HW2 Econometrics 3

Matthew Aaron Looney

10/04/2017

Problem 2

Problem 2. Censoring/Truncation. Greene (2007) analyzed the de- fault behavior and monthly behavior of a large sample of credit card users (13,444).

(2.1)

Estimate the following model

$$\log spend = \beta_1 + \beta_2 \ln income + \beta_3 Age + \beta_4 Adepcnt + \beta 5 ownrent + \varepsilon$$

Table 1: Regression output used to answer Problem 2

	Dependent variable:		
	LOGSPEND		NA
	OLS	*censored regression*	*Heckman selection*
	(1)	(2)	(3)
Ln_income	1.121***	1.117***	0.907***
	(0.033)	(0.033)	(0.162)
AGE	−0.015***	−0.014***	−0.014***
	(0.001)	(0.001)	(0.002)
ADEPCNT	−0.027**	−0.027**	0.016
	(0.011)	(0.011)	(0.034)
OWNRENT	−0.203***	−0.201***	−0.281***
	(0.030)	(0.030)	(0.065)
logSigma		0.296***	
		(0.007)	
Constant	−3.363***	−3.340***	−1.419
	(0.243)	(0.246)	(1.458)
Observations	10,499	10,499	13,444
R^2	0.105		0.105
Adjusted R^2	0.104		0.104
Log Likelihood		−18,012.210	
Akaike Inf. Crit.		36,036.430	
Bayesian Inf. Crit.		36,079.980	
ρ			−0.608
Inverse Mills Ratio			−0.878 (0.646)
Residual Std. Error	1.330 (df = 10494)		
F Statistic	306.358*** (df = 4; 10494)		
Note:			*p<0.1; **p<0.05; ***p<0.01

(2.1.a)

Using OLS. What is the effect of 10% increase in income on credit card expenditure?

- Since we are dealing with log-log we can simply multiply the paramater estimate on income by ten, which gives 11.2120776. So a 10% increase in income is estimated to increase credit card spending by 11.2120776%.

(2.1.b)

Using Censored regression. What is the effect of 10% increase in income on credit card expenditure?

We will need to employ a Censored (Tobit) Regression and calculate the Partial Effects.

The general formulation for the Tobit Model (Greene 7th. ed., pg 848):

$$
\begin{aligned}
{y_i}^* &= {x_i}'\beta + \varepsilon_i \\
\begin{matrix} y_i = 0 \\ y_i = {y_i}^* \end{matrix} & \begin{cases} \text{if } {y_i}^* \leqslant 0 \\ \text{if } {y_i}^* \geqslant 0 \end{cases}
\end{aligned}
$$

The censored regression model is a generalisation of the standard Tobit model. The dependent variable can be either left-censored, right-censored, or both left-censored and right-censored, where the lower and/or upper limit of the dependent variable can be any number:

$$y_i^* = x_i'\beta + \varepsilon_i \tag{1}$$

$$y_i = \begin{cases} a & \text{if } y_i^* \leq a \\ y_i^* & \text{if } a < y_i^* < b \\ b & \text{if } y_i^* \geq b \end{cases} \tag{2}$$

Here a is the lower limit and b is the upper limit of the dependent variable. If $a = -\infty$ or $b = \infty$, the dependent variable is not left-censored or right-censored, respectively.

Censored regression models (including the standard Tobit model) are usually estimated by the Maximum Likelihood (ML) method. Assuming that the disturbance term ε follows a normal distribution with mean 0 and variance σ^2, the log-likelihood function is

$$
\begin{aligned}
\log L = \sum_{i=1}^{N} \Bigg[& I_i^a \log \Phi\left(\frac{a - x_i'\beta}{\sigma}\right) + I_i^b \log \Phi\left(\frac{x_i'\beta - b}{\sigma}\right) \\
& + \left(1 - I_i^a - I_i^b\right)\left(\log \phi\left(\frac{y_i - x_i'\beta}{\sigma}\right) - \log \sigma\right)\Bigg],
\end{aligned} \tag{3}
$$

where $\phi(\cdot)$ and $\Phi(\cdot)$ denote the probability density function and the cumulative distribution function, respectively, of the standard normal distribution, and I_i^a and I_i^b are indicator functions with

$$I_i^a = \begin{cases} 1 & \text{if } y_i = a \\ 0 & \text{if } y_i > a \end{cases} \tag{4}$$

$$I_i^b = \begin{cases} 1 & \text{if } y_i = b \\ 0 & \text{if } y_i < b \end{cases} \tag{5}$$

The log-likelihood function of the censored regression model~(3) can be maximised with respect to the parameter vector $(\beta', \sigma)'$ using standard non-linear optimisation algorithms.

The proper Marginal (Partial) Effects formula:

$$\frac{\partial E[y|x]}{\partial x} = \beta \Pr ob[a < y^* < b]$$

The marginal effects of an explanatory variable on the expected value of the dependent variable is (Greene 7th. ed., pg 849):

$$ME_j = \frac{\partial E[y|x]}{\partial x_j} = \beta_j \left[\Phi\left(\frac{b - x'\beta}{\sigma}\right) - \Phi\left(\frac{a - x'\beta}{\sigma}\right) \right] \tag{6}$$

In order to compute the approximate variance covariance matrix of these marginal effects using the Delta method, we need to obtain the Jacobian matrix of these marginal effects with respect to all estimated parameters (including~σ):

$$\frac{\partial ME_j}{\partial \beta_k} = \Delta_{jk} \left[\Phi\left(\frac{b - x'\beta}{\sigma}\right) - \Phi\left(\frac{a - x'\beta}{\sigma}\right) \right] - \frac{\beta_j \, x_k}{\sigma} \left[\phi\left(\frac{b - x'\beta}{\sigma}\right) - \phi\left(\frac{a - x'\beta}{\sigma}\right) \right] \tag{7}$$

and

$$\frac{\partial ME_j}{\partial \sigma} = -\beta_j \left[\phi\left(\frac{b - x'\beta}{\sigma}\right) \frac{b - x'\beta}{\sigma^2} - \phi\left(\frac{a - x'\beta}{\sigma}\right) \frac{a - x'\beta}{\sigma^2} \right], \tag{8}$$

where Δ_{jk} is "Kronecker's Delta"

with $\Delta_{jk} = 1$ for $j = k$ and $\Delta_{jk} = 0$ for $j \neq k$. If the upper limit of the censored dependent variable (b) is infinity or the lower limit of the censored dependent variable (a) is minus infinity, the terms in the square brackets in equation~(8) that include b or a, respectively, have to be removed.

- Where I compute the partial effect at each observation and then compute the mean.
- The parginal effect of Ln_income on LOGSPEND is 1.1169911. Therefore, a 10% increase of income is estiimated to increase credit card spending by 11.169911.

(2.1.c)

Using Heckman Two-Step Estimator. What the is effect of 10% increase in income on credit card expenditure?

Heckman's standard sample selection model is also called "Tobit-2" model (Amemiya 1984, Amemiya 1985). It consists of the following (unobserved) structural process:

$$y_i^{S*} = \vec{\beta}^{S\prime} \vec{x}_i^S + \varepsilon_i^S \tag{9}$$
$$y_i^{O*} = \vec{\beta}^{O\prime} \vec{x}_i^O + \varepsilon_i^O, \tag{10}$$

where y_i^{S*} is the realisation of the the latent value of the selection "tendency" for the individual i, and y_i^{O*} is the latent outcome. $\vec{x}_i^S$ and $\vec{x}_i^O$ are explanatory variables for the selection and outcome equation, respectively. $\vec{x}^S$ and $\vec{x}^O$ may or may not be equal. We observe

$$y_i^S = \begin{cases} 0 & \text{if } y_i^{S*} < 0 \\ 1 & \text{otherwise} \end{cases} \tag{11}$$

$$y_i^O = \begin{cases} 0 & \text{if } y_i^S = 0 \\ y_i^{O*} & \text{otherwise,} \end{cases} \tag{12}$$

i.e. we observe the outcome only if the latent selection variable y^{S*} is positive. The observed dependence between y^O and x^O can now be written as

$$E[y^O|\vec{x}^O = \vec{x}_i^O, \vec{x}^S = \vec{x}_i^S, y^S = 1] = \vec{\beta}^{O\prime}\vec{x}_i^O + E[\varepsilon^O|\varepsilon^S \geq -\vec{\beta}^{S\prime}\vec{x}_i^S]. \tag{13}$$

Estimating the model above by OLS gives in general biased results, as $E[\varepsilon^O|\varepsilon^S \geq -\vec{\beta}^{S\prime}\vec{x}_i^S] \neq 0$, unless ε^O and ε^S are mean independent (in this case $\varrho = 0$).

Assuming the error terms follow a bivariate normal distribution:

$$\begin{pmatrix} \varepsilon^S \\ \varepsilon^O \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \varrho \\ \varrho & \sigma^2 \end{pmatrix} \right), \tag{14}$$

we may employ the following simple strategy: find the expectations $E[\varepsilon^O|\varepsilon^S \geq -\vec{\beta}^{S\prime}\vec{x}_i^S]$, also called the *control function*, by estimating the selection equations and by probit, and thereafter insert these expectations as additional covariates (see Greene 2002 for details). Accordingly, we may write:

$$y_i^O = \vec{\beta}^{O\prime}\vec{x}_i^O + E[\varepsilon^O|\varepsilon^S \geq -\vec{\beta}^{S\prime}\vec{x}_i^S] + \eta_i \equiv \vec{\beta}^{O\prime}\vec{x}_i^O + \varrho\sigma\lambda(\vec{\beta}^{S\prime}\vec{x}_i^S) + \eta_i \tag{15}$$

where $\lambda(\cdot) = \phi(\cdot)/\Phi(\cdot)$ is commonly referred to as inverse Mill's ratio, $\phi(\cdot)$ and $\Phi(\cdot)$ are standard normal density and cumulative distribution functions and η is a new disturbance term, independent of $\vec{x}^O$ and $\vec{x}^S$. The unknown multiplicator $\varrho\sigma$ can be estimated by OLS ($\hat{\beta}^\lambda$). Essentially, we describe the selection problem as an omitted variable problem, with $\lambda(\cdot)$ as the omitted variable. Since the true $\lambda(\cdot)$s are generally unknown, they are replaced by estimated values based on the probit estimation in the first step.

The relations also reveal the interpretation of ϱ. If $\varrho > 0$, the third term in the right hand side is positive as the observable observations tend to have above average realizations of ε^O. This is usually referred to as positive selection in a sense that the observed outcomes are better than the average. In this case, the OLS estimates are upward biased.

An estimator of the variance of ε^O can be obtained by

$$\hat{\sigma}^2 = \frac{\hat{\vec{\eta}}'\hat{\vec{\eta}}}{n^O} + \frac{\sum_i \hat{\delta}_i}{n^O}\,\hat{\beta}^{\lambda\,2} \tag{16}$$

where $\hat{\vec{\eta}}$ is the vector of residuals from the OLS estimation, n^O is the number of observations in this estimation, and $\hat{\delta}_i = \hat{\lambda}_i(\hat{\lambda}_i + \hat{\vec{\beta}}^{S\prime}\vec{x}_i^S)$. Finally, an estimator of the correlation between ε^S and ε^O can be obtained by $\hat{\varrho} = \hat{\beta}^\lambda/\hat{\sigma}$. Note that $\hat{\varrho}$ can be outside of the $[-1, 1]$ interval.

Since the estimation is not based on the true but on estimated values of $\lambda(\cdot)$, the standard OLS formula for the coefficient variance-covariance matrix is not appropriate [p.~157]{heckman79}. A consistent estimate of the variance-covariance matrix can be obtained by

$$\widehat{VAR}\left[\hat{\vec{\beta}}^O, \hat{\vec{\beta}}^\lambda\right] = \hat{\sigma}^2 \left[\mathbf{X}_\lambda^{\mathbf{O}\prime}\mathbf{X}_\lambda^{\mathbf{O}}\right]^{-1} \left[\mathbf{X}_\lambda^{\mathbf{O}\prime}\left(\mathbf{I} - \hat{\boldsymbol{\varrho}}^{\mathbf{2}}\hat{\boldsymbol{\Delta}}\right)\mathbf{X}_\lambda^{\mathbf{O}} + \mathbf{Q}\right] \left[\mathbf{X}_\lambda^{\mathbf{O}\prime}\mathbf{X}_\lambda^{\mathbf{O}}\right]^{-1} \tag{17}$$

where

$$\mathbf{Q} = \hat{\varrho}^{\mathbf{2}} \left(\mathbf{X}_{\lambda}^{\mathbf{O}\prime} \hat{\mathbf{\Delta}} \mathbf{X}^{\mathbf{S}} \right) \widehat{\mathbf{VAR}} \left[\hat{\vec{\beta}}^{\mathbf{S}} \right] \left(\mathbf{X}^{\mathbf{S}\prime} \hat{\mathbf{\Delta}} \mathbf{X}_{\lambda}^{\mathbf{O}} \right), \tag{18}$$

$\mathbf{X^S}$ is the matrix of all observations of $\vec{x}^S$, $\mathbf{X_\lambda^O}$ is the matrix of all observations of $\vec{x}^O$ and $\hat{\lambda}$, $\mathbf{I}$ is an identity matrix, $\hat{\mathbf{\Delta}}$ is a diagonal matrix with all $\hat{\delta}_i$ on its diagonal, and $\widehat{VAR}\left[\hat{\vec{\beta}}^S\right]$ is the estimated variance covariance matrix of the probit estimate (Greene 1981, Greene 2002).

This is the original idea by (Heckman 1976). As the model is fully parametric, it is straightforward to construct a more efficient maximum likelihood (ML) estimator. Using the properties of a bivariate normal distribution, it is easy to show that the log-likelihood can be written as

$$L = \sum_{\{i:y_i^S=0\}} \log \Phi(-\vec{\beta}^{S\prime}\vec{x}_i^S) + \tag{19}$$

$$+ \sum_{\{i:y_i^S=1\}} \left[\log \Phi \left(\frac{\vec{\beta}^{S\prime}\vec{x}_i^S + \dfrac{\varrho}{\sigma}(y_i^O - \vec{\beta}^{O\prime}\vec{x}_i^O)}{\sqrt{1-\varrho^2}} \right) - \frac{1}{2}\log 2\pi - \log\sigma - \frac{1}{2}\frac{(y_i^O - \vec{\beta}^{O\prime}\vec{x}_i^O)^2}{\sigma^2} \right]. \tag{20}$$

The original article suggests using the two-step solution for exploratory work and as initial values for ML estimation. This was a result of the high costs of estimation. Nowadays, costs are no longer an issue, however, the two-step solution allows certain generalisations more easily than ML, and is more robust in certain circumstances.

This model and its derivations were introduced in the 1970s and 1980s. The model is well identified if the exclusion restriction is fulfilled, i.e. if $\vec{x}^S$ includes a component with a substantial explanatory power but which is not present in $\vec{x}^O$. This means essentially that we have a valid instrument. If this is not the case, the identification is related to the non-linearity of the inverse Mill's ratio $\lambda(\cdot)$. The exact form of it stems from the distributional assumptions. During the recent decades, various semiparametric estimation techniques have been increasingly used in addition to the Heckman model.

- Having run the Heckman two-step estimation procedure and calculated the marginal effect of income on credi card spending we see that a 10% increase in income is estimated to increase credit card spending by 11.240879%.