Econometrica, Vol. 45, No. 7 (October, 1977)

A NOTE ON THE COMPUTATION OF THE TOBIT ESTIMATOR[1]

BY RAY C. FAIR[2]

TOBIT ESTIMATES [4] are generally computed by some version of Newton's method.[3] In this note an alternative procedure is proposed for computing these estimates. Some experimental results are presented that indicate that the procedure may be considerably faster than Newton's method for many problems. For ease of reference, the notation in this note corresponds closely to the notation in Amemiya [1].

1. THE PROCEDURE

The model is

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

where β_0 is a $K\times 1$ vector of unknown coefficients, x_t is a $K\times 1$ vector of values of the explanatory variables for observation t, and u_t is an independently distributed error term with distribution $N(0, \sigma_0^2)$.[4] Assume for simplicity that the sample is ordered so that all of the observations for which the value of the dependent variable is nonzero occur first. Let R be the number of these observations, and let $S = T - R$ be the number of observations for which the value of the dependent variable is zero.

Define the following:

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

$$(3)\qquad f_t = f(\beta' x_t, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}(\beta' x_t/\sigma)^2},$$

$$(4)\qquad \Phi_t = F_t = \int_{-\infty}^{\beta' x_t/\sigma} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\lambda^2}\, d\lambda,$$

$$(5)\qquad \phi_t = \sigma f_t = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(\beta' x_t/\sigma)^2},$$

$$(6)\qquad \gamma_t = \frac{\phi_t}{1-\Phi_t},$$

$$(7)\qquad y' = (y_1, y_2, \ldots, y_R),$$

$$(8)\qquad X' = (x_1, x_2, \ldots, x_R),$$

$$(9)\qquad \bar{X}' = (x_{R+1}, x_{R+2}, \ldots, x_T),$$

$$(10)\qquad \bar{\gamma}' = (\gamma_{R+1}, \gamma_{R+2}, \ldots, \gamma_T).$$

[1] The research described in this paper was undertaken by grants from the National Science Foundation and from the Ford Foundation.

[2] I am indebted to two referees for helpful comments on this note.

[3] In addition to Tobin's original paper [4], see Amemiya [1] for a discussion of the Tobit estimator.

[4] If for a particular model the threshold is not zero, but is instead some known number α, and if x_t contains 1, then this model can be reduced to the model in (1) by redefining the dependent variable to be the original dependent variable minus α. The constant term in this transformed model is then the original constant term minus α. Also, if for a particular model the threshold is an upper bound rather than a lower bound, this model can be reduced to the model in (1) by multiplying all of the data by minus one.

Φ_t is the distribution functon and ϕ_t is the density function of the standard normal variable evaluated at $\beta' x_t/\sigma$. The vector y' is $1\times R$, the matrix X' is $K\times R$, the matrix $\bar{X}'$ is $K\times S$, and the vector $\bar{\gamma}'$ is $1\times S$.

The likelihood function is

$$L=\prod_S (1-F_t)\prod_R \frac{1}{\sqrt{2\pi\sigma^2}}e^{-(1/2\sigma^2)(y_t-\beta' x_t)^2}, \tag{11}$$

where Π_S and Π_R denote multiplication over the S zero and R nonzero observations respectively. L is to be maximized with respect to β and σ^2. The logarithm of L is

$$\log L=\sum_S \log(1-F_t)-\frac{R}{2}\log\sigma^2-\frac{R}{2}\log 2\pi-\frac{1}{2\sigma^2}\sum_R (y_t-\beta' x_t)^2, \tag{12}$$

where Σ_S and Σ_R denote addition over the S zero and R nonzero observations respectively.

The first order conditions for a maximum are

$$\frac{\partial \log L}{\partial \beta}=-\sum_S \frac{x_t f_t}{1-F_t}+\frac{1}{\sigma^2}\sum_R (y_t-\beta' x_t)x_t=0, \tag{13a}$$

$$\frac{\partial \log L}{\partial \sigma^2}=\frac{1}{2\sigma^2}\sum_S \frac{\beta' x_t f_t}{1-F_t}-\frac{R}{2\sigma^2}+\frac{1}{2\sigma^4}\sum_R (y_t-\beta' x_t)^2=0. \tag{13b}$$

Pre-multiplying (13a) by $\beta'/2\sigma^2$ and adding the result to (13b) yields the following equation determining σ^2:

$$\sigma^2=\frac{1}{R}\sum_R (y_t-\beta' x_t)y_t=\frac{y'(y-X\beta)}{R}. \tag{14}$$

After multiplication by σ, (13a) can be written as follows:

$$-\bar{X}'\bar{\gamma}+\frac{1}{\sigma}X'(y-X\beta)=0. \tag{15}$$

Solving (15) for β then yields

$$\begin{aligned}\beta&=(X'X)^{-1}X'y-\sigma(X'X)^{-1}\bar{X}'\bar{\gamma}\\ &=\beta_R^{LS}-\sigma(X'X)^{-1}\bar{X}'\bar{\gamma},\end{aligned} \tag{16}$$

where β_R^{LS} is the ordinary least squares estimate of β_0 for the nonzero observations. Formula (16) shows explicitly the relationship between the ordinary least squares estimator for the nonzero observations and the Tobit estimator.

The procedure proposed in this note for computing the Tobit estimates is as follows:

(1) Compute β_R^{LS} and $(X'X)^{-1}\bar{X}'$.

(2) Choose a value of β, say $\beta^{(1)}$, and compute σ^2 from (14). If this value of σ^2 is less than or equal to zero, take for the value of σ^2 some small positive number. Let $\sigma^{(1)}$ denote the square root of the chosen value of σ^2.

(3) Compute the vector $\bar{\gamma}$ using $\beta^{(1)}$ and $\sigma^{(1)}$. Denote this vector as $\bar{\gamma}^{(1)}$. (A standard FORTRAN function is available for computing Φ_t.)

(4) Compute β from (16) using $\sigma^{(1)}$ and $\bar{\gamma}^{(1)}$. Denote this value as $\tilde{\beta}^{(1)}$. Let $\beta^{(2)}=\beta^{(1)}+\lambda(\tilde{\beta}^{(1)}-\beta^{(1)})$, where $0<\lambda\leq 1$.

(5) Using $\beta^{(2)}$ as the new value of β, go to step (2) and repeat the process. Stop when successive estimates of β are within some prescribed tolerance level.

In carrying out this procedure, the computations in step (1) need only be done once. The parameter λ in step (4) is a damping factor. It is many times useful in procedures of this sort (in order to increase the chance that the procedure will converge) to damp the iteration process by taking λ to be less than one. Olsen [3] has shown that the Tobit likelihood function has a single maximum, so that if the above procedure converges, it converges to the maximum likelihood estimator.

After convergence has been reached, the variance-covariance matrix of $(\beta\,\sigma^2)$ can be estimated as V^{-1}, where[5]

$$V=\begin{bmatrix}\sum_{t=1}^{T} a_t x_t x_t' & \sum_{t=1}^{T} b_t x_t \\ \sum_{t=1}^{T} b_t x_t' & \sum_{t=1}^{T} c_t\end{bmatrix}, \tag{17}$$

and

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

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

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

$$z_t=\frac{\beta' x_t}{\sigma}.$$

It should be stressed that, as with many nonlinear iterative procedures, there is no guarantee that the above procedure will converge. The convergence properties of the present procedure depend, among other things, on the values chosen for $\beta^{(1)}$ and λ. Two obvious choices for $\beta^{(1)}$ are the zero vector and β_R^{LS}. Some experimental results will now be presented that compare, for two problems, the speed of the present procedure for these two choices of $\beta^{(1)}$ and for various choices of λ to the speed of Newton's method.

2. SOME EXPERIMENTAL RESULTS

The two test problems are taken from Fair [2]. The first problem consists of 9 explanatory variables and 601 (150 nonzero and 451 zero) observations, and the second problem consists of 9 explanatory variables and 6,366 (2,053 nonzero and 4,313 zero) observations. The dependent variable in both cases is a measure of time spent in extramarital affairs, a variable that takes on a value for each individual of either zero or some positive number. There are, as one might expect, a large number of zeros in both samples (75.0 and 67.8 per cent respectively), and it turned out in each case that the ordinary least squares estimate of β_0 for the nonzero observations (β_R^{LS}) was not very close to the Tobit estimate. The two samples are from two independent surveys.

The Tobit estimates based on Newton's method were computed using a program called LIMDEP, a program that appears to be fairly widely used. The starting value of β used by this program is zero. The Tobit estimates based on the present procedure were computed using a program written by the author. All computations were done on the IBM 370-158 computer at Yale University. The results are in Table I.

For the present procedure, values of λ between 0.1 and 1.0 were tried in steps of 0.1. The convergence properties of the procedure are clearly sensitive to the value of λ chosen. The best value of λ is 0.4 for the first problem and 0.5 for the second problem. The procedure effectively breaks down for values of λ greater than 0.6 for the first problem and 0.7 for the second problem. The convergence properties are not sensitive to the choice for $\beta^{(1)}$, although for the two problems the zero vector is a slightly better starting point. As mentioned above, β_R^{LS} is not very close to the Tobit estimate for each problem, and so it is not particularly surprising that the zero vector is as good a starting point as β_R^{LS}. For

[5] See Amemiya [1, pp. 1007 and 1010].

problems in which most of the observations are nonzero, β_R^{LS} is likely to be a better starting point.

TABLE I

Present Procedure λ	First Problem (601 observations) $\beta^{(1)}=0$ No. of Iterations	Time (sec.)	$\beta^{(1)}=\beta_R^{LS}$ No. of Iterations	Time (sec.)	Second Problem (6,366 observations) $\beta^{(1)}=0$ No. of Iterations	Time (sec.)	$\beta^{(1)}=\beta_R^{LS}$ No. of Iterations	Time (sec.)
0.1	33	7.5	42	8.2	34	66.2	40	78.4
0.2	16	4.6	21	4.7	16	42.9	23	54.1
0.3	10	3.2	12	3.1	12	34.2	15	37.0
0.4	8	2.6	9	2.5	9	27.8	10	26.3
0.5	17	4.3	20	4.4	8	22.8	8	22.9
0.6	95	19.5	90	16.5	14	32.0	12	27.7
0.7	a	—	a	—	28	56.1	29	55.3
0.8	a	—	a	—	b	—	b	—
Newton's Method	5	18.3			4	127.3		

Notes: a = not converged after 100 iterations; b = not converged after 50 iterations. The time estimates are subject to some measurement error.

Newton's method converged in 5 iterations for the first problem and 4 iterations for the second problem. (Both Newton's method and the present procedure converged to the same answer for each problem within the prescribed tolerance level.) Newton's method required fewer iterations to converge than did the present procedure, but this gain was substantially offset by the much larger number of calculations required per iteration. For the best value of λ, the present procedure appears to be about 6 times faster than Newton's method.[6] Although one must be careful in comparing the speeds of different methods because of the possibility that some methods have been more efficiently programmed than others, the above results are clearly encouraging regarding possible time savings by using the present procedure with a value of λ between about 0.3 and 0.5 over Newton's method. There is no guarantee, of course, that the best value of λ will always lie between 0.3 and 0.5, or even that values of λ within this range will always lead to convergence. When using the present procedure for the first time on a problem, a good strategy would be to start with, say, $\lambda = 0.4$ and an iteration limit of about 20. If after 20 iterations the procedure does not appear to be converging, then smaller values of λ can be tried.

Yale University

Manuscript received October, 1976; revision received January, 1977.

[6] The time estimates presented above do not include set up time, such as time taken to load the data into the program. They include iteration time (including for the present procedure the time taken in step (1) above) and the time taken to compute the variance-covariance matrix of $(\beta\ \sigma^2)$.

REFERENCES

[1] AMEMIYA, TAKESHI: "Regression Analysis When the Dependent Variable is Truncated Normal," *Econometrica*, 41 (1973), 997–1016.

[2] FAIR, RAY C.: "A Theory of Extramarital Affairs," *Journal of Political Economy*, forthcoming 1978.

[3] OLSEN, RANDALL J.: "Note on the Uniqueness of the Maximum Likelihood Estimator for the Tobit Model," mimeo, April, 1976.

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