EC708 Discussion 10
LDV

Yan Liu

Department of Economics
Boston University

April 1, 2022

Outline

Mixed Logit
Definition

Specify the utility of individual t from alternative j as

$$U_{tj} = V_{tj}(\beta) + u_{tj},$$

where $u_{tj}, j = 1, \ldots, J$ is i.i.d. type-I extreme value across t. Evaluated at parameter value β, the choice probability is the standard logit probability:

$$L_{tj}(\beta) = \frac{e^{V_{tj}(\beta)}}{\sum_{\ell=1}^{J} e^{V_{t\ell}(\beta)}}.$$

A mixed logit model is any model whose choice probabilities can be expressed as integrals of $L_{tj}(\beta)$ over a density $f(\cdot)$ of parameters:

$$P_{tj} = \int L_{tj}(\beta) f(\beta) d\beta.$$

Mixed Logit
Definition

- If $f(\beta)$ is discrete, mixed logit becomes latent class model.
- In most applications, $f(\beta)$ is continuous, e.g. normal, lognormal, uniform, triangular, gamma, etc.
- Usually, researchers are interested in estimating parameters that describe $f(\beta)$, denoted by θ. We can write $f(\beta|\theta)$.
- β's are similar to u_{tj}: both are random terms that are integrated out to obtain the choice probability.

Mixed Logit

Random Coefficients

The mixed logit probability can be derived from random coefficients:

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

where $u_{tj}, j = 1, \ldots, J$ is i.i.d. type-I extreme value across t 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

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

Mixed Logit

Random Coefficients

The coefficients β_t can be decomposed into their mean α and deviations μ_t, so that

$$U_{tj} = X'_{tj}\alpha + X'_{tj}\mu_t + u_{tj}.$$

The random portion of utility is $\eta_{tj} = X'_{tj}\mu_t + u_{tj}$, which can be correlated over alternatives:

$$\text{Cov}(\eta_{tj}, \eta_{tk}) = E(X'_{tj}\mu_t + u_{tj})(X'_{tk}\mu_t + u_{tk}) = X'_{tj}E[\mu_t\mu'_t]X_{tk}.$$

An analog to nested logit is obtained by setting $X'_{tj}\mu_t = \sum_{k=1}^{K} \mu_{tk}d_{jk}$, where $d_{jk} = 1$ if j is in nest k and zero otherwise.

Mixed Logit

Substitution Patterns

Mixed logit does not exhibit

- independence of irrelevant alternatives (IIA): P_{tj}/P_{tk} depends on attributed of all alternatives.
- restrictive substitution patterns of logit
    - The elasticity of P_{tj} with respect to mth attribute of X_{tk} is

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

    - $E_{tjX_{tk}^m}$ 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:

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

Mixed Logit

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

$$U_{tj} = X'_{tj}\alpha_t, \quad \alpha_t \sim f(\alpha).$$

Choice probability conditional on α_t is

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

The unconditional choice probability is

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

Mixed Logit

Approximation to Any Random Utility Model

We can approximate the true Q_{tj} with a mixed logit.

1. Scale utility by λ so that $U_{tj}^* = X_{tj}'(\alpha_t/\lambda)$.

2. Add an i.i.d. extreme value term u_{tj}.

3. The mixed logit probability based on this utility is

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

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}.

Outline

BLP Approach

Endogeneity

In many situations, explanatory variables X_{tj} that enter a discrete choice model are not independent of unobserved factors u_{tj}.

- 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.

BLP Approach

Endogeneity

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

- Control function approach: two-step procedure
- Full maximum likelihood approach
- **BLP approach** developed by Berry (1994) and Berry, Levinsohn, and Pakes (1995): take endogeneity out of the nonlinear choice model and put into a linear regression model

BLP Approach

Specification

Assume that utility takes the form

$$U_{tj} = X_j'\beta_t - \alpha_t p_j + \xi_j + u_{tj},$$

- X_j: observed nonprice attributes of product j;
- p_j: price of product j;
- ξ_j: average/common utility from unobserved attributes of product j;
- u_{tj}: i.i.d. extreme value.

Basic issue: p_j depends on $\xi_j \Rightarrow$ endogeneity.

Additional issue: allow for random coefficients using aggregate data (product-level).

BLP Approach

Aggregate Market Shares

Assume $(\alpha_t, \beta_t')' \sim$ i.i.d. $N((\bar{\alpha}, \bar{\beta}')', \Sigma)$. Decompose

$$U_{tj} = \underbrace{X_j'\bar{\beta} - \bar{\alpha}p_j + \xi_j}_{\delta_j} + X_j'(\beta_t - \bar{\beta}) - (\alpha_t - \bar{\alpha})p_j + u_{tj}$$

- 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.

- Let $\theta = (\bar{\alpha}, \bar{\beta}, \Sigma)$. Then, market shares are given by a mixed logit

$$S_j(\delta, \theta) = \int \left[\frac{e^{\delta_j + X_j'(\beta_t - \bar{\beta}) - (\alpha_t - \bar{\alpha})p_j}}{\sum_\ell e^{\delta_\ell + X_j'(\beta_t - \bar{\beta}) - (\alpha_t - \bar{\alpha})p_j}} \right] dF(\alpha_t, \beta_t | \theta).$$

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

BLP Approach

Inner Loop: The Contraction

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

1. Approximate $S_j(\delta, \theta)$ using S simulation draws from $N((\bar{\alpha}, \bar{\beta}')', \Sigma)$:

$$\hat{S}_j(\delta, \theta) = \frac{1}{S} \sum_{s=1}^{S} \frac{e^{\delta_j + X_j'(\beta^s - \bar{\beta}) - (\alpha^s - \bar{\alpha})p_j}}{\sum_\ell e^{\delta_\ell + X_j'(\beta^s - \bar{\beta}) - (\alpha^s - \bar{\alpha})p_j}}.$$

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 et al. (1995) showed that the iterative adjustment process is a contraction that guarantees convergence.

BLP Approach

Outer Loop: Estimation by GMM

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

$$\left(\frac{1}{J} \sum_{j=1}^{J} \xi_j(\theta) Z_j \right)' W \left(\frac{1}{J} \sum_{j=1}^{J} \xi_j(\theta) Z_j \right)$$

where $\xi_j(\theta) = \delta_j(\theta) - (X_j'\bar{\beta} - \bar{\alpha}p_j)$.

- Berry et al. (1995) propose to use as Z_j characteristics of rival products.
- Intuition:
  - Oligopolistic competition makes firms set price as a function of characteristics of rival products.
  - Characteristics of rival products should not affect households' valuation.
- Can concentrate out $(\bar{\alpha}, \bar{\beta})$ and search over Σ using the Nelder-Mead nonderivative "simplex" search routine (fminsearch in Matlab).

Outline

Semiparametric Estimation

Single-Index Model

We consider binary response models of the form

$$P(Y_t = 1 | X_t) = G(X_t'\beta).$$

When G is a CDF, index models can be derived from an underlying latent variable model:

$$Y_t^* = X_t'\beta + u_t, \quad Y_t = 1\{Y_t^* > 0\}.$$

Can we estimate β if we assume u_t is independent of X_t but the CDF G of u_t is unknown?

Semiparametric Estimation

Single-Index Model

- Without knowledge of G, the model is semiparametric with
    - a finite-dimensional parameter of interest (β) and
    - an infinite-dimensional nuisance parameter (function G).
- There are several semiparametric estimators of β, up to scale, that are consistent and $\sqrt{T}$-asymptotically normal. E.g. Powell, Stock, and Stoker (1989), Ichimura (1993), and Klein and Spady (1993).
- Once $\hat{\beta}$ is obtained, G can be consistently estimated using nonparametric regression of Y_t on $X_t'\hat{\beta}$.
- Thus, the response probabilities and the partial effects on these probabilities can be consistently estimated.

Semiparametric Estimation

Maximum Score Estimator

What if we further relax the assumption that u_t is independent of X_t?
Manski (1975) defines $\hat{\beta}_T$ to maximize the predictive score function

$$S_T(\beta) = \sum_{t=1}^{T} Y_t 1\{X_t'\beta > 0\} + (1 - Y_t)1\{X_t'\beta \leq 0\}.$$

- To ensure consistency, need the median of u_t given X_t to be zero.
- $\hat{\beta}_T$ can be interpreted as a least absolute deviations estimator:

$$\hat{\beta}_T = \arg \min_{\beta} \sum_{t=1}^{T} |Y_t - 1\{X_t'\beta > 0\}|.$$

This led to the extension of the maximum score idea to more general quantile estimation of β.

Semiparametric Estimation

Maximum Score Estimator

Caveats:

- The asymptotic distribution of $\hat{\beta}_T$ is nonnormal because the median regression function is flat except at its discontinuity points.

- In fact, the convergence rate of $\hat{\beta}_T$ is $T^{1/3}$ (Kim and Pollard, 1990).

- Maximum score estimation does not allow estimation of the response probabilities and the APEs because the unconditional distribution of u_t is not identified.

Semiparametric Estimation

Maximum Score Estimator

To obtain a faster convergence rate, Horowitz (1992) replaces the conditional median function $1\{X_t'\beta > 0\}$ by a "smoothed" version:

$$S_T^*(\beta) = \sum_{t=1}^{T} Y_t K(X_t'\beta/h_T) + (1 - Y_t)(1 - K(X_t'\beta/h_T)).$$

- $K(\cdot)$ is continuous on $[0, 1]$ with $K(u) \to 0$ or 1 as $u \to -\infty$ or ∞.

- h_T is a sequence of bandwidths which tends to zero as T increases.

Horowitz (1992) shows the minimizer of $S_T^*(\beta)$ is asymptotically normal. The convergence rate can be made at least $T^{2/5}$ and arbitrarily close to $T^{1/2}$.

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.

Horowitz, J. L. (1992), "A smoothed maximum score estimator for the binary response model," *Econometrica: journal of the Econometric Society*, 505–531.

Ichimura, H. (1993), "Semiparametric least squares (SLS) and weighted SLS estimation of single-index models," *Journal of econometrics*, 58, 71–120.

Kim, J. and Pollard, D. (1990), "Cube root asymptotics," *The Annals of Statistics*, 191–219.

Klein, R. W. and Spady, R. H. (1993), "An efficient semiparametric estimator for binary response models," *Econometrica: Journal of the Econometric Society*, 387–421.

Manski, C. F. (1975), "Maximum score estimation of the stochastic utility model of choice," *Journal of econometrics*, 3, 205–228.

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

Powell, J. L., Stock, J. H., and Stoker, T. M. (1989), "Semiparametric estimation of index coefficients," *Econometrica: Journal of the Econometric Society*, 1403–1430.