Estimation of Demand Model 2

BLP Approach

Yuta Toyama

Last updated: November 26, 2018

Overview

- Last week: Discrete choice model when individual data is available.
- This week: Discrete choice model when only market-level data is available.
  - So-called "BLP" approach.

Data Structure

- Consumer level data
  - cross section of consumers
  - sometimes: panel (i.e., repeated choices)
  - sometimes: second choice data
- Market-level data
  - cross section/time series/panel of markets
- Combination
  - sample of consumers plus market-level data
  - quantity/share by demographic groups
  - average demographics of purchasers of good j

Consumer-level Data

- See match between consumers and their choices
- Data include:
  - consumer choices (including choice of outside good)
  - prices/characteristics/advertising of all options
  - consumer demographics
- Advantages:
  - impact of demographics
  - identication and estimation
  - dynamics (especially if we have panel)
- Disadvantages
  - harder/more costly to get
  - sample selection and reporting error

Market-level Data

- We see product-level quantity/market shares by "market"
- Data include:
  - aggregate (market-level) quantity
  - prices/characteristics/advertising
  - denition of market size
  - distribution of demographics
    - sample of actual consumers
    - data to estimate a parametric distribution
- Advantages:
  - easier to get
  - sample selection less of an issue
- Disadvantages:
  - estimation often harder and identication less clear

Estimation of Discrete Choice Model using Aggregate Data

- Rough idea:
  - Discrete choice model implies choice probability.
  - (Integral of) choice probability = market share
- Key difference: econometric error term:
  - With consumer-level data: ϵ_{ij} as econometric error term
  - With market-level data: ξ_{jt} as econometric error term
    - ξ_{jt} rationalizes the observed market share s_{jt}.
- We will see three cases
  - multinomial logit model (Berry 1994)
  - Nested logit model (Berry 1994)
  - Random coefficient logit model (BLP 1995, Nevo 2001)

Berry (1994) Multinomial Logit

- The utility for individual i in market t from purchasing product $j = 1, \cdots J_t$:

$$\begin{aligned} u_{ij} &= \underbrace{X_{jt}\beta - \alpha p_{jt} + \xi_{jt}} + \varepsilon_{ijt}, \\ &= \delta_{jt} + \varepsilon_{ijt}. \end{aligned}$$

  - X_{jt}: a vector of product characteristics for product j in market t
  - β: a vector of coefficients
  - α: a price coefficient
  - p_{jt}: price for product j in market t
  - ξ_{jt}: consumers' valuation of an **unobserved** product characteristics
  - ε_{ijt}: i.i.d. utility shock across consumers and choices
  - δ_{jt}: the mean utility of product j
  - Normalize $u_{i0t} = \epsilon_{i0t}$ (meaning $\delta_{0t} = 0$)
- Note: No consumer heterogeneity except for ϵ_{ijt}

- Each consumer purchases at most one product which gives the highest utility.
- Let M denote the market size (= the number of consumers).
- Given the density $f(\varepsilon)$, the market demand for product j is

$$
\begin{aligned}
q_{jt} &= Ms_{jt} \\
s_{jt}(\delta_{jt}) &= \int_{A_j(\delta)} f(\varepsilon)\varepsilon.
\end{aligned}
$$

where $A_j(\delta) = \{\epsilon | \delta_j + \varepsilon_{ij} > \delta_k + \varepsilon_{ik}, \forall k \neq j\}$

Market shares

- Assume that ϵ_{ijt} follows i.i.d. type I extreme value distribution
- The market share:

$$s_{jt}(\delta) = \frac{\exp(\delta_{jt})}{1 + \sum_{k=1}^{J_t} \exp(\delta_{kt})}$$

$$s_{0t}(\delta) = \frac{1}{1 + \sum_{k=1}^{J_t} \exp(\delta_{kt})}$$

Berry (1994)'s Inversion

- Take the log of the ratio:

$$\log\left(\frac{s_{jt}}{s_{0t}}\right) = \delta_{jt}$$
$$\iff \log\left(\frac{s_{jt}}{s_{0t}}\right) = \alpha p_{jt} + \beta X_{jt} + \xi_{jt}$$

- We can run linear regression to estimate (α, β)!
- Important note: Endogeneity of p_{jt}. Discuss IV later.

- General idea: Regardless of distributional assumption on ϵ_{jt}, there exists a unique $\{\delta_{jt}\}_j$ that rationalizes the observed market share $\{s_{jt}\}_j$.

$$s_{jt} = s_j(\{\delta_{jt}\}_j)\forall j \Longrightarrow \exists\{s_j^{-1}(\cdot)\}\ s.t.\ \delta_{jt} = s_j^{-1}(\{s_{jt}\}_j)$$

Recap of Price Elasticity under Logit

- Own- and cross- price elasticities are give by:

$$\frac{\partial s_j}{\partial p_k}\frac{p_k}{s_j} = \begin{cases} -\alpha p_j(1-s_j), & \text{if } k=j, \\ \alpha p_k s_k, & \text{otherwise} \end{cases}$$

Nested Logit

- The notation is based on Cardell (1991), *Econometric Theory*.
- Group the products into $G+1$ exhaustive and mutually exclusive sets, $g = 0, 1, \cdots, G$.
- Denote the set of products in group g as $\mathcal{G}_g$.
- The outside good, $j = 0$, is assumed to be the only member of group 0.

Setup

- Utility for consumer i choosing product j is given by

$$\begin{aligned} u_{ijt} &= \alpha p_j + \beta X_{jt} + \xi_{jt} + \xi_{igt} + (1-\sigma)\varepsilon_{ijt}, \\ &= \delta_j + \xi_{igt} + (1-\sigma)\varepsilon_{ijt} \end{aligned}$$

  - For consumer i, ξ_{ig} is common to all products in group g.
  - Assuming extreme value distribution for ε_{ij}, $\xi_{ig} + (1-\sigma)\varepsilon_{ij}$ is also an extreme value random variable.
- Role of σ:
  - As $\sigma \to 1$, the within group correlation of utility levels goes to one.
  - As $\sigma \to 0$, the within group correlation of utility levels goes to zero.

Market Share

- The market share of product j within the group g is given by:

$$s_{j/g}(\delta,\sigma) = \exp(\delta_j/(1-\sigma))/D_g$$

where D_g is given by $D_g \equiv \sum_{j \in \mathcal{G}_g} \exp(\delta_j/(1-\sigma))$.

- Similarly, the probability of choosing one of the group g products is

$$s_g(\delta,\sigma) = \frac{D_g^{(1-\sigma)}}{\left[\sum_k D_k^{(1-\sigma)}\right]}$$

- Therefore the market share for product j is given by:

$$s_j(\delta,\sigma) = s_{j/g}(\delta,\sigma)s_g(\delta,\sigma) = \frac{\exp(\delta_j/(1-\sigma))}{D_g^{\sigma}\left[\sum_k D_k^{(1-\sigma)}\right]}$$

- The market share for the outside options

$$s_0(\delta,\sigma) = 1/\left[\sum_k D_k^{(1-\sigma)}\right].$$

- Berry inversion

$$\begin{aligned}\ln(s_j) - \ln(s_0) &= \delta_j/(1-\sigma) - \sigma \ln(D_g)\\ \ln(s_{j/g}) &= \delta_j/(1-\sigma) - \ln(D_g)\end{aligned}$$

- Finally, we have

$$\begin{aligned}\ln(s_{jt}) - \ln(s_{0t}) &= \delta_{jt} + \sigma \ln(s_{jt/g}),\\ &= \beta X_{jt} + \alpha p_{jt} + \sigma \ln(s_{jt/g}) + \xi_{jt}.\end{aligned}$$

Estimation

- Linear regression model again!

$$\ln(s_{jt}) - \ln(s_{0t}) \quad = \quad \beta X_{jt} + \alpha p_{jt} + \sigma \ln(s_{jt/g}) + \xi_{jt}$$

- Instruments needed for both p_{jt} and $\ln(s_{jt|g})$
  - p_{jt} due to price endogeneity
  - inside-share $s_{jt|g}$ depends on ξ_{jt}.
- We will discuss instruments later.

Recap on Price Elasticity

- Remember that

$$\frac{\partial s_j}{\partial p_k}\frac{p_k}{s_j} = \begin{cases} -\alpha p_j(1-\sigma s_{j/g} - (1-\sigma)s_j)/(1-\sigma), & \text{if } k = j, \\ -\alpha p_k(\sigma s_{j/g} + (1-\sigma)s_k)/(1-\sigma), & \text{if } j, k \in g \\ \alpha p_k s_k, & \text{otherwise} \end{cases}$$

Overview of BLP approach

- In a nutshell: Estimation of random coefficient logit model with consumer heterogeneity using market-level data
- Difficulty
  1. Inversion is non-linear. Because
     $$s_j = \int \int P(j|x_i, \nu_i) dF(x_i) dG(\nu_i),$$
     so that parametric transformation does not give us linear regression equation.
  2. Identification. More on this later.
- I will follow the notation by Nevo (2001). Next week, I will talk about the empirical results of BLP (1995).

Setup

- Consider the random coefficient logit model

$$u_{ijt} = \alpha_i p_{jt} + \beta_i X_{jt} + \xi_{jt} + \epsilon_{ijt}$$

  - Random coefficient

$$\begin{pmatrix} \alpha_i \\ \beta_i \end{pmatrix} = \begin{pmatrix} \alpha \\ \beta \end{pmatrix} + \Pi D_i + \Sigma \nu_i$$

    - D_i is observed demographic characteristics
    - $\nu_i \sim N(0, I)$
  - Coefficients have both observed and unobserved heterogeneity (ΠD_i and $\Sigma \nu_i$)

- Rewrite the model

$$u_{ijt} = \delta_{jt}(\theta_1) + \mu_{ijt}(\theta_2) + \epsilon_{ijt}$$

where

$$\delta_{jt}(\theta_1) = \alpha p_{jt} + \beta X_{jt} + \xi_{jt}$$
$$\mu_{ijt}(\theta_2) = (\Pi D_i + \Sigma\nu_i)(p_{jt}, X_{jt})$$

  - $\theta_1 = (\alpha, \beta)$: linear parameters
  - $\theta_2 = (\Pi, \Sigma)$:non-linear parameters

- The market share

$$s_{jt} = \int\int \underbrace{\frac{\exp((\delta_{jt}(\theta_1) + \mu_{ijt}(\theta_2))}{1 + \sum_{j=1}^{J} \exp(\delta_{jt}(\theta_1) + \mu_{ijt}(\theta_2))}}_{=Prob(j|D_i,\nu_i)} dF(D_i)dG(\nu_i)$$
$$= \sigma_{jt}(\{\delta\}; \theta_1, \theta_2)$$

Naive idea for estimation

- Estimate (θ_1, θ_2) by minimizing the distance between observed and predicted market share

$$\min_{(\theta_1,\theta_2)} \sum_t \sum_j \left(s_{jt} - s_{jt}^{pred}(\theta_1, \theta_2) \right)^2$$

- Issues:
  - Computation. All parameters enter in a non-linear way.
  - More importantly: Price p_{jt} correlated with error term ξ_{jt}. Standard IV methods do not work.

Moment Conditions

- We form moment conditions w.r.t. ξ_{jt}

$$E\left[\xi_{jt}\begin{pmatrix} x_{jt} \\ z_{jt} \end{pmatrix}\right] = 0$$

  where z_{jt} is a vector of instruments

- In logit/nested logit, ξ_{jt} can be appeared in a linear way:

$$\ln(s_{jt}) - \ln(s_{0t}) = \beta X_{jt} + \alpha p_{jt} + \sigma \ln(s_{jt/g}) + \xi_{jt}$$

  so that you can runs linear 2SLS.

- In random coefficient logit, we do not have such analytical inversion.

Inversion in a nonlinear setting

- The model

$$\sigma_{jt}(\{\delta_{jt}\}_j, \{x_{jt}\}, \{p_{jt}\}; \theta_2) = \int\int \frac{\exp((\delta_{jt}(\theta_1) + \mu_{ijt}(\theta_2))}{1 + \sum_{j=1}^{J} \exp(\delta_{jt}(\theta_1) + \mu_{ijt}(\theta_2))} dF(D_i)$$

- Under weak conditions, there exists a unique vector of $\{\delta_{jt}\}_j$ that rationalizes the observed share $\{s_{jt}\}_j$

$$\delta_t = \sigma_t^{-1}(\{s_{jt}\}_j, \{x_{jt}\}, \{p_{jt}\}; \theta_2)$$

- Thus,

$$\xi_{jt} = \sigma_t^{-1}(\{s_{jt}\}_j, \{x_{jt}\}, \{p_{jt}\}; \theta_2) - (\alpha p_{jt} + \beta X_{jt})$$

- We can form moment conditions!!
    - Note: $\sigma_t^{-1}()$ is numerically calculated (more on this later).

BLP Estimation Algorithm (Nested Fixed Point Algorithm)

- Preliminary 1: Given δ and θ_2 (and observed covariates x_{jt} and p_{jt}), compute the market share $\sigma_{jt}(\{\delta\}_j, \{x_{jt}\}, \{p_{jt}\}; \theta_2)$.
- Estimation: Given θ_2,
  - Step 1 (inversion): search for δ that equates $\sigma_{jt}(\delta)$ to the observed market share s_{jt} for all j and t.
  - Step 2 (objective function): use the computed δ_t to compute ξ_{jt} and form the GMM objective function.
  - Search for the value of θ that minimizes the GMM objective function.

Preliminary: Computation of Market Share

- The market share

$$\sigma_{jt}(\{\delta_{jt}\}_j;\theta_2) = \int\int \frac{\exp((\delta_{jt}(\theta_1) + \mu_{ijt}(\theta_2))}{1+\sum_{j=1}^{J}\exp(\delta_{jt}(\theta_1)+\mu_{ijt}(\theta_2))} dF(D_i)dG(\nu_i)$$

- Use simulation to calculate the integrals
  - Let $r = 1, \cdots, R$. Consider the draw (D^r, ν^r) from $F(D_i)$ and $G(\nu_i)$
  - $F(D_i)$ is observed demographic distribution. $G(\nu_i)$ is Normal distribution.
- Then,

$$\sigma_{jt}(\{\delta_{jt}\}_j, \theta_2) = \frac{1}{R}\sum_{r=1}^{R} \frac{\exp((\delta_{jt} + \mu_{ijt}(D^r, \nu^r, \theta_2))}{1+\sum_{j=1}^{J}\exp(\delta_{jt}+\mu_{ijt}(D^r, \nu^r, \theta_2))}$$

- Note:
  - Other simulations (importance sampling, Halton sequence, etc).
  - Quadrature.

Step 1: Inversion

- Given θ_2, we want to obtain $\{\delta_{jt}\}_j$ such that

$$s_{jt} = \sigma_{jt}(\{\delta_{jt}\}_j, \theta_2) \ \forall j$$

- Generally, can use a contraction mapping for each market t

$$\delta_t^{h+1} = \delta_t^h + \log(S_t) - \log(\sigma_t(\delta_t, \theta_2))$$

  - repeat this untill $||\delta_t^{h+1} - \delta_t||$ is sufficiently small (like $1e-12$)

- Note: In logit/nested logit, this inversion can be done analytically!

Step 2: Compute the GMM objective

- The error term

$$\xi_{jt}(\theta_1, \theta_2) = \sigma^{-1}(S_t, \theta_2) - \beta x_{jt} - \alpha p_{jt}$$

  - Note: $\theta_1 = (\alpha, \beta)$ enters in a linear way, while θ_2 non-linear.

- Moment conditions

$$E\left[\xi_{jt}\begin{pmatrix} x_{jt} \\ z_{jt} \end{pmatrix}\right] = 0$$

  where z_{jt} is a vector of instruments

- The GMM objective function

$$\xi(\theta)' Z W Z' \xi(\theta)$$

  where
  - $\xi(\theta)$ is $(JT \times 1)$ vector,
  - $Z = (JT \times L)$ matrix (including instruments and exogenous variables)
  - W: weight matrix

Search for Parameters

- The search is non-linear.
  - Try different initial values / Use different optimizers
- Trick: Concentrate out the linear parameters θ_1
  - Given $\sigma^{-1}(S_t, \theta_2)$, can estimate (α, β) in a linear regression (with instrument for p_{jt})
- Computationally demanding
  - For each evaluation of θ, we have to do a contraction mapping to back out δ.
  - Structural estimation methods require a similar kind of computation.
  - There is an alternative approach to reduce computation costs: MPEC (Dube, Fox, and Su 2012 EMA)

Next Week

- Instruments (identification)
- Applications:
  - BLP (1995, EMA)
  - Nevo (2001, EMA)