ECON 626: Applied Microeconomics

Lecture 12:

Conditional Logit

Professors: Pamela Jakiela and Owen Ozier

Department of Economics
University of Maryland, College Park

The Logit Model: Binary Choice Edition

A standard latent variable model of a binary outcome, y:

$$y^* = X'\beta + \varepsilon \text{ and } y = 1\left[y^* > 0\right]$$

The Logit Model: Binary Choice Edition

A standard latent variable model of a binary outcome, y:

$$y^* = X'\beta + \varepsilon \text{ and } y = 1\left[y^* > 0\right]$$

For any symmetric distribution of ε, we can write:

$$\begin{aligned}
\Pr\left[y = 1|X\right] &= \Pr\left[y^* > 0|X\right] \\
&= \Pr\left[\varepsilon > -X'\beta|X\right] \\
&= 1 - F\left(-X'\beta\right) \\
&= F\left(X'\beta\right)
\end{aligned}$$

The Logit Model: Binary Choice Edition

A standard latent variable model of a binary outcome, y:

$$y^* = X'\beta + \varepsilon \text{ and } y = 1\left[y^* > 0\right]$$

For any symmetric distribution of ε, we can write:

$$\begin{aligned}
\Pr\left[y = 1|X\right] &= \Pr\left[y^* > 0|X\right] \\
&= \Pr\left[\varepsilon > -X'\beta|X\right] \\
&= 1 - F\left(-X'\beta\right) \\
&= F\left(X'\beta\right)
\end{aligned}$$

When ε has a standard logistic distribution:

$$F\left(X'\beta\right) = \frac{e^{X'\beta}}{1 + e^{X'\beta}}$$

The Logit Model: Binary Choice Edition

Translates into the likelihood function:

$$\mathcal{L}(\beta) = [F(X'\beta)]^y [1 - F(X'\beta)]^{1-y}$$

$$\ell(\beta) = y \ln [F(X'\beta)] + (1 - y) \ln [1 - F(X'\beta)]$$

The Logit Model: Binary Choice Edition

Translates into the likelihood function:

$$\mathcal{L}(\beta) = \left[F(X'\beta)\right]^{y} \left[1 - F(X'\beta)\right]^{1-y}$$

$$\ell(\beta) = y \ln\left[F(X'\beta)\right] + (1-y)\ln\left[1 - F(X'\beta)\right]$$

Possible interpretations:

- X indicates the characteristics of people/units
- X indicates the difference(s) in attributes between alternatives

The Logit Model: Binary Choice Edition

Translates into the likelihood function:

$$\mathcal{L}(\beta) = [F(X'\beta)]^y [1 - F(X'\beta)]^{1-y}$$

$$\ell(\beta) = y \ln [F(X'\beta)] + (1-y) \ln [1 - F(X'\beta)]$$

Possible interpretations:

- X indicates the characteristics of people/units
- X indicates the difference(s) in attributes between alternatives

Logit model can be extended to cases with $J > 2$ outcomes

The Additive Random Utility Model

In an additive random utility model, realized utility is the sum of the modeled component ("representative utility") and a random component:

$$U_{nj} = V_{nj} + \varepsilon_{nj}$$

where V_{nj} is often assumed to be $X'\beta$, ε_{nj} are IID

The Additive Random Utility Model

In an additive random utility model, realized utility is the sum of the modeled component ("representative utility") and a random component:

$$U_{nj} = V_{nj} + \varepsilon_{nj}$$

where V_{nj} is often assumed to be $X'\beta$, ε_{nj} are IID

We assume that the highest utility alternative is chosen:

$$\begin{aligned} P_{nj} &= \Pr\left[V_{nj} + \varepsilon_{nj} > V_{nk} + \varepsilon_{nk} \forall k \neq j\right] \\ &= \Pr\left[\varepsilon_{nj} > V_{nk} - V_{nj} + \varepsilon_{nk} \forall k \neq j\right] \\ &\dots \\ &= \frac{e^{V_{nj}}}{\sum_{k \in J} e^{V_{nk}}} \end{aligned}$$

whenever ϵ_{nj} is EV1-distributed

The Additive Random Utility Model

Translates into the log-likelihood function:

$$\ell_n(\theta) = \sum_j z_{nj} \cdot \ln(P_{nj})$$

where

- z_{nj} is an indicator equal to one if n chooses option j
- P_{nj} is the probability of choosing alternative j

Representative Utility

A standard formulation of utility: $V_{nj} = X'\beta$

- Example 1: choosing a vacation destination
  - Utility is defined over attributes: Beach? Pool? Nightlife? Spa?
  - When all utility parameters reflect valuation of specific attributes, assumption of linear representative utility is essentially costless

Representative Utility

A standard formulation of utility: $V_{nj} = X'\beta$

- Example 1: choosing a vacation destination
  - Utility is defined over attributes: Beach? Pool? Nightlife? Spa?
  - When all utility parameters reflect valuation of specific attributes, assumption of linear representative utility is essentially costless
- Example 2: deciding how to get to work
  - Utility is defined over money, time, plus other attributes
  - Assumption of (approximately) linear utility over some amounts of money (or time) is standard in some circles, less accepted in others
  - **Including price as one of the attributes allows the researcher to put dollar values on the utility derived from other attributes**

Representative Utility

Linear representative utility is not a requirement

- Example 3: deciding how to get to work, revisited
  - Train and McFadden (1978) assume $V = (1-\beta)\ln G + \beta \ln L$
    - $G = w \cdot h$ is consumption of a numeraire good
    - $L = 24 - h$ is leisure time
  - Commuting by car, bus takes time and costs money; consumers choose optimal amount of labor conditional on mode of transport
  - Implies: $V_{nj} \propto -\left[\left(c_j/w^{\beta}\right) + w^{1-\beta} t_j\right]$

Representative Utility

Linear representative utility is not a requirement

- Example 3: deciding how to get to work, revisited
  - Train and McFadden (1978) assume $V = (1-\beta)\ln G + \beta \ln L$
    - $G = w \cdot h$ is consumption of a numeraire good
    - $L = 24 - h$ is leisure time
  - Commuting by car, bus takes time and costs money; consumers choose optimal amount of labor conditional on mode of transport
  - Implies: $V_{nj} \propto -\left[\left(c_j/w^{\beta}\right) + w^{1-\beta} t_j\right]$
- Example 4: choosing between risky lotteries
  - Von Gaudecker *et al* (2011) assume V_{nj} takes CARA form

The Scale Parameter

Utility of alternative $j \in J$ is given by:

$$U_{nj} = V_{nj} + \varepsilon_{nj}$$

When ε_{nj} is EV1-distributed, the choice probabilities are given by:

$$P_{nj} = \Pr\left[\varepsilon_{nj} > V_{nk} - V_{nj} + \varepsilon_{nk} \forall k \neq j\right]$$

$$= \frac{e^{V_{nj}}}{\sum_{k \in J} e^{V_{nk}}}$$

The Scale Parameter

Utility of alternative $j \in J$ is given by:

$$U_{nj} = V_{nj} + \varepsilon_{nj}$$

When ε_{nj} is EV1-distributed, the choice probabilities are given by:

$$P_{nj} = \Pr\left[\varepsilon_{nj} > V_{nk} - V_{nj} + \varepsilon_{nk} \forall k \neq j\right]$$

$$= \frac{e^{V_{nj}}}{\sum_{k \in J} e^{V_{nk}}}$$

Why? Magical algebra. Critically, the CDF of the standard EV1 is

$$F\left(\varepsilon_{nj})\right) = e^{-e^{-\varepsilon_{nj}}}$$

The difference of two EV1-distributed variables has a logistic distribution

The Scale Parameter

Variance of random component of utility (relative to magnitude/scale of representative utility) indicates importance of noise in decision-making

- Variance of a standard EV1-distributed variable: $\pi^2/6$
  - Obviously no reason to assume this is variance of ε_{nj}
- A more "true" model of utility:

$$U_{nj} = V_{nj}/\sigma_\varepsilon + \varepsilon_{nj}$$

where variance of ε_{nj} is normalized to $\pi^2/6$

The Scale Parameter

Variance of random component of utility (relative to magnitude/scale of representative utility) indicates importance of noise in decision-making

- Variance of a standard EV1-distributed variable: $\pi^2/6$
  - Obviously no reason to assume this is variance of ε_{nj}
- A more "true" model of utility:

$$U_{nj} = V_{nj}/\sigma_\varepsilon + \varepsilon_{nj}$$

  where variance of ε_{nj} is normalized to $\pi^2/6$

- With linear representative utility linear, β, σ not separately identified:

$$U_{nj} = X'\beta/\sigma_\varepsilon + \varepsilon_{nj} \Rightarrow P_{nj} = \frac{e^{X'_{nj}(\beta/\sigma)}}{\sum_{k \in J} e^{X'_{nk}(\beta/\sigma)}}$$

The Scale Parameter

Variance of random component of utility (relative to magnitude/scale of representative utility) indicates importance of noise in decision-making

- Variance of a standard EV1-distributed variable: $\pi^2/6$
  - Obviously no reason to assume this is variance of ε_{nj}
- A more "true" model of utility:

$$U_{nj} = V_{nj}/\sigma_{\varepsilon} + \varepsilon_{nj}$$

  where variance of ε_{nj} is normalized to $\pi^2/6$

- With linear representative utility linear, β, σ not separately identified:

$$U_{nj} = X'\beta/\sigma_{\varepsilon} + \varepsilon_{nj} \Rightarrow P_{nj} = \frac{e^{X'_{nj}(\beta/\sigma)}}{\sum_{k \in J} e^{X'_{nk}(\beta/\sigma)}}$$

Does this matter? Utility is robust to positive, affine transformations.

The Scale Parameter

Magnitude of structural parameters is **relative** to random error term

- True interpretation of estimated parameters is β/σ

The Scale Parameter

Magnitude of structural parameters is **relative** to random error term

- True interpretation of estimated parameters is β/σ

Caveat 1: we may wish to estimate differences in error variance

- Example: literacy/ability, attribute salience

The Scale Parameter

Magnitude of structural parameters is **relative** to random error term

- True interpretation of estimated parameters is β/σ

Caveat 1: we may wish to estimate differences in error variance

- Example: literacy/ability, attribute salience

Caveat 2: variance of random component of utility is (typically) identified when representative utility is non-linear (e.g. labor/leisure tradeoffs)

- "Identification based on functional form assumptions"
    - Are your simplifying assumptions driving your results?
    - Are your results robust to other specifications?

The Scale Parameter

Magnitude of structural parameters is **relative** to random error term

- True interpretation of estimated parameters is β/σ

Caveat 1: we may wish to estimate differences in error variance

- Example: literacy/ability, attribute salience

Caveat 2: variance of random component of utility is (typically) identified when representative utility is non-linear (e.g. labor/leisure tradeoffs)

- "Identification based on functional form assumptions"
  - Are your simplifying assumptions driving your results?
  - Are your results robust to other specifications?

Likelihood of choosing "randomly" is often of inherent interest

Estimation in Stata

Example: choosing a vacation destination

Destination	Cost	Travel Time	Beach?	Warm?
Annapolis	\$1000	1 hour	Yes	No
Bahamas	\$8000	4 hours	Yes	Yes
Costa Rica	\$4000	8 hours	Yes	Yes
Durango	\$2000	6 hours	No	Yes

You observe a data set on the destination choices of 1000 people

- Need to observe the full menu of choices

Estimation in Stata

Specify the utility function:

$$U_{nj} = -\alpha Price_{nj} - \beta TravelTime_j + \gamma Warm_j + \delta Beach_j + \lambda Warm \times Beach_j$$

Preference parameters are NOT heterogeneous across individuals:

- Tastes can only vary with observable characteristics
  - $Price_{nj}$ depends on number of people traveling
  - Could also interact price with household income

Estimation in Stata

Estimation in Stata uses `asclogit` command

- Data set needs to be at the **alternative** level

Estimation in Stata

Estimation in Stata uses `asclogit` command

- Data set needs to be at the **alternative** level

```
asclogit chosen price time warm beach warmxbeach, ///
 case(id) alternative(destination) noconstant altwise
```

where:

- `case()` indicates the decision situation
- `alternative()` indicates the alternative (a unique # within case)
- `noconstant` means do not include alternative-specific constants
- `altwise` means drop alternatives, not entire decision situations

The Independence of Irrelevant Alternatives

The relative odds of choosing alternative j over alternative i:

$$\frac{P_{nj}}{P_{ni}} = \frac{e^{V_{nj}} / \sum_{k \in J} e^{V_{nk}}}{e^{V_{ni}} / \sum_{k \in J} e^{V_{nk}}}$$

$$= \frac{e^{V_{nj}}}{e^{V_{ni}}}$$

$$= e^{V_{nj} - V_{ni}}$$

Odds do not depend on the other elements of the choice set

- Luce (1959) argues that IIA is critical to correct choice probabilities

Of Red Buses and Blue Buses

Suppose a commuter faces a choice between driving, taking the (red) bus

- At the outset, she is two times more likely to drive: $P_{car}/P_{rb} = 2$
- (So, $P_{car} = \frac{2}{3}$ and $P_{rb} = \frac{1}{3}$)

Of Red Buses and Blue Buses

Suppose a commuter faces a choice between driving, taking the (red) bus

- At the outset, she is two times more likely to drive: $P_{car}/P_{rb} = 2$
- (So, $P_{car} = \frac{2}{3}$ and $P_{rb} = \frac{1}{3}$)

Now suppose that her local transit authority introduces a third option — a blue bus — that is exactly identical to the red bus in every way

- By construction: $V_{rb} = V_{bb}$, so $P_{rb} = P_{bb}$

Of Red Buses and Blue Buses

Suppose a commuter faces a choice between driving, taking the (red) bus

- At the outset, she is two times more likely to drive: $P_{car}/P_{rb} = 2$
- (So, $P_{car} = \frac{2}{3}$ and $P_{rb} = \frac{1}{3}$)

Now suppose that her local transit authority introduces a third option — a blue bus — that is exactly identical to the red bus in every way

- By construction: $V_{rb} = V_{bb}$, so $P_{rb} = P_{bb}$

How should the probability of taking the red bus change?

- Common sense prediction: $P_{car} = \frac{2}{3}$ and $P_{rb} = P_{bb} = \frac{1}{6}$
- Logit prediction: $P_{car} = \frac{1}{2}$ and $P_{rb} = P_{bb} = \frac{1}{4}$

Of Red Buses and Blue Buses

Suppose a commuter faces a choice between driving, taking the (red) bus

- At the outset, she is two times more likely to drive: $P_{car}/P_{rb} = 2$
- (So, $P_{car} = \frac{2}{3}$ and $P_{rb} = \frac{1}{3}$)

Now suppose that her local transit authority introduces a third option — a blue bus — that is exactly identical to the red bus in every way

- By construction: $V_{rb} = V_{bb}$, so $P_{rb} = P_{bb}$

How should the probability of taking the red bus change?

- Common sense prediction: $P_{car} = \frac{2}{3}$ and $P_{rb} = P_{bb} = \frac{1}{6}$
- Logit prediction: $P_{car} = \frac{1}{2}$ and $P_{rb} = P_{bb} = \frac{1}{4}$

More flexible model can accommodate different substitution patterns

Mixed Logit

In a mixed logit model, choice probabilities takes the form:

$$P_{nj} = \int \left(\frac{e^{V_{nj}(\beta)}}{\sum_{k \in J} e^{V_{nk}(\beta)}} \right) f(\beta)\, d\beta$$

where $f(\beta)$ is the density of the (unobserved) utility parameter, β

Three examples:

- Random component of utility correlated across alternatives (e.g. correlation between likelihood of choosing red vs. blue bus)
- Random coefficients
- Latent class (i.e. finite mixture) models

Random Coefficients Logit

Utility parameters vary across individuals:

$$U_{nj} = X'_{nj}\beta_n + \varepsilon_{nj}$$

where $\beta \sim \mathcal{N}\left(\mu, \sigma^2\right)$

Random Coefficients Logit

Utility parameters vary across individuals:

$$U_{nj} = X_{nj}'\beta_n + \varepsilon_{nj}$$

where $\beta \sim \mathcal{N}\left(\mu, \sigma^2\right)$

Choice probabilities are given by:

$$P_{nj} = \int \left(\frac{e^{X_{nj}'\beta}}{\sum_{k \in J} e^{X_{nj}'\beta}} \right) \phi\left(\beta | \mu, \sigma^2\right) d\beta$$

Random Coefficients Logit

Utility parameters vary across individuals:

$$U_{nj} = X'_{nj}\beta_n + \varepsilon_{nj}$$

where $\beta \sim \mathcal{N}\left(\mu, \sigma^2\right)$

Choice probabilities are given by:

$$P_{nj} = \int \left(\frac{e^{X'_{nj}\beta}}{\sum_{k \in J} e^{X'_{nj}\beta}} \right) \phi\left(\beta | \mu, \sigma^2\right) d\beta$$

The choice probabilities don't depend on β

- μ, σ are the parameters to be estimated
- Assumption of normality can be replace with any other distribution

Mixed Logit Estimation

Estimation of mixed logit is done via simulated log likelihood

$$SLL_n = \sum_j z_{nj} \cdot \ln\left(\tilde{P}_{nj}\right)$$

where $\tilde{P}_{nj}$ is averaged over R random draws from $f\left(\beta\right)$

$$\tilde{P}_{nj} = \frac{1}{R}\sum_{r=1}^{R} \frac{e^{X'_{nj}\beta^r}}{\sum_{k\in J} e^{X'_{nj}\beta^r}}$$

Mixed Logit Estimation

Estimation of mixed logit is done via simulated log likelihood

$$SLL_n = \sum_j z_{nj} \cdot \ln\left(\tilde{P}_{nj}\right)$$

where $\tilde{P}_{nj}$ is averaged over R random draws from $f(\beta)$

$$\tilde{P}_{nj} = \frac{1}{R}\sum_{r=1}^{R} \frac{e^{X'_{nj}\beta^r}}{\sum_{k\in J} e^{X'_{nj}\beta^r}}$$

Simulation proceeds in three intuitive steps:

- Take R draws from (say) a standard normal
  - Any distribution is possible (e.g. multivariate normal w/ correlation)
- At each step, scale using current values of parameters (e.g. μ, σ)
- Calculate simulated log-likelihood

Mixed Logit Estimation in Stata

Estimation in Stata uses `mixlogit` command

- Same data structure as conditional logit

```
mixlogit chosen warm beach warmxbeach, ///
 group(id) rand(price time) ln(0)
```

Results and interpretation:

- `rand()` indicates variables associated with random coefficients
- `ln()` indicates that the last # are log-normally distributed
- Stata reports means, standard deviations, and associated standard errors for all variables associated with random coefficients