EC708 Discussion 9
Mixed Logit

Yan Liu[1]

Department of Economics
Boston University

April 2, 2021

[1] Parts of the materials are from Train (2009).

Outline

Mixed Logit

A mixed logit model is any model whose choice probabilities can be expressed as integrals of standard logit probabilities over a density of parameters:

$$P_{tj} = \int L_{tj}(\beta) f(\beta) d\beta, \quad L_{tj}(\beta) = \frac{e^{V_{tj}(\beta)}}{\sum_{\ell=1}^{J} e^{V_{t\ell}(\beta)}}.$$

- If utility is linear in β, then $V_{tj}(\beta) = X_{tj}'\beta$, and

$$P_{tj} = \int \left(\frac{X_{tj}'\beta}{\sum_{\ell=1}^{J} X_{t\ell}'\beta} \right) f(\beta) d\beta.$$

- If $f(\beta)$ is discrete, mixed logit becomes latent class model.
- In most applications, $f(\beta)$ is continuous, e.g. normal, lognormal, etc.

Random Coefficients

The mixed logit probability can be derived from random coefficients:

$$U_{tj} = X'_{tj}\beta_t + u_{tj},$$

where u_{tj} is i.i.d. extreme value and $\beta_t \sim f(\beta)$.

The choice probability conditional on β_t is

$$L_{tj}(\beta_t) = \frac{e^{X'_{nj}\beta_t}}{\sum_{\ell=1}^{J} e^{X'_{n\ell}\beta_t}}.$$

The unconditional choice probability is

$$P_{tj} = \int \left(\frac{e^{X'_{nj}\beta}}{\sum_{\ell=1}^{J} e^{X'_{n\ell}\beta}} \right) f(\beta) d\beta.$$

Substitution Patterns

Mixed logit does not exhibit independence of irrelevant alternatives (IIA).

The elasticity of P_{tj} with respect to mth attribute of X_{tk} is

$$E_{tjX_{tk}^m} = -X_{tk}^m \int \beta^m \left[\frac{L_{tj}(\beta)}{P_{tj}} \right] L_{tk}(\beta) f(\beta) d\beta.$$

- $E_{tjX_{tk}^m}$ depends on the correlation between $L_{tj}(\beta)$ and $L_{tk}(\beta)$ over different values of β.
- Recall that in standard logit, the elasticity is the same for all j:

$$E_{tjX_{tk}^m} = -X_{tk}^m \beta^m P_{tk}.$$

Such proportional substitution is a manifestation of IIA property.

Approximation to Any Random Utility Model

McFadden and Train (2000) show that any random utility model (RUM) can be approximated by a mixed logit. Suppose the true RUM is

$$U_{tj} = Z'_{tj}\alpha_t.$$

- Z_{tj}: variables related to alternative j;
- α follows any distribution $f(\alpha)$.

Choice probability conditional on α is

$$q_{tj}(\alpha) = 1\{Z'_{tj}\alpha_t > Z'_{t\ell}\alpha_t \; \forall j \neq \ell\}.$$

The unconditional choice probability is

$$Q_{tj} = \int 1\{Z'_{tj}\alpha_t > Z'_{t\ell}\alpha_t \; \forall j \neq \ell\} f(\alpha) d\alpha.$$

Approximation to Any Random Utility Model

We can approximate the unconditional choice probability with a mixed logit.

1. Scale utility by λ so that $U^*_{tj} = Z'_{tj}(\alpha_t/\lambda)$.
2. Add an i.i.d. extreme value term u_{tj}.
3. The mixed logit probability based on this utility is

$$P_{tj} = \int \frac{e^{Z'_{tj}(\alpha_t/\lambda)}}{\sum_\ell e^{Z'_{t\ell}(\alpha_t/\lambda)}} f(\alpha) d\alpha.$$

As $\lambda \to 0$, α_t/λ grow large, and P_{tj} approaches a 1-0 indicator for the alternative with the highest utility, i.e. the true Q_{tj}.

Endogeneity

In many situations, explanatory variables X_t and unobserved factors u_t are not independent.

- Unobserved attributes of a product can affect its price.
  - Unobserved attributes are costly or can affect demand.
- Marketing efforts can be related to prices.
  - Advertising and sales promotions are not measured by researchers.
- Interrelated choices of decision makers.
  - Example: choices of travel mode and housing location. Observed travel time by public transit and unobserved attitudes toward public transit (reflected in housing location) are negatively correlated.

Endogeneity

Several methods have been developed to estimate choice models in the presence of endogeneity.

- Control function approach: two-step procedure
- Full maximum likelihood approach
- **BLP approach** developed by Berry, Levinsohn, and Pakes (1995): instrumental variables estimation
  - Initially designed to deal with endogenous prices and aggregate (product-level) data.
  - Use the contraction to take endogeneity out of the nonlinear choice model and put into a linear regression model.

The BLP Approach

Assume that utility takes the form

$$U_{tj} = V(p_j, x_j, s_t, \beta_t) + \xi_j + u_{tj},$$

- s_t: a vector of demographic characteristics;
- $V(\cdot)$: function of observed variables and consumer tastes β_t;
- ξ_j: average/common utility that consumers obtain from unobserved attributes of product j;
- u_{tj}: i.i.d. extreme value.

Basic issue: p_j depends on ξ_j.

The BLP Approach

Decompose $V(\cdot) = \bar{V}(p_j, x_j, \bar{\beta}) + \tilde{V}(p_j, x_j, s_t, \tilde{\beta}_t)$. Then

$$U_{tj} = \underbrace{\bar{V}(p_j, x_j, \bar{\beta}) + \xi_j}_{\delta_j} + \tilde{V}(p_j, x_j, s_t, \tilde{\beta}_t) + u_{tj}$$

Endogeneity ξ_j is subsumed into a product-specific constant δ_j such that it is no longer part of the unobserved component of utility.

Given $\tilde{\beta}_t \sim f(\tilde{\beta}|\theta)$, market shares are given by a mixed logit

$$S_j(\delta, \theta) = \int \left[\frac{e^{\delta_j + \tilde{V}(p_j, x_j, s_t, \tilde{\beta})}}{\sum_\ell e^{\delta_\ell + \tilde{V}(p_\ell, x_\ell, s_t, \tilde{\beta})}} \right] f(\tilde{\beta}|\theta) d\tilde{\beta}.$$

$S_j(\delta, \theta)$ does not have a closed form!

The Contraction

BLP provided an algorithm for estimating the constants $\delta_j \; \forall j$ quickly.

1. Replace $f(\tilde{\beta}|\theta)$ with empirical distribution from ns random draws:

$$\hat{S}_j(\delta, \theta) = \frac{1}{ns} \sum_{i=1}^{ns} \frac{e^{\delta_j + \tilde{V}(p_j, x_j, s_t, \tilde{\beta}_i)}}{\sum_\ell e^{\delta_\ell + \tilde{V}(p_\ell, x_\ell, s_t, \tilde{\beta}_i)}}, \quad (\tilde{\beta}_1, \ldots, \tilde{\beta}_{ns}) \sim f(\tilde{\beta}|\theta).$$

2. At each trial value of θ, adjust the constants iteratively by

$$\delta_j^{t+1} = \delta_j^t + \ln(S_j) - \ln[\hat{S}_j(\delta^t, \theta)]$$

   where S_j are actual market shares.

- Berry (1994) showed uniqueness of δ_j's.
- Berry et al. (1995) showed that the iterative adjustment process is a contraction that guarantees convergence.
- To increase efficiency, can use importance sampling for $\hat{S}_j(\delta, \theta)$.

Estimation by GMM

Denote instruments as z_j. Then $E[\xi_j|z_j] = 0$. The GMM objective is

$$\left(\frac{1}{J} \sum_{j=1}^{J} \xi_j(\bar{\beta}, \theta) z_j \right)' W \left(\frac{1}{J} \sum_{j=1}^{J} \xi_j(\bar{\beta}, \theta) z_j \right)$$

where $\xi_j(\bar{\beta}, \theta) = \delta_j(\theta) - \bar{V}(p_j, x_j, \bar{\beta})$.

- Berry et al. (1995) propose to use as z_j the average nonprice attributes of own products and rival products.
- Under linearity $\bar{V}(p_j, x_j, \bar{\beta}) = x_j'\gamma - \alpha p_j$ and normality $\tilde{\beta} \sim N(0, \sigma^2)$, can concentrate out (α, γ) and search over σ using the Nelder-Mead nonderivative "simplex" search routine (fminsearch in Matlab).

Bibliography

Berry, S., Levinsohn, J., and Pakes, A. (1995), "Automobile prices in market equilibrium," *Econometrica: Journal of the Econometric Society*, 841–890.

Berry, S. T. (1994), "Estimating discrete-choice models of product differentiation," *The RAND Journal of Economics*, 242–262.

McFadden, D. and Train, K. (2000), "Mixed MNL models for discrete response," *Journal of applied Econometrics*, 15, 447–470.

Train, K. E. (2009), *Discrete Choice Methods with Simulation*, Cambridge University Press, 2nd ed.