

**COWLES FOUNDATION FOR RESEARCH IN ECONOMICS**

**AT YALE UNIVERSITY**

**Box 2125, Yale Station**

**New Haven, Connecticut 06520**

**COWLES FOUNDATION DISCUSSION PAPER NO. 434**

**Note:** Cowles Foundation Discussion Papers are preliminary materials circulated to stimulate discussion and critical comment. Requests for single copies of a Paper will be filled by the Cowles Foundation within the limits of the supply. References in publications to Discussion Papers (other than mere acknowledgment by a writer that he has access to such unpublished material) should be cleared with the author to protect the tentative character of these papers.

**A NOTE ON THE COMPUTATION OF THE TOBIT ESTIMATOR\***

**Ray C. Fair**

**October 1, 1976**

## A NOTE ON THE COMPUTATION OF THE TOBIT ESTIMATOR\*

by

Ray C. Fair

Tobit estimates [4] are generally computed by some version of Newton's method.<sup>1</sup> 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].

The model is

$$(1) \quad y_t = \beta_0' x_t + u_t \quad \text{if } \text{RHS} > 0, \\ = 0 \quad \text{if } \text{RHS} \leq 0, \quad (t = 1, 2, \dots, T),$$

where  $\beta_0$  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  $N(0, \sigma_0^2)$ . 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

---

\*The research described in this paper was undertaken by grants from the National Science Foundation and from the Ford Foundation.

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

variable is zero.

Define the following:

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

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

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

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

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

$$(7) \quad y' = (y_1, y_2, \dots, y_R),$$

$$(8) \quad X' = (x_1, x_2, \dots, x_R),$$

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

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

$\Phi_t$  is the distribution function and  $\phi_t$  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

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

where  $\prod_S$  and  $\prod_R$  denote multiplication over the  $S$  zero and  $R$  nonzero observations respectively.  $L$  is to be maximized with respect to  $\beta$  and  $\sigma^2$ . The logarithm of  $L$  is

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

where  $\sum_S$  and  $\sum_R$  denote addition over the  $S$  zero and  $R$  nonzero observations respectively.

The first order conditions for a maximum are

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

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

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

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

After multiplication by  $\sigma$ , (13a) can be written as follows:

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

Solving (15) for  $\beta$  then yields

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

where  $\beta_R^{LS}$  is the ordinary least squares estimate of  $\beta_0$  for the non-zero 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 the note for computing the Tobit estimates is as follows:

- 1) Compute  $\beta_R^{LS}$  and  $(X'X)^{-1}\bar{X}'$ .
- 2) Choose a value of  $\beta$ , say  $\beta^{(1)}$ , and compute  $\sigma^2$  from (14). Denote the square root of this value of  $\sigma^2$  as  $\sigma^{(1)}$ .
- 3) Compute the vector  $\bar{y}$  using  $\beta^{(1)}$  and  $\sigma^{(1)}$ . Denote this vector as  $\bar{y}^{(1)}$ . (A standard FORTRAN function is available for computing  $\Phi_t$ .)
- 4) Compute  $\beta$  from (16) using  $\sigma^{(1)}$  and  $\bar{y}^{(1)}$ . 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  $\beta$ , go to step 2) and repeat the process. Stop when successive estimates of  $\beta$  are within some prescribed tolerance level.

In carrying out this procedure, the computations in step 1) need only be done once. The parameter  $\lambda$  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  $\lambda$  to be less than one. Olsen [3] has shown that the Tobit likelihood func-

tion 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 computed as described in Amemiya [1], p. 1010.

The above procedure was tested against Newton's method on two problems. The first problem consisted of 10 explanatory variables and 601 observations (150 nonzero observations and 451 zero observations). The second problem consisted of 10 explanatory variables and 6366 observations (2053 nonzero observations and 4313 zero observations). Both of these problems are discussed in Fair [2]. The Tobit estimates based on Newton's method were computed using the LIMDEP program, a program that appears to be fairly widely used. 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. In all cases the initial value of  $\beta$  was taken to be zero, and in all cases using the present procedure a value of  $\lambda$  of 0.33 was used. No experimentation was done to see if other values of  $\lambda$  led to faster convergence.

For the first problem Newton's method converged in 5 iterations and took 24.1 seconds of computer time, compared to 9 iterations and 5.9 seconds of computer time for the present procedure. For the second problem Newton's method converged in 4 iterations and took 201.2 seconds of computer time, compared to 14 iterations and 62.1 seconds of computer time for the present procedure. (Both methods converged to the same answer for each problem, within, of course, the prescribed tolerance level.) Newton's method required fewer iterations to converge, but this gain was substantially offset by the much larger number of calculations required per iteration.

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 results just cited are clearly encouraging regarding possible time savings by using the present procedure over Newton's method. The present procedure also has the advantage that it is much easier to program than is Newton's method.

#### REFERENCES

- [1] Amemiya, Takeshi, "Regression Analysis When the Dependent Variable is Truncated Normal," Econometrica, XLI (November 1973), 997-1016.
- [2] Fair, Ray C., "A Theory of Extramarital Affairs," Cowles Foundation Discussion Paper No.436 (1976).
- [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, XXVI (January 1958), 24-36.