ECONOMETRICA

VOLUME 41 November 1973 NUMBER 6

REGRESSION ANALYSIS WHEN THE DEPENDENT VARIABLE IS TRUNCATED NORMAL

BY TAKESHI AMEMIYA[1]

The paper considers the estimation of the parameters of the regression model where the dependent variable is normal but truncated to the left of zero. Tobin [8] first considered the problem and proposed an iterative solution of the maximum likelihood equations. The paper proves the strong consistency and the asymptotic normality of the maximum likelihood estimator, proves the inconsistency of Tobin's initial estimator, proposes a computationally simple estimator that is consistent and asymptotically normal, and proves the asymptotic efficiency of the second-round estimator in the method of Newton. An extension to the case where the observations are not available at the truncation point is briefly indicated in the Conclusions.

1. INTRODUCTION

WE CONSIDER THE regression model defined by

$$
\begin{aligned}
(1.1)\qquad y_t &= \beta_0' x_t + u_t \quad \text{if RHS} > 0,\\
&= 0 \qquad\qquad\quad \text{if RHS} \leqslant 0 \qquad (t = 1, 2, \ldots, T),
\end{aligned}
$$

where β_0 is a K-component column vector of unknown constants, x_t is a K-component vector of known constants, and $\{u_t\}$ is independent with the distribution $N(0, \sigma_0^2)$.[2] Our problem is how to estimate β_0 and σ_0^2 on the basis of observations $\{y_1, y_2, \ldots, y_T\}$.

The model was first studied in a pioneering paper [8] by Tobin, who defined the maximum likelihood estimator (more precisely, a root of the normal equations) and proposed an iterative procedure starting from a certain initial estimator.

[1] This research was supported by National Science Foundation Grant GS 2635 to Stanford University. The author wishes to thank his colleague Michael Boskin for arousing his interest in the subject. He also wishes to thank Frederick Nold, a graduate student, for checking some of the derivations as well as for improving the English.

[2] This model is not as restrictive as it may seem. If we assume a more general model

$$
\begin{aligned}
(1.1')\qquad y_t &= \beta' x_t + u_t \quad \text{if RHS} > \alpha_t,\\
&= 0 \qquad\qquad\quad \text{if RHS} \leqslant \alpha_t,
\end{aligned}
$$

where α_t is a known constant, the transformed variable $y_t^* = y_t - \alpha_t$ follows:

$$
\begin{aligned}
(1.1'')\qquad y_t^* &= (\beta', -1)\begin{pmatrix} x_t \\ \alpha_t \end{pmatrix} + u_t \quad \text{if RHS} > 0,\\
&= 0 \qquad\qquad\qquad\qquad\quad \text{if RHS} \leqslant 0.
\end{aligned}
$$

Thus, (1.1″) can be handled easily by a slight modification of the subsequent analysis in this paper. If, moreover, $\alpha_t = \alpha$ for all t, and x_t contains 1, α is absorbed into one element of β, and such a model is completely reduced to the model in the text.

The model is widely applicable in economics. The dependent variable of the regression often has a lower or upper limit; it takes on the limiting value for a number of observations and takes a wide range of values for the remaining observations. For example, the household expenditure on major durable goods for a given year may be zero until the household income exceeds a certain level. Similarly, the supply of labor may be zero until the wage rate exceeds a certain level. Despite its usefulness, the model has not been widely used in the past. It is our hope that the present paper will help revive the econometrician's interest in this important area.

Without detracting from the importance of Tobin's original contribution, we would like to point out a number of aspects of his paper that need to be improved and extended. First, he did not cite the literature on the estimation of the parameters of the truncated normal distribution (obviously unintentionally), which is the more natural precursor of his paper than the probit literature he cited. Thus, people have come to regard his contribution as occupying a singular position separated from the rest of the statistical literature, whereas it should belong naturally in the above-mentioned literature on the truncated normal distribution. Second, his paper is somewhat hard to read because of his unusual notation and the unnecessary complication caused by making the truncation point different for each observation. Third, Tobin does not prove the consistency and the asymptotic normality of the maximum likelihood estimator he proposes. Fourth, the initial estimator he proposes is not consistent. It is the purpose of the present paper to fill in these gaps.

There is a vast literature on the estimation of the parameters of the truncated normal distribution. The simplest model dealt with is the non-regression model obtained by putting $K = 1$, $x_t = 1$ in the model (1.1). Hald [**2**] seems to have sparked interest in the subject in recent years, although there are a few older papers cited in his paper. He defined the maximum likelihood estimator, obtained its asymptotic covariance matrix, and suggested a computational procedure in the simplest non-regression model. His was the standard work, and it was improved and extended by many other people. For example, A. C. Cohen [**1**] extended the model to cases involving (i) known or unknown truncation point, (ii) known or unknown number of truncated observations, and (iii) single or double truncation. Halperin [**3**] proved the consistency and asymptotic normality of the maximum likelihood estimator in the non-regression model where the truncation point is unknown, the number of truncated observations is a fixed proportion of the total number of observations, and the distribution satisfies regularity conditions but is not necessarily normal. For other references, the reader may consult a good summary in Kendall and Stuart [**7**, p. 522]. So far as we are aware, Tobin's paper [**8**] was the first to consider the extension to the regression model. Since then, no important theoretical work has been done on the model.

The present paper proves the consistency and asymptotic normality of the maximum likelihood estimator (more precisely, a root of the normal equations) in the model (1.1). There are two reasons why this is important. One is that the existing theorems about the consistency and asymptotic normality of the maximum likelihood estimator in general are not applicable to our model because of the peculiar form of our likelihood function (3.1). (Halperin's work [**3**] is not applicable either, because his model is different from ours in many ways.) The other reason is that, because ours is a regression model, it is important to spell out the conditions on the regressors that insure consistency and asymptotic normality.

We show that Tobin's initial estimator is not consistent and propose an alternative initial estimator which is consistent and asymptotically normal. The advantage of using a consistent initial estimator in the iteration is two-fold. First, since the normal equations generally have multiple roots, we can hope to attain the convergence of the iteration to the global maximum only if we start from a consistent estimator. Second, in the method of Newton which we use, the second-round estimator of the iteration is asymptotically the same as the maximum likelihood estimator if the initial estimator is consistent with the probabilistic order of the error being the inverse of the square root of the sample size. However, we do not mean to imply that our initial estimator is always preferred to Tobin's. For a small sample, Tobin's initial estimator, if appropriately defined, may approximate the maximum likelihood estimator well and turn out to be superior to our initial estimator. Thus, the two estimators could be complementary.

The order of presentation in the paper is as follows: Section 2 states the assumptions. Section 3 presents the likelihood function and its first and second derivatives. Section 4 gives the first four moments of $y_t - \beta_0' x_t$ conditional upon $y_t > 0$. These moments are utilized in Sections 6, 7, and 8. Section 5 states four lemmas which are used in proving the five theorems in Sections 6, 8, and 9. Two of these lemmas are borrowed from an excellent paper [**5**] by Jennrich on non-linear regression. Section 6 proves the strong consistency (Theorem 1) and the asymptotic normality (Theorem 2) of the maximum likelihood estimator. Section 7 discusses the method of Newton and Tobin's initial estimator, which is shown to be inconsistent. Section 8 proposes a consistent and asymptotically normal initial estimator (Theorems 3 and 4). Section 9 proves the asymptotic efficiency of the second-round estimator (Theorem 5). Finally, Section 10 states the conclusions.

A part of the paper repeats the results obtained by Tobin [**8**], for we need to develop the basic equations in our own notation to prove the four theorems. Therefore, the paper may be read without reference to Tobin's paper.

2. ASSUMPTIONS

ASSUMPTION 1: *Let $\theta = (\beta', \sigma^2)'$ and $\theta_0 = (\beta_0', \sigma_0^2)'$. The parameter space Θ is compact, does not contain the region $\sigma^2 \leqslant 0$, and contains an open neighborhood of θ_0.*

ASSUMPTION 2: *x_t is bounded, and the empirical distribution function G_n, defined by $G_n(x) = j/n$, where j is the number of points $x_1, x_2, \ldots, x_n$ less than or equal to x, converges to a distribution function (designated G).*

ASSUMPTION 3: *$\lim_{T\to\infty} (1/T) \Sigma_{t=1}^{T} x_t x_t'$ is positive definite.*

3. LIKELIHOOD FUNCTION AND ITS DERIVATIVES

Let ψ be the subset of integers $\{1, 2, \ldots, T\}$ such that $y_t = 0$ for $t \in \psi$. Let S be the number of elements in ψ. Let $\bar{\psi}$ be the complement of ψ in the set of integers $\{1, 2, \ldots, T\}$. Obviously, $\bar{\psi}$ has $T - S$ elements.

From (1.1) we will, like Tobin [**8**], define the likelihood function as

$$L = \prod_{t \in \psi} [1 - F(\beta' x_t, \sigma^2)] \cdot \prod_{t \in \bar{\psi}} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(1/2\sigma^2)(y_t - \beta' x_t)^2} \tag{3.1}$$

defined over Θ, where

$$F(\beta' x_t, \sigma^2) = \int_{-\infty}^{\beta' x_t} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}(\lambda/\sigma)^2} \, d\lambda .$$

Henceforth, we will write $F(\beta' x_t, \sigma^2)$ simply as F_t. Equation (3.1) is an unusual likelihood function because it is the product of discrete probabilities and density functions. This is the obvious generalization of the likelihood function defined by Hald [**2**], Cohen [**1**], and others in the non-regression model. The logarithm of (3.1) is

$$\log L = \sum_{\psi} \log(1 - F_t) - \frac{T - S}{2} \log \sigma^2 - \frac{1}{2\sigma^2} \sum_{\bar{\psi}} (y_t - \beta' x_t)^2 . \tag{3.2}$$

Let

$$f_t = f(\beta' x_t, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}(\beta' x_t/\sigma)^2} .$$

Then the first and second derivatives of $\log L$ are[3]

$$\frac{\partial \log L}{\partial \beta} = -\sum_{\psi} \frac{f_t}{1 - F_t} x_t + \frac{1}{\sigma^2} \sum_{\bar{\psi}} (y_t - \beta' x_t) x_t , \tag{3.3}$$

$$\frac{\partial \log L}{\partial \sigma^2} = \frac{1}{2\sigma^2} \sum_{\psi} \frac{\beta' x_t f_t}{1 - F_t} - \frac{T - S}{2\sigma^2} + \frac{1}{2\sigma^4} \sum_{\bar{\psi}} (y_t - \beta' x_t)^2 , \tag{3.4}$$

$$\frac{\partial^2 \log L}{\partial \beta \, \partial \beta'} = -\sum_{\psi} \frac{f_t}{(1 - F_t)^2} \left[f_t - \frac{1}{\sigma^2}(1 - F_t)\beta' x_t \right] x_t x_t' - \frac{1}{\sigma^2} \sum_{\bar{\psi}} x_t x_t' , \tag{3.5}$$

$$\frac{\partial^2 \log L}{\partial \sigma^2 \, \partial \beta} = -\frac{1}{2\sigma^2} \sum_{\psi} \frac{f_t}{(1 - F_t)^2} \left[\frac{1}{\sigma^2}(1 - F_t)(\beta' x_t)^2 - (1 - F_t) - \beta' x_t f_t \right] x_t - \frac{1}{\sigma^4} \sum_{\bar{\psi}} (y_t - \beta' x_t) x_t , \tag{3.6}$$

[3] In obtaining these derivatives we will frequently use the following results:

$$\frac{\partial F_t}{\partial \beta} = f_t x_t , \qquad \frac{\partial F_t}{\partial \sigma^2} = -\frac{1}{2\sigma^2} \beta' x_t f_t , \qquad \frac{\partial f_t}{\partial \beta} = -\frac{1}{\sigma^2} \beta' x_t f_t x_t ,$$

and

$$\frac{\partial f_t}{\partial \sigma^2} = \frac{(\beta' x_t)^2 - \sigma^2}{2\sigma^4} f_t .$$

and

$$\frac{\partial^2 \log L}{\partial(\sigma^2)^2} = \frac{1}{4\sigma^4}\sum_{\psi}\frac{f_t}{(1-F_t)^2}\left[\frac{1}{\sigma^2}(1-F_t)(\beta' x_t)^3 - 3(1-F_t)\beta' x_t - (\beta' x_t)^2 f_t\right] + \frac{T-S}{2\sigma^4} - \frac{1}{\sigma^6}\sum_{\bar{\psi}}(y_t - \beta' x_t)^2. \tag{3.7}$$

Sometimes it will be convenient to make the following substitutions in the above equations:

$$F_t = \Phi_t \stackrel{\text{def}}{=} \int_{-\infty}^{\beta' x_t/\sigma} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\lambda^2}\, d\lambda \tag{3.8}$$

and

$$\sigma f_t = \phi_t \stackrel{\text{def}}{=} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(\beta' x_t/\sigma)^2}. \tag{3.9}$$

Note that Φ_t and ϕ_t are the distribution and density function respectively of the standard normal variable evaluated at $\beta' x_t/\sigma$.

4. DEFINITION AND MOMENTS OF u_t^*

Let u_t^* be the random variable with the density $h(\cdot)$ given by

$$h(\lambda) = \frac{1}{F_{0t}}\frac{1}{\sqrt{2\pi\sigma_0^2}} e^{-\frac{1}{2}(\lambda/\sigma_0)^2}, \qquad -\beta_0' x_t < \lambda < \infty, \tag{4.1}$$
$$= 0 \qquad \text{elsewhere,}$$

where $F_{0t} = F(\beta_0' x_t, \sigma_0^2)$. Then we have

$$y_t = \beta_0' x_t + u_t^*, \qquad \text{for } t \in \bar{\psi}. \tag{4.2}$$

Thus, the conditional moments of y_t, given $t \in \bar{\psi}$, can be easily calculated from the moments of u_t^*.

Since F_{0t} is bounded away from 0 by Assumption 2, all the moments of u_t^* are bounded uniformly in t. The first four moments of u_t^* are

$$Eu_t^* = \sigma_0^2 \frac{f_{0t}}{F_{0t}}, \tag{4.3}$$

where $f_{0t} = f(\beta_0' x_t, \sigma_0^2)$,

$$Eu_t^{*2} = \sigma_0^2 - \sigma_0^2 \beta_0' x_t \frac{f_{0t}}{F_{0t}}, \tag{4.4}$$

$$Eu_t^{*3} = \sigma_0^2 \frac{f_{0t}}{F_{0t}}[(\beta_0' x_t)^2 + 2\sigma_0^2], \tag{4.5}$$

and

$$Eu_t^{*4} = \sigma_0^2\left[3\sigma_0^2 - 3\sigma_0^2\beta_0' x_t \frac{f_{0t}}{F_{0t}} - (\beta_0' x_t)^3 \frac{f_{0t}}{F_{0t}}\right]. \tag{4.6}$$

5. PRELIMINARY LEMMAS

LEMMA 1 (Theorem 1 of Jennrich [**5**, p. 635]): *Let X be Euclidean space and Θ be a compact subset of a Euclidean space. If h is a bounded and continuous function on $X \times \Theta$, and if $\{G_T\}$ is a sequence of distribution functions on X which converge to a distribution function G, then*

$$\int h(x, \theta)\, dG_T(x) \to \int h(x, \theta)\, dG(x)$$

uniformly for all θ in Θ.

LEMMA 2: *Let $\{\xi_t\}$ be independent with $E\xi_t = \mu_t$ and its first four moments be uniformly bounded in t. Let $\{h_t\}$ be a sequence of continuous functions on a compact set Θ such that*

$$\frac{1}{T}\sum_1^T h_t(\theta_1)h_t(\theta_2)$$

converges uniformly for θ_1 and θ_2 in Θ. Then

$$\frac{1}{T}\sum_1^T h_t(\theta)(\xi_t - \mu_t)$$

converges to 0 a.e. uniformly for all θ in Θ.

PROOF: This lemma differs from Theorem 4 of Jennrich [**5**, p. 636] only in that in Jennrich's theorem $\{\xi_t\}$ is assumed to be i.i.d. with finite mean and variance. Every step in Jennrich's proof still holds because both $\{\xi_t\}$ and $\{\xi_t^2\}$ satisfy the conditions of Kolmogorov's strong law of large numbers for non-identically distributed random variables.

LEMMA 3: *Let $Q_T(\omega, \theta)$ be a measurable function on a measurable space Ω and for each ω in Ω a continuous function for θ in a compact set Θ. Then there exists a measurable function $\hat{\theta}_T(\omega)$ such that*

$$Q_T[\omega, \hat{\theta}_T(\omega)] = \sup_{\theta \in \Theta} Q_T(\omega, \theta) \qquad \textit{for all } \omega \textit{ in } \Omega.$$

If $Q_T(\omega, \theta)$ converges to $Q(\theta)$ a.e. uniformly for all θ in Θ, and if $Q(\theta)$ has a unique maximum at $\theta_0 \in \Theta$, then $\hat{\theta}_T$ converges to θ_0 a.e.

PROOF: That there exists the measurable function $\hat{\theta}_T$ is proved in Lemma 2 of Jennrich [**5**, p. 637]. Let N be an open neighborhood around θ_0. Then $\bar{N}$, the complement of N in Θ, is compact. Therefore, $\max_{\theta \in \bar{N}} Q(\theta)$ exists.

Denote $\varepsilon = Q(\theta_0) - \max_{\theta \in \bar{N}} Q(\theta)$. Then $|Q_T(\omega, \theta) - Q(\theta)| < \varepsilon/2$ implies $\hat{\theta}_T \in N$. Therefore, $\hat{\theta}_T$ converges to θ_0 a.e. *Q.E.D.*

LEMMA 4: *Let $f_T(\omega, \theta)$ be a measurable function on a measurable space Ω and for each ω in Ω a continuous function for θ in a compact set Θ. If $f_T(\omega, \theta)$ converges to $f(\theta)$ a.e. uniformly for all θ in Θ, and if $\hat{\theta}_T(\omega)$ converges to θ_0 a.e., then $f_T[\omega, \hat{\theta}_T(\omega)]$ converges to $f(\theta_0)$ a.e.*

PROOF: Because of uniform convergence, we have

$$|f_T[\omega, \hat{\theta}_T(\omega)] - f[\hat{\theta}_T(\omega)]| < \frac{\varepsilon}{2} \qquad \text{a.e. for } T > T_1(\varepsilon).$$

Since f is continuous,

$$|f[\hat{\theta}_T(\omega)] - f(\theta_0)| < \frac{\varepsilon}{2} \qquad \text{a.e. for } T > T_2(\varepsilon).$$

Therefore,

$$|f_T[\omega, \hat{\theta}_T(\omega)] - f(\theta_0)| < \varepsilon \qquad \text{a.e. for } T > \max[T_1, T_2]. \qquad \textit{Q.E.D.}$$

We might note that both Lemmas 3 and 4 will still hold if we change the expression "a.e." to "with probability approaching one." One merely makes the same change in the proofs.

6. STRONG CONSISTENCY AND ASYMPTOTIC NORMALITY OF A ROOT OF THE NORMAL EQUATIONS

The normal equations are defined in vector notation as

$$\frac{\partial \log L}{\partial \theta} = 0 \tag{6.1}$$

where $\theta = (\beta', \sigma^2)'$. We say $\hat{\theta}$ is a root of the normal equations if it is a solution of (6.1) corresponding to a local maximum of the logarithmic likelihood function (3.2).[4] Thus,

$$\frac{\partial^2 \log L(\hat{\theta})}{\partial \theta \, \partial \theta'}$$

is a negative definite matrix. The root is not unique unless we restrict the parameter space, and it may or may not correspond to the global maximum. Many authors would call a root a maximum likelihood estimator even if it did not correspond to the global maximum. We shall use the word *root* to avoid ambiguity. A root may not even exist, for the global maximum may occur on the borderline of Θ.

[4] In customary usage, any solution of (6.1) may be called a root. But since we are not interested in a solution corresponding to a local minimum or a saddle point, we will use the word in this special sense.

However, the consistency theorem developed below implies that it will exist with probability one as T goes to infinity. So, in the large sample theory, we do not have to worry about a borderline global maximum. Even for small samples, a root is most likely to exist if the parameter space Θ is taken to be large.

Tobin [**8**] proposed $\hat{\theta}$ defined above as the estimator of the parameters of the model (1.1).[5] We will study its asymptotic properties below.

Strong Consistency

An estimator is said to be strongly consistent if it converges to the true value a.e. Define the random variables

$$\begin{aligned} v_t &= 1 \qquad \text{with probability } 1 - F_{0t}, \\ &= 0 \qquad \text{with probability } F_{0t}, \end{aligned} \tag{6.2}$$

and

$$w_t = 1 - v_t.$$

Then, from (3.2) and (4.2), $(1/T) \log L \ (\stackrel{\text{def}}{=} Q_T)$ can be written as

$$\begin{aligned} Q_T = {} & \frac{1}{T}\sum_1^T [\log(1 - F_t)]v_t - \frac{1}{2}\log\sigma^2 \frac{1}{T}\sum_1^T w_t - \frac{1}{2\sigma^2}\frac{1}{T}\sum_1^T [(\beta_0 - \beta)'x_t]^2 w_t \\ & - \frac{1}{\sigma^2}\frac{1}{T}\sum_1^T (\beta_0 - \beta)'x_t u_t^* w_t - \frac{1}{2\sigma^2}\frac{1}{T}\sum_1^T u_t^{*2} w_t \end{aligned} \tag{6.3}$$

defined over Θ. We will consider the limit a.e. of each of the five terms above.

By Assumptions 1 and 2 and Lemma 1,

$$\frac{1}{T}\sum_1^T \log[1 - F_t(\theta_1)] \cdot \log[1 - F_t(\theta_2)]$$

converges uniformly for all θ_1 and θ_2 in Θ. All the moments of v_t are uniformly bounded by Assumptions 1 and 2. Therefore, by Lemma 2,

$$\frac{1}{T}\sum_1^T [\log(1 - F_t)](v_t - 1 + F_{0t})$$

converges to 0 a.e. uniformly for all θ in Θ. But by Assumptions 1 and 2 and Lemma 1,

$$\lim_{T\to\infty} \frac{1}{T}\sum_1^T [\log(1 - F_t)](1 - F_{0t})$$

[5] Tobin's model is actually the model (1.1′) of Footnote 1. But as we explain in the same footnote, his model can be analyzed easily by a slight modification of our results.

exists. Therefore

$$(6.4)\qquad \frac{1}{T}\sum_1^T [\log(1-F_t)]v_t \to \lim_{T\to\infty}\frac{1}{T}\sum_1^T [\log(1-F_t)](1-F_{0t})$$

a.e. uniformly for all θ in Θ.

By Assumption 2 and Lemma 1, $\lim_{T\to\infty}(1/T)\,\Sigma_1^T F_{0t}$ exists. Hence, by Kolmogorov's strong law of large numbers,

$$(6.5)\qquad \frac{1}{T}\sum_1^T w_t \to \lim_{T\to\infty}\frac{1}{T}\sum_1^T F_{0t}\qquad \text{a.e.}$$

The convergence of the third and fourth terms can be proved in a way analogous to the proof of (6.4) because w_t and u_t^* are independent and the moments of both are uniformly bounded. Thus,

$$(6.6)\qquad \frac{1}{T}\sum_1^T [(\beta_0-\beta)'x_t]^2 w_t \to \lim_{T\to\infty}\frac{1}{T}\sum_1^T [(\beta_0-\beta)'x_t]^2 F_{0t}$$

a.e. uniformly for all θ in Θ, and

$$(6.7)\qquad \frac{1}{T}\sum_1^T (\beta_0-\beta)'x_t u_t^* w_t \to \sigma_0^2 \lim_{T\to\infty}\frac{1}{T}\sum_1^T (\beta_0-\beta)'x_t f_{0t}$$

a.e. uniformly for all θ in Θ.

Finally, using Kolmogorov's strong law of large numbers as in (6.5), we have

$$(6.8)\qquad \frac{1}{T}\sum_1^T u_t^{*2} w_t \to \sigma_0^2 \lim_{T\to\infty}\frac{1}{T}\sum_1^T (F_{0t}-\beta_0'x_t f_{0t})\qquad \text{a.e.}$$

Thus, combining (6.4) through (6.8), we have shown that Q_T converges a.e. uniformly for all θ in Θ to Q given by

$$(6.9)\qquad Q = \lim_{T\to\infty}\frac{1}{T}\sum_1^T\Big[\,[\log(1-F_t)](1-F_{0t}) - \frac{1}{2}(\log\sigma^2)F_{0t} - \frac{1}{2\sigma^2}\{[(\beta_0-\beta)'x_t]^2F_{0t} + 2\sigma_0^2(\beta_0-\beta)'x_t f_{0t} + \sigma_0^2(F_{0t}-\beta_0'x_t f_{0t})\}\Big].$$

We will evaluate the first and second derivatives of Q. In doing so, we note that we can interchange the derivative and the summation operation because the series of both the first and second derivatives of all the terms of Q converge uniformly by Assumptions 1 and 2 and Lemma 1. Therefore, we have

$$(6.10)\qquad \frac{\partial Q}{\partial\beta} = \lim_{T\to\infty}\frac{1}{T}\sum_1^T\left\{-\frac{1-F_{0t}}{1-F_t}f_t x_t + \frac{1}{\sigma^2}[(\beta_0-\beta)'x_t F_{0t}x_t + \sigma_0^2 f_{0t}x_t]\right\},$$

$$(6.11)\qquad \frac{\partial Q}{\partial \sigma^2} = \frac{1}{2\sigma^2} \lim_{T\to\infty} \frac{1}{T} \sum_1^T \Bigg[\frac{1-F_{0t}}{1-F_t} \beta' x_t f_t - F_{0t} + \frac{1}{\sigma^2} \{ [(\beta_0 - \beta)' x_t]^2 F_{0t} + 2\sigma_0^2 (\beta_0 - \beta)' x_t f_{0t} + \sigma_0^2 (F_{0t} - \beta_0' x_t f_{0t}) \} \Bigg],$$

$$(6.12)\qquad \frac{\partial^2 Q}{\partial \beta\, \partial \beta'} = \lim_{T\to\infty} \frac{1}{T} \sum_1^T \left[\frac{\beta' x_t}{\sigma^2} f_t \frac{1-F_{0t}}{1-F_t} - f_t^2 \frac{1-F_{0t}}{(1-F_t)^2} - \frac{F_{0t}}{\sigma^2} \right] x_t x_t',$$

$$(6.13)\qquad \frac{\partial^2 Q}{\partial \sigma^2 \partial \beta} = -\frac{1}{2\sigma^4} \lim_{T\to\infty} \frac{1}{T} \sum_1^T \Bigg\{ \frac{1-F_{0t}}{1-F_t} (\beta' x_t)^2 f_t x_t - \sigma^2 \frac{1-F_{0t}}{1-F_t} f_t x_t - \sigma^2 \frac{1-F_{0t}}{(1-F_t)^2} \beta' x_t f_t^2 x_t + 2[(\beta_0 - \beta)' x_t F_{0t} x_t + \sigma_0^2 f_{0t} x_t] \Bigg\},$$

and

$$(6.14)\qquad \frac{\partial^2 Q}{\partial (\sigma^2)^2} = \frac{1}{4\sigma^6} \lim_{T\to\infty} \frac{1}{T} \sum_1^T \Bigg[\frac{1-F_{0t}}{1-F_t} (\beta' x_t)^3 f_t - 3\sigma^2 \frac{1-F_{0t}}{1-F_t} \beta' x_t f_t - \sigma^2 \frac{1-F_{0t}}{(1-F_t)^2} (\beta' x_t f_t)^2 + 2\sigma^2 F_{0t} - 4\{ [(\beta_0 - \beta)' x_t]^2 F_{0t} + 2\sigma_0^2 (\beta_0 - \beta)' x_t f_{0t} + \sigma_0^2 (F_{0t} - \beta_0' x_t f_{0t}) \} \Bigg].$$

From (6.10) and (6.11) it is clear that

$$(6.15)\qquad \frac{\partial Q(\theta_0)}{\partial \theta} = 0.$$

From (6.12), (6.13), and (6.14), the second derivatives evaluated at θ_0 are

$$(6.16)\qquad \frac{\partial^2 Q(\theta_0)}{\partial \beta\, \partial \beta'} = \frac{1}{\sigma_0^2} \lim \frac{1}{T} \sum_1^T \left[\beta_0' x_t f_{0t} - \frac{\sigma_0^2 f_{0t}^2}{1 - F_{0t}} - F_{0t} \right] x_t x_t',$$

$$(6.17)\qquad \frac{\partial^2 Q(\theta_0)}{\partial \sigma^2\, \partial \beta} = -\frac{1}{2\sigma_0^3} \lim \frac{1}{T} \sum_1^T \left[\frac{(\beta_0' x_t)^2 f_{0t}}{\sigma_0} + \sigma_0 f_{0t} - \frac{\sigma_0 \beta_0' x_t f_{0t}^2}{1 - F_{0t}} \right] x_t,$$

and

$$(6.18)\qquad \frac{\partial^2 Q(\theta_0)}{\partial (\sigma^2)^2} = \frac{1}{4\sigma_0^4} \lim \frac{1}{1} \sum_1^T \left[\frac{(\beta_0' x_t)^3 f_{0t}}{\sigma_0^2} + \beta_0' x_t f_{0t} - \frac{(\beta_0' x_t f_{0t})^2}{1 - F_{0t}} - 2F_{0t} \right].$$

If we define $z_t = \beta_0' x_t / \sigma_0$ and denote the standard normal density and distribution function evaluated at z_t by ϕ_t and Φ_t, respectively, (6.16), (6.17), and (6.18) can be written as

$$(6.19)\qquad \frac{\partial^2 Q(\theta_0)}{\partial\theta\,\partial\theta'} = -\lim\frac{1}{T}\begin{bmatrix}\sum_1^T a_t x_t x_t' & \sum_1^T b_t x_t \\ \sum_1^T b_t x_t' & \sum_1^T c_t\end{bmatrix} \overset{\text{def}}{=} -A,$$

where

$$a_t = -\frac{1}{\sigma_0^2}\left(z_t\phi_t - \frac{\phi_t^2}{1-\Phi_t} - \Phi_t\right),$$

$$b_t = \frac{1}{2\sigma_0^3}\left(z_t^2\phi_t + \phi_t - \frac{z_t\phi_t^2}{1-\Phi_t}\right),$$

and

$$c_t = -\frac{1}{4\sigma_0^4}\left(z_t^3\phi_t + z_t\phi_t - \frac{z_t^2\phi_t^2}{1-\Phi_t} - 2\Phi_t\right).$$

We have, for any K-component vector p and a scalar q,

$$(6.20)\qquad (p', q)A\binom{p}{q} = \lim\frac{1}{T}\sum_1^T\,[a_t(p'x_t)^2 + c_t q^2 + 2b_t p'x_t q]$$

$$\geqslant \inf_t(\lambda_t)\cdot\lim\frac{1}{T}\sum_1^T[(p'x_t)^2 + q^2],$$

where λ_t is the smallest characteristic root of

$$\begin{bmatrix} a_t & b_t \\ b_t & c_t\end{bmatrix}.$$

By Assumption 3,

$$\lim\frac{1}{T}\sum_1^T[(p'x_t)^2 + q^2]$$

is positive unless all the elements of p and q are zero. We obtain

$$(6.21)\qquad \begin{vmatrix} a_t & b_t \\ b_t & c_t\end{vmatrix} = \frac{1}{4\sigma_0^6}\left[\phi_t\left(\frac{\phi_t}{1-\Phi_t} - z_t\right)(2\Phi_t + z_t\phi_t + z_t^2\Phi_t) + 2\Phi_t^2 - z_t\phi_t\Phi_t - \phi_t^2\right].$$

But if z_t is bounded, we have[6]

$$(6.22)\qquad \frac{\phi_t}{1-\Phi_t} > z_t,$$

[6] To prove (6.22), note that $\phi - z(1-\Phi)$ converges to 0 as $z \to \infty$ and its derivative $\Phi - 1$ is negative. To prove (6.23), note that $2\Phi + z\phi + z^2\Phi$ converges to 0 as $z \to -\infty$ and its derivative $3\phi + 2z\Phi$ is positive. To prove (6.24), note that $2\Phi^2 - z\phi\Phi - \phi^2$ converges to 0 as $z \to -\infty$ and its derivative $\phi(3\Phi + z\phi + z^2\Phi)$ is positive because of (6.23).

(6.23) $$2\Phi_t + z_t\phi_t + z_t^2\Phi_t > 0,$$

and

(6.24) $$2\Phi_t^2 - z_t\phi_t\Phi_t - \phi_t^2 > 0.$$

Therefore, since z_t is uniformly bounded by Assumption 2,

$$\begin{vmatrix} a_t & b_t \\ b_t & c_t \end{vmatrix}$$

is positive and uniformly bounded away from 0. Inequality (6.22) also implies that a_t is positive. Hence, we have $\inf_t (\lambda_t) > 0$. Therefore, we have proved

(6.25) $$\frac{\partial^2 Q(\theta_0)}{\partial\theta\,\partial\theta'} \quad \text{is negative definite.}$$

Since $\partial^2 Q/\partial\theta\,\partial\theta'$ is continuous, (6.25) implies that there is a closed set

$$B(\theta_0) = \{\theta |\, \|\theta - \theta_0\| \leqslant \delta\}, \qquad B \subset \Theta,$$

such that $\partial^2 Q/\partial\theta\,\partial\theta'$ is negative definite for all θ in $B(\theta_0)$. Let θ_1 be an arbitrary point in $B(\theta_0)$ and $\theta_1 \neq \theta_0$. Then

(6.26) $$Q(\theta_1) = Q(\theta_0) + \frac{1}{2}(\theta_1 - \theta_0)'\frac{\partial^2 Q(\theta^*)}{\partial\theta\,\partial\theta'}(\theta_1 - \theta_0) < Q(\theta_0),$$

where θ^* lies on the line segment between θ_1 and θ_0. Therefore, $Q(\theta)$ attains a unique maximum at θ_0 in $B(\theta_0)$.

Let $\tilde{\theta}_T$ be a point in $B(\theta_0)$ such that $Q_T(\tilde{\theta}_T) \geqslant Q_T(\theta)$ for all θ in $B(\theta_0)$. Then, by Lemma 3, $\tilde{\theta}_T$ converges to θ_0 a.e. But since $Q_T(\theta)$ converges to $Q(\theta)$ uniformly for all θ in $B(\theta_0)$ and $Q(\theta)$ attains its unique maximum at θ_0, $\tilde{\theta}_T$ is a root of the normal equations (6.1). Thus we have proved the following theorem.

THEOREM 1: *Under Assumptions* 1, 2, *and* 3, *the normal equations* (6.1) *have a strongly consistent root.*

In proving Theorem 1, we only needed the existence of $B(\theta_0)$, however small, in which $Q(\theta)$ attains its unique maximum at θ_0. In practice, the larger $B(\theta_0)$ is, the easier to locate the root that corresponds to the global maximum. The ideal situation is that $B(\theta_0)$ could be any compact set specified in Assumption 1. However, that is not possible. From (6.14), we see that by taking $\beta' x_t$ to be negative and small and by taking σ_0 small enough we can make $\partial^2 Q/\partial(\sigma^2)^2$ positive. Nevertheless, from (6.12) we can easily see that $\partial^2 Q/\partial\beta\,\partial\beta'$ is negative definite in any compact parameter space specified in Assumption 1. That means, if σ_0^2 were known, the value of β that maximizes Q_T globally in any compact space of β is strongly consistent.

Asymptotic Normality

Let $\hat{\theta}_T$ be a strongly consistent root of the normal equations (6.1). Then, by a Taylor expansion,

$$(6.27) \qquad \sqrt{T}(\hat{\theta}_T - \theta_0) = \left[\frac{\partial^2 Q_T(\theta^*)}{\partial\theta\,\partial\theta'}\right]^{-1} \sqrt{T}\frac{\partial Q_T(\theta_0)}{\partial\theta},$$

where θ^* lies in the line segment between $\hat{\theta}_T$ and θ_0.

First, we want to prove that $\partial^2 Q_T/\partial\theta\,\partial\theta'$ converges to a function a.e. uniformly for all θ in Θ. An inspection of equations (3.5), (3.6), and (3.7) shows that we can prove the a.e. uniform convergence of each term of the right-hand side of these equations in a way analogous to the proof of (6.4) through (6.8). We know θ^* converges to θ_0 a.e. because $\hat{\theta}_T$ converges to θ_0 a.e. by Theorem 1. Therefore, by Lemma 4,

$$(6.28) \qquad \frac{\partial^2 Q_T(\theta^*)}{\partial\theta\,\partial\theta'} \to \frac{\partial^2 Q(\theta_0)}{\partial\theta\,\partial\theta'} \qquad \text{a.e.},$$

where the latter is given in equations (6.16), (6.17), and (6.18).

Next we consider the asymptotic distribution of $\sqrt{T}\,\partial Q_T(\theta_0)/\partial\theta$. We define l as an arbitrary non-zero vector of dimension $K + 1$ and partition l' as (l'_1, l_2) where l_1 is of dimension K and l_2 is a scalar. Then, from (3.3) and (3.4), we have

$$(6.29) \qquad \sqrt{T}l'\frac{\partial Q_T(\theta_0)}{\partial\theta} = \frac{1}{\sqrt{T}}\sum_1^T \left[l'_1 x_t\left(\frac{u_t^* w_t}{\sigma_0^2} - \frac{f_{0t} v_t}{1 - F_{0t}}\right) + \frac{l_2}{2\sigma_0^2}\left(\frac{\beta'_0 x_t f_{0t} v_t}{1 - F_{0t}} - w_t + \frac{u_t^{*2} w_t}{\sigma_0^2}\right)\right],$$

where u_t^* is defined in Section 4, v_t and w_t are as in (6.2). We denote the tth term of the summation of the right-hand side of (6.29) by ξ_t. Then, using (4.3) and (4.4), we can easily show $E\xi_t = 0$. By Assumptions 1 and 2, the second moment and the third absolute moment of ξ_t are uniformly bounded. From (3.7) and (3.8) we can easily verify

$$(6.30) \qquad \lim_{T\to\infty} TE\left[\frac{\partial Q_T(\theta_0)}{\partial\theta}\cdot\frac{\partial Q_T(\theta_0)}{\partial\theta'}\right] = -\frac{\partial^2 Q(\theta_0)}{\partial\theta\,\partial\theta'},$$

which was earlier proved to be positive definite. Therefore, we have

$$\lim_{T\to\infty}\frac{1}{T}\sum_1^T E\xi_t^2 > 0.$$

Then, by Liapounoff's central limit theorem, the distribution of $\sqrt{T}l'\,\partial Q_T(\theta_0)/\partial\theta$ converges to the univariate normal. Hence, we obtain

$$(6.31) \qquad \sqrt{T}\frac{\partial Q_T(\theta_0)}{\partial\theta} \to N\left(0, -\frac{\partial^2 Q(\theta_0)}{\partial\theta\,\partial\theta'}\right).$$

Thus, (6.27), (6.28), and (6.31) imply the following theorem.

THEOREM 2: *Let $\hat{\theta}_T$ be a strongly consistent root of* (6.1). *Then, under Assumptions* 1, 2, *and* 3,

$$\sqrt{T}(\hat{\theta}_T - \theta_0) \rightarrow N\left(0, \left[-\frac{\partial^2 Q(\theta_0)}{\partial\theta\,\partial\theta'}\right]^{-1}\right).$$

The practical importance of Theorem 2 is that for finite samples the distribution of $\hat{\theta}_T$ may be approximated by

$$N\left(\theta_0, \left[-T\frac{\partial^2 Q(\theta_0)}{\partial\theta\,\partial\theta'}\right]^{-1}\right).$$

Since θ_0 is unknown, the covariance matrix must be estimated. We may estimate $\partial^2 Q(\theta_0)/\partial\theta\,\partial\theta'$ simply by replacing the true values with the estimates in the right-hand side of equations (6.16), (6.17), and (6.18) after removing the limit sign.

7. THE ITERATIVE PROCEDURE

As one can see from (3.3) and (3.4), the partial derivatives $\partial \log L/\partial\theta$ are non-linear. Therefore, a root of the normal equations (6.1) can be found only by an iterative procedure. A well-known iterative procedure is the so-called method of Newton where the second-round estimate $\tilde{\theta}_2$, given an initial estimate $\tilde{\theta}_1$, is defined by

$$\tilde{\theta}_2 = \tilde{\theta}_1 - \left[\frac{\partial^2 \log L(\tilde{\theta}_1)}{\partial\theta\,\partial\theta'}\right]^{-1}\frac{\partial \log L(\tilde{\theta}_1)}{\partial\theta}. \tag{7.1}$$

After obtaining $\tilde{\theta}_2$, one may iterate again to obtain $\tilde{\theta}_3$. The process may be continued indefinitely. Tobin [**8**] suggested the same method. It is computationally feasible, since the only non-trivial function of θ involved in the first and second derivatives, (3.3) to (3.7), is $\phi_t/(1 - \Phi_t)$. This is the inverse of Mills's ratio and can be calculated using an expansion. For a full discussion see Kendall and Stuart [**6**, p. 137].

The method of Newton has both an advantage and a disadvantage. The advantage is that if the initial estimate $\tilde{\theta}_1$ is consistent and $\sqrt{T}(\tilde{\theta}_1 - \theta_0)$ has a proper limit distribution, the second-round estimate $\tilde{\theta}_2$ has the same asymptotic distribution as a consistent root of the normal equations under general conditions. This is shown in Section 9. The disadvantage is that there is no guarantee of the convergence of the iterative procedure to the root that corresponds to the global maximum of the likelihood function. Hartley [**4**] proposed a modified Gauss-Newton method which assures the convergence under fewer assumptions than the unmodified method. However, a similar modification of the method of Newton would not necessarily work in our model because $B(\theta_0)$, the set in which $Q(\theta)$ attains its unique maximum at θ_0, may not be large enough. It is for these reasons that it is important to have a consistent initial estimator. We show below that the initial estimator proposed by Tobin [**8**] is not consistent, and in the next section we propose a consistent and asymptotically normal estimator.

Tobin [8] suggests the following initial estimator: Equate the right-hand side of (3.3) and (3.4) to 0, and pre-multiply (3.3) by $\beta'/2\sigma^2$ and add it to (3.4). Then we get

$$\sigma^2 = \frac{\Sigma_{\bar{\psi}}(y_t - \beta' x_t)y_t}{T - S}. \tag{7.2}$$

From (3.3), $\partial \log L/\partial\beta = 0$ means

$$-\frac{1}{\sigma}\sum_{\psi}\frac{\phi_t}{1-\Phi_t}x_t + \frac{1}{\sigma^2}\sum_{\bar{\psi}}(y_t - \beta' x_t)x_t = 0. \tag{7.3}$$

Approximate $\phi_t/(1 - \Phi_t)$ by a linear function $a + b(\beta' x_t/\sigma)$ and substitute into (7.3), obtaining

$$-\frac{1}{\sigma}\sum_{\psi}\left(a + b\frac{\beta' x_t}{\sigma}\right)x_t + \frac{1}{\sigma^2}\sum_{\bar{\psi}}(y_t - \beta' x_t)x_t = 0. \tag{7.4}$$

Solving (7.4) for β and substituting into (7.2), we can solve the resulting quadratic equation for σ. If the roots are imaginary, one must try something else. If the roots are real, one of them must arbitrarily be chosen. Once an estimate of σ is determined, an estimate of β can be determined linearly from (7.4).

Tobin's estimator may approximate the maximum likelihood estimator very well for a given sample size if the appropriate values for a and b are chosen. However, since the optimal values of a and b depend upon the true value θ_0, it will in general be difficult to approximate them. Moreover, there is no guarantee that one can choose better and better values of a and b as the sample size increases. In fact, Tobin's initial estimator can be easily shown to be inconsistent unless the right values of a and b are accidentally chosen. Suppose the estimator is consistent. Then, taking the probability limit of $1/T$ times the equation (7.4), a and b must satisfy

$$a\sigma_0 \lim \frac{1}{T}\sum_1^T (1 - F_{0t})x_t + b \lim \frac{1}{T}\sum_1^T (1 - F_{0t})x_t x_t'\beta_0 = \sigma_0^2 \lim \frac{1}{T}\sum_1^T f_{0t}x_t. \tag{7.5}$$

But (7.5) can be satisfied only accidentally.

8. A CONSISTENT INITIAL ESTIMATOR

Using the results of Section 4, we obtain

$$Ey_t = \beta_0' x_t + \sigma_0^2\frac{f_{0t}}{F_{0t}}, \qquad t \in \bar{\psi}, \tag{8.1}$$

and

$$Ey_t^2 = (\beta_0' x_t)^2 + \sigma_0^2\beta_0' x_t\frac{f_{0t}}{F_{0t}} + \sigma_0^2, \qquad t \in \bar{\psi}. \tag{8.2}$$

Therefore, from (8.1) and (8.2),

$$Ey_t^2 = \beta_0' x_t Ey_t + \sigma_0^2, \qquad t \in \bar{\psi}. \tag{8.3}$$

Hence,

$$(8.4) \qquad y_t^2 = \beta_0' x_t y_t + \sigma_0^2 + \eta_t, \qquad t \in \bar{\psi},$$

where

$$\eta_t = \beta_0' x_t (E y_t - y_t) + y_t^2 - E y_t^2.$$

We define

$$(8.5) \qquad \hat{y}_t = x_t' \left(\sum_{\bar{\psi}} x_t x_t' \right)^{-1} \sum_{\bar{\psi}} x_t y_t,$$

if $\Sigma_{\bar{\psi}} x_t x_t'$ is positive definite. We use $(x_t'\hat{y}_t, 1)$ as $K + 1$ instrumental variables in the regression equation (8.4), and define the estimator

$$(8.6) \qquad \hat{\theta}_1 = \begin{bmatrix} \hat{\beta}_1 \\ \hat{\sigma}_1^2 \end{bmatrix} = \left[\sum_{\bar{\psi}} \begin{bmatrix} x_t \hat{y}_t \\ 1 \end{bmatrix} (x_t' y_t, 1) \right]^{-1} \sum_{\bar{\psi}} \begin{bmatrix} x_t \hat{y}_t \\ 1 \end{bmatrix} y_t^2,$$

if the matrix to be inverted is non-singular. We will prove below the weak consistency (the convergence to the true value in probability) and the asymptotic normality of $\hat{\theta}_1$.

We have

$$(8.7) \qquad \operatorname{plim} \frac{1}{T} \sum_{\bar{\psi}} \begin{bmatrix} \hat{y}_t x_t \\ 1 \end{bmatrix} [y_t x_t', 1]$$
$$= \operatorname{plim} \frac{1}{T} \sum_{1}^{T} F_{0t} \begin{bmatrix} x_t'\gamma_0 \cdot x_t \\ 1 \end{bmatrix} \left[\left(x_t', \frac{f_{0t}}{F_{0t}} \right) \theta_0 \cdot x_t', 1 \right]$$
$$\stackrel{\text{def}}{=} C,$$

where

$$\gamma_0 = \operatorname{plim} \left(\sum_{\bar{\psi}} x_t x_t' \right)^{-1} \left(\sum_{\bar{\psi}} x_t y_t \right)$$
$$= \beta_0 + \sigma_0^2 \left(\lim \frac{1}{T} \sum_{1}^{T} F_{0t} x_t x_t' \right)^{-1} \left(\lim \frac{1}{T} \sum_{1}^{T} f_{0t} x_t \right).$$

Note that

$$\lim \frac{1}{T} \sum_{1}^{T} F_{0t} x_t x_t'$$

is positive definite because of Assumptions 1, 2, and 3.

From (8.6) we have

$$(8.8) \qquad \hat{\theta}_1 - \theta_0 = \left[\frac{1}{T} \sum_{\bar{\psi}} \begin{bmatrix} x_t \hat{y}_t \\ 1 \end{bmatrix} (x_t' y_t, 1) \right]^{-1} \frac{1}{T} \sum_{\bar{\psi}} \begin{bmatrix} x_t \hat{y}_t \\ 1 \end{bmatrix} \eta_t.$$

But we have

$$\text{(8.9)} \qquad \operatorname{plim} \frac{1}{T} \sum_{\psi} \begin{bmatrix} x_t \hat{y}_t \\ 1 \end{bmatrix} \eta_t = \operatorname{plim} \frac{1}{T} \sum_{1}^{T} \begin{bmatrix} x_t'\gamma_0 \cdot x_t \eta_t \\ \eta_t \end{bmatrix}.$$

Since $E\eta_t$ is zero and $E\eta_t^2$, $x_t'\gamma_0$, and x_t are uniformly bounded by Assumptions 1 and 2, the right-hand side of (8.9) is 0 by Kolmogorov's law of large numbers. Thus, we have proved the next theorem.

THEOREM 3: *Under Assumptions* 1, 2, *and* 3, $\hat{\theta}_1$ *defined in* (8.6) *is weakly consistent* (*i.e., converges to* θ_0 *in probability*), *provided that the matrix* C *defined in* (8.7) *is non-singular.*

The assumption of the non-singularity of C is not restrictive in general. Two important necessary conditions for it are that

$$\text{(8.10)} \qquad \lim \frac{1}{T} \sum_{1}^{T} \begin{bmatrix} x_t'\gamma_0 \cdot x_t \\ 1 \end{bmatrix} [x_t'\gamma_0 \cdot x_t', 1] \quad \text{is positive definite,}$$

and that

$$\text{(8.11)} \qquad \lim \frac{1}{T} \sum_{1}^{T} \begin{bmatrix} \left(x_t', \dfrac{f_{0t}}{F_{0t}}\right)\theta_0 \cdot x_t \\ 1 \end{bmatrix} \left[\left(x_t', \frac{f_{0t}}{F_{0t}}\right)\theta_0 \cdot x_t', 1\right] \quad \text{is positive definite.}$$

Condition (8.10) is not satisfied if x_t is a scalar constant for all t. Then our initial estimator fails to be consistent. However, for this simplest case the estimation problem has been extensively studied in the literature cited in the Introduction, and therefore in this paper we are primarily concerned with a proper regression model where there is at least one regressor that varies with t. Then we can expect that (8.10) and (8.11) hold and that C is non-singular except for an unlikely accident.

From (8.8), we have

$$\text{(8.12)} \qquad \sqrt{T}(\hat{\theta}_1 - \theta_0) = \left[\frac{1}{T} \sum_{\psi} \begin{bmatrix} x_t \hat{y}_t \\ 1 \end{bmatrix} (x_t' y_t, 1)\right]^{-1} \frac{1}{\sqrt{T}} \sum_{\psi} \begin{bmatrix} x_t \hat{y}_t \\ 1 \end{bmatrix} \eta_t .$$

We define l as an arbitrary non-zero vector of dimension $K + 1$ and partition l' as (l_1', l_2) where l_1 is of dimension K and l_2 is a scalar. The limit distribution of

$$\frac{1}{\sqrt{T}} l' \sum_{\psi} \begin{bmatrix} x_t \hat{y}_t \\ 1 \end{bmatrix} \eta_t$$

is the same as that of

$$\text{(8.13)} \qquad \frac{1}{\sqrt{T}} \sum_{1}^{T} (l_1' x_t \cdot x_t'\gamma_0 + l_2)\eta_t w_t .$$

Each term of the summation of (7.12) has mean 0, and its variance and third absolute moment are uniformly bounded by Assumptions 1 and 2. We have

$$(8.14)\qquad \lim_{T\to\infty}\frac{1}{T}E\left[\sum_1^T\begin{bmatrix}x_t'\gamma_0x_t\\ 1\end{bmatrix}\eta_tw_t\cdot\sum_1^T(x_t'\gamma_0x_t',1)\eta_tw_t\right]$$
$$=\lim_{T\to\infty}\frac{1}{T}\sum_1^T\left\{\sigma_0^2[(\beta_0'x_t)^2F_{0t}+2\sigma_0^2F_{0t}+\sigma_0^2\beta_0'x_tf_{0t}]\times\begin{bmatrix}x'\gamma_0x_t\\ 1\end{bmatrix}(x_t'\gamma_0x_t',1)\right\}$$
$$\stackrel{\text{def}}{=} D.$$

If C is non-singular, then (8.10) holds. Therefore, since the variance term is positive, by (6.23), D is positive definite. Then the limit distribution of (8.13) is normal by Liapounoff's central limit theorem, and hence so is $\sqrt{T}(\hat\theta_1-\theta_0)$. We have proved Theorem 4:

THEOREM 4: *Under Assumptions* 1, 2, and 3, $\sqrt{T}(\hat\theta_1-\theta_0)\to N(0,C^{-1}DC^{-1})$, *provided that* C *is non-singular.*

9. THE ASYMPTOTIC EFFICIENCY OF THE SECOND-ROUND ESTIMATOR

Following the Newton method of iteration defined in (7.1), we define $\hat\theta_2$ by

$$(9.1)\qquad \hat\theta_2=\hat\theta_1-\left[\frac{\partial^2\log L(\hat\theta_1)}{\partial\theta\,\partial\theta'}\right]^{-1}\frac{\partial\log L(\hat\theta_1)}{\partial\theta},$$

where $\hat\theta_1$ is the weakly consistent estimator defined in Section 8. We will obtain the asymptotic distribution of $\hat\theta_2$ and show that it is the same as that of a strongly consistent root of the normal equations (6.1).

Substituting a Taylor expansion of $\partial\log L(\hat\theta_1)/\partial\theta$ around θ_0 into (9.1), we obtain

$$(9.2)\qquad \sqrt{T}(\hat\theta_2-\theta_0)=\left\{I-\left[\frac{\partial^2Q_T(\hat\theta_1)}{\partial\theta\,\partial\theta'}\right]^{-1}\left[\frac{\partial^2Q_T(\theta^*)}{\partial\theta\,\partial\theta'}\right]\right\}\sqrt{T}(\hat\theta_1-\theta_0)-\left[\frac{\partial^2Q_T(\hat\theta_1)}{\partial\theta\,\partial\theta'}\right]^{-1}\sqrt{T}\frac{\partial Q_T(\theta_0)}{\partial\theta},$$

where θ^* lies on the line segment between θ_0 and $\hat\theta_1$. In Lemma 4, we may change the expression "a.e." to "with probability approaching one." Then, since both $\hat\theta_1$ and θ^* converge to θ_0 in probability, we have

$$(9.3)\qquad \operatorname{plim}\frac{\partial^2Q_T(\hat\theta_1)}{\partial\theta\,\partial\theta'}=\operatorname{plim}\frac{\partial^2Q_T(\theta^*)}{\partial\theta\,\partial\theta'}=\frac{\partial^2Q(\theta_0)}{\partial\theta\,\partial\theta'}.$$

Therefore, from (9.2), (9.3), and Theorem 4, it follows that the limit distribution of $\sqrt{T}(\hat{\theta}_2 - \theta_0)$ is the same as that of

$$-\left[\frac{\partial^2 Q_T(\theta_0)}{\partial\theta\,\partial\theta'}\right]^{-1}\sqrt{T}\frac{\partial Q_T(\theta_0)}{\partial\theta}.$$

Thus, we have proved the next theorem.

THEOREM 5: *If Assumptions* 1, 2, *and* 3 *hold, and if* C *is nonsingular,*

$$\sqrt{T}(\hat{\theta}_2 - \theta_0) \to N\left(0, \left[-\frac{\partial^2 Q(\theta_0)}{\partial\theta\,\partial\theta'}\right]^{-1}\right).$$

10. CONCLUSIONS

We have considered the estimation of the parameters of the regression model where the dependent variable is normal but truncated to the left of zero. We have proved the strong consistency and asymptotic normality of the maximum likelihood estimator (Theorems 1 and 2), proposed an initial estimator to be used in the method of Newton and proved its weak consistency and asymptotic normality (Theorems 3 and 4), and proved that the second-round estimator obtained in the iteration has the same asymptotic distribution as the maximum likelihood estimator (Theorem 5).

A Fortran (IBM 360) program of the consistent initial estimator, Tobin's initial estimator, and the iteration may be obtained from Professor Michael Boskin, Department of Economics, Stanford University.

For some data it is possible that $\hat{\sigma}_1^2$ defined in (8.6) could turn out to be negative. This happens either because the sample size is too small for the consistency of $\hat{\sigma}_1^2$ to take effect or because the model is incorrectly specified. If it is believed that the former is the case, one should try Tobin's initial estimator. In the event that this is also negative (or imaginary $\hat{\sigma}$), one could only guess the initial value of σ^2. If, on the other hand, the sample size is large, a negative $\hat{\sigma}_1^2$ is an indication that the validity of the model specification should be questioned.

In some applications, data on x_t may be available only for t in $\bar{\psi}$. Renumber the t's in $\bar{\psi}$ as $1, 2, \ldots, M$. Then the density of $y_t, t = 1, 2, \ldots, M$, is given by

$$(10.1) \qquad \frac{1}{F_t}\frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{1}{2}((y_t-\beta' x_t)/\sigma)^2}.$$

Therefore, the logarithmic likelihood function is given by

$$(10.2) \qquad -\sum_1^M \log F_t - \frac{M}{2}\log\sigma^2 - \frac{1}{2\sigma^2}\sum_1^M (y_t - \beta' x_t)^2.$$

Note that (4.2) holds for the y_t above. Because of the similarity of (10.2) to (3.2), the consistency and the asymptotic normality of this model can be easily proved using the results obtained in the paper. Clearly, the initial estimator proposed in

Section 8 of the paper is consistent and asymptotically normal under this model as well, for it uses only the observations in $\bar{\psi}$. The asymptotic efficiency of the second-round estimator in the method of Newton can be similarly proved.

Stanford University

Manuscript received June, 1972.

REFERENCES

[1] COHEN, A. C., JR.: "Estimating the Mean and Variance of Normal Populations from Singly Truncated and Doubly Truncated Samples," *Annals of Mathematical Statistics*, 21 (1950), 557–569.

[2] HALD, A.: "Maximum Likelihood Estimation of the Parameters of a Normal Distribution Which is Truncated at a Known Point," *Skandinavisk Aktuarietidskrift*, 32 (1949), 119–134.

[3] HALPERIN, MAX: "Maximum Likelihood Estimation in Truncated Samples," *Annals of Mathematical Statistics*, 23 (1952), 226–238.

[4] HARTLEY, H. O.: "The Modified Gauss-Newton Method for the Fitting of Non-Linear Regression Functions by Least Squares," *Technometrics*, 3 (1961), 269–280.

[5] JENNRICH, ROBERT I.: "Asymptotic Properties of Non-Linear Least Squares Estimators," *Annals of Mathematical Statistics*, 40 (1969), 633–643.

[6] KENDALL, MAURICE G., AND ALAN STUART: *The Advanced Theory of Statistics*, Vol. 1, 3rd ed. London: Charles Griffin and Co., Ltd., 1963.

[7] ———: *The Advanced Theory of Statistics*, Vol. 2, 3rd ed. London: Charles Griffin and Co., Ltd., 1961.

[8] TOBIN, JAMES: "Estimation of Relationships for Limited Dependent Variables," *Econometrica*, 26 (1958), 24–36.