Economics, Univ. Minnesota, Staff paper 90–8.

WU, C. F. J. (1983). On the convergence of the EM algorithm. *Ann. Statist.* **11** 95–103.

LEO BREIMAN
STATISTICS DEPARTMENT
UNIVERSITY OF CALIFORNIA
BERKELEY, CALIFORNIA 94720

YACOV TSUR
DEPARTMENT OF AGRICULTURAL
AND APPLIED ECONOMICS
UNIVERSITY OF MINNESOTA
1994 BUFORD AVENUE
ST. PAUL, MINNESOTA 55108

AMOS ZEMEL
JACOB BLAUSTEIN INSTITUTE
FOR DESERT RESEARCH
BEN-GURION UNIVERSITY
SEDE BOQER CAMPUS, 84993
ISRAEL