Agenda

- estimating demands in differentiated product markets
  - traditional models: CES, multi-stage budgeting
  - discrete choice models
    - Generation I
    - Generation II

Estimating Demand in Differentiated Product Markets

- goal: estimate markups, this requires estimates of demand and supply
- suppose there are J products, the demand system is given by

$$q = D(p, x)$$

  where
  - q is a J-dimensional vector of quantities demanded
  - p is J dimensional vector of prices
  - x is a vector of demand shifters
- need to specify a functional form for the demand system that is consistent with choice theory and flexible enough to fit the data
  - examples: translog, Almost Ideal Demand System (AIDS), linear expenditure system, etc

Dimensionality Problem

- basic problem in IO: J is typically quite large
  - brands of beer are at least 50
  - number of models of cars is over 100
- the number of parameters to be estimated (i.e., price elasticities) is on the order of J^2 - way too many parameters

Solutions

- avoid the problem by focusing on an aggregate (e.g., Porter aggregates all eastbound shipments rather than differentiating across destination cities) or on a small subset of the products (Baker and Bresnahan study a particular segment of the beer industry)
  - issue: do you need to estimate the full demand system to answer the question in which you are interested?
- impose structure on preferences such as symmetry or separability to restrict the substitution patterns across products
- define products as bundles of a limited number of characteristics and define preferences on characteristics rather than products
- we shall consider the second approach first

Product Space Models: CES Model

- the CES utility function

$$U(q_1, ..., q_J) = \left(\sum_{j=1}^{J} q_j^{\sigma} \right)^{1/\sigma}$$

  where σ measures rate of substitution across products

- maximizing U subject to a linear budget constraint yields demand

$$q_j = \frac{p_j^{-1/(1-\sigma)} I}{\sum_{k=1}^{J} p_k^{-\sigma/(1-\sigma)}}, \; j = 1, ..., J$$

  where I is consumer income

- cost of using CES is that it imposes strong and implausible restrictions on own and cross price elasticities

$$\frac{\partial q_i}{\partial p_j} \frac{p_j}{q_i} = \frac{\partial q_k}{\partial p_j} \frac{p_j}{q_k}, \text{ for all } i, j, k$$

Anderson, de Palma and Thisse (1992)

- utility function

$$U(q_1, ..., q_J) = \sum_{j=1}^{J} \delta_j q_j - \sum_{j=1}^{J} q_j \log q_j$$

- the above utility function yields the Logit demand

$$s_j = \frac{\exp(\delta_j - p_j)}{\sum_{k=1}^{J} \exp(\delta_k - p_k)}$$

  where s_j is the budget share of good j

- estimation involves only J parameters: $(\delta_1, ..., \delta_J)$

Anderson, de Palma and Thisse (1992)

- more flexible but still quite restrictive: ratio $\frac{q_j}{q_k}$ is independent of p_i
- intuition from the utility function

$$U(q_1, ..., q_J) = \sum_{j=1}^{J} \delta_j q_j - \sum_{j=1}^{J} q_j \log q_j$$

  - first term implies that consumer only consumes good with highest $\delta_j - p_j$
  - second term captures desire for diversity and implies that consumer purchases a positive amount of every good, but all products are treated symmetrically

More General Approach: Deaton and Mullbauer (1980)

- divide products into smaller groups and allow for flexible functional form within each group
- two ideas: separability and multi-stage budgeting
- (weakly) separable preferences: q can be partitioned into $\left(q^1, \ldots, q^N\right)$

$$U = f\left(v_1\left(q^1\right), \ldots., v_N\left(q^N\right)\right)$$

  where $v_K\left(q^K\right)$ is a sub-utility function (i.e., represents a preference ordering over q^K) and f is an increasing function in all of its arguments

Separable Preferences

- implication: maximizing U subject to linear budget constraint is equivalent to following two step optimization program
  - step 1: fix allocation of income across the commodity groups $\left(I^1, \ldots, I^N\right)$ and solve N optimization problems of the form

$$\max v_K\left(q^K\right) \ s.t. \ p^K q^K = I^K, \ K = 1, .., N$$

    the solution to each of these subproblems is a set of subgroup demands of the form

$$q^K = g^K\left(p^K, I^K\right)$$

  - step 2: substitute subgroup demands into subgroup utility functions to obtain the indirect utility functions $\psi^K\left(p^K, I^K\right)$, the choose allocation of income to solve

$$\max_{(I^1, ..., I^N)} f\left(\psi^1\left(p^1, I^1\right), ..., \psi^N\left(p^N, I^N\right)\right) \ s.t. \ \sum_{K=1}^{N} I^K = x$$

Separable Preferences (Cont.)

- weak separability is a necessary and sufficient condition for the first step
  - preferences are typically assumed to be additively (strong) separable over time
  - preferences over goods are typically assumed to be separable from leisure
- separability implies a reduction in the number of parameters
  - suppose each group consists of m products, number of parameters is $Nm^2/2$ rather than $(Nm)^2/2$
  - it can be tested because it imposes restrictions on the substitution matrix

Multi-stage Budgeting

- in the second step, we often would like to treat the optimization problem as one in which the consumer chooses quantities of the composite commodity groups to maximize utility
- for example, consumer demands for food, clothing, shelter, and entertainment are often expressed as functions of price indices for these commodities and income
- question: when is this procedure legitimate?

Case I: Homothetic Preference

- suppose preferences over q^K are homothetic, then

$$\psi^K\left(p^K, I^K\right) = \frac{I^K}{b_K\left(p^K\right)}$$

  where $b_K\left(p^K\right)$ can be interpreted as the price index for commodity K

- defining $v^K \equiv \psi^K\left(p^K, I^K\right)$, the optimization problem in step 2 can be expressed as

$$\max_{(v_1,...,v_N)} f\left(v_1, ..., v_N\right) \ s.t. \ \sum_{K=1}^{N} b\left(p^K\right) v_K$$

- problem: requiring preferences at the group level to be homothetic imposes a lot of structure on within group demands
  - e.g., budget shares are independent of group income, rules out groups that possess both luxuries and necessities

Case II: Additively Separable

- assume preferences are additively separable across groups, i.e.,

$$u = \sum_{K=1}^{N} v^K \left(q^K\right)$$

- we can allow group preferences to be quasi-homothetic, which yields indirect utility functions of the form

$$\psi^K \left(p^K, I^K\right) = f^K \left(\frac{I^K}{b_K \left(p^K\right)}\right) + a_K \left(p^K\right)$$

  where f^K is monotone increasing

  - this permits very general forms of Engel curves for the individual commodities comprising each group

Application: Hausman Papers on Demand in Beer and Cereal Market

- Hausman, Leonard and Zona (1994) and Hausman (1996)
- data: brand j's prices and shares, by city c and quarter t
- multi-level demand system with three levels
  - top level: overall demand for the product, e.g., beer or ready-to-eat cereal
  - middle level: demand for different market segments, e.g., in beer, lager, pilsner and ale; in cereals, family, kids and adult cereals
  - bottom level: a system of demands for different brands in each segment

Bottom Level: AIDS

- demand for brand j within segment g in city c in quarter t is

$$s_{jct} = \alpha_{jc} + \beta_j \log \left(\frac{I_{gct}}{P_{gct}}\right) + \sum_{k=1}^{J_g} \gamma_{jk} \log \left(p_{kct}\right) + \varepsilon_{jct}$$

  where s_{jct} is segment expenditure share of brand j

- here P_{gct} is the price index for segment g in city c in quarter t and J_g is the number of brands in segment g
  - Stone logarithmic price index

$$P_{gct} = \sum_{j \in g} s_{jct} \log \left(p_{jct}\right)$$

– Deaton and Mullbauer exact price index

$$P_{gct} = \alpha_0 + \sum_{j \in g} \alpha_j p_{jct} + \sum_{j \in g} \sum_{k \in g} \gamma_{jk} \log(p_{kct}) \log(p_{jct})$$

- with stone price index, brands demand can be estimated using linear methods; the Deaton and Mullbauer price index requires non-linear methods

Middle and Top Level Demands

- middle level

$$\log(q_{gct}) = \beta_g \log(I_{ct}) + \sum_{g=1}^{G} \delta_g \log(\pi_{gct}) + \alpha_{gc} + \varepsilon_{gct}$$

  where q_{gct} is the composite quantity of the g segment in city c in quarter t, π_{gct} are the segment price indices computed as above

- top level

$$\log(q_{ct}) = \beta_0 + \beta_1 \log(I_{ct}) + \beta_2 \log(\pi_{ct}) + \theta Z_{ct} + \varepsilon_{ct}$$

  where q_{ct} is the quantity of beer (or cereal) in city c in quarter t, I_{ct} is the expenditure on beer in city$c$ in quarter t, π_{ct} is the price index for beer, and Z_{ct} are demand shifters

Discrete Choice Models

- products are bundles of characteristics: $j = 0, .., J$ where 0 is the outside good
- consumer preferences are defined on space of characteristics
- each consumer $i = 1, .., N$ chooses at most one unit of one of the inside goods, the choice maximizes utility
- consumers are heterogeneous: they have different preferences for different characteristics, the distribution of heterogeneity is parameterized

Discrete Choice Models (Cont.)

- in this model, we are estimating the joint distribution of preferences over characteristics
  - the number of parameters is primarily determined by the dimensionality of the characteristics and independent of the number of products
- the parameter estimates measure the consumers' marginal valuation of characteristics: if a new good is introduced, we can value that good since it is simply a bundle of characteristics
  - we can also predict outcomes demand for the old and new goods in the expanded choice set
- caveat: the new good cannot be too "new" – i.e., possess a new characteristic

Basic Model

- utility of consumer i for product j is given by

$$u_{ij} = U\left(x_j, p_j, \xi_j, \upsilon_i; \theta\right)$$

where

  - x_j: a vector of observed characteristics of product j
  - ξ_j: unobserved characteristic of product j
  - p_j: price of product j
  - υ_i: unobserved preference characteristics of consumer i
  - θ: vector of utility parameters to be estimated
  - υ_i: distribution is denoted by F

Choice Probability

- define the subset of "consumers" (preferences) that lead to choice j as

$$A_j\left(\theta\right) = \left\{\upsilon \left| u_{ij} > u_{ik}, \forall k \right.\right\}$$

- then the probability that consumer i chooses product j is

$$\sigma_j\left(x, p, \xi; \theta\right) = \int_{\upsilon \in A_j(\theta)} f\left(\upsilon\right) d\upsilon$$

where $x = \left(x_1, ..., x_J\right)$, $p = \left(p_1, ..., p_J\right)$ and f is the density associated with F

- under the assumption that "market size" M is very large and υ_i's are i.i.d. across consumers, the Law of Large Numbers implies that market demand converges to $M\sigma_j\left(x, p, \xi; \theta\right)$

Remarks

- if there is no outside good, then the market is covered and aggregate demand is M, the number of consumers in the market $\rightarrow$ inelastic market demand, no market expansion effects

- normalizations: choices of individual consumers are invariant to affine transformation of utilities

  - invariance to additive shifts implies normalizing mean utility of outside good to zero $\rightarrow$ deduct u_{i0} from each u_{ij} for $j = 1, .., J$
  - invariance to scale leads to normalizing one of the other parameters (typically variance of F) to one

Generation I Models

- data: $\{s_j, p_j, x_j\}$ where s_j is the observed market share of product j
- basic idea is to estimate θ (which includes the parameters of F) by minimizing the distance between the predicted choice probabilities and observed market shares
  - the choice model determines $(\sigma_0(\theta), \sigma_1(\theta), ..., \sigma_J(\theta))$
  - each consumer is an independent draw from F, then the distribution of product choices is given by a multinomial distribution

Estimation

- let q_j denote the number of consumers who choose product j, the likelihood function for the data is

$$L = \Pi_{j=0}^{J} [\sigma_j(\theta)]^{q_j}$$

- taking logs, choose θ to

$$\max_{\theta} M \sum_{j=0}^{J} s_j \log [\sigma_j(\theta)] \Leftrightarrow \min_{\theta} \sum_{j=0}^{J} \frac{[s_j - \sigma_j(\theta)]^2}{\sigma_j(\theta)}$$

  - last statement follows from taking a Taylor series approximation of $\sigma_j(\theta)$ around the data point s_j
  - the latter statistic is called a minimum χ^2, if we use observed shares in denominator, then it is modified χ^2

Example: Logit Model

- utility function

$$u_{ij} = x_j\beta - p_j + \epsilon_{ij}$$

  where ϵ_{ij} is distributed i.i.d. with mean zero across products and consumers and its distribution is Type I extreme value, i.e., $F(\epsilon) = \exp[-\exp(-\epsilon)]$

- predicted choice probabilities (market shares)

$$\sigma_j(\theta) = \frac{\exp(x_j\beta - p_j)}{\sum_{k=0}^{J} \exp(x_k\beta - p_k)}, \; j = 0, 1, ..., J$$

  - remark: McFadden (1974) shows that the multinomial logit model is derived from utility maximization if and only if $\{\epsilon_{ij}\}$ are independent across products and distributed Type I extreme value

Example: Logit Model (Cont.)

- normalize the mean utility of the outside good to zero implies that

$$\sigma_0(\theta) = \frac{1}{\sum_{k=0}^{J} \exp(x_k\beta - p_k)}$$

- match predicted market shares to observed ones: the model is correctly specified

$$s_j = \sigma_j(\theta),\ j = 0, 1, ..., J$$

- therefore

$$\log(s_j) - \log(s_0) = x_j\beta - p_j$$

  - no additional taste parameter so $\theta = \beta$

Vertical Model: Bresnahan (1981)

- utility function

$$u_{ij} = \upsilon_i\varphi_j - p_j,\ \upsilon_i > 0$$

  where φ_j measures the quality of good j and is assumed to be strictly increasing in j

  - no unobserved product characteristics: $\varphi_j = x_j\beta$

- product choice probabilities (product demands)

  - necessary condition of positive demand

$$\upsilon\varphi_j - p_j > \upsilon\varphi_{j+1} - p_{j+1}$$

$$\upsilon\varphi_j - p_j > \upsilon\varphi_{j-1} - p_{j-1}$$

    which implies

$$\frac{p_j - p_{j-1}}{\varphi_j - \varphi_{j-1}} < \upsilon < \frac{p_{j+1} - p_j}{\varphi_{j+1} - \varphi_j}$$

Market Shares

- define

$$\Delta_j = \frac{p_j - p_{j-1}}{(x_j - x_{j-1})\beta},\ J > j > 0$$

  and $\Delta_0 = -\infty$, $\Delta_J = \infty$

- necessary and sufficient condition for demands for all J goods to be positive is that Δ_j is strictly increasing in j

- market share of product j is

$$\sigma_j(\theta) = F(\Delta_{j+1}) - F(\Delta_j)$$

  where F is the distribution of υ

Remarks

- normalizations: (i) $\varphi_0 = 0$, (ii) $p_0 = 0$
- here $\theta = (\beta, \lambda)$ where λ is the parameter of F: choose θ to minimize the difference between the observed and predicted market shares
- differences between actual market shares and the choice probabilities can only be due to sampling error
  - as $M \to \infty$, $s_j \to \infty$, model should fit exactly
  - lack of prediction error means model is certain to be rejected by the data: no value for β such that actual shares = predicted shares

Generation II: Berry (1994)

- theory of the error is unobserved product characteristics: ξ_j
- mean utility for product j is

$$u_{ij} = x_j\beta - \alpha p_j + \xi_j \equiv \delta_j$$

- vertical model is a special case: $\varphi_j = x_j\beta + \xi_j$, $\alpha = 1$ and the mean of υ is normalized to 1

Endogeneity

- main problem: firms know ξ when they set prices so prices are correlated with the error which in turn is buried in a highly non-linear set of equations
- solution: assume M is large enough that

$$s_j = \sigma_j(\delta^*)$$

  where $\delta^* \equiv (\delta_1^*, ..., \delta_J^*)$ solves the system of J independent equations
- Berry shows that δ^* is unique and BLP provide a contraction mapping to find it

Estimation: F is known

- if we assume that F is known, then δ^*can be treated as a known nonlinear transformation of the market share data
- using δ^* as data, run the regression

$$\delta^*(s) = x_j\beta - \alpha p_j + \xi_j$$

  - ξ is correlated with p, we need to find instruments: cost shifters, (exogenous) characteristics of other products

Estimation: General Case

- more generally, when F is not known, then δ^* depends on the unknown parameter λ of F
- for each value of (β, λ), there exists a unique solution of ξ that makes the predicted market shares equal to actual shares
- let $\xi(\theta)$ denote this solution and then use the moment conditions

$$E\left[\xi\left(\theta_0\right) Z\right] = 0$$

  to estimate θ_0, where Z is a set of instrumental variables

Examples

- Logit model

$$\log\left(s_j\right) - \log\left(s_0\right) = \delta_j$$

  - no need to numerically compute δ's, simply run 2SLS of difference in log shares on (x_j, p_j) with instruments for price
- vertical model: $\delta_j = \varphi_j - p_j$
  - $s_j = F\left(\Delta_{j+1}\right) - F\left(\Delta_j\right)$ implies $\Delta_j = F^{-1}\left(F\left(\Delta_{j+1}\right) - s_j\right)$ with initial condition $\Delta_J = F^{-1}\left(1 - s_J\right)$
  - the values of φ_j can be obtained from the recursion $\varphi_j = \varphi_1 + \frac{p_j - p_{j-1}}{\Delta_j}$
  - treat φ_j as data and regress δ on x (use IVs if necessary)