ECN 565: Data Science & Econometrics II

Alvin Murphy

Lecture 8b: Limited Dependent Variables – Discrete Choice

Discrete Choice Models

- Goals
  - Extend our knowledge of binary models to multinomial discrete choice
  - Maintain (and extend) the tight link to economics

Discrete Choice Models

- Agents select from a finite set of differentiated alternatives
- Each alternative provides them with a certain payoff
- Agents select the alternative that gives them the highest payoff
  - McFadden (1974) transportation mode choice
  - Smartphone brand choice, e.g., Apple vs. Samsung vs. Google
- Latent variable framework takes on a specific economic interpretation.
  - Individuals maximize utility
  - Firms maximize profit

Discrete Choice Models

- Heterogeneous agents choose among differentiated alternatives
- $i \in \{1, \ldots, N\}$ agents
- $j \in \{1, \ldots, J\}$ alternatives
- x: alternative characteristics
- z: agent characteristics

Discrete Choice Models

- Choice set:
- Mutually Exclusive
  - Not restrictive
- Exhaustive
  - Not restrictive
- Finite
  - Only slightly restrictive

Discrete Choice Models

- Choice Set example
- Berry, Levinsohn, and Pakes (JPE 2004)
  - "Micro-BLP"
- $i \in \{1, \ldots, 37500\}$ vehicle buyers
- $j \in \{1, \ldots, 203\}$ vehicles
- x: price, HP, seats, MPG, safety, model, minivan, pickup, luxury
- z: kids, rural, income

Random Utility Model

- Discrete-choice models often derived from utility maximization
  - Methods can also be used to model other behavior
  - Or to statistically describe outcomes
- RUM Random Utility Model (Marshak (1960) McFadden (1974))
- Decision maker, i chooses alternative j to max utility/payoff/profit
- Utility i receives from j is U_{ij}
- Choose alternative j if $U_{ij} > U_{ik}\ \forall k \neq j$

Random Utility Model

- Don't observe U_{ij}
- Observe alternative attributes, x_{ij} $\forall j$
  - different agents may face different attributes
- Observe agent attributes, z_i
- Observe agent choice $d_{ij} \in \{0, 1\}$ $\forall j$
- e.g. electric vehicle adoption based on prices, incentives, and charging availability.

Random Utility Model

- Specify function relating observables to (part of) utility
    - $V_{ij} = V(x_{ij}, z_i; \theta)$
    - Sometimes called "representative utility"
    - $V_{ij} \neq U_{ij}$
    - $U_{ij} = V_{ij} + \epsilon_{ij}$
    - ϵ typically has an economic interpretation

Random Utility Model

- As usual, econometrician doesn't observe ϵ
- $f(\epsilon_i)$ is joint density of $\epsilon_i = [\epsilon_{i,1}, \ldots, \epsilon_{i,J}]'$
- Now have ingredients to make probabilistic statements about agent's choice:

$$
\begin{aligned}
P_{ij} &= Prob(U_{ij} > U_{ik} \ \forall k \neq j) \\
&= Prob(V_{ij} + \epsilon_{ij} > V_{ik} + \epsilon_{ik} \ \forall k \neq j) \\
&= Prob(\epsilon_{ij} - \epsilon_{ik} > V_{ik} - V_{ij} \ \forall k \neq j) \\
&= Prob(\epsilon_{ik} - \epsilon_{ij} < V_{ij} - V_{ik} \ \forall k \neq j)
\end{aligned}
$$

Random Utility Model

- This probability is a cumulative distribution
- Probability that each $\epsilon_{ik} - \epsilon_{ij}$ is below $V_{ij} - V_{ik}$
  - Recall that V_{ij} is observed/specified by econometrician

$$
\begin{aligned}
P_{ij} &= Prob(\epsilon_{ik} - \epsilon_{ij} < V_{ij} - V_{ik} \; \forall k \neq j) \\
&= \int_{\epsilon} I(\epsilon_{ik} - \epsilon_{ij} < V_{ij} - V_{ik} \; \forall k \neq j) f(\epsilon) d\epsilon
\end{aligned}
$$

- multidimensional integral over density of ϵ

Random Utility Model

- Specification of $f(\epsilon)$ determines which discrete model we specify.
- Multidimensional integral may or may not have closed form solution
- Closed form: Logit, GEV (including nested logit)
- Not closed form: Probit, Mixed Logit
- Simulate $\int_{\epsilon} I(\epsilon_{ik} - \epsilon_{ij} < V_{ij} - V_{ik} \; \forall k \neq j) f(\epsilon) d\epsilon$

Quick preview of models

- Logit: Quick, easy, popular
- Assumes $f(\epsilon)$ is i.i.d. type 1 extreme value (Gumbel)
  - Difference in type 1 extreme values is logistic
  - Max of type 1 extreme values is type 1 extreme value
- Unobserved factors are uncorrelated across alternatives.
  - Also have the same variance across alternatives
- What if alternatives are bus, train, and car?
  - Seems likely that $\epsilon_{i,bus}$ and $\epsilon_{i,train}$ are correlated.
- Other approaches (GEV, Probit) mainly used to avoid independence assumption

Identification

- Closely related to your theory classes
- Meaningless properties of utility are not identified
  - not surprising!

1. Only differences in utility matter
2. Scale of utility is arbitrary

Identification – Differences in Utility

- Only differences in utility matter
- If a constant is added to U_{ij} $\forall j$:
  - Alternative with highest utility doesn't change
  - Choice probabilities don't change
  - $P_{ij} = Prob(\epsilon_{ik} - \epsilon_{ij} < V_{ij} + c - V_{ik} - c \; \forall k \neq j)$
  - $P_{ij} = Prob(\epsilon_{ik} - \epsilon_{ij} < V_{ij} - V_{ik} \; \forall k \neq j)$
- only identified parameters are those that capture differences across alternatives

Identification – Differences in Utility

- Agent characteristics, z_i
- Only enter the model if they shift utility differences
- Can't identify the impact of z on U
- Can identify the impact of z on $U_j - U_k$
- Does education affect overall utility?
- Does education affect the choice among cars?

Identification – Scale of Utility

- Multiplying each alternative utility by constant will not change decision
  - multiplying U_{ij} by constant will not change optimal decision
  - multiplying V_{ij} by constant could change optimal decision
- Address this (usually) by normalizing the variance of the error term.
- $U_{ij} = V_{ij} + \epsilon^*_{ij} = V_{ij} + \sigma\epsilon_{ij}$
  - where ϵ^*_{ij} has scale $= \sigma$ and ϵ_{ij} has scale $= 1$
- Normalize $\sigma = 1$
- Equivalent to dividing through by σ
- $U_{ij}/\sigma = V_{ij}/\sigma + \epsilon_{ij}$

Identification – Scale of Utility

- $U_{ij} = x'_{ij}\beta + \sigma\epsilon_{ij}$
- $U_{ij}/\sigma = x'_{ij}\beta/\sigma + \epsilon_{ij}$
- β is not identified; β/σ is identified;
- Typically, normalize $\sigma = 1$.
- $U_{ij} = x'_{ij}\beta + \epsilon_{ij}$
    - important to remember what units β is measured in
    - β is really β/σ

Identification – Scale of Utility

- Ratio of β coefficients is invariant to scaling
  - Important as this is what we are usually interested in
  - MRS between two attributes
  - When one attribute is price, this is MWTP

Logit Model

- Make assumption that ϵ i.i.d. type 1 extreme value
- Difference between extreme value RVs is distributed logistic
- Very similar to **independent** Normal

Logit Model

- Choice Probabilities

$$
\begin{aligned}
P_{ij} &= Prob(U_{ij} > U_{ik} \; \forall k \neq j) \\
&= Prob(V_{ij} + \epsilon_{ij} > V_{ik} + \epsilon_{ik} \; \forall k \neq j) \\
&= Prob(\epsilon_{ij} - \epsilon_{ik} > V_{ik} - V_{ij} \; \forall k \neq j) \\
&= \int_{\epsilon} I(\epsilon_{ik} - \epsilon_{ij} < V_{ij} - V_{ik} \; \forall k \neq j) f(\epsilon) d\epsilon \\
&= \frac{e^{V_{ij}}}{\sum_k e^{V_{ik}}}
\end{aligned}
$$

Logit Model

- If assume $V_{ij} = x_{ij}'\beta$
- $P_{ij} = \frac{e^{x_{ij}'\beta}}{\sum_k e^{x_{ik}'\beta}}$

Logit Model

- Desirable properties of these choice probabilities:
- P_{ik} is between 0 and 1
    - Any sensible model derived from utility maximization would yield this
- The probabilities sum to one
    - Any sensible model derived from utility maximization would yield this

Logit Model

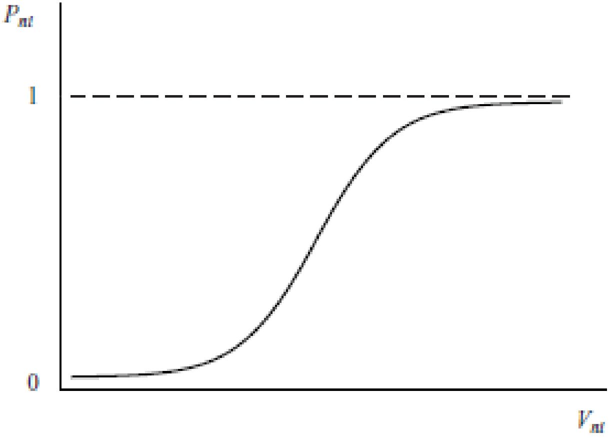

Figure 3.1. Graph of logit curve.

Logit Model

- Taste Variation
- Substitution patterns

Logit Model – Taste Variation

- Logit model is restrictive in the type of taste variation it can handle
- Heterogeneity in tastes for attributes that varies with observables
- No heterogeneity in tastes for attributes that varies with unobservables
    - Only have heterogeneity is taste for overall alternative, ϵ_{ij}

Logit Model – Taste Variation

- Start with general model: $U_{ij} = x'_{ij}\beta_i + \epsilon_{ij}$
  - note i subscript on β_i
- for element l of x_{ij}, let $\beta_{il} = \alpha_{0l} + z'_i\alpha_{1l} + \eta_{il}$
- Logit model naturally allows for α_1 (and α_0)
- Logit model assumes $\eta_i = 0$
- Without $\eta_i = 0$, error term is $x'_{ij}\eta_i + \epsilon_{ij}$
  - Error term will not be independently distributed
  - Also won't be identically distributed
- Extent to which $\eta_i = 0$ is restrictive depends on how informative z is

Logit Model – Substitution Patterns

- If we improve an attribute of one alternative, its probability must rise
- Offsetting reduction in the sum of the probabilities of other alternatives.
- Logit places strong restrictions of the nature of this offsetting effect

Logit Model – Substitution Patterns

- Independence from Irrelevant Alternatives (IIA):

$$
\begin{aligned}
\frac{P_{ij}}{P_{ij'}} &= \frac{e^{V_{ij}} / \sum_k e^{V_{ik}}}{e^{V_{ij'}} / \sum_k e^{V_{ik}}} \\
&= \frac{e^{V_{ij}}}{e^{V_{ij'}}} \\
&= e^{V_{ij} - V_{ij'}}
\end{aligned}
$$

- odds ratio does not depend on alternatives other than j and j'
- proportional substitution patterns
- Red bus, blue bus, car

Logit Model – Substitution Patterns

- Consider the following model (with no observed z_i):
- $U_{ij} = \beta_1 * HP_j + \beta_2 * Price_j + \epsilon_{ij}$
- $J = 3$; one sports car and two regular cars
- HP$=$ [400, 100, 100]; Price $=$ [40, 10, 10]
- $\beta_1 = 2$; $\beta_2 = -20$
- $V_{i1} = V_{i2} = V_{i3} = 0 \; \forall i$
- $P_1 = P_2 = P_3 = 1/3$

Logit Model – Substitution Patterns

- What happens if we remove one of the regular cars from the choice set?
- $U_{ij} = \beta_1 * HP_j + \beta_2 * Price_j + \epsilon_{ij}$
- $J = 2$; one sports car and one regular car
- HP$=$ [400, 100]; Price $=$ [40, 10]
- $\beta_1 = 2$; $\beta_2 = -20$
- $V_{i1} = V_{i2} = 0 \ \forall i$
- $P_1 = P_2 = 1/2$

Logit Model – Substitution Patterns

- What happens if we observe (informative) z_i
- $U_{ij} = \beta_{i1} * HP_j + \beta_{i2} * Price_j + \epsilon_{ij}$
- $J = 3$; one sports car and two regular cars
- HP$=$ [400, 100, 100]; Price $=$ [40, 10, 10]
- $z_i \in \{0, 1\}$. $P(z_i = 1) = 2/3$
- $\beta_{i1} = 6 - 6 * z_i$; $\beta_2 = 0 - 30 * z_i$
  - $z_i = 1 \rightarrow \beta_1 = 0, \; \beta_2 = -30$
  - $z_i = 0 \rightarrow \beta_1 = 6, \; \beta_2 = 0$

Logit Model – Substitution Patterns

- $V_1(z_i = 1) = -1200, \;\; V_2(z_i = 1) = V_3(z_i = 1) = -300 \;\forall i$
- $P_1(z_i = 1) \approx 0, \;\; P_2(z_i = 1) = P_3(z_i = 1) \approx 1/2$
- $V_1(z_i = 0) = 2400, \;\; V_2(z_i = 0) = V_3(z_i = 0) = 600 \;\forall i$
- $P_1(z_i = 0) \approx 1, \;\; P_2(z_i = 0) = P_3(z_i = 0) \approx 0$
- $P_1 = P_2 = P_3 = 1/3$

Logit Model – Substitution Patterns

- What happens if we remove one of the regular cars from the choice set?
- IIA holds separately for each value of z_i
- $V_1(z_i = 1) = -1200, \; V_{i2}(z_i = 1) = -300 \; \forall i$
- $P_1(z_i = 1) \approx 0, \; P_2(z_i = 1) \approx 1$
- $V_1(z_i = 0) = 2400, \; V_{i2}(z_i = 0) = 600 \; \forall i$
- $P_1(z_i = 0) \approx 1, \; P_2(z_i = 0) \approx 0$
- $P_1 \approx 1/3, \; P_2 \approx 2/3$
- I chose extreme numbers to make a point.

Logit Model – Characterizing alternatives

- It is important to carefully characterize set of alternatives
- This is especially true for Logit model
- Example 1: Is the Honda CRV LX a different alternative to Honda CRV EX
  - If so, Logit model says errors are independent
- Example 2: How to determine what is a “neighborhood”
  - Block or Block group?
- Logit models can have poor properties when J is large
  - Independence is the source of the problems

Logit Model – Consumer Surplus

- CS is the utility (in dollar terms) that an individual receives from a given situation.
- Convert utils to dollars using marginal utility of income
    - Coefficient on price, income, numeraire consumption
- Let $\partial U_i / \partial Income_i = \alpha_i$
    - For simple case let α_i be constant over income
- $CS_i = (1/\alpha_i) max_j(U_{ij})$

Logit Model – Consumer Surplus

- Don't observe U_{ij}
- Integrate out over unobservable, ϵ_i
- $E(CS_i) = (1/\alpha_i)E[max_j(V_{ij} + \epsilon_{ij})]$
- $E(CS_i) = (1/\alpha_i)log(\sum_j^J e^{V_{ij}})$
- $\Delta E(CS_i) = (1/\alpha_i)\left(log(\sum_j^{J^1} e^{V_{ij}^1}) - log(\sum_j^{J^0} e^{V_{ij}^0})\right)$
- e.g., difference in consumer surplus when commute times drop by 10 minutes.
- e.g., difference in consumer surplus when one car is removed from choice set.

Logit Model – Estimation

- Assume that x_{ij}, z_i are exogenous
- Likelihood contribution of individual i: $\prod_j P_{ij}^{d_{ij}}$
- Likelihood Function: $L(\beta) = \prod_i \prod_j P_{ij}^{d_{ij}}$
- Log-likelihood Function: $\mathcal{L}(\beta) = \sum_i \sum_j d_{ij} log(P_{ij})$
- $\hat{\beta}_{mle} = \text{argmax}_{\beta} \mathcal{L}(\beta)$
- McFadden (1974): $\mathcal{L}(\beta)$ is globally concave in β when $V_{ij} = x_{ij}'\beta_i$
  - Pre-canned routines exist for this case

Logit Model – More examples

- Marketing: Brand choice modeling for new product launches.
- Urban economics: Housing location or school choice.
- Environmental economics: Estimating willingness-to-pay for clean air or EVs.
- Health economics: Patient choice of hospital or treatment plan.

GEV Model

- IIA property can be restrictive
  - This is especially true when we don't observe z_i
- Main alternatives to Logit are:

1. GEV Generalized Extreme Value
   - Logit is a special case of GEV
   - Most commonly used GEV structure is Nested Logit
2. Probit
3. Mixed Logit
   - Logit is a special case of Mixed-Logit