The 'lobit' Model

Miles D. Williams

June 11, 2019

Censored Regression

It is well known that in the context of a censored outcome variable, the least squares estimator for the linear association between a set of explanatory variables and said outcome is biased. To counteract this problem, James Tobin devised what would come to be affectionately called the "Tobit" model. This model blends the components of a probit and Gaussian model (OLS under restrictive assumptions) to fit a linear predictor, not to the observed outcome, but to an unobserved (*latent*) outcome. In the decades since Tobit was developed, many new generalized forms and adaptations of the original model have been created; though, each abides by similar principles, chief of which is a reliance on the probit link function to model the "censored" component of the outcome variable in question. With relative ease, I replace the probit link function with the logit link function. Much as logit and probit models are viewed as viable alternatives for fitting binary data, so can this new logistic Tobit, or *lobit* model be viewed as an alternative or fitting censored data.

The Likelihood Function

The likelihood function for a basic binary outcome model is given as

$$L = \prod_i [\Pr(y_i|X)^{y_i}(1 - \Pr(y_i))^{1-y_i}] \tag{1}$$

where for a probit model,

$$\Pr(y_i|X) \equiv \Phi\left(\frac{X'\beta}{\sigma}\right),$$

where $\Phi(\cdot)$ denotes the cumulative density function where β and σ a parameters to be estimated. Meanwhile, for a logit model,

$$\Pr(y_i|X) \equiv \frac{e^{X'\beta}}{1 + e^{X'\beta}}.$$

The likelihood function for a Gaussian (normal) model is given as

$$L = \prod_i \sigma^{-1}\varphi\left(\frac{y_i - X'\beta}{\sigma}\right)$$

where $\varphi(\cdot)$ is the probability density function and β and σ are again parameters to be estimated.

The Tobit likelihood function is simply a combination of the likelihood functions for a binary outcome with a probit link and for a normal model. For the classic case where y is a censored outcome with a lower bound of 0 and upper bound of ∞, that is:

$$L = \prod_i \left[1 - \Phi\left(\frac{X'\beta}{\sigma}\right)\right]^{1-D_i} \left[\sigma^{-1}\varphi\left(\frac{y_i - X'\beta}{\sigma}\right)\right]^{D_i} \tag{3}$$

where D_i is a dummy that takes the value 1 when $y_i > 0$, 0 otherwise.

Adopting a logit link function in lieu of a probit link function is straightforward, simply requiring specifying the likelihood function as

$$L = \prod_i \left[1 - \frac{e^{X'\beta}}{1 + e^{X'\beta}}\right]^{1-D_i} \left[\sigma^{-1}\varphi\left(\frac{y_i - X'\beta}{\sigma}\right)\right]^{D_i} \tag{4}$$