A TEST OF THE TOBIT SPECIFICATION AGAINST AN ALTERNATIVE SUGGESTED BY CRAGG

Tsai-Fen Lin and Peter Schmidt*

Abstract—In this paper we present a specification test for the Tobit model. Specifically, we test the Tobit model against the alternative of a two-part model in which one set of parameters determines the probability of a limit observation while a second set of parameters determines the distribution of the non-limit observations. An advantage of our test is that it is easily calculated from the Tobit residuals.

I. Introduction

The Tobit model proposed by Tobin (1958) has been widely used in empirical work for some time, but only recently has much attention been paid to testing the specification of the model (for example, see Nelson (1981), Lee (1981), Ruud (1982) and Olsen (1980)). Nelson's test is a general specification test which compares restricted and unrestricted estimates of various moments of the dependent variable; it is not directed at any specific alternative. The same is true of Olsen's suggestion (p. 1101) to compare actual and predicted numbers of limit observations, and of Ruud's suggestion to test the significance of the difference between the Tobit and Probit estimates. Lee's test is, on the other hand, a test of normality against the alternative of a more general member of the Pearson family.

In this paper we propose a test of the Tobit model against an alternative specification due to Cragg (1971). Cragg's model allows one set of parameters to determine the probability of a limit observation, and a second set of parameters to determine the density of the non-limit observations. The Tobit model is the special case in which the same set of parameters does both. Cragg's model is in some cases a natural alternative to the Tobit model. We suggest using the Lagrange multiplier (LM) test, which is a reasonable test in this case since it is based on the results of estimating the restricted (Tobit) model only, whereas one would have to estimate Cragg's model to perform the Wald or likelihood ratio test.

II. Cragg's Model

The model we consider as an alternative to the Tobit model is the model given by equations (7) and (9) of Cragg (1971, p. 831). The model basically assumes two things. First the probability of a limit observation (a zero) is given by a probit model with parameter vector β_1. That is,

$$P(y_t = 0) = \Phi(-X_t\beta_1), \tag{1}$$

where y_t is the dependent variable, X_t is a row vector of K explanatory variables, β_1 is a column vector of K parameters, and Φ is the standard normal cumulative distribution function (cdf). Here $t = 1, 2, \ldots, T$ indexes observations. Second, it is assumed that the density of y_t, conditional on being a non-limit (positive) observation, is that of $N(X_t\beta_2, \sigma^2)$, truncated at zero. Thus,

$$f(y_t|y_t > 0) = \frac{1}{\Phi(X_t\beta_2/\sigma)} \frac{1}{\sqrt{2\pi}\,\sigma} \times \exp\left\{\frac{-1}{2\sigma^2}(y_t - X_t\beta_2)^2\right\}. \tag{2}$$

If we define the indicator function $I_t = 1$ if $y_t > 0$, $I_t = 0$ if $y_t = 0$, we obtain the log likelihood function

$$\mathscr{L} = \sum_{t=1}^{T} \left\{(1 - I_t)\ln\Phi(-X_t\beta_1) + I_t\left[\ln\Phi(X_t\beta_1) - \ln\Phi(X_t\beta_2/\sigma) - 1/2\ln(2\pi\sigma^2) - \frac{1}{2\sigma^2}(y_t - X_t\beta_2)^2\right]\right\}. \tag{3}$$

This model contains the usual Tobit model as the special case corresponding to $\beta_1 = \beta_2/\sigma$. In that case (3) is easily seen to reduce to the usual Tobit log likelihood.

An alternative to the Tobit model is often well worth considering because the Tobit model is very restrictive. For one thing, in the Tobit model any variable which increases the probability of a non-zero value must also increase the mean of the positive values; a positive element of β means that an increase in the corresponding variable (element of X_t) increases both $P(y_t > 0)$ and $E(y_t|y_t > 0)$. This is not always reasonable. As an example, consider a hypothetical sample of buildings, and suppose that we wish to analyze the dependent variable "loss due to fire," during some time period. Since this is often zero but otherwise positive, the Tobit model might be an obvious choice. However, it is not hard to imagine that newer (and more valuable) buildings might be less likely to have fires, but might have greater average losses when a fire did occur. The Tobit model can not accommodate this possibility.

Another problem with the Tobit model is that it links the shape of the distribution of the positive observations

Received for publication July 6, 1982. Revision accepted for publication March 4, 1983.

* University of Cincinnati and Michigan State University, respectively.

and the probability of a positive observation. For rare events (like fires), the shape of the distribution of the positive observations would have to resemble the extreme upper tail of a normal, which would imply a continuous and faster than exponential decline in density as one moved away from zero. Conversely, when zero occurs less than half of the time, the Tobit model necessarily implies a non-zero mode for the non-zero observations.

Cragg's model avoids both of the above problems with the Tobit model. A reasonably strong case can be made for it as a general alternative to the Tobit model, for analysis of data sets to which Tobit is typically applied—namely, data sets in which zero is a common (and meaningful) value of the dependent variable, and the non-zero observations are all positive. The distribution of such a dependent variable is characterized by the probability that it equals zero and by the (conditional) distribution of the positive observations, both of which Cragg's model parameterizes in a general way.

Of course, there are other existing models which contain the Tobit model as a (testable) special case, but these models are not entirely suitable for application to data sets with the above characteristics. For example, consider the sample selection model of Heckman (1976, 1979), which we write as

$$y_{t1} = X_t\beta_1 + u_{t1}$$
$$y_{t2} = X_t\beta_2 + u_{t2}.$$

Here we observe only the sign of y_{t2}, and we observe y_{t1} if and only if $y_{t2} > 0$.[1] This collapses to the Tobit model if $\beta_1 = \beta_2$ and $u_{t1} = u_{t2}$. However, except in this special case the model does not apply well to data sets with the characteristics listed above: the observed values of y_{t1} need not be positive, in the sense that the model implies a non-zero probability of observed $y_{t1} < 0$; and the unobserved y_{t1} are literally unobserved, rather than observed as equal to zero. The first of these problems can be circumvented, for example, by measuring y_{t1} in logarithms, but then the model no longer includes the Tobit model as a special case, and the second problem is in any case fundamental. To return to the example of fire losses, Heckman's model applies when for some buildings fire losses are literally unknown, while the Tobit model or Cragg's model applies when for some buildings losses are known to equal zero.

Similar comments apply to the models of Heckman (1978) and Lee (1979). We will discuss only the latter model because it includes the former as a special case. Here there are two separate regimes, plus an equation that determines which regime is observed. If the dependent variable in one of the regimes is defined to be identically zero, we have a generalization of the Tobit model; the model reduces to the Tobit model when the regime-switching equation is restricted to be the same as the equation for the non-trivial regime. However, except in this restricted case, nothing in the model guarantees that observations in the non-trivial regime be positive with probability one. As above, this problem can be circumvented, for example, by measuring the dependent variable in logarithms. In this case the resulting model is very similar in spirit to Cragg's model,[2] but it no longer contains the Tobit model as a special case.

III. Derivation of the LM Test

A test of the Tobit specification against Cragg's alternative is a test of the restriction $\beta_1 = \beta_2/\sigma$ in Cragg's model. The LM test is the natural test to use in this case, since it will be based on the Tobit estimates only; it does not require estimation of Cragg's model.

Let θ be the vector of parameters of the model, and let $\tilde{\theta}$ be the maximum likelihood estimates (MLEs) subject to the restriction being tested. Then the LM test statistic is of the form

$$LM = D(\tilde{\theta})'[\mathscr{I}(\tilde{\theta})]^{-1}D(\tilde{\theta}), \tag{4}$$

where $D(\tilde{\theta})$ is the first partial of the log likelihood, evaluated at $\tilde{\theta}$, and $\mathscr{I}(\tilde{\theta})$ is the information matrix, also evaluated at $\tilde{\theta}$. The test statistic is asymptotically distributed as χ^2_K (K being the number of explanatory variables), under the null hypothesis that the Tobit model is correct. (See, e.g., Breusch and Pagan (1980) for more details on the general case.)

To derive the test statistic, it is convenient to reparameterize in a way similar to that suggested by Olsen (1978). Thus we let

$$\theta' = (\S', \beta', h) \tag{5}$$

where

$$\beta = \beta_2/\sigma$$
$$\S = \beta_1 - \beta = \beta_1 - \beta_2/\sigma$$
$$h = 1/\sigma. \tag{6}$$

We then wish to test H_0: $\S = 0$. Note that given H_0, the restricted MLEs are

$$\tilde{\theta}' = [0, \tilde{\beta}', \tilde{h}] \tag{7}$$

where $\tilde{\beta}$ and $\tilde{h}$ are the Tobit estimates. Also, all elements of $D(\tilde{\theta})$ will equal zero except for the first K, which correspond to $\partial L/\partial \tilde{\S}$ (with $\tilde{\S} = 0$); see Breusch and Pagan (1980).

[1] The ordering of the equations is chosen to conform to the previously used notation for Cragg's model, and is the reverse of Heckman's ordering. Obviously this is not a matter of substance.

[2] The difference is only whether one assumes the conditional distribution of the positive observations to be truncated normal, or the unconditional distribution to be lognormal.

If we reparameterize the likelihood (3) in accordance with (6), differentiate with respect to §, and then evaluate at $\tilde{\theta}$, we obtain

$$\frac{\partial L}{\partial \tilde{\S}} = \sum_t [I_t m(X_t\tilde{\beta}) - (1 - I_t) m(-X_t\tilde{\beta})] X_t' \quad (8)$$

where $m(\cdot) = \phi(\cdot)/\Phi(\cdot)$, ϕ and Φ being, respectively, the standard normal density and cdf. However, the Tobit first order conditions imply

$$\sum_t (1 - I_t) m(-X_t\tilde{\beta}) X_t' = \sum_t I_t(\tilde{h}y_t - X_t\tilde{\beta}) X_t' \quad (9)$$

and thus we obtain

$$\frac{\partial L}{\partial \tilde{\S}} = \sum_{t=1}^{T} I_t [m(X_t\tilde{\beta}) - (\tilde{h}y_t - X_t\tilde{\beta})] X_t'. \quad (10)$$

This expression (10) gives the first K elements of $D(\tilde{\theta})$; as noted above, the last $K + 1$ elements of $D(\tilde{\theta})$ equal zero.

Incidentally, (10) is interpretable as a vector of cross products between the explanatory variables and the Tobit "residuals" for the non-limit observations. This is so in the sense that

$$E[(hy_t - X_t\beta) | y_t > 0] = m(X_t\beta) \quad (11)$$

and thus the term in brackets in (10) can be regarded as the Tobit "residual."

The information matrix evaluated at $\tilde{\theta}$, $\mathcal{I}(\tilde{\theta})$, takes the form[3]

$$\mathcal{I}(\tilde{\theta}) = \begin{bmatrix} \mathcal{I}_{\S\S'} & \mathcal{I}_{\S\beta'} & \mathcal{I}_{\S h} \\ \mathcal{I}_{\beta\S'} & \mathcal{I}_{\beta\beta'} & \mathcal{I}_{\beta h} \\ \mathcal{I}_{h\S'} & \mathcal{I}_{h\beta'} & \mathcal{I}_{hh} \end{bmatrix} = \begin{bmatrix} \sum_t a_t X_t'X_t & \sum_t a_t X_t'X_t & 0 \\ & \sum_t (a_t + b_t) X_t'X_t & \sum_t c_t X_t' \\ & & \sum_t d_t \end{bmatrix} \quad (12)$$

where a_t, b_t, c_t and d_t are defined by

$$\begin{aligned} a_t &= m(-X_t\tilde{\beta}) m(X_t\tilde{\beta}) \\ b_t &= \Phi(X_t\tilde{\beta})[1 - X_t\tilde{\beta} m(X_t\tilde{\beta}) - m^2(X_t\tilde{\beta})] \\ c_t &= \frac{1}{\tilde{h}} \Phi(X_t\tilde{\beta})[X_t\tilde{\beta} + m(X_t\tilde{\beta})] \\ d_t &= \frac{1}{\tilde{h}^2} \Phi(X_t\tilde{\beta})\left[2 + (X_t\tilde{\beta})^2 + (X_t\tilde{\beta}) m(X_t\tilde{\beta})\right]. \end{aligned} \quad (13)$$

Given the form of $D(\tilde{\theta})$, we require for LM only the upper left $K \times K$ submatrix of $[\mathcal{I}(\tilde{\theta})]^{-1}$, say $\mathcal{I}^{\S\S'}$. Using the partitioned inverse rule and some algebra yields

$$\mathcal{I}^{\S\S'} = \left[\sum_t a_t X_t'X_t\right]^{-1} + \left[\sum_t b_t X_t'X_t - \left(\sum_t c_t X_t\right)'\left(\sum_t c_t X_t\right) / \sum_t d_t\right]^{-1}. \quad (14)$$

Finally, then, the LM statistic is

$$LM = \left(\frac{\partial L}{\partial \tilde{\S}}\right)' \mathcal{I}^{\S\S'} \left(\frac{\partial L}{\partial \tilde{\S}}\right), \quad (15)$$

where $\partial L/\partial \tilde{\S}$ is given by (10) and $\mathcal{I}^{\S\S'}$ is given by (14). Although $\mathcal{I}^{\S\S'}$ has no obvious interpretation, it is easily calculated from one's Tobit results. On the other hand, $\partial L/\partial \tilde{\S}$ is both easily calculated, and also easily interpreted, as the vector of cross-products of the explanatory variables with the Tobit "residuals."

IV. Concluding Remarks

Our LM test is a relatively simple test of the hypothesis that the Tobit model is correctly specified, against the alternative that different sets of parameters determine the probability of a limit observation and the density of the non-limit observations. Its chief advantage is that it can be performed from the results of estimating the Tobit model. This avoids estimation of the more complicated alternative model of Cragg, at least in cases in which the Tobit model is not rejected.

While Cragg's model is more complicated to estimate than the Tobit model, we should not overstate the difficulty involved in estimating it. The main difficulty is probably the lack of readily available software, whereas programs for Tobit estimation are widespread. In our opinion, the Tobit model is typically used with more faith than it warrants, but this is of course an empirical question, upon which this test (and the other specification tests listed in section I) should shed some light. If

[3] The information matrix is calculated as the expected value of the matrix of second partials of the log likelihood with respect to the parameters. Evaluation of these expected values requires the facts:

$$\begin{aligned} E(I_t) &= \Phi[X_t(\beta + \S)] \\ E(I_t y_t^2) &= h^{-1}[X_t\beta + m(X_t\beta)]\Phi[X_t(\beta + \S)] \\ E(I_t y_t^2) &= h^{-2}\Phi[X_t(\beta + \S)] \\ &\quad \times \left[1 + (X_t\beta) m(X_t\beta) + (X_t\beta)^2\right]. \end{aligned}$$

The first of these is trivial, while the last two follow directly from standard formulae for the moments of a truncated normal variable; see, e.g., Johnson and Kotz (1970, pp. 81–83). The information matrix is then evaluated at $\tilde{\beta}$, $\tilde{h}$, and $\tilde{\S}$ (where of course $\tilde{\S} = 0$).

the Tobit model fails to do well when confronted with such tests, Cragg's model (or other possible generalizations of Tobit) should see more use.

Finally, it should be emphasized that while our test is a test against the alternative of Cragg's model, it will likely show some power against a variety of possible misspecifications of Tobit. Thus a rejection of Tobit by this test need not imply that Cragg's model is indicated by the data. However, as we have argued in section II, Cragg's model is a very plausible alternative to Tobit in many cases, and thus may be an obvious candidate to consider if Tobit is rejected. The most important restrictive assumption remaining in Cragg's model is probably that (like Tobit) it assumes normality. Thus it is presumably worthwhile to extend the recent work on (distributionally) robust versions of Tobit and the sample selection model to Cragg's model as well.

REFERENCES

Breusch, Trevor, and Adrian R. Pagan, "The Lagrange Multiplier Test and Its Application to Model Specification in Econometrics," *Review of Economic Studies* 47 (Mar. 1980), 239–253.

Cragg, John G., "Some Statistical Models for Limited Dependent Variables with Application to the Demand for Durable Goods," *Econometrica* 39 (Sept. 1971), 829–844.

Heckman, James J., "The Common Structure of Statistical Models of Truncation, Sample Selection and Limited Dependent Variables and a Simple Estimator for Such Models," *Annals of Economic and Social Measurement* 5 (Fall 1976), 475–495.

______, "Dummy Endogenous Variables in a Simultaneous Equation System," *Econometrica* 46 (July 1978), 931–960.

______, "Sample Selection Bias as a Specification Error," *Econometrica* 47 (Jan. 1979), 153–162.

Johnson, Norman L., and Samuel Kotz, *Distributions in Statistics: Continuous Univariate Distributions-I* (New York: John Wiley, 1970).

Lee, Lung Fei, "Identification and Estimation in Binary Choice Models with Limited (Censored) Dependent Variables," *Econometrica* 47 (July 1979), 977–996.

______, "A Specification Test for Normality Assumption for the Truncated and Censored Tobit Models," Discussion Paper 44, Center for Econometrics and Decision Sciences, University of Florida (1981).

Nelson, Forrest D., "A Test for Misspecification in the Censored Normal Model," *Econometrica* 49 (Sept. 1981), 1317–1330.

Olsen, Randall J., "A Note on the Uniqueness of the Maximum Likelihood Estimator for the Tobit Model," *Econometrica* 46 (Sept. 1978), 1211–1216.

______, "Approximating a Truncated Normal Regression with the Method of Moments," *Econometrica* 48 (July 1980), 1099–1106.

Ruud, Paul A., "A Score Test of Consistency," unpublished manuscript, University of California, Berkeley (1982).

Tobin, James, "Estimation of Relationships for Limited Dependent Variables," *Econometrica* 26 (Jan. 1958), 24–36.

DUMMY VARIABLES: MECHANICS V. INTERPRETATION

Daniel B. Suits*

Abstract—Regressions containing dummy variables are easily estimated by the familiar expedient of "dropping out" one of the categories but the result is often awkward to interpret. Since coefficients of dummy variables are determined only up to an additive constant, however, the equation can be transformed into a more easily interpretable form by adding on an appropriately chosen constant to each coefficient.

For most regressions the constants should be chosen to force the mean of the transformed coefficients to equal 0. For logarithmic regressions the constants should be chosen to force the sum of the antilogs of the coefficients to equal 1. With logarithmic demand curves fitted to monthly data the resulting antilogs become monthly seasonal indexes.

The technical procedure by which dummy variables are used to capture the influence of categorical variables in regression equations is generally familiar (see Goldberger (1964), Kmenta (1971), Johnston (1960), or, to go back near the beginning of things, Suits (1957)). In many cases, particularly where only two classes of observation are involved, results presented in the usual way involve no special problems of interpretation. For example, use of a dummy variable to distinguish pre-war from post-war behavior, or to measure the shift in a relationship during the period of a strike is readily understood by any reader. But where a set of several dummy variables is employed to measure the variation in behavior among a number of classes—regions, education groups, age brackets, and the like—there is often an important difference between the purely mechanical problem of fitting the regression and the quite different problem of presenting the results in the most effective fashion. The purpose of this paper is to call attention to this distinction, and to illustrate by simple examples.

Received for publication June 1, 1982. Revision accepted for publication November 30, 1982.

* Michigan State University and East-West Population Institute.

Regional Variation

To examine the variation of household expenditure for gasoline with income and region of residence, a least-squares regression is fitted with the result (figures