Discrete Choice Models II

MIT
Department of Economics
14.385

Alberto Abadie

Demand Estimation with Aggregate Data

The first part of this handout provides a brief introduction to demand estimation in differentiated product markets.

The estimation methods use the random utility model setting of the previous handout.

Seminal contributions to this literature are:

- Berry (1994): "Estimating Discrete-Choice Models of Product Differentiation," *RAND Journal of Economics*
- Berry, Levinsohn, and Pakes (BLP, 1995): "Automobile Prices in Market Equilibrium," *Econometrica*

We will restrict ourselves to a simple setting where we observe **aggregate data** on prices and other product characteristics (X_j) and market shares (p_j). Also for simplicity, we will assume one market and many products. See references for other more complicated settings.

Demand Estimation with Aggregate Data

Consider the random utility model for product j and consumer i,

$$U_{ij} = X_j' \beta + \alpha_j + u_{ij} \quad (j = 0, \ldots, J).$$

α_j represents mean consumer valuation of the unobserved characteristics of product j (i.e., unobserved quality), and u_{ij} represents variation in consumer's valuation around that mean.

As a result, α_j is correlated with other observed product characteristics in X_j, in particular prices.

The mean utility level of product j is

$$\delta_j = X_j' \beta + \alpha_j.$$

Normalize $\delta_0 = 0$. If we observed $\delta_1, \ldots, \delta_J$ and a vector of instruments, Z_j, correlated with prices but uncorrelated with α_j (e.g., product specific cost shifters), we could estimate the previous equation by GMM.

Berry's Inversion Mechanism

Berry (1994) proposes an inversion mechanism to calculate $\delta_1, \ldots, \delta_J$ from the observed market shares $p_0, \ldots, p_J$.

The simplest possible case arises when the u_{ij} have a type I extreme value distribution. Then, the market share of product i is

$$p_j = \frac{e^{\delta_j}}{1 + \sum_{k=1}^{m} e^{\delta_k}},$$

and

$$\ln(p_j) - \ln(p_0) = \delta_j.$$

Therefore, an instrumental variable regression of $\ln(p_j) - \ln(p_0)$ on X_j identifies β (e.g., price elasticities).

However, as we have seen in the previous handout, type I extreme value errors produce unrealistic substitution patterns (IIA).

Berry, Levinsohn, and Pakes (BLP)

BLP propose a random coefficient model:

$$U_{ij} = X_j'\beta_i + \alpha_j + u_{ij} \quad (j = 0, \ldots, J),$$

where $\beta_i \sim N(\beta_0, \Sigma_0)$.

We obtain

$$U_{ij} = \delta_j + v_{ij},$$

where $\delta_j = X_j'\beta_0 + \alpha_j$, and $v_{ij} = X_j'(\beta_i - \beta_0) + u_{ij}$.

For each value of (β, Σ), we can approximate $\delta_1, \ldots, \delta_J$ using Monte-Carlo simulation:

(1) Obtain a large number $\beta_i^{(1)}, \ldots, \beta_i^{(R)}$ of computer generated values of a $N(\beta, \Sigma)$.

(2) Approximate $p_j(\beta, \Sigma, \delta_1, \ldots, \delta_J)$ as

$$\widehat{p}_j(\beta, \Sigma, \delta_1, \ldots, \delta_J) = \frac{1}{R} \sum_{r=1}^{R} \frac{e^{\delta_j + X_j'(\beta_i^{(r)} - \beta)}}{1 + \sum_{k=1}^{m} e^{\delta_k + X_k'(\beta_i^{(r)} - \beta)}}.$$

Berry, Levinsohn, and Pakes (BLP)

(3) Recover $\widehat{\delta}_1(\beta, \Sigma), \ldots, \widehat{\delta}_J(\beta, \Sigma)$ solving the set of equations

$$p_1 = \widehat{p}_1(\beta, \Sigma, \delta_1, \ldots, \delta_J)$$

$$\vdots$$

$$p_J = \widehat{p}_J(\beta, \Sigma, \delta_1, \ldots, \delta_J)$$

This is Berry's inversion mechanism. BLP show that the solution to this problem can be found iteratively, with iteration n

$$\delta_{jn} = \delta_{jn-1} + \ln(p_j) - \ln(\widehat{p}_j(\beta, \Sigma, \delta_{1n-1}, \ldots, \delta_{Jn-1}))$$

until convergence to $\widehat{\delta}_j(\beta, \Sigma)$.

Ignoring simulation error (which disappears as $R \to \infty$) and inversion error, we obtain

$$\widehat{\delta}_j(\beta_0, \Sigma_0) = X_j'\beta_0 + \alpha_j.$$

Given the availability of instruments Z_j, we can assemble a GMM estimator based on last equation.

Berry, Levinsohn, and Pakes (BLP)

Estimation of (β_0, Σ_0) is carried out using a nested optimization procedure:

- **Outer loop:** Iterates over (β, Σ) to minimize the GMM objective function
- **Inner loop:** Inside each iteration of the outer loop, iterates

$$\delta_{jn} = \delta_{jn-1} + \ln(p_j) - \ln(\widehat{p}_j(\beta, \Sigma, \delta_{1n-1}, \ldots, \delta_{Jn-1}))$$

  to compute $(\widehat{\delta}_1(\beta, \Sigma), \ldots, \widehat{\delta}_J(\beta, \Sigma))$.

  Use $(\widehat{\delta}_1(\beta, \Sigma), \ldots, \widehat{\delta}_J(\beta, \Sigma))$ to calculate the value of the GMM objective function at (β, Σ).

Dynamic Discrete Choice

Discrete choice in a dynamic setting, where today's decisions affect future values of state variables and agents maximize expected intertemporal utility.

Seminal contributions to this literature are:

- Rust (1987): "Optimal Replacement of GMC Bus Engines: An Empirical Model of Harold Zurcher," *Econometrica*
- Hotz and Miller (1993): "Conditional Choice Probabilities and the Estimation of Dynamic Models," *Review of Economic Studies*

These notes draw heavily from:

- Aguirregabiria and Mira (2010): "Dynamic Discrete Choice Structural Models: A Survey," *Journal of Econometrics*

Dynamic Discrete Choice

Consider an agent or set of agents choosing actions, a_{it} from a discrete set $A = \{0, 1, \ldots, J\}$ over an infinite horizon.

Agents observe state variables $S_{it} = (X_{it}, \varepsilon_{it})$. X_{it} is observed by the agent and by the econometrician. ε_{it} is observed by the agent but not observed by the econometrician.

In Rust (1987):

- The agent is Harold Zurcher, superintendent of maintenance at the Madison (WI) Metropolitan Bus Company
- X_{it} is engine mileage for bus i at month t
- ε_{it} are other characteristics of bus i at month t, which affect Zurcher's decisions, but unobserved by the econometrician.
- $a_{it} \in \{0, 1\}$ codes Zurcher's bus engine replacement decision

Other applications: Retirement decisions, occupational choice, dynamic discrete games.

Dynamic Discrete Choice

Agents' beliefs about future states follow a Markov transition process with transition probability function

$$P(X_{it+1}, \varepsilon_{it+1} | a_{it}, X_{it}, \varepsilon_{it}, \theta_p).$$

The value function $V_\theta(X_{it}, \varepsilon_{it})$ is the solution to the Bellman equation

$$\begin{aligned} V_\theta(X_{it}, \varepsilon_{it}) = \max_{a \in A} \Big[& U(a, X_{it}, \varepsilon_{it}, \theta_u) \\ & + \beta \int V_\theta(X_{it+1}, \varepsilon_{it+1}) dP(X_{it+1}, \varepsilon_{it+1} | a, X_{it}, \varepsilon_{it}, \theta_p) \Big] \end{aligned}$$

where

- U is the instantaneous utility function,
- β is the discount factor, (typically imputed, not estimated)
- $\theta = (\theta_p, \theta_u)$.

The optimal decision rule solves the Bellman equation.

Dynamic Discrete Choice

Some usual assumptions that make the problem tractable:

Additive separability + Logit:

$$U(a, X_{it}, \varepsilon_{it}, \theta_u) = u(a, X_{it}, \theta_u) + \varepsilon_{it}(a),$$

and $\varepsilon_{it} = (\varepsilon_{it}(0), \varepsilon_{it}(1), \ldots, \varepsilon_{it}(J))'$ is i.i.d. across i and t, with mutually independent (centered) type I extreme value components.

Conditional independence:

$$P(X_{it+1} | X_{it}, \varepsilon_{it}, a_{it}, \theta_p) = P(X_{it+1} | X_{it}, a_{it}, \theta_p)$$

Discrete support: X_{it} has discrete and finite support $\mathcal{X}$.

Dynamic Discrete Choice

For the engine replacement application in Rust (1987):

- Instantaneous utility is

$$u(a, X_{it}, \theta_u) + \varepsilon(a) = \begin{cases} -c(X_{it}, \theta_{u1}) + \varepsilon(0) & \text{if } a_{it} = 0 \\ -\theta_{u2} - c(0, \theta_{u1}) + \varepsilon(1) & \text{if } a_{it} = 1 \end{cases}$$

  where
  - $c(X_{it}, \theta_{u1})$: operating cost of a bus with X_{it} mileage (could be, e.g., polynomial), normalize $c(0, \theta_{u1}) = 0$,
  - θ_{u2}: engine replacement cost.

- Mileage (X_{it}) is discretized in 90 intervals of length 5000.

- The transition probabilities $P(X_{it+1}|X_{it}, a_{it}, \theta_p)$ are given by a Multinomial with three values corresponding to $[0, 5000)$, $[5000, 10000)$, $[10000, \infty)$ for mileage between t and $t+1$:

$$X_{it+1} - (1 - a_{it})X_{it},$$

  so θ_p has two parameters (probabilities sum to one).

Dynamic Discrete Choice

The additive separability + Logit and conditional independence assumptions produce an integrated version of the Bellman equation with closed form (see supplementary notes):

$$\bar{V}_\theta(X_{it}) = \ln\left(\sum_{a=0}^{J}\exp\left\{u(a, X_{it}, \theta_u) + \beta\sum_{x\in\mathcal{X}}\bar{V}_\theta(x)P(X_{it+1} = x|a, X_{it}, \theta_p)\right\}\right).$$

Moreover, the conditional choice probabilities are

$$P(a_{it} = a|X_{it}, \theta) = \frac{\exp\{\bar{v}_\theta(a, X_{it})\}}{\sum_{j=0}^{J}\exp\{\bar{v}_\theta(j, X_{it})\}}$$

where

$$\bar{v}_\theta(a, X_{it}) = u(a, X_{it}, \theta_u) + \beta\sum_{x\in\mathcal{X}}\bar{V}_\theta(x)P(X_{it+1} = x|a, X_{it}, \theta_p).$$

Dynamic Discrete Choice

The log-likelihood is

$$\sum_{i=1}^{N}\sum_{t=1}^{T}\ln P(a_{it}|X_{it},\theta)+\sum_{i=1}^{N}\sum_{t=2}^{T}\ln P(X_{it}|a_{it-1},X_{it-1},\theta_p)$$

Transition probabilities are specified as primitives of the model, which makes it easy to evaluate the second term of the log-likelihood. Typically, θ_p is estimated separately in a first step by maximizing that term.

Evaluating the first term of the likelihood is more difficult because it involves the integrated value function, $\bar{V}_\theta(x)$.

Rust (1987) proposes a *nested fixed point algorithm* (NFXP):

- **Outer loop:** Iterates over θ to maximize the likelihood
- **Inner loop:** Inside each iteration of the outer loop, iterates the Bellman equation until convergence to find $\bar{V}_\theta(x)$ (this is facilitated by the discrete nature of x)

Dynamic Discrete Choice

Suppose $\mathcal{X} = \{x_1, \ldots, x_k\}$ (in Rust's paper, $k = 90$). Let

$$\boldsymbol{u}(a, \theta_u) = \begin{pmatrix} u(a, x_1, \theta_u) \\ \vdots \\ u(a, x_k, \theta_u) \end{pmatrix}$$

and

$$\boldsymbol{F}(a) = \begin{pmatrix} p(x_1|a, x_1, \theta_p) & \cdots & p(x_k|a, x_1, \theta_p) \\ \vdots & \ddots & \vdots \\ p(x_1|a, x_k, \theta_p) & \cdots & p(x_k|a, x_k, \theta_p)) \end{pmatrix}$$

where

$$p(x'|a, x, \theta_p) = P(X_{it+1} = x'|a_{it} = a, X_{it} = x, \theta_p).$$

The matrix $\boldsymbol{F}(a)$ can be estimated in the first step from the second term in the log-likelihood function. So consider it as known for the rest of the argument.

Dynamic Discrete Choice

Let

$$\bar{\boldsymbol{V}}_\theta = \begin{pmatrix} \bar{V}_\theta(x_1) \\ \vdots \\ \bar{V}_\theta(x_k) \end{pmatrix}.$$

be the vectorized version of the value function, $\bar{V}_\theta(x)$.

Then, for any given θ_u, $\bar{\boldsymbol{V}}_\theta$ is given by the fixed point

$$\bar{\boldsymbol{V}}_\theta = \ln\left(\sum_{a=0}^{J} \exp\left\{\boldsymbol{u}(a,\theta_u) + \beta \boldsymbol{F}(a)\bar{\boldsymbol{V}}_\theta\right\}\right).$$

And the conditional choice probabilities are given by

$$P(a_{it} = a | X_{it} = x, \theta) = \frac{\exp\left(u(a,x,\theta_u) + \beta \boldsymbol{F}(a,x)\bar{\boldsymbol{V}}_\theta\right)}{\sum_{j=0}^{J} \exp\left(u(j,x,\theta_u) + \beta \boldsymbol{F}(j,x)\bar{\boldsymbol{V}}_\theta\right)},$$

where $\boldsymbol{F}(a,x) = (p(x_1|a,x,\theta_p), \ldots, p(x_k|a,x,\theta_p))$.

Dynamic Discrete Choice

A problem with the NFXP is its computational cost. The Bellman equation is solved iteratively in each optimization step.

The **conditional choice probability** (CCP) algorithm of Hotz and Miller (1993) does not require solving the Bellman equation, reducing computational cost drastically.

Hotz and Miller notice that, for given values of θ_u, it is often possible to obtain estimates $\widehat{v}(a, X_{it}, \theta_u)$ of $\bar{v}_\theta(a, X_{it})$ from preliminary estimates of θ_p and $P(a_{it} = a | X_{it})$.

Then, θ_u can be estimated in a second step by GMM. For example, MLE maximizes

$$\sum_{i=1}^{N} \sum_{t=1}^{T} \ln \left(\frac{\exp(\widehat{v}(a_{it}, X_{it}, \theta_u)}{\sum_{j=0}^{J} \exp(\widehat{v}(j, X_{it}, \theta_u)} \right)$$

with respect to θ_u.

Dynamic Discrete Choice

Example: Utility function linear in θ_u (e.g., polynomial):

$$u(a, x, \theta_u) = z(a, x)'\theta_u + \varepsilon(a).$$

Then, it can be shown

$$\bar{v}_\theta(a, x) = \tilde{z}(a, x, \theta)'\theta_u + \tilde{e}(a, x, \theta),$$

where $\tilde{z}(a, x, \theta)$ and $\tilde{e}(a, x, \theta)$ depend on θ only through θ_p, which can be estimated in a first step, and $P(a_{it}|X_{it} = x)$, which can be estimated non-parametrically.

Therefore, we can obtain estimates $\widehat{z}(a, x)$ and $\widehat{e}(a, x)$ of $\tilde{z}(a, x, \theta)$ and $\tilde{e}(a, x, \theta)$, so $\widehat{v}(a, x, \theta_u) = \widehat{z}(a, x)'\theta_u + \widehat{e}(a, x)$.

Then, the conditional choice probabilities are approximated as

$$P(a_{it} = a|X_{it} = x, \theta_u) = \frac{\exp(\widehat{z}(a, x)'\theta_u + \widehat{e}(a, x))}{\sum_{j=0}^{J} \exp(\widehat{z}(j, x)'\theta_u + \widehat{e}(j, x))}.$$

See Aguirregabiria and Mira (2010) for further details.