Econometrics

Binary Choice Models

Paul P. Momtaz

The Anderson School
UCLA

Outline

Linear Probability Model (LPM)

The Model

$$Pi = Pr[y_i = 1|x_i] = x_i'\beta \qquad \text{(by OLS)}$$

since $x_i'\beta = \mathbb{E}[y_i|x_i] = 1 \cdot P(y_i = 1|x) + 0 \cdot P(y_i = 0|x_i) = P(y_i = 1|x_i)$

Linear Probability Model (LPM)

Problems with the LPM

Problems with the LPM:

- Pi can be < 0 or > 1.
- Error distribution not normal $\varepsilon_i = -x_i'\beta$ or $\varepsilon_i = 1 = x_i'\beta$
- ε_i heteroskedastic

$$\begin{aligned}\mathbb{E}[\varepsilon_i|x_i] &= P(\varepsilon_i = 1 - x_i'\beta|x_i)(1 - x_i'\beta|x_i)(-x_i'\beta) \\ &= Pi(1 - x_i'\beta) + (1 - Pi)(-x_i'\beta) = 0\end{aligned}$$

$$\begin{aligned}Var[\varepsilon_i|x_i] &= P(\varepsilon_i = 1 - x_i'\beta|x_i)(1 - x_i'\beta)^2 + P(\varepsilon_i = -x_i'\beta|x_i)(-x_i'\beta)^2 \\ &= x_i'\beta(1 - x_i'\beta)^2 + (1 - x_i'\beta)(-x_i'\beta)^2 \\ &= x_i\beta(1 - x_i'\beta)\end{aligned}$$

$Var[\varepsilon_i|x_i]$ depends on x_i

Non-linear Transformations

Probit

Probit (If CDF Standard Normal):

$$P(y_i = 1|x_i) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x_i'\gamma} \exp(-u^2/2)du = \Phi(x_i'\gamma) = F(x_i'\gamma)$$

Non-linear Transformations
Logit

Logit (If CDF Logistic):

$$P(y_i = 1|x_i) = \frac{e^{x_i'\gamma}}{1 + e^{x_i'\gamma}} = \frac{1}{1 + e^{-x_i'\gamma}} = \bigwedge(x_i'\gamma) = F(x_i'\gamma)$$

Note:

$$F(z) = \bigwedge(z) = \frac{e^z}{1 - e^z} = \Delta f(z) = \bigwedge'(z) = \frac{e^z}{1 + e^z} \frac{1}{1 + e^z} = \bigwedge(z)[1 - \bigwedge($$

Latent Variable Threshold Model

Introduction

$$y_i^* = x_i'\beta + \varepsilon_i \qquad y_i^* \text{ unobserved} = \text{latent variable}$$

$$y_i = \begin{cases} 1 & y_i^* > \lambda \\ 0 & y_i^* \leq \lambda \end{cases} \qquad \lambda \text{ a threshold}$$

Identification Problems:

$$y_i^* > \lambda \leftrightarrow x_i'\beta + \varepsilon_i > \lambda \leftrightarrow (x_i'\beta - \lambda) + \varepsilon_i > 0 \Rightarrow \text{set } \lambda = 0$$

$$P(y_i = 1|x_i) = Q(\varepsilon_i \leq x_i'\beta|x_i) = P\left(\frac{\varepsilon_i}{\sigma} \leq \frac{x_i'\beta}{\sigma}|x_i\right) = P\left(\frac{\varepsilon_i}{\sigma} \leq x_i'\beta^*|x_i\right)$$

So, β identified up to a scale factor $\Rightarrow$ Normalize ε_i distribution, assume σ^2 known

Latent Variable Threshold Model

Estimation: MLE

$$P(y_i = 1|x_i) = F(x_i'\beta), \qquad P(y_i = 0|x_i) = 1 - F(x_i'\beta)$$

$$L(\beta) = \prod_{i=1}^{n} F(x_i'\beta)^{y_i}[1 - F(x_i'\beta)]^{1-y_i}$$

$$\log(\beta) = \sum_{i=1}^{n} y_i \log(F(x_i'\beta)) + (1 - y_i)\log(1 - F(x_i'\beta))$$

Score Function:

$$\frac{\partial \log L(\beta)}{\partial \beta} = \sum_{i=1}^{n} \left[\frac{y_i f_i}{F_i} - \frac{(1-y_i) f_i}{1 - F_i}\right] x_i = \sum_{i=1}^{n} \underbrace{\left[\frac{y_i - F_i}{F_i(1-F_i)} f_i\right]}_{\text{generalized residual}} x_i = 0$$

Logit Model

$$\frac{\partial \log L(\beta)}{\partial \beta} = \sum_{i=1}^{n}(y_i - \bigwedge i)x_i = 0, \qquad \text{since}$$

$$f(z) = \bigwedge(z)(1 - \bigwedge(z)) \text{ cancels out}$$

Remember:

$$\bigwedge i = \bigwedge(x_i'\beta) = (1 + e^{-x_i'\beta})^{-1} \qquad \text{so} \qquad \frac{\partial \log L(\beta)}{\partial \beta} = \sum_{i=1}^{n}[y_i - \mathbb{E}[y_i|x_i]]x_i =$$

Hessian Matrix:

$$\frac{\partial^2 \log L(\beta)}{\partial \beta \partial \beta'} = -\sum_{i=1}^{n} \bigwedge i(1 - \bigwedge i)x_i x_i' \Rightarrow \text{ globally concave, } \hat{\beta} \text{ unique}$$

Interpretation: Marginal Effects

Interpretation β in linear model ($y_i = x_i'\beta + \varepsilon_i$) : $\beta = \frac{\partial y_i}{\partial x_i}$

Interpretation in BCM ($\rho_i = F(x_i'\beta)$):

$$\frac{\partial \rho[y_i = 1|x_i]}{\partial x_i} = f(x_i'\beta)\beta \Rightarrow \text{Marginal effect}$$

Marginal Effects $\neq \beta$ (but sign is the same)

Logit: $$\frac{\partial \rho(y_i = 1|x_i)}{\partial x_i} = \bigwedge(x_i'\beta)[1 - \bigwedge(x_i'\beta)]\beta = \rho_i(1-\rho_i)\beta$$

Probit: $$\frac{\partial \rho(y_i = 1|x_i)}{\partial x_i} = \phi(x_i'\beta)\beta = \phi(\Phi^{-1}(\rho_i))\beta$$

Marginal effects often evaluated at $\rho_i = \overline{\rho}$ or $x_i = \overline{x}$

Generalized Residuals

$$e_i^G = \frac{y_i - F_i}{F_i(1 - F_i)} f_i$$

$$\text{ML Requires: } \sum_{i=1}^{n} \hat{e}_i^G x_i = 0$$

$$\hat{e}_i^G = \begin{cases} f_i/F_i & y_i = 1 \\ -f_i/(1 - F_i) & y_i = 0 \end{cases}$$

$$\text{Logit: } \hat{e}_i^G = y_i - \bigwedge(x_i'\beta) = y_i - \hat{\rho}_i$$

$$\text{Probit: } \hat{e}_i^G = \begin{cases} \frac{\phi(x_i'\hat{\beta})}{\Phi(x_i'\hat{\beta})} & y_i = 1 \\ \frac{-\phi(x_i'\hat{\beta})}{1-\Phi(x_i'\hat{\beta})} & y_i = 0 \end{cases}$$

LM Tests for BCM

LM Tests for BCM:

- Outer-product gradient (OPG) form of LM test.
- LM test for omitted regressors.
- LM test for heteroskedasticity

Tobit Model

Introduction

Truncated moments:

$$z \sim N(0,1), \qquad \phi(z) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{z^2}{2}\right) \qquad \Phi(z) = \int_{-\infty}^{z} \phi(u)du$$

$$\begin{aligned}
\mathbb{E}[z|z<k] &= \int_{-\infty}^{k} z\phi(z|z<k)dz \\
&= \int_{-\infty}^{k} z\phi(z)/\rho(z<k)dz \\
&= \frac{1}{\Phi(k)} \int_{-\infty}^{k} z\phi(z)dz \\
&= \frac{\phi(k)}{\Phi(k)} < 0, \qquad \text{where } \int_{-\infty}^{k} z\phi(z)dz = -\phi(k)
\end{aligned}$$

$$\mathbb{E}[z|z>k] = \frac{\phi(k)}{1-\Phi(k)} > 0$$

Tobit Model

Example and Inverse Mills Ratio

Example: Latent variable model $y_i^* = x_i'\beta + \varepsilon_i$, $\varepsilon_i \sim N(0, \sigma^2)$

$$\mathbb{E}[\varepsilon_i \varepsilon_i > -x_i'\beta, x_i] = \sigma \mathbb{E}\left[\frac{\varepsilon_i}{\sigma} | \frac{\varepsilon_i}{\sigma} > -\frac{x_i'\beta}{\sigma}, x_i\right] = \sigma \frac{\phi\left(\frac{x_i'\beta}{\sigma}\right)}{\Phi\left(\frac{x_i'\beta}{\sigma}\right)} = \sigma\lambda_i$$

λ_i is inverse Mills Ratio

Sample Tobit Model:

$$y_i^* = x_i'\beta + \underbrace{\varepsilon_i \sim N(0, \sigma^2)}_{\text{Tobit Assumption}} \qquad y_i = \begin{cases} y_i^* = x_i'\beta + \varepsilon_i & \text{if } y_i^* > 0 \\ 0 & \text{if } y_i^* = 0 \end{cases}$$

Tobit Model

OLS Estimator Bias

OLS Estimator Bias:

$$\mathbb{E}[y_i|\text{interior}] = \mathbb{E}[y_i|y_i^* > 0, x_i] = x_i'\beta + \mathbb{E}[\varepsilon_i|\varepsilon_i > -x_i'\beta, x_i]$$

$$\mathbb{E}[y_i|\text{interior} + \text{corner} = x_i] = \mathbb{E}[y_i|y_i > 0, x_i] \cdot P[y_i > 0|x_i] + P(y_i = 0|x_i) = $$
$$= P(y_i > 0, x_i) \cdot [x_i'\beta + \mathbb{E}[\varepsilon_i|\varepsilon_i > -x_i'\beta, x_i]]$$

$$\mathbb{E}[y_i|x_i] \neq x_i'\beta \neq \mathbb{E}[y_i|\text{interior}]$$

Correcting OLS estimator bias (when $y_i > 0$, i.e. interior solutions)

$$\mathbb{E}[y_i|y_i > 0, x_i] = x_i'\beta + \mathbb{E}[\varepsilon_i|\varepsilon_i > -x_i'\beta, x_i]$$
$$= x_i'\beta + \sigma \frac{\phi\left(\frac{x_i'\beta}{\sigma}\right)}{\Phi\left(\frac{x_i'\beta}{\sigma}\right)}$$

Outline

Two-Step Heckman Sample Selection Bias

The Idea

Define

$$D_i = \begin{cases} 1 & \text{if } y_i > 0 \text{ (interior solution)} \\ 0 & \text{if } y_i = 0 \text{ (corner solution)} \end{cases}$$

$$P(D_i = 1|x_i) = P(y_i > 0|x_i) = P(y_i^* > 0|x_i) = P(\varepsilon_i \leq x_i'\beta|x_i) = \Phi\left(\frac{x_i'\beta}{\sigma}\right)$$

Step 1: Estimate β/σ using Probit for $P(D_i = 1|x_i)$ on full sample and construct $\hat{\lambda}_i$ for each observation of interior solution

Step 2: OLS regression of y_i on x_i and $\hat{\lambda}_i$ using interior cases

Two-Step Heckman Sample Selection Bias

Drawbacks

Drawbacks of this procedure:

- OLS s.e. in the step 2 wrong
- Identification only through fact that λ_i non-linear
  - Problematic if λ_i little variation and close to linear in x_i
- Monte Carlo shows additional variable in step 1 often relevant for identification in step 2, but not available.

Outline

ML Estimation of Tobit Model

Intuition

$$\log L = \sum_{y_i=0} \log P(y_i = 0|x_i) + \sum_{y_i>0} \log(f(y_i|y_i > 0, x_i)P(y_i > 0|x_i))$$
$$= \sum_{y_i=0} \log P(y_i = 0|x_i) + \sum_{y_i>0} \log(f(y_i|x_i))$$

Intuition:

- For $y_i = 0$: likelihood contribution given by having proba mass $P(y_i = 0|x_i)$
- For $y_i > 0$: likelihood contribution given by conditional clustering given $y_i > 0$, $f(y_i|y_i > 0|x_i)$ times proba mass $P(y_i > 0|x_i)$

ML Estimation of Tobit Model

Mechanics

For $y_i = 0$ $\quad P(y_i = 0|x_i) = P(x_i'\beta + \varepsilon_i \leq 0|x_i) = \Phi\left(-\frac{x_i'\beta}{\sigma}\right) = 1 - \Phi\left(\frac{x_i'\beta}{\sigma}\right)$

For $y_i > 0$ $\quad f(y_i|x_i) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2}\frac{(y_i - x_i'\beta)^2}{\sigma^2}\right) = \frac{1}{\sigma}\phi\left(\frac{y_i - x_i'\beta}{\sigma^2}\right)$

$$\log L(\beta, \sigma, y) = \sum_{y_i > 0} \log\left[\frac{1}{\sigma}\phi\left(\frac{y_i - x_i'\beta}{\sigma}\right)\right] + \sum_{y_i = 0} \log\left[1 - \Phi\left(\frac{x_i'\beta}{\sigma}\right)\right]$$

$$= \sum_{i=1}^{n} D_i \left[\log\phi\left(\frac{y_i - x_i'\beta}{\sigma}\right) - \log\sigma\right] + \sum_{i=1}^{n}(1 - D_i)\log\left[1 - \Phi\right.$$

Transform $FOC[\beta]$ to get generalized residual

$\hat{\varepsilon}_i^G = D_i\frac{y_i - x_i'\beta}{\sigma} - (1 - D_i)\hat{\lambda}_{0i}$, so far $D_i = 1, \varepsilon_i^G$ is scaled $\frac{\hat{\varepsilon}_i}{\beta}$

ML Estimation of Tobit Model

Interpretation

Interpretation of Tobit Coefficients:

- $\frac{\partial P(y_i=0|x_i)}{\partial x_i} = -\phi\left(\frac{x_i'\beta}{\sigma}\right)\frac{\beta}{\sigma}$ which is scaled version of Probit without normalized restrictions
- $\frac{\partial \mathbb{E}[y_i|x_i]}{\partial x_i} = \beta\Phi\left(\frac{x_i'\beta}{\sigma}\right)$ where sign determined by β as per Probit (total effects)

Outline

Decomposition of Tobit Model Effects

Total effects have two parts

$$\frac{\partial \mathbb{E}[y_i|x_i]}{\partial x_i} = \underbrace{P(y_i > 0|x_i)\frac{\partial \mathbb{E}[y_i|y_i > 0, x_i]}{\partial x_i}}_{\underbrace{\beta\Phi\left(\frac{x_i'\beta}{\sigma}\right) - \beta\phi\left(\frac{x_i'\beta}{\sigma}\right)\left[\frac{x_i'\beta}{\sigma} + \frac{\phi\left(\frac{x_i'\beta}{\sigma}\right)}{\Phi\left(\frac{x_i'\beta}{\sigma}\right)}\right]}_{\frac{\partial \mathbb{E}[y_i|y_i>0,x_i]}{\partial x_i} = \beta\gamma\left(\frac{x_i'\beta}{\sigma}\right)}} + \underbrace{\mathbb{E}[y_i|y_i > 0, x_i]\frac{\partial P}{\ldots}}_{\phi\left(\frac{x_i'\beta}{\sigma}\right)\left[\frac{x_i'\beta}{\sigma} + \ldots\right]}$$

Where $\gamma()$ is an adjustment factor $\in (0, 1)$

Outline

Endogenous Selection (or Tobit II) Model

Deviate from assumption that same variables x_i affecting probability of $y_i > 0$ observation also determine the level of $y_i > 0$ observation.

$$y_{1i} = \begin{cases} y_{1i}^* & \text{if } y_{2i}^* > 0 \\ \text{not obsereved} & \text{if } y_{2i}^* \leq 0 \end{cases} \qquad y_{2i}^* = x_{2i}'\beta + \varepsilon_{2i}, \qquad y_{2i}^* \text{ not observed}$$

y_{2i}^* observed if $y_{2i}^* = 0$ (observation rule) $\varepsilon_{1i}, \varepsilon_{2i} \sim$ Joint N

$$\begin{pmatrix} \varepsilon_{1i} \\ \varepsilon_{2i} \end{pmatrix} \sim N\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{pmatrix}\right] \Rightarrow \sigma_{12} \neq 0 \Rightarrow \text{endogenous selection}$$

Endogenous Selection (or Tobit II) Model

Selection Indicator: $$D_i = \begin{cases} 1 & \text{if } y_{2i}^* > 0 \\ 0 & \text{else} \end{cases}$$

Can only estimate β_2/σ_2, so set $\sigma_2 = 1$ (as in Probit)

Example: Level of wage depends on x_{1i}. But level of wage only observed for workers. Prosperity to worker depends on exogenous x_{2i}. $\sigma_{12} \neq 0$ since sample of wages growth from people that work.

Random sample assumption violated

Outline

Selection Bias

$$\mathbb{E}[\varepsilon_{2i}|y^*_{2i} > 0, x_{2i}] = \mathbb{E}[\varepsilon_{2i}|\varepsilon_{2i} > -x'_{2i}\beta_{2i}x_{2i}]$$
$$= \frac{\phi(x'_{2i}\beta)}{\Phi(x'_{2i}\beta)} = \lambda_i$$

$$\mathbb{E}[\varepsilon_{2i}|y^*_{2i} > 0, x_{2i}] = \sigma_{12}\mathbb{E}[\varepsilon_{2i}|y^*_{2i} > 0, x_{2i}] = \sigma_{12}\lambda_i$$

$$\mathbb{E}[y^*_{2i}|y^*_{2i} > 0, x'_{2i}\beta_1 + \sigma_{12}\lambda_i] \neq x'_{1i}\beta_1$$

- Endogenous selection or selectivity bias
- λ_i is equivalent to Medeman's lambda, Heckman correction, inverse Mills ratio.

Outline

Heckman's Two Step Estimator

Step 1:

Estimate Probit $P(D_i = 1|x_{2i}) = \Phi(x_{2i}'\beta_2)$ by ML to get $\hat{\beta}_2$. Construct $\hat{\lambda}_i = [\phi(x_{2i}'\hat{\beta}_2)/\Phi(x_{2i}'\hat{\beta}_2)]$ in sample where y_{2i}^* observable.

Step 2:

Run OLS $y_i = x_{2i}'\beta_1 + \sigma_{12}\hat{\lambda}_i +$ error for selected sample

Outline

Stochastic Threshold Model (Verbeck)

$$y_{1i}^* \equiv \text{ level of wage}, \qquad S_i^* = z_i'\gamma + \eta_i \equiv \text{ reservation wage}$$

$$y_{2i}^* \equiv \text{ propensity towards work}$$

$$y_{2i}^* = y_{1i}^* - S_i^* = x_{2i}'\beta_2 + \varepsilon_{2i} \qquad \text{where} \qquad \varepsilon_{2i} = \varepsilon_{1i} - \eta_i \text{ and } x_{2i}'\beta_2 = x_{1i}'\beta_1 -$$

Implication:

- $\sigma_{12} = cov(\varepsilon_{1i}, \varepsilon_{2i}) = Var(\varepsilon_{1i}) - cov(\varepsilon_{1i}, \eta_i)$
- If $cov(\eta_i, \varepsilon_{1i}) = 0 \Rightarrow \sigma_{12} > 0$
- x_{2i} contains all variables in x_{1i} plus additional from z_i identification when if linear combinations of x_{1i} and x_{2i} since λ non-linear in contrast to linear model.