Beyond the IIA Assumption

Matteo Paradisi
(EIEF)

Applied Micro - Lecture 11

Independence of Irrelevant Alternative (IIA)

- IIA is a common assumption to all models we have seen so far
- The critical assumption is that error terms are i.i.d.
- This was not assumed in multinomial probit, where we estimated covariances
- In multinomial logit, we assumed that the variance-covariance matrix is diagonal
- What is the implication of this assumption?

$$\frac{\Pr\left(y_i = j | x_i\right)}{\Pr\left(y_i = k | x_i\right)} = \frac{e^{x_{ij}\beta_j}}{e^{x_{ik}\beta_k}}$$

- The relative probability does not depend on any other alternative

Independence of Irrelevant Alternative (IIA)

Example (McFadden, 1974):

- Suppose we have three equally distributed transportation categories:
  - Blue bus ($P = 33\%$), Car ($P = 33\%$), Red bus ($P = 33\%$)
- Now, we paint the red busses blue. We now have two choices
- With IIA: Blue bus ($P = 50\%$), Car ($P = 50\%$)
- However, it seems more natural to have: Blue bus ($P = 66\%$), Car ($P = 33\%$)
- Economic interpretation: if alternative categories can serve as substitutes, then the results of MNL may not be very realistic.

The IIA in the Conditional Logit

- Let's see McFadden example in conditional logit
- Assume utility for choices

$$U_{ij} = X_{ij}'\beta + \varepsilon_{ij}$$

- Now, let's assume that people are indifferent bw the two buses
- In the model

$$U_{i,\text{red bus}} = U_{i,\text{blue bus}}$$

- How do we break the tie?
- We assume that choice between the two is random
- Explicitly $X_{i,\text{red bus}} = X_{i,\text{blue bus}} = X_{i,\text{bus}}$

The IIA in the Conditional Logit

- The probability of bus over car is

$$\Pr\left(y_i = \text{Bus}\right) = \frac{e^{X'_{i,bus}\beta}}{e^{X'_{i,bus}\beta} + e^{X'_{i,car}\beta}}$$

- Also,

$$\Pr\left(y_i = \text{RedBus} | y_i = \text{Bus}\right) = \frac{1}{2}$$

- Conditional Logit Implies

$$\Pr\left(y_i = \text{Car} | y_i = \text{Car or RedBus}\right) = \frac{e^{X'_{i,car}\beta}}{e^{X'_{i,redbus}\beta} + e^{X'_{i,car}\beta}}$$

- it does not depend on the presence of BlueBus
- However, presumably taking away the blue bus choice would lead all the current blue bus users to shift to the red bus, and not to cars.

The IIA and Unrealistic Substitution Patterns

- IIA can be thought of as the presence of unrealistic substitution patterns
- Example: restaurant choice bw Chez Panisse (C), Lalime's (L), and the Bongo Burger (B)
- Two characteristics: price ($P_C = 95$, $P_L = 80$, $P_B = 5$), quality $Q_C = 10$, $Q_L = 9$, $Q_B = 2$
- Market shares: $S_C = 0.10$, $S_L = 0.25$, $S_B = 0.65$
- Utility associated with i and j is

$$U_{ij} = -0.2P_j + 2Q_j + \varepsilon_{ij}$$

The IIA and Unrealistic Substitution Patterns

- Suppose L exits the market, or its price goes to infinity
- Prediction of conditional logit: $S'_C = 0.13$, $S'_B = 0.87$
- That seems implausible!
- People planning to go to Lalime's more likely go to Chez Panisse if Lalime's is closed
- Hence, one would expect $S'_C = 0.35$, $S'_B = 0.65$
- While model predicts most of those who would have gone to L will now dine at B

IIA: Possible Solutions

Many different ways to relax IIA have been proposed

- Multinomial Probit
  - allows for covariance in residuals
- Nested Logit
  - create nests: groups of choices
- Ordered multinomial choice models
  - only work when can order choices
- Random Coefficients Logit
  - allow for different βs across individuals
- Extension of Random Coefficients Logit: BLP
  - allows for unobservable choice characteristics

The Nested Logit

One Solution: Nested Logit

- Think about the choice of a product
- You can model it as a sequence of choices
- e.g. first, decide whether to buy, then decide category, then single product
- Draw a decision tree to guide model construction

One Solution: Nested Logit

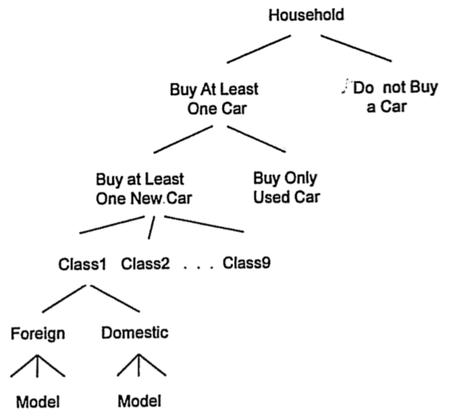

FIGURE 1.—Automobile choice model.

One Solution: Nested Logit

- Suppose we have data on car purchases
- There are 6 models: A, B, C, D, E, F
- Divide them in 3 categories
  - sport cars (A and B)
  - station wagons (C and D)
  - off-roads (E and F)
- Assume that people first choose the type of car, and then the model

Nested Logit

- Suppose there are J choices
- Divided in K groups N_k
- The nested logit assumes

$$F\left(\xi_{i,1},\xi_{i,2},\ldots,\xi_{i,J}\right)=\exp\left[-\sum_{q=1}^{K}\left(\sum_{j\in N_q}e^{\frac{\xi_{i,j}}{\sigma_q}}\right)^{\sigma_q}\right]$$

- σ_q is a dissimilarity parameter: it tells us the degree of dissimilarity between alternatives within the nest

Nested Logit

- Let's write down the probability of choice $j \in N_k$ conditional on being within the nest

$$\Pr\left(y_i = j | j \in N_k\right) = \frac{e^{\frac{1}{\sigma_k} x_{ij}\beta_j}}{\sum_{s\in N_k} e^{\frac{1}{\sigma_k} x_{is}\beta_s}}$$

- Now, the probability of choosing nest N_k is

$$\Pr\left(y_i \in N_k\right) = \frac{\left(\sum_{s\in N_k} e^{\frac{1}{\sigma_k} x_{is}\beta_s}\right)^{\sigma_k}}{\sum_{m=1}^{K}\left(\sum_{s\in N_m} e^{\frac{1}{\sigma_m} x_{is}\beta_s}\right)^{\sigma_m}}$$

- Hence, the unconditional probability of alternative j is

$$\Pr\left(y_i = j\right) = \Pr\left(y_i \in N_k\right)\Pr\left(y_i = j | j \in N_k\right)$$
$$= \frac{\left(\sum_{s\in N_k} e^{\frac{1}{\sigma_k} x_{is}\beta_s}\right)^{\sigma_k}}{\sum_{m=1}^{K}\left(\sum_{s\in N_m} e^{\frac{1}{\sigma_m} x_{is}\beta_s}\right)^{\sigma_m}} \frac{e^{\frac{1}{\sigma_k} x_{ij}\beta_j}}{\sum_{s\in N_k} e^{\frac{1}{\sigma_k} x_{is}\beta_s}}$$

Nested Logit

- The model also allows for variables affecting all alternatives within a nest in the same way
- We call these variables w with coefficient γ
- We have

$$\Pr\left(y_i = j\right) = \frac{e^{\sigma_k w \gamma_k}\left(\sum_{s \in N_k} e^{\frac{1}{\sigma_k} x_{is}\beta_s}\right)^{\sigma_k}}{\sum_{m=1}^{K} e^{\sigma_m w \gamma_m}\left(\sum_{s \in N_m} e^{\frac{1}{\sigma_m} x_{is}\beta_s}\right)^{\sigma_m}} \frac{e^{\frac{1}{\sigma_k} x_{ij}\beta_j}}{\sum_{s \in N_k} e^{\frac{1}{\sigma_k} x_{is}\beta_s}}$$

Nested Logit: IIA

- IIA holds within nest

$$\frac{\Pr(y_i = j | j \in N_k)}{\Pr(y_i = h | h \in N_k)} = \left(\frac{e^{x_{ij}\beta_j}}{e^{x_{ih}\beta_h}}\right)^{\frac{1}{\sigma_k}}$$

- But does not hold across nests

$$\frac{\Pr(y_i = j | j \in N_k)}{\Pr(y_i = h | h \in N_m)} = \frac{\left(\sum_{s \in N_k} e^{\frac{1}{\sigma_k} x_{is}\beta_s}\right)^{\sigma_k} \sum_{s \in N_m} e^{\frac{1}{\sigma_m} x_{is}\beta_s}}{\left(\sum_{s \in N_m} e^{\frac{1}{\sigma_m} x_{is}\beta_s}\right)^{\sigma_m} \sum_{s \in N_k} e^{\frac{1}{\sigma_k} x_{is}\beta_s}} \frac{e^{\frac{1}{\sigma_k} x_{ij}\beta_j}}{e^{\frac{1}{\sigma_m} x_{ih}\beta_h}}$$

Ordered Response Models

Ordered Response Models

- One alternative way to deal with the strong IIA is to change the assumptions about the data generating process
- Hence, the process that leads observations into the various categories
- We study a class of models that can be used when multinomial data can be ordered
- Obviously, this is implementable only if the order is meaningful
- Example: survey data where satisfaction goes from (1=extremely unsatisfied) to (5=very satisfied)

Ordered Response Models

- Advantage: we have a single latent variable

$$y^* = x\beta + u$$

- We observe $y = \{0, 1, 2, \ldots, J\}$

- Assumption is that data generating process follows

$$y = \begin{cases} 0 & \text{if } y^* \leq \alpha_1 \\ 1 & \text{if } \alpha_1 < y^* \leq \alpha_2 \\ \vdots & \\ J & \text{if } y^* > \alpha_J \end{cases}$$

- $\alpha_1, \alpha_2, \ldots, \alpha_J$ is the vector of parameters to estimate

Ordered Response Models

- Write the probability of each choice

$$
\begin{aligned}
\Pr(y_i = 0|x_i) &= \Pr(x_i\beta + u_i \leq \alpha_1|x_i) \\
&= \Pr(u_i \leq \alpha_1 - x_i\beta|x_i) \\
\Pr(y_i = 1|x_i) &= \Pr(\alpha_1 - x_i\beta < u_i \leq \alpha_2 - x_i\beta|x_i) \\
&\vdots \\
\Pr(y_i = 1|x_i) &= \Pr(\alpha_J - x_i\beta < u_i|x_i)
\end{aligned}
$$

- In order to estimate the model we must make assumptions on the distribution of u
- Important feature is that there is only one error term
- Hence, we do not need to worry about correlation between alternatives

Ordered Probit

- Let's assume $u \sim N(0,1)$
- Probabilities are

$$
\begin{aligned}
\Pr(y_i = 0 | x_i) &= \Phi(\alpha_1 - x_i\beta) \\
\Pr(y_i = 1 | x_i) &= \Phi(\alpha_2 - x_i\beta) - \Phi(\alpha_1 - x_i\beta) \\
&\vdots \\
\Pr(y_i = J | x_i) &= 1 - \Phi(\alpha_J - x_i\beta)
\end{aligned}
$$

Setup the Likelihood

- The individual log-likelihood contribution is

$$\begin{aligned}\ell_i(\alpha,\beta) = {} & 1(y_i=0)\ln\Phi(\alpha_1 - x_i\beta) \\ & + 1(y_i=1)\ln[\Phi(\alpha_2 - x_i\beta) - \Phi(\alpha_1 - x_i\beta)] \\ & + \ldots + 1(y_i=J)[1 - \ln\Phi(\alpha_J - x_i\beta)]\end{aligned}$$

- Hence the log-likelihood is

$$L(\alpha,\beta) = \sum_{i=1}^{N} \ell_i(\alpha,\beta)$$

Ordered Probit: Marginal Effects

- Like in the other probit models

$$\frac{\partial \Pr(y_i = 0|x_i)}{\partial x_{ik}} = -\beta_k \varphi(\alpha_1 - x_i\beta)$$

$$\frac{\partial \Pr(y_i = j|x_i)}{\partial x_{ik}} = -\beta_k \left[\varphi(\alpha_{j-1} - x_i\beta) - \varphi(\alpha_j - x_i\beta)\right]$$

$$\vdots$$

$$\frac{\partial \Pr(y_i = J|x_i)}{\partial x_{ik}} = \beta_k \varphi(\alpha_J - x_i\beta)$$

- So for the first and last choice the sign of β is the sign of the marginal effect
- For intermediate choices, the sign is ambiguous
- Important: β is the effect on a latent variable, which is more interpretable here than in the standard probit

Random Coefficients Logit and BLP

Random Coefficients Logit: Intuition

- allow for unobserved heterogeneity in the slope coefficients
- Why we think that if Lalime's price goes up, its clients will go to Chez Panisse?
- We think individuals with taste for L, likely to have a taste for close substitutes in terms of observables
- Chez Panisse as well, rather than for the Bongo Burger.

Random Coefficients Logit

- Model utilities as

$$U_{ij} = X_{ij}'\beta_i + \varepsilon_{ij}$$

- ε_{it} independent of everything else, i.i.d., and either extreme value, or normal
- Rewrite as

$$U_{ij} = X_{ij}'\bar{\beta} + \nu_{ij}$$

- with

$$\nu_{ij} = \varepsilon_{ij} + X_{ij} \cdot (\beta_i - \bar{\beta})$$

Random Coefficients Logit

$$\nu_{ij} = \varepsilon_{ij} + X_{ij} \cdot (\beta_i - \bar{\beta})$$

- Notice that this term is not independent across choices!
- Hence, we have relaxed the IIA assumption

Random Coefficients Logit

- How do we estimate the model?

$$\nu_{ij} = \epsilon_{ij} + X_{ij} \cdot (\beta_i - \bar{\beta})$$

- Solution 1: assume finite number of individual types

$$\beta_i \in \{b_0, b_1, \ldots, b_K\}$$

with

$$\Pr(\beta_i = b_k | Z_i) = p_k \text{ or } \Pr(\beta_i = b_k | Z_i) = \frac{e^{Z_i' \gamma_k}}{1 + \sum_{h=1}^{K} e^{Z_i' \gamma_h}}$$

- Solution 2: specify distribution

$$\beta_i | Z_i \sim N(Z_i' \gamma, \Sigma)$$

BLP Models

- This is a class of models introduced by Berry Levinsohn and Pakes
- Extends random coefficients to allow for
  - unobserved product characteristics,
  - endogeneity of choice characteristics,
  - allows for consistent estimation without individual level choice data, needs market shares
- This is mostly used in Industrial Organization
- Model demand for differentiated products when there is a large number of products

BLP Intuition

- Three dimensions: products j, markets t, and individuals i
- Only one purchase per individual
- Random coefficients utility

$$U_{ijt} = X'_{jt}\beta_i + \zeta_{jt} + \varepsilon_{ij}$$

- ζ_{jt}: unobserved product characteristic, can vary by product and market
- ζ_{jt} can include product and market dummy $\zeta_{jt} = \zeta_j + \zeta_t$
- Assume ε_{ij} has type I extreme distribution, iid across i, j, t

BLP Intuition

- Assume the β_is follow the structure

$$\beta_i = \beta + Z_i' \Gamma + \eta_i$$

- where

$$\eta_i | Z_i \sim N(0, \Sigma)$$

- Z_is are normalized to have mean 0 so that βs are average marginal utilities
- For estimation, need estimates of distribution of Z_i and market share for j, t combinations

BLP Intuition

- Write utility as
$$U_{ijt} = \delta_{jt} + \nu_{ijt} + \epsilon_{ijt}$$
where
$$\delta_{jt} = \beta' X_{jt} + \zeta_{jt} \text{ and } \nu_{ijt} = \left(Z_i' \Gamma + \eta_i\right)' X_{jt}$$

- simple case
$$s_{jt}\left(\delta_{jt}, \Gamma = 0, \Sigma = 0\right) = \frac{e^{\delta_{jt}}}{\sum_{h=0}^{J} e^{\delta_{ht}}}$$

- When restrictions do not hold, draw Z_i from empirical distribution in market t, draw from distribution of η_i, compute purchase probability. Repeat to find market share.
- Find δ_{jt} to match observed market shares
- Unobserved product characteristics are $\zeta_{jt} = \delta_{jt}\left(s, \Gamma, \Sigma\right) - \beta' X_{jt}$
- Exploit exogeneity of ζ_{jt} to then estimate βs with GMM