The Tobit Model

Econ 674

Purdue University

Estimation

- In this lecture, we address estimation and application of the *tobit* model.
- The tobit model is a useful specification to account for mass points in a dependent variable that is otherwise continuous.
- For example, our outcome may be characterized by lots of zeros, and we want our model to speak to this incidence of zeros.
  1.
  2.
  3.

The tobit

Like the probit and ordered probit, the tobit model can be given a latent variable interpretation. We write this as follows:

-

We observe data on (x_i, y_i) but not on z_i. Note that z_i is *partially observed*.

Note that, unlike the probit and ordered probit, the scale parameter is not fixed at unity (why)?

In some cases, application of the tobit is, perhaps, not ideal while in others, the tobit can be applied more credibly. Two examples illustrate.

The tobit

Case #1:

Suppose we seek to model expenditures on automobiles during the calendar year. We apply a *tobit* to model this data. How would you interpret your model in terms of this specific application?

Many would give z_i an interpretation like *desired* expenditure. If this is positive, then the person buys a car and spends the desired amount. If this is negative or zero, then we simply see that the person did not buy the car.

Are there any problems here?

1.

2.

The tobit

Case #2:

Suppose that you seek to model expenditures on tobacco products during the calendar year. The observed variable y_i represents the fraction of income spent on such products during the calendar year. The data is likely characterized by lots of zeros.

In this case,

1. It is quite likely to see y_i values very close to zero, given its construction.
2. Perhaps negative values of z_i make more sense in the context of this application. Specifically, people may contribute to anti-smoking campaigns, which we might interpret as a type of negative expenditure.

The tobit

An important and often overlooked point is that, although it might seem natural to assert that the "censoring" point is at zero, it may, in fact, be something different from zero. [Zuehlke (2003)].

That is, there may be some minimum level of expenditure that is possible.

For this reason, we might consider a variant of the tobit with an *unknown censoring point*:

-

for some constant c that is to be estimated from the data.

Estimation in the (Standard) Tobit

$$z_i = x_i\beta + u_i, \quad u_i|x_i \overset{iid}{\sim} \mathcal{N}(0, \sigma^2)$$

$$y_i = \max\{0, z_i\}.$$

To derive the log likelihood in the tobit, (though it is not necessary to do so), we first consider the c.d.f. :

$$\Pr(Y_i \leq c|X).$$

It is convenient to express this probability in the following way:

-

where D_i can be any binary variable, yet it is convenient to define it here as

-

Estimation in the (Standard) Tobit

$$\begin{aligned}\Pr(Y_i \leq c|x_i) &= \Pr(Y_i \leq c|x_i, D_i = 1)\Pr(D_i = 1|x_i) \\ &+ \Pr(Y_i \leq c|x_i, D_i = 0)\Pr(D_i = 0|x_i),\end{aligned}$$

With respect to the components of the above, some of these are straightforward:

-

and hence, $\Pr(D_i = 0|x_i) = 1 - \Phi(x_i\beta/\sigma)$. What about $\Pr(Y_i \leq c|x_i, D_i = 0)$? Intuitively,

-

Estimation in the (Standard) Tobit

As for the remaining conditional density, note for $c > 0$:

-

Estimation in the (Standard) Tobit

Thus, we obtain the following "density" function for Y_i:

$$f(y_i|x_i) = \frac{1}{\sigma}\phi([y_i - x_i\beta]/\sigma)I(y_i > 0) + I(y_i = 0)\left[1 - \Phi(x_i\beta/\sigma)\right].$$

From here, it is not hard to get to the log likelihood:

-

In the above $n_1 = \sum_{i=1}^{n} D_i$, or the number of uncensored observations.

Estimation in the (Standard) Tobit

- From here, a standard tobit analysis can be carried out.
- That is, the score vector can be obtained, as can the Hessian matrix.
- However, these are quite messy, particularly the Hessian.
- Moreover, it turns out that a reparameterization of the problem simplifies these expressions considerably and, furthermore, that we can prove global concavity for the reparameterized model.

Estimation in the (Standard) Tobit

We employ the reparameterization suggested by Olson (1978). Specifically, we let

-

Then we obtain

$$L(\delta, \theta; y) = -\frac{n_1}{2}\log(2\pi) + n_1 \log\delta - \frac{1}{2}\sum_{i:y_i>0}(\delta y_i - x_i\theta)^2 + \sum_{i:y_i=0}\log[1-\Phi(x_i\theta)].$$

From this, we obtain the score:

-

Estimation in the (Standard) Tobit

With a bit of work, the components of the Hessian matrix can also be obtained:

$$L_{\theta\theta'} = \sum_{i:y_i=0} \frac{\phi(x_i\theta)}{1-\Phi(x_i\theta)} \left(x_i\theta - \frac{\phi(x_i\theta)}{1-\Phi(x_i\theta)} \right) x_i'x_i - \sum_{i:y_i>0} x_i'x_i.$$

Estimation in the (Standard) Tobit

Let

$$\gamma = [\theta' \;\; \delta]', \quad X = \left[\begin{array}{c} X_0 \\ X_1 \end{array} \right], \quad y = [y_0 \;\; y_1]',$$

where X_0 consists of the X observations with $y_i = 0$ (and similarly for X_1, etc.). That is, we first arrange the data with the $y_i = 0$ outcomes appearing first, followed by those with $y_i = 1$.

With this notation in hand, one can show that the Hessian can be written as:

-

Estimation in the (Standard) Tobit

In the last slide, D_i is an $n_0 \times n_0$ matrix with diagonal element

$$-\frac{\phi(x_i\theta)}{1-\Phi(x_i\theta)}\left(x_i\theta - \frac{\phi(x_i\theta)}{1-\Phi(x_i\theta)}\right).$$

Furthermore, one can show that the Hessian is always negative semidefinite (and thus the log likelihood is globally concave) provided the elements of D are positive. (why?)

Given the form of these elements above, this is true *iff*

$$x_i\theta - \frac{\phi(x_i\theta)}{1-\Phi(x_i\theta)} < 0.$$

This is indeed true, but in order to prove it, we must digress a little bit.

Mean of Truncated Normal

To obtain the density function for any truncated random variable w, we apply the formula:

-

That is, we keep the shape of the marginal density, chop off the tail, and scale it up to make sure it integrates to unity. Thus,

-

For the case of a standard normal random variable w, with $c = x_i\theta$, we get:

-

Mean of Truncated Normal

Now, clearly, it must be the case that

-

In the case of a standard normal random variable w, then, we have

-

or

-

Note that this is exactly the term we needed to prove was negative in order to verify that the Hessian is negative semidefinite.

Mean of Truncated Normal

- This result motivates use of the reparameterization in practice.
- An iterative maximization routine should converge quickly to the maximum given the uniqueness of this maximum.
- *Invariance* can be applied to estimate β and σ. Specifically,

$$\hat{\sigma} = \hat{\delta}^{-1}, \qquad \hat{\beta} = \hat{\theta}/\hat{\delta}.$$

- The *Delta method* can be used to obtain large sample standard errors.

A note on discarding the zeros

It is somewhat common, though unfortunate, practice in the applied literatures to simply discard the zero responses when estimating the tobit. Of course, this is not a valid procedure since:

$$\begin{aligned} E(y_i|x_i, y_i > 0) &= x_i\beta + E(u_i|u_i > -x_i\beta, x_i) \\ &= x_i\beta + \sigma\frac{\phi(x_i\beta)}{\Phi(x_i\beta)}. \end{aligned}$$

Thus, the conditional mean function, given that positive values occur, is not simply the population conditional mean $x_i\beta$. As such, OLS results will be biased and inconsistent.

Marginal Effects

We now describe a method for calculating marginal effects in the tobit. Though several of these have been discussed, we focus our attention on effects with respect to the mean of the *observed* y outcome: First, note (similar to our previous discussion):

$$\begin{aligned} E(y|x) &= E(y|x, z > 0)\Pr(z > 0|x) + E(y|x, z \leq 0)\Pr(z \leq 0|x) \\ &= E(y|x, z > 0)\Pr(z > 0|x) \end{aligned}$$

(why?) Hence, we have:

$$\frac{\partial E(y|x)}{\partial x_j} = \frac{\partial E(y|x, z > 0)}{\partial x_j}\Pr(z > 0|x) + E(y|x, z > 0)\frac{\partial \Pr(z > 0|x)}{\partial x_j}.$$

Marginal Effects

To make things a bit simpler notationally, let $\phi \equiv \phi(x_i\beta/\sigma)$ and define Φ analogously.

To put together all of the pieces of the marginal effect expression, we first note:

$$\begin{aligned} E(y|x, z > 0) &= E(z|x, z > 0) \\ &= x\beta + E(u|x, z > 0) \\ &= x\beta + E(u|u > -x\beta, x). \end{aligned}$$

The last term, again, is the mean of a truncated normal random variable, though in this case the variance of u is σ^2 rather than unity. It follows by similar reasoning that

-

Marginal Effects

In order to completely characterize the marginal effect, we must differentiate the normal density function. That is, we seek:

$$\begin{aligned}
\frac{\partial \phi}{\partial x_j} &= \frac{\partial \left[(2\pi)^{-1/2} \exp\left(-[1/2](x\beta/\sigma)^2\right)\right]}{\partial x_j} \\
&= (2\pi)^{-1/2} \exp\left(-[1/2](x\beta/\sigma)^2\right)(-x\beta/\sigma)(\beta_j/\sigma) \\
&= \phi[-x\beta/\sigma](\beta_j/\sigma)
\end{aligned}$$

Therefore,

-

Marginal Effects

Putting this together with the other pieces comprising our marginal effect, we obtain:

$$\frac{\partial E(y|x)}{\partial x_j} = \left[\beta_j + \sigma \frac{-\phi\Phi[x\beta/\sigma](\beta_j/\sigma) - \phi^2(\beta_j/\sigma)}{\Phi^2}\right]\Phi + \left(x\beta + \sigma\frac{\phi}{\Phi}\right)\phi[\beta_j/\sigma]$$

Rather conveniently, terms cancel to produce:

-

Any intuition here?

- As $\Phi \to 1$, the probability associated with the mass point at zero approaches zero. In this limiting case, we are essentially back into linear regression framework, whence the marginal effect reduces to β_j.

Tobit: Application

- Using the female labor supply data on the course website, we fit a tobit model to account for the censoring at zero weeks of work.
- We work in the (δ, θ) parameterization.
- The fsolve command is used in MATLAB, so the score vector is programmed into the maximization routine.
- The following slide gives results from this exercise.

Tobit: Application

Variable	MATLAB			STATA	
	Pt. Est	Std. Err	Marg Eff	Pt. Est	Stderr
Constant	31.93	3.83	—-	31.93	3.66
Ability	.061	.0221	.056	.061	.022
SpouseInc.	-.123	.0254	-.114	-.123	.025
Kids	-13.52	1.22	-12.55	-13.52	1.16
Education	.932	.292	.865	.932	.291
σ	23.24	.383	—-	23.24	.383

References

- Tobin, J. (1958). “Estimation of Relationships for Limited Dependent Variables” *Econometrica*.
- Olsen, R.J. (1978). “ANote on the Uniqueness of the Maximum Likelihood Estimator for the tobit model” *Econometrica*.
- Zuehlke, T. (2003). "Estimation of a Tobit Model with Unknown Censoring Threshold.” *Applied Economics*.