2. Consumer Demand

2.1 Introduction

The estimation of demand equations is a fundamental component in most empirical applications in IO.[1] 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 prices or quantities. 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. 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. Demand functions are a representation of consumers' valuation of products. Because we cannot observe consumer utility or satisfaction directly, we obtain consumer preferences by estimating demand equations. As such, the estimation of demand is fundamental in the evaluation of the consumer welfare gains or losses associated with taxes, subsidies, new products, or mergers. Finally, demand estimation can also be used to improve our measures of Cost-of-Living indices (see Hausman, 2003, and Pakes, 2003).[2]

[1] Ackerberg et al. (2007) and Nevo (2011) are survey papers on demand estimation.

[2] 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. CPIs 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

Most products that we find in today's markets are differentiated products: automobiles; smartphones; laptop computers; or supermarket products such as ketchup, soft drinks, 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, which was the standard approach in EIO until the 1990s, and demand systems in characteristics space. We will see in this chapter that the model in characteristics space has several advantages over the model in product space, which has made it the predominant approach in empirical IO over the last two decades.

2.2 Demand systems in product space

2.2.1 Model

In this model, consumer preferences are defined over products themselves. Consider J different products that we index by $j \in \{1, 2, ..., J\}$. These J products may include all the product categories that an individual consumer may consume (for instance, food, transportation, clothing, entertainment) and all the varieties of products within each category (for instance, every possible variety of computers, or of automobiles). This means that the number of products J can be of the order of millions. Section 2.2.2 shows how to use multi-stage budgeting to deal with this high dimensionality problem. For this purpose, it is convenient to introduce "product zero" that we denote as the *outside product*, which represents all the other products that are not products 1 to J.

Let q_j denote the quantity of product j that a consumer buys and consumes, and let $(q_0, q_1, ..., q_J)$ be the vector with the purchased quantities of all the products, including the outside good. The price of the outside good is normalized to one, such that q_0 represents the dollar expenditure in goods other than 1 to J. The consumer has a utility function $U(q_0, q_1, ..., q_J)$ defined over the vector of quantities. The consumer's problem consists of choosing the vector $(q_0, q_1, ..., q_J)$ which maximizes her utility subject to her budget constraint.

$$\begin{array}{cc} \max\limits_{\{q_0, q_1, ..., q_J\}} & U(q_0, q_1, ..., q_J) \\ \\ subject\ to: & q_0 + p_1\, q_1 + ... + p_J\, q_J \leq y \end{array} \tag{2.1}$$

where p_j is the price of product j, and y is the consumer's disposable income. We can define the Lagrangian problem:

$$\max_{\{q_0, q_1, ..., q_J\}} \quad U(q_0, q_1, ..., q_J) + \lambda \left[y - q_0 - p_1\, q_1 - ... - p_J\, q_J \right] \tag{2.2}$$

The first order conditions are:

$$\begin{array}{rcl} U_j - \lambda p_j & = & 0 \text{ for } j = 0, 1, ..., J; \\ \\ y - q_0 - p_1 q_1 - ... - p_J q_J & = & 0 \end{array} \tag{2.3}$$

in the CPI as a cost of living index.Hausman (2003) and Pakes (2003) argue that the estimation of demand systems provides a possible solution to these sources of bias in the CPI.

where U_j represents the marginal utility of product j. The demand system is the solution to this optimization problem. We can represent this solution in terms of J functions, one for each product. These are the *Marshallian demand equations*:

$$\begin{array}{rcl} q_0 & = & f_0\left(p_1, p_2, ..., p_J, y\right) \\ q_1 & = & f_1\left(p_1, p_2, ..., p_J, y\right) \\ ... & & ... \\ q_J & = & f_J\left(p_1, p_2, ..., p_J, y\right) \end{array} \tag{2.4}$$

Function f_j provides the optimal consumption of product j as a function of prices and income.

The form of these functions depends on the form of the utility function U. Different utility functions imply different demand systems. Not every system of equations that relates quantities and prices is a demand system. It should come from the solution to the consumer problem for a given utility function. This has two clear implications on a demand system. First, a demand system should satisfy the *adding up condition* $\sum_{j=0}^{J} p_j\, f_j\left(p_1, p_2, ..., p_J, y\right) = y$. And second, it should be homogeneous of degree zero in prices and income: for any scalar $\delta \geq 0$, we have that $f_j\left(\delta p_1, \delta p_2, ..., \delta p_J, \delta y\right) = f_j\left(p_1, p_2, ..., p_J, y\right)$ for any product j.

A substantial part of the empirical literature on demand deal with finding utility functions that generate demand systems with two important practical features. First, they are simple enough to be estimable using standard econometric methods such as linear regression. And second, they are flexible in the sense of allowing for rich patterns in the elasticities of substitution between products. In general, there is a trade-off between these two features. The following are some examples of models that have been considered in the literature. They are shorted chronologically.

The Linear Expenditure demand system

Consider the Stone-Geary utility function:

$$U = (q_0 - \gamma_0)^{\alpha_0}\, (q_1 - \gamma_1)^{\alpha_1}\, ...\, (q_J - \gamma_J)^{\alpha_J} \tag{2.5}$$

where $\{\alpha_j, \gamma_j : j = 1, 2, ..., J\}$ are parameters. The parameter γ_j can be interpreted as the minimum amount of consumption of good j that a consumer needs to "survive". Parameter α_j represents the intensity of product j in generating utility. More formally, α_j is the elasticity of utility with respect to the amount of product j. Without loss of generality, given the ordinality of the utility function, we consider that $\sum_{i=0}^{J} \alpha_i = 1$. This utility function was first proposed by Geary (1950), and Stone (1954) was the first to estimate the Linear Expenditure System. In the Appendix to this chapter, section 2.5.1, we derive the expression for the demand equations of the Linear Expenditure System. They have the following form:

$$q_j = \gamma_j + \alpha_j \left[\frac{y - P_\gamma}{p_j}\right] \tag{2.6}$$

where P_γ is the aggregate price index $\sum_{i=0}^{J} p_i\, \gamma_i$.

This system is convenient because of its simplicity. Suppose that we have data on individual purchases and prices over T periods of time ($t = 1, 2, ..., T$): $\{q_{0t}, q_{1t}, ..., q_{Jt}\}$

and $\{p_{1t}, p_{2t}, ..., p_{Jt}\}$.[3] The model implies a system of J linear regressions. For product j:

$$q_{jt} = \gamma_j + \alpha_j \, \frac{y_t}{p_{jt}} + \beta_{j0} \, \frac{p_{0t}}{p_{jt}} + ... + \beta_{jJ} \, \frac{p_{Jt}}{p_{jt}} + \xi_{jt} \tag{2.7}$$

with $\beta_{jk} = -\alpha_j \, \gamma_k$. Variable ξ_{jt} is an error term that can come, for instance, from measurement error in purchased quantity q_{jt}, or from time variation in the coefficient γ_j. The intercept and slope parameters in these linear regression models can be estimated using instrumental variable methods.

However, the model is also very restrictive. Note that for any $j \neq k$, we have that $\dfrac{\partial q_j}{\partial p_k} = -\alpha_j \gamma_k / p_j < 0$, such that all the cross-price elasticities are negative. This implies that all the products are complements in consumption. This is not realistic in most applications, particularly when the goods under study are varieties of a differentiated product.

Constant Elasticity of Substitution demand system

Consider the Constant Elasticity of Substitution (CES) utility function:

$$U = \left(\sum_{j=0}^{J} q_j^{\sigma} \right)^{1/\sigma} \tag{2.8}$$

where $\sigma \in [0,1]$ is a parameter that represents the degree of substitution between the $J+1$ products. The marginal utilities are:

$$U_j = q_j^{\sigma-1} \frac{U}{\sum_{i=0}^{J} q_i^{\sigma}} \tag{2.9}$$

For any two pairs of products, j and k, we have that $\dfrac{\partial^2 U}{\partial q_j \partial q_k} < 0$, such that all the products are substitutes in consumption.

Given the CES utility function, we derive in the Appendix the following expression for the demand equations:

$$q_j = \frac{y}{P_{\sigma}} \, \left[\frac{p_j}{P_{\sigma}} \right]^{-1/(1-\sigma)} \tag{2.10}$$

where P_{σ} is the following aggregate price index:

$$P_{\sigma} = \left(\sum_{j=0}^{J} p_j^{-\sigma/(1-\sigma)} \right)^{-(1-\sigma)/\sigma} \tag{2.11}$$

The CES model is also very convenient because of its simplicity. Suppose that we have data of individual purchases and prices over T periods of time. The model implies the following log-linear regression model:

$$\ln \left(\frac{q_{jt}}{y_t} \right) = \beta_0 + \beta_1 \, \ln(p_{jt}) + \beta_2 \, \ln\left(P_{\sigma t}\right) + \xi_{jt} \tag{2.12}$$

[3] Given information on household income, y_t, the consumption of product zero can be obtained using the budget constraint, $q_{0t} = y_t - \sum_{j=1}^{J} p_j \; q_j$.

where $\beta_1 = -1/(1-\sigma)$, and $\beta_2 = \sigma/(1-\sigma)$. The error term ξ_{jt} can be interpreted as measurement error in quantities. The construction of the true price index $P_{\sigma t}$ requires knowledge of the parameter σ. To deal with this issue several approaches have been used in the literature: (a) approximating the true price index with a conjecture about σ; (b) controlling for the term $\beta_2 \ln(P_{\sigma t})$ y including time dummies; (c) estimating the model in deviations with respect to the equation for the outside product, $\ln\left(\frac{q_{jt}}{y_t}\right) - \ln\left(\frac{q_{0t}}{y_t}\right) = \beta_1$ $\ln(p_{jt}) + \xi_{jt} - \xi_{0t}$; and (d) taking into account the structure of the price index as a function of prices and σ and estimating the model using nonlinear least squares.

The demand elasticity β_1 can be estimated using a standard method for linear regression models. For instance, if the number of products in our dataset is large relative to the number of time periods, one can control for the time-effects using time dummies, and β can be estimated using OLS or IV methods.

This model imposes strong restrictions on consumer behavior. In particular, the elasticity of substitution between any pair of products is exactly the same. For any three products, say j, k, and i:

$$Elasticity_{k,j} = \frac{\partial \ln q_k}{\partial \ln p_j} = \frac{-\sigma}{1-\sigma} = \frac{\partial \ln q_i}{\partial \ln p_j} = Elasticity_{i,j}. \tag{2.13}$$

Having the same degree of substitution for all pairs of products can be quite unrealistic in most applications in IO. In fact, there are many industries that have both products that are close substitutes as well as products that are relatively unique. In such industries, we would expect an increase in the price of a product with many close substitutes to generate a substantial reduction in quantity, while this would not be the case if the product was relatively unique (i.e. did not have close substitutes).

Deaton and Muellbauer "Almost Ideal" demand system

The "Almost Ideal" demand system (AIDS) was proposed by Deaton and Muellbauer (1980, 1980). Because of its flexibility, it is the most popular specification in empirical applications where preferences are defined on the product space. The standard derivation of the AIDS does not start from the utility function but from the Expenditure Function of the model. The *expenditure function* of a demand system, $E(u, \mathbf{p})$ is defined as the minimum consumer expenditure required to achieve a level of utility u given the vector of prices $\mathbf{p} = (p_1, p_2, ..., p_J)$.

$$E(u, \mathbf{p}) = \min_{q_0, q_1, \dots, q_J} \textstyle\sum_{j=0}^{J} p_j \; q_j \quad \text{subject to: } U(q_0, q_1, ..., q_J) = u \tag{2.14}$$

Given its definition, the expenditure function is non-decreasing in all its arguments, and it is homogeneous of degree one in prices: for any $\delta \geq 0$, $E(u, \delta\ \mathbf{p}) = \delta\ E(u, \mathbf{p})$. Shephard's Lemma establishes that the derivative of the expenditure function with respect to the price of product j is the Hicksian or compensated demand function, $h_j(u, \mathbf{p})$:

$$q_j = h_j(u, \mathbf{p}) = \frac{\partial E(u, \mathbf{p})}{\partial p_j} \tag{2.15}$$

Similarly, combined with the condition that income is equal to total expenditure, $y = E(u, \mathbf{p})$, Shephard's Lemma implies that the partial derivative of the log-expenditure function with respect to the log-price of product j is equal to the expenditure share of the product, $w_j \equiv p_j q_j / y$.

$$w_j = \frac{p_j\, q_j}{y} = \frac{\partial \ln E(u, \mathbf{p})}{\partial \ln p_j} \tag{2.16}$$

Therefore, given a expenditure function that is consistent with consumer theory (non-decreasing and homogeneous of degree one), we can derive the demand system. Deaton and Muellbauer propose the following log-expenditure function:

$$\ln E(u, \mathbf{p}) = a(\mathbf{p}) + b(\mathbf{p})\ u \tag{2.17}$$

with

$$\begin{array}{rcl} a(\mathbf{p}) & = & \sum_{j=1}^{J} \alpha_j\ \ln p_j + \frac{1}{2} \sum_{j=1}^{J} \sum_{k=1}^{J} \gamma_{jk}^{*}\ \ln p_j\ \ln p_k \\ b(\mathbf{p}) & = & \prod_{j=1}^{J} p_j^{\beta_j} \end{array} \tag{2.18}$$

Homogeneity of degree of the expenditure function requires the following restrictions on the parameters:

$$\sum_{j=1}^{J} \alpha_j = 1;\ \sum_{j=1}^{J} \gamma_{jk}^{*} = 0;\ \sum_{k=1}^{J} \gamma_{jk}^{*} = 0;\ \sum_{j=1}^{J} \beta_j = 0. \tag{2.19}$$

Applying Shephard's Lemma to this log-expenditure function, we can derive the following demand system represented in terms of expenditure shares:

$$w_j = \alpha_j + \beta_j\ \left[\ln(y) - \ln(P_{\alpha,\gamma})\right] + \sum_{k=1}^{J} \gamma_{jk}\ \ln p_k \tag{2.20}$$

where $\gamma_{jk} \equiv (\gamma_{jk}^{*} + \gamma_{kj}^{*})/2$ such that the model implies the symmetry condition $\gamma_{jk} = \gamma_{kj}$; and $P_{\alpha,\gamma}$ is a price index with the following form:

$$\ln P_{\alpha,\gamma} = \sum_{j=1}^{J} \alpha_j\ \ln p_j + \frac{1}{2} \sum_{j=1}^{J} \sum_{k=1}^{J} \gamma_{jk}\ \ln p_j\ \ln p_k \tag{2.21}$$

The number of free parameters in this demand system is $2J + \frac{J(J+1)}{2}$, which increases quadratically with the number of products.

Suppose that we have data on individual purchases, income, and prices over T periods of time. For each product j, we can estimate the regression equation:

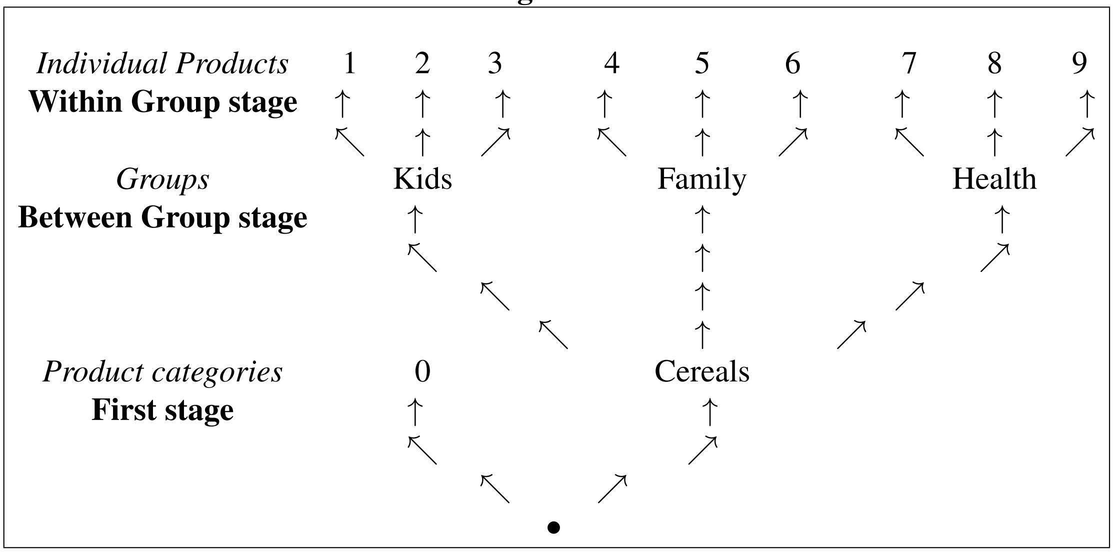
$$w_{jt} = \alpha_j + \beta_j\ \ln(y_t) + \gamma_{j1}\ \ln(p_{1t}) + ... + \gamma_{jJ}\ \ln(p_{Jt}) + \xi_{jt} \tag{2.22}$$

Since the number of parameters increases quadratically with the number of products, the estimation of this model (without restrictions on the parameters) requires that the number of observations T (either time periods or geographic markets) is substantially larger than the number of products J. For differentiated products with many varieties, say $J > 100$ (such as most differentiated products like automobiles, smartphones, cereals, beer, etc), the number of parameters can be of the order of several thousands such that this condition does not hold. Increasing the number of observations by using data from many consumers does not help in the estimation of price elasticities because consumers in the same market face the same prices, that is, prices do not have variation across consumers, only over time and geographic markets.

2.2.2 Multi-stage budgeting

To reduce the number of parameters when J is relatively large, Deaton and Muellbauer propose a multi-stage budgeting approach. Suppose that the $J+1$ products can be classified in G groups or segments. For instance, in the ready-to-eat cereal industry, some empirical studies distinguish three segments: *Kids*, *All family*, and *Health*. The following diagram presents the nested structure of the demand system.

Figure 2.1



Suppose that the utility function is:

$$U = v_0(q_0) + v_1(\widetilde{\mathbf{q}}_1) + \ldots + v_G(\widetilde{\mathbf{q}}_G) \tag{2.23}$$

where $\widetilde{\mathbf{q}}_g$ is the vector of quantities of product varieties in group g; and $v_g(\widetilde{\mathbf{q}}_g)$ is the sub-utility from group g. Then, the demand system at the lower stage, the **within-group stage**, is:

$$w_{jt} = \alpha_j^{(1)} + \beta_j^{(1)} \ln\left(\frac{e_{gt}}{P_{gt}}\right) + \sum_{k \in \mathscr{J}_g} \gamma_{jk}^{(1)} \ln(p_{kt}) \tag{2.24}$$

where e_{gt} is the expenditure from all the products in group g, and P_{gt} is a price index for group g. According to the model, this price index depends on the parameters of the model in group g. The number of parameters increases quadratically with J_g instead of with J. The demand system at the **group stage** is:

$$\frac{e_{gt}}{e_t} = \alpha_g^{(2)} + \beta_g^{(2)} \ln\left(\frac{e_t}{P_t}\right) + \sum_{g'=1}^{G} \gamma_{g,g'}^{(2)} \ln(P_{gt}) \tag{2.25}$$

where e_t is the total expenditure in the large category (for instance, cereals), and P_t is the price index for the category (for instance, cereals). Finally, at the top-stage, the **demand for the category** is:

$$\frac{e_t}{y_t} = \alpha^{(3)} + \beta^{(3)} \left[\ln(y_t) - \ln(P_t)\right] \tag{2.26}$$

This multi-stage budgeting model can reduce substantially the number of parameters. For instance, suppose that a differentiated product category, say cereals, has 50 products such that the number of parameters in the unrestricted model is $2*50+\frac{50(50+1)}{2}=1,325$. Now, suppose that we can divide the 50 products into 10 groups with 5 products each. This implies that at the within-group stage we have 250 parameters (25 for each group), in the group stage we have 75 parameters, and in the category stage we have 3 parameters, for a total of 328 parameters. Using one year of monthly data over 500 geographic markets, we have 6,000 observations. If these data have enough (exogenous) variation in prices, it seems possible to estimate this restricted system. This is the approach in Hausman (1996) that we describe in more detail in section **??**.

2.2.3 Estimation

In empirical work, the most commonly used demand systems are the Rotterdam Model (Theil, 1975), the Translog Model (Christensen, Jorgenson, and Lau, 1975), and the Almost Ideal Demand System (AIDS) (Deaton and Muellbauer, 1980). 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 (for instance, 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 models.

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 between 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:

$$w_{jt} = \mathbf{X}_t\ \alpha_j + \beta_j\ \ln(y_t/P_t) + \sum_{k=1}^{J} \gamma_{jk}\ \ln(p_{kt}) + \xi_{jt} \tag{2.27}$$

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

2.2.4 Some limitations

Demand systems in product space suffer of some limitations for the empirical work and questions that we stusy in IO.

(1) Every consumer purchases/consumes each of the J products. The system of demand equations that we have derived above is based on the assumption that the marginal conditions of optimality hold for every product. This means that the optimal bundle for a consumer is an interior solution such that $q_j > 0$ for every product j. This condition is very unrealistic when we consider the demand of differentiated products within a product category, for instance, the demand of automobiles. In this context, a consumer buys only one unit of a single variety (for instance, one Toyota Corola) or of a few varieties (for instance, one Toyota Corola, and one KIA Sorento minivan). To account for this type of consumer decisions, we need to model the consumer problem as a discrete choice model.

(2) Representative consumer. 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, for instance, 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.

In principle, demand systems in product space could be applied to household level data. Suppose that we have this type of data. Let us use the subindex h for households. The demand system becomes:

$$w_{jht} = \mathbf{X}_{ht}\ \alpha_j + [\mathbf{X}_{ht}\ \beta_j] \ln(y_{ht}/P_t) + \sum_{k=1}^{J} [\mathbf{X}_{ht}\ \gamma_{jk}] \ln(p_{kt}) + \xi_{jht} \tag{2.28}$$

where $\mathbf{X}_{ht}$ represents a vector of exogenous household characteristics, other than income. Now, α_j, β_j, and γ_{jk} are vectors of parameters with the same dimension as $\mathbf{X}_{ht}$. 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 elasticities are identified only from the time-series (or market) 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 clearly illustrates the problem of observed zero consumption of some products, that we have mentioned in point (1) above. Some households do not consume all the product categories, even when these categories are quite broad. For instance, vegetarian househods do not consume any meat. This class of model predicts that the household consumes a positive amount of every product category. This prediction is typically rejected when using household level data.

(3) The problem of too many parameters. In the standard model, the number of parameters is $2J + \frac{J(J+1)}{2}$, that is, 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.

(4) Finding instruments for prices. Most empirical applications of this class of models have ignored the potential endogeneity of prices.[4] However, it is well known that simultaneity and endogeneity are potentially important issues in any demand estimation. Prices are determined in the equilibrium of the market and depend on all the exogenous variables affecting demand and supply. Therefore, we expect prices to be correlated with the error terms ξ in the demand equations. Correlation between regressors and the error term implies that the OLS method is an inconsistent estimator of the parameters in demand equation. 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 would be to have information on production costs for each individual good. However, this type of information is rarely available.

(5) Predicting the 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 good $J+1$, requires estimates of the parameters associated with that good: α_{J+1}, β_{J+1} and $\{\gamma_{J+1,j} : j = 1,2,...,J+1\}$. Of course, this makes it impossible to make counterfactual predictions, that is, predicting the demand of a product that has not yet been introduced in any market. 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.

2.2.5 Dealing with limitations

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 was among the most prominent in the introduction of new brands within U.S. industries, with approximately 190 new brands 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 geographic 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 hundreds of brands, the model and the estimation concentrates on 20 brands classified into three segments: adult (7 brands), child (4 brands), and family

[4] 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.

(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 the data can be described as $\{p_{jmt}, q_{jmt} : j = 1, 2, ..., 20; m = 1, 2, ..., 7; t = 1, 2, ..., 137\}$. Quantities are measured in physical units. This dataset does not contain information on firms' costs, such as input prices or wholesale prices.

(2) Model. Hausman estimates an Almost-Ideal-Demand-System combined with a nested three-level structure. The nested structure is similar to the one described in the diagram of Figure 2.1. 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). The estimates are then used to construct the next level's price indexes, and to implement the estimation at the next level. At the lowest level, within a segment, the demand system is:

$$s_{jmt} = \alpha^1_{jm} + \alpha^2_t + \beta_j \ \ln(y_{gmt}) + \sum_{k=1}^{J} \gamma_{jk} \ \ln(p_{kmt}) + \xi_{jmt} \tag{2.29}$$

where y_{gmt} is overall expenditure in segment/group g. 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:

$$\ln(p_{jmt}) = \delta_j \ c_{jt} + \tau_{jm} + \kappa_{j1} \ \xi_{1mt} + ... + \kappa_{jJ} \ \xi_{Jmt} \tag{2.30}$$

All the components in the right-hand-side, δ_j, c_{jt}, τ_{jm}, κ's, and ξ's, are unobservable to the researcher. Variable c_{jt} represents a cost at the product level that is common to all the city markets. Variable τ_{jm} is a city-brand fixed effect that captures differences in transportation costs. The terms $\kappa_{j1} \ \xi_{1mt} + ... + \kappa_{jJ} \ \xi_{Jmt}$ captures how the price of product j in market m responds to local demand shocks, $\xi_{1mt}, \xi_{2mt}, ..., \xi_{Jmt}$. The identification assumption is that these demand shocks are not (spatially) correlated across markets:

$$\mathbb{E}(\xi_{jmt} \ \xi_{km't}) = 0 \qquad \text{for any } j,k \text{ and } m' \neq m \tag{2.31}$$

The assumption implies that, after controlling for brand-city fixed effects, all the correlations between prices at different locations come from correlations 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:

$$\overline{P}_{j(-m)t} = \frac{1}{M-1} \sum_{m' \neq m} p_{jm't} \tag{2.32}$$

(3) Evaluating the effects of new goods. Suppose that we define product J as being a "new" product in the market, although it is a product in our sample for which we have data on prices and quantities, and for which we can estimate all the parameters of the model including α_J^0, $\{\beta_{Jk}\}$ and γ_J. The expenditure function $e(\mathbf{p},u)$ for the Deaton and Muellbauer demand system is:

$$e(\mathbf{p},u) = \sum_{j=1}^{J} \alpha_j \ln(p_j) + \frac{1}{2} \sum_{j=1}^{J} \sum_{k=1}^{J} \gamma_{jk} \ln(p_j) \ln(p_k) + u \prod_{j=1}^{J} p_j^{\beta_j} \tag{2.33}$$

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. Suppose that we have estimated the demand parameters after the introduction of good J in the market, 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. The *virtual price* of good J – represented as p_J^* – is the price of this product that makes its demand equal to zero. That is, the virtual price of product J in market m at quarter t is the value p_{Jmt}^* that solves the following equation:

$$\hat{D}_{jmt}(p_{1mt}, p_{2mt}, ..., p_{Jmt}^*) = 0 \tag{2.34}$$

Note that this virtual price depends on the prices of the other products, as well as on other exogenous variables affecting demand.

Given the virtual price p_{Jmt}^*, we have a counterfactual scenario in which consumers do not buy product J in market m in period t. Hausman compares the factual situation with this counterfactual scenario. Let u_{mt} be the utility of the representative consumer in market m at period t with the new product: that is, $u_{mt} = \hat{V}(\mathbf{p}_{mt}, y_{mt})$. 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, we can measure the change in welfare associated to the introduction of the new product using the following *Equivalent Variation* measure:

$$EV_{mt} = \hat{e}(p_{1mt}, p_{2mt}, ..., p_{Jmt}^*, u_{mt}) - y_{mt} \tag{2.35}$$

Hausman considers this measure of consumer welfare.

This approach uses a market with prices and income $(p_{1mt}, p_{2mt}, ..., p_{Jmt}^*, y_{mt})$ as the counterfactual to measure the value of good J in a market with actual prices and income $(p_{1mt}, p_{2mt}, ..., p_{Jmt}, y_{mt})$. However, this choice of counterfactual does not account for the potential effect on prices of the introduction of the new product. In some applications, the welfare gains from these competition effects can be substantial and we are interested in measuring them. To measure these effects we should calculate equilibrium prices before and after the introduction of the new good. This requires the estimation not only of demand parameters but also of firms' marginal costs, as well as an assumption about competition (competitive market, Cournot, Bertrand). Though the Equivalent Variation presented above does not account for competition effects, it has some attractive features. First, it has a clear economic interpretation as the welfare gain *in the absence of competition effects*. Second, since it only depends on demand estimation, it is robust to misspecification of the supply side of the model.

2.3 Demand in characteristics space

2.3.1 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: for instance, 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, that is, 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 the "amount" of attribute k in brand j. For instance, in the case of laptops we could define the variables as follows: 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 ξ_j, that contains the "amounts" of the different unobservable attributes of variety 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 $(\mathbf{X}, \xi)$. The utility of consumer h if she consumes that product is $V_h(\mathbf{X}, \xi)$. Importantly, note that the utility function V_h is defined over any possible bundle of attributes $(\mathbf{X}, \xi)$ that may or may not exist in the market. For a product j that exists in the market and has attributes $(\mathbf{X}_j, \xi_j)$, this utility is $V_{hj} = V_h(\mathbf{X}_j, \xi_j)$. The total utility of a consumer has two additive components: the utility from this product, and the utility from other goods: $U_h = u_h(C) + V_h(\mathbf{X}, \xi)$, where C represents the amount of a composite good, and $u_h(C)$ is the utility from the composite good.

Consumers differ 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 υ_h that may be completely unobservable to the researcher. Therefore, we can write the utility of consumer h as:

$$U_h = u(C; \upsilon_h) + V(\mathbf{X}, \xi; \upsilon_h) \tag{2.36}$$

We also assume that there is continuum of consumers with measure H, such that υ_h has a well-defined density function f_υ in the market.

Each consumer buys at most one variety of the product (per period). Given her income, y_h, and the vector of product prices $\mathbf{p} = (p_1, p_2, ..., p_J)$, a consumer decides which variety 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:

$$\begin{array}{ll} \max\limits_{\{d_{h1}, d_{h2}, ..., d_{hJ}\}} & u(C; \upsilon_h) + \sum\limits_{j=1}^{J} d_{hj}\, V(\mathbf{X}_j, \xi_j; \upsilon_h) \\ subject\ to: & C + \sum_{j=1}^{J} d_{jh}\, p_j \leq y_h \\ & d_{hj} \in \{0,1\} \text{ and } \sum_{j=1}^{J} d_{jh} \in \{0,1\} \end{array} \tag{2.37}$$

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

$$\{d_j^*(\mathbf{X},\mathbf{p},y_h;\upsilon_h) = 1\} \Leftrightarrow$$

$$\left\{u(y_h - p_j;\upsilon_h) + V(\mathbf{X}_j,\xi_j;\upsilon_h) > u(y_h - p_k;\upsilon_h) + V(\mathbf{X}_k,\xi_k;\upsilon_h) \text{ for any } k \neq j\right\} \tag{2.38}$$

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

$$q_j(\mathbf{X},\mathbf{p},f) = \int d_j^*(\mathbf{p},y_h;\upsilon_h)\ ,\beta)\ f(\upsilon_h,y_h)\ d\upsilon_h\ d\mathbf{y}_h \tag{2.39}$$

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

Now, we provide specific examples of this general model. Each example is based on specific assumptions about the form of the utility function and the probability distribution of consumer heterogeneity. These examples are also important models which are workhorses in the literature on the estimation of demand of differentiated products.

2.3.2 Logit model

Consider the following restrictions on the general model presented above. First, the utility from the outside product is linear and the same for all the consumers: $u(C;\upsilon_h) = \alpha\ C$, where α is a parameter that represents the marginal utility of the composite good C. Second, the utility of purchasing product j is:

$$V(\mathbf{X}_j,\widetilde{\xi}_j,\upsilon_h) = \mathbf{X}_j\beta + \xi_j + \varepsilon_{hj}, \tag{2.40}$$

where ε's are unobservable random variables (for the researcher) which are independently and identically distributed (i.i.d.) over consumers and products with an Extreme Value Type 1 distribution.[5] Then, $U_{hj} = -\alpha\ p_j + \mathbf{X}_j\ \beta + \xi_j + \varepsilon_{hj}$ and the the Extreme Value assumption on the ε variables implies that the market shares have the following closed-form logit structure.

$$s_j = \frac{q_j}{H} = \frac{\exp\{\delta_j\}}{1+\sum_{k=1}^{J}\exp\{\delta_k\}} \tag{2.41}$$

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

The parameter α represents the marginal utility of income and it is measured in utils per dollar. In the vector β, the parameter β_k associated to characteristic X_{jk} (the $k-th$ element of vector $\mathbf{X}_j$) represents the marginal utility of this characteristic and it is measured in utils per unit of X_{jk}. Therefore, for any product attribute k, the ratio of parameters β_k/α is measured in dollars per unit of X_{jk} such that it is a monetary measure of the marginal utility of the attribute.

[5] For more information about the Extreme Value type I distribution see Appendix 11 on Random Utlity discrete choice logit models.

2.3.3 Nested Logit model

As explained below, the logit model imposes strong restrictions on the own and cross price elasticities of products. The Nested Logit model relaxes these restrictions.

Suppose that we partition $J+1$ products (including the outside product) in $G+1$ groups. We index groups of products by $g \in \{0,1,...,G\}$. Let $\mathscr{J}_g$ represent the set of products in group g. The utility function has the same structure as in the Logit model with the only (important) difference that the variables ε_{hj} have the structure of a nested logit model:

$$\varepsilon_{hj} = \lambda \ \varepsilon_{hg}^{(1)} + \varepsilon_{hj}^{(2)} \tag{2.42}$$

where $\varepsilon_{hg}^{(1)}$ and $\varepsilon_{hj}^{(2)}$ are i.i.d. Extreme Value type 1 variables, and λ is a parameter. This model implies the following closed-form expression for the market shares:

$$s_j = \frac{\exp\left\{\lambda \ I_g\right\}}{\sum_{g'=0}^{G} \exp\left\{\lambda \ I_{g'}\right\}} \frac{\exp\left\{\delta_j\right\}}{\sum_{k\in\mathscr{J}_g} \exp\left\{\delta_k\right\}} \tag{2.43}$$

where I_g is denoted the *inclusive value of group g* and it is defined as follows:

$$I_g \equiv \ln\left(\sum_{j\in\mathscr{J}_g} \exp\left\{\delta_j\right\}\right) \tag{2.44}$$

This inclusive value can be interpreted as the expected utility of a consumer who chooses group g knowing the δ values of the products in that group but before knowing the realization of the random variables $\varepsilon_{hj}^{(2)}$. That is,

$$I_g \equiv \mathbb{E}_{\varepsilon^{(2)}}\left(\max_{j\in\mathscr{J}_g} \ \left[\delta_j + \varepsilon_{hj}^{(2)}\right]\right)$$

where $\mathbb{E}_{\varepsilon^{(2)}}(.)$ represents the expectation over the distribution of the random variables $\varepsilon_{hj}^{(2)}$. Because of this interpretation, inclusive values are also denoted as *Emax values*. When the variables $\varepsilon_{hj}^{(2)}$ have a Extreme Value type 1 distribution, this Emax or inclusive value has the simple form presented above as the logarithm of the sum of the exponential of δ's.

The equation for the market shares in the nested Logit model has an intuitive interpretation as the product of between-groups and within-groups market shares. Let $s_g^* \equiv \sum_{j\in\mathscr{J}_g} s_j$ be the aggregate market share of all the products that belong to group g. And let $s_{j|g} \equiv s_j / \sum_{k\in\mathscr{J}_g} s_k$ be the market share of product j *within* its group g. By definition, we have that $s_j = s_g^* \ s_{j|g}$. The nested Logit model implies that within-group market shares have the logit structure $s_{j|g} = \exp\left\{\delta_j\right\} / \exp\{I_g\}$, and the group market shares have the logit structure $\exp\left\{\lambda \ I_g\right\} / \sum_{g'=0}^{G} \exp\left\{\lambda \ I_{g'}\right\}$.

Goldberg and Verboven (2001) estimate a nested logit model for the demand of automobiles in European car markets.

2.3.4 Random Coefficients Logit

Suppose that the utilities $V(\mathbf{X}_j, \widetilde{\xi}_j; \upsilon_h)$ and $u(C; \upsilon_h)$ are linear in parameters, but these parameters are household specific. That is, $U_{hj} = -\alpha_h \ p_j + \mathbf{X}_j \ \beta_h + \xi_j + \varepsilon_{hj}$ where ε's are still *i.i.d.* Extreme Value Type 1, and

$$\begin{bmatrix} \alpha_h \\ \beta_h \end{bmatrix} = \begin{bmatrix} \alpha \\ \beta \end{bmatrix} + v_h \quad \text{with } v_h \sim i.i.d.\ N(0,\Sigma) \tag{2.45}$$

Then, we can write utilities as:

$$U_{hj} = -\alpha \ p_j + \mathbf{X}_j \ \beta + \xi_j + \widetilde{v}_{hj} + \varepsilon_{hj} \tag{2.46}$$

where $\widetilde{v}_{hj} = -v_h^{\alpha} \ p_j + v_h^{\beta_1} \ X_{1j} + ... + v_h^{\beta_K} \ X_{Kj}$ has a heteroskedastic normal distribution. Then, the expression for the market shares is:

$$s_j = \frac{q_j}{H} = \int \frac{\exp\left\{\delta_j + \widetilde{v}_{hj}\right\}}{1 + \sum_{k=1}^{J} \exp\left\{\delta_k + \widetilde{v}_{hk}\right\}} \ f(\widetilde{\mathbf{v}}_h \mid \mathbf{p}, \mathbf{X}, \Sigma) \ d\widetilde{\mathbf{v}}_h \tag{2.47}$$

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

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

$$s_j = \sigma_j(\delta \mid \mathbf{p}, \mathbf{X}, \Sigma) \qquad \text{for } j = 1,2,...,J \tag{2.48}$$

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

Berry, Levinsohn, and Pakes (1995) estimate a random coefficients logit model to study the demand of automobiles in the US.

The importance of allowing for random coefficients. 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 strong, and typically unrealistic, restrictions on demand elasticities. The random coefficients model can generate more flexible elasticities.

In discrete choice models, the *Independence of Irrelevant Alternative* (IIA) is a property of consumer choice that establishes that the ratio between the probabilities that a consumer chooses two alternatives, say j and k, should not be affected by the availability or the attributes of other alternatives:

$$IIA : \frac{\Pr(d_{hj} = 1)}{\Pr(d_{hk} = 1)} \text{depends only on attributes of} j \text{ and } k \tag{2.49}$$

While IIA may be a reasonable assumption when we study the demand of single individual, it is quite restrictive when we look at the demand of multiple individuals because these individuals are heterogeneous in their preferences. The logit model implies IIA. In the logit model:

$$\frac{\Pr(d_{hj} = 1)}{\Pr(d_{hk} = 1)} = \frac{s_j}{s_k} = \frac{\exp\left\{-\alpha \ p_j + \mathbf{X}_j \ \beta + \xi_j\right\}}{\exp\left\{-\alpha \ p_k + \mathbf{X}_k \ \beta + \xi_k\right\}} \ \Rightarrow \ \text{IIA} \tag{2.50}$$

This property implies a quite restrictive structure for the cross demand elasticities. In the logit model, for $j \neq k$, we have that $\frac{\partial \ln s_j}{\partial \ln p_k} = -\alpha \ p_k \ s_k$, which 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.

2.3.5 Berry's Inversion Property

Berry (1994) shows that, under some regularity conditions (more later), the demand system $\mathbf{s} = \sigma(\delta \mid \mathbf{p}, \mathbf{X}, \Sigma)$ is invertible in δ such that there is an inverse function σ^{-1}, and:

$$\delta = \sigma^{-1}(\mathbf{s} \mid \mathbf{p}, \mathbf{X}, \Sigma) \tag{2.51}$$

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

This inversion property has important implications for the estimation of the demand system. Under this inversion, the unobserved product characteristics ξ_j enter additively in the equation $\delta_j = \sigma_j^{-1}(\mathbf{s} \mid \mathbf{p}, \mathbf{X}, \Sigma)$. Under this additivity, and the mean independence of the unobservables ξ_j conditional on the exogenous product characteristics $\mathbf{X}$, we can construct moment conditions and obtain GMM estimators of the structural parameters that deal with the endogeneity of prices.

Example: Logit model (Manski, 1983; Berkovec and Rust, 1985)**.** In the logit model, the demand system is $s_j = \exp\{\delta_j\}/D$, where $D \equiv 1 + \sum_{k=1}^{J} \exp\{\delta_k\}$, such that $\ln(s_j) = \delta_j - \ln(D)$. Let s_0 be the market share of the outside good such that, $s_0 = 1 - \sum_{k=1}^{J} s_k$. For the outside good, $s_0 = 1/D$, such that $\ln(s_0) = -\ln(D)$. Combining the equations for $\ln(s_j)$ and $\ln(s_0)$ we have that:

$$\delta_j = \ln\left(s_j\right) - \ln\left(s_0\right) \tag{2.52}$$

and this equation is the inverse mapping $\sigma_j^{-1}(\mathbf{s} \mid \mathbf{p}, \mathbf{X}, \Sigma)$ for the logit model. ■

Example: Nested Logit model. In the nested Logit model, the demand system is $s_j = s_g^*\ s_{j|g}$ where $s_g^* = \exp\{\lambda\ I_g\}/D$ and $s_{j|g} = \exp\{\delta_j\}/\exp\{I_g\}$, such that $\ln(s_j) = \ln(s_g^*) + \ln(s_{j|g})$ with $\ln(s_g^*) = \lambda\ I_g - \ln(D)$ and $\ln(s_{j|g}) = \delta_j - I_g$. For the outside alternative, we have that $\ln(s_0) = -\ln(D)$. Combining these expressions we can obtain that $\ln\left(s_j\right) = (\lambda - 1)\ I_g + \ln\left(s_0\right) + \delta_j$. And taking into account that $I_g = [\ln(s_g^*) - \ln(s_0)]/\lambda$, we have that:

$$\delta_j = [\ln(s_j) - \ln(s_0)] + \left(\frac{1-\lambda}{\lambda}\right)\ [\ln(s_g^*) - \ln(s_0)] \tag{2.53}$$

This equation is the inverse mapping σ_j^{-1} for the nested Logit model. ■

We also have a closed-form expression for σ_j^{-1} in the case of the Nested Logit model. However, in general, for the Random Coefficients model we do not have a closed form expression for the inverse mapping σ_j^{-1}. Berry (1994) and Berry, Levinsohn, and Pakes (1995) propose a fixed point algorithm to compute the inverse mapping for the Random Coefficients logit model. They propose the following fixed point mapping: $\delta = F(\delta \mid \mathbf{s}, \mathbf{p}, \mathbf{X}, \Sigma)$ or $\delta_j = F_j(\delta \mid \mathbf{s}, \mathbf{p}, \mathbf{X}, \Sigma)$ where:

$$F_j(\delta \mid \mathbf{s}, \mathbf{p}, \mathbf{X}, \Sigma) \equiv \delta_j + \ln(s_j) - \ln\left(\sigma_j(\delta | \mathbf{p}, \mathbf{X}, \Sigma)\right) \tag{2.54}$$

It is straightforward to see that δ is a fixed point of the mapping $F(\delta \mid \mathbf{s}, \mathbf{p}, \mathbf{X}, \Sigma)$ if and only if $\delta = \sigma^{-1}(\mathbf{s} \mid \mathbf{p}, \mathbf{X}, \Sigma)$. Therefore, finding a solution (fixed point) in δ to the system of equations $\delta = F(\delta \mid \mathbf{s}, \mathbf{p}, \mathbf{X}, \Sigma)$ is equivalent to finding the inverse function $\sigma^{-1}(\mathbf{s} \mid \mathbf{p}, \mathbf{X}, \Sigma)$ at a particular value of $(\mathbf{s}, \mathbf{p}, \mathbf{X}, \Sigma)$.

Defnition: Contraction. Let $\mathscr{X}$ be a set in $\mathbb{R}^n$, let $\|.\|$ be the Euclidean distance, and let $f(x)$ be a function from $\mathscr{X}$ into $\mathscr{X}$. We say that $f(x)$ is a *contraction* (with respect to $\mathscr{X}$ and $\|.\|$) if and only if there is a constant $\lambda \in [0,1)$ such that for any pair of values x and x' in $\mathscr{X}$ we have that $\|f(x) - f(x')\| \leq \lambda \, \|x - x'\|$. ■

Contraction mapping Theorem. If $f : \mathscr{X} \to \mathscr{X}$ is a contraction, then the following results hold. *(A)* there is only one solution in $\mathscr{X}$ to the fixed point problem $x = f(x)$. *(B)* Let x^* be this unique solution such that $x^* = f(x^*)$. For any arbitrary value $x_0 \in \mathscr{X}$ define the sequence $\{x_k : k \geq 1\}$ such that $x_k = f(x_{k-1})$. Then, $\lim_{k\to\infty} x_k = x^*$. ■

Berry (1994) shows that this mapping $F(\delta \mid \mathbf{s}, \mathbf{p}, \mathbf{X}, \Sigma)$ is a contraction as long as the values of δ are not too small. As established by the *Contraction Mapping Theorem*, this implies that the mapping has a unique fixed point and we can find it by using the *fixed point iteration algorithm*. For this model, the algorithm proceeds as follows.

Fixed Point algorithm

- Start with an initial guess δ^0.

- At iteration $R \geq 1$, we calculate $\sigma_j(\delta^{R-1} \mid \mathbf{p}, \mathbf{X}, \Sigma)$ for every product j by evaluating the multiple integration expression in equation (2.47) and then we update the vector δ using the updating equation:

$$\delta^R = F_j(\delta^{R-1} \mid \mathbf{s}, \mathbf{p}, \mathbf{X}, \Sigma) = \delta_j^{R-1} + \ln(s_j) - \ln\left(\sigma_j(\delta^{R-1} | \mathbf{p}, \mathbf{X}, \Sigma)\right) \tag{2.55}$$

- Given δ^R, we check for convergence. If $\left\|\delta^R - \delta^{R-1}\right\|$ is smaller than a pre-specified small constant (for instance, 10^{-6}), we stop the algorithm and take δ^R as the solution to the fixed point of the algorithm. Otherwise, we proceed with iteration $R+1$. ■

2.3.6 Dealing with limitations

Discrete choice demand models can deal with some limitations of demand systems in product space.

[1] Representative consumer assumption. The model is micro founded. It takes into account that the shape of demand and price sensitivity is intimately related to consumer heterogeneity in tastes. Therefore, we can estimate demand systems with precision when J is large. In fact, for these models, large J implies more precise estimates.

[2] Too many parameters problem. The number of parameters does not increase with the number of products J but with the number of observable product attributes K.

[3] Instruments for prices. As we describe below, in the regression equation $\sigma^{-1}(\mathbf{s} \mid \mathbf{p}, \mathbf{X}, \Sigma) = -\alpha \, p_j + \mathbf{X}_j \, \beta + \xi_j$ we can use the observable exogenous characteristics of other products, $\mathbf{X}_k : k \neq j$, as instruments for price. In the equation for product j, the characteristics of other products, $\{\mathbf{X}_k : k \neq j\}$, are valid instruments for the price of product j. To see this, note that the variables $\{\mathbf{X}_k : k \neq j\}$ are not correlated with the error term ξ_j but they are correlated with the price p_j. The later condition may not be obvious because it depends on an assumption about pricing decisions. Suppose that product prices are the result of price competition between the firms that produce these products. To provide a simple intuition, suppose that there is one firm per product and consider the Logit model of demand. The profit function of firm j is $p_j \, q_j - C_j(q_j) - F_j$ where $C_j(q_j)$ and F_j are the variable and the fixed costs of producing j, respectively. For the Logit model, $\partial q_j / \partial p_j = -\alpha q_j (1 - s_j)$ and the marginal condition of optimality for the price of product j is:

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

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 $\mathbf{X}_k \beta$ (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 by IV using as instruments for prices the characteristics of the other products. We provide further details in the next section.

[4] Problems to predict the demand of new products. Predicting the demand of new products does not require knowing additional parameters. Given the structural parameters β, α, and Σ, we can predict the demand of a new hypothetical product which has 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:

$$q_{J+1} = H \int \frac{\exp\{-\alpha \; p_{J+1} + \mathbf{X}_{J+1} \; \beta + \xi_{J+1} + \widetilde{v}_{hJ+1}\}}{1 + \sum_{k=1}^{J+1} \exp\{-\alpha \; p_k + \mathbf{X}_k \; \beta + \xi_k + \widetilde{v}_{hk}\}} \; f(\widetilde{\mathbf{v}}_h | \mathbf{p}, \mathbf{X}, \Sigma) \tag{2.56}$$

Note that to obtain this prediction we need also to use the residuals $\{\xi_k\}$ that can be obtained from the estimation of the model. Given any hypothetical new product with characteristics $(X_{J+1},\ p_{J+1},\ \xi_{J+1})$, the model provides the market share of this new product, its demand elasticity, and the effect of introducing this new product on the market share of any pre-existing product.

2.3.7 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 J is large (for instance, 100 varieties or more). The researcher observes the dataset $\left\{q_j, \mathbf{X}_j, p_j : j = 1, 2, ..., J\right\}$. 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 econometric model is:

$$s_j = \sigma_j(\mathbf{X}, \mathbf{p}, \xi; \theta) \tag{2.57}$$

Unobserved product characteristics ξ are correlated with prices $\mathbf{p}$ (endogeneity). Dealing with endogeneity in nonlinear models where unobservables do not enter additively is complicated. In principle, we would like to avoid using a Maximum Likelihood approach because it requires the specification of how the vector of prices $\mathbf{p}$ depends on the exogenous variables $(\mathbf{X}, \xi)$, and an assumption about the probability distribution of the vector of unobservables ξ. If these assumptions on the supply side of the model are incorrect, the maximum likelihood estimator provides inconsistent estimates of demand parameters. We would prefer using a method that does not require these additional assumptions.

In this context, an important contribution of Berry (1994) and Berry, Levinsohn, and Pakes (1995) was to show that there is a general class of models with the **invertibility property** described above. This property implies that we can represent the model using a equation where the unobservables ξ enter additively:

$$\sigma_j^{-1}(\mathbf{s}\mid\mathbf{p},\mathbf{X},\Sigma) = -\alpha\ p_j + \mathbf{X}_j\ \beta + \xi_j \tag{2.58}$$

Given this representation of the model, we can estimate the structural parameters $\theta = \{\alpha,\ \beta,\ \Sigma\}$ using GMM. The key identification assumption is the mean independence of the unobserved product characteristics and the exogenous product characteristics.

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

Generalized Method of Moments (GMM) estimation. Under the previous assumption, we can use the characteristics of other products $(\mathbf{X}_k : k \neq j)$ to construct moment conditions to estimate structural parameters in equation (2.58). For instance, we can use the average characteristics of other products as the vector of instruments, $\frac{1}{J-1}\sum_{k\neq j}\mathbf{X}_k$. It is clear that $\mathbb{E}\left(\frac{1}{J-1}\sum_{k\neq j}\mathbf{X}_k\ \xi_j\right) = 0$, and we can estimate θ using GMM. Suppose that we have a vector of instruments $\mathbf{Z}_j$ (for instance, $\mathbf{Z}_j = [\mathbf{X}_j, \frac{1}{J-1}\sum_{k\neq j}\mathbf{X}_k]$) such that the following identification conditions hold:

(ID.1) $\mathbb{E}(\mathbf{Z}_j\ \xi_j) = 0$;

(ID.2) $\dim(\mathbf{Z}_j) \geq \dim(\theta)$;

(ID.3) $\mathbb{E}\left[\left(\frac{\partial\sigma_j^{-1}(\mathbf{s}|\mathbf{p},\mathbf{X},\Sigma)}{\partial\Sigma}, p_j, \mathbf{X}_j\right)'\left(\frac{\partial\sigma_j^{-1}(\mathbf{s}|\mathbf{p},\mathbf{X},\Sigma)}{\partial\Sigma}, p_j, \mathbf{X}_j\right) \mid \mathbf{Z}_j\right]$ is non-singular.

Under conditions (ID.1) to (ID.3), the moment restrictions $\mathbb{E}(\mathbf{Z}_j\ \xi_j) = 0$ can identify the vector of parameters θ.

To obtain the GMM estimator of θ, we replace the population moment restrictions $\mathbb{E}(\mathbf{Z}_j\ \xi_j) = 0$ with their sample counterpart. To do this, we replace the population expectation $\mathbb{E}(.)$ with the sample mean $\frac{1}{J}\sum_{j=1}^{J}(.)$, and the unobservable ξ_j with its expression in terms of observables and parameters of the model. Then, the sample moment conditions becomes:

$$\frac{1}{J}\sum_{j=1}^{J}\mathbf{Z}_j\ \left(\sigma_j^{-1}(\mathbf{s}|\mathbf{p},\mathbf{X},\Sigma) + \alpha\ p_j - \mathbf{X}_j\ \beta\right) = 0 \tag{2.59}$$

If the number of these restrictions (that is, the number of instruments in the vector $\mathbf{Z}_j$) is equal to the number of parameters in θ, then the model is just identified and the GMM estimator is defined as the value of θ that solves exactly this system of sample moment conditions. When the number of restrictions is greater than the number of parameters, the model is over-identified, and the GMM estimator is defined as the value of θ that minimizes a quadratic form of the moment restrictions. Let $m(\theta)$ be function that represents in a compact form the sample moments $\frac{1}{J}\sum_{j=1}^{J}\mathbf{Z}_j\ [\sigma_j^{-1}(\mathbf{s}|\mathbf{p},\mathbf{X},\Sigma) + \alpha\ p_j - \mathbf{X}_j\ \beta]$ as a function of θ. The GMM estimator is defined as:

$$\widehat{\theta} = \arg\min_{\theta}\left[m(\theta)'\ \mathbf{W}\ m(\theta)\right] \tag{2.60}$$

where $\mathbf{W}$ is a weighting matrix.

Choice of instruments

When J is large, a possible concern with the instruments $\frac{1}{J-1}\sum_{k\neq j} X_k$ is that they may have very little sample variability across j. To deal with this problem we can define instruments that take into account some intuitive features on price competition between differentiated products. Product j faces stronger competition if there are other products with similar characteristics. Therefore, we expect that the price of product j declines with the number of its *close neighbors*, where these close neighbors are defined as other products with similar characteristics as product j. To implement this idea, define d^* as the average distance between the observable characteristics of all the products in the market. That is, $d^* = \frac{1}{J(J-1)/2}\sum_{j=1}^{J}\sum_{k>j}\left\|\mathbf{X}_k - \mathbf{X}_j\right\|$. Let $\tau \in (0,1)$ be a small constant such that when the distance between two products is smaller than $\tau\, d^*$ we can say that the two products are very similar (for instance, $\tau = 0.10$). We can define a set of *close neighbors* for product j as:

$$\mathscr{N}_j = \left\{k \neq j : \left\|\mathbf{X}_k - \mathbf{X}_j\right\| \leq \tau d^*\right\} \tag{2.61}$$

Let $\left|\mathscr{N}_j\right|$ represents the number of elements in the set $\mathscr{N}_j$. We can construct the vector of instruments,

$$\mathbf{Z}_j = [\mathbf{X}_j,\ \left|\mathscr{N}_j\right|,\ \frac{1}{\left|\mathscr{N}_j\right|}\sum_{k\in\mathscr{N}_j}\mathbf{X}_k] \tag{2.62}$$

This vector of instruments can have more sample variability than $\frac{1}{J-1}\sum_{k\neq j} X_k$ and it can be also more correlated with p_j.

The vector of instruments $\mathbf{Z}_j$ should have at least as many variables as the number of parameters $\theta = \{\alpha,\ \beta,\ \Sigma\}$. Without further restrictions we have that $\dim(\Sigma) = \frac{K(K+1)}{2}$ where $\dim(X_j) = K$, such that $\dim(\theta) = (K+1) + \frac{K(K+1)}{2}$. Note that the vector of instruments suggested above, $\mathbf{Z}_j = [\mathbf{X}_j,\ \left|\mathscr{N}_j\right|,\ \frac{1}{\left|\mathscr{N}_j\right|}\sum_{k\in\mathscr{N}_j}\mathbf{X}_k]$, has only $2K+1$ elements such that the order condition of identification (ID.2) does not hold, that is, $\dim(\mathbf{Z}_j) = 2K+1 < (K+1) + \frac{K(K+1)}{2} = \dim(\theta)$.

Several solutions have been applied to deal with this under-identification problem. A common approach is to impose restrictions on the variance matrix of the random coefficients Σ. The most standard restriction is that Σ is a diagonal matrix (that is, zero correlation between the random coefficients of different product attributes) and there is (at least) one product attribute without random coefficients (i.e. one element in the diagonal of Σ is equal to zero). Under these restrictions, we have that $\dim(\theta) = 2K$ and the order condition of identification holds.

Another approach which has been used in some papers is including additional moments restrictions that come from "micro-moments" or more precisely, market shares for some demographic groups of consumers. This is the approach in Petrin (2002). A third possible approach is to extend the set of instruments beyond $\frac{1}{\left|\mathscr{N}_j\right|}\sum_{k\in\mathscr{N}_j}\mathbf{X}_k$. In this case, one could use the two step method in Newey (1990) to obtain the vector of optimal instruments.

Weak instruments problem

Armstrong (2016) points out a potential inconsistency in this GMM estimator when the number of products is large but the number of markets and firms is small. BLP instruments affect prices only through price-cost margins. If price-cost margins converge fast enough to a constant as $J \to \infty$, then the GMM-BLP estimator is inconsistent because – asymptotically – the BLP instruments do not have any power to explain prices. This is an extreme case of weak instruments. This is also a potential issue in small samples: the bias and variance of the estimator can be very large in small samples. Armstrong (2016) studies this issue under different data structures.

Suppose that the dataset consists of J products (indexed by j) which belong to N firms (indexed by n) over T markets or time periods (indexed by t), such that we observe prices and quantities p_{jnt} and q_{jnt}. Armstrong studies the predictive power of BLP instruments and the potential inconsistency of the GMM-BLP estimator under different scenarios according to: (a) the form of the demand system, a more specifically whether it is a standard Logit or a random coefficients Logit; and (b) the structure of the panel data, or more precisely, whether the number of products per firm J/N converges to a constant, to zero, or to infinity when J goes to infinity. In the analysis here, we consider that the number of markets T is fixed.

First, consider the case of the standard logit model with single product firms such that $J/N = 1$. Under Bertrand competition, the price equation has the following form (see section 4.3 in chapter 4):

$$p_j = MC_j + \frac{1}{\alpha}\alpha(1 - s_j) \tag{2.63}$$

BLP instruments affect price p_j only through the term $1/(1 - s_j)$. If $\sqrt{J}/(1 - s_j)$ converges to a constant as $J \to \infty$, then the GMM-BLP estimator is inconsistent: it is asymptotically equivalent to using instrumental variables that are independent of prices and do not have any identification power. This is an extreme case of a problem of weak instruments.

In contrast, when firms are multiproduct the GMM-BLP estimator can be consistent as $J \to \infty$. A multiproduct firm maximizes the joint profit from all its products and obtains price-cost margins which are above those when the products are sold by single-product firms. With this industry data, price-cost margins are larger and converge to a constant more slowly than $\sqrt{J}$. More precisely, as $J \to \infty$, keeping constant the number of firms N, there is a constant $\alpha < 1/2$ such that $J^{\alpha}/(1 - s_j)$ converges to a constant. This implies that Logit or random coefficients Logit are consistent as $J \to \infty$.

Armstrong (2016) extends this analysis to the Random Coefficients Logit model. In this model, the GMM-BLP estimator is also inconsistent when the number of markets T and the number of products per firm J/N are fixed. The estimator is consistent when T and N are fixed and firms are asymmetric in their characteristics. Consistency can be also achieved if T goes to infinity and N and J are fixed.

Alternatives to BLP instruments

An alternative to BLP instruments are Hausman-Nevo instruments and Arellano-Bond or Dynamic Panel Data instruments.

Hausman-Nevo instruments. The dataset includes T geographic markets and J products. The T markets belong to R regions where cost shocks are spatially correlated within region, but demand shocks are not. Suppose that the unobservable ξ_{jt} has the following fixed-effects structure:

$$\xi_{jt} = \xi_j^{(1)} + \xi_t^{(2)} + \xi_{jt}^{(3)} \tag{2.64}$$

and $\xi_{jt}^{(3)}$ is not spatially correlated, that is, for any pair of markets t and t', $\mathbb{E}(\xi_{jt}^{(3)}\ \xi_{jt'}^{(3)}) = 0$. Under these conditions, we can control for $\xi_j^{(1)}$, and $\xi_t^{(2)}$ using product and market fixed effects, respectively. Furthermore, it is is possible to use prices in other markets to construct valid instrumental variables: that is, variables correlated with price but uncorrelated with the unobservable $\xi_{jt}^{(3)}$. More precisely, let Z_{jt} be the average price in markets in region R (where market t belongs) excluding market t:

$$Z_{jt} = \frac{1}{T_R - 1} \sum_{t' \in \mathcal{T}_R,\, t' \neq t} p_{jt'} \tag{2.65}$$

where T_R is the number of markets in region R, and $\mathcal{T}_R$ is the set of these markets. Since $\xi_{jt}^{(3)}$ is not spatially correlated, we have that $\mathbb{E}(Z_{jt}\ \xi_{jt}^{(3)}) = 0$, and Z_{jt} is correlated with p_{jt} because cost shocks are spatially correlated within the region.

Arellano-Bond or Dynamic Panel Data instruments. Now, consider that the subindex t represents time such that the dataset consists of J products over T periods of time, where T is small and J is large. The demand error term ξ_{jt} has the structure in equation (2.64) and $\xi_{jt}^{(3)}$ is not serially correlated: for any two time periods t and t', $\mathbb{E}(\xi_{jt}^{(3)}\ \xi_{jt'}^{(3)}) = 0$. Consider the demand equation in first differences, that is, the equation at period t minus the equation at $t-1$:

$$\sigma_j^{-1}(\mathbf{s}_t\ |\mathbf{p}_t, \mathbf{X}_t, \Sigma) - \sigma_j^{-1}(\mathbf{s}_{t-1}\ |\mathbf{p}_{t-1}, \mathbf{X}_{t-1}, \Sigma) = -\alpha\ \Delta p_{jt} + \Delta \mathbf{X}_{jt}\ \beta + \Delta \xi_{jt} \tag{2.66}$$

where $\Delta p_{jt} \equiv p_{jt} - p_{jt-1}$, $\Delta \mathbf{X}_{jt} \equiv \mathbf{X}_{jt} - \mathbf{X}_{jt-1}$, and $\Delta \xi_{jt} \equiv \xi_{jt} - \xi_{jt-1}$. Consider the instruments $\mathbf{Z}_{jt} = \{s_{jt-2},\ p_{jt-2}\}$. Under these assumptions, $\mathbf{Z}_{jt}$ are valid instrumental variables, $\mathbb{E}(\mathbf{Z}_{jt}\ \Delta\xi_{jt}^{(3)}) = 0$. If shocks in marginal costs are serially correlated, then $\mathbf{Z}_{jt}$ is correlated with the price difference Δp_{jt} after controlling for the exogenous regressors $\Delta \mathbf{X}_{jt}$.

Computation of the GMM estimator

Berry, Levinsohn, and Pakes (1995) propose a Nested Fixed Point (NFXP) algorithm to compute the GMM estimator of θ.[6] As indicated by its name, this method can be described in terms of two nested fixed point algorithms: an inner algorithm that consists of fixed point iterations to calculate the values $\sigma_j^{-1}(\mathbf{s}\ |\mathbf{p}, \mathbf{X}, \Sigma)$ for a given value of Σ; and an outer Newton algorithm that minimizes the GMM criterion function with respect to θ.

Let $Q(\theta) = m(\theta)'\ \mathbf{W}\ m(\theta)$ be the GMM criterion function, where $\mathbf{W}$ is a weighting matrix that can give different weights to the moments in the vector $m(\theta)$.[7] The GMM

[6] The term Nested Fixed Point (NFXP) algorithm was coined by Rust (1987) in the context of the estimation dynamic discrete choice structural models. In chapter 7, we describe the algorithm in the context of dynamic discrete choice models.

[7] The GMM estimator is consistent and asymptotically normal for any weighting matrix $\mathbf{W}$ that is semi-positive definite. For instance, the identity matrix is a possible choice. However, there is an optimal weighting matrix that minimizes the variance of the GMM estimator.

estimator is the value $\widehat{\theta}$ that maximizes $Q(\theta)$. The first order conditions of optimality are $\partial Q(\widehat{\theta})/\partial\theta = 0$. Newton's method is based on a first order Taylor's approximation to the condition $\partial Q(\widehat{\theta})/\partial\theta = 0$ around some value θ^0 such that, by the Mean Value Theorem, there exists a scalar $\lambda \in [0,1)$ such that for $\theta^* = (1-\lambda)\theta^0 + \lambda\widehat{\theta}$ we have that:

$$\frac{\partial Q(\widehat{\theta})}{\partial\theta} = \frac{\partial Q(\theta^*)}{\partial\theta} + \frac{\partial^2 Q(\theta^*)}{\partial\theta\partial\theta'}[\widehat{\theta} - \theta^0] \tag{2.67}$$

Therefore, we have that $\partial Q(\theta^*)/\partial\theta + \partial^2 Q(\theta^*)/\partial\theta\partial\theta' \, [\widehat{\theta} - \theta^0] = 0$, and solving for $\widehat{\theta}$, we get:

$$\widehat{\theta} = \theta^0 - \left[\frac{\partial^2 Q(\theta^*)}{\partial\theta\partial\theta'}\right]^{-1} \left[\frac{\partial Q(\theta^*)}{\partial\theta}\right] \tag{2.68}$$

If we knew the value θ^*, then we could obtain the estimator $\widehat{\theta}$ using this expression. However, note that $\theta^* = (1-\lambda)\theta^0 + \lambda\widehat{\theta}$ such that it depends on $\widehat{\theta}$ itself. We have a "chicken and egg" problem. To deal with this problem, Newton's method proposes an iterative procedure.

Newton's algorithm. We start with an initial candidate for estimator, θ^0. At every iteration $R \geq 1$, we update the value of θ, from θ^{R-1} to θ^R, using the following formula:

$$\theta^R = \theta^{R-1} - \left[\frac{\partial^2 Q(\theta^{R-1})}{\partial\theta\partial\theta'}\right]^{-1} \left[\frac{\partial Q(\theta^{R-1})}{\partial\theta}\right] \tag{2.69}$$

iven θ^R and θ^{R-1}, we check for convergence. If $\left\|\theta^R - \theta^{R-1}\right\|$ is smaller than a pre-specified small constant (for instance, 10^{-6}), we stop the algorithm and take θ^R as the estimator $\widehat{\theta}$. Otherwise, we proceed with iteration $R+1$. ■

The Nested Fixed Point algorithm makes it explicit that the evaluation of the criterion function $Q(\theta)$ at any value of θ, and of its first and second derivatives, requires the solution of another fixed point problem to evaluate the inverse mapping $\sigma_j^{-1}(\mathbf{s}|\,\mathbf{p},\mathbf{X},\Sigma)$. Therefore, there are two fixed point algorithms which are nested: the outer algorithm minimizes function $Q(\theta)$ using Newton's method; and the inner algorithm that obtains a fixed point for δ.

Nested Fixed Point algorithm. We start with an initial guess $\theta^0 = (\alpha^0,\beta^0,\Sigma^0)$. At every iteration $R \geq 1$ of the outer (Newton) algorithm, we take Σ^{R-1} and apply the Fixed Point described above (equation (2.55)) to compute the inverse mapping $\sigma_j^{-1}(\mathbf{s}|\,\mathbf{p},\mathbf{X},\Sigma^{R-1})$. We also apply the same Fixed Point algorithm to calculate numerically the gradient vector $\partial\sigma_j^{-1}(\mathbf{s}|\,\mathbf{p},\mathbf{X},\Sigma^{R-1})/\partial\Sigma$ and the Hessian matrix $\partial^2\sigma_j^{-1}(\mathbf{s}|\,\mathbf{p},\mathbf{X},\Sigma^{R-1})/\partial\Sigma\partial\Sigma'$. Given these objects, we can obtain the gradient vector $\frac{\partial Q(\theta^{R-1})}{\partial\theta}$ and the Hessian matrix $\frac{\partial^2 Q(\theta^{R-1})}{\partial\theta\partial\theta'}$. Then, we apply one Newton iteration as described in equation (2.69).

Given θ^R and θ^{R-1}, we check for convergence. If $\left\|\theta^R - \theta^{R-1}\right\|$ is smaller than a pre-specified small constant (for instance, 10^{-6}), we stop the algorithm and take θ^R as the estimator $\widehat{\theta}$. Otherwise, we proceed with iteration $R+1$. ■

The Nested Fixed Point algorithm may be computationally intensive because it requires the repeated solution of the fixed point problem that calculates the inverse mapping $\sigma_j^{-1}(\mathbf{s}|\ \mathbf{p}, \mathbf{X}, \Sigma)$, which itself requires Monte Carlo simulation methods to approximate the multiple dimension integrals that define the market shares. Some alternative algorithms have been proposed to reduce the number of times that the inner algorithm is called to computed the inverse mapping. Dubé, Fox, and Su (2012) propose the MPEC algorithm. Lee and Seo (2015) propose a Nested Pseudo Likelihood method in the same spirit as Aguirregabiria and Mira (2002).

2.3.8 Nonparametric identification

Empirical applications of discrete choice models of demand make different parametric assumptions such as the normal distribution of the random coefficients and the additive separability of observable and unobservable product characteristics in the utility function. Berry and Haile (2014) show that these parametric restrictions are not essential for the identification of this type of demand system.

Let v_{hj} be the utility of consumer h for purchasing product j. Define $v_h = (v_{h1}, ..., v_{hJ})$ that has CDF $F_v(v_h|\mathbf{X}, \mathbf{p}, \xi)$, where $(\mathbf{X}, \mathbf{p}, \xi)$ are the vectors of characteristics of all the products. In this general model, the interest is in the nonparametric identification of the distribution function $F_v(v_h|\mathbf{X}, \mathbf{p}, \xi)$. The following assumption plays a key role in the identification results by Berry and Haile (2014).

Assumption BH-1. Unobserved product characteristics, ξ_j, enter in the distribution of consumers' preferences v_h through the term $X_{1j} + \xi_j$, where X_{1j} is one of the observable product attributes.

$$F_v(v_h \mid \mathbf{X}, \mathbf{p}, \xi) = F_v(v_h \mid \mathbf{X}_{(-1)}, \mathbf{p},\ \mathbf{X}_1 + \xi) \tag{2.70}$$

where $\mathbf{X}_{(-1)}$ represents the observable product characteristics other than X_1, and $\mathbf{X}_1 + \xi$ represents the vector $(X_{11} + \xi_1,\ ...,\ X_{1J} + \xi_J)$.

Assumption BH-1 implies that the marginal rate of substitution between the observable characteristic X_{1j} and the unobservable ξ_j is constant. The restriction that it is equal to one is without lost of generality. Under this assumption, it is clear that:

$$s_j = \Pr\left(j = \arg\max_k v_{ik} \mid \mathbf{X}_{(-1)}, \mathbf{p},\ \mathbf{X}_1 + \xi \right) = \sigma_j\left(\mathbf{X}_{(-1)}, \mathbf{p},\ \mathbf{X}_1 + \xi\right) \tag{2.71}$$

For notational convenience, we use ξ_j^* to represent $X_{j1} + \xi_j$ and the vector ξ^* to represent $\mathbf{X}_1 + \xi$ such that we can write the market share function as $\sigma_j\left(\mathbf{X}_{(-1)}, \mathbf{p},\ \xi^*\right)$.

Assumption BH-2. The mapping $\mathbf{s} = \sigma_j\left(\mathbf{X}_{(-1)}, \mathbf{p},\ \xi^*\right)$ is invertible in ξ^* such that we have, $\xi_j^* = X_{j1} + \xi_j = \sigma_j^{-1}\left(\mathbf{s} \mid \mathbf{X}_{(-1)}, \mathbf{p}\right)$.

What are the economic conditions that imply this inversion property? **Connected substitutes.** The assumption of **Connected substitutes** can be described in terms of two conditions.

Condition (i). All goods are weak gross substitutes, that is, for any $k \neq j$, $\sigma_j\left(\mathbf{X}_{(-1)}, \mathbf{p},\ \xi^*\right)$ is weakly decreasing in ξ_k^*. A sufficient condition is that, as in the parametric model, higher values of ξ_j^* raise the utility of good j without affecting the utilities of other goods.

Condition (ii). "Connected strict substitution". Starting from any inside good, there is a chain of substitution [that is, σ_j is strictly decreasing in ξ_k^*] leading to the outside good.

Connected strict substitution requires only that there is not a subset of products that substitute only among themselves, that is, all the goods must belong in one demand system.

Suppose that we have data from T markets, indexed by t. We can write the inverse demand system as:

$$X_{jt}^{(1)} = \sigma_j^{-1}\left(\mathbf{s}_t \mid \mathbf{X}_{(-1)t}, \mathbf{p}_t\right) - \xi_{jt} \tag{2.72}$$

Let $\mathbf{Z}_t$ be a vector of instruments [we explain below how to obtain these instruments] and suppose that: (a) $\mathbb{E}[\xi_{jt} \mid \mathbf{Z}_t] = 0$; (b) [completeness] if $\mathbb{E}[B(\mathbf{s}_t, \mathbf{X}_{(-1)t}, \mathbf{p}_t) \mid \mathbf{Z}_t] = 0$, then $B(\mathbf{s}_t, \mathbf{X}_{(-1)t}, \mathbf{p}_t) = 0$ *almost surely*. **Important:** Completeness requires that $\dim(\mathbf{Z}_t) \geq \dim(\mathbf{s}_t, \mathbf{X}_{(-1)t}, \mathbf{p}_t)$, that is, instruments for all $2J$ endogenous variables $(\mathbf{s}_t, \mathbf{p}_t)$

Under these conditions, all the inverse functions σ_j^{-1} are nonparametrically identified. Then, ξ_{jt} is identified and we can again invert σ_j^{-1} to identify the demand system σ.

Sources of instruments. Note that we need not only J instruments for prices but also J instruments for for market shares $\mathbf{s}_t$. Instruments for $\mathbf{s}_t$ must affect quantities not only through prices. For instance, supply/marginal cost shifters or Hausman-Nevo are IVs for prices but they are not useful for $\mathbf{s}_t$ because they affect quantities only through prices. The vector $\mathbf{X}_{1t}$ is a natural candidate as IV for $\mathbf{s}_t$. By the implicit function theorem, $\dfrac{\partial \sigma^{-1}(\mathbf{s}_t, \mathbf{p}_t)}{\partial \mathbf{s}_t'} = \left[\dfrac{\partial \sigma(\delta_t, \mathbf{p}_t)}{\partial \delta_t'}\right]$. Identifying the effects of $\mathbf{s}_t$ on σ^{-1} is equivalent to identifying the effects of ξ^* on market shares σ. The vector $\mathbf{X}_{1t}$ directly shifts the indices ξ^* , so these are natural instruments for market shares in the inverse demand function.

Identification of Utility [Welfare analysis]. Without further restrictions, identification of the system of demand equations σ does not imply identification of the distribution of random utilities F_v. In general, to identify changes in consumer welfare we need F_v.

Assumption BH-3 [Quasi-linear preferences]. $v_{hj} = \mu_{hj} - p_j$ where the variables μ_{hj} are independent of p_j conditional on $(\xi^*, \mathbf{X}_{(-1)})$.

Under Assumption 3, the distribution F_v is identified from the demand system σ.

2.4 Valuation of product innovations

Product innovation is ubiquitous in most industries, and a key strategy for differentiation. During the last decades we have witnessed a large increase in the number of varieties of different products. Evaluating consumer value of new products, and of quality improvements in existing products, has received substantial attention in the context of: improving Cost of Living Indexes (COLI); costs and benefits of firms' product differentiation; and social value of innovations.

The common approach consists of the following steps: estimating a demand system of differentiated products; using the estimated system to obtain consumers' indirect utility functions (or surplus functions); and comparing consumers' utilities with and without the new product. Typically, one of the two scenarios (with or without the new product) is a counterfactual.

2.4.1 Hausman on cereals

Hausman (1996) presents an application of demand in product space to an industry with many varieties: ready-to eat (RTE) cereals in US. This industry has been characterized by the proliferation of many varieties. We have described the Hausman's data, model, and estimation method in section 2.5 above. Here we describe Hausman's evaluation of the welfare effects of the introduction of a new brand, Cinnamon Cheerios.

Hausman uses the estimated demand system to evaluate the value of a new variety that was introduced during this period: apple-cinamon cheerios (*ACC*). He first obtains the value of the price *ACC* that makes the demand of this product equal to zero. He obtains a virtual price of \$7.14 (double the actual observed price \$3.5). Given this price, he calculates the consumer surplus (alternatively the CV or the EV).

He obtains estimated welfare gains of \$32,268 per city and weekly average with a standard error of \$3,384. Aggregated at the level of US and annually, the consumer-welfare gain is \$78.1 million (or \$0.31 per person per year) which is a sizable amount of consumer's surplus.

Valuation of new products. Consider an individual with preference parameters $(\alpha_h, \beta_h, \varepsilon_h)$ facing a set of products $\mathscr{J}$ with vector of prices $\mathbf{p}$. The indirect utility function is defined as (income effects are assumed away because linearity):

$$v(\mathbf{p}, \alpha_h, \beta_h, \varepsilon_h) = \max_{j \in \mathscr{J}} \left[-\alpha_h p_j + \mathbf{x}_j \beta_h + \xi_j + \varepsilon_{hj} \right] \tag{2.73}$$

To measure aggregate consumer welfare, Hausman uses the money-metric welfare function in McFadden (1974) and Small and Rosen (1981). As indicated by its name, an attractive feature of this welfare measure is that its units are monetary units. To obtain this money metric, we divide utility by the marginal utility of income. The money-metric welfare for consumer h is defined as $\frac{1}{\alpha_h}\, v(\mathbf{p}, \alpha_h, \beta_h, \varepsilon_h)$. The money metric at the aggregate market level is $W(\mathbf{p}) = \int \frac{1}{\alpha_h}\, v(\mathbf{p}, \alpha_h, \beta_h, \varepsilon_h)\; dF(\alpha_h, \beta_h, \varepsilon_h)$. For the random coefficients logit model:

$$W(\mathbf{p}) = \int \frac{1}{\alpha_h} \ln \left[\sum_{j=0}^{J} \exp\left\{ -\alpha_h p_j + \mathbf{x}_j \beta_h + \xi_j \right\} \right] \, dF(\alpha_h, \beta_h) \tag{2.74}$$

We can include $\mathbf{x}$ and $\mathscr{J}$ as explicit arguments of the welfare function: $W(\mathbf{p}, \mathbf{x}, \mathscr{J})$. We can use $W$to measure the welfare effects of a change in: Prices, $W(\mathbf{p}^1, \mathbf{x}, \mathscr{J}) - W(\mathbf{p}^0, \mathbf{x}, \mathscr{J})$; Products characteristics: $W(\mathbf{p}, \mathbf{x}^1, \mathscr{J}) - W(\mathbf{p}, \mathbf{x}^0, \mathscr{J})$; Set of products: $W(\mathbf{p}, \mathbf{x}, \mathscr{J}^1) - W(\mathbf{p}, \mathbf{x}, \mathscr{J}^0)$.

Some limitations.

[1] Problems to evaluate **radical innovations**. That is, innovations that introduce a product with new characteristics that previous products did not have in any amount.

[2] With **logit errors**, there is very limited "crowding" of products such that the welfare function goes infinity when $J \to \infty$ even if all J have exactly the same attributes X and ξ. See section 2.4.4 below on the Logit model and the value of new products.

[3] **Outside alternative**. Unobserved "qualities" ξ_{jt} are relative to the outside alternative. For instance, if there exist quality improvements in the outside alternative, then this approach underestimates the welfare improvements in this industry.

2.4.2 Trajtenberg (1989)

Trajtenberg (1989) on computed tomography scanners. The computed tomography (CT) scanner is considered a key innovation in imaging diagnosis in medicine during the 1970s. The first was installed in the US in 1973, and soon after 20 firms entered in this market with different varieties, General Electric being the leader. Clients are hospitals. Three characteristics are key to scanner quality: scan time, image quality, and reconstruction time.

Data. The data contains 55 products and covers the period 1973 to 1981. Observable product characteristics are price, scan speed, resolution, and reconstruction speed. The quantity variable is sales in the US market. The data includes also information on the identity and attributes of the buying hospital. It is hospital-year level data and the dependent variable is the product choice of hospital h at year t.

Model. The model is a nested logit where scanners are divided in two groups depending on the part of the body for which the scanner is designed to scan. Then, the groups are "head scanners" and "body scanners". The utility function is quadratic in the three product attributes (other than price).

Estimation results. The estimation method does not account for potential endogeneity of prices. The estimated elasticity of substitution between the two groups is very close to zero. That is, it seems that head scanners and body scanners are very different products and there is almost zero substitution between these two groups. The estimated coefficient for price – parameter α – is significantly positive (correct sign) for head scanners but it is negative (wrong sign) for body scanners. The wrong sign of this parameter estimate is likely an implication of the endogeneity of price: that is, the positive correlation between price and unobserved product quality.

Welfare effects. The counterfactual experiment consists of eliminating all CT scanner products, keeping only the outside product. The estimated welfare effect of CT scanners during this period is \$16 million of 1982. Using data on firms' R&D investment, Trajtenberg obtains a social rate of return of 270%. That is, every dollar of investment in the R&D of CT scanners generate 2.7 dollars in return. This is a very substantial rate of return.

2.4.3 Petrin (2002) on minivans

The aim of this study was to evaluate the consumer welfare gains from the introduction of a new type of car, the minivan. Estimation of a BLP demand system of automobiles. Combining market level and micro moments. Observing average family size conditional on the purchase of a minivan and asking the model to match this average helps to identify parameters that capture consumer taste for the characteristics of minivans.

In 1984, Chrysler introduced its own minivan: the Dodge caravan. It was an immediate success. GM and Ford responded by quickly introducing their own minivans in 1985. By 1998, there were 6 firms selling a total of 13 different minivans, with Chrysler being the leader with a market share of 44% within the minivan market segment.

Data. It is a product-year panel dataset during the period 1981-1993 and with $J =$ 2407 products. Variables in the dataset include quantity sold in the US market, price, acceleration, dimensions, drive type, fuel efficiency, and indicators for luxury car, SUV, minivans, and full-size vans.

The dataset also includes consumer level information from the Consumer expenditure survey (CEX). The CEX links demographics of purchasers of new vehicles to the type of vehicles they purchase. In the CEX, we observe 2,660 new vehicle purchases over the period and sample. This micro-level data is used to estimate the probabilities of new vehicle purchases for different income groups. Observed purchases of minivans (120), station wagons (63), SUVs (131), and full-size vans (23). Used to estimate average family size and age of purchasers of each of these vehicle types.

FAMILY VEHICLE SALES AS A PERCENTAGE OF TOTAL VEHICLE SALES: U.S. AUTOMOBILE MARKET, 1981–93

Year	Minivans (1)	Station Wagons (2)	Sport-Utilities (3)	Full-Size Vans (4)	Minivans and Station Wagons (5)	U.S. Auto Sales (Millions) (6)
1981	.00	10.51	.58	.82	10.51	7.58
1982	.00	10.27	.79	1.17	10.27	7.05
1983	.00	10.32	3.51	1.04	10.32	8.48
1984	1.58	8.90	5.51	1.20	10.48	10.66
1985	2.32	7.33	6.11	1.05	9.65	11.87
1986	3.63	6.70	5.73	.85	10.43	12.21
1987	4.86	6.47	6.44	.73	11.33	11.21
1988	5.97	5.14	7.18	.69	11.11	11.76
1989	6.45	4.13	7.47	.61	10.58	11.06
1990	7.95	3.59	7.78	.27	11.54	10.51
1991	8.29	3.05	7.80	.29	11.34	9.75
1992	8.77	3.07	9.33	.39	11.84	10.12
1993	9.93	3.02	11.66	.29	12.95	10.71

Figure 2.1: Petrin (2002) Market shares by type of automobile

Estimates. Tables 2.2 to 2.4 present estimates of demand parameters separated in three groups: price coefficients, marginal utilities of product characteristics, and random coefficients.

Price effects of the introduction of minivans. Petrin used the estimated model to implement the counterfactual experiment of eliminating minivan cars from consumers' choice set. This experiment takes into account that in the counterfactual scenario without minivans the equilibrium prices of all the products will change. Table 2.6 presents equilibrium prices with and without minivans. The introduction of minivans (particularly, Dodge caravan) had an important negative effect on the prices of many substitutes that were top-selling vehicles in the large-sedan and wagon segments of the market. There were also some price increases due to cannibalization of own products.

Welfare effects. The preferred estimates are those of the model with random coefficients, using BLP instruments, and using micro moments. Based on these estimates, the mean

TABLE 4
PARAMETER ESTIMATES FOR THE DEMAND-SIDE EQUATION

Variable	OLS Logit (1)	Instrumental Variable Logit (2)	Random Coefficients (3)	Random Coefficients and Microdata (4)
	A. Price Coefficients (α's)			
α_1	.07	.13	4.92	7.52
	(.01)**	(.01)**	(9.78)	(1.24)**
α_2			11.89	31.13
			(21.41)	(4.07)**
α_3			37.92	34.49
			(18.64)**	(2.56)**

Figure 2.2: Petrin (2002) Parameter estimates. Price coefficients

per capita Compensated Variation of introducing minivans is \$1247. This is a very substantial welfare gain. Petrin compares this estimated welfare gain with the ones using other models and estimates of the model: OLS logit; IV logit; and IV BLP without micro moments. These other models and methods imply estimated welfare gains that are substantially smaller than the preferred model. This is mainly because these methods under-estimate the marginal utility of income parameters.

Petrin also provides a decomposition of the welfare gains in the contribution of product characteristics x_j and ξ_j, and of the logit errors ε_{hj}. For the preferred model, the mean per capita welfare gain of \$1247 is decomposed into a contribution of \$851 from product characteristics, and a contribution of \$396 from the logit errors. The other models and methods imply very implausible contributions from the logit errors.

2.4.4 Logit and new products

The Logit errors can have unrealistic implications on the evaluation of welfare gains. Because of these errors, welfare increases unboundedly (though concavely) with J. To illustrate this, consider the simpler case where all the products are identical except for the logit errors. In this case, the aggregate welfare function is $W = \ln(\sum_{j=0}^{J} \exp\{\delta\}) = \delta \ln(J+1)$, which is an increasing and concave function of the number of products J. Though the BLP or Random Coefficients-Logit model limits the influence of the logit errors, this model is still subject to this problem.

Ackerberg and Rysman (2005) propose a simple modification of the logit model that can contribute to correct for this problem. Consider a variation of the BLP model where the dispersion of the logit errors depends on the number of products in the market. For $j > 0$, $U_{hj} = -\alpha_h p_j + \mathbf{x}_j \beta_h + \xi_j + \sigma(J)\ \varepsilon_{hj}$. The parameter $\sigma(J)$ is strictly decreasing in J and it goes to 0 as J goes to ∞. As J increases, the differentiation from the $\varepsilon' s$

Random Coefficient Parameter Estimates

Variable	Random Coefficients (γ's) Uses No Microdata (1)	Uses CEX Microdata (2)
Constant	1.46	3.23
	(.87)*	(.72)**
Horsepower/weight	.10	4.43
	(14.15)	(1.60)**
Size	.14	.46
	(8.60)	(1.07)
Air conditioning standard	.95	.01
	(.55)*	(.78)
Miles/dollar	.04	2.58
	(1.22)	(.14)**
Front wheel drive	1.61	4.42
	(.78)**	(.79)**
γ_{mi}	.97	.57
	(2.62)	(.10)**
γ_{sw}	3.43	.28
	(5.39)	(.09)**
γ_{su}	.59	.31
	(2.84)	(.09)**
γ_{pv}	4.24	.42
	(32.23)	(.21)**

Figure 2.3: Petrin (2002) Parameter estimates. Random coefficients

becomes less and less important. Function $\sigma(J)$ can be parameterized and its parameters can be estimated together with the rest of the model. Though Ackerberg and Rysman consider this approach, they favor a similar approach that is simpler to implement. They consider the model:

$$U_{hj} = -\alpha_h p_j + \mathbf{x}_j \beta_h + \xi_j + f(J, \gamma) + \varepsilon_{hj} \tag{2.75}$$

where $f(J,\gamma)$ is a decreasing function of J parameterized by γ. For instance, $f(J,\gamma) = \gamma \ln(J)$. It can be also extended to a nested logit version. For group g: $f_g(J,\gamma) = \gamma_g \ln(J_g)$. The reason for the specification $f(J,\gamma)$ instead of $\sigma(J,\gamma)$ is simplicity in estimation.

2.4.5 Product complementarity

The class of discrete choice demand models we have considered so far rules out complementarity between products. This is an important limitation in some relevant contexts. For instance, this is a significant limitation in the evaluation of the merger between two firms producing complements, such as Pepsico and Frito-Lay, or in the evaluation of the welfare effects of new products that may complement with existing products. This is also a significant limitation when considering industries with both substitution and complementarity effects. Examples include radio stations that play recorded music, movies based on a book novel and the book itself, or, arguably, Uber and taxis. Gentzkow (2007) extends the McFadenn / BLP framework to allow for complementarity, and studies the demand and welfare effect of online newspapers.

Model

Consumers can choose bundles of products. We start with a simple example. There are two products A and B. The set of possible choices for a consumer is $\{0, A, B, AB\}$. The

PARAMETER ESTIMATES FOR THE COST SIDE
Dependent Variable: Estimated (Log of) Marginal Cost

Variable (τ's)	Parameter Estimate	Standard Error
Constant	1.50	.08
ln(horse power/weight)	.84	.03
ln(weight)	1.28	.04
ln(MPG)	.23	.04
Air conditioning standard	.24	.01
Front wheel drive	.01	.01
Trend	−.01	.01
Japan	.12	.01
Japan × trend	−.01	.01
Europe	.47	.03
Europe × trend	−.01	.01
ln(q)	−.05	.01

Figure 2.4: Petrin (2002) Parameter estimates. Marginal Cost

utilities of these choice alternatives are 0, u_A, u_B, and $u_{AB} = u_A + u_B + \Gamma$, respectively. The parameter Γ measures the degree of demand complementarity between products A and B. The indirect utilities u_A and u_A have the following form: $u_A = \beta_A - \alpha p_A$ and $u_B = \beta_B - \alpha p_B$, where p_A and p_B are prices. Let $\mathbb{P}_j = \Pr(u_j = \max\{0, u_A, u_B, u_{AB}\})$ be the probability or proportion of consumers that choose alternative j. Importantly, in contrast to the discrete choice demand models considered above, in this model $\mathbb{P}_A$ and $\mathbb{P}_B$ are not the market shares of products A and B, respectively. To obtain the market shares of these products we should take into account the share of consumers buying bundle AB. Let s_A and s_B be the market shares of products A and B, respectively. We have that:

$$\begin{aligned} s_A &= \mathbb{P}_A + \mathbb{P}_{AB} \\ s_B &= \mathbb{P}_B + \mathbb{P}_{AB} \end{aligned} \tag{2.76}$$

The substitutability or complementarity between products A and B depends on the cross-price derivatives in the demand of these products. Products A and B are substitutes if $\partial s_A / \partial p_B > 0$, and they are complements if $\partial s_A / \partial p_B < 0$. Complements / substitutes is closely related to the sign of Γ. Figure 2.7 illustrates this relationship.

We represent u_A on the horizontal axis and u_B on the the vertical axis. We consider three cases: case 1 with $\Gamma = 0$; case 2 with $\Gamma > 0$; and case 3 with $\Gamma < 0$. For each case, we partition the space in four regions where each region represents the values (u_A, u_B) for which an alternative is the optimal choice. Consider the effect of a small increase in the price of product B. Because it is a marginal increase, it affects only those consumers who are in the frontier of the choice sets. More precisely, the increase in p_B implies that all the frontiers shift vertically and upward. That is, to keep the same choice as with previous prices the utility u_B should be larger.

EQUILIBRIUM PRICES WITH AND WITHOUT THE MINIVAN, 1984: 1982–84 CPI-ADJUSTED DOLLARS

	Price			%
	With Minivan	Without Minivan	ΔPrice	ΔPrice
	A. Largest Price Decreases on Entry			
GM Oldsmobile Toronado (large sedan)	15,502	15,643	−141	.90
GM Buick Riviera (large sedan)	15,379	15,519	−139	.89
GM Buick Electra (large sedan)	12,843	12,978	−135	1.04
GM Chevrolet Celebrity (station wagon)	8,304	8,431	−127	1.51
Ford Cadillac Eldorado (large sedan)	19,578	19,704	−126	.64
Ford Cadillac Seville (large sedan)	21,625	21,749	−125	.57
GM Pontiac 6000 (station wagon)	9,273	9,397	−123	1.31
GM Oldsmobile Ciera (station wagon)	9,591	9,714	−123	1.27
GM Buick Century (station wagon)	8,935	9,056	−121	1.34
GM Oldsmobile Firenza (station wagon)	7,595	7,699	−104	1.35
	B. Largest Price Increases on Entry			
Chrysler LeBaron (station wagon)	9,869	9,572	297	3.10
Volkswagen Quattro (station wagon)	13,263	13,079	184	1.41
Chrysler (Dodge) Aries K (station wagon)	7,829	7,659	170	2.22
AMC Eagle (station wagon)	10,178	10,069	109	1.08

Figure 2.5: Petrin (2002) Prices with and without minivans

In case 1 with $\Gamma = 0$, this implies that P_{AB} declines and P_A increases but they do it by the same absolute magnitude such that $s_A = P_A + P_{AB}$ does not change. Therefore, with $\Gamma = 0$, we have that $\partial s_A / \partial p_B = 0$ and products A and B are neither substitutes nor complements.

In case 2 with $\Gamma > 0$, we can distinguish two different types of marginal consumers: those located in a point like m and those in a point like o in Figure 2.7 panel 2. For consumers in point m, an increase in p_B makes them switch from choosing the bundle AB to choosing A. As in case 1, this change does not affect the demand of product A. However, we have now also the consumers in point o. These consumers switch from buying the bundle AB to buying nothing. This implies a reduction in P_{AB} without an increase in P_A such that it has a negative effect on the demand of product A. Therefore, with $\Gamma > 0$, we have that $\partial s_A / \partial p_B < 0$ and products A and B are complements in demand.

In case 3 with $\Gamma < 0$, we can also distinguish two different types of marginal consumers: those located in a point like m and those in a point like o in Figure 2.7 panel 3. Similarly to the previous two cases, for consumers in point m, an increase in p_B does not have any effect on the demand of product A. For consumers in point o, an increase in p_B makes them switch from buying product B to buying product A. This implies an increase in the demand of product A. Therefore, with $\Gamma < 0$, we have that $\partial s_A / \partial p_B > 0$ and products A and B are substitutes in demand.

Suppose that: $u_{hA} = \beta_A - \alpha \ p_A + v_{hA}$; and $u_{hB} = \beta_B - \alpha \ p_B + v_{hB}$. Allowing for correlation between unobservables v_{hA} and v_{hB} is very important. Observing that frequent online readers are also frequent print readers might be evidence that the products in question are complementary, or it might reflect the correlation between unobservable tastes for goods. Suppose that (v_{hA}, v_{hB}) are standard normals with correlation ρ. The

Average Compensating Variation Conditional on Minivan Purchase, 1984: 1982–84 CPI-Adjusted Dollars

	OLS Logit	Instrumental Variable Logit	Random Coefficients	Random Coefficients and Microdata
Compensating variation:				
Median	9,573	5,130	1,217	783
Mean	13,652	7,414	3,171	1,247
Welfare change from difference in:				
Observed characteristics ($\delta_j+\mu_{ij}$)	−81,469	−44,249	−820	851
Logit Error (ϵ_{ij})	95,121	51,663	3,991	396
Income of minivan purchasers:				
Estimate from model	23,728	23,728	99,018	36,091
Difference from actual (CEX)	−15,748	−15,748	59,542	−3,385

Figure 2.6: Petrin (2002) Consumer welfare effects of minivans

parameters of the model are: β_A, β_B, α, ρ, Γ. The researcher (with consumer level data) observes prices and bundles market shares: $\mathbb{P}_A$, $\mathbb{P}_B$, $\mathbb{P}_{AB}$.

Identification

Even with micro-level data with information on shares $\mathbb{P}_A$, $\mathbb{P}_B$, $\mathbb{P}_{AB}$, the parameters (β_A, β_B, α, ρ, Γ) are not identified. Even if α is known, we have 3 data points and 4 parameters. Without further restrictions, a high value of P_{AB} can be explained by either high Γ or high ρ. We want to distinguish between these two interpretations because they have different economic and policy implications. Gentzkow considers two sources of identification: (1) Exclusion restrictions; and (2) Panel data and restrictions on the structure of the unobservables.

(a) Exclusion restrictions. Suppose that there is an exogenous consumer characteristic (or vector) z that enters in consumer valuation of product A but not of product B: $\beta_A(z)$, but β_B does not depend on z. For instance, if B is a print newspaper and A is its online version, z could be Internet access at work (at home could be more endogenous). Suppose z is binary for simplicity. Now, the data $[\mathbb{P}_A(z), \mathbb{P}_B(z), \mathbb{P}_{AB}(z)$: $z \in \{0,1\}]$ can identify $\beta_A(0)$, $\beta_A(1)$, β_B, Γ, and ρ. Intuition: if $\Gamma > 0$ (complementarity), then $z = 1$ should increase $P_A(z)$ and $P_{AB}(z)$. Otherwise, if $\Gamma = 0$, then $z = 1$ should increase $P_A(z)$ but not $P_{AB}(z)$.

(2) Panel Data. Suppose that we observe consumer choices at different periods of time, and suppose that: $v_{jht} = \eta_{jh} + \varepsilon_{jht}$. The time-invariant effects η_{Ah} and η_{Bh} are correlated with each other; but ε_{Aht} and ε_{Aht} are independent and i.i.d. over h,t. Preference parameters are assumed to be time invariant. Suppose that $T = 2$. We have 8 possible choice histories, 7 probabilities, and 4 parameters: β_A, β_B, Γ, and ρ. Identification intuition: if $\Gamma > 0$, changes over time in demand should be correlated

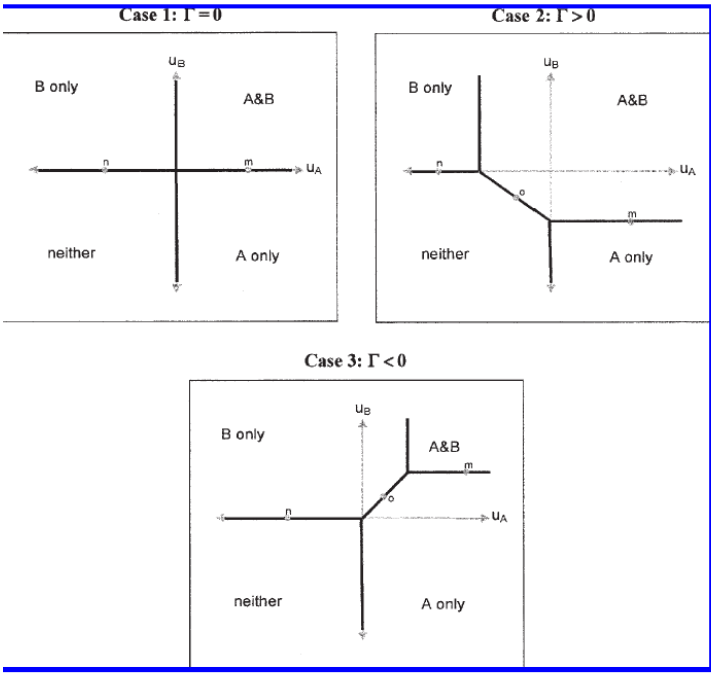


Figure 2.7: Gentzkow (2007) Choice regions and effect of Γ

between the two goods. If $\Gamma = 0$, changes over time should be uncorrelated between goods.

Data

Survey: 16,179 individuals in Washington DC, March-2000 and Feb. 2003. Information on individual and household characteristics, and readership of: print local newspapers read over last week; major local online newspapers over last week. Two main local print newspapers: Times and Post. One main online newspaper: post.com. Three products: Times, Post, and post.com. Outside alternative being all the other local papers.

Empirical results

Estimation results from reduced-form OLS regressions and from a structural model without heterogeneity suggest that the print and online editions of the Post are strong complements. According to those estimates, the addition of the post.com to the market increases profits from the Post print edition by \$10.5 million per year. However, **properly accounting for consumer heterogeneity changes the conclusions substantially**. Estimates of the model with both observed and unobserved heterogeneity show that the print and online editions are **significant substitutes**. Figure 2.8 presents estimates of the Γ parameters.

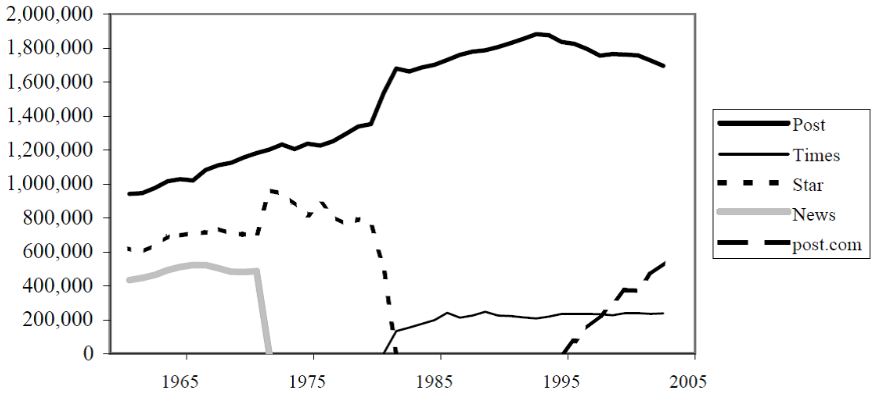


Figure 2.8: Gentzkow (2007) Time series of readers

Table 2.9 presents estimates of the effect of the online edition on the print edition. Raising the price of the Post by $0.10 would increase post.com readership by about 2%. Removing the post.com from the market entirely would increase readership of the Post by 27,000 readers per day, or 1.5%. The estimated $33.2 million of revenue generated by the post.com comes at a cost of about $5.5 million in lost Post readership. For consumers, the online edition generated a per-reader surplus of $0.30 per day, implying a total welfare gain of $45 million per year. Reduced-form OLS regressions and a structural model without heterogeneity suggest that the print and online editions of the Post are strong complements.

TABLE 6—PARAMETER ESTIMATES FROM FULL MODEL: OTHER

Interaction terms		*Excluded variables (coefficient in utility of post.com)*	
Post-post.com	−1.285**	Internet at work	1.357**
	(0.2307)		(0.180)
Post-*Times*	0.0809	Fast connection	0.146
	(0.2479)		(0.193)
post.com-*Times*	−1.231**	Use for education-related	0.361
	(0.4832)		(0.212)
Nonlinear parameters		Use for work	0.582**
τ	6.846**		(0.222)
	(0.5027)		
γ	0.0454**		
	(0.0179)		

Figure 2.9: Gentzkow (2007) Estimates of Γ parameters

2.5 Appendix

2.5.1 Derivation of demand systems

The Linear Expenditure System

The utility function has the Stone-Geary form:

$$U = (q_0 - \gamma_0)^{\alpha_0}\,(q_1 - \gamma_1)^{\alpha_1} \ldots (q_J - \gamma_J)^{\alpha_J} \tag{2.77}$$

The marginal utility of product j is $U_j = \alpha_j \dfrac{U}{q_j - \gamma_j}$. Therefore, the marginal condition of optimality $U_j - \lambda p_j = 0$ implies that $\alpha_j \dfrac{U}{q_j - \gamma_j} = \lambda p_j$, or equivalently,

$$p_j\, q_j = \alpha_j \frac{U}{\lambda} + p_j\, \gamma_j \tag{2.78}$$

Adding up this expression over the $J+1$ products and using the budget constraint and the restriction $\sum_{j=0}^{J} \alpha_j = 1$, we have that $y = \frac{U}{\lambda} + \sum_{j=0}^{J} p_j\, \gamma_j$ such that:

$$\frac{U}{\lambda} = y - \sum_{j=0}^{J} p_j \gamma_j \tag{2.79}$$

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

$$q_j = \gamma_j + \alpha_j \left[\frac{y - P_\gamma}{p_j}\right] \tag{2.80}$$

where P_γ is the aggregate price index $\sum_{i=0}^{J} p_i\, \gamma_i$.

TABLE 8—IMPACT OF THE ONLINE EDITION ON DEMAND FOR PRINT

Case 1: Full model	
Cross-price derivative	8,358
	(1,436)
Change in print readership	−26,822
	(4,483)
Change in print profits	−$ 5,466,846
	(913,699)
Case 2: Model with observable characteristics only	
Cross-price derivative	−8,421
	(752)
Change in print readership	25,655
	(2,270)
Change in print profits	$ 5,229,009
	(462,771)
Case 3: Model with no heterogeneity	
Cross-price derivative	−16,143
	(702)
Change in print readership	51,897
	(2,254)
Change in print profits	$10,577,720
	(459,464)

Figure 2.10: Gentzkow (2007) Effect of Online on Print

Constant Elasticity of Substitution demand system

The utility function is:

$$U = \left(\sum_{j=0}^{J} q_j^{\sigma} \right)^{1/\sigma} \tag{2.81}$$

The marginal utility is $U_j = \dfrac{q_j^{\sigma-1}\, U}{\sum_{i=0}^{J} q_i^{\sigma}}$ such that the marginal condition of optimality for product j is $\dfrac{q_j^{\sigma-1}\, U}{\sum_{i=0}^{J} [\alpha_i\, q_i]^{\sigma}} - \lambda\ p_j = 0$. We can re-write this condition as:

$$\frac{q_j^{\sigma-1}}{\sum_{i=0}^{J} [\alpha_i\, q_i]^{\sigma}} \frac{U}{\lambda} = p_j\, q_j \tag{2.82}$$

Adding the expression over the $J+1$ products, we have that:

$$y = \sum_{j=0}^{J} p_j\, q_j = \frac{U}{\lambda} \tag{2.83}$$

That is, $\dfrac{U}{\lambda} = y$. Plugging this result into the marginal condition for product j above, and taking into account that $\sum_{i=0}^{J} q_i^{\sigma} = U^{\sigma}$, we have that:

$$\frac{q_j^{\sigma}}{U^{\sigma}} y = p_j\, q_j \tag{2.84}$$

This equation can be re-written as:

$$q_j = \left[\frac{y}{p_j}\right]^{1/(1-\sigma)} \left[\frac{1}{U^\sigma}\right]^{1/(1-\sigma)} \tag{2.85}$$

Plugging this expression in the definition of the utility function, we can get:

$$U = \left(\sum_{j=0}^{J}\left[\frac{y}{p_j}\right]^{\sigma/(1-\sigma)}\right)^{1/\sigma} \left[\frac{1}{U^\sigma}\right]^{1/(1-\sigma)} \tag{2.86}$$

Solving for U, we have:

$$U = \left(\sum_{j=0}^{J}\left[\frac{y}{p_j}\right]^{\sigma/(1-\sigma)}\right)^{(1-\sigma)/\sigma} = \frac{y}{P_\sigma} \tag{2.87}$$

where P_σ is the price index:

$$P_\sigma = \left(\sum_{j=0}^{J}\left[p_j\right]^{-\sigma/(1-\sigma)}\right)^{-(1-\sigma)/\sigma} \tag{2.88}$$

Finally, plugging these results into the expression $q_j = \left[\frac{y}{p_j}\right]^{1/(1-\sigma)} \left[\frac{1}{U^\sigma}\right]^{1/(1-\sigma)}$, we get the CES demand equations:

$$q_j = \frac{y}{P_\sigma} \left[\frac{p_j}{P_\sigma}\right]^{-1/(1-\sigma)} \tag{2.89}$$

2.6 Exercises

2.6.1 Exercise 1

To answer the questions in this exercise you need to use the dataset `verboven_cars.dta` Use this dataset to implement the estimations describe below. Please, provide the STATA code that you use to obtain the results. For all the models that you estimate below, impose the following conditions:

- For market size (number of consumers), use Population/4, that is, `pop/4`

- Use prices measured in euros (`eurpr`).

- For the product characteristics in the demand system, include the characteristics: `hp, li, wi, cy, le,` and `he.`

- Include also as explanatory variables the market characteristics: ln(pop) and log(gdp).

- In all the OLS estimations include fixed effects for market (`ma`), year (`ye`), and brand (`brd`).

- Include the price in logarithms, that is, ln(`eurpr`).

- Allow the coefficient for log-price to be different for different markets (countries). That is, include as explanatory variables the log price, but also the log price interacting (multiplying) each of the market (country) dummies except one country dummy (say the dummy for Germany) that you use as a benchmark.

Question 1.1 Obtain the OLS-Fixed effects estimator of the Standard logit model. Interpret the results.

Question 1.2 Test the null hypothesis that all countries have the same price coefficient.

Question 1.3 Based on the estimated model, obtain the average price elasticity of demand for each country evaluated at the mean values of prices and market shares for that country.

2.6.2 Exercise 2

The STATA datafile `eco2901_problemset_01_2012_airlines_data.dta` contains a panel dataset of the US airline industry in 2004. A market is a *route* or directional city-pair, for instance, round-trip Boston to Chicago. A product is the combination of route (m), airline (f), and the indicator of stop flight or nonstop flight. For instance, a round-trip Boston to Chicago, non-stop, with American Airlines is an example of product. Products compete with each other at the market (route) level. Therefore, the set of products in market m consists of all the airlines with service in that route either with nonstop or with stop flights. The dataset contains $2,950$ routes, 4 quarters, and 11 airlines (where the airline "Others" is a combination of multiple small airlines). The following table includes the list of variables in the dataset and a brief description.

Variable name		Description
`route_city`	:	Route: Origin city to Destination City
`route_id`	:	Route: Identification number
`airline`	:	Airline: Name (Code)
`direct`	:	Dummy of Non-stop flights
`quarter`	:	Quarter of year 2004
`pop04_origin`	:	Population Origin city, 2004 (in thousands)
`pop04_dest`	:	Population Destination city, 2004 (in thousands)
`price`	:	Average price: route, airline, stop/nonstop, quarter (in dollars)
`passengers`	:	Number of passengers: route, airline, stop/nonstop, quarter
`avg_miles`	:	Average miles flown for route, airline, stop/nonstop, quarter
`HUB_origin`	:	Hub size of airline at origin (in million passengers)
`HUB_dest`	:	Hub size of airline at destination (in million passengers)

In all the models of demand that we estimate below, we include time-dummies and the following vector of product characteristics:

{ `price, direct dummy, avg_miles, HUB_origin, HUB_dest, airline dummies`

In some estimations we also include market (route) fixed effects. For the construction of market shares, we use as measure of market size (total number of consumers) the average population in the origin and destination cities, in number of people, that is, `1000*(pop04_origin + pop04_dest)/2`.

Question 2.1. Estimate a Standard Logit model of demand: (a) by OLS without route fixed effects; (b) by OLS with route fixed effects. Interpret the results. What is the average consumer willingness to pay (in dollars) for a nonstop flight (relative to a stop flight), ceteris paribus? What is the average consumer willingness to pay for one million more people of hub size in the origin airport, ceteris paribus? What is the average consumer willingness to pay for Continental relative to American Airlines, ceteris paribus? Based on the estimated model, obtain the average elasticity of demand for Southwest products. Compare it with the average elasticity of demand for American Airline products.

Question 2.2. Consider a Nested Logit model where the first nest consists of the choice between groups "Stop", "Nonstop", and "Outside alternative", and the second nest consists in the choice of airline. Estimate this Nested Logit model of demand: (a) by OLS without route fixed effects; (b) by OLS with route fixed effects. Interpret the results. Answer the same questions as in Question 2.1.

Question 2.3. Consider the Nested Logit model in Question 2.2. Propose and implement an IV estimator that deals with the potential endogeneity of prices. Justify your choice of instruments, for instance, BLP, or Hausman-Nevo, or Arellano-Bond, ... Interpret the results. Compare them with the ones from Question 2.2.

Question 2.4. Given your favorite estimation of the demand system, calculate price-cost margins for every observation in the sample. Use these price cost margins to estimate a marginal cost function in terms of all the product characteristics, except price. Assume constant marginal costs. Include also route fixed effects. Interpret the results.

Question 2.5. Consider the route Boston to San Francisco ("BOS to SFO") in the fourth quarter of 2004. There are 13 active products in this route-quarter, and 5 of them are non-stop products. The number of active airlines is 8: with both stop and non-stop flights, America West (HP), American Airlines (AA), Continental (CO), US Airways (US), and United (UA); and with only stop flights, Delta (DL), Northwest (NW), and "Others". Consider the "hypothetical" merger (in 2004) between Delta and Northwest. The new airline, say DL-NW, has airline fixed effects, in demand and costs, equal to the average of the fixed effects of the merging companies DL and NW. As for the characteristics of the new airline in this route: `avg_miles` is equal to the minimum of `avg_miles` of the two merging companies; `HUB_origin = 45`; `HUB_dest = 36`; and the new airline still only provides stop flights in this route.

(a) Using the estimated model, obtain airlines profits in this route-quarter before the hypothetical merger.
(b) Calculate equilibrium prices, number of passengers, and profits , in this route-quarter after the merger. Comment the results.
(c) Suppose that, as the result of the merger, the new airline decides also to operate non-stop flights in this route. Calculate equilibrium prices, number of passengers, and profits , in this route-quarter after the merger. Comment the results.