Lecture 16

Discrete Choice Models
Limited Dependent Models

Lecture 16: Objectives

- Discrete Choice Models
  - Binary Choice (lecture 15)
  - Multiple Choice
    - Unordered choice sets
    - Ordered choice sets
- Count Data
- Limited Dependent Models
  - Tobit Models
  - Sample Selection

Lecture 16: Multiple Choice

- Multiple Choice:
  One decision; Many choices
  - Unordered choice sets
    - Urban transport (car, train, bus, bike)
    - Occupational choice
  - Ordered choice sets
    - Bond ratings
    - Exam grades
- Multivariate Choice: Several decisions
  - Where to live (city/suburb); type of housing (house/flat)

Unordered Choice Sets

- Choices: $j = 1, 2, \ldots, J$
  We want to model $P(Y = j)$
  Explanatory variables: $\mathbf{z}_{ij} = [\mathbf{x}_{ij} \quad \mathbf{w}_i]$
  - $\mathbf{x}_{ij}$: choice attributes
  - $\mathbf{w}_i$: individual characteristics
- Define an index function: $\theta_{ij} = \mathbf{x}'_{ij}\beta + \mathbf{w}'_i\gamma_j$
  Maximising a random utility $U_{ij} = \theta_{ij} + \varepsilon_{ij}$ across choices allows us to model $P(Y_i = j)$
- Probit models get complicated if $J > 2$

Multinomial Logit

- Most work is done using logit models

$$P(Y_i = j) = \frac{e^{\theta_{ij}}}{\sum_{j=1}^{J} e^{\theta_{ij}}} = \frac{e^{\mathbf{x}'_{ij}\beta + \mathbf{w}'_i\gamma_j}}{\sum_{j=1}^{J} e^{\mathbf{x}'_{ij}\beta + \mathbf{w}'_i\gamma_j}}$$

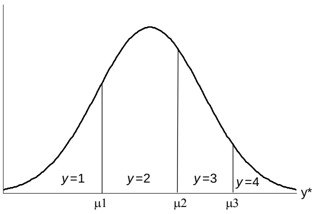
- Increasing every γ with a constant doesn't affect this probability. We normalise $\gamma_1 = 0$
- We are comparing choices with a baseline decision, in this case choice #1

Multinomial Logit, cont

- Some authors distinguish between *multinomial models* ($\beta = 0$), *conditional models* ($\gamma = 0$) and *mixed models* ($\beta,\gamma \neq 0$)
- These models can all be estimated using maximum likelihood
- Logit models imply that ratios of probabilities are independent of other choices:
  IIA (independence of irrelevant alternatives).
  Probit or nested logit models can be used

Ordered Choice Sets

- The ordered probit (or logit) uses a latent variable representation



$$y_i^* = \mathbf{x}_i'\beta + \varepsilon_i$$

$$P(Y_i = j) = P(\mu_{j-1} < y_i^* < \mu_j)$$

- $\boldsymbol{\beta}$ and $\boldsymbol{\mu}$ can be estimated using maximum likelihood

Count Data

- In some cases the numbers we assign to the "choices" mean something. That is, *y* can be interpreted as a discrete random variable.
- We can use OLS regression (normal QML)
- **COUNT DATA** models are to be preferred if *y* comes from so called "rare events" with many zeroes and small values. For example:
  - number of patents registered in a period
  - number of sick days per year

The Poisson Model

- The simplest count model assumes that no more than one event can occur in a short period of time, and that there is independence between these time periods

  ⇒ the **POISSON** model: $E(y_i \mid \mathbf{x}_i) = e^{\mathbf{x}_i'\beta}$

  where $y|\mathbf{x}$ has a Poisson distribution
- This is usually estimated by ML from

  $\ln L = \sum \left[-e^{\mathbf{x}_i'\beta} + y_i \mathbf{x}_i'\beta - \ln y_i! \right]$

Problems with Poisson models

- The Poisson distribution has variance equal to mean. In empirical studies we often observe **overdispersion**: (variance > mean)
- There are several tests for overdispersion
- Robust standard errors exist (Poisson QML)
- We can extend the model to account for heterogeneity and/or dependence. One example is the **NEGATIVE BINOMIAL** model

Censoring vs Truncation

- Difference between censoring and truncation
  - Example: 5 shots of which 4 hit the target
    - Censoring: misses are recorded as 0
    - Truncating: only shots on target are recorded
- Modeling y with the aid of $\boldsymbol{x}$

CENSORING: y_i is measured only if $y_i > \tau$
$\boldsymbol{x}_i$ is measured for all i

TRUNCATION: $\boldsymbol{x}_i$ and y_i measured only if $y_i > \tau$

Truncated Regression

- A truncated regression $y_i = \boldsymbol{x}_i'\beta + \varepsilon_i$ exists for all i but is only observed for $y_i > \tau$
- $E(y_i \mid \boldsymbol{x}_i, y_i > \tau) > E(y_i \mid \boldsymbol{x}_i) = \boldsymbol{x}_i'\beta$ since we are excluding small values of y
- It can be shown that

$$E(y_i \mid \boldsymbol{x}_i, y_i > \tau) = \boldsymbol{x}_i'\beta + \sigma\lambda\left((\tau - \boldsymbol{x}_i'\beta)/\sigma\right)$$

$$\lambda(z) = \frac{f(z)}{1 - F(z)} \quad \textbf{INVERSE MILLS RATIO}$$

where f and F are the p.d.f. and c.d.f. of ε

Truncated Regression, cont

- The truncated regression can thus be written
  $y_i = \boldsymbol{x}_i'\beta + \sigma\lambda\left((\tau - \boldsymbol{x}_i'\beta)/\sigma\right) + v_i$
  where $E(v|\boldsymbol{x})=0$
- This can be estimated using ML
- Truncated regression is rarely used directly, since we usually have observations on $\boldsymbol{x}_i$ when $y_i \le \tau$, or τ is stochastic.

Tobit Model: Censored Regression

- The model: $y_i = \begin{cases} \tau & y_i^* \le \tau \\ y_i^* & y_i^* > \tau \end{cases}$
  $y_i^* = \boldsymbol{x}_i'\beta + \varepsilon_i$ with $\varepsilon_i \sim \mathrm{N}(0, \sigma^2)$
- OLS will be inconsistent since
  $$E(y_i \mid \boldsymbol{x}_i) = \mathrm{P}(y_i \le \tau)E(y_i \mid \boldsymbol{x}_i, y_i \le \tau) + \mathrm{P}(y_i > \tau)E(y_i \mid \boldsymbol{x}_i, y_i > \tau) = \Phi(\alpha)\tau + (1 - \Phi(\alpha))(\boldsymbol{x}_i'\beta + \sigma\lambda(\alpha))$$
  where $\alpha = (\tau - \mathbf{x}_i'\beta)/\sigma$

Tobit Model; cont

- The Tobit model is estimated using ML
- The marginal effects are given by
  $\psi_j = \beta_j \Phi(\boldsymbol{x}'\beta/\sigma)$
- Problems:
  - Heteroscedasticity
  - Non-normality
  - The same model assumed for y^* both above and below the threshold

Sample Selection Models

- We have two processes.
  Selection Process: $z_i^* = \mathbf{w}_i'\gamma + v_i$
  Regression Process: $y_i = \mathbf{x}_i'\beta + \varepsilon_i \quad \text{if } z_i^* > 0$
  (ε, v) is bivariate normally distributed with $\rho = \text{Cor}(\varepsilon, v)$
- This is sometimes called Tobit type II model
- Examples
  - Advantages of moving only observed for movers
  - Purchases of seldom bought goods only observed for purchasers

Sample Selection, cont

- We can estimate with ML, but usually done with Heckman's Two-Step method (Heckit).

1) Estimate the probit model

$$\Rightarrow \hat{\lambda}_i = \frac{\phi(-\mathbf{w}_i'\hat{\gamma})}{1-\Phi(-\mathbf{w}_i'\hat{\gamma})} = \frac{\phi(\mathbf{w}_i'\hat{\gamma})}{\Phi(\mathbf{w}_i'\hat{\gamma})}$$

2) Estimate the regression $y_i = \mathbf{x}_i'\beta + (\rho\sigma_\varepsilon)\hat{\lambda}_i + u_i$

- To identify the model empirically we need some elements of $\mathbf{w}$ not to be present in $\mathbf{x}$

Sample Selection, cont

- The standard errors of the Heckit procedure can be calculated in a number of ways. We have to take account of the fact that we are estimating λ
- The estimated marginal effects are given by
  $\hat{\psi}_j = \hat{\beta}_j[1-\hat{\lambda}_i(\hat{\lambda}_i - \mathbf{w}_i'\hat{\gamma})]$