Econ 673: Microeconometrics

Chapter 4: Properties of Discrete Choice Models

Fall 2008

Outline

Readings

Required

- Train, K., (2003), *Discrete Choice Methods with Simulation*, Cambridge, MA: Cambridge University Press, Ch. 2.

Recommended

- McFadden, D. (1981), "Econometric Models of Probabilistic Choice", in C. Manski and D. McFadden (eds.) *Structural Analysis of Discrete Data with Econometric Applications*, Cambridge MA: MIT Press, pp. 198-272.

Discrete Choice Models

Discrete choice models characterize decision making in which the available alternatives are:

- Mutually exclusive
- Exhaustive
- Finite

Defining the relevant choice set is a crucial step in the analysis

Mutually Exclusive

- Not very restrictive - one can usually redefine the choice set to satisfy this restriction.
- Example #1 (Train): Home Heating Fuel Choice
  - Electric
  - Natural Gas
  - Oil
  - Wood
  - Other

  Issue: Some households have duel-fuel or multiple systems.

Possible Solutions

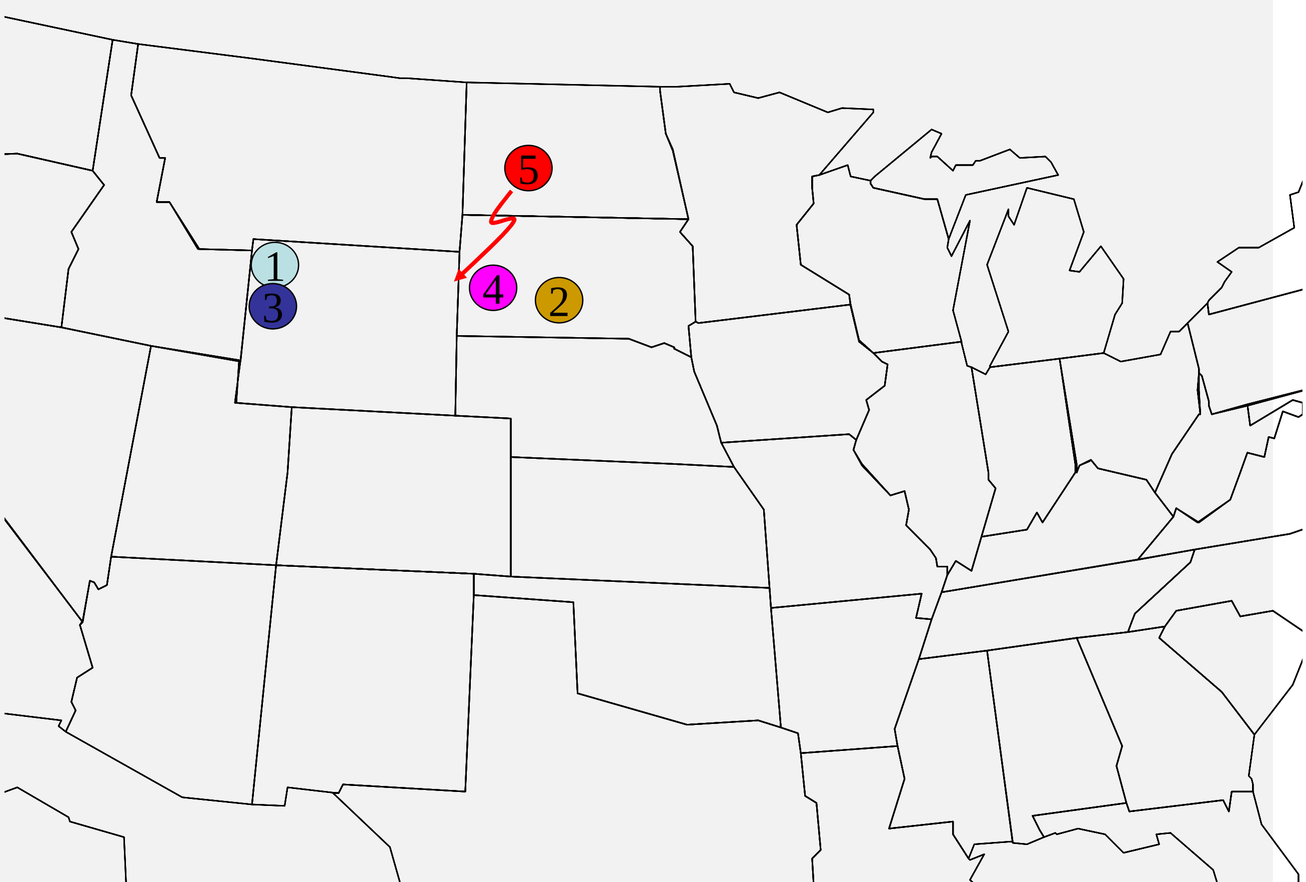
1. Explicitly include combinations
   - Electric only
   - Natural gas only
   - oil only
   - Wood only
   - Natural gas with electric room heaters, etc.
2. Focus on primary heating systems
   The potential problems here are:
   - It may be difficult to identify primary system from available data
   - Secondary heaters may be important users of energy (e.g., electric space heaters)

Example #2: Recreation Demand

Consider the following choice set for a summer vacation:

1. Yellowstone National Park (Idaho, Wyoming, South Dakota)
2. Badlands National Park (South Dakota)
3. Grand Tetons National Park (Wyoming)
4. Mount Rushmore National Memorial (South Dakota)
5. Devil's Tower National Monument (Wyoming)

For many vacationers, the alternatives in this choice set are not mutually exclusive.



Devil's Tower

Alternative Solutions

1. Redefine alternatives in terms of portfolios
   - Yellowstone only
   - Yellowstone and Grand Tetons only
   - Grand Tetons and Mount Rushmore only
   - Etc.
2. Focus on Primary Destination
   The potential problems:
   - Identifying primary destination – even in minds of recreationists
   - onstructing corresponding explanatory variables (e.g. travel costs)

Exhaustive Criteria

Readily satisfied by including a "none of the above" alternative

- Heating: "no heating"
- Recreation demand: "stay at home" or "all other sites"

Relevance

- Choice set should be defined not only to be mutually exclusive and exhaustive, but also relevant
- Choices must be ones from which agent actually chooses, not just the universe of alternatives

Example: Recreation Demand

- Should the choice set consist of
  - All feasible sites?
    - Nothing is revealed about not visiting an unknown site
    - Computationally intractable
  - Only sites visited in past five years?
    - Useful information is revealed about sites that are not visited.
  - Sites within a given distance?
    - Limits applicability of results to, say, day trips.
- Similar issues emerge in job and career search literature
- Ideally, one would model both choices and accumulation of information about choice set.

Finite or Countable Choice Sets

- This is a restrictive characteristic
  - Excludes the "how much" decision
  - Focuses on the "which" or "how many" decision
- These decisions may be tied
  - Electric power: Choice of rate structure and quantity of electricity consumed
  - Labor: Choice of participation decision and reservation wage or how much to work
  - Recreation: Choice of where to visit and how many trips to take or trip duration

Deriving Choice Probabilities

- Most discrete choice models assume individual agents make choices so as to maximize their utility – the Random Utility Maximization (RUM) hypothesis
- Encompasses a wide variety of concepts
  - Producers: profit or costs
  - Consumers: typically conditional indirect utility, but also willingness-to-pay

RUM Model

- The utility individual i receives from choosing alternative j is given by U_{ij}.
- Key: U_{ij} is assumed known to the agent, but unobserved by the analyst.
- Individual i is assumed to choose alternative j if

$$U_{ij} > U_{ik} \forall k \neq j \quad (1)$$

- The probabilistic nature of the problem is from the perspective of the analyst who does not know or observe:
  - All of the factors influencing U_{ij}
  - The functional form of U_{ij}

RUM Model (Cont'd)

The researcher specifies a functional relationship between utility and the observable characteristics of the individual and alternative:

$$U_{ij} = V(\mathbf{x}_{ij}, \mathbf{s}_i) + \epsilon_{ij} \tag{2}$$

- $\mathbf{x}_{ij}$ denotes observable characteristics of alternative j for individual i
- $\mathbf{s}_i$ denotes observable individual characteristics
- $V_{ij} = V(\mathbf{x}_{ij}, \mathbf{s}_i)$ denotes the "representative utility"
- ϵ_{ij} captures variations in preferences in the population due to unobservables.

RUM Model (Cont'd)

- The characteristics of ϵ_{ij} (i.e., its distribution) depend critically on what is or is not observable and the functional form of V_{ij}.
- Suppose that

$$U_{ij} = U(\mathbf{x}_{ij}, \mathbf{z}_{ij}, \mathbf{s}_i, \mathbf{t}_i) \tag{3}$$

where

- $\mathbf{z}_{ij}$ denotes unobservable characteristics of alternative j for individual i
- $\mathbf{t}_i$ denotes unobservable individual characteristics

Then

$$\epsilon_{ij} = U(\mathbf{x}_{ij}, \mathbf{z}_{ij}, \mathbf{s}_i, \mathbf{t}_i) - V(\mathbf{x}_{ij}, \mathbf{s}_i) \tag{4}$$

Example: Recreation Demand

Site Selection may depend upon the equipment owned by the individual (e.g., boats)

- If, say, boat ownership is unobserved, it can induce correlation across alternatives requiring a boat or across choices over time
- If boat ownership is controlled for this source correlation disappears

Measurement Error

- ϵ_{ij} may also capture measurement errors.
- Let $\tilde{\mathbf{x}}_{ij}$ denote the observed characteristics and $\mathbf{x}_{ij}$ denote its true value from the individual's perspective.
- Then

$$\begin{aligned}\epsilon_{ij} &= [U(\mathbf{x}_{ij}, \mathbf{z}_{ij}, \mathbf{s}_i, \mathbf{t}_i) - V(\mathbf{x}_{ij}, \mathbf{s}_i)] + [V(\mathbf{x}_{ij}, \mathbf{s}_i) - V(\tilde{\mathbf{x}}_{ij}, \mathbf{s}_i)] \\ &= \eta_{ij} + \tilde{\eta}_{ij}\end{aligned}$$

where
$\eta_{ij} = U(\mathbf{x}_{ij}, \mathbf{z}_{ij}, \mathbf{s}_i, \mathbf{t}_i) - V(\mathbf{x}_{ij}, \mathbf{s}_i)$ denotes true heterogeneity.
$\tilde{\eta}_{ij} = V(\mathbf{x}_{ij}, \mathbf{s}_i) - V(\tilde{\mathbf{x}}_{ij}, \mathbf{s}_i)$ arises due to measurement error.
Interpretation of errors is important in welfare analysis

Choice Probabilities

The researcher assumes a distribution for $\boldsymbol{\epsilon}_{i\cdot} = (\epsilon_{i1}, \ldots, \epsilon_{iJ})'$

$$f(\boldsymbol{\epsilon}_{i\cdot}|\mathbf{x}_{i\cdot}, \mathbf{s}_i) \tag{5}$$

From the analysts perspective, given the observables $\mathbf{x}_{i\cdot}$ and $\mathbf{s}_i$, the decision makers probability of choosing alternative j becomes:

$$\begin{aligned} P_{ij} &= Pr(U_{ij} > U_{ik} \forall k \neq j | \mathbf{x}_{i\cdot}, \mathbf{s}_i) \\ &= Pr(V_{ij} + \epsilon_{ij} > V_{ik} + \epsilon_{ik} \forall k \neq j | \mathbf{x}_{i\cdot}, \mathbf{s}_i) \\ &= Pr(\epsilon_{ik} - \epsilon_{ij} < V_{ij} - V_{ik} \forall k \neq j | \mathbf{x}_{i\cdot}, \mathbf{s}_i) \\ &= \int_{D_\epsilon} I(\epsilon_{ik} - \epsilon_{ij} < V_{ij} - V_{ik} \forall k \neq j | \mathbf{x}_{i\cdot}, \mathbf{s}_i) f(\boldsymbol{\epsilon}_{i\cdot}|\mathbf{x}_{i\cdot}, \mathbf{s}_i) \, d\boldsymbol{\epsilon}_{i\cdot} \end{aligned}$$

where D_ϵ denotes the support for the residual's distribution.

Choice Probabilities (cont'd)

- Alternative models appearing in the literature stem from choices of $f(\boldsymbol{\epsilon}_{i\cdot}|\mathbf{x}_{i\cdot}, \mathbf{s}_i)$; e.g., logit, nested logit, and probit)
- Historically, these distributions were chosen for convenience.
- Simulation methods allow for more realistic assumptions.

Alternative Specific Constants

- Alternative specific constants (α_j) are used to capture the mean of unobserved characteristics.
- Specifically, suppose we start with unobserved components capture by ϵ^*_{ij}. Then

$$\begin{aligned}\alpha_j &\equiv E\left(\epsilon^*_{ij}|\mathbf{x}_{ij},\mathbf{s}_i\right)\\ &= E\left[U(\mathbf{x}_{ij},\mathbf{z}_{ij},\mathbf{s}_i,\mathbf{t}_i)-V(\mathbf{x}_{ij},\mathbf{s}_i)|\mathbf{x}_{ij},\mathbf{s}_i\right]\end{aligned}$$

and

$$\epsilon_{ij}\equiv\epsilon^*_{ij}-\alpha_j.$$

Then for

$$U_{ij}=V(\mathbf{x}_{ij},\mathbf{s}_i)+\alpha_j+\epsilon_{ij} \quad (6)$$

we have that

$$E\left(\epsilon_{ij}|\mathbf{x}_{ij},\mathbf{s}_i\right)=0. \quad (7)$$

Identification in Choice Models

1. Only Differences in Utility Matter
   Suppose that

$$V_{ij}=V(\mathbf{x}_{ij},\mathbf{s}_i)+\alpha_j \quad (8)$$

Then

$$\begin{aligned}V_{ij}-V_{ik} &= [V(\mathbf{x}_{ij},\mathbf{s}_i)+\alpha_j]-[V(\mathbf{x}_{ik},\mathbf{s}_i)+\alpha_k]\\ &= [V(\mathbf{x}_{ij},\mathbf{s}_i)-V(\mathbf{x}_{ik},\mathbf{s}_i)]+[\alpha_j-\alpha_k]\end{aligned}$$

Notice that this implies that the alternative specific constants are identified only up to an additive constant.

Identification in Choice Models (cont'd)

2. Factors Affecting Alternatives Equally are Irrelevant
   If
   $$V_{ij} = V(\mathbf{x}_{ij}, \mathbf{s}_i) = \mathbf{x}'_{ij}\beta + \mathbf{s}'_i\gamma \tag{9}$$
   Then the characteristics vector $\mathbf{s}_i$ becomes irrelevant, since
   $$\begin{aligned} V_{ij} - V_{ik} &= \left[\mathbf{x}'_{ij}\beta + \mathbf{s}'_i\gamma\right] - \left[\mathbf{x}'_{ik}\beta + \mathbf{s}'_i\gamma\right] \\ &= \mathbf{x}'_{ij}\beta - \mathbf{x}'_{ik}\beta \\ &= \left(\mathbf{x}_{ij} - \mathbf{x}_{ik}\right)'\beta \end{aligned}$$

Identification in Choice Models (cont'd)

On the other hand, if
$$V_{ij} = V(\mathbf{x}_{ij}, \mathbf{s}_i) = \mathbf{x}'_{ij}\beta + \mathbf{s}'_i\gamma_j \tag{10}$$
Then the characteristics vector $\mathbf{s}_i$ is still relevant, since
$$\begin{aligned} V_{ij} - V_{ik} &= \left[\mathbf{x}'_{ij}\beta + \mathbf{s}'_i\gamma_j\right] - \left[\mathbf{x}'_{ik}\beta + \mathbf{s}'_i\gamma_k\right] \\ &= \left(\mathbf{x}_{ij} - \mathbf{x}_{ik}\right)'\beta + \mathbf{s}'_i\left(\gamma_j - \gamma_k\right) \end{aligned}$$
Though again the alternative specific parameters are identified only up to an additive constant.

Identification in Choice Models (cont'd)

3. Only Error Differences Matter, Not Their Levels

Note that:

$$\begin{aligned} P_{ij} &= Pr\left(U_{ij} > U_{ik} \forall k \neq j | \mathbf{x}_{i\cdot}, \mathbf{s}_i\right) \quad j = 1, \ldots, J-1 \\ &= Pr\left(\epsilon_{ik} - \epsilon_{ij} < V_{ij} - V_{ik} \forall k \neq j | \mathbf{x}_{i\cdot}, \mathbf{s}_i\right) \\ &= Pr\left(\eta_{ik} - \eta_{ij} < \Delta_J V_{ij} - \Delta_J V_{ik} \forall k \neq j | \mathbf{x}_{i\cdot}, \mathbf{s}_i\right) \\ &= \int_{D_\epsilon} I\left(\eta_{ik} - \eta_{ij} < \Delta_J V_{ij} - \Delta_J V_{ik} \forall k \neq j | \mathbf{x}_{i\cdot}, \mathbf{s}_i\right) g\left(\boldsymbol{\eta}_{i\cdot} | \mathbf{x}_{i\cdot}, \mathbf{s}_i\right) d\boldsymbol{\eta}_{i\cdot} \end{aligned}$$

where $\eta_{ij} \equiv \epsilon_{ij} - \epsilon_{iJ}$ and $\Delta_J V_{ij} \equiv V_{ij} - V_{iJ}; j = 1, \ldots, J$

Identification in Choice Models (cont'd)

4. The Scale of Utility is Irrelevant

$$\begin{aligned} P_{ij} &= Pr\left(\epsilon_{ik} - \epsilon_{ij} < V_{ij} - V_{ik} \forall k \neq j | \mathbf{x}_{i\cdot}, \mathbf{s}_i\right) \\ P_{ij} &= Pr\left(\lambda\left[\epsilon_{ik} - \epsilon_{ij}\right] < \lambda\left[V_{ij} - V_{ik}\right] \forall k \neq j | \mathbf{x}_{i\cdot}, \mathbf{s}_i\right) \end{aligned}$$

The original utility specification is indistinguishable from

$$\tilde{U}_{ij} = \lambda U_{ij} = \lambda V(\mathbf{x}_{ij}, \mathbf{s}_i) + \lambda \epsilon_{ij} \tag{11}$$

One must employ a normalizing restriction; e.g., for multivariate probit one often uses

$$\sigma_{11} \equiv Var(\eta_{11}) = 1 \tag{12}$$

Implications

- Estimation techniques will identify characteristics of $g\left(\boldsymbol{\eta}_{i\cdot}|\mathbf{x}_{i\cdot},\mathbf{s}_i\right)$.
- Without strong distributional assumptions. $f\left(\boldsymbol{\eta}_{i\cdot}|\mathbf{x}_{i\cdot},\mathbf{s}_i\right)$ will not be fully characterized.
- At best, one can hope to identify $J-1$ means (i.e., the $\alpha_j - \alpha_J$'s) and $\frac{(J-1)(J-2)}{2} - 1$ variance-covariance terms (in addition to any marginal effects).