Estimation of Differentiated Products Demand Models 1

Introduction

Yuta Toyama

Last update: November 5, 2018

Why estimate demand function?

- Demand model is a key building block in applied microeconomics.
- Combined with supply-side model, we can do so many things!
  - Measuring market power or merger analysis (Nevo 2000, 2001)
  - Welfare gains from new products (Petrin 2002)
  - Effects of tax or subsidy (Goldberg and Hellerstein 2013)
- We focus on estimation of **differentiated products demand**
  - automobile, appliances, insurance plans, etc...
  - Estimation of homogenous goods is relatively straightforward.

Lecture Plan

- Introduction
  - Pricing model in differentiated product markets (Bertrand competition)
  - Issues in estimation of differentiated products demand.
- Characteristics approach (discrete choice modelling) with individual data
  - Reference: Train (2009). Available online.
- Discrete choice model with aggregate data: So-called “BLP approach”
  - Reference: Berry, Levinsohn, and Pakes (1995), Nevo (2000, JEMS).
  - Model (logit, nested logit, random coefficient logit)
  - Estimation and computation
  - Instruments
- Applications (e.g., welfare analysis)
- PS 4:
  - Estimation of simple logit model with aggregate data
  - Calculate elasticity matrix and markups based on demand estimates.

Multiproduct Oligopoly for Differentiated Products Markets

- Firm f produces J_f products. The profit for firm f

$$\pi_f = \sum_{j \in J_f} p_j q_j(\mathbf{p}) - C_j(q_j(\mathbf{p}))$$

  - The demand for product j, $q_j(\mathbf{p})$, depends on prices (and characteristics) of other products.

- Solution concept: **Bertrand-Nash equilibrium**

- Given prices of products offered by competitors, FOCs are

$$q_j(\mathbf{p}) + \sum_{r \in J_f} (p_r - mc_r) \frac{\partial q_r(\mathbf{p})}{\partial p_j} = 0, \forall j \in J_f$$

Vector Representation of Equilibrium Condition

- J: the number of products in the market. Define $\mathbf{q} = (q_1, \cdots, q_J)'$,
- $S(\mathbf{p})$: $J \times J$ matrix such that (i, j) element $S_{i,j}(\mathbf{p}) = -\partial q_j / \partial p_i$
- Ω^*: $J \times J$ matrix such that (i, j) element is given by

$$\Omega^*_{i,j} = \begin{cases} 1 & \text{if } \exists f : \{i, j\} \in J_f \\ 0 & \text{otherwise} \end{cases}$$

  - called **ownership matrix**.
- $\Omega(\mathbf{p}) = \Omega^* * S(\mathbf{p})$,where $*$ is element-by-element product
  - i.e., $\Omega_{i,j} = \Omega^*_{i,j} \times S_{i,j}$
- Then

$$\mathbf{q}(\mathbf{p}) - \underbrace{\Omega(\mathbf{p})}_{(J \times J)} \underbrace{(\mathbf{p} - \mathbf{mc})}_{(J \times 1)} = \mathbf{0}$$

What can we do?

- Estimation: markup and marginal costs:

$$\mathbf{p} - \mathbf{mc} = \Omega(\mathbf{p})^{-1}\mathbf{q}(\mathbf{p})$$

  - With (1) demand estimates and (2) ownership matrix, can obtain **mc**.
  - We use the observed ownership structure of products in data.

- Simulation: Merger analysis
  - Consider a counterfactual ownership matrix in which two firms are merged and jointy-maximizing their profits.
  - Solve the new FOC to simulate the outcome when two firms merged.

- Other simulation: effects of subsidy
  - With subsidy, $\mathbf{q}\left((1-\tau)\mathbf{p}\right) - \Omega\left((1-\tau)\mathbf{p}\right) \cdot (\mathbf{p} - \mathbf{mc}) = \mathbf{0}$

Importance of demand model (and estimates) in pricing decisions

- Consider firm f with $j = 1, 2$. FOC for product 1:

$$q_1 + (p_1 - mc_1)\frac{\partial q_1}{\partial p_1} + (p_2 - mc_2)\frac{\partial q_2}{\partial p_1} = 0$$

- Rewrite this

$$p_1 - mc_1 = \underbrace{\left(-\frac{\partial q_1}{\partial p_1}\right)^{-1} q_1}_{term1} + \underbrace{\left(-\frac{\partial q_1}{\partial p_1}\right)^{-1}\left(-\frac{\partial q_2}{\partial p_1}\right)(p_2 - mc_2)}_{term2}$$

- term 1: **own-price**. More elastic demand $\Rightarrow$ lower markup.
- term 2: **cross-price**. **Cannibalization between own products**
  - Higher substitution $\Rightarrow$ lower price of product 1 (to avoid cannibalization)
  - Mergers between firms with closer substitutes lead to higher price.

- Demand estimate is a key input for pricing decisions.

Issues in Estimation of Differentiated Products Demand

- Demand system with J products

$$Q_j = D_j(P_1, \cdots P_J) + \epsilon_j \; \forall j = 1, \cdots, J$$

  - Q_j: sales, P_j: price, ϵ_j: error term.
- Example: Constant elasticity demand system

$$\begin{pmatrix} \log Q_1 \\ \vdots \\ \log Q_J \end{pmatrix} = \begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_J \end{pmatrix} + \begin{pmatrix} \theta_{11} & \cdots & \theta_{1J} \\ \vdots & & \vdots \\ \theta_{J1} & \cdots & \theta_{JJ} \end{pmatrix} \begin{pmatrix} \log P_1 \\ \vdots \\ \log P_J \end{pmatrix} + \begin{pmatrix} \epsilon_1 \\ \vdots \\ \epsilon_J \end{pmatrix}$$

  - θ_{jj} is price-elasticity
- Issue 1: Endogeneity of price. Need instruments.
- Issue 2: Too many parameters!!
  - # of price-elasticity parameters: J^2 problem

Two major approaches for J^2 problem

- Product-Space approach with **economic restrictions**
  - Reduce # of parameters using economic restrictions.
  - AIDS model by Deaton and Muellbaur (1980)
  - Hausman (1996), Chadhuri, Goldberg, and Jia (2008)
  - I will not talk about this approach in the course.
- **(Main focus)** Characteristic-Space approach
  - Consider preference on the characteristics of products.
  - Demand is derived from discrete choice decisions.
  - Discrete choice with individual data: Train (20110
  - Discrete choice with aggregate data: BLP

Overview of Characteristics Approach

- Each product is a bundle of characteristics.
- Consumer drives utility from these characteristics.
- Each consumer chooses the product which gives the highest utility (discrete choice model).
- With stochastic structure, we can obtain the choice probability of a particular consumer for each product.
- Aggregating across (heterogenous) consumers, we can obtain the market demand for that product

Setup

- Consumer $i = 1, \cdots, N$
- Product $j \in \mathbf{J} = \{0, 1, \ldots, J\}$
  - also called "alternative", "choice", etc...
  - $j = 0$ is called "outside goods" (i.e., not obtaining anything).
- The utility of consumer i from obtaining product $j \geq 1$

$$u_{ij} = \beta_0 + \alpha p_j + \beta X_j + \epsilon_{ij}$$

  and $u_{i0} = \epsilon_{i0}$
  - p_j: price of product j
  - X_{ij}: characteristics of product j for person i
    - example: interaction between family size and car size
  - (α, β): preference parameters.
  - ϵ_{ij}: idyosyncratic preference shock **(random shock)**.
- The model is also called **random utility model.**

- Note: We will generalize this specification later
  - Random coefficients of (α, β)
  - More consumer heterogeneity, etc.

- Consumer chooses the product j that gives the highest utility

$$d_i = \arg \max_{j \in \{0,1,\cdots,J\}} u_{ij}$$

where

$$u_{ij} = \beta_0 + \alpha p_j + \beta X_{ij} + \epsilon_{ij}$$

Multinomial logit model

- Assume that $\epsilon_{i,j}$ follows i.i.d. type I extreme value distribution

$$F(x) = e^{e^{-x}}$$

- This error structure leads to the following choice probability

$$\Pr\left(d_i = j | \{p_j, X_j\}_j\right) = \frac{exp(\beta_0 + \alpha p_j + \beta X_j)}{1 + \sum_{k=1}^{J} exp(\beta_0 + \alpha p_k + \beta X_k)}$$

  - Note: the utility from good 0 is $u_{i0} = 0 + \epsilon_{i0}$.

- Market demand can be given by

$$D_j(\mathbf{p}) = N \times \Pr\left(d_i = j | \{p_j, X_j\}_j\right)$$

Estimation by Maximum Likelihood

- Data: $\{X_j, p_j, d_{ij}\}_j$ for each $i = 1, \cdots, N$.
  - $d_{ij} = 1$ if consumer i chooses product j, and 0 otherwise.
- The likelihood function can be written as

$$L(\theta) = \prod_{i=1}^{N} \left[\prod_{j=0}^{J} (Pr(d_i = j))^{d_{ij}} \right]$$

where $\theta = (\beta_0, \alpha, \beta)$

The IIA property

- Key assumption: independence of the error term: $\epsilon_{i,j}$ and $\epsilon_{i,j'}$ are independently randomly distributed if $j \neq j$.
- The model implies that

$$\frac{P(j|X)}{P(j'|X)} = \frac{\exp(\beta X_j)}{\exp(\beta X_{j'})}$$

- The MNP model satisfies Luce's **independence from irrelevant alternative axiom (IIA)**
  - The odds of selecting one alternative j relative to another alternative j' is independent from the size/composition of the choice set $\mathbf{J}$ or the utilities of other choices not equal to j or j'.
- The IIA property puts strong restrictions on substitution patterns between products.

Limitation of Logit Model –Blue/Red Bus Problem–

- Suppose that $J = \{car, red\ bus\}$.
- Assume $\beta X_{car} = \beta X_{redbus}$, so that both have 50% shares.
- Now consider a third artificial alternative "blue bus",
  - $X_{bluebus} = X_{redbus}$ (imagine half of the existing red bus are painted in blue).
- You would expect $Pr(bluebus) = Pr(redbus) = 1/4, P(car) = 1/2$ as a "reasonable" prediction.
- However, the multinomial logit implies

$$Pr(bluebus) = Pr(car) = Pr(redbus) = 1/3.$$

- Key assumption: independence of random shocks $\epsilon_{redbus}, \epsilon_{bluebus}, \epsilon_{car}$.
  - $\epsilon_{bluebus}$ should be more correlated with ϵ_{redbus}. If so, *bluebus* would steal more consumers from *redbus*.

Implications on Substitution Patterns

- The price elasticity under the multinomial logit

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

where $s_j = Pr(d = j)$. (interpret this as market share). Note: $\alpha < 0$

- Issue 1: Own price elasticity is (almost) proportional to own price
  - Higher price leads to higher price elasticity.
- Issue 2: Cross-price elasticity is proportional to s_k
  - Proportional substitution from k to all products (not depending on j!!)
  - Direct implication from the IIA property!
    - Change in $p_k \rightarrow$ change in s_j and $s_{j'}$, but $s_j/s_{j'}$ should be constant by the IIA.

One way to avoid IIA: Random Coefficients

- Suppose that

$$u_{ij} = \beta_i X_j + \epsilon_{ij}$$

where $\beta_i \sim N(\beta, \sigma^2)$. Assume the logit error ϵ_{ij} and independence between β_i and ϵ_{ij}.

- Let $\beta_i = \beta + \nu_i$ where $\nu_i \sim N(0, \sigma^2)$.
- Then, the utility is

$$u_{ij} = \beta X_j + \underbrace{(X_j \nu_i + \epsilon_{ij})}_{\equiv \eta_{ij}}$$

- Now,

$$Cov(\eta_{ij}, \eta_{ij'}) = \sigma^2 X_j X_{j'}$$

so that two alternatives j and j' are correlated.

- We will see this in more detail later.

Three models

1. Nested logit model
2. Random coefficient logit model
3. Multinomial probit model (to be skipped today)

Nested Logit Model

- Denote $u_j = v_j + \epsilon_j$
- Consider that we can group alternatives (products) in "nests.
- (See whiteboard)
- K nests with J_k choices within each nest.
- Joint CDF for $(\epsilon_1, \cdots, \epsilon_J)$

$$F(\epsilon) = \exp\left[-G\left(e^{-\epsilon_{11}}, \cdots, e^{-\epsilon_{K,J_K}}\right)\right]$$

$$G() = \sum_{k=1}^{K}\left(\sum_{j=1}^{J_k} \epsilon_{jk}^{1/\lambda_k}\right)^{\lambda_k}$$

where $\lambda_k \approx \sqrt{1 - Corr(\epsilon_{k,j}, \epsilon_{k,j'})}$

  - Notation: ϵ_{jk} is error term for product j in nest k.
  - $Corr(\epsilon_{jk}, \epsilon_{j'k}) \neq 0$ and $Corr(\epsilon_{jk}, \epsilon_{j'k'}) = 0$ for $k \neq k'$.
    - Correlation within group, no correlation across groups.
  - $\lambda_k = 1\ \forall k$ leads to the multinomial logit model.

- Choice probability for product j in nest k is given by

$$
\begin{aligned}
Pr(d=j) &= \frac{exp(\frac{v_j}{\lambda_k})\left(\sum_{k\in J_k}\exp(\frac{v_j}{\lambda_k})\right)^{\lambda_k-1}}{\sum_{l=1}^{K}\left(\sum_{j\in J_l}\exp(\frac{v_j}{\lambda_l}\right)^{\lambda_l}} = \\
&= Pr(d=j|nest\ \ k)\times P(choose\ \ nest\ \ k)
\end{aligned}
$$

where

$$
\begin{aligned}
Pr(d=j|nest\ \ k) &= \frac{exp(\frac{v_j}{\lambda_k})}{\sum_{j\in J_k}\exp(\frac{v_j}{\lambda_k})} \\
Pr(nest\ \ k) &= \frac{\exp(\lambda_k I_k)}{\sum_{l=1}^{K}\exp(\lambda_l I_l)}
\end{aligned}
$$

and

$$
I_k = \log\left(\sum_{j\in J_k}\exp(\frac{v_j}{\lambda_k})\right)
$$

Breaking the IIA property

- IIA still holds within each nest (group)
  - If two products j and j' are in the same group,

$$\frac{P(j|k)}{P(j'|k)} = \frac{\exp(v_j/\lambda_k)}{\exp(v_{j'}/\lambda_k)}$$

- IIA breaks across different nests (groups)
  - Two products: j in nest k and j' in nest k'

$$\frac{P(j|k)}{P(j'|k')} = \frac{\exp(v_j/\lambda_k)\left(\sum_{j \in J_k} \exp(\frac{v_j}{\lambda_k})\right)^{\lambda_k - 1}}{\exp(v_{j'}/\lambda_{k'})\left(\sum_{j \in J_{k'}} \exp(\frac{v_j}{\lambda_{k'}})\right)^{\lambda_{k'} - 1}}$$

Elasticity (To be discussed in detail next week)

- Denote $\lambda_k = \lambda \forall k$ and define $\sigma = 1 - \lambda$.
  - $\sigma = 0$ means the multinomial logit.
- Price elasticity is

$$\frac{\partial s_j}{\partial p_k}\frac{p_k}{s_j} = \begin{cases} \alpha p_j(1-\sigma s_{j|g} - (1-\sigma)s_j)/(1-\sigma) & \textit{if } k = j \\ \alpha p_k(\sigma s_{j|g} + (1-\sigma)s_k)/(1-\sigma) & \textit{if } k \neq j, \textit{same group} \\ -\alpha p_k s_k & \textit{if } k \neq j \end{cases}$$

  - remember: $u_{ij} = \alpha p_j + ... + \epsilon_{ij}$
  - $s_{j|g} = Prob(j|g)$
- Stronger substitution within group.

Notes

- Estimation can be done by MLE.

- More flexible substitution patterns.
  - Two products in the same nest (group) are closer substitutes than those in the different groups.
  - The parameter λ_k governs the degree of substitution.
  - We discuss the elasticity pattern in a more simplified setting next week.

- Cons: Group structure is pre-determined.

Random Coefficient Logit Model (or Mixed Logit Model)

- Consider the model with random coefficients

$$u_{ij} = \beta_i X_{ij} + \epsilon_{ij}$$

  where $\beta_i \sim f(\beta)$. Assume the logit error ϵ_{ij}.

- With coefficient β_i, the choice probability is logit-form

$$Pr(d_i = j|\beta_i) = \frac{\exp\left(\beta_i X_{ij}\right)}{\sum_{j=1}^{J} \exp\left(\beta_i X_{ij}\right)}$$

- But, the researcher does not know β_i. Thus, we need to integrate it out.
- The unconditional choice probability

$$Pr(d_i = j) = \int \frac{\exp\left(\beta_i X_{ij}\right)}{\sum_{j=1}^{J} \exp\left(\beta_i X_{ij}\right)} dF(\beta_i)$$

- The stochastic terms of different alternatives are correlated.
- If $X_{i,j}$ and $X_{i,j'}$ are close, then j and j' are more close substitutes.

- The elasticity

$$\frac{\partial s_j}{\partial p_k}\frac{p_k}{s_j} = \begin{cases} \frac{p_j}{s_j}\int \beta_i^{price} s_j(\beta_i)\left(1 - s_j(\beta_i)\right) dF(\beta_i) & if \;\; k = j \\ \frac{p_j}{s_j}\int \beta_i^{price} s_j(\beta_i) s_k(\beta_i) dF(\beta_i) & if \;\; k \neq j \end{cases}$$

where $s_j = Pr(d_i = j)$ and $s_j(\beta_i) = Pr(d_i = j|\beta_i)$.

- Higher correlation between $s_j(\beta_i)$ and $s_k(\beta_i)$ over different β_i leads to more substitution.

Estimation using Simulation Method

- Consider specification: $\beta \sim N(\mu, \sigma^2)$. Need to estimate (μ, σ^2).
- Remember that the individual choice probability

$$Pr(d_i = j) = \int \underbrace{\frac{\exp\left((\mu + \sigma\nu_i)X_{ij}\right)}{\sum_{j=1}^{J} \exp\left((\mu + \sigma\nu_i)X_{ij}\right)}}_{Pr(d_i = j|\nu_i, \theta)} dG(\nu_i)$$

where $G(\nu_i)$ is the standard normal distribution.

- The likelihood function is not analytical because of the integral over ν_i!!

Simulated MLE

- Idea: Approximate the likelihood through simulation!
- Given $\theta = (\mu, \sigma)$,
  - Step 1: Draw ν from $G(\nu)$. Label ν^r.
  - Step 2: Calculate the logit formula $Pr(d_i = j|\nu_i)$ with ν^r.

$$Pr(d_i = j|\nu^r) = \frac{\exp\left((\mu + \sigma\nu^r)X_{ij}\right)}{\sum_{j=1}^{J} \exp\left((\mu + \sigma\nu^r)X_{ij}\right)}$$

  - Step 3: Repeat these R times, and average the results

$$\hat{Pr}(d_i = j; \theta) = \frac{1}{R}\sum_{r=1}^{R} Pr(d_i = j|\nu^r, \theta)$$

  - Step 4: Maximize the simulated log likelihood (MSL)

$$SLL(\theta) = \sum_{i=1}^{N}\sum_{j=1}^{J} d_{ij} \log \hat{Pr}(d_i = j)$$

Dealing with Endogeneity

- So far, I have ignored the endogeneity issue of covariates.
- In general, dealing with endogeneity in discrete choice model is tough.
- I focus on a particular type of endogeneity that matters in demand estimation: **unobserved quality of products at the market level.**
- Idea:
  - Product quality may not be fully captured by observed X_j
  - Firms's pricing decisions depends on such unobserved product quality, leading to endogeneity issue.

Model

- Introduce a new index t that denotes "market".
  - "market" can represent both "geography" and "time".

- Consider the random coefficient logit model

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

  - Random coefficient

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

    and D_i is observed demographic characteristics
  - ξ_{jt}: unobserved product characteristics

- Note: This is more general than necessary to explain how to deal with endogeneity issues. But, this setup is very close to what we see in the next week.

- Rewrite the model

$$u_{ijt} = \delta_{jt} + \mu_{ijt} + \epsilon_{ijt}$$

where

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

  - δ_{jt} is called "**mean utility**", which captures both observed and unobserved quality.

- The choice probability for consumer i:

$$Pr(d = j|D_i) = \int \frac{\exp((\delta_{jt} + \mu_{ijt}(\Pi, \Sigma))}{1 + \sum_{j=1}^{J} \exp(\delta_{jt} + \mu_{ijt}(\Pi, \Sigma))} dF(\nu_i)$$

Estimation with Microdata (individual data)

- With individual data, we can treat δ_{jt} as a product-market FE and estimate it in MLE!
  - For identification, need to observe many consumers within a market.
- First step: estimate $(\{\delta_{jt}\}_{j,t}, \Pi, \Sigma)$ by MLE.
- Second step: Use estimated $\hat{\delta}_{jt}$ to recover (α, β) by regression

$$\hat{\delta}_{jt} = \alpha p_{jt} + \beta X_{jt} + \xi_{jt}$$

with an IV for p_{jt}.

Next Step

- How should we do if we do not have microdata?
  - Relatedly, what if we have too many products? (Goolsbee and Petrin 2004, BLP 2004)
- What are the potential IVs for price p_{jt}?
- We will discuss these issues in the next lectures.