

Lecture 4

Static Discrete Choice Models

Tyler Ransom

ECON 6343, University of Oklahoma

Attribution

Many of these slides are based on slides written by Peter Arcidiacono. I use them with his permission.

These slides also follow Chapters 1-3 of Train (2009)

Plan for the day

1. Describe static discrete choice models
2. Derive logit probabilities
3. Show how to estimate logit and multinomial logit models
4. Cover the IIA property
5. Discuss expected utility and consumer surplus

What are discrete choice models?

- Discrete choice models are one of the workhorses of structural economics
- Deeply tied to economic theory:
 - utility maximization
 - revealed preference
- Used to model "utility" (broadly defined), for example:
 - consumer product purchase decisions
 - firm market entry decisions
 - investment decisions

Example of a discrete choice model

- Cities in the Bay Area are interested in how the introduction of rideshare services will impact ridership on Bay Area Rapid Transit (BART)
- Questions that cities need to know the answers to:
 - Is rideshare a substitute for public transit or a complement?
 - How inelastic is demand for BART? Should fares be \uparrow or \downarrow ?
 - Should BART services be scaled up to compete with rideshares?
 - Will the influx of rideshare vehicles increase traffic congestion / pollution?
- Each of these questions requires making a counterfactual prediction
- In particular, need a way to make such a prediction confidently and in a way that is easy to understand

Properties of discrete choice models

1. Agents choose from among a **finite** set of alternatives (called the choice set)
2. Alternatives in choice set are **mutually exclusive**
3. Choice set is **exhaustive**

Notation

- Let d_i indicate the choice individual i makes where $d_i \in \{1, \dots, J\}$. categorical variable
- Individuals choose d to maximize their utility, U , which generally is written as:

$$U_{ij} = u_{ij} + \epsilon_{ij} \tag{1}$$

where:

assume that the components are additively separable.

- u_{ij} relates observed factors to the utility individual i receives from choosing option j
- ϵ_{ij} are unobserved to the econometrician but observed to the individual

$$d_{ij} = 1 \text{ if } u_{ij} + \epsilon_{ij} > u_{ij'} + \epsilon_{ij'} \text{ for all } j' \neq j \tag{2}$$

Probabilities

- With the ϵ 's unobserved, the probability of i making choice j is given by:

$$\begin{aligned} P_{ij} &= \Pr(u_{ij} + \epsilon_{ij} > u_{ij'} + \epsilon_{ij'} \quad \forall j' \neq j) \\ &= \Pr(\epsilon_{ij'} - \epsilon_{ij} < u_{ij} - u_{ij'} \quad \forall j' \neq j) \\ &= \int_{\epsilon} I(\epsilon_{ij'} - \epsilon_{ij} < u_{ij} - u_{ij'} \quad \forall j' \neq j) f(\epsilon) d\epsilon \end{aligned}$$

- Note that, regardless of what distributional assumptions are made on the ϵ 's, the probability of choosing a particular option does not change when we:
 1. Add a constant to the utility of all options (i.e. **only differences in utility matter**)
 2. Multiply by a positive number (need to **scale something**; e.g. the variance of the ϵ 's)

Utility is ordinal.

Monotonic transformation preserves the rank of the utility.

Variables

- Suppose we have:

$$u_{i1} = \alpha Male_i + \beta_1 X_i + \gamma Z_1$$

$$u_{i2} = \alpha Male_i + \beta_2 X_i + \gamma Z_2$$

- Since only differences in utility matter:

$$u_{i1} - u_{i2} = (\beta_1 - \beta_2)X_i + \gamma(Z_1 - Z_2)$$

- We can't tell whether men are happier than women, but can tell whether they more strongly prefer one option
- We can only obtain differenced coefficient estimates on X 's

Number of Error Terms

- Similar to the X 's, there are restrictions on the number of error terms
- This is because only differences in utility matter
- Recall that the probability i will choose j is given by:

$$\begin{aligned} P_{ij} &= \Pr(u_{ij} + \epsilon_{ij} > u_{ij'} + \epsilon_{ij'} \quad \forall j' \neq j) \\ &= \Pr(\epsilon_{ij'} - \epsilon_{ij} < u_{ij} - u_{ij'} \quad \forall j' \neq j) \\ &= \int_{\epsilon} I(\epsilon_{ij'} - \epsilon_{ij} < u_{ij} - u_{ij'} \quad \forall j' \neq j) f(\epsilon) d\epsilon \end{aligned} \tag{3}$$

where the integral is J -dimensional

Number of Error Terms

- Rewriting the last line of (3) as a $J - 1$ dimensional integral over the differenced ϵ 's:

$$P_{ij} = \int_{\tilde{\epsilon}} I(\tilde{\epsilon}_{ij'} < \tilde{u}_{ij'} \ \forall \ j' \neq j) g(\tilde{\epsilon}) d\tilde{\epsilon} \quad (4)$$

- This means one dimension of $f(\epsilon)$ is not identified and must therefore be normalized
- Arises from only differences in utility mattering (**location normalization**)
- The scale of utility also doesn't matter (**scale normalization**)
 - The scale normalization implies we must place restrictions on the variance of ϵ 's

More on the scale normalization

- The need to normalize scale means that we can never estimate the variance of $F(\tilde{\epsilon})$
- This contrasts with linear regression models, where we can easily estimate MSE
- The scale normalization means our β 's and γ 's are implicitly divided by an unknown variance term:

$$\begin{aligned}u_{i1} - u_{i2} &= (\beta_1 - \beta_2)X_i + \gamma(Z_1 - Z_2) \\&= \tilde{\beta}X_i + \gamma\tilde{Z} \\&= \frac{\beta^*}{\sigma}X_i + \frac{\gamma^*}{\sigma}\tilde{Z}\end{aligned}$$

- $\tilde{\beta}$ is what we estimate, but we will never know β^* because utility is scale-invariant

beta in marginal effects of a logit model is sort of difference-in-difference, as it is relative to baseline alternative.

Cannot compare coefficients from one logit model with other because the variance of unobserved factors would be different. Marginal effects, however, are comparable as it incorporates the unobserved variance.

Where does the logit formula come from?

- Consider a binary choice set $\{1, 2\}$. The Type 1 extreme value CDF for ϵ_2 is:

$$F(\epsilon_2) = e^{-e^{(-\epsilon_2)}}$$

- To get the probability of choosing 1, substitute in for ϵ_2 with $\epsilon_1 + u_1 - u_2$:

$$\Pr(d_1 = 1 | \epsilon_1) = e^{-e^{-(\epsilon_1 + u_1 - u_2)}} \quad (5)$$

- But ϵ_1 is unobserved so we need to integrate it out

Logit derivation

- Taking the integral over what is random (ϵ_1):

$$\begin{aligned}\Pr(d_1 = 1) &= \int_{-\infty}^{\infty} \left(e^{-e^{-(\epsilon_1 + u_1 - u_2)}} \right) f(\epsilon_1) d\epsilon_1 \\ &= \int_{-\infty}^{\infty} \left(e^{-e^{-(\epsilon_1 + u_1 - u_2)}} \right) e^{-\epsilon_1} e^{-e^{-\epsilon_1}} d\epsilon_1 \\ &= \int_{-\infty}^{\infty} \exp\left(-e^{-\epsilon_1} - e^{-(\epsilon_1 + u_1 - u_2)}\right) e^{-\epsilon_1} d\epsilon_1 \\ &= \int_{-\infty}^{\infty} \exp\left(-e^{-\epsilon_1} [1 + e^{u_2 - u_1}]\right) e^{-\epsilon_1} d\epsilon_1\end{aligned}$$

- We can simplify by U-substitution where $t = \exp(-\epsilon_1)$ and $dt = -\exp(-\epsilon_1)d\epsilon_1$

And adjusting bounds of integration accordingly, $\exp(-\text{inf})=\text{infinity}$ and $\exp(\text{inf})=0$

Logit Derivation

- Substituting in then yields:

$$\begin{aligned}\Pr(d_1 = 1) &= \int_{-\infty}^0 \exp\left(-t \left[1 + e^{(u_2 - u_1)}\right]\right) (-dt) \\ &= \int_0^{\infty} \exp\left(-t \left[1 + e^{(u_2 - u_1)}\right]\right) dt \\ &= \frac{\exp\left(-t \left[1 + e^{(u_2 - u_1)}\right]\right)}{-\left[1 + e^{(u_2 - u_1)}\right]} \Bigg|_0^{\infty} \\ &= 0 - \frac{1}{-\left[1 + e^{(u_2 - u_1)}\right]} \\ &= \frac{\exp(u_1)}{\exp(u_1) + \exp(u_2)}\end{aligned}$$

(this is still incomplete, need to incorporate difference in utility)

Logit Estimation

- Consider our model from before:

$$u_{i1} - u_{i2} = (\beta_1 - \beta_2)X_i + \gamma(Z_1 - Z_2)$$

- We observe X_i , Z_1 , Z_2 , and d_i
- Assuming $\epsilon_1, \epsilon_2 \stackrel{iid}{\sim} T1EV$ gives the likelihood of choosing 1 and 2 respectively as:

$$P_{i1} = \frac{\exp(u_{i1} - u_{i2})}{1 + \exp(u_{i1} - u_{i2})}$$

Because the choice set is exhaustive, $P_{i1} = 1 - P_{i2}$

$$P_{i2} = \frac{1}{1 + \exp(u_{i1} - u_{i2})}$$

- Note: if $\epsilon_1, \epsilon_2 \stackrel{iid}{\sim} T1EV$ then $\tilde{\epsilon}_1 \sim Logistic$, where $\tilde{\epsilon}_1 := \epsilon_1 - \epsilon_2$

Likelihood function

- We can view the event $d_i = j$ as a weighted coin flip
- This gives us a random variable that follows the Bernoulli distribution
- Supposing our sample is of size N , the likelihood function would then be

$$\begin{aligned}\mathcal{L}(X, Z; \beta, \gamma) &= \prod_{i=1}^N P_{i1}^{d_{i1}} P_{i2}^{d_{i2}} \\ &= \prod_{i=1}^N P_{i1}^{d_{i1}} [1 - P_{i1}]^{(1-d_{i1})}\end{aligned}\tag{6}$$

where P_{i1} and P_{i2} are both functions of X, Z, β, γ

Log likelihood function

- For many reasons, it's better to maximize the log likelihood function
- Taking the log of (6) gives

Multinomial Logit Estimation

- Adding more choices with i.i.d. Type I extreme value errors yields the **multinomial logit**
- Normalizing with respect to alternative J we have (for $j \in \{1, \dots, J-1\}$)

$$u_{ij} - u_{iJ} = (\beta_j - \beta_J)X_i + \gamma(Z_j - Z_J) \quad (8)$$

- We observe X_i, Z_1, \dots, Z_J , and d_i . The likelihood of choosing j and J respectively is:

$$P_{ij} = \frac{\exp(u_{ij} - u_{iJ})}{1 + \sum_{j'=1}^{J-1} \exp(u_{ij'} - u_{iJ})}, \quad P_{iJ} = \frac{1}{1 + \sum_{j'=1}^{J-1} \exp(u_{ij'} - u_{iJ})} \quad (9)$$

- The log likelihood function we maximize is then:

$$\ell(X, Z; \beta, \gamma) = \sum_{i=1}^N \left[\sum_{j=1}^{J-1} (d_{ij} = 1)(u_{ij} - u_{iJ}) \right] - \ln \left(1 + \sum_{j'=1}^{J-1} \exp(u_{ij'} - u_{iJ}) \right) \quad (10)$$

Independence of Irrelevant Alternatives (IIA)

- One of the properties of the multinomial logit model is **IIA**
- P_{ij}/P_{ik} does not depend upon what other alternatives are available:

$$\begin{aligned}\frac{P_{ij}}{P_{ik}} &= \frac{e^{u_{ij}} / \sum_{j'} e^{u_{ij'}}}{e^{u_{ik}} / \sum_{j'} e^{u_{ij'}}} \\ &= \frac{e^{u_{ij}}}{e^{u_{ik}}} \\ &= e^{u_{ij} - u_{ik}}\end{aligned}$$

Advantage of IIA

- IIA can simplify estimation. Instead of using as our likelihood

$$P_{ij} = \frac{\exp(u_{ij})}{\sum_{j'}^J \exp(u_{ij'})}, \quad (11)$$

we can use the conditional likelihood $P_i(j|j \in K)$ where $K < J$.

Use pair-wise binary logits.

- The log likelihood function is then (Begg and Gray, 1984):

$$L(\beta, \gamma | d_i \in K) = \sum_{i=1}^N \left[\sum_{j=1}^{K-1} (d_{ij} = 1)(u_{ij} - u_{iK}) \right] - \ln \left(1 + \sum_{j'=1}^K \exp(u_{ij'} - u_{iK}) \right)$$

Disadvantage of IIA

Most famously illustrated by the "red bus/blue bus problem"

- Consider a commuter with the choice set {ride a blue-colored bus, drive a car}
- Now add a red-colored bus to the choice set
- Assume that the only difference in utility between a red bus and a blue bus is in ϵ
- This will **double** the probability of taking a bus
- Why? $P(\text{blue bus})/P(\text{car})$ does not depend upon whether the red bus is available

We'll talk later about how to address this disadvantage

Expected Utility

- It is possible to move from the estimates of the utility function to expected utility
- (or at least differences in expected utility)
- Individual i is going to choose the best alternative
- Thus, expected utility from the best choice, V_i , is given by:

$$V_i = E \max_j (u_{ij} + \epsilon_{ij}) \quad = \text{log of denominator from previous slides.}$$

where the expectation is over all possible values of ϵ_{ij}

Expected Utility

- For the multinomial logit, this has a closed form:

$$V_i = \ln \left(\sum_{j=1}^J \exp u_{ij} \right) + C \quad (12)$$

where C is Euler's constant (a.k.a. Euler-Mascheroni constant)

because T1EV distribution is a limiting form of Gamma distribution

- We will use this later when we discuss dynamic discrete choice models

Alternative expression for expected utility

- Note that we can also express V_i as:

$$\begin{aligned} V_i &= \ln \left(\sum_{j=1}^J \frac{\exp(u_{iJ}) \exp(u_{ij})}{\exp(u_{iJ})} \right) + C \\ &= \ln \left(\sum_{j=1}^J \exp(u_{ij} - u_{iJ}) \right) + u_{iJ} + C \\ &= \ln \left(1 + \sum_{j=1}^{J-1} \exp(u_{ij} - u_{iJ}) \right) + u_{iJ} + C \end{aligned}$$

- This representation will become useful later in the course

From Expected Utility to Consumer Surplus

- We may want to transform utility into dollars to get consumer surplus
- We need something in the utility function (such as price) that is measured in dollars
- Suppose $u_{ij} = \beta_j X_i + \gamma Z_j - \delta p_j$
- The coefficient on price, δ then gives the utils-to-dollars conversion:

$$E(CS_i) = \frac{1}{\delta} \left[\ln \left(\sum_{j=1}^J \exp u_{ij} \right) + C \right] \quad (13)$$

Assuming delta does not change over policies.

- We can calculate the change in consumer surplus after a policy change as $E(CS_{i2}) - E(CS_{i1})$ where the C 's cancel out => gives us welfare effect

References

Begg, C. B. and R. Gray (1984). "Calculation of Polychotomous Logistic Regression Parameters Using Individualized Regressions". In: *Biometrika* 71.1, pp. 11-18. DOI: [10.1093/biomet/71.1.11](https://doi.org/10.1093/biomet/71.1.11).

Train, K. (2009). *Discrete Choice Methods with Simulation*. 2nd ed. Cambridge; New York: Cambridge University Press. ISBN: 9780521766555.