CHAPTER 2

Demand Estimation

1. Introduction

The estimation of demand equations is a fundamental component in most empirical applications in IO. It is also important in many other fields in empirical economics. There are important reasons why economists in general, and IO economists in particular, are interested in demand estimation. Knowledge of the demand function, and of the corresponding marginal revenue function, is crucial for the determination of a firm's optimal choice of prices or quantities. Any measure of consumer welfare, in one way or the other, is based on the estimation of demand. In many applications in empirical IO, demand estimation is also a necessary first step to measure market power. In the absence of direct information about firms' costs, the estimation of demand and marginal revenue is key for the identification of marginal costs (using the marginal cost equals marginal revenue condition) and firms' market power. The estimation of demand of differentiated products is very helpful in the prediction of the demand of a new product. Similarly, the estimation of the degree of substitution between the products of two competing firms is a fundamental factor in evaluating the profitability of a merger between these firms. Economists are also interested in demand estimation in order to improve our measures of consumer welfare, and more specifically, Cost-of-Living indices (COLI). For instance, the Boskin commission (Boskin et al., 1997 and 1998) concluded that the US Consumer Price Index (CPI) overstated the change in the cost of living by about 1.1 percentage points per year.[1] As Hausman (2003) and Pakes (2003) explain, the estimation of demand systems provides a solution to each of these sources of bias in the CPI.

Most products that we find in today's markets are differentiated products: automobiles; smartphones; laptop computers; or supermarket products such as ketchup, soft drinks,

[1] Consumer prices indexes (CPI) are typically constructed using weights which are obtained from a consumer expenditure survey. For instance, the Laspeyres index for a basket of n goods is $CPI_L = \sum_{i=1}^{n} w_i^0 \left(\frac{P_i^1}{P_i^0} \right)$, where P_i^0 and P_i^1 are the prices of good i at periods 0 and 1, respectively, and w_i^0 is the weight of good i in the total expenditure of a representative consumer at period 0. A source of bias in this index is that it ignores that the weights w_i^0 change over time as the result of changes in relative prices of substitute products, or the introduction of new products between period 0 to period 1. The Boskin Commission identifies the introduction of new goods, quality improvements in existing goods, and changes in relative prices as the main sources of bias in the CPI as a cost of living index.

breakfast cereals, or laundry detergent. A differentiated product consists of a collection of varieties such that each variety is characterized by some attributes that distinguishes it from the rest. A variety is typically produced by a single manufacturer, but a manufacturer may produce several varieties.

We distinguish two approaches to model demand systems of differentiated products: demand systems in product space; and demand systems in characteristics space. In empirical applications, the 'product space' model was the standard approach until the 1990s. However, we will see in this chapter that the 'characteristics space' approach has several advantages that made it the predominant model in empirical IO over the last two decades.

2. Demand systems in product space

2.1. Model. In this model, consumer preferences are defined over goods (or varieties) themselves. Consider a product with J varieties that we index by $j \in \{1, 2, ..., J\}$. Let q_j the quantity that a consumer buys and consumes of variety j, and let $(q_1, q_2, ..., q_J)$ be the vector with the purchased quantities of all the varieties. The consumer has a utility function $U(q_1, q_2, ..., q_J)$ defined over the vector of quantities. The consumer problem consists in choosing the vector $(q_1, q_2, ..., q_J)$ to maximize his utility subject to his budget constraint.

$$\begin{array}{cc} \max\limits_{\{q_1, q_2, ..., q_J\}} & U(C, q_1, q_2, ..., q_J) \\ \\ subject\ to: & C + p_1\ q_1 + p_2\ q_2 + ... + p_J\ q_J \leq y \end{array} \tag{2.1}$$

where C represents consumption of the outside-good or numerarie, $(p_1, p_2, ..., p_J)$ is the vector of prices, and y is the consumer's disposable income. The demand system is the solution to this optimization problem. We can represent this solution in terms of J functions, one for each variety, that give us the optimal quantity of each variety as a function of prices and income. These are the *Marshallian demand equations*:

$$\begin{array}{rcl} q_1 & = & f_1\left(p_1, p_2, ..., p_J, y\right) \\ q_2 & = & f_2\left(p_1, p_2, ..., p_J, y\right) \\ ... & & ... \\ q_J & = & f_J\left(p_1, p_2, ..., p_J, y\right) \end{array} \tag{2.2}$$

The form of the functions f_1, f_2, ..., f_J depends on the form of the utility function $U(.)$. The following are some examples.

Example 1 (Linear Expenditure System). Consider the Stone-Geary utility function:

$$U = C\ (q_1 - \gamma_1)^{\alpha_1}\ (q_2 - \gamma_2)^{\alpha_2} ...\ (q_J - \gamma_J)^{\alpha_J} \tag{2.3}$$

where $\{\alpha_j,\ \gamma_j : j = 1, 2, ..., J\}$ are parameters. This utility function was first proposed by Geary (1950), and Richard Stone (1954) was the first to estimate the Linear Expenditure System. Solving the budget constraint into the utility function, we have that $U = [y - p_1$

q_1- ... $-p_J\ q_J]$ $(q_1-\gamma_1)^{\alpha_1}$ ... $(q_J-\gamma_J)^{\alpha_J}$. The marginal conditions of optimality of the consumer problem are $dU/dq_j = 0$ for every variety j, and this implies:

$$\alpha_j \frac{U}{q_j-\gamma_j} - p_j \frac{U}{C} = 0 \tag{2.4}$$

Solving for q_j, we get $q_j = \gamma_j + \alpha_j \frac{C}{p_j}$. Plugging this expression into the buget constraint we have that:

$$C = \frac{y - \sum_{j=1}^{J} p_j\ \gamma_j}{1+\sum_{j=1}^{J} \alpha_j} \tag{2.5}$$

And plugging this expression into equation $q_j = \gamma_j + \alpha_j \frac{C}{p_j}$, we obtain the equations of the Linear Expenditure System:

$$q_j = \gamma_j + \frac{\alpha_j^*}{p_j}\left(y - \sum_{i=1}^{J} p_i\ \gamma_i\right) \tag{2.6}$$

where $\alpha_j^* = \alpha_j/[1+\sum_{i=1}^{J}\alpha_i]$. This system is convenient for its simplicity. However, it is also very restrictive. For instance, it imposes the restriction that all the goods are complements in consumption. This is not realistic in most applications, particularly when the goods under study are varieties of a differentiated product.

Example 2 ("Almost Ideal Demand System"). The most popular specification of demand system when preferences are defined on the product space is the "Almost Ideal Demand System" proposed by Deaton and Muellbauer (1980a, 1980b). The utility function has the following form:

$$U = \left[\prod_{j=1}^{J} q_j^{\alpha_j}\right] + \sum_{j=1}^{J}\sum_{k=1}^{J} \delta_{jk}\ q_j\ q_k \tag{2.7}$$

For this model the system of Marshallian demand equations is:

$$w_j = \alpha_j + \gamma_j\ \ [\ln(y) - \ln(P)] + \sum_{k=1}^{J} \beta_{jk}\ \ \ln(p_k) \tag{2.8}$$

where $w_j \equiv p_j q_j / y$ is the expenditure share of product j, $\{\alpha_j, \beta_{jk}, \gamma_j\}$ are parameters which are known functions of the utility parameters $\{\alpha_j, \delta_{jk}\}$, and P is a price index that is defined as $\ln(P) \equiv \sum_{j=1}^{J} \alpha_j \ln(p_j) + \frac{1}{2}\sum_{j=1}^{J}\sum_{k=1}^{J} \beta_{jk} \ln(p_j)\ln(p_k)$. The model implies the symmetry conditions $\beta_{jk} = \beta_{kj}$. Therefore, the number of free parameters is: $2J + \frac{J(J+1)}{2}$, that increases quadratically with the number of products.

2.2. Estimation. In empirical work, the most commonly used demand systems are the Rotterdam Model (Theil, 1975), the Translog Model (Christensen, Jorgensen and Lau, 1975), and the Almost Ideal Demand System (AIDS) (Deaton and Muellbauer, 1980a). Since Deaton and Muellbauer proposed their Almost Ideal Demand System in 1980, this model has been estimated in hundreds of empirical applications. In most of the applications, a "good" is an aggregate product category (e.g., beef meat, or chicken meat). However, there are also some applications for varieties of a differentiated product, such as the one in Hausman (1996) that we examine later in this chapter. In this section we describe the typical application of this class of model.

The typical dataset consists of aggregate market level data for a single market, over T time periods, with information on consumption and prices for a few product categories. For instance, Verbeke and Ward (2001) use monthly data from January 1995 to December 1998 ($T = 48$ data points) from a consumer expenditure survey in Belgium. They estimate a demand system for fresh meat products that distinguishes three product categories: Beef/veal, Pork, and Poultry. We index time by t. For each period t we observe aggregate income y_t, and prices and quantities of the J product categories: $\{y_t, q_{jt}, p_{jt} : t = 1, 2, ..., T;$ $j = 1, 2, ..., J\}$. We want to estimate the demand system:

$$w_{jt} = \alpha_j^0 + \mathbf{X}_t \ \boldsymbol{\alpha}_j^1 + \gamma_j \ \ln(y_t/P_t) + \textstyle\sum_{k=1}^{J} \beta_{jk} \ \ln(p_{kt}) + \varepsilon_{jt} \tag{2.9}$$

where $\mathbf{X}_t$ is a vector of exogenous characteristics that may affect demand, e.g., demographic variables. We want to estimate the vector of structural parameters $\theta = \{\alpha_j, \beta_{jk}, \gamma_j : \forall j, k\}$. Typically, this system is estimated by OLS or by Nonlinear Least Squares (NLLS) to incorporate the restriction that $\ln(P_t)$ is equal to $\sum_{j=1}^{J} \left[\alpha_j^0 + \mathbf{X}_t \ \boldsymbol{\alpha}_j^1\right] \ln(p_{jt}) + \frac{1}{2} \sum_{j=1}^{J} \sum_{k=1}^{J} \beta_{jk} \ln(p_{jt})$ $\ln(p_{kt})$, and the symmetry restrictions on $\beta' s$. These estimation methods assume that prices are not correlated with the error terms $\varepsilon' s$. We discuss this and other assumptions in the section.

2.3. Some limitations and extensions of this approach. (1) Representative consumer assumption. The representative consumer assumption is a very strong one and it does not hold in practice. The demand of certain goods depends not only on aggregate income but also on the distribution of income and on the distribution of other variables affecting consumers' preferences, e.g., age, education, etc. The propensity to substitute between different products can be also very heterogeneous across consumers. Therefore, ignoring consumer heterogeneity is a very important limitation of the actual applications in this literature. However, there is nothing that avoids the estimation of this model using household level data. Suppose that we have this type of data, and let use the subibdex h for households. The demand system becomes:

$$w_{jht} = \alpha_j^0 + \mathbf{Z}_{ht} \ \boldsymbol{\alpha}_j^1 + [\gamma_j^0 + \gamma_j^1 \mathbf{Z}_{ht}] \ln(y_{ht}/P_t) + \textstyle\sum_{k=1}^{J} [\beta_{jk}^0 + \beta_{jk}^1 \mathbf{Z}_{ht}] \ln(p_{kt}) + \varepsilon_{jht} \tag{2.10}$$

where $\mathbf{Z}_{ht}$ represents a vector of exogenous household characteristics, other than income. This model incorporates household observed heterogeneity in a flexible way: in the level of demand, in price elasticities, and in income elasticities.

Note that (typically) prices do not vary across households. Therefore, price elasiticities are identified only from the time-series variation in prices, and not from the cross-sectional variation across households. In this context, household level data is useful to allow for consumer heterogeneity in price responses, but it does not provide additional sample variation to improve the precision in the estimation of price elasticities.

Household level data introduces also a new issue in this class of models: some households do not consume all the product categories, even when when these categories are quite aggregate, e.g., beef meat. However, this class of model predicts that the household consumes a positive amount of every product category.

(2) Too many parameters problem. In the standard model, the number of parameters is $2J + \frac{J(J+1)}{2}$, i.e., J intercept parameters (α); J income elasticities (γ); and $\frac{J(J+1)}{2}$ free price elasticities (β). The number of parameters increases quadratically with the number of goods. Note also that, in most applications, the sample variation in prices comes only from time series, and the sample size T is relatively small. This feature of the model implies that the number of products, J, should be quite small. For instance, even if J is as small as 5, the number of parameters to estimate is 25. Therefore, with this model and data, it is not possible to estimate demand systems for differentiated products with many varieties. For instance, suppose that we are interested in the estimation of a demand system for different car models, and the number of car models is $J = 100$. Then, the number of parameters in the AIDS model is $5,250$, and we need many thousands of observations (markets or/and time periods) to estimate this model. This type of data is typically not available.

(3) Finding instruments for prices. Most empirical applications of this class of models have ignored the potential endogeneity of prices. [2] However, it is well known and simultaneity and endogeneity are potentially important issues in any demand estimation. If part of the unobservable ε_{jt} are known to firms when the chose prices, we expect that prices will be correlated with this error terms and the OLS method will provide inconsistent estimates of demand parameters. The typical solution to this problem is using instrumental variables. In the context of this model, the researcher needs at least as many instruments as prices, that is J. The ideal case is when we have information on production costs for each individual good. However, that information is very rarely available.

(4) Problems to predict demand of new goods. In the literature of demand of differentiated products, a class of problem that has received substantial attention is the evaluation or prediction of the demand of a new product. Trajtenberg (1989), Hausman (1996), and Petrin (2002) are some of the prominent applications that deal with this empirical question. In a demand system in product space, estimating the demand of a new good, say

[2] An exception is, for instance, Eales and Unnevehr (1993) who find strong evidence on the endogeneity of prices in a system of meat demand in US. They use livestock production costs and technical change indicators as instruments.

$J+1$, requires estimates of the parameters associated with that good: α_{J+1}, γ_{J+1} and $\{\beta_{J+1,j} : j = 1, 2, ..., J+1\}$. Of course, this makes it impossible to make counterfactual predictions, i.e., predict the demand of a product that has not been introduced in any market yet. But it also limits the applicability of this model in cases where the new product has been introduced very recently or in very few markets, because we may not have enough data to estimate these parameters.

2.4. Dealing with some of the limitations: Hausman on cereals. Hausman (1996) studies the demand for ready-to eat (RTE) cereals in US. This industry has been characterized by the dominant position of six multiproduct firms and by the proliferation of many varieties. During the period 1980-92, the RTE cereal industry has been among the most prominent introducers of new brands within U.S. industries, with approximately 190 new brands that were added to the pool of existing 160 brands. Hausman shows that using panel data from multiple geographic markets, together with assumptions on the spatial structure of unobserved demand shocks and costs, it is possible to deal with some of the problems mentioned above within the framework of demand systems in product space. He applies the estimated system to evaluate the welfare gains from the introduction of Apple-Cinnamon Cheerios by General Mills in 1989.

(1) Data. The dataset comes from supermarket scanner data collected by Nielsen company. It covers 137 weeks ($T = 137$) and seven geoagraphic markets ($M = 7$) or standard metropolitan statistical areas (SMSAs), including Boston, Chicago, Detroit, Los Angeles, New York City, Philadelphia, and San Francisco. Though the data includes information from hundred of brands, the model and the estimation concentrates in 20 brands classified in three segments: adult (7 brands), child (4 brands), and family (9 brands). Apple-Cinnamon Cheerios are included in the family segment. We index markets by m, time by t, and brands by j, such that data can be described as $\{p_{jmt}, q_{jmt} : j = 1, 2, ..., 20;\ m = 1, 2, ..., 7;\ t = 1, 2, ..., 137\}$. Quantities are measured is phisycal units. There are not observable cost shifters.

(2) Model. Hausman estimates an Almost-Ideal-Demand-System combined with a nested three-level structure. The top level is the overall demand for cereal using a price index for cereal relative to other goods. The middle level of the demand system estimates demand among the three market segments, adult, child, and family, using price indexes for each segment. The bottom level is the choice of brand within a segment. For instance, within the family segment the choice is between the brands Cheerios, Honey-Nut Cheerios, Apple-Cinnamon Cheerios, Corn Flakes, Raisin Bran (Kellogg), Wheat Rice Krispies, Frosted Mini-Wheats, Frosted Wheat Squares, and Raisin Bran (Post). Overall price elasticities are then derived from the estimates in all three segments. The estimation is implemented in reverse order, beginning at the lowest level (within segment), and then using those estimates to construct price indexes at the next level, and implementing the estimation at the next level. At the lowest level, within a segment, the demand system is:

$$s_{jmt} = \alpha^1_{jm} + \alpha^2_t + \gamma_j \ \ln(y_{Smt}) + \sum_{k=1}^{J} \beta_{jk} \ \ln(p_{kmt}) + \varepsilon_{jmt} \tag{2.11}$$

where y_{Smt} is overal segment expenditure. The terms α^1_{jm} and α^2_t represent product, market and time effects, respectively, which are captured using dummies.

(2) Instruments. Suppose that the supply (pricing equation) is:

$$\ln(p_{jmt}) = \delta_j \ c_{jt} + \tau_{jm} + u_j(\boldsymbol{\varepsilon}_{mt}) \tag{2.12}$$

where c_{jt} represents a common cost shifter (unobservable to the researcher) which is consistent with the national level production in this industry, τ_{jm} is city-brand fixed effect that captures differences in transportation costs, and $u_j(\boldsymbol{\varepsilon}_{mt})$ captures the response of prices to local demand shocks, with $\boldsymbol{\varepsilon}_{mt} \equiv \{\varepsilon_{1mt}, \varepsilon_{2mt}, ..., \varepsilon_{Jmt})$. The identification assumption is that these demand shocks are not (spatially) correlated across markets: for any pair of markets $m \neq m$ it is assumed that:

$$E(u_j(\boldsymbol{\varepsilon}_{mt}) \ u_k(\boldsymbol{\varepsilon}_{m't})) = 0 \qquad \text{for any } j, k \tag{2.13}$$

The assumption implies that after controlling for brand-city fixed effects, all the correlation between prices at different locations comes from correlation in costs, and not from spatial correlation in demand shocks. Under these assumptions we can use average prices in other local markets, $\overline{P}_{j(-m)t}$, as instruments, where:

$$\overline{P}_{j(-m)t} = \frac{1}{M-1} \sum_{m' \neq m} p_{jm't} \tag{2.14}$$

(3) Approach to evaluate the effects of new goods. Suppose that product J is a "new" product, though it is a product in our sample and we have data on prices and quantities of this product such that we can estimate all the parameters of the model including α^0_J, $\{\beta_{Jk}\}$ and γ_J. The expenditure function $e(\mathbf{p}, u)$ for Deaton & Muellbauer demand system is:

$$e(\mathbf{p}, u) = \sum_{j=1}^{J} \alpha^0_j \ln(p_j) + \frac{1}{2} \sum_{j=1}^{J} \sum_{k=1}^{J} \beta_{jk} \ln(p_j) \ln(p_k) + u \prod_{j=1}^{J} p_j^{\gamma_j}$$

And let $V(\mathbf{p}, y)$ be the indirect utility associated with the demand system, that we can easily obtain by solving the demand equations into the utility function. The functions $e(\mathbf{p}, u)$ and $V(\mathbf{p}, y)$ corresponding to the situation where the new product J is already in the market. Suppose that we have estimated the demand parameters after the introduction of the good and let $\hat{\theta}$ be the vector of parameter estimates. We use $\hat{e}(\mathbf{p}, u)$ and $\hat{V}(\mathbf{p}, y)$ to represent the functions $e(\mathbf{p}, u)$ and $V(\mathbf{p}, y)$ when we use the parameter estimates $\hat{\theta}$. Similarly, we use $\hat{D}_j(\mathbf{p}, y)$ to represent the estimated Marshallian demand of product j.

The concept of *virtual price* plays a key role in Hausman's approach to obtain the value of a new good. Hausman defines the *virtual price* of the new good J (represented as p_J^*) as the price of this product that makes its demand just equal to zero. Of course, this virtual price depends on the prices of the other goods and on the level of income. We can define a virtual price of product J for each market and quarter in the data. That is, p_{Jmt}^* is implicitly defined as the price of product J that solves the equation:

$$\hat{D}_j(p_{1mt}, p_{2mt}, ..., p_{Jmt}^*) = 0$$

Hausman compares the factual situation with the new product with the counterfactual situation where everything is equal except that the price of product J is p_{Jmt}^* such that nobody buys this product. Let u_{mt} be the utility of the representative consumer in market m at period t with the new product: i.e., $u_{mt} = \hat{V}(\mathbf{p}_{mt}, y_{mt})$. By construction, it should be the case that $\hat{e}(\mathbf{p}_{mt}, u_{mt}) = y_{mt}$. To reach the same level of utility u_{mt} without the new product, the representative consumer's expenditure should be $\hat{e}(p_{1mt}, p_{2mt}, ..., p_{Jmt}^*, u_{mt})$. Therefore, the *Equivalent Variation* (in market m at period t) associated to the introduction of the new product is:

$$EV_{mt} = \hat{e}(p_{1mt}, p_{2mt}, ..., p_{Jmt}^*, u_{mt}) - y_{mt}$$

Hausman consider this measure of consumer welfare.

(4) Limitations of this approach. A key issue in Hausman's approach is the consideration of a market with prices and income $(p_{1mt}, p_{2mt}, ..., p_{Jmt}^*, y_{mt})$ as the relevant counterfactual to measure the value of good J in a market with actual prices and income $(p_{1mt}, p_{2mt}, ..., p_{Jmt}, y_{mt})$. This choice of counterfactual has some important limitations. In particular, it does not take into account that the introduction of the new good can change the prices of other goods. In many cases we are interested in estimating the reaction of different firms to the introduction of a new good. To obtain these effects we should calculate equilibrium prices before and after the introduction of the new good. Therefore, we should estimate both demand and firms' costs under an assumption about competition (e.g., competitive market, Cournot, Bertrand).

3. Demand systems in characteristics space

3.1. Model. The model is based on three basic assumptions. First, a product, say a laptop computer, can be described as a bundle of physical characteristics: e.g., CPU speed, memory, screen size, etc. These characteristics determine a *variety* of the product. Second, consumers have preferences on bundles of characteristics of products, and not on the products per se. And third, a product has J different varieties and each consumer buys at most one variety of the product per period, i.e., all the varieties are substitutes in consumption.

We index varieties by $j \in \{1, 2, ..., J\}$. From an empirical point of view, we can distinguish two sets of product characteristics. Some characteristics are observable and measurable to the researcher. We represent with them using a vector of K attributes $\mathbf{X}_j \equiv (X_{1j}, X_{2j}, ..., X_{Kj})$, where X_{kj} represents that "amount" of attribute k in brand j. For instance, in the case of laptops we could have that X_{1j} represents CPU speed; X_{2j} is RAM memory; X_{3j} is hard disk memory; X_{4j} is weight; X_{5j} is screen size; X_{6j} is a dummy (binary) variable that indicates whether the manufacturer of the CPU processor s Intel or not; etc. Other characteristics are not observable, or at least measurable, to the researcher but they are known and valuable to consumers. There may be many of these unobservable attributes, and we describe these attributes using a vector $\boldsymbol{\xi}_j$, that contains the "amounts" that variety j has of the different unobservable attributes. The researcher does not even know even the number of unobservable attributes, i.e., he does not known the dimension and the space of $\boldsymbol{\xi}_j$.

We index households by $h \in \{1, 2, ..., H\}$ where H represents the number of households in the market. A household has preferences defined over bundles of attributes. Consider a product with arbitrary attributes $(\mathbf{X}, \boldsymbol{\xi})$. The utility of consumer h if he consumes that product is $V_h(\mathbf{X}, \boldsymbol{\xi})$. Importantly, note that the utility function V_h is defined over any possible bundle of attributes $(\mathbf{X}, \boldsymbol{\xi})$ that may or may not exist in the market. For a product j that exists in the market and has attributes $(\mathbf{X}_j, \boldsymbol{\xi}_j)$, this utility is $V_{hj} = V_h(\mathbf{X}_j, \boldsymbol{\xi}_j)$. The total utility of a consumer is additively separable in the utility from this product, and the utility from other goods: $U_h = u_h(C) + V_h(\mathbf{X}, \boldsymbol{\xi})$, where C represents the amount of a composite good, and $u_h(C)$ is the utility from the composite good.

Consumers are differences in their levels of income, y_h, and in their preferences. Consumer heterogeneity in preferences can be represented in terms of a vector of consumer attributes $\boldsymbol{\upsilon}_h$ that may be completely unobservable to the researcher. Therfore, we can write the utility of consumer h as:

$$U_h = u(C; \boldsymbol{\upsilon}_h) + V(\mathbf{X}, \boldsymbol{\xi}; \boldsymbol{\upsilon}_h) \tag{3.1}$$

We also assume that there is continuum of consumers with measure H, such that $\boldsymbol{\upsilon}_h$ has a well-defined density function $f_{\boldsymbol{\upsilon}}$ in the market.

Each consumer buys at most one variety of the product (per period). Given his income, y_h, and the vector of product prices $\mathbf{p} = (p_1, p_2, ..., p_J)$, a consumer decides which varity to buy, if any. Let $d_{hj} \in \{0, 1\}$ be the indicator of the event "consumer h buys product j". A consumer decision problem is:

$$\begin{array}{ll} \max_{\{d_{h1},d_{h2},\ldots,d_{hJ}\}} & u(C;\boldsymbol{\upsilon}_h)+\sum_{j=1}^{J} d_{hj}\ V(\mathbf{X}_j,\boldsymbol{\xi}_j;\boldsymbol{\upsilon}_h) \\ \textit{subject to}: & C+\sum_{j=1}^{J} d_{jh}\ p_j \leq y_h \\ & d_{hj}\in\{0,1\} \text{ and } \sum_{j=1}^{J} d_{jh}\in\{0,1\} \end{array} \tag{3.2}$$

A consumer should choose between $J+1$ possible choice alternatives, each of the J products, and alternative $j=0$ that represents not to buy any product. The solution to this consumer decision problem provides the consumer-level demand equations $d_j^*(\mathbf{X},\mathbf{p},y_h;\boldsymbol{\upsilon}_h)\in\{0,1\}$ such that:

$$\begin{array}{l} \{d_j^*(\mathbf{X},\mathbf{p},y_h;\boldsymbol{\upsilon}_h)=1\}\Leftrightarrow \\ \left\{u(y_h-p_j;\boldsymbol{\upsilon}_h)+V(\mathbf{X}_j,\boldsymbol{\xi}_j;\boldsymbol{\upsilon}_h)>u(y_h-p_k;\boldsymbol{\upsilon}_h)+V(\mathbf{X}_k,\boldsymbol{\xi}_k;\boldsymbol{\upsilon}_h) \text{ for any } k\neq j\right\} \end{array} \tag{3.3}$$

where $k=0$ the alternative of not buying any variety (i.e., outside alternative), that has indirect utility $u(y_h;\boldsymbol{\upsilon}_h)$. Given consumers demands $d_j^*(\mathbf{X},\mathbf{p},y_h;\boldsymbol{\upsilon}_h)$ and the joint density function $f(\boldsymbol{\upsilon}_h,y_h)$, we can obtain the aggregate demand functions:

$$q_j(\mathbf{X},\mathbf{p},f)=\int d_j^*(\mathbf{p},y_h;\boldsymbol{\upsilon}_h)\ ,\beta)\ f(\boldsymbol{\upsilon}_h,y_h)\ d\boldsymbol{\upsilon}_h\ d\mathbf{y}_h) \tag{3.4}$$

and the market shares $s_j(\mathbf{X},\mathbf{p},f)\equiv\dfrac{q_j(\mathbf{X},\mathbf{p},f)}{H}$.

Example. Logit model of product differentiation.

Suppose that:

$$V(\mathbf{X}_j,\widetilde{\boldsymbol{\xi}}_j;\boldsymbol{\upsilon}_h)=\mathbf{X}_j\beta+\xi_j+\varepsilon_{hj}$$

where ε's are *i.i.d.* Extreme Value Type 1. And

$$u(C;\boldsymbol{\upsilon}_h)=\alpha\ C$$

Then,

$$U_{hj}=-\alpha\ p_j+\mathbf{X}_j\ \beta+\xi_j+\varepsilon_{hj}$$

And

$$s_j=\frac{q_j}{H}=\frac{\exp\left\{-\alpha\ p_j+\mathbf{X}_j\ \beta+\xi_j\right\}}{1+\sum_{k=1}^{J}\exp\left\{-\alpha\ p_k+\mathbf{X}_k\ \beta+\xi_k\right\}}$$

where $\delta_j\equiv-\alpha\ p_j+\mathbf{X}_j\ \beta+\xi_j$ represents the mean utility of buying product j.

Example. Random Coefficients Logit and BLP

Consider the random coefficients Logit model where:

$$U_{hj}=-\alpha\ p_j+\mathbf{X}_j\ \boldsymbol{\beta}+\xi_j+\widetilde{v}_{hj}+\varepsilon_{hj}$$

- where

$$\widetilde{v}_{hj} = -v_h^{\alpha}\ p_j + v_h^{\beta_1}\ X_{1j} + ... + v_h^{\beta_K}\ X_{Kj}$$

that has an heteroscedastic normal distribution. Then,

$$s_j = \frac{q_j}{H} = \int \frac{\exp\left\{-\alpha\ p_j + \mathbf{X}_j\ \beta + \xi_j + \widetilde{v}_{hj}\right\}}{1 + \sum_{k=1}^{J} \exp\left\{-\alpha\ p_k + \mathbf{X}_k\ \beta + \xi_k + \widetilde{v}_{hk}\right\}}\ f(\widetilde{\mathbf{v}}_h|\Sigma)\ d\widetilde{\mathbf{v}}_h$$

where $\delta_j \equiv -\alpha\ p_j + \mathbf{X}_j\ \beta + \xi_j$ represents the mean utility of buying product j.

In general, for any distribution of consumer heterogeneity $\boldsymbol{v}_h$, the model implies a mapping between the $J \times 1$ vector of mean utilities $\boldsymbol{\delta} = \{\delta_j : j = 1, 2, ..., J\}$ and the $J \times 1$ vector of market shares $\mathbf{s} = \{s_j : j = 1, 2, ..., J\}$.

$$s_j = \sigma_j(\boldsymbol{\delta}, \Sigma) \qquad \text{for } j = 1, 2, ..., J$$

or in vector form $\mathbf{s} = \sigma(\boldsymbol{\delta}, \Sigma)$.

3.2. Berry's Inversion Property. Under some regularity conditions (more later) the system $\mathbf{s} = \sigma(\boldsymbol{\delta}, \Sigma)$ is invertible in $\boldsymbol{\delta}$ such that there is an inverse function $\sigma^{-1}(.)$ and:

$$\boldsymbol{\delta} = \sigma^{-1}\left(\mathbf{s};\ \Sigma\right)$$

or for a product j, $\delta_j = \sigma_j^{-1}\left(\mathbf{s};\ \Sigma\right)$. The form of the inverse mapping σ^{-1} depends on the PDF $f_{\widetilde{v}}$.

Example: Logit model. In the logit model:

$$s_j = \frac{\exp\left\{\delta_j\right\}}{1 + \sum_{k=1}^{J} \exp\left\{\delta_k\right\}}$$

Let s_0 be the market share of the "outside good" such that, $s_0 = 1 - \sum_{k=1}^{J} s_k$. Then,

$$\delta_j = \ln\left(\frac{s_j}{s_0}\right)$$

That is, $\sigma_j^{-1}\left(\mathbf{s};\ \Sigma\right) = \ln\left(\frac{s_j}{s_0}\right)$ and we have a closed form expression for the inverse mapping σ_j^{-1}.

We also have a closed-form expression for σ_j^{-1} in the case of the Nested Logit model (without the IIA). However, in general, for the RC or BLP model we do not have a closed form expression for σ_j^{-1}. Berry and BLP propose a fixed point algorithm to compute the δ's They propose the following fixed point mapping: $\boldsymbol{\delta} = F\left(\boldsymbol{\delta};\ \mathbf{s},\ \Sigma\right)$ or $\delta_j = F_j\left(\boldsymbol{\delta};\ \mathbf{s},\ \Sigma\right)$ where:

$$F_j\left(\boldsymbol{\delta};\ s_j,\ \Sigma\right) \equiv \delta_j + \ln(s_j) - \ln\left(\sigma_j\left(\boldsymbol{\delta};\ \Sigma\right)\right)$$

This mapping is a contraction as long as the values of $\boldsymbol{\delta}$ are not too small. Under this condition, the mapping has a unique fixed point and we can find it by using fixed point iteration algorithm.

Iterative procedure: Start with initial $\boldsymbol{\delta}^0$. At iteration $R+1$:

$$\boldsymbol{\delta}^{R+1} = F\left(\boldsymbol{\delta}^R;\, \mathbf{s},\, \Sigma\right) = \delta_j^R + \ln(s_j) - \ln\left(\sigma_j\left(\boldsymbol{\delta}^R;\, \Sigma\right)\right)$$

Iterate until convergence.

3.3. Estimation. Suppose that the researcher has a dataset from a **single market** at only **one period** but for a product with many varieties: $M = T = 1$ but large J (e.g., 100 varieties). The researcher observes:

$$Data = \{q_j,\, X_j,\, p_j : j = 1, 2, ..., J\}$$

Given these data, the researcher is interested in the estimation of the parameters of the demand system: $\theta = \{\alpha,\ \beta,\ \Sigma\}$. For the moment, we assume that market size H is known to the researcher. But it can be also estimated as a parameter. For the asymptotic properties of the estimators, we consider that $J \to \infty$.

The model is:

$$s_j = \frac{q_j}{H} = \sigma_j\left(\mathbf{X},\, \mathbf{p},\, \boldsymbol{\xi};\, \theta\right)$$

Unobserved characteristics $\boldsymbol{\xi}$ are correlated with $\mathbf{p}$ (endogeneity). Dealing with endogeneity in nonlinear models is complicated. Without further restrictions, we need full MLE: an specification of the model of $\mathbf{p}$ and a parametric specification of the distribution of $\boldsymbol{\xi}$. BLP contribution was to show that there is a general class of models (BLP models) with an **invertibility property**. This property implies that we can represent the model using a equation where the unobservables $\boldsymbol{\xi}$ enter additively and linearly, and then we can estimate these equations using GMM.

Consider the system represented using the inverse mapping:

$$\sigma_j^{-1}\left(\mathbf{s};\, \Sigma\right) = -\alpha\ p_j + \mathbf{X}_j\ \boldsymbol{\beta} + \xi_j$$

We want to estimate $\theta = \{\alpha,\ \beta,\ \Sigma\}$ in this model. For instance, in the logit model:

$$\ln\left(\frac{s_j}{s_0}\right) = -\alpha\ p_j + \mathbf{X}_j\ \boldsymbol{\beta} + \xi_j$$

Assumption: $E\left(\xi_j \mid \mathbf{X}_1,\, ...,\, \mathbf{X}_J\right) = 0$

BLP Instruments. Under the previous assumption, we can use the characteristics of other products $(\mathbf{X}_k : k \neq j)$ as instruments for p_j. For instance, we can use as vector of instruments:

$$\mathbf{Z}_j = \frac{1}{J-1}\sum_{k \neq j} \mathbf{X}_k$$

It is clear that $E\left(\mathbf{Z}_j\ \xi_j\right) = 0$, and we can estimate θ using GMM based on the sample moment conditions:

$$m(\theta) = \frac{1}{J}\sum_{j=1}^{J}\left[\begin{array}{c}\mathbf{X}_j \\ \mathbf{Z}_j\end{array}\right]\left(\sigma_j^{-1}\left(\mathbf{s};\ f_{\widetilde{v}}\right) + \alpha\ p_j - \mathbf{X}_j\ \boldsymbol{\beta}\right) = 0$$

GMM estimator:

$$\widehat{\theta} = \arg\min_{\theta}\left[m(\theta)'\ \mathbf{W}\ m(\theta)\right]$$

When J is large, a possible concern with the instruments $\mathbf{Z}_j = \frac{1}{J-1}\sum_{k\neq j}\mathbf{X}_k$ is that they may have very little sample variability across j. Then, instead we can define a set of "neighbors" for each product j:

$$N_j = \{k \neq j : \|\mathbf{X}_k - \mathbf{X}_j\| \leq \tau\}$$

And construct the instruments:

$$\mathbf{Z}_j = \frac{1}{|N_j|}\sum_{k\in N_j}\mathbf{X}_k$$

This $\mathbf{Z}_j$ has more sample variability but it is also more correlated with $\mathbf{X}_j$. Trade-off. Optimal instruments: two step method in Newey (1990).

3.3.1. *Estimation of the Logit model.* Some of the first applications of the Logit model to demand systems with aggregate data were Manski (*Transportation Research*, 1983) and Berkovec and Rust (*Transportation Research*, 1985). Consider the logit model:

$$\log\left(\frac{s_j}{s_0}\right) = x_j\beta - \alpha\ p_j + \xi_j$$

This model solves three of the problems associated to the estimation of demand systems in product space.

First, **the number of parameters to estimate does not increase with the number of products** J. It increases only with the number of observed characteristics. Therefore, we can estimate with precision demand systems where J is large.

Second, the parameters are not product-specific but characteristic-specific. Therefore, given β and α **we can predict the demand of a new hypothetical product which have never been introduced in the market**. Suppose that the new product has observed characteristics $\{x_{J+1}, p_{J+1}\}$ and $\xi_{J+1} = 0$. For the moment, assume also that: (1) incumbent firms do not change their prices after the entry of the new product; and (2) incumbent firms do not exit or introduce new products after the entry of the new product. Then, the demand of the new product is:

$$q_{J+1} = S\ \frac{\exp\{x_{J+1}\beta - \alpha p_{J+1}\}}{1 + \sum_{k=1}^{J+1}\exp\{x_k\beta - \alpha p_k + \xi_k\}}$$

Note that to obtain this prediction we need also to use the residuals $\{\xi_k\}$

Third, but not less important, **the model provides valid instruments for prices which do not require one to observe cost shifters**. In the equation for product j, the characteristics of other products, $\{x_k : k \neq j\}$, are valid instruments for the price of product j. To see this note that the variables $\{x_k : k \neq j\}$: (1) do not enter in the equation for $\log(s_j/s_0)$; (2) are not correlated with the error term ξ_j; and (3) they are correlated with the price p_j. Condition (3) is not obvious, and in fact it depends on an assumption about price decisions. Suppose that product prices are the result of price competition between the firms that produce these products. For simplicity, suppose that there is on firm per product. The profit function of firm j is:

$$\pi_j = p_j \ q_j - C_j(q_j) - F_j$$

where $C_j(q_j)$ and F_j are the variable and the fixed costs of producing j, respectively. The first order conditions for firm j's best response price is:

$$q_j + \left[p_j - C_j'(q_j)\right] \frac{\partial q_j}{\partial p_j} = 0$$

For the Logit model, $\partial q_j / \partial p_j = -\alpha q_j (1 - s_j)$. Then,

$$p_j = C_j'(q_j) + \frac{1}{\alpha(1 - s_j)}$$

Though this is just an implicit equation, it makes it clear that p_j depends (through s_j) on the characteristics of all the products. If $x_k \beta$ (for $k \neq j$) increases, then s_j will go down, and according to the previous expression the price p_j will also decrease. Therefore, we can estimate the demand parameters β and α by IV using as instruments of prices the characteristics of the other products.

(a) Estimation of pricing equation

- Estimation of price-cost margins
- Counterfactuals with margins.
- Contribution of product characteristics to price-cost margins.
- Estimation of variable costs and of returns to scale.

3.3.2. *The IIA Property of the Logit Model.* In general, the more flexible is the structure of the unobserved consumer heterogeneity, the more flexible and realistic can be the elasticities of substitution between products that the model can generate. The logit model imposes very strong, and typically unrealistic, restrictions on demand elasticities. The random coefficients model generate much more flexible elasticities.

In discrete choice models, the IIA can be considered as an axiom of consumer choice that establishes that the relatively likelihood that a consumer chooses two alternative, say j and

k, should not be affected by the availability or the attributes of other alternatives:

$$\frac{\Pr(d_{hj}=1)}{\Pr(d_{hk}=1)} \text{ depends only on attributes of } j \text{ and } k$$

While IIA may be a reasonable assumption when we study **demand of single individual,** it is very restrictive when we look at the demand of multiple individuals because these individuals are heterogeneous in their preferences. The logit model implies IIA. But the RC model does not impose this property.

In the logit model:

$$\frac{\Pr(d_{hj}=1)}{\Pr(d_{hk}=1)} = \frac{s_j}{s_k} = \frac{\exp\left\{-\alpha\ p_j + \mathbf{X}_j\ \beta + \xi_j\right\}}{\exp\left\{-\alpha\ p_k + \mathbf{X}_k\ \beta + \xi_k\right\}} \Rightarrow \text{ IIA}$$

This implies that cross demand elasticities:

$$\frac{\partial \ln q_j}{\partial \ln p_k} = -\alpha\ p_k\ s_k$$

that is the same for any product j.

A 1% increase in the price of product k implies the same % increase in the demand of any product other than j.

This is very unrealistic: Example demand of automobiles.

Let H_k^0 be the group of consumers who were buying product k but decide to substitute product k by other products when p_k increases in a certain amount.

How is the pattern of substitution? It depends on consumers heterogeneity in preferences.

Consumers in H_k^0 have a relatively high value of ε_{hk}. But the logit implies that for consumers in H_k^0 the probability distribution of $\{\varepsilon_{hj} : j \neq k\}$ is the same that for the whole population of consumers in the market.

Therefore, the pattern of substitution is proportional to the market shares:

$$\frac{\partial s_j}{\partial p_k} = -\alpha\ s_j\ s_k \qquad \text{such that } \frac{\partial \ln s_j}{\partial \ln p_k} = -\alpha\ p_k\ s_k$$

This restriction does not appear in a model where the unobserved heterogeneity does not have an iid extreme value distribution. In particular, in the RC model, consumers in group H_k^0 have a relatively high values of the v_{hk} that correspond to high attributes of product k. Therefore, the consumers in this group will be more likely to choose other products that have similar attributes as product k.

4. Recommended Exercises

4.1. Exercise 1.

Bibliography

[1] Ackerberg, D., L. Benkard, S. Berry, and A. Pakes (2007): "Econometric Tools for Analyzing Market Outcomes," Chapter 63 in Handbook of Econometrics, vol. 6A, James J. Heckman and Ed Leamer, eds. North-Holland Press.

[2] Ackerberg, D. and Rysman, M. "Unobserved Product Differentiation in Discrete Choice Models: Estimating Price Elasticities and Welfare Effects," RAND Journal of Economics, 36, 771-788.

[3] Aguirregabiria, V. and C-Y. Ho (2011): "A dynamic oligopoly game of the US airline industry: Estimation and policy experiments," Journal of Econometrics, forthcoming.

[4] Arellano, M. and S. Bond (1991): "Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations," Review of Economic Studies, 58, 277-297.

[5] Berry, S. (1994): "Estimating Discrete Choice Models of Product Differentiation," RAND Journal of Economics, 25, 242-262.

[6] Berry, S., J. Levinsohn and A. Pakes (1995): "Automobile Prices in Market Equilibrium," Econometrica, 60(4), 889-917.

[7] Berry, S., J. Levinsohn and A. Pakes (1999): "Voluntary Export Restraints on Automobiles: Evaluating a Strategic Trade Policy," American Economic Review, 89(3), 400-430.

[8] Berry, S., J. Levinsohn and A. Pakes (2004): "Differentiated Products Demand Systems from a Combination of Micro and Macro Data: The New Vehicle Market," Journal of Political Economy, 112(1), 68-104.

[9] Berry, S., O. Linton, and A. Pakes (2004): "Limit Theorems for Estimating the Parameters of Differentiated Product Demand Systems," Review of Economic Studies, 71(3), 613-654.

[10] Berry, S. and A. Pakes (2007): "The Pure Characteristics Demand Model," International Economics Review, 48(4), Special Issue Economics to Econometrics: Contributions in Honor of Daniel L. McFadden, Charles F. Manski and Whitney Newey, eds., pp. 1193-1225.

[11] Boskin, M., E. Dulberger, R. Gordon, Z. Griliches, and D. Jorgenson (1998): "Consumer Prices, the Consumer Price Index and the Cost of Living," Journal of Economic Perspectives, 12, 3–26.

[12] Boskin, M., E. Dulberger, R. Gordon, Z. Griliches, and D. Jorgenson (1997): "The CPI Commission: Findings and Recommendations," American Economic Review, 87, 78–83.

[13] Christensen, L., D. Jorgenson, and L. Law (1975): "Transcendental Logarithmic Utility Functions," American Economic Review, 65, 367-383.

[14] Deaton; A. and J. Muellbauer (1980a): "An Almost Ideal Demand System," American Economic

[15] Review, 70(3), 312-326.

[16] Deaton, A. and J. Muellbauer (1980b): "Economics and Consumer Behavior," Cambridge University Press. Cambridge, UK.

[17] Draganska, M., M. Mazzeo, and K. Seim. (2009): "Beyond Plain Vanilla: Modeling Joint Product Assortment and Pricing Decisions," Quantitative Marketing & Economics 7(2).

[18] Eales, J. and L. Unnevehr (1993): "Simultaneity and Structural Change in U.S. Meat Demand," American Journal of Agricultural Economics , 75(2), 259-268.

[19] Geary, R. (1950): "A note on "A constant-utility index of the cost of living"." The Review of Economic Studies, 18, 65-66.

[20] Goldberg, P. (1995): "Product Differentiation and Oligopoly in International Markets: The Case of the U.S. Automobile Industry," Econometrica, 63, 891-951.

[21] Hausman, J. (1996): "Valuation of new goods under perfect and imperfect competition." in T. Bresnahan and R. Gordon (Eds.), The Economics of New Goods, Studies in Income and Wealth Vol. 58, 207-248. University of Chicago Press.

[22] Hausman, J. (2003): "Sources of Bias and Solutions to Bias in the Consumer Price Index," Journal of Economic Perspectives, 17(1), 23-44.

[23] Hausman, J., G. Leonard, and J. Zona (1994): "Competitive Analysis with Differentiated Products," Annales D'Economie et de Statistique, 34.

[24] Lerman, S. and C. Manski, 1981, On the Use of Simulated Frequencies to Approximate Choice Probabilities In C. Manski and D. McFadden (eds.), Structural Analysis of Discrete Data with Econometric Applications, MIT Press, Cambridge, MA.

[25] McFadden, D. (1974): "Conditional Logit Analysis of Qualitative Choice Behavior," in P. Zarembka (ed.), *Frontiers in Econometrics*, 105-142, Academic Press. New York.

[26] Melnikov, O., 2000, Demand for Differentiated Durable Products: The Case of the U.S. Computer Printer Market. Manuscript. Department of Economics, Yale University.

[27] Nevo, A. (2001): "Measuring Market Power in the Ready-to-Eat Cereal Industry," Econometrica,69(2), 307-342.

[28] Nevo, A. (2011): "Empirical Models of Consumer Behavior," *Annual Review of Economics*,

[29] Pakes, A. and D. Pollard (1989): "Simulation and the Asymptotics of Optimization Estimators," Econometrica, 57, 1027-1057.

[30] Pakes, Ariel (2003): "A Reconsideration of Hedonic Price Indexes with an Application to PC's," Ameri can Economic Review, 93(5), 1578-1596.

[31] Petrin, A. (2002): "Benefits of New Products," Journal of Political Economy, 110, 705–729.

[32] Shum, M. (2004): "The Effects of Advertising in Brand Loyal Markets: the Case of Breakfast Cereals." Journal of Economics and Management Strategy, 13, 241–272.

[33] Stern, S. (1997): "Simulation-based estimation," Journal of Economic Literature, 35, 2006–2039.

[34] Stone, R. (1954): "Linear expenditure systems and demand analysis: an application to the pattern of British demand," The Economic Journal, 64, 511-527.

[35] Su et al on MPEVC applied to BLP. Econometrica.

[36] Theil, H. (1975): "Theory and Measurement of Consumer Demand," Vols. I and 11, North-Holland, Amsterdam.

[37] Trajtenberg, M. (1989): "The Welfare Analysis of Product Innovations, with an Application to Computed Tomography Scanners," Journal of Political Economy, 97, 444–479.

[38] Verbeke, W. and R. Ward (2001): "A fresh meat almost ideal demand system incorporating negative TV press and advertising impact," Agricultural Economics, 25, 359-374.