MECT Microeconometrics
Blundell Lecture 3
Selection Models

Richard Blundell
http://www.ucl.ac.uk/~uctp39a/

University College London

February-March 2016

The Selectivity Model

Generalises the censored regression model by specifying mixture of discrete and continuous processes.

- Extends the 'corner solution' model to cover models with fixed costs.
- Extends to cover the case of the heterogeneous treatment effect models.

Write the latent process for the variable of interest as

$$y_{1i}^* = x_{1i}'\beta_1 + u_{1i}$$

with $E(u_1|x_1) = 0$. The observation rule for y_1 is given by

$$y_{1i} = \begin{cases} y_{1i}^* & \text{if } y_{2i}^* > 0 \\ 0 & \text{otherwise} \end{cases}$$

where

$$y_{2i}^* = x_{2i}'\beta_2 + u_{2i}$$

and

$$y_{2i} = \begin{cases} 1 & \text{if } y_{2i}^* > 0 \\ 0 & \text{otherwise} \end{cases}$$

as in the Probit model.

Consider the selected sample with $y_{2i}^* > 0$, OLS is biased as we know

$$\begin{aligned} E(u_{1i}|y_{2i}^* &> 0) = E(u_{1i}|x_{2i}'\beta_2 + u_{2i}) \\ &= E(u_{1i}|u_{2i} > -x_{2i}'\beta_2) \\ &\neq 0, \text{ if } u_1 \text{ and } u_2 \text{ are correlated.} \end{aligned}$$

- Suppose to begin with we assume (u_1, u_2) are jointly normally distributed with mean zero and constant covariance matrix,

$$\begin{pmatrix} u_1 \\ u_2 \end{pmatrix} \backsim \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & 1 \end{pmatrix}\right).$$

- We can write the orthogonal decomposition of u_1 given u_2 as

$$u_{1i} = \sigma_{12} u_{2i} + \varepsilon_{1i}$$

where ε_1 is distributed independently of u_2 and has a marginal normal distribution.

Substituting we have

$$\begin{aligned} E(u_{1i}|y_{2i}^* &> 0) = E(\sigma_{12}u_{2i} + \varepsilon_{1i}|u_{2i} > -x_{2i}'\beta_2) \\ &= \sigma_{12}E(u_{2i}|u_{2i} > -x_{2i}'\beta_2) + E(\varepsilon_{1i}|u_{2i} > -x_{2i}'\beta_2) \\ &= \sigma_{12}E(u_{2i}|u_{2i} > -x_{2i}'\beta_2) \end{aligned}$$

- From last lecture we have the conditional mean for the truncated normal

$$\begin{aligned} E(w|w > c) &= \int_c^\infty wf(w|w > c)dw \\ &= \frac{\sigma}{1 - \Phi\left(\frac{c}{\sigma}\right)}\left[-\phi\left(\frac{w}{\sigma}\right)\right]_c^\infty = \sigma\frac{\phi\left(\frac{c}{\sigma}\right)}{1 - \Phi\left(\frac{c}{\sigma}\right)} \end{aligned}$$

Noting that $\sigma_{22} \equiv 1$, we have

$$\begin{aligned} E(u_{1i}|y_{2i}^* &> 0) = \sigma_{12} E(u_{2i}|u_{2i} > -x_{2i}'\beta_2) \\ &= \sigma_{12} \frac{\phi\left(-x_{2i}'\beta_2\right)}{1 - \Phi\left(-x_{2i}'\beta_2\right)} \\ &= \sigma_{12} \frac{\phi\left(x_{2i}'\beta_2\right)}{\Phi\left(x_{2i}'\beta_2\right)} \\ &= \sigma_{12}\lambda\left(x_{2i}'\beta_2\right). \end{aligned}$$

- In general provided we have this linear index specification

$$E(u_{1i}|y_{2i}^* > 0) = g\left(x_{2i}'\beta_2\right).$$

- Implying that selection is simply a function of the single index in the selection equation $x_{2i}'\beta_2$, even when joint normality can not be assumed. However, note the restrictiveness of the single linear index specification.

▶ Given this result for the joint normal linear index selection model we can easily derive the familiar Heckman and Maximum Likelihood estimators. The selection model can now be rewritten:

$$y_{1i}^* = x_{1i}'\beta_1 + \sigma_{12}\lambda\left(x_{2i}'\beta_2\right) + \varepsilon_{1i}$$

with $E(\varepsilon_1|x_1, x_2) = 0$ and $E(\varepsilon_1^2|x_1, x_2) = \omega_{11}$.
The observation rule for y_1 is given by

$$y_{1i} = \begin{cases} y_{1i}^* & \text{if } y_{2i}^* > 0 \\ 0 & \text{otherwise} \end{cases}$$

where

$$y_{2i}^* = x_{2i}'\beta_2 + u_{2i}$$

as before.

▶ We can write the log-likelihood to mirror this conditional specification as and the loglikelihood contribution for observation i is

$$\ln l_i(\beta_1, \beta_2, \omega_{11}, \sigma_{12}) = \left\{ \begin{array}{c} D_i \ln \left(\frac{1}{\sqrt{2\pi\omega_{11}}} \exp \left(-\frac{(y_{1i} - x_{1i}'\beta_1 - \sigma_{12}\lambda(x_{2i}'\beta_2))^2}{2\omega_{11}} \right) \right) + \\ D_i \ln \Phi \left(x_{2i}'\beta_2 \right) + (1 - D_i) \ln \left[1 - \Phi \left(x_{2i}'\beta_2 \right) \right] \end{array} \right\}$$

$$\ln \mathcal{L}_N(\beta_1, \beta_2, \omega_{11}, \sigma_{12}) = \sum_{i=1}^{N} \left\{ \begin{array}{c} D_i \ln \left(-\frac{(y_{1i} - x_{1i}'\beta_1 - \sigma_{12}\lambda(x_{2i}'\beta_2))^2}{\omega_{11}} \right) + \\ D_i \ln \Phi \left(x_{2i}'\beta_2 \right) + (1 - D_i) \ln \left[1 - \Phi \left(x_{2i}'\beta_2 \right) \right] \end{array} \right.$$

- Notice that $\beta_1, \omega_{11}, \sigma_{12}$ do not occur in the second part of this expression so there is a natural partition of the loglikelihood into the binary model for selection that estimates β_2 and the conditional model on the selected sample.
- Thus we have the Heckman selectivity estimator or Heckit.

- The Heckit estimator is the first round of a full MLE estimation which produces consistent but not fully efficient estimators.
  - First estimate β_2 by Probit.
  - Then, condition on β_2, estimate $\beta_1, \omega_{11}, \sigma_{12}$ from the least squares estimation of the conditional model on the selected sample.
  - Can clearly go on to produce the MLE estimators. Stata allows either option.
- Note that the LM or Score test can be constructed directly by including $\lambda\left(x_{2i}'\beta_2\right)$ in the selected regression and testing the coefficient.
  - This is a one degree of freedom score test so that a t-test can be used.

- **Advantages of the Normal Selection Model:**

(i) avoids the Tobit assumption.
(ii) 2-step Heckit estimator is straightforward.
(iii) t-test of the null hypothesis $H_0 : \sigma_{12} = 0$, i.e. no selectivity bias, can be constructed easily.

- **Disadvantages:**

(i) assumes joint normality
(ii) need to allow for the estimated β_2 in $\lambda\left(x'_{2i}\beta_2\right)$. Typically easiest to compute full MLE and use the usual formula for correct standard errors. Note that the t-test of selectivity bias can be carried out without this extra computation because the test statistic is valid under the null hypothesis H_0.
(iii) need $\lambda\left(x'_{2i}\beta_2\right)$ to vary independently of $x'_{1i}\beta_1$.

- The requirement that $\lambda\left(x_{2i}'\beta_2\right)$ varies independently of $x_{1i}'\beta_1$ is strictly one of nonparametric identification since, in the parametric joint normal case for example, λ is a nonlinear function given by $\frac{\phi(x_{2i}'\beta_2)}{\Phi(x_{2i}'\beta_2)}$ and is **not** perfectly collinear with $x_{1i}'\beta_1$ even if exactly the same variables are in x_1 and x_2.
- However, even in the joint normal case $\frac{\phi(x_{2i}'\beta_2)}{\Phi(x_{2i}'\beta_2)}$ can be approximately linear over large ranges of $x_{2i}'\beta_2$. In general, identification requires an exclusion restriction just as in the standard endogenous regressor case.
  - This is really a triangular structure for a simultaneous model.
  - D_i is a single endogenous variable in the structural model for y_1.
  - The order condition requires that at least one exogenous variable is excluded for each included rhs endogenous variable.

When we are unwilling to assume a parametric distribution for u_1 and u_2 then the identification arguments becomes even more clear.

- As we noted above, given the linear index structure, the selection model can still be written:

$$y_{1i} = x_{1i}'\beta_1 + g\left(x_{2i}'\beta_2\right) + \varepsilon_{1i}$$

for y_{1i} observed and with $E(\varepsilon_1|x_1, x_2) = 0$ and (maybe) $E(\varepsilon_1^2|x_1, x_2) = \omega_{11}$.

- But if we do not know the form of g, perfect collinearity can occur if there is no exclusion restriction.

Indeed, in general we will need to exclude a continuous 'instrumental' variable.

- Often this lines up well with the economic problem being addressed.
  - For example, wages and employment. In this case the excluded instrument is nonlabour income. This determines employment but not wages, at least in the static competitive model.

Think of other cases: prices firms set across different markets, the instrument maybe local costs; occupational choice and earnings?

- Notice the Tobit structure did not need such an exclusion restriction even when nonlinearity was relaxed.
- Does selection matter? Empirical examples include Blundell, Reed and Stoker, AER 2003.
- Try the Mroz data?
- Does relaxing joint normality matter? Some evidence it does......see the Newey, Powell and Walker AER (1990) and references therein. But need relatively large sample sizes to provide precision in semiparametric extensions.

Semiparametric Methods:

$$y_{1i} = x_{1i}'\beta_1 + g\left(x_{2i}'\beta_2\right) + \varepsilon_{1i}$$

for y_{1i} observed.

- two-step methods (analogous to the Heckit estimator)
- Quasi-maximum likelihood estimators (analogous to Klien-Spady)

(i) Two-Step methods?

1. Estimate β_2, say by maximum score.
2. Estimate β_1, given $\widehat{\beta}_2$.

At the second stage there are also a number of possibilities. One attractive approach is simply to use a **series approximation** to $g\left(x_{2i}'\beta_2\right)$

$$y_{1i} = x_{1i}'\beta_1 + \sum_{j=1}^{J} \eta_j \rho_j \left(x_{2i}'\widehat{\beta}_2\right) + \epsilon_{1i}$$

where

$$\rho_j\left(x_{2i}'\widehat{\beta}_2\right) = \lambda\left(x_{2i}'\beta_2\right).\left(x_{2i}'\beta_2\right)^{j-1}$$

e.g. for $J = 3$, estimate on the selected sample only:

$$\begin{aligned} y_{1i} &= x_{1i}'\beta_1 + \eta_1\lambda\left(x_{2i}'\widehat{\beta}_2\right) + \eta_2\lambda\left(x_{2i}'\widehat{\beta}_2\right).x_{2i}'\widehat{\beta}_2 \\ &\quad + \eta_2\lambda\left(x_{2i}'\widehat{\beta}_2\right).\left(x_{2i}'\widehat{\beta}_2\right)^2 + \epsilon_{1i}. \end{aligned}$$

Semiparametric Methods:

An alternative is to use **Kernel regression**.
Note that for the selected observations we have a partially (or semi) linear structure:

$$y_{1i} = \beta_1' x_{1i} + g\left(x_{2i}'\beta_2\right) + \varepsilon_{1i}$$

so that

$$E(y_{1i}|x_{2i}'\beta_2) = \beta_1' E(x_{1i}|x_{2i}'\beta_2) + g\left(x_{2i}'\beta_2\right)$$

now subtract the latter expression from the former

$$y_{1i} - E(y_{1i}|x_{2i}'\beta_2) = \beta_1'(x_{1i} - E(x_{1i}|x_{2i}'\beta_2)) + \varepsilon_{1i}$$

which no longer depends on g at all!

Semiparametric Methods:

Suggests an estimator.
Starting with:

$$y_{1i} - E(y_{1i}|x_{2i}'\beta_2) = \beta_1'(x_{1i} - E(x_{1i}|x_{2i}'\beta_2)) + \varepsilon_{1i}$$

- Replace $E(y_{1i}|x_{2i}'\beta_2)$ and $E(x_{1i}|x_{2i}'\beta_2)$ by their Kernel regression counterparts, then estimate β_1. Note that x_2 **must** contain some excluded continuous instrument otherwise $x_{1i} - E(x_{1i}|x_{2i}'\beta_2)$ will be null.
- Newey, Powell and Walker (1990) show that $\sqrt{N}(\widehat{\beta}_1 - \beta_1) \sim^a N(0, \Omega)$.
- They present some results for the Mroz data.

Semiparametric Methods:

Ahn and Powell (1993) present another similar and very intuitive 'differencing' or 'matching' style estimator.
They note that

$$y_{1i} = x_{1i}'\beta_1 + g\left(x_{2i}'\beta_2\right) + \varepsilon_{1i}$$

and consider two observations i and j with $x_{2i}'\beta_2$ 'close':

$$y_{1i} - y_{1j} = \left(x_{1i} - x_{1j}\right)'\beta_1 + g\left(x_{2i}'\beta_2\right) - g\left(x_{2j}'\beta_2\right) + \varepsilon_{1i} - \varepsilon_{1j}$$

or

$$y_{1i} - y_{1j} = \left(x_{1i} - x_{1j}\right)'\beta_1 + \left(g_i - g_j\right) + \varepsilon_{1ij}$$

they suggest finding j observations as close to i as is possible and then eliminate g by regression. They use a Kernel estimator to define observations that are 'close'.

Semiparametric Methods:

Key structure for this estimator is:

$$y_{1i} - y_{1j} = (x_{1i} - x_{1j})' \beta_1 + g\left(x_{2i}'\beta_2\right) - g\left(x_{2j}'\beta_2\right) + \varepsilon_{1i} - \varepsilon_{1j}$$

- Note that we can effectively use $x_{2i}'\beta_2$ in place of g_i or any other monotonic function of $x_{2i}'\beta_2$.
- Note also that there is no requirement to have a single(linear) index for the selection rule. Could replace this purely with a 'propensity score'. That is some selection or assignment equation as a general function of the x_2 variables.

Alternative Bivariate Models for Selected Samples

1. Double-Hurdle models
2. Infrequency of purchase models