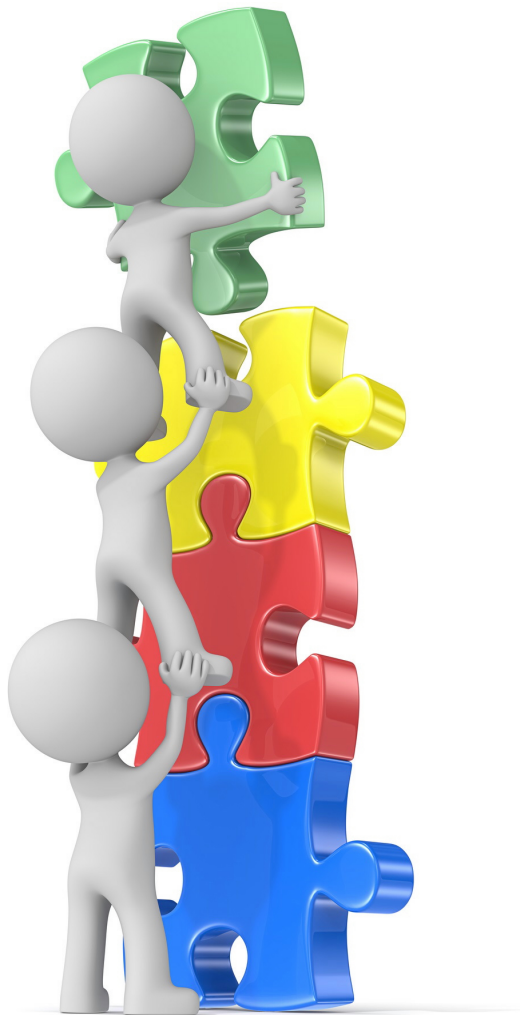# Empirical Industrial Organization:
Models, Methods, and Applications

Victor Aguirregabiria

(University of Toronto)

This version: December 30th, 2018

Contents

CHAPTER 1

Introduction

1. Some general ideas on Empirical Industrial Organization

Industrial Organization (IO) studies the behavior of firms in markets. We are interested in understanding how firms interact strategically in markets, and how their actions affect the market allocation. IO economists are particularly interested in three aspects related to a market allocation: *market structure*, *firms' market power*, and *firms' strategies*. These are key concepts in IO. *Market structure* is a description of the number of firms in the market and of their respective market shares. A monopoly is an extreme case of market structure where a single firm concentrates the total output in the market. At the other extreme we have an atomist market structure where industry output is equally shared by a very large number of very small firms. Between these two extremes, we have a whole spectrum of possible oligopoly market structures. *Market power* (or *monopoly power*) is the ability of a firm, or group of firms, to gain extraordinary profits, to get rents above those needed to remunerate all the inputs at market prices. A *firm's strategy* is a description of the firms' actions (e.g., selling price, production, market entry) contingent on the state of the market (e.g., demand and cost conditions). We say that a firm behaves strategically if it takes into account that its actions affect other firms' profits and that a change in its own strategy can generate as a response a change in the strategies of competing firms.

A significant part of the research in IO deals with understanding the determinants of market power, market structure, and firms' strategies in actual markets and industries. IO economists propose models where these variables are determined endogenously and depend on multiple exogenous factors such as consumer demand, input supply, technology, regulation, and firms' beliefs about the behavior of competitors and the nature of competition. The typical model in IO treats demand, technology, and institutional features as given, and postulates some assumptions about how firms compete in a market. Based on these assumptions, we study firms' strategies. In particular, we are interested in finding what is the strategy that maximize a firm's profit given its beliefs about the behavior of other firms, and what is the set of firms' strategies that are consistent with profit maximization and with rational beliefs about each other behavior, i.e., equilibrium strategies. We use Game Theory

tools to find these equilibrium strategies, and to study how changes in exogenous factors affect firms' strategies, market structure, firms' profits, and consumer welfare.

The models of Perfect Competition and of Perfect Monopoly are two examples of IO models. However, they are extreme cases and they do not provide a realistic description of many markets and industries in our today's economy. Many interesting markets are characterized by a concentrated market structure, market power, and firms' strategic behavior. We refer to these markets as *oligopoly markets*, and they are the focus of IO.

Most of the issues that we study in IO have an important empirical component. To answer questions related to competition between firms in an industry, we typically need information on consumer demand, firms' costs, and firms' strategies or actions in that industry. *Empirical Industrial Organization* (EIO) deals with the combination of data, models, and econometric methods to answer empirical questions related to the behavior of firms in markets. The tools of EIO are used in practice by firms, government agencies, consulting companies, and academic researchers. Firms use these tools to improve their strategies, decision making, and profits. For instance, EIO methods are useful tools to determine a firm's optimal prices, to evaluate the value added of a merger, to predict the implications of introducing a new product in the market, or to measure the benefits of price discrimination. Government agencies use the tools of industrial organization to evaluate the effects of a new policy in an industry (e.g., an increase in the sales tax, or an environmental policy), or to identify anti-competitive practices such as collusion, price fixing, or predatory conducts. Academic researchers use the tools of EIO to improve our understanding of industry competition. The following are some examples of this type of questions.

Example 1. A company considers launching a new product, e.g., a new smartphone. To estimate the profits that the new product will generate to the company, and to decide the initial price that maximizes these profits, the company needs to predict the demand for this new product, and the response (i.e., price changes) of the other firms competing in the market of smartphones. Data on sales, prices, and product attributes from firms and products that are already active in the market can be used together models and methods in EIO to estimate the demand and the profit maximizing price of the new product, and to predict the response of competing firms.

Example 2. A government has introduced a new environmental policy that imposes new restrictions on the emissions of pollutants from factories in an industry. The new policy encourages firms in this industry to adopt a new technology that is environmentally cleaner. This alternative technology reduces variable costs but increases fixed costs. These changes in the cost structure affect competition. In particular, we expect a decline in the number of firms and an increase output-per-firm in the industry. The government wants to know

how this new policy has affected competition and welfare in the industry. Using data on prices, quantities, and number of firms in the industry, together with a model of oligopoly competition, we can evaluate the effects of this policy change.

Example 3. For many years, the industry of micro-processors for personal computers has been characterized by the duopoly of Intel and AMD, with a clear leadership by Intel that enjoys more than two-thirds of the world market and a substantial degree of market power. There are multiple factors that may contribute to explain this market power and its persistence over time. For instance, large entry costs, economies of scale, learning-by-doing, consumer brand loyalty, or predatory conduct and entry deterrence, are potential, not mutually exclusive, explanations. What is the relative contribution of each of these factors to explain the observed market structure and market power? Data on prices, quantities, product characteristics, and firms' investment in capacity can help us to understand and to measure the contribution of these factors.

2. Data in Empirical IO

Early research in empirical IO between the 1950s and 1970s was based on aggregate industry level data from multiple industries (Bain, 1951 and 1954, Demsetz, 1973). Studies in this literature looked at the empirical relationship between a measure of market structure and a measure of market power. The typical study estimated a linear regression model

$$\frac{P_n - MC_n}{P_n} = \beta_0 + \beta_1 HHI_n + \varepsilon_n \tag{2.1}$$

where each observation n in the sample represents an industry, the dependent variable $\frac{P_n - MC_n}{P_n}$ is the *Lerner Index* that measures market power and is defined as price minus marginal cost divided by price, and the key explanatory variable is the *Herfindahl-Hirschman Index* that measures market concentration, i.e., $HHI_n = \sum_{i=1}^{N_n} s_{ni}^2$, where s_{ni} is the share of firm i's output in total industry n output. These linear regression models were estimated using industry-level cross-sectional data from very diverse industries, and they typically found a positive and statistically significant relationship between concentration and market power, i.e., the OLS estimate of β_1 was statistical greater than zero. A main purpose of these empirical studies was to identify a relationship between market concentration and market structure that could be applied to most industries. Furthermore, the interpretation of the estimated regression functions was causal and not just as an equilibrium relationship, i.e., *ceteris paribus*, an increase in the degree of market concentration implies an increase in market power.

In the 1980s, the seminal work of Bresnahan (1981, 1982, 1987), Porter (1983), Schmalensee (1989), and Sutton (1991), among others, configured the basis for the so called *New Empirical*

IO. These authors pointed out at the serious limitations in the previous empirical literature based on aggregate industry-level data. One of the criticisms to the previous literature was that industries, even those apparently similar, can be very different in their exogenous or primitive characteristics such as demand, technology, and regulation. This heterogeneity implies that the relationship between market concentration and price-cost margins can be also very different across industries. The parameters of the linear regression models estimated in previous literature are heterogeneous across industries. A second important criticism to the old EIO literature was that industry concentration, or market structure, cannot be considered as an exogenous explanatory variable. Market power and market structure are both endogenous variables that are jointly determined in an industry. The regression equation of market power on market structure should be interpreted as an equilibrium condition where there are multiple exogenous factors, both observable and unobservable to the researcher, that simultaneously affect these two endogenous variables. Not taking into account the correlation between the explanatory variable (market structure) and the error term (unobserved heterogeneity in industry fundamentals) in this regression model implies an spurious estimation of *causal effect* or *ceteris paribus effect* of market structure on market power.

Given these limitations of the old EIO, the proponents of the *New Empirical IO* emphasized the need to study competition by looking at each industry separately using richer data at a more disaggregate level and combining these data with game theoretical models of oligopoly competition. Since then, the typical empirical application in IO has used data of a single industry, with information at the level of individual firms, products, and markets, on prices, quantities, number of firms, and exogenous characteristics affecting demand and costs.

In the old EIO, sample variability in the data came from looking at multiple industries. This source of sample variation is absent in the typical empirical study in the New EIO. Furthermore, given that most studies look at oligopoly industries with a few firms, sample variation across firms is also very limited and it is not enough to obtain consistent and precise estimates of parameters of interest. What are the main sources of sample variability in empirical studies in EIO? Most of the sample variation in these studies come from observing multiple products and local markets within the same industry. For instance, in some industries the existence of transportation costs implies that firms compete for consumers at the level of local geographic markets. The particular description of a geographic local market (e.g., a city, a county, a census tract, or a census block) depends on the specific industry under study. Prices and market shares are determined at the local market level. Therefore, having data from many local markets can help to identify the parameters of our models. Sample variation at the product level is also extremely helpful. Most industries in today's

economies are characterized by product differentiation. Firms produce and sell many varieties of a product. Having data at the level of very specific individual products and markets is key to identify and estimate most IO models that we study in this book.

The typical dataset in EIO consists of cross-sectional or panel data of many products and/or local markets from the same industry, with information on selling prices, produced quantities, product attributes, and local market characteristics. Ideally, we would like to have data on firms' costs. However, this information is very rare. Firms are very secretive about their costs and strategies. Therefore, we typically have to infer firms' costs from our information on prices and quantities. When we have information on firms' inputs, inference on firms' costs can take the form of estimating production functions. When information on firms' inputs is not available, or not rich enough, we exploit our models of competition and profit maximization to infer firms' costs. Similarly, we will have to estimate price-cost margins (market power) and firms' profits using this information.

3. Specification of a structural model in Empirical IO

To study competition in an industry, EIO researchers propose and estimate structural models of demand and supply where firms behave strategically. These models typically have the following components or submodels: a model of consumer behavior or demand; a specification of firms' costs; a static equilibrium model of firms' competition in prices or quantities; a dynamic equilibrium model of firms' competition in some form of investment such as capacity, advertising, quality, or product characteristics; and a model of firm entry (and exit) in a market. The parameters of the model are structural in the sense that they describe consumer preferences, production technology, and institutional constraints. This class of econometric models are useful tools to understand competition, business strategies, and the evolution of an industry, to identify collusive and anti-competitive behavior, or to evaluate the effects of public policies in oligopoly industries, to mention some of their possible applications.

To understand the typical structure of an EIO model, and to illustrate and discuss some important economic and econometric issues in this class of models, this section presents a simple empirical model of oligopoly competition. Though simple, this model incorporates already some important issues related to the specification, testable predictions, endogeneity, identification, estimation, and policy experiments, that we will see and study in detail throughout the rest of the book. This example is inspired in Ryan (2012), and the model can seen as a simplified version of the model in that paper.

We start with an empirical question. Suppose that we want to study competition in the cement industry of a country or region. It is well-known that this industry is energy

intensive and generates a large amount of air pollutants. For these reasons, the government or regulator in this example is evaluating whether to pass a new law that restricts the amount of emissions a cement plant can make. This law would imply the adoption of a type of technology that it is already available but that few plants currently use. The "new" technology implies lower marginal costs but larger fixed costs than the "old" technology. The government would like to evaluate the implications of the new environmental regulation on firms' profits, competition, consumer welfare, and air pollution. As we discuss below, this evaluation can be *ex-ante* (i.e., before the new policy is actually implemented) or *ex-post* (i.e., after the implementation of the policy change).

The next step is to specify a model that incorporates the **key features of the industry** that are important to answer our empirical question. The researcher needs to have some knowledge about competition in this industry, and about the most important features in demand and technology that characterize the industry. The model that I propose here incorporates four basic but important features of the cement industry. First, it is an homogeneous product industry. There is very little differentiation in the cement product. Nevertheless, the existence of large transportation costs per dollar value of cement makes spatial differentiation a potentially important dimension for firm competition. In this simple example, we abstract from product differentiation when modelling competition between firms, though, as explained below, we take into it account to a certain extent when we define local markets. Second, there are substantial fixed costs of operating a cement plant. The cost of buying (or renting) cement kilns, and the maintenance of this equipment, does not depend on the amount of output the plant produces and it represents a substantial fraction of the total cost of a cement plant. Third, variable production costs increase more than proportionally when output approaches the maximum installed capacity of the plant. Fourth, transportation of cement per dollar value is very high. This explains why the industry is very local. Cement plants are located nearby the point of demand (i.e., construction places in cities or small towns) and they do not compete with cement plants located in other towns. For the moment, the simple model that we present here, ignores an important feature of the industry that will turn relevant for our empirical question. Installed capacity is a dynamic decision that depends on the plant's capacity investments and on depreciation.

The specification of the model depends importantly on the **data** that is available for the researcher. The level of aggregation of the data (e.g., consumer and firm level vs. market level data), its frequency, or the availability or not of panel data are important factors that the researcher should take into account when she specifies the model. Model features that are important to explain firm-level data might be quite irrelevant, or they may be underidentified, when using market level data. In this example, we consider a panel (longitudinal) dataset

with aggregate information at the level of local markets. Later in this chapter we discuss the advantages of using richer firm-level data. The dataset consists of M local markets (e.g., towns) observed over T consecutive quarters.[1] We index markets by m and quarters by t. For every market-quarter observation, the dataset contains information on the number of plants operating in the market (N_{mt}), aggregate amount of output produced by all the plants (Q_{mt}), market price (P_{mt}), and some exogenous market characteristics ($\mathbf{X}_{mt}$) such as population, average income, etc.

$$Data = \{ \ P_{mt} \ , \ Q_{mt} \ , \ N_{mt}, \ \mathbf{X}_{mt} : m = 1, 2, ..., M; \ t = 1, 2, ..., T \ \} \tag{3.1}$$

Note that the researcher does not observe output at the plant level. Though the absence of data at the firm level is not ideal it is not uncommon either, especially when using publicly available data from census of manufacturers or businesses. Without information on output at the firm-level, our model has to impose strong restrictions on the degree heterogeneity in firms' demand and costs. Later in this chapter, we discuss potential biases generated by these restrictions and how we can avoid them when we have firm-level data.

Our **model of oligopoly competition** has four main components: (a) demand equation; (b) cost function; (c) model of Cournot competition; and (d) model of market entry. An important aspect in the construction of an econometric model is the specification of unobservables. Including unobservable variables in our models is a way to acknowledge the rich amount of heterogeneity in the real world (between firms, markets, or products, and over time), as well as the limited information of the researcher relative to the information available to actual economic agents in our models. Unobservables also account for measurement errors in the data. In general, the richer the specification of unobservables in a model, the more robust the empirical findings. Of course, there is a limit to the degree of unobserved heterogeneity that we can incorporate in our models, and this limit is given by the identification of the model.

3.1. Demand equation. In this simple model we assume cement is an homogeneous product. We also abstract from spatial differentiation of cement plants.[2] We postulate a demand equation that is linear in prices and in parameters.

$$Q_{mt} = S_{mt} \ \left(\mathbf{X}_{mt}^{D} \ \boldsymbol{\beta}_{X} - \beta_{P} \ P_{mt} + \varepsilon_{mt}^{D} \right) \tag{3.2}$$

$\boldsymbol{\beta}_{X}$ and $\beta_{P} \geq 0$ are parameters. S_{mt} represents demand size or population size. $\mathbf{X}_{mt}^{D}$ is a subvector of $\mathbf{X}_{mt}$ that contains observable variables that affect the demand of cement in a

---

[1]The definition of what is a local market represents an important modelling decision for this type of data and empirical application. We will examine this issue in detail in chapter 5.

[2]See Miller and Osborne (2013) for an empirical study of spatial differentiation and competition of cement plants.

market, such as average income, population growth, or age composition of the population. ε_{mt}^D is an unobservable shock in demand per capita. This shock implies vertical parallel shifts in the demand curve.[3] There are at least two ways to generate this demand equation from consumer utility maximizing behavior. A first interpretation is that $\mathbf{X}_{mt}^D \, \boldsymbol{\beta}_X - \beta_P \, P_{mt} + \varepsilon_{mt}^D$ is the downward sloping demand curve of a *representative consumer* in market m at period t. According to this interpretation, $\mathbf{X}_{mt}^D \, \boldsymbol{\beta}_X + \varepsilon_{mt}^D$ is the willingness to pay of this representative consumer for the first unit that he buys of the product, and β_P captures the decreasing marginal utility from additional units. An alternative interpretation is based on the assumption that there is a continuum of consumers in the market with measure S_{mt}. Each consumer can buy at most one unit of the product. A consumer with willingness to pay v has a demand equal to one unit if $(v - P_{mt}) \geq 0$ and his demand is equal to zero if $(v - P_{mt}) < 0$. Then, the aggregate market demand is $Q_{mt} = S_{mt}^* \, (1 - G_{mt}(P_{mt}))$ where $G_{mt}(v)$ is the distribution function of consumers' willingness to pay in market m at period t, such that $\Pr\left(v \geq P_{mt}\right) = 1 - G_{mt}(P_{mt})$. Suppose that the distribution function G_{mt} is uniform with support $[(A_{mt}-1)/\beta_P \, , \, A_{mt}/\beta_P]$ and $A_{mt} \equiv \mathbf{X}_{mt}^D \, \boldsymbol{\beta}_X + \varepsilon_{mt}^D$. Then, the aggregate market demand has the form in equation (3.2).

For some of the derivations below, it is convenient to represent the demand using the *inverse demand curve*:

$$P_{mt} = A_{mt} - B_{mt} \, Q_{mt} \tag{3.3}$$

where the intercept A_{mt} has the same definition as above, and the slope B_{mt} is $1/(\beta_P \, S_{mt})$. Using the standard representation of the demand curve in the plane, with Q in the horizontal axis and P in the vertical axis, we have that this curve moves upward when A_{mt} increases (vertical parallel shift) or when B_{mt} declines (counter-clockwise rotation).[4]

3.2. Cost function. The **cost function** of a firm has two components, variable cost and fixed cost: $C(q) = VC(q) + FC$, where q is the amount of output produced by a single firm, $C(q)$ is the total cost of a firm active in the market, and $VC(q)$ and FC represent variable cost and fixed cost, respectively.

If we had firm-level data on output, inputs, and input prices, we could estimate a production function and then use the dual approach to construct the variable cost and fixed cost function. For instance, with a Cobb-Douglas PF in terms of labor (L) and capital (K), we can estimate the production function in logs, $\ln q = \alpha_L \ln L + \alpha_K \ln K + \varepsilon$, and under the

---

[3]A more general specification of the linear demand equation includes an unobservable shock that affects the slope of the demand curve.

[4]In principle, market size S_{mt}^* could enter the vector $\mathbf{X}_{mt}^D$ to take into account that the distribution of consumers willingness to pay may change with the size of the population in the market. In that case, an increase in market size implies both a vertical shift and a rotation in the demand curve.

assumption that labor is a variable input and capital is a fixed input, we have that:

$$VC(q) = p_L \left[\frac{q}{\exp\{\varepsilon\}K^{\alpha_K}}\right]^{1/\alpha_L} \tag{3.4}$$

and

$$FC = p_K\ K \tag{3.5}$$

where p_L and p_K represent input prices.

Here we consider a common situation where the researcher does not have data on inputs at the firm level. Costs cannot be identified/estimated from a production function. We will estimate costs using revealed preference. Firms' choice of output reveal information about their marginal costs. And firms' decision of being active or not in the market reveal information about their fixed costs. *** Comment that there are some restrictions also involved in the PF approach: why labor the only variable input? why observed capital the only fixed input? There may be inputs unobservable to the research. The reveal preference approach is robust to these unobserved inputs. But the PF approach, if feasible, provides useful identifying information ***

We start assuming that every firm, either an incumbent or a potential entrant, has the same cost function. For convenience, we specify a quadratic variable cost function:

$$VC_{mt}(q) = \left(\mathbf{X}_{mt}^{MC}\ \boldsymbol{\gamma}_X^{MC} + \varepsilon_{mt}^{MC}\right)\ q + \frac{\gamma_2^{MC}}{2}q^2 \tag{3.6}$$

$\boldsymbol{\gamma}_X^{MC}$ and γ_2^{MC} are parameters. $\mathbf{X}_{mt}^{MC}$ is a subvector of $\mathbf{X}_{mt}$ that contains observable variables that affect the marginal cost of cement production, including the prices of variable inputs such as limestone, energy, or labor. ε_{mt}^{MC} is a market shock in marginal cost that is unobserved to the researcher but observable to firms. Given this variable cost function, the marginal cost is $MC_{mt}(q) = \overline{MC}_{mt} + \gamma_2^{MC}q$, where $\overline{MC}_{mt} \equiv \mathbf{X}_{mt}^{MC}\ \boldsymbol{\gamma}_X^{MC} + \varepsilon_{mt}^{MC}$ represents the exogenous part of the marginal cost, as well as the minimum possible value of the the marginal cost. The increasing component of the marginal cost, $\gamma_2^{MC}q$, captures the industry feature that this cost increases when output approaches the maximum capacity of a plant.

The fixed cost is specified as $FC_{mt} = \mathbf{X}_{mt}^{FC}\ \boldsymbol{\gamma}_X^{FC} + \varepsilon_{mt}^{FC}$, where $\boldsymbol{\gamma}_X^{FC}$ is a vector of parameters. $\mathbf{X}_{mt}^{FC}$ is a vector of observable variables that affect fixed costs such as the rental price of fixed capital equipment. ε_{mt}^{FC} is an unobservable market specific shock that captures the deviation of market m at quarter t from the conditional mean value $\mathbf{X}_{mt}^{FC}\ \boldsymbol{\gamma}_X^{FC}$. By including the market-specific shocks ε_{mt}^{MC} and ε_{mt}^{FC} we allow for market heterogeneity in costs that is unobservable to the researcher.

3.3. Cournot competition. Suppose that there are N_{mt} plants active in local market m at quarter t. For the moment, we treat the number of active firms as given, though this variable is endogenous in the model and we explain later how it is determined in the

equilibrium of the model. We assume that firms active in a local market compete with each other ala Cournot. The assumption of Cournot competition is far of being innocuous for the predictions of the model, and we reexamine this assumption at the end of this chapter.

The profit function of firm i is:

$$\Pi_{mt}(q_i, \widetilde{Q_i}) = P_{mt}(q_i + \widetilde{Q_i}) \, q_i - VC_{mt}(q_i) - FC_{mt} \tag{3.7}$$

where q_i is firm i's own output, and $\widetilde{Q_i}$ represents the firm i's beliefs about the total amount of output of the other firms in the market. Under the assumption of *Nash-Cournot* competition, each firm i takes as given the quantity produced by the rest of the firms, $\widetilde{Q_i}$, and chooses his own output q_i to maximize his profit. The profit function $\Pi_{mt}(q_i, \widetilde{Q_i})$ is globally concave in q_i for any positive value of $\widetilde{Q_i}$. Therefore, there is a unique value of q_i that maximizes the firm's profit, i.e., a firm best response is a function. This best response output is characterized by the following condition of optimality that establishes that marginal revenue equals marginal cost:

$$P_{mt} + \frac{\partial P_{mt}(q_i + \widetilde{Q_i})}{\partial q_i} \, q_i = MC_{mt}(q_i) \tag{3.8}$$

Taking into account that in our linear demand model $\partial P_{mt}/\partial q = -B_{mt}$, and that the equilibrium is symmetric ($q_i = q$ for every firm i) such that $Q_{mt} = q + \widetilde{Q} = N_{mt} \, q$, we can get the following expression for output-per-firm in the Cournot equilibrium with N active firms:

$$q_{mt}(N) = \frac{A_{mt} - \overline{MC}_{mt}}{B_{mt} \, (N + 1) + \gamma_2^{MC}} \tag{3.9}$$

This equation shows that, keeping the number of active firms fixed, output per firm increases with demand, declines with marginal cost, and it does not depend on fixed cost. This is a general result that does not depend on the specific functional form that we have chosen for demand and variable costs: by definition, fixed costs do not have any influence on marginal revenue or marginal costs when the number of firms in the market is fixed. As we show below, fixed costs do have an indirect effect on output per firm through its effect on the number of active firms: the larger the fixed cost, the lower the number of firms, and the larger output per firm.

Price over average variable cost is $P_{mt} - AVC_{mt} = [A_{mt} - B_{mt} \, N_{mt} \, q_{mt}(N)] - [\overline{MC}_{mt} + \gamma_2^{MC}/2 \, q_{mt}(N)] = [A_{mt} - \overline{MC}_{mt}] - [B_{mt} \, N_{mt} + \gamma_2^{MC}/2] \, q_{mt}(N)$. Plugging expression (3.9) into this equation, we get the following relationship between price-cost margin and output-per-firm in the Cournot equilibrium:

$$P_{mt} - AVC_{mt} = \frac{\left(B_{mt} + \gamma_2^{MC}/2\right) \, \left(A_{mt} - \overline{MC}_{mt}\right)}{B_{mt} \, (N_{mt} + 1) + \gamma_2^{MC}} = \left(B_{mt} + \gamma_2^{MC}/2\right) \, q_{mt}(N) \tag{3.10}$$

As the number of plants goes to infinity, the equilibrium price-cost margin converges to zero, and price becomes equal to the minimum marginal cost, $\overline{MC}_{mt}$, that is achieved by having

infinite plants each with an atomist size. Plugging this expression into the profit function
we get that in a Cournot equilibrium with N firms, the profit of an active firm is:

$$
\begin{aligned}
\Pi^*_{mt}(N) &= (P_{mt} - AVC_{mt})\ q_{mt}(N) - FC_{mt} \\
&= \left(B_{mt} + \gamma_2^{MC}/2\right) \left(\frac{A_{mt} - \overline{MC}_{mt}}{B_{mt}\,(N+1) + \gamma_2^{MC}}\right)^2 - FC_{mt}
\end{aligned}
\tag{3.11}
$$

This Cournot equilibrium profit function is continuous and strictly decreasing in the number
of active firms, N. These properties of the equilibrium profit function are important for the
determination of the equilibrium number of active firms that we present in the next section.

3.4. Model of market entry. Now, we specify a model for how the number of active
firms in a local is determined in equilibrium. Remember that the profit of a firm that is
not active in the industry is zero.[5] The equilibrium entry condition establishes that every
active firm and every potential entrant is maximizing profits. Therefore, active firms should
be making non-negative profits, and potential entrants are not leaving positive profits on
the table. Active firms should be better off in the market than in the outside alternative.
That is, the profit of every active firms should be non-negative: $\Pi^*_{mt}(N_{mt}) \geq 0$. Potential
entrants should be better off in the outside alternative than in the market. That is, if a
potential entrant decides to enter in the market, it gets negative profits. Additional entry
implies negative profits: $\Pi^*_{mt}(N_{mt} + 1) < 0$.

Figure 1.1 presents the Cournot equilibrium profit of a firm as a function of the number
of firms in the market, N, for an example where the demand function is $P = \$100 - 0.1Q$,
the variable cost function is $VC(q) = \$20q + q^2/2$, and the fixed cost is $\$1,400$. As shown
in equation (3.11), the equilibrium profit function is continuous and strictly decreasing in
N. These properties imply that there is a unique value of N that satisfies the equilibrium
conditions $\Pi^*_{mt}(N) \geq 0$ and $\Pi^*_{mt}(N + 1) < 0$.[6] In the example of Figure 1.1, the equilibrium
number is 5 firms. In general, solving for the equilibrium number of firms is straightforward.
Let N^*_{mt} be the real number that (uniquely) solves the condition $\Pi^*_{mt}(N) = 0$. Given the
form of the equilibrium profit function $\Pi^*_{mt}(N)$, we have that:

$$
N^*_{mt} \equiv -\left(1 + \frac{\gamma_2^{MC}}{B_{mt}}\right) + \left(A_{mt} - \overline{MC}_{mt}\right) \sqrt{\frac{1 + \gamma_2^{MC}/2B_{mt}}{FC_{mt}\ B_{mt}}}
\tag{3.12}
$$

---

[5]In this model, the normalization to zero of the value of the outside option is innocuous. This normalization means that the 'fixed cost' FC_{mt} is actually the sum of the fixed cost in this market and the firm's profit in the best outside alternative.

[6]Suppose that there are two different integer values N_A and N_B that satisfy the entry equilibrium conditions $\Pi^*_{mt}(N) \geq 0$ and $\Pi^*_{mt}(N + 1) < 0$. Without loss of generality, suppose that $N_B > N_A$. Since $N_B \geq N_A + 1$, strict monotonicity of Π^* implies that $\Pi^*(N_B) \leq \Pi^*(N_A + 1) < 0$. But $\Pi^*(N_B) < 0$ contradicts the equilibrium condition for N_B.

The equilibrium number of firms is the largest integer that is smaller than N_{mt}^*. We represent this relationship as $N_{mt} = int(N_{mt}^*)$ where $int(.)$ is the integer function, i.e., largest integer that is smaller or equal than the argument. This expression shows that the number of active firms increases with demand and declines with marginal and fixed costs.

Figure 1.1: Cournot equilibrium profit as function of number of firms
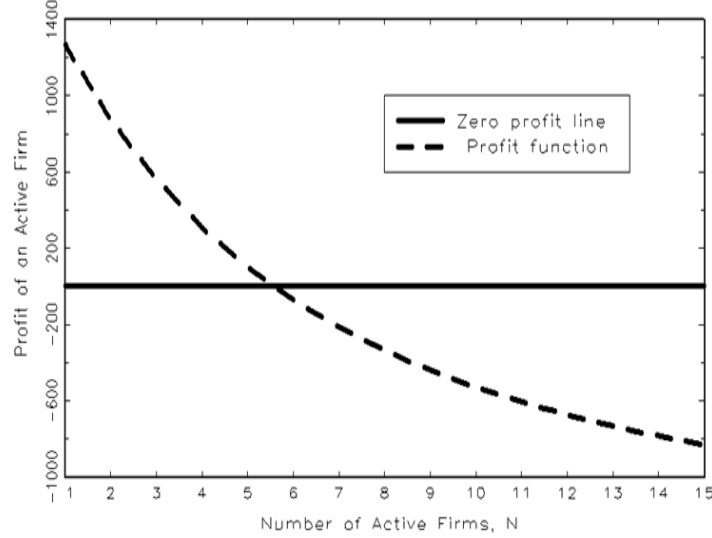
Given the formulas for the equilibrium output per firm (equation 3.9) and profit (equation 3.11), we can obtain the following expression for the Cournot equilibrium profit: $\Pi_{mt}^*(N) = (B_{mt} + \gamma_2^{MC}/2) \, q_{mt}(N)^2 - FC_{mt}$. Therefore, the entry equilibrium condition, represented as $\Pi_{mt}^*(N_{mt}^*) = 0$, is equivalent to:

$$\left(\frac{Q_{mt}}{N_{mt}}\right)^2 = \left(\frac{N_{mt}^*}{int(N_{mt}^*)}\right)^2 \frac{FC_{mt}}{B_{mt} + \gamma_2^{MC}/2} \tag{3.13}$$

For the sake of interpretation, we can treat $N_{mt}^*/int(N_{mt}^*)$ as a constant approximately equal to one.[7] This expression shows how taking into account the endogenous determination of the number of firms in a market has important implications on firm size (output per firm). Firm size increases with fixed costs and declines with the slope of the demand curve, and with the degree of increasing marginal costs. Industries with large fixed costs, inelastic demand curves, and rapidly increasing marginal costs, have larger firms and small number of them. In the extreme case, we can have a *natural monopoly*. The opposite case, in terms of market structure, is an industry with small fixed costs, very elastic demand, and constant marginal costs. An industry with these exogenous demand and cost characteristics will have an atomist market structure with many and very small firms. It is clear that exogenous demand and cost determine in the equilibrium of the industry market structure and market power.

---

[7]For $N_{mt}^* > 1$, the ratio $N_{mt}^*/int(N_{mt}^*)$ lies always within the interval $(1, 2)$.

3.5. Structural equations, equilibrium, and reduced form equations. For simplicity, in some of the discussions in this chapter, we treat the number of firms N_{mt} as a continuous variable: $N_{mt} \equiv int(N_{mt}^*) = N_{mt}^*$. Then, we can replace the two inequalities $\Pi_{mt}^*(N_{mt}) \geq 0$ and $\Pi_{mt}^*(N_{mt} + 1) < 0$ by the equality condition $\Pi_{mt}^*(N_{mt}) = 0$. This approximation is not necessarily innocuous, and we do not use it later in the book. For the moment, we keep it because it provides simple, linear in parameters, expressions for the equilibrium values of endogenous variables, and this simplifies our analysis of identification and estimation of the model in this introductory chapter. In this subsection, we omit the market and time subindexes.

The model can be described as a system of three equations: the demand equation; the Cournot equilibrium condition; and the entry equilibrium condition. The system has three endogenous variables: the number of firms in the market, N; the market price, P; and output per-firm, $q \equiv Q/N$,

$$\textbf{Demand equation:} \quad P \;=\; A - B\,N\,q$$

$$\textbf{Cournot Equilibrium Condition:} \quad q \;=\; \frac{A - \overline{MC}}{B\,(N+1) + \gamma_2^{MC}} \tag{3.14}$$

$$\textbf{Entry Equilibrium Condition:} \quad q^2 \;=\; \frac{FC}{B + \gamma_2^{MC}/2}$$

This is a system of simultaneous equations. The system of equations in (3.14) is denoted as the *structural equations* of the model. Given a value of the exogenous variables, $\mathbf{X}$ and $\boldsymbol{\varepsilon} \equiv (\varepsilon^D,\ \varepsilon^{MC},\ \varepsilon^{FC})$, and of the structural parameters, $\boldsymbol{\theta} \equiv \{\boldsymbol{\beta}_X,\ \beta_P,\ \boldsymbol{\gamma}_X^{MC},\ \gamma_2^{MC},\ \boldsymbol{\gamma}^{FC}\}$, an *equilibrium* of the model if a value of the vector of endogenous variables $\{N,\ P,\ q\}$ that solves this system of equations.

In this model, we can show that an equilibrium always exists and it is unique. To show this, notice that the entry equilibrium condition determines output per firm as a function of the exogenous variables.

$$q = \sqrt{\frac{FC}{B + \gamma_2^{MC}/2}} \tag{3.15}$$

This expression provides the equilibrium value for output per-firm. Plugging this expression for q into the Cournot equilibrium condition and solving for N, we can obtain the equilibrium value for the number of firms as:

$$N = -\left(1 + \frac{\gamma_2^{MC}}{B}\right) + \left(A - \overline{MC}\right)\sqrt{\frac{1 + \gamma_2^{MC}/2B}{FC\,B}} \tag{3.16}$$

Finally, plugging the equilibrium expressions for N and q into the demand equation, we can obtain the equilibrium price as:

$$P \;=\; \overline{MC} + (\gamma_2^{MC} + B)\sqrt{\frac{FC}{B + \gamma_2^{MC}/2}} \tag{3.17}$$

Equations (3.15), (3.16), and (3.17) present the equilibrium values of the endogenous variables as functions of exogenous variables and parameters only. These three equations are called the *reduced form equations* of the model. In this model, where the equilibrium is always unique, the reduced form equations are functions. More generally, in models with multiple equilibria, reduced form equations are correspondences such that for a given value of the exogenous variables there are multiple values of the vector of endogenous variables, each value representing a different equlibria.

4. Identification and estimation

Suppose that the researcher has access to a panel dataset that follows M local markets over T quarters. For every market-quarter the dataset includes information on market price, aggregate output, number of firms, and some exogenous market characteristics such as population, average household income, and input prices: $\{P_{mt}, Q_{mt}, N_{mt}, \mathbf{X}_{mt}\}$. The researcher wants to use these data and the model described above to learn about different aspects of competition in this industry and to evaluate the effects of the policy change described above. Before we study the identification and estimation of the structural parameters of the model, it is interesting to examine some empirical predictions of the model that can be derived from the reduced form equations.

4.1. Empirical evidence from reduced form equations. From an empirical point of view, the reduced form equations establish relationships between exogenous market characteristics, such as market size, and the observable endogenous variables of the model: price, number of firms, and firm size. Can we learn about competition in this industry, and about some of the structural parameters, by estimating the reduced form equations? As we show below, there is very important evidence that can be obtained from the estimation of these equations. However, to provide answers to some other questions, and in particular our policy question, requires the estimation of the structural model.

4.1.1. *Relationship between market size and firm size.* The reduced form equation for output-per-firm in (3.15), implies the following relationship between firm size (or output per firm) q and market size S:

$$\ln(q) = \frac{1}{2}\left[\ln(\beta_P FC) + \ln(S) - \ln\left(1 + 0.5\beta_P \gamma_2^{MC} S\right)\right] \tag{4.1}$$

We can distinguish three different possible cases for this relationship. When fixed cost is zero ($FC = 0$) there is not any relationship between firm size and market size. The model becomes one of perfect competition and the equilibrium is characterized by a very large number of firms ($N = \infty$) each with an atomistic size ($q = 0$). When the fixed cost is strictly positive ($FC > 0$) there is a positive relationship between market size and firm size. Markets with more product demand have larger firms. We can distinguish two different cases when the fixed cost is strictly positive: (a) constant marginal cost, $\gamma_2^{MC} = 0$; and (b) increasing marginal cost, $\gamma_2^{MC} > 0$. With constant marginal cost, the relationship between firm size and market size is $\ln(q) = \frac{1}{2}\left[\ln\left(\beta_P FC\right) + \ln\left(S\right)\right]$ such that firm size always increases proportionally with market size. When the marginal cost is increasing ($\gamma_2^{MC} > 0$), the limit of firm size when market size goes to infinite is equal to $\sqrt{2FC/\gamma_2^{MC}}$, and this constant represents the maximum size of a firm in the industry. Note that $\sqrt{2FC/\gamma_2^{MC}}$ is the level of output-per-firm that minimizes the Average Total Cost, i.e., the *Minimum Efficient Scale* (MES). Figure 1.2 illustrates these two cases for the relationship between firm size and market size. The values of the parameters that generate these curves are: $FC = 1,400$, $\beta_P = 1$, $\gamma_2^{MC} = 1$ (when positive), and values of market size S between 0.1 and 20.

Figure 1.2: Relationship between firm size and market size
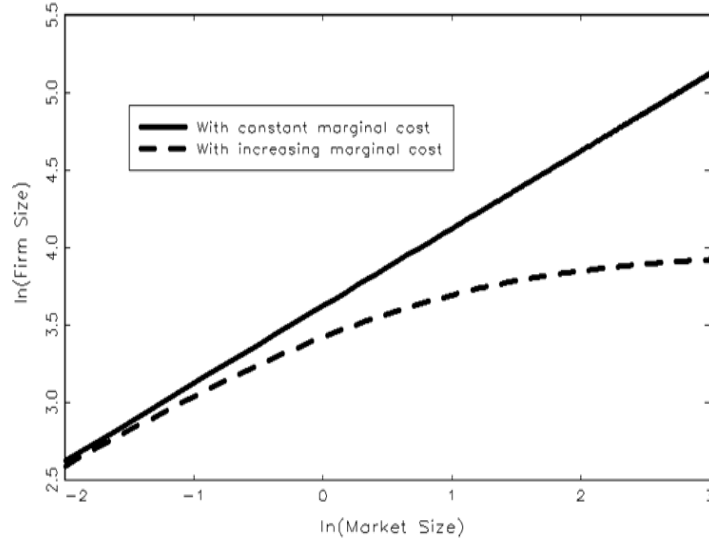
Equation (4.1) and figure 1.2 show that the shape of the relationship between market size and firm size reveals information on the relative magnitude of the fixed cost and the convexity of the variable cost. Given a cross-section of local markets in an homogenous product industry, the representation of the scatterplot of sample points of (S_{mt}, q_{mt}) in the plane, and the estimation of a nonlinear (or nonparametric) regression of q_{mt} on S_{mt} provides empirical evidence on this aspect of cost structure. Campbell and Hopenhayn (2005) look

at this empirical relationship in thirteen retail industries using a sample of 225 US cities. Figure 1.3 presents the scatterplot and the estimation regression line for the logarithm of firm size on the logarithm of market size in Women Clothing retail industry. In this example, the relationship in logarithms is linear and this is consistent with $FC > 0$ and $\gamma_2^{MC} = 0$ for this industry. In logarithms, for small γ_2^{MC}, we have that $\ln(q_{mt}) = \alpha_0 + \alpha_1 \ln(S_{mt}) + \alpha_2 S_{mt} + u_{mt}$, where $\alpha_1 \equiv 1/2$, and $\alpha_2 \equiv -\beta_1 \gamma_2^{MC}/2$. Therefore, testing the null hypothesis $\alpha_2 = 0$ is equivalent to test for non-convexity in the variable cost, i.e., $\gamma_2^{MC} = 0$. Note that market size is measured with error and this creates an endogeneity problem in the estimation of this relationship. Campbell and Hopenhayn take into account this issue and try to correct for endogeneity bias using Instrumental Variables.

Figure 1.3: 'Market size matters' (Campbell & Hopenhayn, 2005)
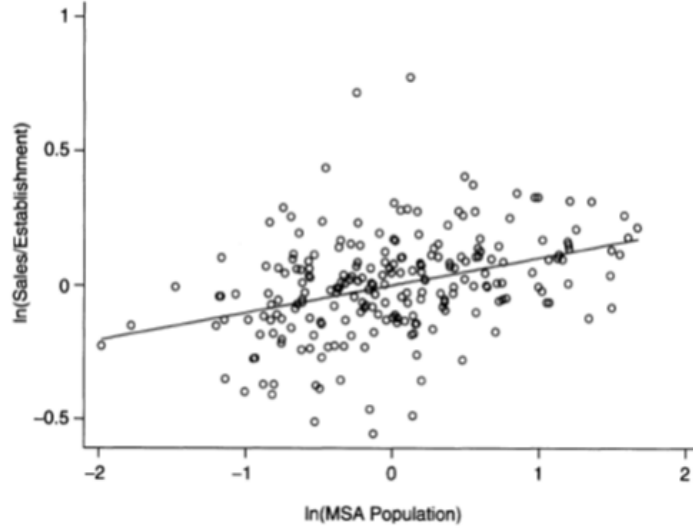
This testable prediction on the relationship between market size and firm size is not shared by other models of firm competition such as models of monopolistic competition or models of perfect competition, where market structure, market power, and firm size do not depend on market size. In all the industries studied by Campbell and Hopenhayn, this type of evidence is at odds with models of monopolistic and perfect competition.

4.1.2. *Relationship between market size and price.* Are prices higher in small or in larger markets? This is an interesting empirical question per se. The model shows that the relationship between price and market size can reveal some interesting information about competition in an industry. We can distinguish three cases depending on the values of FC and γ_2^{MC}. If the industry is such that the fixed cost is zero or negligible, then the model predicts that there should not be any relationship between market size and price. In fact, price should be always equal to the minimum marginal cost, $\overline{MC}_{mt}$. When the fixed cost is strictly positive and the variable cost is linear in output, the reduced form equation for

price becomes $P = \overline{MC} + \sqrt{FC/\beta_1 S^*}$. In this case, an increase in market size always has a negative effect on price, though the marginal effect is decreasing. When market size goes to infinity, price converges to the minimum marginal cost $\overline{MC}$. This is also the relationship that we have between market size and price when the variable cost function is strictly convex, with the only difference that now as market size goes to infinity the price converges to $\overline{MC} + \sqrt{2\gamma_2^{MC} FC}$, which is the marginal cost when output-per-firm is at the *Minimum Efficient Scale* (MES).

———————————
FIGURE 1.4

Price vs. Market size: reduced form equation

———————————

As in the case of firm size, we can use cross-sectional data on prices and market size to test for the relationship between these variables. Finding a significant negative effect of market size on price implies the rejection of monopolistic and perfect competition models in favor of oligopoly competition.

4.1.3. ***Policy Question and Reduced Form Equations.*** Remember that we want to evaluate the effects of a policy that generates an increase in the fixed cost and a reduction in the marginal cost. What do the reduce form equations say about the effects of this policy? Could the estimation of the reduced form equation provide enough information to answer our policy questions?

Equation (4.1) shows that an increase in the fixed cost FC and a reduction in the marginal cost parameter γ_2^{MC} imply a larger firm size. Therefore, the model predicts that the new policy will transform the industry in one with larger firms. However, without further information about the values of the parameters of the model, the reduced form equations do not provide a prediction about the effects on the number of firms, aggregate output, price, and consumer welfare. Not only the magnitude but even the sign of these effects depend on the values of the structural parameters. A larger fixed cost reduces the number of firms and aggregate output, increases price, and it has a negative effect on consumer welfare. A reduction in the marginal cost has exactly the opposite effects, in terms of sign, on all the endogenous variables. The net effects are ambiguous and they depend on the values of the demand and cost parameters and on the magnitude of the change in fixed cost and marginal cost.

Interestingly, the sign of the effect of the policy on number of firms, output, prices, and consumer welfare depends on market size. The effect of a reduction in marginal cost is quantitatively more important in large markets than in small ones. Therefore, in large markets this positive effect dominates the negative effect of the increase in the fixed costs.

We may have that in large markets the policy increases the number of firms, reduces prices, and increases consumer welfare, and the effects on small markets are just the opposite. The welfare effects of this policy are not neutral with respect to market size.

It is relevant to distinguish two cases or scenarios in terms of the information for the researcher about the policy change. In the first case, that we denote as a *factual policy change*, the sample includes observations both before and after the policy change. The second case represents a *counterfactual policy change*, and the data contains only observations without the new policy. The distinction is relevant because the identification assumptions are different in each. In the case of a factual policy change, and under some conditions, we may need only the identification of the parameters in the reduced form equations. Identification of reduced form parameters require weaker assumptions than identification of structural parameters.

Many empirical questions in IO deal with predicting the effects of changes that have not yet occurred. When an industry regulator recommends to approve or not a merger between two companies, he has to predict the effects of a merger that has not yet taken place. Similarly, a company that decides whether to introduce or not a new product in a market, or that designs the features of that new product, needs to predict the demand for that hypothetical product before it has been introduced in the market. In our example here, we consider first the case that the regulator has not implemented the new environmental regulation yet, and he wants to predicts the effects of this regulation. To evaluate the effects of our policy change in a counterfactual setting we make use of our structural model and a two step approach. First, we use our data to estimate the structural parameters of the model. And second, we use the estimated model to predict the responses to changes in some parameters or/and exogenous variables implied by the counterfactual policy change, under the assumption that the rest of the parameters remain constant. We now turn to problem of identification of the structural parameters.[8]

4.2. Estimation of structural parameters. The researcher wants to use the available data to estimate the vector of structural parameters $\boldsymbol{\theta} = \{\beta_0, \ \beta_1, \ \beta_S, \ \gamma_1^{MC}, \ \gamma_2^{MC}, \ \gamma^{FC}\}$. Given an estimate of the true $\boldsymbol{\theta}$, we can use our model to evaluate/predict the effects of and hypothetical change in the cost parameters γ_1^{MC}, γ_2^{MC}, and γ^{FC} implied by the policy. For simplicity, we start considering a version of the model without measurement error in the observable measure of market size, i.e., $\exp\{\varepsilon_{mt}^S\} = 1$ for every market and period (m, t).

---

[8]Sometimes, for some counterfactual policy questions we need to know only some of the structural parameters. This idea goes back at least to the origins of the Cowles Foundation in the 1950s, and more specifically to the work of Jacob Marschak (1953), and it has been exploited recently in different studies. See also Chetty (2009) and Aguirregabiria (2010).

The econometric model can be represented using the following system of simultaneous equations:

$$\frac{Q_{mt}}{S_{mt}} = \beta_X \mathbf{X}^D_{mt} - \beta_1 P_{mt} + \varepsilon^D_{mt}$$

$$\left(P_{mt} - \frac{1}{\beta_1} \frac{q_{mt}}{S_{mt}} \right) = \gamma^{MC}_X \mathbf{X}^{MC}_{mt} + \gamma^{MC}_2 q_{mt} + \varepsilon^{MC}_{mt} \qquad (4.2)$$

$$\frac{q^2_{mt}}{S_{mt}} + \beta_1 \gamma^{MC}_2 q_{mt} = \gamma^{FC}_X \mathbf{X}^{FC}_{mt} + \varepsilon^{FC}_{mt}$$

We complete the econometric model with an assumption about the distribution of the unobservables. It is standard to assume that the unobservables ε_{mt} are mean independent of the observable exogenous variables.

Assumption 1: The vector of unobservable variables in the structural model, $\boldsymbol{\varepsilon}_{mt}$, is mean independent of S_{mt}: $\mathbb{E}(\boldsymbol{\varepsilon}_{mt}|S_{mt}) = 0$.

We say the parameters of the model are identify if there is a feasible estimator of $\boldsymbol{\theta}$ that is *consistent* in a statistical or econometric sense.[9] A standard approach to prove that the vector of parameters is identified consists in using the moment moment restrictions implied by the model to show that we can **uniquely** determine the value of $\boldsymbol{\theta}$ as a function of moments that include only observable variables. For instance, in a classical linear regression model $Y = X\beta + \varepsilon$ under the assumptions of no correlation between the error term and the regressors (i.e., $\mathbb{E}(X\ \varepsilon) = 0$) and not perfect collinearity between the regressors (i.e., $\mathbb{E}(X\ X')$ is non-singular), we have that $\beta = \mathbb{E}(X\ X')^{-1}\ \mathbb{E}(X\ Y)$ and this expression shows that the vector of parameters β is identified using data of Y and X. In our model, Assumption 1, provides moment restrictions, but we show below that these restrictions are not sufficient to identify the parameters of the model.

4.2.1. *Endogeneity.* The key identification problem in our model is that the regressors in the three equations are endogenous variables that are correlated with the unobservables or

---

[9]Given our sample with large M and small T, and an estimator $\widehat{\boldsymbol{\theta}}_M$ we say that $\widehat{\boldsymbol{\theta}}_M$ is a consistent estimator of the true value $\boldsymbol{\theta}$ if $\widehat{\boldsymbol{\theta}}_M$ converges in probability to $\boldsymbol{\theta}$ as the sample size M goes to infinity: $p\lim_{M\to\infty} \widehat{\boldsymbol{\theta}}_M = \boldsymbol{\theta}$, or using the definition of the limit in probability operator: for any scalar $\delta > 0$,

$$\lim_{M\to\infty} \Pr\left(\left\|\widehat{\boldsymbol{\theta}}_M - \boldsymbol{\theta}\right\| > \delta \right) = 0$$

A sufficient condition for the consistency of the estimator $\widehat{\boldsymbol{\theta}}_M$ is that the bias and variance of the estimator ($\mathbb{E}(\widehat{\boldsymbol{\theta}}_M - \boldsymbol{\theta})$ and $Var(\widehat{\boldsymbol{\theta}}_M)$) converge to zero as M goes to infinity.

error terms. It is convenient to write the system of equations as:

$$\frac{Q_{mt}}{S_{mt}} = \beta_0 + \beta_S\ S_{mt} - \beta_1\ P_{mt} + \varepsilon_{mt}^D$$

$$\left[P_{mt} - \frac{1}{\beta_1}\frac{q_{mt}}{S_{mt}}\right] = \gamma_1^{MC} + \gamma_2^{MC}\ q_{mt} + \varepsilon_{mt}^{MC} \tag{4.3}$$

$$\left[\frac{1}{\beta_1}\frac{q_{mt}^2}{S_{mt}} + \gamma_2^{MC}q_{mt}^2\right] = \gamma^{FC} + \varepsilon_{mt}^{FC}$$

In the second equation, the left-hand-side is the price minus the price-cost-margin and this should be equal to the marginal cost in the right-hand-side. In the third equation, the left-hand-side is total profit minus variable profit, and this should be equal to the fixed cost in the right-hand-side.

Given this representation of the system of equations, it is clear that we can follow a sequential approach to identify and estimate the model. First, we consider the identification of demand parameters. Given identification of the demand slope parameter β_1, the variable in right-hand-side of the Cournot equilibrium equation is known, and we consider the identification of parameters in the variable cost. Finally, given β_1 and γ_2^{MC} the variable in the right-hand-side of entry-equilibrium equation is known and therefore the identification of the fixed cost parameter follows trivially from the moment condition $\mathbb{E}(\varepsilon_{mt}^{FC}) = 0$. Following this sequential approach, it should be clear that there are two endogeneity or identification problems: (1) in the demand equation, price is an endogenous regressor, i.e., $\mathbb{E}(P_{mt}\ \varepsilon_{mt}^D) \neq 0$; and (2) in the Cournot equilibrium equation, output per firm is an endogenous regressor, i.e., $\mathbb{E}(q_{mt}\ \varepsilon_{mt}^{MC}) \neq 0$.

How can we deal with this endogeneity problem? There is not such a thing as "the" method or approach to deal with endogeneity problems. There are different approaches, each with their relative advantages and limitations. These approaches are based on different assumptions that may be more or less plausible depending on the application. The advantages and plausibility of an approach should be judged in the context of an specific application.

We now use our simple model to illustrate some of the identification assumptions and strategies that have been used in many applications in empirical IO and that we will see throughout this book: (a) randomized experiments; (b) exclusion restrictions; (c) "natural experiments" as exclusion restrictions; and (d) restrictions on the covariance structure of the unobservables.

4.2.2. *Randomized experiments.* The implementation of an adequate randomized experiment is an ideal situation for the identification of an econometric model. The careful design of a useful randomized experiment is not a trivial problem. We illustrate some of the issues

in the context of our model. We also want to emphasize here that the structural model is a useful tool in the design of the randomized experiment.

Suppose that we want to estimate first the demand equation. We need to design an experiment that generates sample variation in price that is not perfectly correlated with market size and it is independent of the unobserved demand shock ε_{mt}^D. Suppose that experiment consists in a firm subsidy per unit of output produced and sold in the market. In marker-quarter (m, t) this subsidy is of τ_{mt} dollars per unit of output, and τ_{mt} is randomly distributed over (m, t) and independently distributed of any market characteristic, e.g., it is determined as random draw from some distribution. We need also to assume that the implementation of the experiment does not introduce any change in the behavior of consumers. Under this condition, we have that the following conditions hold: the subsidy is not correlated with the demand shock $\mathbb{E}(\tau_{mt} \, \varepsilon_{mt}^D) = 0$; the subsidy is not correlated with market size $\mathbb{E}(\tau_{mt} \, S_{mt}) = 0$; and the subsidy is correlated with price $\mathbb{E}(\tau_{mt} \, P_{mt}) = 0$. These conditions imply that we can use the amount of subsidy in a market τ_{mt} as an instrument for the P_{mt} in the demand equation, to identify all the parameters in the demand. More precisely, the moment conditions $\mathbb{E}(\varepsilon_{mt}^D) = 0$, $\mathbb{E}(S_{mt} \, \varepsilon_{mt}^D) = 0$, and $\mathbb{E}(\tau_{mt} \, \varepsilon_{mt}^D) = 0$ identify the parameters β_0, β_S, and β_1 in the demand equation. Given the estimated demand parameters, we can use also the moment conditions $\mathbb{E}(\varepsilon_{mt}^{MC}) = 0$, $\mathbb{E}(S_{mt} \, \varepsilon_{mt}^{MC}) = 0$, and $\mathbb{E}(\tau_{mt} \, \varepsilon_{mt}^{MC}) = 0$ to identify variable cost parameters in the Cournot equation, and the moment conditions $\mathbb{E}(\varepsilon_{mt}^{FC}) = 0$, $\mathbb{E}(S_{mt} \, \varepsilon_{mt}^{FC}) = 0$, and $\mathbb{E}(\tau_{mt} \, \varepsilon_{mt}^{FC}) = 0$ to identify the fixed cost parameter in the entry equation.

A well known concern in any experiment, either in the lab or in the field, is that agents behavior may change if they know that they are the subjects of an experiment. In the experiment that we have here, that is a potential concern for the behavior of firms. Firms' involve in the experiment may change the way they compete during the time that experiment is implemented. For instance, they may decide to agree not to change their levels of output such that the subsidy will not be pass-through to the price and they will keep the subsidy as a pure transfer. However, as long as the subsidy has some effect price (i.e., there is at least a partial pass-through of the subsidy to price), this concern does not affect the identification of the demand parameters. What is most important in this experiment is that consumers are not aware of this experiment and therefore they do not change their demand behavior. In contrast, if some consumers are aware of the existence of this experiment, and given the temporary nature of the experiment, they may decide to buy cement for inventory. In that case, the experiment will affect the demand and the estimates of the demand parameters based on this randomized experiment will be biased.

4.2.3. *Exclusion restrictions (Instrumental Variables).* In econometrics, and in empirical micro fields in particular, the most common approach to deal with endogeneity problems is using instrumental variables. An instrumental variable is an observable variable that satisfies three restrictions in the equation we want to estimate: (i) it does not appear explicitly in the equation; (ii) it is correlated with the endogenous regressor(s); and (iii) it is not correlated with the error term (unobservables) of the equation. In the context of our model, for the estimation of demand parameters we need a variable that is not included in the demand equation, is not correlated with the demand shock, and is correlated with price.

According to our model, input prices are a variables that may satisfy these conditions. For instance, limestone and coal are two important variable inputs in the production of cement. The prices of limestone and coal are potential instruments because the affect marginal cost, they should be correlated with price, but they do not enter in the demand equation. What it is not so obvious is whether these variables are not correlated with the unobserved demand shock. If the demand of coal and limestone from the cement industry represents a small fraction of the total demand of these inputs in the local market, it seems plausible to argue that shocks in the demand of cement may not be correlated with the price of these inputs. However, if the cement industry represents 90% of the demand of limestone in a local market, this independence assumption seems completely implausible.

4.2.4. *'Natural experiments' as exclusion restrictions.* Consider an unexpected natural shock that affected the production cost of some markets in a particular period of time. Let I_{mt} be the indicator of the event "affected by the natural experiment". This variable is zero for every market before period t^* when the natural event occurred; it is always zero for markets that do not experience the event, i.e., the control group; and it goes from zero to one for markets in the experimental group. Since there are good reasons to believe that the natural event affected costs, it is clear that price depends of the dummy variable I_{mt}. The key identification assumption to use I_{mt} as an instrument for price is that the natural event did not affected demand. Under this assumption, the moment condition $\mathbb{E}(I_{mt}\ \varepsilon_{mt}^D) = 0$, together with the conditions $\mathbb{E}(\varepsilon_{mt}^D) = 0$ and $\mathbb{E}(S_{mt}\ \varepsilon_{mt}^D) = 0$, identify the demand parameters.

The assumption that the natural event did not affect the demand is a strong assumption. Though the natural event is completely exogenous and unexpected, there is no reason why it may have occurred in markets that have relatively high (or low) levels of demand, or it took place during a period of high (or low) demand. In contrast to the case of the randomized experiment described above, where by the own design of the experiment the subsidy was not correlated with the demand shock, there is nothing in the natural experiment implying that $\mathbb{E}(I_{mt}\ \varepsilon_{mt}^D) = 0$. To try to deal with this issue, most applications exploiting identification

from 'natural experiments' assume a particular structure for the unobserved error.

$$\varepsilon_{mt}^D = \omega_m^D + \delta_t^D + u_{mt}^D, \tag{4.4}$$

We can control for ω_m^D using market dummies, and for δ_t using time dummies. The 'natural experiment' dummy I_{mt} can be correlated with ω_m^D and/or with δ_t^D. The identification assumption is that I_{mt} is not correlated with the shock u_{mt}^D.

4.2.5. *Restrictions on Covariance-Structure of Unobservables.* Suppose that the unobservables in the demand and in the marginal cost have the covariance structure:

$$\begin{aligned} \varepsilon_{mt}^D &= \omega_m^D + \delta_t^D + u_{mt}^D, \\ \varepsilon_{mt}^{MC} &= \omega_m^{MC} + \delta_t^{MC} + u_{mt}^{MC} \end{aligned} \tag{4.5}$$

This components of variance specification of the unobservables, together with restrictions on the serial or/and the spatial correlation of the demand shocks u_{mt}^D, have been exploited to obtain exclusion restrictions and instrumental variables estimators. We distinguish two cases depending on whether the restrictions are on the serial correlation of the shock (i.e., Arellano-Bond Instruments; Arellano and Bond, 1991), or on the spatial correlation (i.e., Hausman-Nevo Instruments; Hausman, 1997, and Nevo, 2000).

Arellano-Bond instruments. Suppose that the shock u_{mt}^D is not serially correlated over time. That is, all the time persistence in unobserved demand comes from the time-invariant effect ω_m^D, and from the common industry shocks δ_t^D, but the idiosyncratic demand shock u_{mt}^D is not persistent over time. Under these conditions, in the demand equation in first-differences, $\Delta Q_{mt}/S_{mt} = \beta_S \ \Delta S_{mt} - \beta_1 \ \Delta P_{mt} + \Delta \delta_t^D + \Delta u_{mt}^D$, the lagged endogenous variables $\{P_{mt-2}, Q_{mt-2}, N_{mt-2}\}$ are not correlated with the error Δu_{mt}^D, and the can be used as instruments to estimate demand parameters. The key identification assumption is that the shocks u_{mt}^{MC} in the marginal cost are more persistent that the demand shocks u_{mt}^D.

Hausman-Nevo instruments. Suppose that we can classify the M local markets in R regions. Local markets in the same region may share similar supply of inputs in the production of cement and similar production costs. However, suppose that the demand shock u_{mt}^D is not spatially correlated, such that local markets in the same region have independent demand shocks. All the spatial correlation in demand comes from observables variables, from correlation between the time-invariant components ω_m^D, or from the common shock δ_t^D. Let $\overline{P}_{(-m)t}$ be the average price of cement in markets that belong to the same region as market m but where the average excludes market m. Under these conditions, and after controlling for ω_m^D using market-dummies and for δ_t^D using time-dummies, the average price $\overline{P}_{(-m)t}$ is not correlated with the demand shock u_{mt}^D and it can be used as an instrument to estimate demand parameters. The key identification assumption is that the shocks u_{mt}^{MC} in the marginal have spatial correlation that is not present in demand shocks u_{mt}^D.

Zero covariance between unobservables. In simultaneous equations models, an assumption of zero covariance between the unobservables of two structural equations provides a moment condition that can be used to identified structural parameters. In the context of our model, consider the restrictions $\mathbb{E}(\varepsilon_{mt}^{FC} \, \varepsilon_{mt}^{D}) = 0$ and $\mathbb{E}(\varepsilon_{mt}^{FC} \, \varepsilon_{mt}^{MC}) = 0$. These conditions imply the moment conditions:

$$\mathbb{E}\left(\left[\frac{1}{\beta_1}\frac{q_{mt}^2}{S_{mt}} + \gamma_2^{MC} q_{mt}^2 - \gamma^{FC}\right]\left[\frac{Q_{mt}}{S_{mt}} - \beta_0 - \beta_S S_{mt} + \beta_1 P_{mt}\right]\right) = 0 \qquad (4.6)$$

and

$$\mathbb{E}\left(\left[\frac{1}{\beta_1}\frac{q_{mt}^2}{S_{mt}} + \gamma_2^{MC} q_{mt}^2 - \gamma^{FC}\right]\left[P_{mt} - \frac{1}{\beta_1}\frac{q_{mt}}{S_{mt}} - \gamma_1^{MC} - \gamma_2^{MC} q_{mt}\right]\right) = 0 \qquad (4.7)$$

These moment restrictions, together with those from the restrictions $\mathbb{E}(\varepsilon_{mt}^{D}) = 0$, $\mathbb{E}(\varepsilon_{mt}^{MC}) = 0$, $\mathbb{E}(\varepsilon_{mt}^{FC}) = 0$, $\mathbb{E}(S_{mt} \, \varepsilon_{mt}^{D}) = 0$, $\mathbb{E}(S_{mt} \, \varepsilon_{mt}^{MC}) = 0$, and $\mathbb{E}(S_{mt} \, \varepsilon_{mt}^{FC}) = 0$, identify the structural parameters of the model.

We can consider a weaker version of this assumption: if $\varepsilon_{mt}^{FC} = \omega_m^{FC} + \delta_t^{FC} + u_{mt}^{FC}$ and $\varepsilon_{mt}^{D} = \omega_m^{D} + \delta_t^{D} + u_{mt}^{D}$, we can allow for correlation between the $\omega's$ and $\delta's$ and assume that only the market specific shocks u_{mt}^{FC} and u_{mt}^{D} are not correlated.

4.2.6. *Multiple equilibria and Identification.* Multiplicity of equilibria is common feature in many models in IO. In our example, for any value of the parameters and exogenous variables the equilibrium in the model is unique. There are three assumptions in our simple model that play an important role in generating this strong equilibrium uniqueness: (a) linearity assumptions, i.e., linear demand; (b) homogeneous firms, i.e., homogeneous product and costs; and (c) no dynamics. Once we relax any of these assumptions, multiple equilibria is the rule more than the exception: for some values of the exogenous variables and parameters, the model has multiple equilibria.

Is multiplicity of equilibria an important issue for estimation? It may or may not, depending on the structure of the model and on the estimation method that we choose. We will examine this issue in detail throughout this book, but let us provide here some general ideas about this issue.

Suppose that the fixed cost of operating a plant in the market FC_{mt} is a decreasing function of the number of firms in the local market. For instance, there are positive synergies between firms in terms of attracting skill labor, etc. Then, $FC_{mt} = \gamma^{FC} - \delta \, N_{mt} + \varepsilon_{mt}^{FC}$, where δ is a positive parameter. Then, the equilibrium condition for market entry becomes:

$$\left(\frac{Q_{mt}}{N_{mt}}\right)^2 = \frac{\gamma^{FC} - \delta \, N_{mt} + \varepsilon_{mt}^{FC}}{B_{mt} + \gamma_2^{MC}/2} \qquad (4.8)$$

This equilibrium equation can imply multiple equilibria for the number of firms in the market. The existence of positive synergies in the entry cost introduces some "coordination" aspects in the game of entry (Cooper, 1999). If δ is large enough, this coordination feature can

generate multiple equilibria. Of course, multiplicity in the number of firms also implies multiplicity in the other endogenous variables, price, and output per firm. Therefore, the reduced form equations are now correspondences, instead of functions, that relate exogenous variables and parameters with endogenous variables.

Does this multiplicity of equilibria generates problems for the identification and estimation of the structural parameters of the model? Not necessarily. Note that, in contrast to the case of the reduced form equations, the three structural equations (demand, Cournot equilibrium, and entry condition) still hold with the only difference that we now have the term $-\delta\ N_{mt}$ in the structural equation for the entry equilibrium condition. That is,

$$\left[\frac{1}{\beta_1}\frac{q_{mt}^2}{S_{mt}} + \gamma_2^{MC}q_{mt}^2\right] \;=\; \gamma^{FC} - \delta\ N_{mt} + \varepsilon_{mt}^{FC} \qquad (4.9)$$

The identification of the parameters in demand and variable costs is not affected. Suppose that those parameters are identified such that the left-hand-side in the previous equation is known variable to the researcher. In the right hand side, we now how the number of firms as a regressor. This variable is endogenous and correlated with the error term ε_{mt}^{FC}. However, dealing with the endogeneity of the number of firms for the estimation of the parameters γ^{FC} and δ is an issue that does not have anything to do with multiple equilibria. We have that endogeneity problem whether or not the model has multiple equilibria, and the way of solving that problem does not depend on the existence of multiple equilibria. For instance, if we have valid instruments and estimate this equation using Instrumental Variables (IV), the estimation will be the same regardless of the multiple equilibria in the model.

In fact, more than a problem for identification, multiple equilibria may help for identification in some cases. For instance, if there is multiple equilibria in the data and equilibrium selection is random and independent of ε_{mt}^{FC}, then multiple equilibria helps for identification because it generates additional sample variation in the number of firms that is independent of the error term.

In some models, multiplicity of equilibria can be a nuisance for estimation. Suppose that we want to estimate the model using the maximum likelihood (ML) method. To use the ML method we need to derive the probability distribution of the endogenous variables conditional on the exogenous variables and the parameters of the model. However, in a model with multiple equilibria there is not such a thing as "the" distribution of the endogenous variables. There are multiple distributions, one for each equilibrium type. Therefore, we do not have a likelihood function but a likelihood correspondence. Is the MLE well define in this case? How to compute it? Is it computationally feasible? Are there alternative methods that are computationally simpler? We will address all these questions later in the course.

4.3. Extensions. The rest of the course deals with empirical models of market structure that relax some of these assumptions. This is a list of assumptions that we will relax:

a) Heterogeneity in firms' costs. Exploiting information on firms' inputs to identify rich cost structures. [Chapter 2: Production Function Estimation]

b) Product differentiation and more general forms of demand. [Chapter 3: Demand Estimation]

c) Relaxing the assumption of Cournot competition, and identification of the "nature of competition" from the data (e.g., Cournot, or Bertrand, or Collusion). [Chapter 4: Static models of competition].

d) Heterogeneity in entry costs and oligopoly games of entry. [Chapter 5: Static games of entry].

e) Spatial differentiation and plant spatial location. [Chapter 5: Games of spatial competition].

f) Competition in quality and other product characteristics [Chapter 5: Games of quality competition].

g) Investment in capacity and fixed inputs [Chapters 6 and 7: Dynamic structural models of firm investment decisions].

h) Consumers intertemporal substitution and dynamic demand of storable and durable products [Chapter 8: Dynamic demand].

i) Dynamic strategic interactions in firms' entry-exit and investment decisions. [Chapter 9: Dynamic games].

j) Mergers [Chapter 9: Dynamic games].

k) Firm networks, chains, and competition between networks [Chapter 9: Dynamic games].

5. Exercises

5.1. Exercise 1. Let me recommend you the following exercise. Write a program (in GAUSS, Matlab, STATA, or your favorite software) where you do the following:

- Fix, as constants in your program, the values of the exogenous cost variables MC_{mt}, and FC_{mt}, and of demand parameters β_0 and β_1. Then, consider 100 types of markets according to their firm size. For instance, a vector of market sizes $\{1, 2, ..., 100\}$.

- For each market type/size, obtain equilibrium values of the endogenous variables, and of output per firm, firm's profit, and consumer surplus. For each of these variables, generate a two-way graph with the endogenous variable in vertical axis and market size in the horizontal index.

- Now, consider a policy change that increases fixed cost and reduces marginal cost. Do the same thing as before but with the new parameters. Obtain two-way graphs of each variable against market size including both the curve before and after the policy change.

5.2. Exercise 2. Write a program (in GAUSS, Matlab, STATA, or your favorite software) where you do the following:

- Fix, as constants in the program, the number of markets, M, time periods in the sample, T, and the values of structural parameters, including the parameters in the distribution of the unobservables and the market size. For instance, you could assume that the four unobservables ε have a join normal distribution with zero mean and a variance-covariance matrix, and that market size is independent of these unobservables and it has a log normal distribution with some mean and variance parameters.

- Generate NT random draws from the distribution of the exogenous variables. For each draw of the exogenous variables, obtain the equilibrium values of the endogenous variables. Now, you have generated a panel dataset for $\{P_{mt}, Q_{mt}, N_{mt}, S_{mt}\}$

- Use this data to estimate the model by OLS, and also try some of the approaches described above to identify the parameters of the model.

5.3. Exercise 3. The purpose of this exercise is to use the estimated model (or the true model) from exercise #2 to evaluate the contribution of different factors to explain the cross-sectional dispersion of endogenous variables such as prices, firm size, or number of firms. Write a program (in GAUSS, Matlab, STATA, or your favorite software) where you do the following:

- For a particular year of your panel dataset, generate figures for the empirical distribution of the endogenous variables, say price.

- Now, we obtain four counterfactual distributions: (i) the distribution of prices if there were not heterogeneity in market size: set all market sizes equal to the one in the median market, and obtain the counterfactual equilibrium price for each market; (ii) the distribution of prices if there were not market heterogeneity in demand shocks: set all demand shocks equal to zero, and obtain the counterfactual equilibrium price for each market; (iii) the distribution of prices if there were not market heterogeneity in marginal costs; and (iv) the distribution of prices if there were not market heterogeneity in fixed costs. Generate figures of each of these counterfactual distributions together with the factual distribution.

Bibliography

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

[2] Aguirregabiria, V., (2010): "Another look at the identification of dynamic discrete decision processes: An application to retirement behavior," Journal of Business and Economic Statistics.

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

[4] Bain, J. (1951): "Relation of Profit Rate to Industry Concentration: American Manufacturing, 1936–1940," The Quarterly Journal of Economics, 65 (3), 293-324.

[5] Bain, J. (1954): "Economies of Scale, Concentration, and the Condition of Entry in Twenty Manufacturing Industries," The American Economic Review , 44 (1), 15-39.

[6] Bresnahan, T. (1981): "Departures from Marginal-Cost Pricing in the American Automobile Industry: Estimates for 1977-1978," Journal of Econometrics, 17, 201-227.

[7] Bresnahan, T. (1982): "The Oligopoly Solution Concept is Identified," Economics Letters, 10, 87-92.

[8] Bresnahan, T. (1987): "Competition and Collusion in the American Automobile Market: The 1955 Price War," Journal of Industrial Economics, 35, 457-482.

[9] Campbell, J. and H. Hopenhayn: (2005): "Market size matters," Journal of Industrial Economics ...

[10] Cooper, R (1999): "Coordination Games: Complementarities and Macroeconomics," Cambridge University Press. Cambridge, UK.

[11] Demsetz, H. (1973): "Industry structure, market rivalry, and public policy," Journal of Law and Economics, 16 (1).

[12] Hausman, J. (1996): "Valuation of new goods under perfect and imperfect competition." in "The economics of new goods". University of Chicago Press, 207-248.

[13] Marschak, J. (1953): "Economic measurements for policy and prediction," in *Studies in Econometric Method*, eds. W. Hood and T. Koopmans. New York. Wiley.

[14] Miller, N. and M. Osborne (2013): "Spatial Differentiation and Price Discrimination in the Cement Industry: Evidence from a Structural Model," manuscript. Rotman School of Management. University of Toronto.

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

[16] Porter, R. (1983): "A Study of Cartel Stability: The Joint Executive Committee, 1880-1886," Bell Journal of Economics, 15, 301-314.

[17] Schmalensee, R. (1989): "Inter Industry Studies of Structure and Performance." In Handbook of Industrial Organization, vol. 2, ed. Richard Schmalensee and Robert D. Willig, 951–1010. Amsterdam: North Holland.

[18] Sutton, J. (1991): "Sunk costs and market structure: Price competition, advertising, and the evolution of concentration." MIT press. Cambridge, MA, USA.

CHAPTER 2

Demand Estimation

1. Introduction

The estimation of demand equations is a fundamental component in most empirical applications in IO. It is also important in many other fields in empirical economics. There are important reasons why economists in general, and IO economists in particular, are interested in demand estimation. Knowledge of the demand function, and of the corresponding marginal revenue function, is crucial for the determination of a firm's optimal 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. As such, they are fundamental in the evaluation of the consumer welfare gains or losses associated to taxes, subsidies, the introduction of a new product, or a merger. Demand estimation can be used to improve our measures of Cost-of-Living indices (see Hausman, 2003, and Pakes, 2003).[1]

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

---

[1]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 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.

the rest. A variety is typically produced by a single manufacturer, but a manufacturer may produce several varieties.

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

2. Demand systems in product space

2.1. Model. In this model, consumer preferences are defined over 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 (e.g., food, transportation, clothing, entertainment) and all the specific variety products within each category (e.g., every possible variety of computers, or of automobiles). This means that the number of products J can be of the order of millions. We will see later how, under some conditions, we can separate the demand of different product categories and reduce the dimensionality of this large product space. For this purpose, it is convenient to introduce **"product zero"** that we denote as **"the outside product"** and it represents all the other products which are nor product 1 to J.

Let q_j the quantity that a consumer buys and consumes of product j, and let $(q_1, q_2, ..., q_J)$ be the vector with the purchased quantities of all the products. Let c be the amount of the outside good. The consumer has a utility function $U(c, q_1, q_2, ..., q_J)$ defined over the vector of quantities. The consumer problem consists in choosing the vector $(c, q_1, q_2, ..., q_J)$ to maximize his utility subject to his budget constraint.

$$\max_{\{c, q_1, q_2, ..., q_J\}} U(c, q_1, q_2, ..., q_J)$$

$$\text{subject to}: \quad c + p_1 \, q_1 + p_2 \, q_2 + ... + p_J \, q_J \leq y \tag{2.1}$$

where p_j is the price of product j and y is the consumer's disposable income. Note that the price of the outside good is normalized to one such that c is the expenditure, in dollars, in all the goods other than 1 to J. We can define the Lagrange problem:

$$\max_{\{c, q_1, q_2, ..., q_J\}} U(c, q_1, q_2, ..., q_J) + \lambda \left[y - c - p_1 \, q_1 - ... - p_J \, q_J \right] \tag{2.2}$$

The first order conditions are: $U_j - \lambda \, p_j = 0$ for $j = 1, 2, ..., J$; $U_0 - \lambda = 0$; and $y - c - p_1 \, q_1 - ... - p_J \, q_J = 0$, 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, that give us the optimal quantity of each variety as a

function of prices and income. These are the *Marshallian demand equations*:

$$
\begin{aligned}
q_1 &= f_1\left(p_1, p_2, ..., p_J, y\right) \\
q_2 &= f_2\left(p_1, p_2, ..., p_J, y\right) \\
... & \qquad\qquad ... \\
q_J &= f_J\left(p_1, p_2, ..., p_J, y\right)
\end{aligned}
\tag{2.3}
$$

The form of the functions f_1, f_2, ..., f_J 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. We should be able to derive it as the solution of the consumer problem for a given utility function (**invertibility of a demand system**). A substantial part of the empirical literature on demand is based on finding utility functions which imply demand systems with the following properties: (i) simple enough to be estimable using standard econometric methods such as linear regression; (ii) flexible enough such that it allows for flexible patterns in the elasticities of substitution between products. The following are some examples of models that have been considered in the literature.

(a) The Linear Expenditure System. Consider the Stone-Geary utility function:

$$
U = c\left(q_1 - \gamma_1\right)^{\alpha_1}\left(q_2 - \gamma_2\right)^{\alpha_2}...\left(q_J - \gamma_J\right)^{\alpha_J}
\tag{2.4}
$$

where $\{\alpha_j,\ \gamma_j : j = 1, 2, ..., J\}$ are parameters. This utility function was first proposed by Geary (1950), and Richard Stone (1954) was the first to estimate the Linear Expenditure System. Solving the budget constraint into the utility function, we have that $U = [y - p_1 q_1 - ... - p_J q_J]\left(q_1 - \gamma_1\right)^{\alpha_1} ... \left(q_J - \gamma_J\right)^{\alpha_J}$. The marginal conditions of optimality of the consumer problem are $dU/dq_j = 0$ for every variety j, and this implies:

$$
\alpha_j \frac{U}{q_j - \gamma_j} - p_j \frac{U}{c} = 0
\tag{2.5}
$$

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

$$
c = \frac{y - \sum_{j=1}^{J} p_j\, \gamma_j}{1 + \sum_{j=1}^{J} \alpha_j}
\tag{2.6}
$$

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

$$
q_j = \gamma_j + \alpha_j^* \frac{y}{p_j} - \sum_{i=1}^{J} \alpha_j^* \gamma_i \frac{p_i}{p_j}
\tag{2.7}
$$

where $\alpha_j^* = \alpha_j / [1 + \sum_{i=1}^{J} \alpha_i]$.

This system is convenient for its simplicity. Suppose that we have data of individual purchases and prices over T periods of time ($t = 1, 2, ..., T$): $\{c_t,\ q_{1t},\ q_{2t},\ ...,\ q_{Jt}\}$ and $\{p_{1t},\ p_{2t},\ ...,\ p_{Jt}\}$. The model implies a system of J linear regressions. For product j:

$$
q_{jt} = \gamma_j + \alpha_j^*\, x_{jt} + \xi_{jt}
\tag{2.8}
$$

with $x_{jt} = c_t/p_{jt}$, and ξ_{jt} is an error term that can come, for instance, from measurement error in purchased quantity q_{jt}, or from a 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 $\dfrac{\partial q_j}{\partial p_k} = \dfrac{\alpha_j}{p_j}\dfrac{\partial c}{\partial p_k} = -\alpha_j \gamma_k/p_j \left[1 + \sum_{i=1}^{J} \alpha_i\right] < 0$, such 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.

(b) Constant Elasticity of Substitution (CES) demand system. Consider the CES utility function:

$$U = c \left(\sum_{j=1}^{J} q_j^{\sigma}\right)^{1/\sigma} \tag{2.9}$$

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

$$U_j = q_j^{\sigma-1}\frac{U}{\sum_{i=1}^{J} q_i^{\sigma}} \quad \text{and} \quad U_0 = \frac{U}{c} \tag{2.10}$$

For any two pairs of products, j and k, we have that $\dfrac{\partial^2 U}{\partial q_j \partial q_k} \equiv \dfrac{\partial U_j}{\partial q_k} < 0$, such that all the products (except the outside product) are substitutes in consumption. From the solution of the consumer problem, we have that: $U_0 - \lambda = 0$ implies that $\dfrac{U}{c} - \lambda = 0$, and $c = \dfrac{U}{\lambda}$; and $U_j - \lambda \ p_j = 0$ implies that $q_j^{\sigma-1}\dfrac{U}{\sum_{i=1}^{J} q_i^{\sigma}} = \lambda \ p_j$, and $\dfrac{q_j^{\sigma-1}}{\sum_{i=1}^{J} q_i^{\sigma}}\dfrac{U}{\lambda} = p_j$. Combining these equations, we have the following equations for $j = 1, 2, ..., J$:

$$q_j = \left[\frac{c^*}{p_j}\right]^{\frac{1}{1-\sigma}}$$

with $c^* = c/\sum_{i=1}^{J} q_i^{\sigma}$.

The CES model is also very convenient because its simplicity. Suppose that we have data of individual purchases and prices over T periods of time: $\{c_t, q_{1t}, q_{2t}, ..., q_{Jt}\}$ and $\{p_{1t}, p_{2t}, ..., p_{Jt}\}$. The model implies the following log-linear regression model:

$$\ln(q_{jt}) = \alpha_t + \beta \ \ln(p_{jt}) + \xi_{jt} \tag{2.11}$$

with $\beta = \dfrac{-1}{1-\sigma}$, $\alpha_t = \ln(c_t^*)$, and the error term ξ_{jt} has the same interpretation as above. The demand elasticity β can be estimated from this model and data using a standard method for linear regression models. For instance, if the number of products is large relative to the number of time periods, one can control for the time-effects $\alpha_t's$ using time dummies, and

β can be estimated using OLS or IV methods (see below for a detail discussion on the endogeneity of prices).

This model also imposes strong restrictions. In particular, the elasticity of substitution between any pair of products is exactly the same. For any three products, say j, k, and i: $Elasticity_{k,j} = \frac{\partial \ln q_k}{\partial \ln p_j} = \frac{\partial \ln q_i}{\partial \ln p_j} = Elasticity_{i,j}$. This is very unrealistic.

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

$$U = \left[\prod_{j=1}^{J} q_j^{\alpha_j} \right] + \sum_{j=1}^{J} \sum_{k=1}^{J} \delta_{jk} \, q_j \, q_k \tag{2.12}$$

where $\alpha's$ and $\delta's$ are parameters. $\alpha's$ are positive, but $\delta's$ can be positive or negative. This utility allows for complementarity and substitutability between products, and for a flexible pattern of substitution between different products. The marginal utilities are:

$$U_j = \alpha_j \frac{U^*}{q_j} + \sum_{k=1}^{J} \delta_{jk} \, q_k \quad \text{and} \quad U_0 = \frac{U^*}{c} \tag{2.13}$$

with $U^* \equiv c \prod_{i=1}^{J} (q_i - \gamma_i)^{\alpha_i}$. For any pair of products, j and k, we have that:

$$U_{jk} = \alpha_j \alpha_k \frac{U^*}{(q_j - \gamma_j)(q_k - \gamma_k)} + \delta_{jk} \gtreqless 0 \tag{2.14}$$

that can be positive (complements) or negative (substitutes) and take different values for each pair of products. Solving the consumer problem, we have that: $U_0 - \lambda = 0$ implies that $\frac{U^*}{c} - \lambda = 0$, and $c = \frac{U^*}{\lambda}$; and $U_j - \lambda \, p_j = 0$ implies that $\alpha_j \frac{U^*}{q_j} + \sum_{k=1}^{J} \delta_{jk} \, q_k - \lambda p_j = 0$. Combining these equations, we can derive the following system of Marshallian demand equations:

$$w_j = \beta_j^{(0)} + \beta_j^{(y)} \left[\ln(y) - \ln(P) \right] + \sum_{k=1}^{J} \beta_{jk}^{(p)} \, \ln(p_k) \tag{2.15}$$

where $w_j \equiv p_j q_j / y$ is the expenditure share of product j; $\{\beta_j^{(0)}, \beta_j^{(y)}, \beta_j^{(p)}\}$ are parameters which are known functions of the utility parameters $\{\alpha_j, \delta_{jk}\}$; and P is a price index that is defined as $\ln(P) \equiv \sum_{j=1}^{J} \alpha_j \ln(p_j) + \frac{1}{2} \sum_{j=1}^{J} \sum_{k=1}^{J} \beta_{jk} \ln(p_j) \ln(p_k)$. The model implies the symmetry conditions $\beta_{jk} = \beta_{kj}$. Therefore, the number of free parameters is: $2J + \frac{J(J+1)}{2}$, that increases quadratically with the number of products.

Suppose that we have data on individual purchases, income, and prices over T periods of time: $\{c_t, q_{1t}, q_{2t}, ..., q_{Jt}\}$, y_t, and $\{p_{1t}, p_{2t}, ..., p_{Jt}\}$. For each product j, we can estimate the regression equation:

$$w_{jt} = \beta_j^{(0)} + \beta_j^{(y)} \, \ln(y_t) + \beta_{j1}^{(p)} \, \ln(p_{1t}) + ... + \beta_{jJ}^{(p)} \, \ln(p_{Jt}) + \xi_{jt} \tag{2.16}$$

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 buy using data from many consumers does not help in the estimation of price elasticities because consumers in the same market face the same prices, i.e., prices do not have variation across consumers, only over time and geographic markets.

[NOTE: Discuss the aggregation of the AIDS over consumers].

(d) Multi-stage budgeting using the Almost Ideal demand system. To reduce the number of parameters when J is relatively large, Deaton and Muellbauer propose using a multi-stage budgeting approach. Suppose that the utility function is separable in the utility from G groups of products:

$$U = v_1 \left(\widetilde{\mathbf{q}}_1 \right) + v_2 \left(\widetilde{\mathbf{q}}_2 \right) + ... + v_G \left(\widetilde{\mathbf{q}}_G \right) \tag{2.17}$$

where $\widetilde{\mathbf{q}}_g$ is the vector of quantities of products varieties in group g; and $v_g \left(\widetilde{\mathbf{q}}_g \right)$ is the sub-utility from group g. Then, the demand system at the lower stage (stage 1 or **within-group stage**) is:

$$w_{jt} = \beta_j^{(0)} + \beta_j^{(y)} \ln \left(\frac{e_{gt}}{P_{gt}} \right) + \sum_{k \in \mathcal{J}_g} \beta_{jk}^{(p)} \ln(p_{kt}) \tag{2.18}$$

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:

$$\frac{e_{gt}}{e_t} = \beta_g^{(0,2)} + \beta_g^{(y,2)} \ln \left(\frac{e_t}{P_t} \right) + \sum_{g'=1}^{G} \beta_{g,g'}^{(p,2)} \ln(P_{gt}) \tag{2.19}$$

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

$$\frac{e_t}{y_t} = \beta^{(0,3)} + \beta^{(y,3)} \ln \left(y_t \right) + \beta^{(p,3)} \ln(P_t) \tag{2.20}$$

This multi-stage budgeting model can reduce substantially the number of parameters. For instance, the 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 that make a total 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 estimated this restricted system. This the approach in Hausman (1998) that we describe below in more detail.

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

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

$$w_{jt} = \alpha_j^0 + \mathbf{X}_t\, \boldsymbol{\alpha}_j^1 + \gamma_j\, \ln(y_t/P_t) + \sum_{k=1}^J \beta_{jk}\, \ln(p_{kt}) + \varepsilon_{jt} \qquad (2.21)$$

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

2.3. Some limitations and extensions of this approach. (1) Every consumer purchases/consumes each of the J products. The system of demand equations that we have derived above is based of 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 $c > 0$ and $q_j > 0$ for every product j. This condition is very unrealistic when consider the demand of differentiated products within a product category, e.g., the demand of automobiles. In this context, a consumer buys only one unit of a single variety (e.g., one Toyota Corola) or of a few varieties (e.g., 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, e.g., age, education, etc. The propensity to substitute between different products can be also very heterogeneous across consumers. Therefore, ignoring consumer heterogeneity is a very important limitation of the actual applications in this literature. However, there is nothing that avoids the estimation of this model using household level data. Suppose that we have this type of data, and let use the subindex h for households. The demand system becomes:

$$w_{jht} = \alpha_j^0 + \mathbf{Z}_{ht} \; \boldsymbol{\alpha}_j^1 + [\gamma_j^0 + \gamma_j^1 \mathbf{Z}_{ht}] \ln(y_{ht}/P_t) + \sum_{k=1}^{J} [\beta_{jk}^0 + \beta_{jk}^1 \mathbf{Z}_{ht}] \ln(p_{kt}) + \varepsilon_{jht} \quad (2.22)$$

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

Note that (typically) prices do not vary across households. Therefore, price elasiticities are identified only from the time-series (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 introduces also a new issue in this class of models: some households do not consume all the product categories, even when these categories are quite aggregate, e.g., beef meat. However, this class of model predicts that the household consumes a positive amount of every product category.

(3) Too many parameters problem. In the standard model, the number of parameters is $2J + \frac{J(J+1)}{2}$, i.e., J intercept parameters (α); J income elasticities (γ); and $\frac{J(J+1)}{2}$ free price elasticities (β). The number of parameters increases quadratically with the number of goods. Note also that, in most applications, the sample variation in prices comes only from time series, and the sample size T is relatively small. This feature of the model implies that the number of products, J, should be quite small. For instance, even if J is as small as 5, the number of parameters to estimate is 25. Therefore, with this model and data, it is not

possible to estimate demand systems for differentiated products with many varieties. For instance, suppose that we are interested in the estimation of a demand system for different car models, and the number of car models is $J = 100$. Then, the number of parameters in the AIDS model is $5,250$, and we need many thousands of observations (markets or/and time periods) to estimate this model. This type of data is typically not available.

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

(5) Problems to predict demand of new goods. In the literature of demand of differentiated products, a class of problem that has received substantial attention is the evaluation or prediction of the demand of a new product. Trajtenberg (1989), Hausman (1996), and Petrin (2002) are some of the prominent applications that deal with this empirical question. In a demand system in product space, estimating the demand of a new good, say $J + 1$, requires estimates of the parameters associated with that good: α_{J+1}, γ_{J+1} and $\{\beta_{J+1,j} : j = 1, 2, ..., J+1\}$. Of course, this makes it impossible to make counterfactual predictions, i.e., predict the demand of a product that has not been introduced in any market yet. But it also limits the applicability of this model in cases where the new product has been introduced very recently or in very few markets, because we may not have enough data to estimate these parameters.

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

---

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

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 hundred of brands, the model and the estimation concentrates in 20 brands classified in three segments: adult (7 brands), child (4 brands), and family (9 brands). Apple-Cinnamon Cheerios are included in the family segment. We index markets by m, time by t, and brands by j, such that data can be described as $\{p_{jmt}, q_{jmt} : j = 1, 2, ..., 20; \ m = 1, 2, ..., 7; \ t = 1, 2, ..., 137\}$. Quantities are measured is physical units. There are not observable cost shifters.

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

$$s_{jmt} = \alpha_{jm}^1 + \alpha_t^2 + \gamma_j \ \ln(y_{gmt}) + \sum_{k=1}^{J} \beta_{jk} \ \ln(p_{kmt}) + \xi_{jmt} \tag{2.23}$$

where y_{gmt} is overall expenditure in segment/group g. The terms α_{jm}^1 and α_t^2 represent product, market and time effects, respectively, which are captured using dummies.

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

$$\ln(p_{jmt}) = \delta_j \ c_{jt} + \tau_{jm} + \kappa_j(\xi_{mt}) \tag{2.24}$$

where c_{jt} represents a common cost shifter (unobservable to the researcher) which is consistent with the national level production in this industry, τ_{jm} is city-brand fixed effect that captures differences in transportation costs, and $\kappa_j(\xi_{mt})$ captures the response of prices to local demand shocks, with $\xi_{mt} \equiv \{\xi_{1mt}, \xi_{2mt}, ..., \xi_{Jmt})$. The identification assumption is that

these demand shocks are not (spatially) correlated across markets: for any pair of markets $m \neq m$ it is assumed that:

$$E(u_j(\xi_{mt})\, u_k(\xi_{m't})) = 0 \qquad \text{for any } j, k \tag{2.25}$$

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

$$\overline{P}_{j(-m)t} = \frac{1}{M-1} \sum_{m' \neq m} p_{jm't} \tag{2.26}$$

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

$$e(\mathbf{p}, u) = \sum_{j=1}^{J} \alpha_j^0 \ln(p_j) + \frac{1}{2} \sum_{j=1}^{J} \sum_{k=1}^{J} \beta_{jk} \ln(p_j) \ln(p_k) + u \prod_{j=1}^{J} p_j^{\gamma_j} \tag{2.27}$$

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

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

$$\hat{D}_j(p_{1mt}, p_{2mt}, ..., p_{Jmt}^*) = 0 \tag{2.28}$$

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

the *Equivalent Variation* (in market m at period t) associated to the introduction of the new product is:

$$EV_{mt} = \hat{e}(p_{1mt}, p_{2mt}, ..., p^*_{Jmt}, u_{mt}) - y_{mt} \qquad (2.29)$$

Hausman consider this measure of consumer welfare.

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

3. Demand systems in characteristics space

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

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

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

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

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

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

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

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

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

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

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

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

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

Example. Logit model of product differentiation. Consider the following specification assumptions on the general model presented above. First, the utility from the outside product is linear and the same form all the consumers: $u(C; \boldsymbol{v}_h) = \alpha \, C$. Second, the utility of purchasing product j is $V(\mathbf{X}_j, \widetilde{\boldsymbol{\xi}}_j; \boldsymbol{v}_h) = \mathbf{X}_j \beta + \xi_j + \varepsilon_{hj}$, where ε's are Extreme Value Type 1 and independently and identically distributed (i.i.d.) over consumers and products. 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.

$$s_j = \frac{q_j}{H} = \frac{\exp\left\{\delta_j\right\}}{1 + \sum_{k=1}^{J} \exp\left\{\delta_k\right\}} \tag{3.5}$$

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

Example. Nested Logit model. 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 $\mathcal{J}_g$ 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:

$$\varepsilon_{hj} = \lambda \, \varepsilon_{hg}^{(1)} + \varepsilon_{hj}^{(2)} \tag{3.6}$$

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:

$$s_j = \frac{\exp\left\{\lambda \, I_g\right\}}{\sum_{g'=0}^{G} \exp\left\{\lambda \, I_{g'}\right\}} \frac{\exp\left\{\delta_j\right\}}{\sum_{k \in \mathcal{J}_g} \exp\left\{\delta_k\right\}} \tag{3.7}$$

where I_g is the inclusive value of group g that has the following form:

$$I_g \equiv \ln\left(\sum_{j \in \mathcal{J}_g} \exp\left\{\delta_j\right\}\right) \tag{3.8}$$

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 $s_g^* \equiv \sum_{j \in \mathcal{J}_g} s_j$ be the aggregate market share of all the products and group g. And let $s_{j|g} \equiv s_j / \sum_{k \in \mathcal{J}_g} s_k$ be the within-group market share of product j in 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 structural $s_{j|g} = \exp\left\{\delta_j\right\} / \exp\{I_g\}$, and the group market shares have the logit structure $\exp\left\{\lambda \, I_g\right\} / \sum_{g'=0}^{G} \exp\left\{\lambda \, I_{g'}\right\}$. ∎

Example. Random Coefficients Logit. Suppose that the utilities $V(\mathbf{X}_j, \widetilde{\boldsymbol{\xi}}_j; \boldsymbol{v}_h)$ and $u(C; \boldsymbol{v}_h)$ are linear in parameters, but these parameters are household specific. That is,

$U_{hj} = -\alpha_h \, p_j + \mathbf{X}_j \, \boldsymbol{\beta}_h + \xi_j + \varepsilon_{hj}$ where ε's are still *i.i.d.* Extreme Value Type 1, and

$$\begin{bmatrix} \alpha_h \\ \boldsymbol{\beta}_h \end{bmatrix} = \begin{bmatrix} \alpha \\ \boldsymbol{\beta} \end{bmatrix} + v_h \quad \text{with } v_h \sim i.i.d. \ N(0, \Sigma) \tag{3.9}$$

Then, we can write utilities as:

$$U_{hj} = -\alpha \, p_j + \mathbf{X}_j \, \boldsymbol{\beta} + \xi_j + \widetilde{v}_{hj} + \varepsilon_{hj} \tag{3.10}$$

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

$$s_j = \frac{q_j}{H} = \int \frac{\exp\{\delta_j + \widetilde{v}_{hj}\}}{1 + \sum_{k=1}^{J} \exp\{\delta_k + \widetilde{v}_{hk}\}} \, f(\widetilde{\mathbf{v}}_h \mid \mathbf{p}, \mathbf{X}, \Sigma) \, d\widetilde{\mathbf{v}}_h \tag{3.11}$$

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

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

$$s_j = \sigma_j(\boldsymbol{\delta} \mid \mathbf{p}, \mathbf{X}, \Sigma) \quad \text{for } j = 1, 2, ..., J \tag{3.12}$$

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

The imporance 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 alternative, say j and k, should not be affected by the availability or the attributes of other alternatives:

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

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:

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

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

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

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

$$\boldsymbol{\delta} = \sigma^{-1}(\mathbf{s} \mid \mathbf{p}, \mathbf{X}, \Sigma) \tag{3.13}$$

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 propery has important implications for the structural estimation of the demand system. Under this inversion, the unobserved product charcteristics 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 (Masnki, 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 $\ln(s_0) = -\ln(D)$. Combinig the equations for $\ln(s_j)$ and $\ln(s_0)$ we have that:

$$\delta_j = \ln(s_j) - \ln(s_0) \tag{3.14}$$

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 $\ln(s_j) = (\lambda - 1)$ $I_g + \ln(s_0) + \delta_j$. And taking into account that $I_g = [\ln(s_g^*) - \ln(s_0)]/\lambda$, we have that:

$$\delta_j = [\ln(s_j) - \ln(s_0)] + \left(\frac{1-\lambda}{\lambda}\right)[\ln(s_g^*) - \ln(s_0)] \tag{3.15}$$

and this equation is the inverse mapping $\sigma_j^{-1}(\mathbf{s} \mid \mathbf{p}, \mathbf{X}, \Sigma)$ 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) abd 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: $\boldsymbol{\delta} = F(\boldsymbol{\delta} \mid \mathbf{s}, \mathbf{p}, \mathbf{X}, \Sigma)$

or $\delta_j = F_j(\boldsymbol{\delta} \mid \mathbf{s}, \mathbf{p}, \mathbf{X}, \Sigma)$ where:

$$F_j(\boldsymbol{\delta} \mid \mathbf{s}, \mathbf{p}, \mathbf{X}, \Sigma) \equiv \delta_j + \ln(s_j) - \ln\left(\sigma_j(\boldsymbol{\delta}|\mathbf{p}, \mathbf{X}, \Sigma)\right) \tag{3.16}$$

It is straightforward to see that $\boldsymbol{\delta}$ is a fixed point of the mapping $F(\boldsymbol{\delta} \mid \mathbf{s}, \mathbf{p}, \mathbf{X}, \Sigma)$ if and only $\boldsymbol{\delta} = \sigma^{-1}(\mathbf{s} \mid \mathbf{p}, \mathbf{X}, \Sigma)$. Therefore, finding a solution (fixed point) in $\boldsymbol{\delta}$ to the system of equation $\boldsymbol{\delta} = F(\boldsymbol{\delta} \mid \mathbf{s}, \mathbf{p}, \mathbf{X}, \Sigma)$ is equivalent to finding the inverse function $\sigma^{-1}(\mathbf{s} \mid \mathbf{p}, \mathbf{X}, \Sigma)$ at a particular value of $(\mathbf{s}, \mathbf{p}, \mathbf{X}, \Sigma)$. Very importantly, Berry (1994) shows that this mapping $F(\boldsymbol{\delta} \mid \mathbf{s}, \mathbf{p}, \mathbf{X}, \Sigma)$ is a contraction as long as the values of $\boldsymbol{\delta}$ are not too small. Under this condition, the mapping has a unique fixed point and we can find it by using *fixed point iteration algorithm*. This algorithm proceeds as follows.

Fixed Point algorithm

- Start with an initial guess $\boldsymbol{\delta}^0$.

- At iteration $R \geq 1$, we calculate $\sigma_j(\boldsymbol{\delta}^{R-1} \mid \mathbf{p}, \mathbf{X}, \Sigma)$ for every product j by evaluating the multiple integration expression in equation (3.11) and then we update the vector $\boldsymbol{\delta}$ using the updating equation:

$$\boldsymbol{\delta}^R = F_j(\boldsymbol{\delta}^{R-1} \mid \mathbf{s}, \mathbf{p}, \mathbf{X}, \Sigma) = \delta_j^{R-1} + \ln(s_j) - \ln\left(\sigma_j(\boldsymbol{\delta}^{R-1}|\mathbf{p}, \mathbf{X}, \Sigma)\right) \tag{3.17}$$

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

3.3. Dealing with liminations of demand models in product space. 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 with precision demand systems where 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 $\sigma^{-1}(\mathbf{s} \mid \mathbf{p}, \mathbf{X}, \Sigma) = -\alpha \, p_j + \mathbf{X}_j \, \boldsymbol{\beta} + \xi_j$ 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 price 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 on 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 of prices the characteristics of the other products. We provide further details in the next section.

[4] Problems to predict demand of new varieties. Predicting the demand of new product 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:

$$q_{J+1} = H \int \frac{\exp\left\{-\alpha\ p_{J+1} + \mathbf{X}_{J+1}\ \boldsymbol{\beta} + \xi_{J+1} + \widetilde{v}_{hJ+1}\right\}}{1 + \sum_{k=1}^{J+1} \exp\left\{-\alpha\ p_k + \mathbf{X}_k\ \boldsymbol{\beta} + \xi_k + \widetilde{v}_{hk}\right\}}\ f(\widetilde{\mathbf{v}}_h | \mathbf{p}, \mathbf{X}, \Sigma) \tag{3.18}$$

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.

3.4. Estimation. Suppose that the researcher has a dataset from a **single market** at only **one period** but for a product with many varieties: $M = T = 1$ but large J (e.g., 100 varieties). The researcher observes the dataset $\{q_j, \mathbf{X}_j, p_j : j = 1, 2, ..., J\}$. Given these data, the researcher is interested in the estimation of the parameters of the demand system: $\theta = (\alpha, \beta, \Sigma)$. For the moment, we assume that market size H is known to the researcher. But it can be also estimated as a parameter. For the asymptotic properties of the estimators, we consider that $J \to \infty$.

The econometric model is:

$$s_j = \sigma_j(\mathbf{X}, \mathbf{p}, \boldsymbol{\xi}; \theta) \tag{3.19}$$

Unobserved product characteristics $\boldsymbol{\xi}$ are correlated with prices $\mathbf{p}$ (endogeneity). Dealing with endogeneity in nonlinear models where unobservables do not enter additively is complicated. Without further restrictions, we need full MLE: an specification of the model of $\mathbf{p}$ and a parametric specification of the distribution of $\boldsymbol{\xi}$. 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 (BLP models) with the **invertibility property** described above. This property implies that we can represent the model using a equation where the unobservables $\boldsymbol{\xi}$ enter additively:

$$\sigma_j^{-1}(\mathbf{s}\ |\mathbf{p}, \mathbf{X}, \Sigma) = -\alpha\ p_j + \mathbf{X}_j\ \boldsymbol{\beta} + \xi_j \tag{3.20}$$

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

GMM estimation. Under the previous assumption, we can use the characteristics of other products $(\mathbf{X}_k : k \neq j)$ to construct moment coidtions to estimate structural parameters in equation (3.20). For instance, we can use as vector of instruments the average characteristics of other products, $\frac{1}{J-1}\sum_{k\neq j}\mathbf{X}_k$. It is clear that $E\left(\frac{1}{J-1}\sum_{k\neq j}\mathbf{X}_k\ \xi_j\right) = 0$, and we can estimate θ using GMM. Suppose that we have a vector of instruments $\mathbf{Z}_j$ (e.g., $\mathbf{Z}_j = [\mathbf{X}_j, \frac{1}{J-1}\sum_{k\neq j}\mathbf{X}_k]$) such that the following conditions hold:
(1) $E(\xi_{jt} \mid \mathbf{Z}_j) = 0;$
(2) $\dim(\mathbf{Z}_j) \geq \dim(\theta);$
(3) $E\left[\left(\frac{\partial\sigma_j^{-1}(\mathbf{s}|\mathbf{p},\mathbf{X},\Sigma)}{\partial\Sigma},\ p_j,\ \mathbf{X}_j\right)'\left(\frac{\partial\sigma_j^{-1}(\mathbf{s}|\mathbf{p},\mathbf{X},\Sigma)}{\partial\Sigma},\ p_j,\ \mathbf{X}_j\right)\ \mid \mathbf{Z}_j\right]$ is non-singular.
Under conditions (1) to (3) we can estimate θ consistently using GMM. Define the sample moment conditions:

$$m(\theta) = \frac{1}{J}\sum_{j=1}^{J}\mathbf{Z}_j\ \left(\sigma_j^{-1}(\mathbf{s}|\mathbf{p},\mathbf{X},\Sigma) + \alpha\ p_j - \mathbf{X}_j\ \boldsymbol{\beta}\right) = 0 \tag{3.21}$$

The GMM estimator is defined as:

$$\widehat{\theta} = \arg\min_{\theta}\left[m(\theta)'\ \mathbf{W}\ m(\theta)\right] \tag{3.22}$$

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

3.4.1. *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. Instead, we can define a set of "neighbors" for each product j, $N_j = \{k \neq j : \|\mathbf{X}_k - \mathbf{X}_j\| \leq \tau\}$, and construct the vector of instruments $\frac{1}{|\mathcal{N}_j|}\sum_{k\in\mathcal{N}_j}X_k$. This vector of instruments can have more sample variability than $\frac{1}{J-1}\sum_{k\neq j}X_k$ but it is also more correlated with X_j.

The vector of instruments 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, $Z_j = [X_j, \frac{1}{|\mathcal{N}_j|} \sum_{k \in \mathcal{N}_j} X_k])$, has only $2K$ such that the order condition of identification (2) does not hold, i.e., $\dim(Z_j) = 2K < (K+1) + \frac{K(K+1)}{2} = \dim(\theta)$. 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 restrictions is that Σ is a diagonal matrix (i.e., 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 with a vector of instruments $Z_j = [X_j, \frac{1}{|\mathcal{N}_j|} \sum_{k \in \mathcal{N}_j} X_k]$. Other approach that 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 $\frac{1}{|\mathcal{N}_j|} \sum_{k \in \mathcal{N}_j} X_k$. In this case, to use the optimal instruments one could use the two step method in Newey (1990).

3.4.2. *Weak instruments problem: Armstrong (2016).* Armstrong (2016) points out a pontential incosistency 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 GMM-BLP estimator is inconsistent [i.e., a extreme case of weak instruments]. This is also a potential issue in small samples: therei is a large bias and variance, and under-covered confidence intervals. Armstrong (2016) studies this issue under different data structures. Suppose that the dataset has variation over products (J products indexed by j), firms (N firms indexed ny n), and markets or time periods (T markets indexed by t), such that we observe prices and quantities p_{jnt} and q_{jnt}. Armstrong shows that the inconsistency depends on the form of the demand system (i.e., standard Logit or random coefficients model), and on the asymptotics of products per firm J/N.

Consider first 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:

$$p_j = MC_j + \frac{1}{\alpha} \frac{1}{1 - s_j} \tag{3.23}$$

BLP instruments affect price p_j only through the term $\frac{1}{1-s_j}$. If $\sqrt{J} \left[\frac{1}{1 - s_j} \right] \to$ constant [zero] as $J \to \infty$, then GMM-BLP estimator is inconsistent, and it is not asymptotically different to using instrumental variables that are independent of prices. This is an extreme case of

weak instruments. In the logit with multiproduct firms, firms internalize the cannibalization effect. In this case, as $J \to \infty$ keeping constant the number of firms N, markups converge to firm-specific constants, and the rate of convergence of markups is slower than $\sqrt{J}$. In this case, Logit or random coefficients Logit are consistent as $J \to \infty$.

Armstrong (2016) shows that the GMM-BLP estimator of the Random Coefficients Logit model is inconsistent when the number of markets T is fixed and the number of products per firm J/N is also 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.

3.4.3. *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 geaographic markets and J prododucts. 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 variance-components structure:

$$\xi_{jt} = \xi_j^{(1)} + \xi_t^{(2)} + \xi_{jt}^{(3)} \tag{3.24}$$

and $\xi_{jt}^{(3)}$ is not spatially correlated, i.e., 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 and we can construct instrumental variables for that are correlated with prices and uncorrelated with $\xi_{jt}^{(3)}$. More precisely, define $Z_{jt} = \frac{1}{T_R - 1} \sum_{t' \in \mathcal{T}_R, t' \neq t} p_{jt'}$ where T_R is the number of markets in region R (where market t belongs) and $\mathcal{T}_R$ is the set of markets in region R. Since is not serially 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 (3.24) 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, i.e., the equation at period t minus the equation at $t - 1$:

$$\sigma_j^{-1}(\mathbf{s}_t \mid \mathbf{p}_t, \mathbf{X}_t, \Sigma) - \sigma_j^{-1}(\mathbf{s}_{t-1} \mid \mathbf{p}_{t-1}, \mathbf{X}_{t-1}, \Sigma) = -\alpha \, \Delta p_{jt} + \Delta \mathbf{X}_{jt} \, \boldsymbol{\beta} + \Delta \xi_{jt} \tag{3.25}$$

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, $\mathbb{E}(\mathbf{Z}_{jt} \, \Delta \xi_{jt}^{(3)}) = 0$. And if shocks in marginal costs are serially correlated, then $\mathbf{Z}_{jt}$ is correlated with the price difference Δp_{jt} after controlling for the exogeneous regressors $\Delta \mathbf{X}_{jt}$.

3.4.4. *Computation of the GMM estimator.* Berry, Levinsohn, and Pakes (1995) proposes a Nested Fixed Point (NFXP) algorithm to compute the GMM estimator of θ.[3] As indicated but its name, this method can be described in terms of two nested fixed point algorithms: a inner algorithm that consists in 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 such that the GMM estimator can be defined as the value $\widehat{\theta}$ that satisfies the conditon $\partial Q(\widehat{\theta})/\partial \theta = 0$. 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, weh have that $\partial Q(\widehat{\theta})/\partial \theta = 0 = \partial Q(\theta^*)/\partial \theta + \partial^2 Q(\theta^*)/\partial\theta\partial\theta' \, [\widehat{\theta} - \theta^0]$ for $\theta^* = (1 - \lambda)\theta^0 + \lambda\widehat{\theta}$ and some $\lambda \in (0, 1)$. Solving for $\widehat{\theta}$, we have that:

$$\widehat{\theta} = \theta^0 - \left[\frac{\partial^2 Q(\theta^*)}{\partial\theta\partial\theta'}\right]^{-1} \left[\frac{\partial Q(\theta^*)}{\partial\theta}\right] \tag{3.26}$$

Given an arbitrary value θ^0 we do not know, at least ex-ante, the mean value θ^* that satisfies this condition and give us the minimizer/estimator $\widehat{\theta}$. Then, 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:

$$\theta^R = \theta^{R-1} - \left[\frac{\partial^2 Q(\theta^{R-1})}{\partial\theta\partial\theta'}\right]^{-1} \left[\frac{\partial Q(\theta^{R-1})}{\partial\theta}\right] \tag{3.27}$$

   - 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 (e.g., 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 other fixed point problem to evaluate the inverse mapping $\sigma_j^{-1}(\mathbf{s}| \, \mathbf{p}, \mathbf{X}, \Sigma)$.

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} an applied the Fixed Point described above (equation (3.17)) 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

---

[3]The term Nested Fixed Point (NFXP) algorithm was coined by Rust (1987) in the context of the estimation dynamic discrete choice structural models.

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 (3.27).

- 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 (e.g., 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. Dube, Fox, and Su (2012) propose the MPEDC algorithm. Lee and Seo (2015) propose a Nested Pseudo Likelihood method in the same spirit as Aguirregabiria and Mira (2002).

3.5. Nonparametric identification. Empirical applications of discrete choice models of demand make different parametric assumptions such us the normal distribution of the random coefficients, and additive separability of observable and unobservable product characteristics in the utility function. Berry and Haile (2014) show that these parametric functional forms and distributional assumptions are not essential for the identification of this type of demand systems. The identification relies primarily on the standard requirement that instruments be available for the endogenous variables.

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

$$F_v(v_h \mid \mathbf{X}, \mathbf{p}, \boldsymbol{\xi}) = F_v(v_h \mid \mathbf{X}_{(-1)}, \mathbf{p}, \ \mathbf{X}_1 + \boldsymbol{\xi}) \tag{3.28}$$

where $\mathbf{X}_{(-1)}$ represents the observable product characteristics other than X_1, and $\mathbf{X}_1 + \boldsymbol{\xi}$ 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:

$$s_j = \Pr\left(j = \arg\max_k v_{ik} \mid \mathbf{X}_{(-1)}, \mathbf{p}, \ \mathbf{X}_1 + \boldsymbol{\xi}\right) = \sigma_j\left(\mathbf{X}_{(-1)}, \mathbf{p}, \ \mathbf{X}_1 + \boldsymbol{\xi}\right) \qquad (3.29)$$

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

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

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

(i) All goods are weak gross substitutes: s; i.e., for any $k \neq j$, $\sigma_j\left(\mathbf{X}_{(-1)}, \mathbf{p}, \ \boldsymbol{\xi}^*\right)$ 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.

(ii) "Connected strict substitution". Starting from any inside good, there is a chain of substitution [i.e., 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, i.e., 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:

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

Let $\mathbf{Z}_t$ be a vector of instruments [we explain below how to obtain these instruments] and suppose that: (a) $E[\xi_{jt} \mid \mathbf{Z}_t] = 0$; (b) [completeness] if $E[B(\mathbf{s}_t, \mathbf{X}_{(-1)t}, \mathbf{p}_t) \mid \mathbf{Z}_t] = 0$, then $B(\mathbf{s}_t, \mathbf{X}_{(-1)t}, \mathbf{p}_t) = 0$ *almost surely*. **Important:** Completeness requires that $\dim(\mathbf{Z}_t) \geq \dim(\mathbf{s}_t, \mathbf{X}_{(-1)t}, \mathbf{p}_t)$, i.e., 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 invert again σ_j^{-1} to identify the demand system σ.

Sources of instruments. Note that we need instruments not only for prices but also for market shares $\mathbf{s}_t$. J instruments for $\mathbf{p}_t$ and J instruments for $\mathbf{s}_t$. **Instruments for** $\mathbf{s}_t$**:** instruments that affect quantities not only through prices. For instance, supply/MC shifters or Hausman-Nevo are IVs for prices but they are not useful for $\mathbf{s}_t$ because the 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, $\dfrac{\partial \boldsymbol{\sigma}^{-1}\left(\mathbf{s}_t, \mathbf{p}_t\right)}{\partial \mathbf{s}_t'} = \left[\dfrac{\partial \boldsymbol{\sigma}\left(\boldsymbol{\delta}_t, \mathbf{p}_t\right)}{\partial \boldsymbol{\delta}_t'}\right]$. Identifying the effects of $\mathbf{s}_t$ on $\boldsymbol{\sigma}^{-1}$ is equivalent to identifying the effects of $\boldsymbol{\xi}^*$ on market shares $\boldsymbol{\sigma}$. The vector $\mathbf{X}_{1t}$ directly shifts

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

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

4. Consumer valuation of product innovations

Product innovation is ubiquous is 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.
- Social value of innovations.

The standard approach is based on: Estimation of a demand system of differentiated products; Constructing consumer indirect utility function (or surplus function) with and without the new product. Typically, one of the two scenarios (with or without) is a counterfactual.

In the definition of the counterfactual scenario the researcher needs to the value of unobservables in the counterfactual scenario. e.g., industry time-trends, unobserved product characteristics, distribution of consumer idiosyncratic product-specific shocks. Trajtenberg (JPE, 1989). Petrin (JPE, 2002). Valuing new goods with product complementarity: Gentzkow (AER, 2007)

4.1. An Application: Hausman (1996) 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. Period 1980-92: 190 new brands were added to the pool of existing 160 brands. He deals with the limitations mentioned above by using: (a) Multi-stage budgeting (and focusing on most popular varieties); (b) Data from many periods (weekly data) and multiple geographic markets (cities), and assuming that parameters are constant across weeks-markets (up to fixed effects in the intercepts). (c) Exploiting assumptions on the geographic

structure of demand/supply shocks to generate instruments for prices. **(d) Evaluates the introduction of a new brand (Cheerios).**

Data. Supermarket scanner data: period 1990-1992. 137 weeks ($T = 137$); 7 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 hundred of brands, the model and the estimation concentrates in 20 brands classified in three segments: adult (7 brands), child (4 brands), and family (9 brands). $\{p_{jmt}, q_{jmt} : j = 1, 2, ..., 20; m = 1, 2, ..., 7; t = 1, 2, ..., 137\}$. Quantities are measured is physical units. There are not observable cost shifters.

Model. Almost-Ideal-Demand-System

$$w_{jmt} = \alpha^1_{jm} + \alpha^2_t + \gamma_j \ln\left(\frac{y_{gmt}}{P_{gmt}}\right) + \sum_{k \in J_g} \beta_{jk} \ln(p_{kmt}) + \varepsilon_{jmt}$$

The terms α^1_{jm} and α^2_t represent brand-city and time "fixed effects". Suppose that the supply (pricing equation) is:

$$\ln(p_{jmt}) = \delta_j \, c_{jt} + \tau_{jm} + u_{jmt}$$

c_{jt} represents a common cost shifter (unobservable to the researcher); τ_{jm} is city-brand fixed effect that captures differences in transportation costs; u_{jmt} captures potential response of prices to local demand shocks.

Instruments. The identification assumption is that demand shocks are not (spatially) correlated across markets: for any pair of markets $m \neq m$ it is assumed that:

$$E(u_{jmt} \, u_{km't}) = 0 \qquad \text{for any } j, k$$

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

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

Value of cheerios. Hausman uses the estimated demand system to evaluate the value of a new varity 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 virtiual 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 an 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 $\mathcal{J}$ with vector of prices $\mathbf{p}$. The indirect utility function is defined as (income effects are assumed away because linearity):

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

Following McFadden (1981) and Small & Rosen (1981), we define the **money-metric welfare function**: $\frac{1}{\alpha_h} v(\mathbf{p}, \alpha_h, \beta_h, \varepsilon_h)$. And its aggregate version: $W(\mathbf{p}) = \int \frac{1}{\alpha_h} v(\mathbf{p}, \alpha_h, \beta_h, \varepsilon_h)$ $dF(\alpha_h, \beta_h, \varepsilon_h)$. For the random coefficients logit model:

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

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

Some limitations

[1] Problems to evaluate **radical innovations** with new types of characteristics.

[2] **Logit errors**. There is very limited "crowding" of products, i.e., $\lim_{J \to \infty} W(\mathcal{J}) = \infty$ (see Ackerberg and Rysman, RAND 2005).

[3] **Outside alternative**. Unobserved "qualities" ξ_{jt} are relative to the outside alternative. Suppose there is quality improvements in the outside alternative. We will underestimate the welfare improvements in this industry.

4.2. Trajtenberg (JPE, 1989). Trajtenberg (1989) on computed tomography scanners. Computed tomography (CT) scanners is considered a key innovation in imaging diagnosis in medicine during the 1970s. The first was installed in 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 for the quality: scan time; image quality; and reconstruction time.

Data. Period: 1973-1981. 55 products. Product characteristics (price, scan speed, resolution, reconstruction speed) and sales in US. Identity and attributes of the buying hospital. Hospital-year level data: the dependent variable is the product choice of hospital h at year t.

Estimated model. Nested logit at the hospital-year level: nests are "head" and "body" scanner. Utility: Quadratic in the three product attributed (other than price). Estimated elasticity of substitution between the two "nests" is very close to zero, i.e., two different

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

products. Estimated $\alpha's$ are significant but with the wrong sign for "body" [Not accounting for endogeneity of prices].

Welfare effects. The counterfactual is 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 of firms' R&D investment, he obtains a social rate of return of **270%**.

4.3. Petrin (JPE, 2002) on minivans. Evaluate consumer welfare gains from the introduction of a new type of car, the minivan. Estimation of a BLP demand system of automobiles. Combine 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 the Dodge caravan (its minivan). It was an immediate success. GM and Ford responded quickly introducing in 1985 their own. By 1998, there were 6 firms selling a total of 13 different minivans, Chrysler being the leader (44%).

Data. Period: 1981-1993. $J = 2407$. Product-year panel. Variables: Quantity sold; price, acceleration, dimensions, drive type, fuel efficiency, a measure of luxury. Consumer expenditure survey (CEX).links demographics of purchasers of new vehicles to the vehicles they purchase. In the CEX, we observe 2,660 new vehicle purchases over the period and sample. 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.

Market shares

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)**
	B. Base Coefficients (β's)			
Constant	−10.03	−10.04	−12.74	−15.67
	(.32)**	(.34)**	(5.65)**	(4.39)**
Horsepower/weight	1.48	3.78	3.40	−2.83
	(.34)**	(.44)**	(39.79)	(8.16)
Size	3.17	3.25	4.60	4.80
	(.26)**	(.27)**	(24.64)	(3.57)*
Air conditioning standard	−.20	.21	−1.97	3.88
	(.06)**	(.08)**	(2.23)	(2.21)*
Miles/dollar	.18	.05	−.54	−15.79
	(.06)**	(.07)	(3.40)	(.87)**
Front wheel drive	.32	.15	−5.24	−12.32
	(.05)**	(.06)**	(3.09)	(2.36)**
Minivan	.09	−.10	−4.34	−5.65
	(.14)	(.15)	(13.16)	(.68)**
Station wagon	−1.12	−1.12	−20.52	−1.31
	(.06)**	(.07)**	(36.17)	(.36)**
Sport-utility	−.41	−.61	−3.10	−4.38
	(.09)**	(.10)**	(10.76)	(.41)**
Full-size van	−1.73	−1.89	−28.54	−5.26
	(.16)**	(.17)**	(235.51)	(1.30)**
% change GNP	.03	.03	.08	.24
	(.01)**	(.01)**	(.02)**	(.02)**

Petrin on minivans: Estimates

Estimates

Random coef.

Marginal costs

Welfare evaluation. Counterfactual: no minivans. It takes into account the counterfactual equilibrium prices without minivans (based on estimates of demand and marginal costs). The introduction of minivans (particularly, Dodge caravan) had an important negative effect on 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.

Price changes with / without minivans

RANDOM COEFFICIENT PARAMETER ESTIMATES

	RANDOM COEFFICIENTS (γ's)	
VARIABLE	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)**

TABLE 8

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

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

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

The mean per capita Compensated Variation for minivans is \$1247. But the estimates are substantially larger with the other methods (OLS logit; IV logit: IV BLP without micro moments). This is mainly because these methods under-estimate the marginal utility of income. Decomposition of the welfare gains in the contribution of: x_j and ξ_j, \$851; logit ε_{hj}: \$396. Other methods imply very implausible contribution of the idiosyncratic ε's.

Welfare gains

4.4. Ackerberg & Rysman (RAND, 2005). The Logit errors can have unrealistic implications on the evaluation of welfare gains. Because these errors, welfare increases un-boundedly (though concavely) with J. No crowding, e.g., identical products except for the ε's: $W = \ln(\sum_{j=0}^{J} \exp\{\delta\}) = \delta \ln(J+1)$. Though the BLP or RC-Logit limits the influence of the logit errors, it is still subject to this problem. Ackerberg & Rysman (2005) propose a simple but interesting modification of the logit model that can account 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 $j > 0$:

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

$\sigma(J)$ is strictly decreasing in J and it goes to 0 as J goes to ∞. As J increases, the differentiation from the $\varepsilon's$ becomes less and less important. Function $\sigma(J)$ can be parameterized and its parameters estimated together with the rest of the model. Though Ackerberg & Rysman consider this approach, they favor a similar approach that is simpler to implement. Instead, the consider the model. For $j > 0$:

$$U_{hj} = -\alpha_h p_j + \mathbf{x}_j \beta_h + \xi_j + f(J, \gamma) + \varepsilon_{hj}$$
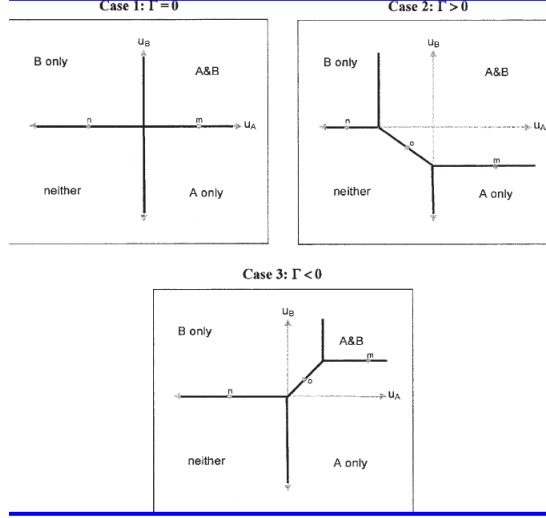
where $f(J, \gamma)$ is a decreasing function of J parameterized by γ. For instance, $f(J, \gamma) = \gamma \ln(J)$. It can be also extended to a nested logit version. For group g: $f_g(J, \gamma) = \gamma_g \ln(J_g)$. The reasons for the specification $f(J, \gamma)$ instead of $\sigma(J, \gamma)$ is simplicity in estimation.

4.5. Valuing new goods with product complementarity: Gentzkow (AER, 2007). The class of discrete choice demand models we have considered so far rule out complementarity between products. This is an important limitation in some relevant contexts:

- Evaluating the merger between two firms producing complements, e.g., Pepsico and Frito-Lay

- Evaluating the welfare effects of new products.

Also, sometimes there are both substitution and complementarity effects: e.g., radio stations on recorded music; movie on the book novel; Uber and taxis. Gentzkow (2007) extends McFadenn / BLP framework to allow for complementarity, and studies the demand and welfare effect of online newspapers.

Model. Now consumers can choose bundles of products. Simple example: Two products A and B. Choice set: $\{0, A, B, AB\}$. Utilities: 0, u_A, u_B, and $u_{AB} = u_A + u_B + \Gamma$. Discrete choice: $P_j = \Pr(u_j = \max\{0, u_A, u_B, u_{AB}\})$. Quantities: $Q_A = P_A + P_{AB}$; $Q_B = P_B + P_{AB}$. Products A and B are: substitutes if $\partial Q_A / \partial p_B > 0$; complements if $\partial Q_A / \partial p_B < 0$. Complements / substitutes is closely related to the sign of Γ.

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, but it might also reflect that unobservable tastes for the goods are correlated.

Suppose that (v_{hA}, v_{hB}) are standard normals with correlation ρ. The parameters of the model are: β_A, β_B, α, ρ, Γ. The researcher (with consumer level data) observes prices and bundles market shares: P_A, P_B, P_{AB}.

Identification. Even with micro-level data on shares P_A, P_B, 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. But these two interpretations have very different economic and policy implications, including the evaluation of new products, e.g., including A when B was present.

Gentzkow considers two sources of identification or additional moment conditions: (1) Exclusion restrictions; and (2) Panel data.

(1) 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.

$$\beta_A(z), \text{ but } \beta_B \text{ does not depend on } z$$

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 $[P_A(z), P_B(z), P_{AB}(z): z \in \{0,1\}]$ can identify $\beta_A(0)$, $\beta_A(1)$, β_B, Γ, and ρ. Intuition: if $\Gamma > 0$ (complementaity), 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 iid over h, t. Preference parameters are assumed 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 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.

Time-series readers

Estimated Gammas

Effects of post.com
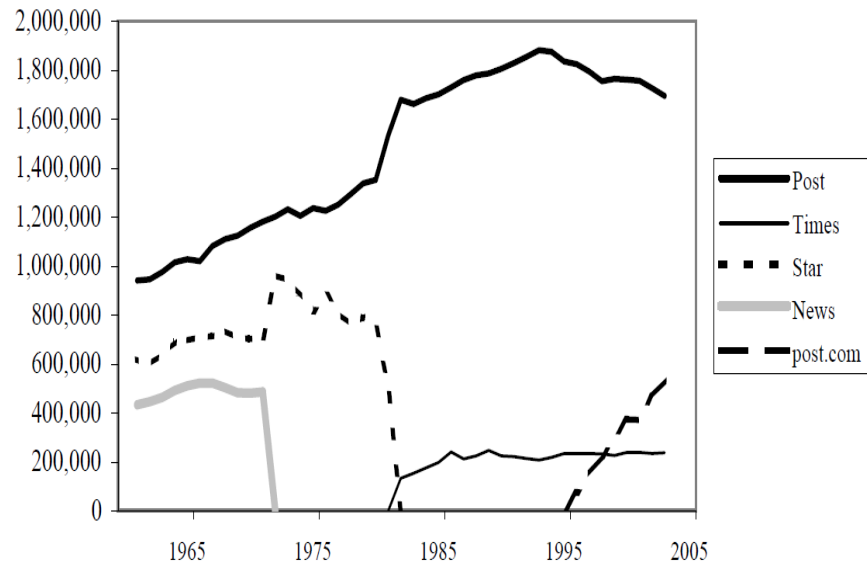
TABLE 6—PARAMETER ESTIMATES FROM FULL MODEL: OTHER

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)		

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)

Main results. Reduced-form OLS regressions and 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**.

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.

5. Exercises

5.1. 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, i.e., `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, i.e., 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.

5.2. 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, e.g., 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, i.e., `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, e.g., 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.

Bibliography

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

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

[3] Aguirregabiria, V. and P. Mira (2002): "Swapping the nested fixed point algorithm: A class of estimators for discrete Markov decision models," Econometrica 70, 1519-1543.

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

[5] Anderson, S., A. De Palma and J-F Thisse (1992), Discrete Choice Theory of Product Differentiation, Cambridge, MA: MIT Press.

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

[7] Armstrong, T. B. (2016): "Large market asymptotics for differentiated product demand estimators with economic models of supply," Econometrica, 84(5), 1961-1980.

[8] Berkovec, J. and J. Rust (1985): "A nested logit model of automobile holdings for one vehicle households." *Transportation Research Part B: Methodological*, 19(4), 275-285.

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

[10] Berry, S., and P. Haile (2014): "Identification in differentiated products markets using market level data," Econometrica, 82(5), 1749-1797.

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

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

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

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

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

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

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

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

[19] Deaton; A. and J. Muellbauer (1980a): "An Almost Ideal Demand System," American Economic Review, 70(3), 312-326.

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

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

[22] Dubé, J. P., Fox, J. T., & Su, C. L. (2012): "Improving the numerical performance of static and dynamic aggregate discrete choice random coefficients demand estimation," Econometrica, 80(5), 2231-2267.

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

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

[25] Gentzkow, M. (2007). Valuing new goods in a model with complementarity: Online newspapers. American Economic Review, 97(3), 713-744.

[26] Gowrisankaran, G., and M. Rysman (2012): "Dynamics of consumer demand for new durable goods," Journal of political Economy, 120(6), 1173-1219.

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

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

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

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

[31] Lee, J., and Seo, K. (2015): "A computationally fast estimator for random coefficients logit demand models using aggregate data," The RAND Journal of Economics, 46(1), 86-102.

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

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

[34] Manski, C, (1983): "Analysis of equilibrium automobile holdings in Israel with aggregate discrete choice models." Transportation Research Part B: Methodological, 17(5), 373-389.

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

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

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

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

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

[40] Petrin, A. (2002): "Quantifying the benefits of new products: The case of the minivan," Journal of political Economy, 110(4), 705-729.

[41] Rust, J. (1987): "Optimal replacement of GMC bus engines: An empirical model of Harold Zurcher," Econometrica 55, 999-1033.

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

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

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

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

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

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

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

CHAPTER 3

Estimation of Production Functions

1. Introduction

Production functions (PF) are important primitive components of many economic models. The estimation of PFs plays a key role in the empirical analysis of issues such as productivity dispersion and misallocation, the contribution of different factors to economic growth, skill-biased technological change, estimation of economies of scale and economies of scope, evaluation of the effects of new technologies, learning-by-doing, or the quantification of production externalities, among many others.

In empirical IO, the estimation of production functions can be used to obtain firms' costs. Cost functions play an important role in any empirical study of industry competition. As explained in chapter 1, data on production costs at the firm-market-product level is rare. For this reason costs functions are often estimated in an indirect way, using first order conditions of optimality for profit maximization (see chapter 4). However, cross-sectional or panel datasets with firm-level information on output and inputs of the production process are more commonly available. Given this information, it is possible to estimate the industry production function and use it to obtain firms' cost functions.

There are multiple issues that should be taken into account in the estimation of productions functions.

(a) Measurement issues. There are important measurement problems such as measurement error in output (e.g., observeing revenue instead of output in physical units), and inputs (e.g., construction of the economic value of capital using information of book value; differences in the quality of labor).

(b) Specification assumptions. The choice of functional form for the production function is an important modelling decision, especially when the model includes different types of labor and capital inputs that may be complements or substitutes.

(c) Simultaneity / endogeneity. It is a key econometric issue in the estimation of production functions. Observed inputs (e.g., labor and capital) can be correlated with unobserved inputs or productivity shocks (e.g., managerial ability, quality of land, materials, capacity utilization). This correlation introduces biases in some estimators of PF parameters.

(d) Multicollinearity between observed inputs is also a relevant issue in some empirical applications. The high correlation between observed labor and capital can seriously reduce the precision in the estimation of PF parameters.

(e) Endogenous exit. In panel datasets, firm exit from the sample is not exogenous and it is correlated with firm size. Smaller firms are more likely to exit than larger firms. Endogenous exit can introduce selection-bias in some estimators of PF parameters.

In this chapter, we concentrate on the problems of simultaneity, multicollinearity, and endogenous exit, and on different solutions that have been proposed to deal with these issues. For the sake of simplicity, we discuss these issues in the context of a Cobb-Douglas PF. However, the arguments and results can be extended to more general specifications of PFs. In principle, some of the estimation approaches can be generalized to estimate nonparametric specifications of PF. Griliches and Mairesse (1998), Bond and Van Reenen (2007), and Ackerberg et al. (2007) include surveys of this literature.

2. Model and Data

2.1. Model. A Production Function (PF) is a description of a production technology that relates the physical output of a production process to the physical inputs or factors of production. A general representation is:

$$Y = F(X_1, X_2, ..., X_J, A) \tag{2.1}$$

where Y is a measure of firm output, X_1, X_2, .., and X_J are measures of J firm inputs, and A represents the firm technological efficiency. The marginal productivity of input j is $MP_j = \partial F / \partial X_j$.

A very common specification is the Cobb-Douglas PF (Cobb and Douglas, 1928, *American Economic Review)*:

$$Y = L^{\alpha_L} \ K^{\alpha_K} \ A \tag{2.2}$$

where L and K represent labor and capital inputs, respectively, and α_L and α_K are technological parameters that are assumed the same for all the firms in the market and industry under study. This Cobb-Douglas PF can be generalized to include explicitly more inputs, e.g., $Y = L^{\alpha_L} \ K^{\alpha_K} \ R^{\alpha_R} \ E^{\alpha_E} \ A$, where R represents R&D and E is energy inputs. We can also distinguish different types of labor (blue collar and white collar labor), and capital (equipment, information technology). For the Cobb-Douglas PF, the productivity term A is denoted the *Total Factor Productivity* (TFP). For this PF the marginal productivity of input j is $MP_j = \alpha_j \frac{Y}{X_j}$. All the inputs are complements in production, i.e., the marginal productivity of any input j increases with the amount of any other input k:

$$\frac{\partial MP_j}{\partial X_k} = \frac{\alpha_j}{X_j} \frac{\alpha_k}{X_k} Y > 0 \tag{2.3}$$

This is not necessarily the case for other production functions such as the Constant Elasticity of Substitution (CES) or the Translog.

Given the production function $Y = F(X_1, X_2, ..., X_J, A)$ and input prices $(W_1, W_2, ..., W_J)$, the cost function $C(Y)$ is defined as the minimum cost of producing the amount of output Y:

$$C(Y) = \min_{\{X_1, X_2, ..., X_J\}} W_1 X_1 + W_2 X_2 + ... + W_J X_J$$

$$\text{subject to: } Y \geq F(X_1, X_2, ..., X_J, A) \tag{2.4}$$

The marginal conditions of optimality imply that for every input j, $W_j - \lambda\, F_j(\mathbf{X}, A) = 0$, where λ is the Lagrange multiplier of the restriction.

Exercise 3.1. Given the Cobb-Douglas PF $Y = X_1^{\alpha_1} ... X_1^{\alpha_J} A$ and input prices W_j, obtain the corresponding cost function. The marginal condition of optimality for input j implies $W_j - \lambda \alpha_j\, (Y/X_j) = 0$, or equivalently:

$$W_j\, X_j = \lambda \alpha_j\, Y \tag{2.5}$$

Therefore, the cost is equal to $\sum_{j=1}^{J} W_j\, X_j = \lambda \alpha\, Y$, where the parameter α is defined as $\alpha \equiv \sum_{j=1}^{J} \alpha_j$. Note that α represents the returns to scale in the production function: constant if $\alpha = 1$, decreasing if $\alpha < 1$, and increasing if $\alpha > 1$. To obtain the expression of the cost function, we still need to obtain the (endogenous) value of the Lagrange multiplier λ. For this, we solve the marginal conditions $X_j = \lambda \alpha_j\, Y/W_j$ into the production function:

$$Y = A \left(\frac{\lambda \alpha_1 Y}{W_1}\right)^{\alpha_1} \left(\frac{\lambda \alpha_2 Y}{W_2}\right)^{\alpha_2} ... \left(\frac{\lambda \alpha_J Y}{W_J}\right)^{\alpha_J} \tag{2.6}$$

Solving in this expression for the Lagrange multiplier we get

$$\lambda = \left(\frac{W_1}{\alpha_1}\right)^{\frac{\alpha_1}{\alpha}} \left(\frac{W_2}{\alpha_2}\right)^{\frac{\alpha_2}{\alpha}} ... \left(\frac{W_J}{\alpha_J}\right)^{\frac{\alpha_J}{\alpha}} Y^{\frac{1-\alpha}{\alpha}}\, A^{\frac{-1}{\alpha}} \tag{2.7}$$

And plugging this multiplier into the expression $\lambda \alpha\, Y$ for the cost, we obtain the cost function:

$$C(Y) = \alpha \left(\frac{Y}{A}\right)^{\frac{1}{\alpha}} \left(\frac{W_1}{\alpha_1}\right)^{\frac{\alpha_1}{\alpha}} \left(\frac{W_2}{\alpha_2}\right)^{\frac{\alpha_2}{\alpha}} ... \left(\frac{W_J}{\alpha_J}\right)^{\frac{\alpha_J}{\alpha}} \quad \blacksquare \tag{2.8}$$

Looking at the Cobb-Douglas cost function in equation (2.8) we can identify some interesting properties. First, the returns to scale parameter α determines the shape of the cost as a function of output. More specifically, the sign of the second derivative $C''(Y)$ is equal to the sign of $\frac{1}{\alpha} - 1$. If $\alpha = 1$ (*constant returns to scale*, CRS) we have $C''(Y) = 0$ such that the cost function is linear in output. If $\alpha < 1$ (*decreasing returns to scale*, DRS) we have $C''(Y) > 0$ and the cost function is strictly convex in output. Finally, if $\alpha > 1$ (*increasing returns to scale*, IRS) we have $C''(Y) < 0$ such that the cost function is concave in output.

An attractive feature of the Cobb-Douglas PF from the point of view of estimation is that it is linear in logarithms:

$$y = \alpha_L \; l + \alpha_K \; k + \omega \tag{2.9}$$

where y is the logarithm of output, l is the logarithm of labor, k is the logarithm of physical capital, and ω is the logarithm of the residual term U. The simplicity of the Cobb-Douglas PF comes also with a price. One of its drawbacks is that it implies that the elasticity of substitution between labor and capital (or between any two inputs) is always one. This implies that all technological changes are neutral for the demand of inputs. For this reason, the Cobb-Douglas PF cannot be used to study topics such as skill-biased technological change. For empirical studies where it is important to have a flexible form for the elasticity of substitution between inputs, the translog PF has been a popular specification:

$$Y = L^{[\alpha_{L0} + \alpha_{LL} l + \alpha_{LK} k]} \; K^{[\alpha_{K0} + \alpha_{KL} l + \alpha_{KK} k]} \; U \tag{2.10}$$

that in logarithms becomes,

$$y = \alpha_{L0} \; l + \alpha_{K0} \; k + \alpha_{LL} \; l^2 + \alpha_{KK} \; k^2 + (\alpha_{LK} + \alpha_{KL}) \; l \; k + \omega \tag{2.11}$$

2.2. Data. The most common type of data that has been used for the estimation of PFs consists of panel data of firms or plants with annual frequency and information on: a measure of output, e.g., number of units, or revenue, or valued added; input measures such as labor, capital, R&D, materials, and energy; and some measures of output and input prices typically at the industry level but sometimes at the firm level. For the US, the most commonly used datasets in the estimation of PFs has been Compustat, and the Longitudinal Research Database from US Census Bureau. In Europe, some country Central Banks (e.g., Bank of Italy, Bank of Spain) collect firm level panel data with rich information on output, inputs, and prices.

For the rest of this chapter we consider that researcher observes a panel dataset of N firms, indexed by i, over several periods of time, indexed by t, with the following information:

$$\textbf{Data} \; = \; \{y_{it}, \; l_{it}, \; k_{it}, \; w_{it}, \; r_{it} : \; i = 1, 2, ...N; \; t = 1, 2, ..., T_i\} \tag{2.12}$$

where y, l, and k have been defined above, and w and r represent the logarithms of the price of labor and the price of capital for the firm, respectively. T_i is the number of periods that the researcher observes firm i.

Throughout this chapter, we consider that all the observed variables are in mean deviations. Therefore, we omit constant terms in all the equations.

3. Econometric Issues

We are interested in the estimation of the parameters α_L and α_K in the Cobb-Douglas PF (in logs):

$$y_{it} = \alpha_L \ l_{it} + \alpha_K \ k_{it} + \omega_{it} + e_{it} \tag{3.1}$$

ω_{it} represents unobserved (for the econometrician) inputs such as managerial ability, quality of land, materials, etc, which are known to the firm when it decides capital and labor. We refer to ω_{it} as the logarithm of *total factor productivity (log-TFP), or unobserved productivity, or productivity shock.* e_{it} represents measurement error in output, or any shock affecting output that is unknown to the firm when it decides capital and labor. We assume that the error term e_{it} is independent of inputs and of the productivity shock. We use y_{it}^e to represent the "true" expected value of output for the firm, $y_{it}^e \equiv y_{it} - e_{it}$.

3.1. Simultaneity Problem. The simultaneity problem in the estimation of a PF establishes that if the unobserved productivity ω_{it} is known to the firm when it decides the amount of inputs to use in production, (k_{it}, l_{it}), then these observed inputs should be correlated with the unobservable ω_{it} and the OLS estimator of α_L and α_K will be biased and inconsistent. This problem was already pointed out in the seminal paper by Marshak and Andrews (1944).

Example 1: Suppose that firms in our sample operate in the same markets for output and inputs. These markets are competitive. Output and inputs are homogeneous products across firms. For simplicity, consider a PF with only one input, say labor: $Y = L^{\alpha_L} \exp\{\omega + e\}$. The first order condition of optimality for the demand of labor implies that the expected marginal productivity should be equal to the price of labor W_L: i.e., $\alpha_L \ Y^e/L = W_L$, where $Y^e = Y/\exp\{e\}$ because the firm's profit maximization problem does not depend on the measurement error or/and non-anticipated shocks in e_{it}. . Note that the price of labor W_L is the same for all the firms because, by assumption, they operate in the same competitive output and input markets. Then, the model can be described in terms of two equations: the production function and the marginal condition of optimality in the demand for labor. In logarithms, and in deviations with respect to mean values (no constant terms), these two equations are:[1]

$$
\begin{aligned}
y_{it} &= \alpha_L \ l_{it} + \omega_{it} + e_{it} \\
y_{it} - l_{it} &= e_{it}
\end{aligned}
\tag{3.2}
$$

---

[1]The firm's profit maximization problem depends on output $\exp\{y_i^e\}$ without the measurement error e_i.

The reduced form equations of this structural model are:

$$y_{it} = \frac{\omega_{it}}{1 - \alpha_L} + e_{it}$$

$$l_{it} = \frac{\omega_{it}}{1 - \alpha_L}$$

(3.3)

Given these expressions for the reduced form equations, it is straightforward to obtain the bias in the OLS estimation of the PF. The OLS estimator of α_L in this simple regression model is a consistent estimator of $Cov(y_{it}, l_{it})/Var(l_{it})$. But the reduced form equations, together with the condition $Cov(\omega_{it}, e_{it}) = 0$, imply that the covariance between log-output and log-labor should be equal to the variance of log-labor: $Cov(y_{it}, l_i) = Var(l_{it})$. Therefore, under the conditions of this model the OLS estimator of α_L converges in probability to 1 regardless the true value of α_L. Even in the hypothetical case that labor has very low productivity and α_L is close to zero, the OLS estimator converges in probability to 1. It is clear that in this case ignoring the endogeneity of inputs can generate a serious in the estimation of the PF parameters.

[Graphical representation of structural equations in space (ℓ, y) and graphical interpretation of the bias.]

Example 2: Consider the similar conditions as in Example 1, but now firms in our sample produce differentiated products and use differentiated labor inputs. In particular, the price of labor R_{it} is an exogenous variable that has variation across firms and over time. Suppose that a firm is a price taker in the market for the type labor input that it demands to produce its product and that the market price R_{it} is independent of the firm's productivity shock ω_{it}. In this version of the model the system of structural equations is very similar to the one in (3.2) with the only difference that the labor demand equation now includes the logarithm of the price of labor: $y_{it} - l_{it} = r_{it} + e_{it}$. The reduced form equations for this model are:

$$y_{it} = \frac{\omega_{it} - r_{it}}{1 - \alpha_L} + r_{it} + e_{it}$$

$$l_{it} = \frac{\omega_{it} - r_{it}}{1 - \alpha_L}$$

(3.4)

Again, we can use these reduced form equations to obtain the asymptotic bias in the estimation of α_L if we ignore the endogeneity of labor in the estimation of the PF. The OLS estimator of α_L converges in probability to $Cov(y_{it}, l_{it})/Var(l_{it})$ and in this case this implies the following expression for the bias:

$$Bias\left(\hat{\alpha}_L^{OLS}\right) = \frac{1 - \alpha_L}{1 + \dfrac{\sigma_r^2}{\sigma_\omega^2}}$$

(3.5)

where σ_ω^2 and σ_r^2 represent the variance of the productivity shock and the logarithm of the price of labor, respectively. The bias is always upward because the firm's labor demand is always positively correlated with the firm's productivity shock. The ratio between the variance of the price of labor and the variance of productivity, $\sigma_r^2/\sigma_\omega^2$, plays a key role in the determination of the magnitude of this bias. Sample variability in input prices, if it is not correlated with the productivity shock, induces exogenous variability in the labor input. This exogenous sample variability in labor reduces the bias of the OLS estimator. The bias of the OLS estimator declines monotonically with the variance ratio $\sigma_r^2/\sigma_\omega^2$. Nevertheless, the bias can be very significant if the exogenous variability in input prices is not much larger than the variability in unobserved productivity.

3.2. Endogenous Exit. Firm or plant panel datasets are unbalanced, with significant amount of firm exits. Exiting firms are not randomly chosen from the population of operating firms. For instance, existing firms are typically smaller than surviving firms.

Let d_{it} be the indicator of the event "firm i stays in the market at the end of period t". Let $V^1(l_{it-1}, k_{it}, \omega_{it})$ be the value of staying in the market, and let $V^0(l_{it-1}, k_{it}, \omega_{it})$ be the value of exiting (i.e., the scrapping value of the firm). Then, the optimal exit/stay decision is:

$$d_{it} = I\left\{ \ V^1(l_{it-1}, k_{it}, \omega_{it}) - V^0(l_{it-1}, k_{it}, \omega_{it}) \geq 0 \ \right\} \tag{3.6}$$

Under standard conditions, the function $V^1(l_{it-1}, k_{it}, \omega_{it}) - V^0(l_{it-1}, k_{it}, \omega_{it})$ is strictly increasing in all its arguments, i.e., all the inputs are more productive in the current firm/industry than in the best alternative use. Therefore, the function is invertible with respect to the productivity shock ω_{it} and we can write the optimal exit/stay decision as a single-threshold condition:

$$d_{it} = I\left\{ \ \omega_{it} \geq \omega^*\left(l_{it-1}, k_{it}\right) \ \right\} \tag{3.7}$$

where the threshold function $\omega^*\left(.,.\right)$ is strictly decreasing in all its arguments.

Consider the PF $y_{it} = \alpha_L\ l_{it} + \alpha_K\ k_{it} + \omega_{it} + e_{it}$. In the estimation of this PF, we use the sample of firms that survived at period t: i.e., $d_{it} = 1$. Therefore, the error term in the estimation of the PF is $\omega_{it}^{d=1} + e_{it}$, where:

$$\omega_{it}^{d=1} \ \equiv \ \{\omega_{it} \mid d_{it} = 1\} \ = \ \{\omega_{it} \mid \omega_{it} \geq \omega^*\left(l_{i,t-1}, k_{it}\right)\} \tag{3.8}$$

Even if the productivity shock ω_{it} is independent of the state variables $(l_{i,t-1}, k_{it})$, the self-selected productivity shock $\omega_{it}^{d=1}$ will not be mean-independent of $(l_{i,t-1}, k_{it})$. That is,

$$
\begin{aligned}
E\left(\omega_{it}^{d=1} \mid l_{i,t-1}, k_{it}\right) \ &= \ E\left(\omega_{it} \mid l_{i,t-1}, k_{it}, d_{it} = 1\right) \\[6pt]
&= \ E\left(\omega_{it} \mid l_{i,t-1}, k_{it}, \omega_{it} \geq \omega^*\left(l_{i,t-1}, k_{it}\right)\right) \tag{3.9} \\[6pt]
&= \ \lambda\left(l_{i,t-1}, k_{it}\right)
\end{aligned}
$$

$\lambda\left(l_{i,t-1}, k_{it}\right)$ is the selection term. Therefore, the PF can be written as:

$$y_{it} = \quad \alpha_L \; l_{it} + \alpha_K \; k_{it} + \lambda\left(l_{i,t-1}, k_{it}\right) + \tilde{\omega}_{it} + e_{it} \qquad (3.10)$$

where $\tilde{\omega}_{it} \equiv \left\{\omega_{it}^{d=1} - \lambda\left(l_{i,t-1}, k_{it}\right)\right\}$ that, by construction, is mean-independent of $\left(l_{i,t-1}, k_{it}\right)$.

Ignoring the selection term $\lambda\left(l_{i,t-1}, k_{it}\right)$ introduces bias in our estimates of the PF parameters. The selection term is an increasing function of the threshold $\omega^*\left(l_{i,t-1}, k_{it}\right)$, and therefore it is decreasing in $l_{i,t-1}$ and k_{it}. Both l_{it} and k_{it} are negatively correlated with the selection term, but the correlation with the capital stock tend to be larger because the value of a firm depends strongly on its capital stock than on its "stock" of labor. Therefore, this selection problem tends to bias downward the estimate of the capital coefficient.

To provide an intuitive interpretation of this bias, first consider the case of very large firms. Firms with a large capital stock are very likely to survive, even if the firm receives a bad productivity shock. Therefore, for large firms, endogenous exit induces little censoring in the distribution of productivity shocks. Consider now the case of very small firms. Firms with a small capital stock have a large probability of exiting, even if their productivity shocks are not too negative. For small firms, exit induces a very significant left-censoring in the distribution of productivity, i.e., we only observe small firms with good productivity shocks and therefore with high levels of output. If we ignore this selection, we will conclude that firms with large capital stocks are not much more productive than firms with small capital stocks. But that conclusion is partly spurious because we do not observe many firms with low capital stocks that would have produced low levels of output if they had stayed.

This type of selection problem has been pointed out also by different authors who have studied empirically the relationship between firm growth and firm size. The relationship between firm size and firm growth has important policy implications. Mansfield (1962), Evans (1987), and Hall (1987) are seminal papers in that literature. Consider the regression equation:

$$\Delta s_{it} = \alpha + \beta \; s_{i,t-1} + \varepsilon_{it} \qquad (3.11)$$

where s_{it} represents the logarithm of a measure of firm size, e.g., the logarithm of capital stock, or the logarithm of the number of workers. Suppose that the exit decision at period t depends on firm size, $s_{i,t-1}$, and on a shock ε_{it}. More specifically,

$$d_{it} = I\left\{\; \varepsilon_{it} \geq \varepsilon^*\left(s_{i,t-1}\right)\; \right\} \qquad (3.12)$$

where $\varepsilon^*\left(.\right)$ is a decreasing function, i.e., smaller firms are more likely to exit. In a regression of Δs_{it} on $s_{i,t-1}$, we can use only observations from surviving firms. Therefore, the regression of Δs_{it} on $s_{i,t-1}$ can be represented using the equation $\Delta s_{it} = \alpha + \beta \; s_{i,t-1} + \varepsilon_{it}^{d=1}$, where $\varepsilon_{it}^{d=1} \equiv \left\{\varepsilon_{it}|d_{it}=1\right\} = \left\{\varepsilon_{it}|\varepsilon_{it} \geq \varepsilon^*\left(s_{i,t-1}\right)\right\}$. Thus,

$$\Delta s_{it} = \quad \alpha + \beta s_{i,t-1} + \lambda\left(s_{i,t-1}\right) + \tilde{\varepsilon}_{it} \qquad (3.13)$$

where $\lambda\left(s_{i,t-1}\right) \equiv E(\varepsilon_{it}|\varepsilon_{it} \geq \varepsilon^*\left(s_{i,t-1}\right))$, and $\tilde{\tilde{\varepsilon}}_{it} \equiv \left\{\varepsilon_{it}^{d=1}-\lambda\left(l_{i,t-1}, k_{it}\right)\right\}$ that, by construction, is mean-independent of firm size at $t-1$. The selection term $\lambda\left(s_{i,t-1}\right)$ is an increasing function of the threshold $\varepsilon^*\left(s_{i,t-1}\right)$, and therefore it is decreasing in firm size. If the selection term is ignored in the regression of Δs_{it} on $s_{i,t-1}$, then the OLS estimator of β will be downward biased. That is, it seems that smaller firms grow faster just because small firms that would like to grow slowly have exited the industry and they are not observed in the sample.

Mansfield (1962) already pointed out to the possibility of a selection bias due to endogenous exit. He used panel data from three US industries, steel, petroleum, and tires, over several periods. He tests the null hypothesis of $\beta = 0$, i.e., Gibrat's Law. Using only the subsample of surviving firms, he can reject Gibrat's Law in 7 of the 10 samples. Including also exiting firms and using the imputed values $\Delta s_{it} = -1$ for these firms, he rejects Gibrat's Law for only for 4 of the 10 samples. Of course, the main limitation of Mansfield's approach is that including exiting firms using the imputed values $\Delta s_{it} = -1$ does not correct completely for selection bias. But Mansfield's paper was written almost twenty years before Heckman's seminal contributions on sample selection in econometrics. Hall (1987) and Evans (1987) dealt with the selection problem using Heckman's two-step estimator. Both authors find that ignoring endogenous exit induces significant downward bias in β. However, they also find that after controlling for endogenous selection a la Heckman, the estimate of β is significantly lower than zero. They reject Gibrat's Law. A limitation of their approach is that their models do not have any exclusion restriction and identification is based on functional form assumptions, i.e., normality of the error term, and linear relationship between firm size and firm growth.

4. Estimation Methods

4.1. Using Input Prices as Instruments. If input prices, r_i, are observable, and they are not correlated with the productivity shock ω_i, then we can use these variables as instruments in the estimation of the PF. However, this approach has several important limitations. First, input prices are not always observable in some datasets, or they are only observable at the aggregate level but not at the firm level. Second, if firms in our sample use homogeneous inputs, and operate in the same output and input markets, we should not expect to find any significant cross-sectional variation in input prices. Time-series variation is not enough for identification. Third, if firms in our sample operate in different input markets, we may observe significant cross-sectional variation in input prices. However, this variation is suspicious of being endogenous. The different markets where firms operate can be also different in the average unobserved productivity of firms, and therefore $cov\left(\omega_i, r_i\right) \neq 0$, i.e., input prices not a valid instruments. In general, when there is cross-sectional variability

in input prices, can one say that input prices are valid instruments for inputs in a PF? Is $cov(\omega_i, r_i) = 0$? When inputs are firm-specific, it is commonly the case that input prices depend on the firm's productivity.

4.2. Panel Data: Fixed-Effects Estimators. Suppose that we have firm level panel data with information on output, capital and labor for N firms during T time periods. The Cobb-Douglas PF is:

$$y_{it} = \alpha_L \, l_{it} + \alpha_K \, k_{it} + \omega_{it} + e_{it} \tag{4.1}$$

Mundlak (1961) and Mundlak and Hoch (1965) are seminal studies in the use of panel data for the estimation of production functions. They consider the estimation of a production function of an agricultural product. They postulate the following assumptions:

Assumption PD-1: ω_{it} has the following variance-components structure: $\omega_{it} = \eta_i + \delta_t + \omega_{it}^*$. The term η_i is a time-invariant, firm-specific effect that may be interpreted as the quality of a fixed input such as managerial ability, or land quality. δ_t is an aggregate shock affecting all firms. And ω_{it}^* is an firm idiosyncratic shock.

Assumption PD-2: The amount of inputs depend on some other exogenous time varying variables, such that $var\left(l_{it} - \bar{l}_i\right) > 0$ and $var\left(k_{it} - \bar{k}_i\right) > 0$, where $\bar{l}_i \equiv T^{-1} \sum_{t=1}^{T} l_{it}$, and $\bar{k}_i \equiv T^{-1} \sum_{t=1}^{T} k_{it}$.

Assumption PD-3: ω_{it}^* is not serially correlated.

Assumption PD-4: The idiosyncratic shock ω_{it}^* is realized after the firm decides the amount of inputs to employ at period t. In the context of an agricultural PF, this shock may be interpreted as weather, or other random and unpredictable shock.

The Within-Groups estimator (WGE) or fixed-effects estimator of the PF is just the OLS estimator in the Within-Groups transformed equation:

$$(y_{it} - \bar{y}_i) = \alpha_L \left(l_{it} - \bar{l}_i\right) + \alpha_K \left(k_{it} - \bar{k}_i\right) + (\omega_{it} - \bar{\omega}_i) + (e_{it} - \bar{e}_i) \tag{4.2}$$

Under assumptions (PD-1) to (PD-4), the WGE is consistent. Under these assumptions, the only endogenous component of the error term is the fixed effect η_i. The transitory shocks ω_{it}^* and e_{it} do not induce any endogeneity problem. The WG transformation removes the fixed effect η_i.

It is important to point out that, for short panels (i.e., T fixed), the consistency of the WGE requires the regressors $x_{it} \equiv (l_{it}, k_{it})$ to be strictly exogenous. That is, for any (t, s):

$$cov\left(x_{it}, \omega_{is}^*\right) = cov\left(x_{it}, e_{is}\right) = 0 \tag{4.3}$$

Otherwise, the WG-transformed regressors $(l_{it} - \bar{l}_i)$ and $(k_{it} - \bar{k}_i)$ would be correlated with the error $(\omega_{it} - \bar{\omega}_i)$. This is why Assumptions (PD-3) and (PD-4) are necessary for the consistency of the OLS estimator.

However, it is very common to find that the WGE estimator provides very small estimates of α_L and α_K (see Grilliches and Mairesse, 1998). There are at least two factors that can explain this empirical regularity. First, though Assumptions (PD-2) and (PD-3) may be plausible for the estimation of agricultural PFs, they are very unrealistic for manufacturing firms. And second, the bias induced by measurement-error in the regressors can be exacerbated by the WG transformation. That is, the noise-to-signal ratio can be much larger for the WG transformed inputs than for the variables in levels. To see this, consider the model with only one input, say capital, and suppose that it is measured with error. We observe k_{it}^* where $k_{it}^* = k_{it} + e_{it}^k$, and e_{it}^k represents measurement error in capital and it satisfies the classical assumptions on measurement error. In the estimation of the PF in levels we have that:

$$Bias(\hat{\alpha}_L^{OLS}) = \frac{Cov(k, \eta)}{Var(k) + Var(e^k)} - \frac{\alpha_L \, Var(e^k)}{Var(k) + Var(e^k)} \tag{4.4}$$

If $Var(e^k)$ is small relative to $Var(k)$, then the (downward) bias introduced by the measurement error is negligible in the estimation in levels. In the estimation in first differences (similar to WGE, in fact equivalent when $T = 2$), we have that:

$$Bias(\hat{\alpha}_L^{WGE}) = -\frac{\alpha_L \, Var(\Delta e^k)}{Var(\Delta k) + Var(\Delta e^k)} \tag{4.5}$$

Suppose that k_{it} is very persistent (i.e., $Var(k)$ is much larger than $Var(\Delta k)$) and that e_{it}^k is not serially correlated (i.e., $Var(\Delta e^k) = 2 * Var(e^k)$). Under these conditions, the ratio $Var(\Delta e^k)/Var(\Delta k)$ can be large even when the ratio $Var(e^k)/Var(k)$ is quite small. Therefore, the WGE may be significantly downward biased.

4.3. Dynamic Panel Data: GMM Estimation. In the WGE described in previous section, the assumption of strictly exogenous regressors is very unrealistic. However, we can relax that assumption and estimate the PF using GMM method proposed by Arellano and Bond (1991). Consider the PF in first differences:

$$\Delta y_{it} = \alpha_L \, \Delta l_{it} + \alpha_K \, \Delta k_{it} + \Delta \delta_t + \Delta \omega_{it}^* + \Delta e_{it} \tag{4.6}$$

We maintain assumptions (PD-1), (PD-2), and (PD-3), but we remove assumption (PD-3). Instead, we consider the following assumption.

Assumption PD-5: There are adjustment costs in inputs (at least in one input). More formally, the reduced form equations for labor and capital are $l_{it} = f_L(l_{i,t-1}, k_{i,t-1}, \omega_{it})$ and

$k_{it} = f_K(l_{i,t-1}, k_{i,t-1}, \omega_{it})$, respectively, where either $l_{i,t-1}$ or $k_{i,t-1}$, or both, have non-zero partial derivatives in f_L and f_K.

Under these assumptions $\{l_{i,t-j}, k_{i,t-j}, y_{i,t-j} : j \geq 2\}$ are valid instruments in the PD in first differences. Identification comes from the combination of two assumptions: (1) serial correlation of inputs; and (2) no serial correlation in productivity shocks $\{\omega_{it}^*\}$. The presence of adjustment costs implies that the shadow prices of inputs vary across firms even if firms face the same input prices. This variability in shadow prices can be used to identify PF parameters. The assumption of no serial correlation in $\{\omega_{it}^*\}$ is key, but it can be tested using an LM test (see Arellano and Bond, 1991).

This GMM in first-differences approach has also its own limitations. In some applications, it is common to find unrealistically small estimates of α_L and α_K and large standard errors. (see Blundell and Bond, 2000). Overidentifying restrictions are typically rejected. Furthermore, the i.i.d. assumption on ω_{it}^* is typically rejected, and this implies that $\{x_{i,t-2}, y_{i,t-2}\}$ are not valid instruments. It is well-known that the Arellano-Bond GMM estimator may suffer of weak-instruments problem when the serial correlation of the regressors in first differences is weak (see Arellano and Bover, 1995, and Blundell and Bond, 1998). First difference transformation also eliminates the cross-sectional variation in inputs and it is subject to the problem of measurement error in inputs.

The weak-instruments problem deserves further explanation. For simplicity, consider the model with only one input, x_{it}. We are interested in the estimation of the PF:

$$y_{it} = \alpha \ x_{it} + \eta_i + \omega_{it}^* + e_{it} \tag{4.7}$$

where ω_{it}^* and e_{it} are not serially correlated. Consider the following dynamic reduced form equation for the input x_{it}:

$$x_{it} = \delta \ x_{i,t-1} + \lambda_1 \ \eta_i + \lambda_2 \ \omega_{it}^* \tag{4.8}$$

where δ, λ_1, and λ_2 are reduced form parameters, and $\delta \in [0, 1]$ captures the existence of adjustment costs. The PF in first differences is:

$$\Delta y_{it} = \alpha \ \Delta x_{it} + \Delta \omega_{it}^* + \Delta e_{it} \tag{4.9}$$

For simplicity, consider that the number of periods in the panel is $T = 3$. In this context, Arellano-Bond GMM estimator is equivalent to Anderson-Hsiao IV estimator (Anderson and Hsiao, 1981, 1982) where the endogenous regressor Δx_{it} is instrumented using $x_{i,t-2}$. This IV estimator is:

$$\hat{\alpha}_N = \frac{\sum_{i=1}^{N} x_{i,t-2} \ \Delta y_{it}}{\sum_{i=1}^{N} x_{i,t-2} \ \Delta x_{it}} \tag{4.10}$$

Under the assumptions of the model, we have that $x_{i,t-2}$ is orthogonal to the error $(\Delta\omega_{it}^* + \Delta e_{it})$. Therefore, $\hat{\alpha}_N$ identifies α if the (asymptotic) R-square in the auxiliary regression of Δx_{it} on $x_{i,t-2}$ is not zero.

By definition, the R-square coefficient in the auxiliary regression of Δx_{it} on $x_{i,t-2}$ is such that:

$$p\lim R^2 = \frac{Cov\left(\Delta x_{it}, x_{i,t-2}\right)^2}{Var\left(\Delta x_{it}\right)\ Var\left(x_{i,t-2}\right)} = \frac{\left(\gamma_2 - \gamma_1\right)^2}{2\left(\gamma_0 - \gamma_1\right)\gamma_0} \tag{4.11}$$

where $\gamma_j \equiv Cov\left(x_{it}, x_{i,t-j}\right)$ is the autocovariance of order j of $\{x_{it}\}$. Taking into account that $x_{it} = \frac{\lambda_1\,\eta_i}{1-\delta} + \lambda_2(\omega_{it} + \delta\,\omega_{i,t-1} + \delta^2\,\omega_{i,t-2} + ...)$, we can derive the following expressions for the autocovariances:

$$
\begin{aligned}
\gamma_0 &= \frac{\lambda_1^2\,\sigma_\eta^2}{(1-\delta)^2} + \frac{\lambda_2^2\,\sigma_\omega^2}{1-\delta^2} \\
\gamma_1 &= \frac{\lambda_1^2\,\sigma_\eta^2}{(1-\delta)^2} + \delta\,\frac{\lambda_2^2\,\sigma_\omega^2}{1-\delta^2} \\
\gamma_2 &= \frac{\lambda_1^2\,\sigma_\eta^2}{(1-\delta)^2} + \delta^2\,\frac{\lambda_2^2\,\sigma_\omega^2}{1-\delta^2}
\end{aligned}
\tag{4.12}
$$

Therefore, $\gamma_0 - \gamma_1 = (\lambda_2^2\sigma_\omega^2)/(1+\delta)$ and $\gamma_1 - \gamma_2 = \delta(\lambda_2^2\sigma_\omega^2)/(1+\delta)$. The R-square is:

$$
\begin{aligned}
R^2 &= \frac{\left(\delta\,\dfrac{\lambda_2^2\sigma_\omega^2}{1+\delta}\right)^2}{2\left(\dfrac{\lambda_2^2\sigma_\omega^2}{1+\delta}\right)\left(\dfrac{\lambda_1^2\,\sigma_\eta^2}{(1-\delta)^2} + \dfrac{\lambda_2^2\,\sigma_\omega^2}{1-\delta^2}\right)} \\[2mm]
&= \frac{\delta^2\left(1-\delta\right)^2}{2\left(1-\delta + (1+\delta)\,\rho\right)}
\end{aligned}
\tag{4.13}
$$

with $\rho \equiv \lambda_1^2\sigma_\eta^2/\lambda_2^2\sigma_\omega^2 \geq 0$. We have a problem of weak instruments and poor identification if this R-square coefficient is very small. It is simple to verify that this R-square is small both when adjustment costs are small (i.e., δ is close to zero) and when adjustment costs are large (i.e., δ is close to one). When using this IV estimator, large adjustments costs are bad news for identification because with δ close to one the first difference Δx_{it} is almost iid and it is not correlated with lagged input (or output) values. What is the maximum possible value of this R-square? It is clear that this R-square is a decreasing function of ρ. Therefore, the maximum R-square occurs for $\lambda_1^2\sigma_\eta^2 = \rho = 0$ (i.e., no fixed effects in the input demand). Then, $R^2 = \delta^2\left(1-\delta\right)/2$. The maximum value of this R-square is $R^2 = 0.074$ that occurs when $\delta = 2/3$. This is the upper bound for the R-square, but it is a too optimistic upper bound because it is based on the assumption of no fixed effects. For instance, a more realistic case for ρ is $\lambda_1^2\sigma_\eta^2 = \lambda_2^2\sigma_\omega^2$ and therefore $\rho = 1$. Then, $R^2 = \delta^2\left(1-\delta\right)^2/4$. The maximum value of this R-square is $R^2 = 0.016$ that occurs when $\delta = 1/2$.

Arellano and Bover (1995) and Blundell and Bond (1998) have proposed GMM estimators that deal with this weak-instrument problem. Suppose that at some period $t_i^* \leq 0$ (i.e., before the first period in the sample, $t = 1$) the shocks ω_{it}^* and e_{it} were zero, and input and output were equal to their firm-specific, steady-state mean values:

$$
\begin{aligned}
x_{it_i^*} &= \frac{\lambda_1 \eta_i}{1 - \delta} \\[2mm]
y_{it_i^*} &= \alpha \frac{\lambda_1 \eta_i}{1 - \delta} + \eta_i
\end{aligned}
\tag{4.14}
$$

Then, it is straightforward to show that for any period t in the sample:

$$
\begin{aligned}
x_{it} &= x_{it_i^*} + \lambda_2 \left(\omega_{it}^* + \delta \omega_{it-1}^* + \delta^2 \omega_{it-2}^* + ... \right) \\[2mm]
y_{it} &= y_{it_i^*} + \omega_{it}^* + \alpha \lambda_2 \left(\omega_{it}^* + \delta \omega_{it-1}^* + \delta^2 \omega_{it-2}^* + ... \right)
\end{aligned}
\tag{4.15}
$$

These expressions imply that input and output in first differences depend on the history of the i.i.d. shock $\{\omega_{it}^*\}$ between periods t_i^* and t, but they do not depend on the fixed effect η_i. Therefore, $cov(\Delta x_{it}, \eta_i) = cov(\Delta y_{it}, \eta_i) = 0$ and lagged first differences are valid instruments in the equation in levels. That is, for $j > 0$:

$$
\begin{aligned}
E \left(\Delta x_{it-j} \left[\eta_i + \omega_{it}^* + e_{it} \right] \right) = 0 &\quad \Rightarrow \quad E \left(\Delta x_{it-j} \left[y_{it} - \alpha x_{it} \right] \right) = 0 \\[2mm]
E \left(\Delta y_{it-j} \left[\eta_i + \omega_{it}^* + e_{it} \right] \right) = 0 &\quad \Rightarrow \quad E \left(\Delta y_{it-j} \left[y_{it} - \alpha x_{it} \right] \right) = 0
\end{aligned}
\tag{4.16}
$$

These moment conditions can be combined with the "standard" Arellano-Bond moment conditions to obtain a more efficient GMM estimator. The Arellano-Bond moment conditions are, for $j > 1$:

$$
\begin{aligned}
E \left(x_{it-j} \left[\Delta \omega_{it}^* + \Delta e_{it} \right] \right) = 0 &\quad \Rightarrow \quad E \left(x_{it-j} \left[\Delta y_{it} - \alpha \Delta x_{it} \right] \right) = 0 \\[2mm]
E \left(y_{it-j} \left[\Delta \omega_{it}^* + \Delta e_{it} \right] \right) = 0 &\quad \Rightarrow \quad E \left(y_{it-j} \left[\Delta y_{it} - \alpha \Delta x_{it} \right] \right) = 0
\end{aligned}
\tag{4.17}
$$

Based on Monte Carlo experiments and on actual data of UK firms, Blundell and Bond (2000) have obtained very promising results using this GMM estimator. Alonso-Borrego and Sanchez-Mangas (2001) have obtained similar results using Spanish data. The reason why this estimator works better than Arellano-Bond GMM is that the second set of moment conditions exploit cross-sectional variability in output and input. This has two implications. First, instruments are informative even when adjustment costs are larger and δ is close to one. And second, the problem of large measurement error in the regressors in first-differences is reduced.

Bond and Soderbom (2005) present a very interesting Monte Carlo experiment to study the actual identification power of adjustment costs in inputs. The authors consider a model with a Cobb-Douglas PF and quadratic adjustment cost with both deterministic and stochastic components. They solve firms' dynamic programming problem, simulate data of

inputs and output using the optimal decision rules, and use simulated data and Blundell-Bond GMM method to estimate PF parameters. The main results of their experiments are the following. When adjustment costs have only deterministic components, the identification is weak if adjustment costs are too low, or too high, or two similar between the two inputs. With stochastic adjustment costs, identification results improve considerably. Given these results, one might be tempted to "claim victory": if the true model is such that there are stochastic shocks (independent of productivity) in the costs of adjusting inputs, then the panel data GMM approach can identify with precision PF parameters. However, as Bond and Soderbom explain, there is also a negative interpretation of this result. Deterministic adjustment costs have little identification power in the estimation of PFs. The existence of shocks in adjustment costs which are independent of productivity seems a strong identification condition. If these shocks are not present in the "true model", the apparent identification using the GMM approach could be spurious because the "identification" would be due to the misspecification of the model. As we will see in the next section, we obtain a similar conclusion when using a control function approach.

Blundell and Bond (2001): Results

509 manufacturing firms; 1982-89				
Parameter	OLS-Levels	WG	AB-GMM	SYS-GMM
β_L	0.538	0.488	0.515	0.479
	(0.025)	(0.030)	(0.099)	(0.098)
β_K	0.266	0.199	0.225	0.492
	(0.032)	(0.033)	(0.126)	(0.074)
ρ	0.964	0.512	0.448	0.565
	(0.006)	(0.022)	(0.073)	(0.078)
Sargan (p-value)	-	-	0.073	0.032
m2	-	-	-0.69	-0.35
Constant RS (p-v)	0.000	0.000	0.006	0.641

4.4. Control Function Approaches. In a seminal paper, Olley and Pakes (1996) propose a control function approach to estimate PFs. Levinshon and Petrin (2003) have extended Olley-Pakes approach to contexts where data on capital investment presents significant censoring at zero investment.

Consider the Cobb-Douglas PF in the context of the following model of simultaneous equations:

$$(PF) \quad y_{it} = \alpha_L \, l_{it} + \alpha_K \, k_{it} + \omega_{it} + e_{it}$$

$$(LD) \quad l_{it} = f_L \left(l_{i,t-1}, k_{it}, \omega_{it}, r_{it} \right) \qquad\qquad (4.18)$$

$$(ID) \quad i_{it} = f_K \left(l_{i,t-1}, k_{it}, \omega_{it}, r_{it} \right)$$

where equations (LD) and (ID) represent the firms' optimal decision rules for labor and capital investment, respectively, in a dynamic decision model with state variables $(l_{i,t-1}, k_{it}, \omega_{it}, r_{it})$. The vector r_{it} represents input prices. Under certain conditions on this system of equations, we can estimate consistently α_L and α_K using a control function method.

Olley and Pakes consider the following assumptions:

Assumption OP-1: $f_K \left(l_{i,t-1}, k_{it}, \omega_{it}, r_{it} \right)$ is invertible in ω_{it}.

Assumption OP-2: There is not cross-sectional variation in input prices. For every firm i, $r_{it} = r_t$.

Assumption OP-3: ω_{it} follows a first order Markov process.

Assumption OP-4: Time-to-build physical capital. Investment i_{it} is chosen at period t but it is not productive until period $t + 1$. And $k_{it+1} = (1 - \delta)k_{it} + i_{it}$.

In Olley and Pakes model, lagged labor, $l_{i,t-1}$, is not a state variable, i.e., there a not labor adjustment costs, and labor is a perfectly flexible input. However, that assumption is not necessary for Olley-Pakes estimator. Here we discuss the method in the context of a model with labor adjustment costs.

Assumption OP-2 implies that the only unobservable variable in the investment equation that has cross-sectional variation across firms is the productivity shock ω_{it}. This restriction is crucial for OP method and for the related Levinshon-Petrin method. This imposes restrictions on the underlying model of market competition and inputs demands. For instance, this assumption implicitly establishes that firms operate in the same input markets and they do not have any monopsony power in these markets, e.g., no internal labor markets. Since a firm's input demand depends also on output price (or on the exogenous demand variables affecting product demand), assumption OP-2 also implies that firms operate in the same output market with either homogeneous goods or completely symmetric product differentiation. Note that these economics restrictions can be relaxed if the researcher has data on inputs prices at the firm level, i.e., r_{it} is observable.

Olley-Pakes method deals both with the simultaneity problem and with the selection problem due to endogenous exit. For the sake of clarity, we start describing here a version

of the method that does not deal with the selection problem. We will discuss later their approach to deal with endogenous exit.

The method proceeds in two-steps. The first step estimates α_L using a control function approach, and it relies on assumptions (*OP-1*) and (*OP-2*). This first step is the same with and without endogenous exit. The second step estimates α_K and it is based on assumptions (*OP-3*) and (*OP-4*). This second step is different when we deal with endogenous exit.

Step 1: Estimation of α_L. Assumptions (*OP-1*) and (*OP-2*) imply that $\omega_{it} = f_K^{-1}(l_{i,t-1}, k_{it}, i_{it}, r_t)$. Solving this equation into the PF we have:

$$
\begin{aligned}
y_{it} &= \alpha_L \; l_{it} + \alpha_K \; k_{it} + f_L^{-1}(l_{i,t-1}, k_{it}, i_{it}, r_t) + e_{it} \\
&= \alpha_L \; l_{it} + \phi_t(l_{i,t-1}, k_{it}, i_{it}) + e_{it}
\end{aligned}
\tag{4.19}
$$

where $\phi_t(l_{i,t-1}, k_{it}, i_{it}) \equiv \alpha_K \; k_{it} + f_L^{-1}(l_{i,t-1}, k_{it}, i_{it}, r_t)$. Without a parametric assumption on the investment equation f_K, equation (4.19) is a semiparametric partially linear model. The parameter α_L and the functions $\phi_1(.), \phi_2(.), ..., \phi_T(.)$ can be estimated using semiparametric methods. A possible semiparametric method is the kernel method in Robinson (1988). Instead, Olley and Pakes use polynomial series approximations for the nonparametric functions ϕ_t.

This method is a control function method. Instead of instrumenting the endogenous regressors, we include additional regressors that capture the endogenous part of the error term (i.e., proxy for the productivity shock). By including a flexible function in $(l_{i,t-1}, k_{it}, i_{it})$, we control for the unobservable ω_{it}. Therefore, α_L is identified if given $(l_{i,t-1}, k_{it}, i_{it})$ there is enough cross-sectional variation left in l_{it}. The key conditions for the identification of α_L are: (a) invertibility of $f_L(l_{i,t-1}, k_{it}, \omega_{it}, r_t)$ with respect to ω_{it}; (b) $r_{it} = r_t$, i.e., no cross-sectional variability in unobservables, other than ω_{it}, affecting investment; and (c) given $(l_{i,t-1}, k_{it}, i_{it}, r_t)$, current labor l_{it} still has enough sample variability. Assumption (c) is key, and it is the base for Ackerberg, Caves, and Frazer (2006) criticism (and extension) of Olley-Pakes approach.

Example 3: Consider Olley-Pakes model but with a parametric specification of the optimal investment equation (*ID*). More specifically, the inverse function f_K^{-1} has the following linear form:

$$
\omega_{it} = \gamma_1 \; i_{it} + \gamma_2 \; l_{i,t-1} + \gamma_3 \; k_{it} + r_{it}
\tag{4.20}
$$

Solving this equation into the PF, we have that:

$$
y_{it} = \alpha_L \; l_{it} + (\alpha_K + \gamma_3) \; k_{it} + \gamma_1 \; i_{it} + \gamma_2 \; l_{i,t-1} + (r_{it} + e_{it})
\tag{4.21}
$$

Note that current labor l_{it} is correlated with current input prices r_{it}. That is the reason why we need Assumption OP-2, i.e., $r_{it} = r_t$. Given that assumption we can control for the

unobserved r_t by including time-dummies. Furthermore, to identify α_L with enough preci-sion, there should not be high collinearity between current labor l_{it} and the other regressors $(k_{it}, i_{it}, l_{i,t-1})$.

Step 2: Estimation of α_K. Given the estimate of α_L in step 1, the estimation of α_K is based on Assumptions *(OP-3)* and *(OP-4)*, i.e., the Markov structure of the productivity shock, and the assumption of time-to-build productive capital. Since ω_{it} is first order Markov, we can write:

$$\omega_{it} = E[\omega_{it} \mid \omega_{i,t-1}] + \xi_{it} = h(\omega_{i,t-1}) + \xi_{it} \tag{4.22}$$

where ξ_{it} is an innovation which is mean independent of any information at $t-1$ or be-fore. $h(.)$ is some unknown function. Define $\phi_{it} \equiv \phi_t(l_{i,t-1}, k_{it}, i_{it})$, and remember that $\phi_t(l_{i,t-1}, k_{it}, i_{it}) = \alpha_K \, k_{it} + \omega_{it}$. Therefore, we have that:

$$
\begin{aligned}
\phi_{it} &= \alpha_K \, k_{it} + h(\omega_{i,t-1}) + \xi_{it} \\
&= \alpha_K \, k_{it} + h(\phi_{i,t-1} - \alpha_K \, k_{i,t-1}) + \xi_{it}
\end{aligned}
\tag{4.23}
$$

Though we do not know the true value of ϕ_{it}, we have consistent estimates of these values from step 1: i.e., $\hat{\phi}_{it} = y_{it} - \hat{\alpha}_L \, l_{it}$.[2]

If function $h(.)$ is nonparametrically specified, equation (4.23) is a partially linear model. However, it is not a "standard" partially linear model because the argument of the h function, $\phi_{i,t-1} - \alpha_K k_{i,t-1}$, is not observable, i.e., it depends on the unknown parameter α_K. To estimate $h(.)$ and α_K, Olley and Pakes propose a recursive version of the semiparametric method in the first step. Suppose that we consider a quadratic function for $h(.)$: i.e., $h(\omega) = \pi_1 \omega + \pi_2 \omega^2$. Then, given an initial value of α_K, we construct the variable $\hat{\omega}_{it}^{\alpha_K} = \hat{\phi}_{it} - \alpha_K k_{it}$, and estimate by OLS the equation $\hat{\phi}_{it} = \alpha_K k_{it} + \pi_1 \hat{\omega}_{it-1}^{\alpha_K} + \pi_2 (\hat{\omega}_{it-1}^{\alpha_K})^2 + \xi_{it}$. Given the OLS estimate of α_K, we construct new values $\hat{\omega}_{it}^{\alpha_K} = \hat{\phi}_{it} - \alpha_K k_{it}$ and estimate again α_K, π_1, and π_2 by OLS. We proceed until convergence. An alternative to this recursive procedure is the following Minimum Distance method. For instance, if the specification of $h(\omega)$ is quadratic, we have the regression model:

$$
\begin{aligned}
\hat{\phi}_{it} &= \alpha_K k_{it} + \pi_1 \hat{\phi}_{i,t-1} + \pi_2 \hat{\phi}_{i,t-1}^2 + (-\pi_1 \alpha_K) k_{i,t-1} + (\pi_2 \alpha_K^2) k_{i,t-1}^2 \\
&\quad + (-2\pi_2 \alpha_K) \hat{\phi}_{i,t-1} k_{i,t-1} + \xi_{it}
\end{aligned}
\tag{4.24}
$$

We can estimate the parameters α_K, π_1, π_2, $(-\pi_1 \alpha_K)$, $(\pi_2 \alpha_K^2)$, and $(-2\pi_2 \alpha_K)$ by OLS. This estimate of α_K can be very imprecise because the collinearity between the regressors. However, given the estimated vector of $\{\alpha_K, \pi_1, \pi_2, (-\pi_1 \alpha_K), (\pi_2 \alpha_K^2), (-2\pi_2 \alpha_K)\}$ and its

---

[2]In fact, $\hat{\phi}_{it}$ is an estimator of $\phi_{it} + e_{it}$, but this does not have any incidence on the consistency of the estimator.

variance-covariance matrix, we can obtain a more precise estimate of (α_K, π_1, π_2) by using minimum distance.

Example 4: Suppose that we consider a parametric specification for the stochastic process of $\{\omega_{it}\}$. More specifically, consider the AR(1) process $\omega_{it} = \rho\, \omega_{i,t-1} + \xi_{it}$, where $\rho \in [0,1)$ is a parameter. Then, $h\left(\omega_{i,t-1}\right) = \rho\omega_{i,t-1} = \rho(\phi_{i,t-1} - \alpha_K\, k_{i,t-1})$, and we can write:

$$\phi_{it} = \alpha_K\, k_{it} + \rho\, \phi_{i,t-1} + (-\rho\alpha_K)\, k_{i,t-1} + \xi_{it} \qquad (4.25)$$

we can see that a regression of ϕ_{it} on k_{it}, $\phi_{i,t-1}$ and $k_{i,t-1}$ identifies (in fact, over-identifies) α_K and ρ.

Time-to build is a key assumption for the consistency of this method. If new investment at period t is productive at the same period, then we have that: $\phi_{it} = \alpha_K\, k_{i,t+1} + h\left(\phi_{i,t-1} - \alpha_K\, k_{it}\right) + \xi_{it}$. Now, the regressor $k_{i,t+1}$ depends on investment at period t and therefore it is correlated with the innovation in productivity ξ_{it}.

Levinshon and Petrin (2003) propose a related control function method. The main difference with OP method is that Levinshon and Petrin (LP) use the demand function for intermediate inputs instead of the investment equation to invert out unobserved productivity. The consider a Cobb-Douglas production function in terms of labor, capital, and intermediate inputs (materials):

$$y_{it} = \alpha_L\, l_{it} + \alpha_K\, k_{it} + \alpha_M\, m_{it} + \omega_{it} + e_{it} \qquad (4.26)$$

The investment equation is replaced with the intermediate input demand:

$$m_{it} = f_M\left(l_{i,t-1}, k_{it}, \omega_{it}, r_{it}\right) \qquad (4.27)$$

Levinshon and Petrin maintain Assumptions OP-2 to OP-4, and replace Assumption OP-1 of monotonicity (invertibility) of the investment equation with a similar assumption for the intermediate input demand.

Assumption LP-1: $f_M\left(l_{i,t-1}, k_{it}, \omega_{it}, r_{it}\right)$ is invertible in ω_{it}.

Similarly as for the Olley-Pakes method, the key identification restriction in Levinshon-Petrin method is that the only unobservable variable in the intermediate input demand equation that has cross-sectional variation across firms is the productivity shock ω_{it}, i.e., *Assumption OP-2:* There is not cross-sectional variation in input prices such that $r_{it} = r_t$ for every firm i.

LP method also proceeds in two-steps. The first step consists in the least squares estimation of the parameter α_L and the nonparametric functions $\{\phi_t(.) : t = 1, 2, ..., T\}$ in the semiparametric regression equation:

$$y_{it} = \alpha_L\, l_{it} + \phi_t(l_{i,t-1}, k_{it}, m_{it}) + e_{it} \qquad (4.28)$$

where $\phi_t(l_{i,t-1}, k_{it}, m_{it}) = \alpha_K\ k_{it} + f_M^{-1}(l_{i,t-1}, k_{it}, m_{it}, r_t)$ and f_M^{-1} represents the inverse function of the intermediate input demand with respect to productivity. The second step is also similar to OP's second step but in the model with the intermediate input. More specifically, the estimates of α_L and ϕ_t are plugged-in, and a least squares is applied to the estimation of the parameters α_K and α_M and function $h(.)$ in the regression equation:

$$\phi_{it} \;=\; \alpha_K\ k_{it} + \alpha_M\ m_{it} + h\left(\phi_{i,t-1} - \alpha_K\ k_{i,t-1} - \alpha_M\ m_{i,t-1}\right) + \xi_{it} \tag{4.29}$$

There are several advantages of using the intermediate input

4.5. Ackerberg-Caves-Frazer Critique. Under Assumptions $(OP\text{-}1)$ and $(OP\text{-}2)$, we can invert the investment equation to obtain the productivity shock $\omega_{it} = f_K^{-1}(l_{i,t-1}, k_{it}, i_{it}, r_t)$. Then, we can solve the expression into the labor demand equation, $l_{it} = f_L\left(l_{i,t-1}, k_{it}, \omega_{it}, r_t\right)$, to obtain the following relationship:

$$l_{it} = f_L\left(l_{i,t-1},\ k_{it},\ f_K^{-1}(l_{i,t-1}, k_{it}, i_{it}, r_t),\ r_t\right) = G_t\left(l_{i,t-1}, k_{it}, i_{it}\right) \tag{4.30}$$

This expression shows an important implication of Assumptions $(OP\text{-}1)$ and $(OP\text{-}2)$. For any cross-section t, there should be a deterministic relationship between employment at period t and the observable state variables $(l_{i,t-1}, k_{it}, i_{it})$. In other words, once we condition on the observable variables $(l_{i,t-1}, k_{it}, i_{it})$, employment at period t should not have any cross-sectional variability. It should be constant. This implies that in the regression in step 1, $y_{it} = \alpha_L\ l_{it} + \phi_t(l_{i,t-1}, k_{it}, i_{it}) + e_{it}$, it should not be possible to identify α_L because the regressor l_{it} does not have any sample variability that is independent of the other regressors $(l_{i,t-1}, k_{it}, i_{it})$.

Example 5: The problem can be illustrated more clearly by using linear functions for the optimal investment and labor demand. Suppose that the inverse function f_K^{-1} is $\omega_{it} = \gamma_1 i_{it} + \gamma_2\ l_{i,t-1} + \gamma_3\ k_{it} + \gamma_4 r_t$; and the labor demand equation is $l_{it} = \delta_1 l_{i,t-1} + \delta_2 k_{it} + \delta_3 \omega_{it} + \delta_4 r_t$. Then, solving the inverse function f_K^{-1} into the production function, we get:

$$y_{it} = \alpha_L\ l_{it} + (\alpha_K + \gamma_3)\ k_{it} + \gamma_1\ i_{it} + \gamma_2\ l_{i,t-1} + (\gamma_4 r_t + e_{it}) \tag{4.31}$$

And solving the inverse function f_K^{-1} into the labor demand, we have that:

$$l_{it} = (\delta_1 + \delta_3\gamma_2)l_{i,t-1} + (\delta_2 + \delta_3\gamma_3)k_{it} + \delta_3\gamma_1 i_{it} + (\delta_4 + \delta_3\gamma_4)r_t \tag{4.32}$$

Equation (4.32) shows that there is perfect collinearity between l_{it} and $(l_{i,t-1}, k_{it}, i_{it})$ and therefore it should not be possible to estimate α_L in equation (4.31). Of course, in the data we will find that l_{it} has some cross-sectional variation independent of $(l_{i,t-1}, k_{it}, i_{it})$. Equation (4.32) shows that if that variation is present it is because input prices r_{it} have cross-sectional variation. However, that variation is endogenous in the estimation of equation (4.31) because

the unobservable r_{it} is part of the error term. That is, if there is apparent identification, that identification is spurious.

After pointing out this important problem in Olley-Pakes model and method, Ackerberg-Caves-Frazer study different that could be combined with Olley-Pakes control function approach to identify the parameters of the PF. For identification, we need some source of exogenous variability in labor demand that is independent of productivity and that does not affect capital investment. Ackerberg-Caves-Frazer discuss several possible arguments/assumptions that could incorporate in the model this kind of exogenous variability.

Consider a model with same specification of the PF, but with the following specification of labor demand and optimal capital investment:

$$
\begin{aligned}
(LD') \quad & l_{it} = f_L\left(l_{i,t-1}, k_{it}, \omega_{it}, r_{it}^L\right) \\
(ID') \quad & i_{it} = f_K\left(l_{i,t-1}, k_{it}, \omega_{it}, r_{it}^K\right)
\end{aligned}
\tag{4.33}
$$

Ackerberg-Caves-Frazer propose to maintain Assumptions (OP-1), (OP-3), and (OP-4), and to replace Assumption (OP-2) by the following assumption.

Assumption ACF: Unobserved input prices r_{it}^L and r_{it}^K are such that conditional on $(t, i_{it}, l_{i,t-1}, k_{it})$: (a) r_{it}^L has cross-sectional variation, i.e., $var(r_{it}^L \mid t, i_{it}, l_{i,t-1}, k_{it}) > 0$; and (b) r_{it}^L and r_{it}^K are independently distributed.

There are different possible interpretations of Assumption ACF. The following list of conditions (a) to (d) is a group of economic assumptions that generate Assumption ACF: (a) the capital market is perfectly competitive and the price of capital is the same for every firm $(r_{it}^K = r_t^K)$; (b) there are internal labor markets such that the price of labor has cross sectional variability; (c) the realization of the cost of labor r_{it}^L occurs after the investment decision takes place, and therefore r_{it}^L does not affect investment; and (d) the idiosyncratic labor cost shock r_{it}^L is not serially correlated such that lagged values of this shock are not state variables for the optimal investment decision. Aguirregabiria and Alonso-Borrego (2008) consider similar assumptions for the estimation of a production function with physical capital, permanent employment, and temporary employment.

4.5.1. *Other identifying conditions: Quasi-fixed inputs.* Consider a CD-PF with labor and capital as only inputs. Suppose that OP assumptions hold such that ℓ_{it} is perfectly collinear with $\phi_t(\ell_{i,t-1}, k_{it}, i_{it})$. If both capital and labor are quasi-fixed inputs, then it is possible to use a control function method in the spirit of OP or LP to identify/estimate β_L and β_K. Or in other words, this model has moment conditions that identify β_L and β_K (Wooldridge, EL 2009).

In the first step we have:

$$y_{it} = \beta_L \, \ell_{it} + \phi_t(\ell_{i,t-1}, k_{it}, i_{it}) + e_{it}$$

$$= \beta_L \, g_t(\ell_{i,t-1}, k_{it}, i_{it}) + \phi_t(\ell_{i,t-1}, k_{it}, i_{it}) + e_{it}$$

$$= \psi_t(\ell_{i,t-1}, k_{it}, i_{it}) + e_{it}$$

In this first step, we estimate $\psi_t(\ell_{i,t-1}, k_{it}, i_{it})$ nonparametrically. In the second step, given ψ_{it}, and taking into account that $\psi_{it} = \beta_L \, \ell_{it} + \beta_K \, k_{it} + \omega_{it}$, and $\omega_{it} = h\,(\omega_{i,t-1}) + \xi_{it}$, we have that:

$$\psi_{it} = \beta_L \, \ell_{it} + \beta_K \, k_{it} + h\,(\psi_{it} - \beta_L \, \ell_{it-1} + \beta_K \, k_{it-1}) + \xi_{it}$$

In this second step, ℓ_{it} is correlated with ξ_{it}, but $(k_{it}, \psi_{it}, \ell_{it-1}, \, k_{it-1})$ are not, and (ℓ_{it-2}, k_{it-2}) can be used to instrument ℓ_{it}. This approach is in the same spirit as the Dynamic Panel Data (DPD) methods of Arellano-Bond and Blundell-Bond. This approach cannot be applied if some inputs (e.g., materials) are perfectly flexible. The PF coefficient parameter of the flexible inputs cannot be identified from the moment conditions in the second step.

4.5.2. *Other identifying conditions: F.O.C. for flexible inputs.* Klette & Grilliches (1996), Doraszelski & Jaumandreu (2013), and Gandhi, Navarro, & Rivers (2013) propose combining conditions from the PF with conditions from the demand of variable inputs. This approach requires the price of the variable input to be observable to the researcher, though this price may not have cross-sectional variation across firms.

Note that in the LP method, the function that relates m_{it} with the state variables is just the condition "VMP of materials equal to price of materials". The parameters in this condition are the same as in the PF. This approach takes these restrictions into account.

For the CD-PF, with materials as flexible input, we have that:

$$(PF) \quad y_{it} = \beta_L \, \ell_{it} + \beta_K \, k_{it} + \beta_M \, m_{it} + \omega_{it} + e_{it}$$

$$(FOC) \quad p_t - p_t^M = \ln(\beta_M) + \beta_L \, \ell_{it} + \beta_K \, k_{it} + (\beta_M - 1)m_{it} + \omega_{it}$$

The difference between these two equations is:

$$\ln(s_{it}^M) = \ln(\beta_M) + e_{it}$$

where $s_{it}^M \equiv \dfrac{P_t^M M_{it}}{P_t Y_{it}}$ is the ratio between material expenditures and revenue. The parameter(s) of the flexible inputs are identified from the expensiture-share equations. The parameter(s) of the quasi-fixed inputs are identified using the dynamic conditions described above.

Gandhi, Navarro, & Rivers (2013) show that this approach can be extended in two important way: (1) To a nonparametric specification of the production function: $y_{it} =$

$f\left(\ell_{it}, k_{it}, m_{it}\right) + \omega_{it} + e_{it}$; (2) - To a model with monopolistic competition (instead of perfect competition) with and isoelastic product demand. This approach relies on two strong assumptions. There is not any bias or missing parameter in the MC of the flexible input. Suppose that $MC_{Mt} = P_t^M \ \tau$, then our estimate of β_M will actually estimate $\beta_M \ \tau$. In its current version, this method cannot incorporate oligopoly competition in the product market, only a limited form of monopolistic competition.

4.6. Olley and Pakes on Endogenous Selection. Olley and Pakes (1996) show that there is a structure that permits to control for selection bias without a parametric assumption on the distribution of the unobservables. Before describing the approach proposed by Olley and Pakes, it will be helpful to describe some general features of semiparametric selection models.

Consider a selection model with outcome equation,

$$y_i = \begin{cases} x_i \ \beta + \varepsilon_i & \text{if} \ d_i = 1 \\ \\ \text{unobserved} & \text{if} \ d_i = 0 \end{cases} \tag{4.34}$$

and selection equation

$$d_i = \begin{cases} 1 & \text{if} \ h(z_i) - u_i \geq 0 \\ \\ 0 & \text{if} \ h(z_i) - u_i < 0 \end{cases} \tag{4.35}$$

where x_i and z_i are exogenous regressors; (u_i, ε_i) are unobservable variables independently distributed of (x_i, z_i); and $h(.)$ is a real-valued function. We are interested in the consistent estimation of the vector of parameters β. We would like to have an estimator that does not rely on parametric assumptions on the function h or on the distribution of the unobservables.

The outcome equation can be represented as a regression equation: $y_i = x_i \ \beta + \varepsilon_i^{d=1}$, where $\varepsilon_i^{d=1} \equiv \{\varepsilon_i | d_i = 1\} = \{\varepsilon_i | u_i \leq h(z_i)\}$. Or similarly,

$$y_i = x_i\beta + E(\varepsilon_i^{d=1}|x_i, z_i) + \tilde{\varepsilon}_i \tag{4.36}$$

where $E(\varepsilon_i^{d=1}|x_i, z_i)$ is the selection term. The new error term, $\tilde{\varepsilon}_i$, is equal to $\varepsilon_i^{d=1} - E(\varepsilon_i^{d=1}|x_i, z_i)$ and, by construction, is mean independent of (x_i, z_i). The selection term is equal to $E\left(\varepsilon_i \mid x_i, z_i, u_i \leq h(z_i)\right)$. Given that u_i and ε_i are independent of (x_i, z_i), it is simple to show that the selection term depends on the regressors only through the function $h(z_i)$: i.e., $E\left(\varepsilon_i \mid x_i, z_i, u_i \leq h(z_i)\right) = g(h(z_i))$. The form of the function g depends on the distribution of the unobservables, and it is unknown if we adopt a nonparametric specification of that distribution. Therefore, we have the following partially linear model: $y_i = x_i\beta + g(h(z_i)) + \tilde{\varepsilon}_i$.

Define the *propensity score* P_i as:

$$P_i \equiv \Pr\left(d_i = 1 \mid z_i\right) = F_u\left(h(z_i)\right) \tag{4.37}$$

where F_u is the CDF of u. Note that $P_i = E\left(d_i \mid z_i\right)$, and therefore we can estimate propensity scores nonparametrically using a Nadaraya-Watson kernel estimator or other nonparametric methods for conditional means. If u_i has unbounded support and a strictly increasing CDF, then there is a one-to-one invertible relationship between the propensity score P_i and $h(z_i)$. Therefore, the selection term $g(h(z_i))$ can be represented as $\lambda(P_i)$, where the function λ is unknown. The selection model can be represented using the partially linear model:

$$y_i = x_i\beta + \lambda(P_i) + \tilde{\varepsilon}_i. \tag{4.38}$$

A sufficient condition for the identification of β (without a parametric assumption on λ) is that $E\left(x_i\, x_i' \mid P_i\right)$ has full rank. Given equation (4.38) and nonparametric estimates of propensity scores, we can estimate β and the function λ using standard estimators for partially linear model such as the kernel estimator in Robinson (1988), or alternative estimators as discussed in Yatchew (2003).

Now, we describe Olley-Pakes procedure for the estimation of the production function taking into account endogenous exit. The first step of the method (i.e., the estimation of α_L) is not affected by the selection problem because we are controlling for ω_{it} using a control function approach. However, there is endogenous selection in the second step of the method. For simplicity consider that the productivity shock follows an AR(1) process: $\omega_{it} = \rho\, \omega_{i,t-1} - \xi_{it}$. Then, the "outcome" equation is:

$$\phi_{it} = \begin{cases} \alpha_K\ k_{it} + \rho\ \phi_{i,t-1} + (-\rho\alpha_K)\ k_{i,t-1} + \xi_{it} & \text{if}\quad d_{it} = 1 \\[2mm] \text{unobserved} & \text{if}\quad d_i = 0 \end{cases} \tag{4.39}$$

The exit/stay decision is: $\{d_{it} = 1\}$ iff $\{\omega_{it} \geq \omega^*(l_{it-1}, k_{it})\}$. Taking into account that $\omega_{it} = \rho\omega_{i,t-1} + \xi_{it}$, and that $\omega_{i,t-1} = \phi_{i,t-1} - \alpha_K\, k_{it-1}$, we have that the condition $\{\omega_{it} \geq \omega^*(l_{it-1}, k_{it})\}$ is equivalent to $\{\xi_{it} \leq \omega^*(l_{it-1}, k_{it}) - \rho(\phi_{i,t-1} - \alpha_K k_{it-1})\}$. Then, it is convenient to represent the exit/stay equation as:

$$d_{it} = \begin{cases} 1 & \text{if}\quad \xi_{it} \leq h(l_{it-1}, k_{it}, \phi_{i,t-1}, k_{it-1}) \\[2mm] 0 & \text{if}\quad \xi_{it} > h(l_{it-1}, k_{it}, \phi_{i,t-1}, k_{it-1}) \end{cases} \tag{4.40}$$

where $h(l_{it-1}, k_{it}, \phi_{i,t-1}, k_{it-1}) \equiv \omega^*(l_{it-1}, k_{it}) - \rho(\phi_{i,t-1} - \alpha_K k_{it-1})$. The propensity score is $P_{it} \equiv E\left(d_{it} \mid l_{it-1}, k_{it}, \phi_{i,t-1}, k_{it-1}\right)$. And the equation controlling for selection is:

$$\phi_{it} = \alpha_K k_{it} + \rho\phi_{i,t-1} + (-\rho\alpha_K)\, k_{i,t-1} + \lambda\left(P_{it}\right) + \tilde{\xi}_{it} \tag{4.41}$$

where, by construction, $\tilde{\xi}_{it}$ is mean independent of k_{it}, k_{it-1}, $\phi_{i,t-1}$, and P_{it}. And we can estimation equation (4.41) using standard methods for partially linear models.

5. Innovation and productivity growth: Production functions

5.1. What Determines Productivity? Large and persistent differences in TFP across firms. Ubiquitous, even within narrowly defined industries and products. **Large:** 90th to 10th percentile TFP ratios: $\frac{A_{90th}}{A_{10th}}$. U.S. manufacturing, **average** within 4-digit SIC industries = **1.92** (Syverson, 2004). Denmark: average = **3.75** (Fox and Smeets, 2011). China or India, **average > 5** (Hsieh & Klenow, 2009). **Persistent:** AR(1) of log-TFP with annual frequency: autoregressive coefficients between 0.6 to 0.8. **It matters:** Higher productivity producers are more likely to survive.

Why firms differ in their productivity levels? What supports such large productivity differences in equilibrium? Can producers control the factors that influence productivity or are they purely external effects of the environment? If firms can partly control their TFP, what type of choices increase it?

Why dispersion is possible in equilibrium? Let the profit of a firm be:

$$\pi_i = R(q_i, A_i, d) - C(q_i, A_i, w) - F$$

$R(q_i, A_i, d)$ = Revenue function. d = State of the industry. $C(q_i, A_i, w)$ = Cost function. w = Input prices. F = fixed costs. **Key condition:** either $R(.)$ is strictly concave in q_i, or $C(.)$ is strictly convex in q_i. [The variable profit function is strictly concave].

Example: Perfect competition [or Bertrand competition with homogeneous product]. $R = P\, q_i$ is linear in q_i. We need $C(.)$ to be strictly convex. i.e., DRS in variable inputs.

Example: Cournot competition or Bertrand competition with differentiated product. R is strictly concave in q_i (downward sloping demand). So we can have either CRS or DRS.

Equilibrium can be described by two types of conditions. At the **intensive margin**, optimal $q_i^* = q^*[A_i, d, w]$ is such that:

$$MR_i \equiv \frac{\partial R(q_i, A_i, d)}{\partial q_i} = \frac{\partial C(q_i, A_i, w)}{\partial q_i} \equiv MC_i$$

At the **extensive margin**, a firm is active in the market iff:

$$R(q^*[A_i, d, w], A_i, d) - C(q^*[A_i, d, w], A_i, w) - F \geq 0$$

If variable profit is strictly concave, this equilibrium can support firms with different TFPs, A_i. It is not optimal for the firm with highest TFP to provide all the output in the industry. Firms with different TFPs (above a certain threshold value) operate in the same market.

How can a firm affect its TFP? (HR) Managerial Practices. (Bloom & Van Reenen, 2007; Ichniowski and Shaw, 2003). Learning-by-Doing (Benkard, 2000). Organizational structure

(vertical integration vs outsourcing). Higher-Quality (Labor and Capital) inputs. **Adoption of new (IT) technologies.** (Brynjolfsson et al., 2008). **Investment in R&D.** Long literature linking R&D investment and productivity. **Innovation**. Many firms undertake both process and product innovation without formally reporting R&D spending.

Innovation, R&D, and TFP. Multiple studies show evidence that R&D and innovation are very (the most?) important factors to explain firm heterogeneity in TFP level and growth. As usual, the main difficulty in these studies comes from separating causation from correlation. For the rest of this lecture, we review models, methods, and datasets in different empirical applications dealing with the causal effect of R&D and innovations on TFP.

But still we will have to address the question **"why firms have different propensities to innovate / invest in R&D?"**, e.g., managerial talent, competition, spillovers, ... What factors determine how large innovative activity will be? Can we predict whether product or process innovation will dominate, based on market features?

5.2. Application: Olley & Pakes on the US Telecom industry. US Telecom. equipment industry: 1974-1987. Technological change and deregulation. Elimination of barriers to entry. Antitrust decisions against AT&T: The Consent Decree (implemented in 1984) –> divestiture of AT&T. Substantial entry/exit of plants.

Data: US Census of manufacturers.

TABLE VI

ALTERNATIVE ESTIMATES OF PRODUCTION FUNCTION PARAMETERS[a]

(STANDARD ERRORS IN PARENTHESES)

Sample:	Balanced Panel		Full Sample[c, d]						
								Nonparametric F_ω	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Estimation Procedure	Total	Within	Total	Within	OLS	Only P	Only h	Series	Kernel
Labor	.851	.728	.693	.629	.628			.608	
	(.039)	(.049)	(.019)	(.026)	(.020)			(.027)	
Capital	.173	.067	.304	.150	.219	.355	.339	.342	.355
	(.034)	(.049)	(.018)	(.026)	(.018)	(.02)	(.03)	(.035)	(.058)
Age	.002	−.006	−.0046	−.008	−.001	−.003	.000	−.001	.010
	(.003)	(.016)	(.0026)	(.017)	(.002)	(.002)	(.004)	(.004)	(.013)
Time	.024	.042	.016	.026	.012	.034	.011	.044	.020
	(.006)	(.017)	(.004)	(.017)	(.004)	(.005)	(.01)	(.019)	(.046)
Investment	—	—	—	—	.13	—	—	—	—
					(.01)				
Other Variables	—	—	—	—	—	Powers of P	Powers of h	Full Polynomial in P and h	Kernel in P and h
# Obs.[b]	896	896	2592	2592	2592	1758	1758	1758	1758

Going from OLS balanced panel to OLS full sample almost doubles β_K and reduces β_L by 20%. [Importance of endogenous exit]. Controlling for simultaneity further increases β_K and reduces β_L.

TABLE XI

DECOMPOSITION OF PRODUCTIVITY[a]

(EQUATION (16))

Year	p_t	$\bar{p}_t$	$\Sigma_i \Delta s_{it} \Delta p_{it}$	$\rho(p_t, k_t)$
1974	1.00	0.90	0.01	−0.07
1975	0.72	0.66	0.06	−0.11
1976	0.77	0.69	0.07	−0.12
1977	0.75	0.72	0.03	−0.09
1978	0.92	0.80	0.12	−0.05
1979	0.95	0.84	0.12	−0.05
1980	1.12	0.84	0.28	−0.02
1981	1.11	0.76 '	0.35	0.02
1982	1.08	0.77	0.31	−0.01
1983	0.84	0.76	0.08	−0.07
1984	0.90	0.83	0.07	−0.09
1985	0.99	0.72	0.26	0.02
1986	0.92	0.72	0.20	0.03
1987	0.97	0.66	0.32	0.10

5.3. Levinshon & Petrin on Chilean manufacturing. The main difference with OP method is that LP use the demand function for intermediate inputs instead of the investment equation to invert out unobserved productivity. Two main motivations. (1) Investment can be responsive to more persistent shocks in TFP; materials is responsive to every shock in TFP. (2) In some datasets **Zero Investment** accounts for a large fraction of the data. At $i_{it} = 0$ (corner solution / extensive margin) there is not invertibility between i_{it} and ω_{it}. Problems: loss of efficiency; missing estimates of TFP for many observations.

They consider a Cobb-Douglas production function in terms of labor, capital, and intermediate inputs (materials):

$$y_{it} = \beta_L \, \ell_{it} + \beta_K \, k_{it} + \beta_M \, m_{it} + \omega_{it} + e_{it}$$

Investment equation is replaced with demand for materials:

$$m_{it} = f_M \left(\ell_{i,t-1}, k_{it}, \omega_{it}, r_{it} \right)$$

Assumption LP-1: $f_M \left(\ell_{i,t-1}, k_{it}, \omega_{it}, r_{it} \right)$ is invertible in ω_{it}.

They maintain OP-2 [No other unobservables; $r_{it} = r_t$], OP-3 [Markov TFP], and OP-4 [Time-to-build]. Least squares estimation of parameter β_L and the nonparametric functions $\{\phi_t(.) : t = 1, 2, ..., T\}$ in regression equation:

$$y_{it} \;=\; \beta_L \, \ell_{it} + \phi_t(\ell_{i,t-1}, k_{it}, m_{it}) + e_{it}$$

$\phi_t(\ell_{i,t-1}, k_{it}, m_{it}) = \beta_K \, k_{it} + \beta_M m_{it} + f_M^{-1}(\ell_{i,t-1}, k_{it}, m_{it}, r_t)$ and f_M^{-1} is the inverse function of f_M with respect to ω_{it}.

The second step is also similar to OP's second step but in the model with the intermediate input. ϕ_{it} is estimated in 1st step; and $\phi_{it} = \beta_K \, k_{it} + \beta_M \, m_{it} + \omega_{it}$. Then,

$$\phi_{it} \;=\; \beta_K \, k_{it} + \beta_M \, m_{it} + h \left(\phi_{i,t-1} - \beta_K \, k_{i,t-1} - \beta_M \, m_{i,t-1} \right) + \xi_{it}$$

Important difference with OP: In this second step $E(m_{it} \ \xi_{it}) \neq 0$, i.e., materials m_{it} is

endogenous.

LP propose two approaches: **"unrestricted method"**: instrument m_{it} with its lagged values [see GNR (2013) criticism]; **"restricted method"**: under statis input, price-taking: $\beta_M =$ Cost of materials/Revenue.

Empirical application. Plant-level data from 8 different Chilean manufacturing industries: 1979-1985 [Pinochet period].

Var input shares

TABLE 3
Average Nominal Revenue Shares (Percentage), 1979-85

Industry	Unskilled	Skilled	Materials	Fuels	Electricity
Metals	15.2	8.3	44.9	1.6	1.7
Textiles	13.8	6.0	48.2	1.0	1.6
Food Products	12.1	3.5	60.3	2.1	1.3
Beverages	11.3	6.8	45.6	1.8	1.5
Other Chemicals	18.9	10.1	37.8	1.7	0.7
Printing & Pub.	19.8	10.7	40.1	0.5	1.3
Wood Products	20.6	5.3	47.0	3.0	2.4
Apparel	14.0	4.9	52.4	0.9	0.3

Zeroes

TABLE 2
Percent of Usable Observations, 1979-85

Industry	Investment	Fuels	Materials	Electricity
Metals	44.8	63.1	99.9	96.5
Textiles	41.2	51.2	99.9	97.0
Food Products	42.7	78.0	99.8	88.3
Beverages	44.0	73.9	99.8	94.1
Other Chemicals	65.3	78.4	100	96.5
Printing & Pub.	39.0	46.4	99.9	96.8
Wood Products	35.9	59.3	99.7	93.8
Apparel	35.2	34.5	99.9	97.2

Zeroes

TABLE 4
Unrestricted and Restricted Parameter Estimates for 8 Industries
(Bootstrapped Standard Errors in Parentheses)

Input	Industry (ISIC Code)							
	311	381	321	331	352	322	342	313
Unskilled labor								
	0.138	0.164	0.138	0.206	0.137	0.163	0.192	0.087
	(0.010)	(0.032)	(0.027)	(0.035)	(0.039)	(0.044)	(0.048)	(0.082)
Skilled labor								
	0.053	0.185	0.139	0.136	0.254	0.125	0.161	0.164
	(0.008)	(0.017)	(0.030)	(0.032)	(0.036)	(0.038)	(0.036)	(0.087)
Materials								
	0.703	0.587	0.679	0.617	0.567	0.621	0.483	0.626
	(0.013)	(0.017)	(0.019)	(0.022)	(0.045)	(0.020)	(0.028)	(0.075)
Fuels								
	0.023	0.024	0.041	0.018	0.004	0.0162	0.053	0.087
	(0.004)	(0.008)	(0.012)	(0.018)	(0.020)	(0.016)	(0.014)	(0.027)
Capital								
unrestricted	0.13	0.09	0.08	0.18	0.17	0.10	0.21	0.08
	(0.032)	(0.027)	(0.054)	(0.029)	(0.034)	(0.024)	(0.042)	(0.050)
restricted	0.14	0.09	0.06	0.11	0.15	0.09	0.21	0.07
	(0.011)	(0.02)	(0.019)	(0.025)	(0.034)	(0.039)	(0.045)	(0.11)
Electricity								
unrestricted	0.038	0.020	0.017	0.032	0.017	0.022	0.020	0.012
	(0.021)	(0.010)	(0.024)	(0.028)	(0.032)	(0.014)	(0.024)	(0.022)
restricted	0.011	0.015	0.014	0.021	0.005	0.008	0.011	0.012
No. Obs.	6051	1394	1129	1032	758	674	507	465

6. Measuring the productivity effects of R&D

Investment in R&D and innovation is expensive. Investors (e.g., firms, policy makers) are interested in measuring its returns, private and social. **Process R&D:** Directed towards invention of new methods of production. **Product R&D:** Directed towards creation of new and improved goods. Both can increase the firm's TFP It can have also **spillover effects in other firms**: competition spillovers, and/or knowledge spillovers.

6.1. Knowledge capital model. Knowledge capital model (Grilliches, 1979). Most studies measuring returns to R&D are based on the estimation production function, i.e., effect of R&D on TFP. Cobb-Douglas in logs:

$$y_{it} = \beta_L \ \ell_{it} + \beta_K \ k_{it} + \beta_M \ m_{it} + \beta_R \ k_{it}^R + \omega_{it} + e_{it}$$

k_{it} = log of **stock of physical capital;** k_{it}^R = log of **stock of knowledge capital**. A major difficulty is the **measurement of the stock of knowledge capital**.

Measurement of knowledge capital. We observe firms' R&D expenses, R_{it}. How to construct K_{it}^R? **Perpetual inventory method.** Given $\{R_{it} : t = 1, 2, ...T_i\}$, the transition rule:

$$K_{it}^R = (1 - \delta_R) \ K_{i,t-1}^R + R_{it}$$

and values for δ_R and K_{i0} we can construct $\{K_{it}^R : t = 1, 2, ...T_i\}$. How to choose δ_R and K_{i0}? It is very difficult to know the true value of the rate of technological obsolescence, δ_R: it can be endogenous, vary across industries and firms, ...

Different studies using patent renewal data (Pakes and Schankerman, 1984; Pakes, 1986) or Tobin's Q model (Hall, 2005) estimate depreciation rates ranging between 10% and 35%. Different authors (e.g., Grilliches and Mairesse, 1984) have performed sensitivity analysis on

the estimates of β_R for different value of δ_R. They report small differences, if any, in the estimate of β_R when δ_R varies between 8% and 25%.

6.2. Extending Knowledge Capital model: Doraszelski & Jaumandreu (2013).

In their model, TFP and Knowledge capital (KC) are unobservables to the researcher. They follow a stochastic process that is endogenous and depends on (observable) R&D investments. The model accounts for **uncertainty and heterogeneity** across firms in the link between R&D and TFP.

It takes into account that the outcome of R&D investments is subject to a high degree of uncertainty. For the estimation of the structural parameters in PF and stochastic process of KC, they exploit first order conditions for variable inputs.

Model. The PF (in logs) is:

$$y_{it} = \beta_L \ \ell_{it} + \beta_K \ k_{it} + \beta_M \ m_{it} + \omega_{it} + e_{it}$$

log-TFP ω_{it} follows a stochastic process with transition probability $p\left(\omega_{it+1} \mid \omega_{it}, \ r_{it}\right)$ where r_{it} is log-R&D expenditure. Every period t a firm chooses static inputs (ℓ_{it}, m_{it}) and investment in physical capital and R&D (i_{it}, r_{it}) to maximize its value.

$$V(s_{it}) = \max_{i_{it}, r_{it}} \left\{\pi(s_{it}) - c^{(1)}(i_{it}) - c^{(2)}(r_{it}) + \rho \mathbb{E}\left[V(s_{it+1})|s_{it}, i_{it}, r_{it}\right]\right\}$$

with $s_{it} = (k_{it}, \omega_{it}, \text{ input prices } [w_{it}], \text{ demand shifters } [d_{it}])$.

TFP stochastic process. The Markov structure of log-TFP implies:

$$\omega_{it} = \mathbb{E}\left[\omega_{it} \mid \omega_{it-1}, \ r_{it-1}\right] + \xi_{it} = g\left(\omega_{it-1}, \ r_{it-1}\right) + \xi_{it}$$

where $\mathbb{E}\left[\xi_{it} \mid \omega_{it-1}, \ r_{it-1}\right] = 0$. The *productivity innovation* ξ_{it} captures to sources of uncertainty for the firm: the naturally linked to the evolution of TFP; the uncertainty inherent to R&D (e.g., chance of discovery, degree of applicability, success in implementation).

Marginal Revenue (Market Power). D&J identification approach exploits marginal conditions (MR = MC) for variable inputs. This requires an assumption about competition/market power. They assume:

$$MR_{it} = P_{it}\left(1 - \frac{1}{\eta(p_{it}, d_{it})}\right)$$

where $\eta(p_{it}, d_{it})$ is price elasticity of demand for firm i, i.e., monopolisitc competition.

VMP of labor = wage. This marginal condition of optimality for labor provides a closed-form expression for labor demand. Solving for log-TFP in the labor demand equation, we

get:

$$\omega_{it} = \lambda - \beta_K \, k_{it} + (1 - \beta_L - \beta_M) \, \ell_{it} + (1 - \beta_M) \, (w_{it} - p_{it})$$

$$+ \beta_M \, (p_{Mit} - p_{it}) - \ln \left(1 - \frac{1}{\eta(p_{it}, d_{it})} \right)$$

We represent the RHS as $h(x_{it}, \beta)$, such that $\omega_{it} = h(x_{it}, \beta)$, with:

$$x_{it} = (k_{it}, \, \ell_{it}, \, w_{it}, \, p_{Mit}, \, p_{it}, \, d_{it})$$

Estimation. Combining the PF equation with the stochastic process for TFP, and the marginal condition for optima labor, we have the equation:

$$y_{it} = \beta_L \, \ell_{it} + \beta_K \, k_{it} + \beta_M \, m_{it} + g \left[h(x_{it-1}, \beta), \, r_{it-1} \right] + \xi_{it} + e_{it}$$

And form the marginal condition for labor we have:

$$h(x_{it}, \beta) = g \left[h(x_{it-1}, \beta), \, r_{it-1} \right] + \xi_{it}$$

The "parameters" in this system of equations are: β_L, β_K, β_M, $g(.)$, and $\eta(.)$. The unobservables ξ_{it} and e_{it} is mean independent of any observable variable at period $t-1$ or before. Therefore, x_{it-1} and r_{it-1} are exogenous w.r.t. $\xi_{it} + e_{it}$. Capital stock k_{it} is also exogenous because time-to-build. **But we need to instrument ℓ_{it} and m_{it}.**

Identification. To see that the parameters of the model are identified, it is convenient to consider a simplified version with: $\beta_K = \beta_M = 1/\eta(.) = 0$ and $g \left[\omega_{t-1}, \, r_{t-1} \right] = \rho_\omega \, \omega_{t-1} + \rho_r \, r_{t-1}$. Them we have:

$$y_{it} = \beta_L \, \ell_{it} + \rho_\omega \left[(1 - \beta_L) \, \ell_{it-1} + w_{it-1} - p_{it-1} \right] + \rho_r \, r_{it-1} + \xi_{it} + e_{it}$$

Using as instruments $Z_{it} = (y_{it-1}, \ell_{it-1}, w_{it-1} - p_{it-1}, r_{it-1})$, moment conditions $\mathbb{E} \left[Z_{it} \, (\xi_{it} + e_{it}) \right] = 0$ identify β_L, ρ_ω, ρ_r. Given the identification of these parameters, we know $\omega_{it} = h(x_{it}, \beta) = (1 - \beta_L) \ell_{it} + (w_{it} - p_{it})$. The model implies, that:

$$\xi_{it} = h(x_{it}, \beta) - \rho_\omega \, h(x_{it}, \beta) - \rho_r \, r_{it-1}$$

such that ξ_{it} is identified, and so its variance $Var(\xi_{it})$ that represents uncertainty in the link between R&D and TFP.

The instrument $w_{it-1} - p_{it-1}$ plays a very important role in the identification of the model. Without variation in lagged (real) input prices the model is NOT identified. But note that the model does not use contemporaneous input prices as instruments because they can be correlated with the innovation ξ_{it}.

Data. Panel of Spanish manufacturing firms ($N = 1,870$). Annual data for period $1990 - 1999$ (max $T_i = 10$). 10 industries (SIC 2-digits). Period of rapid growth in output

and physical capital, coupled with stagnant employment. **R&D intensity = R&D expenditure / Sales.** Average among all firms is 0.6% (smaller than in France, Germany, or UK, $> 2\%$). **R&D intensity** among performers (column 13) is between 1% and 3.5%.

TABLE 1
Descriptive statistics

Industry	Obs.[a]	Firms[a]	Entry[a] (%)	Exit[a] (%)	Output (s. d.)	Capital (s. d.)	Labour (s. d.)	Materials (s. d.)	Price (s. d.)	Obs. (%)	Stable (%)	Occas. (%)	R&D inten. (s. d.)
						Rates of growth[a]						With R&D[b]	
	(1)	(2)	(3)	(4)	(5)	(7)	(6)	(8)	(9)	(10)	(11)	(12)	(13)
1. Metals and metal products	1235	289	88 (30.4)	17 (5.9)	0.050 (0.238)	0.086 (0.278)	0.010 (0.183)	0.038 (0.346)	0.012 (0.055)	420 (34.0)	63 (21.8)	72 (24.9)	0.0126 (0.0144)
2. Non-metallic minerals	621	131	20 (15.3)	15 (11.5)	0.037 (0.208)	0.062 (0.238)	−0.001 (0.141)	0.039 (0.308)	0.010 (0.059)	186 (30.0)	16 (12.2)	41 (31.3)	0.0100 (0.0211)
3. Chemical products	1218	275	64 (23.3)	15 (5.5)	0.068 (0.196)	0.093 (0.238)	0.007 (0.146)	0.054 (0.254)	0.007 (0.061)	672 (55.2)	124 (45.1)	55 (20.0)	0.0268 (0.0353)
4. Agric. and ind. machinery	576	132	36 (27.3)	6 (4.5)	0.059 (0.275)	0.078 (0.247)	0.010 (0.170)	0.046 (0.371)	0.013 (0.032)	322 (55.9)	52 (39.4)	35 (26.5)	0.0219 (0.0275)
6. Transport equipment	637	148	39 (26.4)	10 (6.8)	0.087 (0.354)	0.114 (0.255)	0.011 (0.207)	0.087 (0.431)	0.007 (0.037)	361 (56.7)	62 (41.9)	35 (23.6)	0.0224 (0.0345)
7. Food, drink, and tobacco	1408	304	47 (15.5)	22 (7.2)	0.025 (0.224)	0.094 (0.271)	−0.003 (0.186)	0.019 (0.305)	0.022 (0.065)	386 (27.4)	56 (18.4)	64 (21.1)	0.0071 (0.0281)
8. Textile, leather, and shoes	1278	293	77 (26.3)	49 (16.7)	0.020 (0.233)	0.059 (0.235)	−0.007 (0.192)	0.012 (0.356)	0.016 (0.040)	378 (29.6)	39 (13.3)	66 (22.5)	0.0152 (0.0219)
9. Timber and furniture	569	138	52 (37.7)	18 (13.0)	0.038 (0.278)	0.077 (0.257)	0.014 (0.210)	0.029 (0.379)	0.020 (0.035)	66 (12.6)	7 (5.1)	18 (13.8)	0.0138 (0.0326)
10. Paper and printing products	665	160	42 (26.3)	10 (6.3)	0.035 (0.183)	0.099 (0.303)	−0.000 (0.140)	0.026 (0.265)	0.019 (0.089)	113 (17.0)	21 (13.1)	25 (13.8)	0.0143 (0.0250)

Production Function Estimates. Comparing GMM and OLS estimates, correcting for endogeneity has the expected implications, e.g., β_L and β_M decline, and β_K increases. There are not big differences in the β estimates across industries. Test of OIR from instruments: Cannot be rejected at 5% level. Test of parameter restrictions (in the two equations): Rejected at 5% level only in 2 out of 10 industries.

TABLE 2
Production function estimates and specification tests

Industry	Capital (std. err.)	Labour (std. err.)	Materials (std. err.)	Capital (std. err.)	Labour (std. err.)	Materials (std. err.)	$\chi^2(df)$	p val.	$\chi^2(3)$	p val.
	OLS[a]			GMM[a]			Overidentifying restrictions test		Parameter restrictions test	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
1. Metals and metal products	0.109 (0.013)	0.252 (0.022)	0.642 (0.020)	0.106 (0.014)	0.111 (0.031)	0.684 (0.011)	62.553 (51)	0.129	11.666	0.009
2. Non-metallic minerals	0.096 (0.021)	0.275 (0.034)	0.655 (0.028)	0.227 (0.014)	0.137 (0.016)	0.633 (0.014)	50.730 (47)	0.329	6.047	0.109
3. Chemical products	0.060 (0.010)	0.239 (0.021)	0.730 (0.020)	0.132 (0.015)	0.122 (0.026)	0.713 (0.011)	48.754 (47)	0.402	0.105	0.991
4. Agric. and ind. machinery	0.051 (0.017)	0.284 (0.038)	0.671 (0.027)	0.079 (0.015)	0.281 (0.029)	0.642 (0.013)	45.833 (44)	0.396	1.798	0.615
6. Transport equipment	0.080 (0.023)	0.289 (0.033)	0.636 (0.046)	0.117 (0.015)	0.158 (0.023)	0.675 (0.016)	40.296 (47)	0.745	0.414	0.937
7. Food, drink, and tobacco	0.094 (0.014)	0.177 (0.016)	0.739 (0.016)	0.068 (0.014)	0.129 (0.024)	0.766 (0.008)	61.070 (46)	0.068	8.866	0.031
8. Textile, leather, and shoes	0.059 (0.010)	0.335 (0.024)	0.605 (0.019)	0.057 (0.011)	0.313 (0.016)	0.593 (0.013)	66.143 (51)	0.075	4.749	0.191
9. Timber and furniture	0.079 (0.019)	0.283 (0.029)	0.670 (0.029)	0.131 (0.009)	0.176 (0.017)	0.697 (0.011)	44.951 (43)	0.390	0.618	0.892
10. Paper and printing products	0.092 (0.016)	0.321 (0.029)	0.621 (0.025)	0.121 (0.013)	0.249 (0.025)	0.617 (0.014)	51.371 (42)	0.152	5.920	0.118

Stochastic Process for TFP. The model where TFP is exogenous (doesn't depend on R&D) is clearly rejected. Models with linear effects or without complementarity between ω_{t-1} and r_{t-1} are rejected. $Var(e)$ is approx. equal to $Var(\omega)$ in most industries. $Var(\xi)/Var(\omega)$

is between 30% and 75%. Very significant uncertainty of the effect of R&D on TFP. Significant differences across industries.

Industry	Exogeneity test		Separability test		$\dfrac{Var(e_{jt})}{Var(\omega_{jt})}$	$\dfrac{Var(\xi_{jt})}{Var(\omega_{jt})}$
	$\chi^2(10)$	p val.	$\chi^2(3)$	p val.		
	(1)	(2)	(3)	(4)	(5)	(6)
1. Metals and metal products	65.55	0.000	16.360	0.001	0.735	0.407
2. Non-metallic minerals	92.65	0.000	13.027	0.005	0.842	0.410
3. Chemical products	40.79	0.000	8.647	0.034	0.749	0.244
4. Agric. and ind. machinery	51.88	0.000	11.605	0.009	1.410	0.505
6. Transport equipment	56.85	0.000	18.940	0.000	1.626	0.524
7. Food, drink, and tobacco	38.29	0.000	7.186	0.066	1.526	0.300
8. Textile, leather, and shoes	29.91	0.001	18.417	0.000	1.121	0.750
9. Timber and furniture	118.17	0.000	32.260	0.000	1.417	0.515
10. Paper and printing products	59.73	0.000	23.249	0.000	0.713	0.433

Testing the standard Knowledge Capital model. Testing three versions of the Knowledge capital model. **Basic model:** $\omega_{it} + e_{it} = \beta_R\, k_{it}^R + e_{it}$. Rejected for all industries. **Hall & Hayashi (1989) and Klette (1996) KC model.** $K_{it}^R = \left[K_{it-1}^R\right]^\sigma \left[1 + R_{it-1}^R\right]^{1-\sigma} \exp\{\xi_{it}\}$. Using D&J notation:

$$\omega_{it} = \sigma\,\omega_{it-1} + (1-\sigma)\,r_{it-1} + \xi_{it}$$

Rejected at 5% in 8 industries, and at 7% in all industries. **Model with:** $\beta_R\, k_{it}^R + \omega_{it} + e_{it}$, and ω_{it} with exogenous Markov process. Rejected at 5% in 2 industries, and at 10% in 6 industries.

Knowledge capital model tests

| Basic		Generalization 1		Generalization 2	
$N(0,1)$	p val.	$N(0,1)$	p val.	$N(0,1)$	p val.
(7)	(8)	(9)	(10)	(11)	(12)
-2.815	0.002	-2.431	0.008	-1.987	0.023
-2.041	0.021	-1.541	0.062	-0.784	0.216
-3.239	0.001	-2.090	0.018	-1.400	0.081
-2.693	0.004	-1.588	0.056	-1.493	0.068
-2.317	0.010	-2.042	0.021	-1.821	0.034
-3.263	0.001	-2.499	0.006	-0.901	0.184
-2.770	0.003	-1.788	0.037	-1.488	0.068
-2.510	0.006	-2.097	0.018	-1.028	0.152
-3.076	0.001	-2.210	0.014	-1.595	0.055

R&D and TFP (Counterfactuals). Distribution of TFP with R&D stochastically dominates distribution without R&D. Differences in means are **between 3% and 5%** for all industries and firm sizes, except for small firms in industries with low observed R&D intensity.
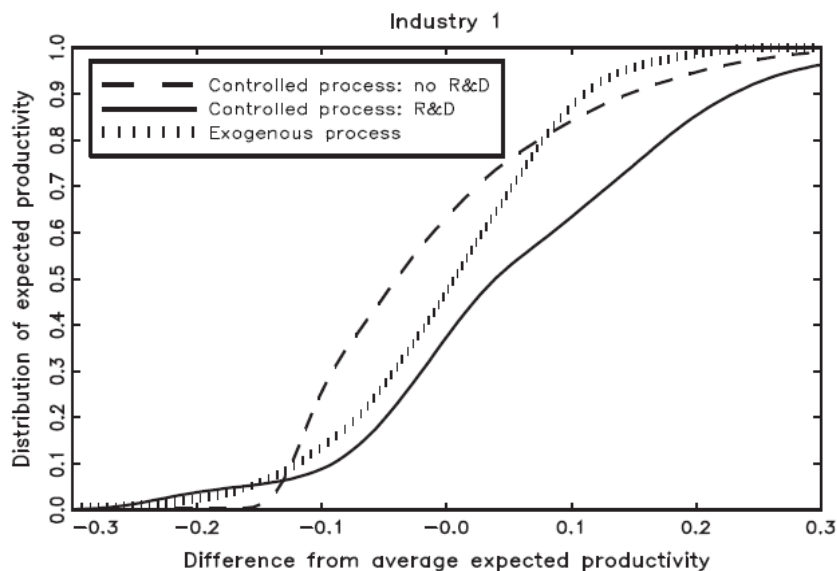
Elasticities of TFP w.r.t. R&D and lagged TFP. **Elasticity w.r.t. R&D:** Considerable variation between and within industries. Average is 0.015. **Degree of persistence:** Considerable between and within industries. Non-performers have a higher degree of persistence than performers. Persistence is negatively related to the degree of uncertainty

Industry	Elasticity wrt. R_{jt-1}[a]			
	Q_1	Q_2	Q_3	Mean
	(1)	(2)	(3)	(4)
1. Metals and metal products	−0.013	0.007	0.021	0.022
2. Non-metallic minerals	−0.018	−0.012	0.000	−0.006
3. Chemical products	0.009	0.011	0.014	0.013
4. Agric. and ind. machinery	−0.017	−0.009	0.021	0.005
6. Transport equipment	−0.034	−0.008	0.010	0.020
7. Food, drink, and tobacco	−0.008	0.010	0.026	0.020
8. Textile, leather, and shoes	−0.003	0.014	0.051	0.046
9. Timber and furniture	−0.031	0.005	0.048	0.004
10. Paper and printing products	−0.036	0.022	0.049	0.013

Elasticity wrt. ω_{jt-1} [b]					
Performers			Non-performers		
Q_1	Q_2	Q_3	Q_1	Q_2	Q_3
(5)	(6)	(7)	(8)	(9)	(10)
0.504	0.619	0.755	0.441	0.759	0.901
0.433	0.477	0.575	0.377	0.646	0.878
0.459	0.523	0.634	0.547	0.815	0.947
0.434	0.721	0.791	0.729	0.894	0.979
0.404	0.615	0.727	0.423	0.513	0.646
0.445	0.705	0.867	0.822	0.930	0.965
0.090	0.325	0.626	0.491	0.605	0.689
0.458	0.585	0.814	0.303	0.430	0.641
0.405	0.676	0.812	0.569	0.644	0.670
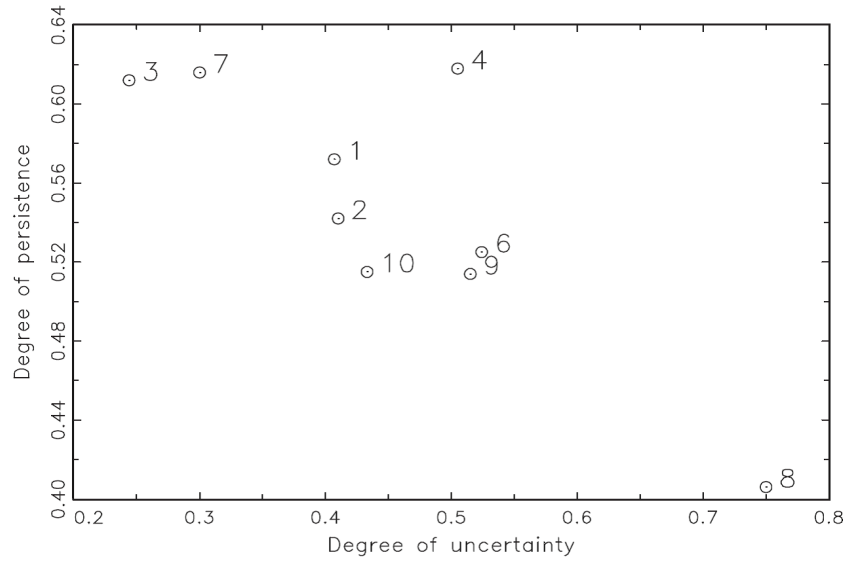
Summary of results. They model TFP growth as the consequence of R&D expenditures with uncertain outcomes. Results show that this model can explain better the relationship between TFP and R&D than standard Knowledge Capital models without uncertainty and nonlinearity. R&D is a major determinant of the differences in TFP across firms and of their evolution. They also find that firm-level uncertainty in the outcome of R&D is considerable. Their estimates suggest that engaging in R&D roughly doubles the degree of uncertainty in the evolution of a producer's TFP.

7. Exercises

7.1. Exercise 1. Consider an industry for an homogeneous product. Firms use capital and labor to produce output according to a Cobb-Douglas technology with parameters α_L and α_K and Total Factor Productivity (TFP) A.

Question 1.1. Write the expression for this Cobb-Douglas production function (PF).

Suppose that firms are price takers in the input markets for labor and capital. Let W_L and W_K be the price of labor and capital, respectively. Capital is a fixed input such that the fixed cost for a firm, say i, is $FC_i = W_K K_i$. The *variable cost function*, $VC(Y)$, is defined as the minimum cost of labor to produce an amount of output Y.

Question 1.2. Derive the expression for the variable cost function of a firm in this industry. Explain your derivation. [Hint: Given that capital is fixed and there is only one variable input, the minimization problem is trivial. The PF implies that there is only one possible amount of labor that give us a certain amount of output].

Question 1.3. Using the expressions for the fixed cost and for the variable cost function in Q1.2:

(a) Explain how an increase in the amount of capital affects the fixed cost and the variable cost of a firm.

(b) Explain how an increase in TFP affects the fixed cost and the variable cost.

Suppose that the output market in this industry is competitive: firms are price takers. The demand function is linear with the following form: $P = 100 - Q$, where P and Q are the industry price and total output, respectively. Suppose that $\alpha_L = \alpha_K = 1/2$, and the value of input prices are $W_L = 1/2$ and $W_K = 2$. Remember that firms' capital stocks are fixed (exogenous), and for simplicity suppose that all the firms have the same capital stock $K = 1$.

Question 1.4. Using these primitives, write the expression for the profit function of a firm (revenue, minus variable cost, minus fixed cost) as a function of the market price, P, the firm's output, Y_i, and its TFP, A_i.

Question 1.5. Using the condition "price equal to marginal cost", obtain the optimal amount of output of a firm as a function of the market price, P, and the firm's TFP, A_i. Explain your derivation.

Question 1.6. A firm is active in the market (i.e., it finds optimal to produce a positive amount of output) only if its profit is greater or equal than zero. Using this condition show that a firm is active in this industry only if its TFP satisfies the condition $A_i \geq 2/P$. Explain your derivation.

Let $(P^*, Q^*, Y_1^*, Y_2^*, ..., Y_N^*)$ the equilibrium price, total output, and individual firms' outputs. Based on the previous results, the market equilibrium can be characterized by the following conditions: (i) the demand equation holds; (ii) total output is equal to the sum of firms' individual outputs; (iii) firm i is active ($Y_i^* > 0$) if and only if its total profit is greater than zero; and (iv) for firms with $Y_i^* > 0$, the optimal amount of output is given by the condition price is equal to marginal cost.

Question 1.7. Write conditions (i) to (iv) for this particular industry.

Question 1.8. Combine conditions (i) to (iv) to show that the equilibrium price can be written as the solution to this equation:

$$P^* = 100 - P^* \left[\sum_{i=1}^{N} A_i^2 \, 1\{A_i \geq 2/P^*\} \right]$$

where $1\{x\}$ is the indicator function that is defined as $1\{x\} = 1$ if condition x is true, and $1\{x\} = 0$ if condition x is false. Explain your derivation.

Suppose that the subindex i sorts firms by their TFP such that firm 1 is the most efficient, then firm 2, etc. That is, $A_1 > A_2 > A_3 > .....$

Question 1.9. Suppose that $A_1 = 7$, $A_2 = 5$, and $A_3 = 1$. Obtain the equilibrium price, total output, and output of each individual firm in this industry. [Hint: Start with the conjecture that only firms 1 and 2 produce in equilibrium. Then, confirm this conjecture. Note that we do not need to know the values of A_4, A_5, etc].

Question 1.10. Explain why the most efficient firm, with the largest TFP, does not produce all the output of the industry.

7.2. Exercise 2. The Stata datafile `blundell_bond_2000_production_function.dta` contains annual information on sales, labor, and capital for 509 firms for the period 1982-1989 (8 years). Consider a Cobb-Douglas production function in terms of labor and capital. Use this dataset to implement the following estimators.

Question 2.1. OLS with time dummies. Test the null hypothesis $\alpha_L + \alpha_K = 1$. Provide the code in Stata and the table of estimation results. Comment the results.

Question 2.2. Fixed Effects estimator with time dummies. Test the null hypothesis of no time-invariant unobserved heterogeneity: $\eta_i = \eta_i$ for every firm i. Provide the code in Stata and the table of estimation results. Comment the results.

Question 2.3. Fixed Effects - Cochrane Orcutt estimator with time dummies. Test the two over-identifying restrictions of the model. Provide the code in Stata and the table of estimation results. Comment the results.

Question 2.4. Arellano-Bond estimator with time dummies and non-serially correlated transitory shock. Provide the code in Stata and the table of estimation results. Comment the results.

Question 2.5. Arellano-Bond estimator with time dummies and AR(1) transitory shock. Provide the code in Stata and the table of estimation results. Comment the results.

Question 2.6. Blundell-Bond system estimator with time dummies and non-serially correlated transitory shock. Provide the code in Stata and the table of estimation results. Comment the results.

Question 2.7. Blundell-Bond system estimator with time dummies and AR(1) transitory shock. Provide the code in Stata and the table of estimation results. Comment the results.

Question 2.8. Based on the previous results, select your preferred estimates of the production function. Explain your choice.

7.3. Exercise 3. The Stata datafile `data_mines_eco2901_2017.dta` contains annual information on output and inputs from 330 copper mines for the period 1992-2010 (19 years). The following is a description of the variables.

Variable name		**Description**
`id`	:	Mine identification number
`year`	:	Year [from 1992 to 2010]
`active`	:	Binary indicator of the event "mine is active during the year"
`prod_tot`	:	Annual production of pure copper of the mine [in thousands of tonnes]
`reserves`	:	Estimated mine reserves [in thousands of ore]
`grade`	:	Average ore grade (in %) of mined ore during the year (% copper / ore)
`labor_n_tot`	:	Total number of workers per year (annual equivalent)
`cap_tot`	:	Measure of capital [maximum production capacity of the mine]
`fuel_cons_tot`	:	Consumption of fuel (in physical units)
`elec_cons_tot`	:	Consumption of electricity (in physical units)
`materials_tot`	:	Consumption of intermediate inputs / materials (in $ value)

Note that some variables have a few missing values even at years when the mine is actively producing.

Question 3.1. Consider a Cobb-Douglas production function in terms of labor, capital, fuel, electricity, and ore grade. Use this dataset to implement the following estimators:
- OLS
- Fixed-Effects
- Arellano-Bond estimator with non-serially correlated transitory shock
- Arellano-Bond estimator with $AR(1)$ transitory shock
- Blundell-Bond estimator with non-serially correlated transitory shock
- Blundell-Bond estimator with $AR(1)$ transitory shock
- Olley-Pakes (Using the first difference in cap_tot as investment)
- Levinshon-Petrin

Question 3.2. Suppose that these mines are price takers in the input markets. Consider that the variable inputs are labor, fuel, and electricity.

(a) Derive the expression for the Variable Cost function for a mine (i.e., the minimum cost to produce an amount of output given input prices).

(b) Let $\ln MC_{it}$ be the logarithm of the realized Marginal Cost of mine i at year t. I have not included data on input prices in this dataset, so we will assume that mines face the same prices for variable inputs, and normalize to zero the contribution of these input prices to $\ln MC_{it}$. Calculate the quantiles 5%, 25%, 50%, 75%, and 95% in the cross-sectional distributions of $\ln MC_{it}$ at each year in the sample. Present a figure with the time-series of these five quantiles over the sample period. Comment the results.

(c) For a particular sample year, say 2005, calculate the contribution of each component of $\ln MC_{it}$ (i.e., total factor productivity, capital, ore grade, and output) to the cross-sectional variance of $\ln MC_{it}$. Present it in a table. Comment your results.

[Note: To measure the contribution of each component, use the following approach. Consider $y = \beta_1 x_1 + \beta_2 x_2 + ... + \beta_K x_K$. A measure of the contribution of x_j to $var(y)$ is $\rho_j \equiv \dfrac{var(y) - var(y \mid x_j = constant)}{var(y)}$. Note that $\rho_j \in (0,1)$ for any variable x_j. However, in general, $\sum_{j=1}^{K} \rho_j$ can be either smaller or greater than one, depending the sign of the covariances between the components.]

(d) Consider the balance panel of mines that are active in the industry every year during the sample period. Repeat exercises (b) and (c) for this balanced panel. Compared your results with those in (c) and (d). Comment the results.

Bibliography

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

[2] Ackerberg, D., K. Caves and G. Frazer (2015): "Identification Properties of Recent Production Function Estimators, " *Econometrica*, 83(6), 2411-2451.

[3] Aguirregabiria, V. and Alonso-Borrego, C. (2014): "Labor Contracts and Flexibility: Evidence from a Labor Market Reform in Spain," manuscript. Department of Economics. University of Toronto.

[4] Alonso-Borrego, C., and R. Sanchez-Mangas (2001): "GMM Estimation of a Production Function with Panel Data: An Application to Spanish Manufacturing Firms," Statistics and Econometrics Working Papers #015527. Universidad Carlos III.

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

[6] Arellano, M., and O. Bover (1995): "Another Look at the Instrumental Variable Estimation of Error-Components Models," Journal of Econometrics, 68, 29-51.

[7] Aw, B., Roberts, M. and Xu, D. (2011), "R&D Investment, Exporting, and Productivity Dynamics", American Economic Review, 101, 1312-1344.

[8] Blundell, R., and S. Bond (1998): "Initial conditions and moment restrictions in dynamic panel data models," Journal of Econometrics, 87, 115-143.

[9] Blundell, R., and S. Bond (2000): "GMM estimation with persistent panel data: an application to production functions," Econometric Reviews, 19(3), 321-340.

[10] Bond, S., and M. Söderbom (2005): "Adjustment costs and the identification of Cobb Douglas production functions," IFS Working Papers W05/04, Institute for Fiscal Studies.

[11] Bond, S. and J. Van Reenen (2007): "Microeconometric Models of Investment and Employment," in J. Heckman and E. Leamer (editors) *Handbook of Econometrics, Vol. 6A*. North Holland. Amsterdam.

[12] Cobb, C. and P. Douglas (1928): "A Theory of Production," American Economic Review, 18(1), 139-165.

[13] Crepon, B., E. Guguet, and J. Mairesse (1998): "Research, Innovation and Productivity: An Econometric Analysis at the Firm Level," Economics of Innovation and new Technology, 7(2), 115-158.

[14] Doraszelski, U., and J. Jaumandreu (2013): "R&D and Productivity" Estimating Endogenous Productivity," Review of Economic Studies, , 80, 1338 - 1383.

[15] Gandhi, A., S. Navarro, and D. Rivers (2017): "On the identification of production functions: How heterogeneous is productivity?" Journal of Political Economy, forthcoming,

[16] Griliches, Z., and J. Mairesse (1998): "Production Functions: The Search for Identification," in *Econometrics and Economic Theory in the Twentieth Century: The Ragnar Frisch Centennial Symposium*. S. Strøm (editor). Cambridge University Press. Cambridge, UK.

[17] Hall, B., J. Mairesse, and P. Mohnen (2010): "Measuring the Returns to R&D", in Hall, B. and Rosenberg, N. (eds), Handbook of the Economics of Innovation. North-Holland.

[18] Kasahara, H. (2009): "Temporary Increases in Tariffs and Investment: The Chilean Case," Journal of Business and Economic Statistics, 27(1), 113-127.

[19] Levinshon, J., and A. Petrin (2003): "Estimating Production Functions Using Inputs to Control for Unobservables," Review of Economic Studies , 70, 317-342.

[20] Marschak, J. (1953): "Economic measurements for policy and prediction," in *Studies in Econometric Method*, eds. W. Hood and T. Koopmans. New York. Wiley.

[21] Marshak, J., and W. Andrews (1944): "Random simultaneous equation and the theory of production," Econometrica, 12, 143–205.

[22] Mundlak, Y. (1961): "Empirical Production Function Free of Management Bias," Journal of Farm Economics, 43, 44-56.

[23] Mundlak, Y., and I. Hoch (1965): "Consequences of Alternative Specifications in Estimation of Cobb-Douglas Production Functions," Econometrica, 33, 814-828.

[24] Olley, S., and A. Pakes (1996): "The Dynamics of Productivity in the Telecommunications Equipment Industry", *Econometrica,* 64, 1263-97.

[25] Pakes, A. (1994): "Dynamic structural models, problems and prospects," in C. Sims (ed.) Advances in Econometrics. Sixth World Congress, Cambridge University Press.

[26] Syverson, C. (2011): "What Determines Productivity?" Journal of Economic Literature, 49:2, 326–365

[27] Wooldridge, J. (2009): "On Estimating Firm-Level Production Functions Using Proxy Variables to Control for Unobservables," Economics Letters, 104, 112-114.

CHAPTER 4

Static Models of Competition in Prices and Quantities

1. Introduction

In most industries, the decisions of how much to produce and the price to charge are fundamental determinants of the firm's profit. These decisions are also main sources of strategic interactions between firms. In the market for an homogeneous good, the market price of the product depends on the total quantity produced by all the firms in the industry. In a differentiated product industry, demand for a firm's product depends on the prices of products sold by other firms in the industry. This type of strategic interactions have first order importance to understand competition and outcomes in most industries. For this reason, models of competition where firms choose prices or quantities are at the core of Industrial Organization.

The answer to many economy questions in IO require not only the estimation of demand and cost functions but also the explicit specification of an equilibrium model of competition. For instance, evaluating the effects of a merger, or a sales tax, or the entry in the market of a new product, require the explicit specification of a model of competition.

From an empirical point view, there are several purposes in the estimation of models of competition in prices or quantities. In many empirical applications, the researcher has information on firms' prices and quantities sold, but information on firms' costs is not always available. The researcher may not observe even the amounts of firms' inputs, such that it is not possible to obtain costs by estimating the production function as described in chapter 3. In this context, empirical models of competition in prices or quantities may provide an approach to obtain estimates of firms' marginal costs and the structure of the marginal cost function. Given an assumption about competition (e.g., perfect competition, Cournot, Bertrand, Stackelberg, collusion), the model predicts that a firm's marginal cost should be equal to a particular (model specific) marginal revenue. This is the key condition that is used to estimate firms' marginal costs in this class of models. Typically, the first step in the econometric analysis of these models consists in the estimation of the demand function or demand system. Given the estimated demand, we can construct an estimate of the realized marginal revenue for every observation in the sample. This measure of marginal revenue provides, directly, an estimate of the realized marginal cost at each sample observation.

113

Finally, we use this sample of realized marginal costs to estimate the marginal cost function, and in particular how the marginal cost depends on firm's output of different products (i.e., economies of scale and scope), and possibly on other firm's characteristics such as historical cumulative output (i.e., learning by doing), installed capacity, or geographic distance between the firm's production plants (i.e., economies of density).

The value of a firm's marginal revenue depends on the form of competition in the industry, or *the nature of competition*. Given the same demand function, the marginal revenue is different under perfect competition, Cournot, Bertrand, or collusion. The researcher's selection of a model of competition typically implies answers the following choices: (a) is the product homogeneous or differentiated; (b) do firms compete in prices or in quantities?; (c) is there collusion between some or all the firms in the industry?; and (d) what does a firm believe about the behavior of other firms in the market? For instance, if the researcher assumes that the product is homogenous, firms compete in quantities, there is no collusion in the industry, and firms choose their levels of output under the belief that the other firms will not change their respective output levels (i.e., Nash assumption), then the form of competition is the one in the Cournot model. In principle, some of these assumptions may be supported by the researcher's knowledge of the industry. However, in general, some of these assumptions are difficult to justify. Ideally. we would like to learn from our data about the nature of competition. Suppose that the researcher has data on firms' marginal costs (or estimates of these costs based on a production function) and an estimation of the demand system. Then, given an assumption about the form of competition in this industry (e.g., perfect competition, Cournot, collusion), the researcher can use the demand to obtain firms' marginal revenues and check whether they are equal to the observed marginal costs. That is, the researcher can test if a particular form of competition is consistent with the data. In this way, it is possible to find the form of competition that is consistent with the data, e.g., identify if there is evidence of collusive behavior. We will see in this chapter that, even if the researcher does not have data on firms' costs, it is still possible to combine the demand system and the equilibrium conditions to jointly identify marginal costs and the *nature of competition* in the industry. This is the main purpose of the so called *conjectural variation approach*.

In this chapter, we describe the specification and estimation of empirical models of Cournot competition in an homogenous product industry, Bertrand competition in a differentiated product industry, and the conjectural variation approach both in homogenous and differentiated product industries.

2. Homogenous product industry

2.1. Estimating marginal costs from equilibrium conditions. First, we consider the situation where the researcher does not have direct measures of marginal costs and uses the equilibrium conditions to estimate these costs.

2.1.1. *Perfect competition.* We first illustrate this approach in the context of a perfectly competitive industry for an homogeneous product. Suppose that the researcher knows, or is willing to assume, that the industry under study is perfectly competitive, and she has data on the market price and on firms' output for T periods of time (or T geographic markets) that we index by t. The dataset consists of $\{p_t, q_{it}\}$ for $i = 1, 2, ..., N_t$ and $t = 1, 2, ..., T$, where N_t is the number of firms active at period t. The variable profit of firm i is $p_t q_{it} - C_i(q_{it})$. Under perfect competition, the marginal revenue of any firm i is the market price, p_t. The marginal condition of profit maximization for firm i is $p_t = MC_i(q_{it})$ where $MC_i(q_{it})$ is the marginal cost, $MC_i(q_{it}) \equiv C_i'(q_{it})$. Under perfect competition, all the firms should have the same marginal costs. This is a clear testable restriction of the assumption of perfect competition with homogeneous product.

Consider a particular specification of the cost function. With a Cobb-Douglas production function, we have that:

$$MC_i(q_{it}) = q_{it}^\theta W_{1it}^{\alpha_1} ...W_{Jit}^{\alpha_J} \exp\{\varepsilon_{it}^{MC}\} \tag{2.1}$$

W_{jit} is the price of variable input j for firm i, and α's are technological parameters in the Cobb-Douglas production function. ε_{it}^{MC} is an unobservable to the researcher that captures the cost (in)efficiency of a firm and that depends on the firm's total factor productivity, unobserved input prices and unobserved fixed inputs. The technological parameter θ is equal to $\frac{1}{\alpha_V} - 1$, and α_V is the sum of the Cobb-Douglas coefficients of all the variable inputs. Therefore, the equilibrium condition $p_t = MC_i(q_{it})$ implies the following regression model in logarithms:

$$\ln(p_t) = \theta \ln(q_{it}) + \alpha_1 \ln(W_{1it}) + ... + \alpha_J \ln(W_{Jit}) + \varepsilon_{it}^{MC} \tag{2.2}$$

We can distinguish three cases for parameter θ. Constant Returns to Scale (CRS): $\alpha_V = 1$ such that $\theta = 0$ and this implies that the marginal cost function is a constant function. Decreasing (Increasing) Returns to Scale: $\alpha_V < 1$ ($\alpha_V > 1$) such that $\theta > 0$ ($\theta < 0$) and the log-marginal cost function is an increasing (decreasing) linear function of log-output. The

Using data on prices and quantities, we can estimate the slope parameter θ in this regression equation. Given estimates of the parameters θ and α's we can estimate ε_{it}^{MC} as a residual from this regression. Therefore, we can estimate the marginal cost function function of each firm. Since the dependent variable of the regression, $\ln(p_t)$, is constant over firms,

then, by construction, firms that produce more are more cost-efficient according to the term $\alpha_1 \ln(W_{1it}) + ... + \alpha_J \ln(W_{Jit}) + \varepsilon_{it}^{MC}$.

Estimation of equation (2.2) by OLS suffers of an endogeneity problem. The equilibrium condition implies that firms with a large value of ε_{it}^{MC} are less cost-efficient and, and all else equal, should have a lower level of output. Therefore, the regressor $\ln(q_{it})$ is negatively correlated with the error term ε_{it}^{MC}. This negative correlation between the regressor and the error term implies that the OLS estimator provides a downward biased estimate of the true θ. For instance, the OLS estimate could show increasing returns to scale, $\theta < 0$, when the true technology has decreasing returns to scale, $\theta > 0$. This endogeneity problem does not disappear if we consider the model in market means.

We can deal with this endogeneity problem by using instrumental variables. Suppose that X_t^D is an observable variable (or vector of variables) that affects the demand of the product but not the marginal costs of the firms. The equilibrium of the model implies that these demand variables should be correlated with firms' outputs, $\ln(q_{it})$. Of course, this condition is testable. Under the assumption that these observable demand variables X_t^D are not correlated with the unobserved term in the marginal cost, $\mathbb{E}(X_t^D \ \varepsilon_{it}^{MC}) = 0$, we can use these variables as instruments for log-output in the regression equation (**??**) for the consistent estimation of θ.

2.1.2. *An application.*

2.1.3.

2.1.4. *Cournot competition.* Now, suppose that the researcher assumes that the market is not perfectly competitive and that firms compete a la Nash-Cournot. The demand can be represented using the inverse demand function $p_t = P\left(Q_t, X_t^D\right)$, where $Q_t \equiv \sum_{i=1}^N q_{it}$ is the market total output, and X_t^D is a vector of exogenous market characteristic that affect demand. Each firm chooses its own output q_{it} to maximize profit. Profit maximization implies the condition of marginal revenue equal to marginal cost, and the marginal revenue function is:

$$MR_{it} = p_t + P'_Q\left(Q_t, X_t^D\right)\left[1 + \frac{dQ_{(-i)t}}{dq_{it}}\right]q_{it} \tag{2.3}$$

$P'_Q\left(Q_t, X_t^D\right)$ is the derivative of the inverse demand function with respect to total output. Variable $Q_{(-i)t}$ is the aggregate output of firms other than i, and the derivative $\frac{dQ_{(-i)t}}{dq_{it}}$ represents the *belief* or *conjecture* that firm i has about how other firms will respond by changing their output when this firm changes marginally its own output. Under the assumption of Nash-Cournot competition, this *belief* or *conjecture is zero*:

$$Nash - Cournot \Leftrightarrow \frac{dQ_{(-i)t}}{dq_{it}} = 0 \tag{2.4}$$

Firm i takes as fixed the quantity produced by the rest of the firms, $Q_{(-i)t}$, and chooses his own output q_{it} to maximize his profit. Therefore, the first order condition of optimality under Nash-Cournot competition is:

$$MR_{it} = p_t + P'_Q\left(Q_t, X^D_t\right)\ q_{it} = MC_i(q_{it}) \tag{2.5}$$

We assume that the profit function is globally concave in q_{it} for any positive value of $Q_{(-i)t}$ such that there is a unique value of q_{it} that maximizes the firm's profit, and it is fully characterized by the marginal condition of optimality that establishes that marginal revenue equals marginal cost.

Consider the same specification of the cost function as before. Suppose that the demand function has been estimated in a first step such that there is a consistent estimate of the demand function. Therefore, the researcher can construct consistent estimates of marginal revenues $p_t + P'_Q\left(Q_t, X^D_t\right)\ q_{it}$ for every firm i. Then, the econometric model can be described in terms of the following linear regression model in logarithms:[1]

$$\ln\left(MR_{it}\right) = \theta\ \ln(q_{it}) + \alpha_1\ \ln(W_{1it}) + ... + \alpha_J\ \ln(W_{Jit}) + \varepsilon^{MC}_{it} \tag{2.6}$$

We are interested in the estimation of the parameter θ and α's and of the firms' relative efficiency, ε^{MC}_{it}.

OLS estimation of this regression function suffers of the same endogeneity problem as in the perfect competition case described above. The model implies that $E(\ln\left(q_{it}\right)\ \varepsilon^{MC}_{it}) \neq 0$, and more specifically there is a negative correlation between a firm's output and its unobserved inefficiency. To deal with this endogeneity problem, we can use instrumental variables. As in the case of perfect competition, we can use observable variables that affect demand but not costs, X^D_t, as instruments. With Cournot competition, we may have additional types of instruments.

Suppose that the researcher observes some exogenous input prices $W_{it} = (W_{1it}, ..., W_{Jit})$ and that at least one of these prices has cross-sectional variation over firms. For instance, suppose that there is information at the firm level on the firm's wage rate, or its capital stock, or its installed capacity. Note that, in equilibrium the input price of the competitors have an effect on the level of output of a firm. That is, given its own input prices W_{it}, log-output $\ln(q_{it})$ still depends on the input prices of other firms competing in the market, W_{jt} for $j \neq i$. A firm's output increases if, all else equal, the wage rates of a competitor increases. Note that the partial correlation between W_{jt} and $\ln(q_{it})$ is a testable condition. Under the assumption that the vector W_{jt} is exogenous, i.e., $E(W_{jt}\ \varepsilon^{MC}_{it}) = 0$, a natural

---

[1]For notational simplicity, here I omit the estimation error from the estimation of the demand function in the first step. Note that, in this case, this estimation error only implies measurement error in the dependent variable and it does not affect the consistency of the instrumental variables estimator described below or the estimation of robust standard errors.

approach to estimate this model is using IV or GMM based on moment conditions that use the characteristics of other firms as an instrument for output. For instance, the moment conditions can be:

$$E\left(\left[\begin{array}{c} \ln(W_{it}) \\ \sum_{j\neq i}\ln(W_{jt}) \end{array}\right] \; [\ln\left(MR_{it}\right) - \theta\;\ln(q_{it}) - W_{it}\;\alpha]\right) = \mathbf{0} \tag{2.7}$$

2.1.5. *An application.*

2.1.6.

2.2. Identification of the nature of competition: Conjectural variation model.

2.2.1. *Model.* Consider an industry where, at period t, the inverse demand curve is $p_t = P\left(Q_t, X_t^D\right)$, and firms, indexed by i, have cost functions $C_i(q_{it})$. Every firm i, chooses its amount of output, q_{it}, to maximize its profit, $p_t\;q_{it} - C_i(q_{it})$. Without further assumptions, the marginal condition for the profit maximization of a firm is marginal revenue equal to marginal cost, where the marginal revenue of firm i is:

$$MR_{it} = p_t + P_Q'\left(Q_t\;X_t^D\right)\;\left[1 + \frac{\partial Q_{(-i)t}}{\partial q_{it}}\right]\;q_{it} \tag{2.8}$$

As mentioned above, the term $\dfrac{\partial Q_{(-i)t}}{\partial q_{it}}$ represents the **belief** that firm i has about how the other firms in the market will respond if he changes its own amount of output marginally. We denote this **conjecture** or **belief** as the *conjectural variation* of firm i at period t, and denote it as $CV_{it} \equiv \dfrac{\partial Q_{(-i)t}}{\partial q_{it}}$.

As researchers, we can consider different assumptions about firms' beliefs or conjectural variations. Different assumptions on CVs imply different models of competition with their corresponding equilibrium outcomes. John Nash (1951) proposed the following conjecture: when a player constructs her best response, she believes that the other players will not response to a change in her decision. In the Cournot model, Nash conjecture implies that $CV_{it} = 0$. For every firm i, the "perceived" marginal revenue is $MR_{it} = p_t + P_Q'\left(Q_t\;X_t^D\right)\;q_{it}$, and the condition $p_t + P_Q'\left(Q_t\;X_t^D\right)\;q_{it} = MC_i(q_{it})$ implies the Cournot equilibrium.

There are CVs that generate the perfect competition equilibrium and the collusive or cartel equilibrium.

Perfect competition. For every firm i, $CV_{it} = -1$. Note that this conjecture implies that: $MR_{it} = p_t + P_Q'\left(Q_t\;X_t^D\right)\;[1 - 1]\;q_{it} = p_t$, and the conditions $p_t = MC_i(q_{it})$ imply the perfect competition equilibrium.

Collusion (Cartel). For every firm i, $CV_{it} = N_t - 1$. This conjecture implies, $MR_{it} = p_t + P_Q'\left(Q_t\;X_t^D\right)\;N_t\;q_{it}$, that generates that equilibrium conditions $p_t + P_Q'\left(Q_t\;X_t^D\right)\;N_t\;q_{it} = MC_i(q_{it})$. When firms have constant and homogeneous MCs, this condition implies

$p_t + P'_Q \left(Q_t \ X^D_t \right) Q_t = MC_t$, which is the equilibrium condition for the Monopoly (collusive or cartel) outcome.

The value of the beliefs / CV parameters are related to the nature of competition.

$$\text{Perfect competition:} \qquad CV_{it} = -1; \quad MR_{it} = p_t$$

$$\text{Nash-Cournot:} \qquad CV_{it} = 0; \quad MR_{it} = p_t + P'_Q \left(Q_t \right) \ q_{it} \tag{2.9}$$

$$\text{Cartel all firms:} \quad CV_{it} = N_t - 1; \quad MR_{it} = p_t + P'_Q \left(Q_t \right) \ Q_t$$

Given this result, one can argue that CV is closely related to the nature of competition, and therefore with equilibrium price and quantities. If CV is negative, the degree of competition is stronger than Cournot. The closer to -1, the more competitive. If CV is positive, the degree of competition is weaker than Cournot. The closer to $N_t - 1$, the less competitive.

Interpreting CV_{it} as an exogenous parameter is not correct. Conjectural variations represent firms' beliefs, and as such they are **endogenous outcomes** from the model.

2.2.2. *Estimation with information on marginal costs.* Consider an homogeneous product industry and a researcher with data on firms' quantities and marginal costs, and market prices over T periods of time: $\{p_t, MC_{it}, q_{it}\}$ for $i = 1, 2, ..., N_t$ and $t = 1, 2, ..., T$. Under the assumption that every firm chooses the amount of output that maximizes its profit given its belief CV_{it}, we have that the following condition holds:

$$p_t + P'_Q \left(Q_t \ X^D_t \right) \ [1 + CV_{it}] \ q_{it} = MC_{it} \tag{2.10}$$

And solving for the conjectural variation, we have:

$$CV_{it} = \frac{p_t - MC_{it}}{-P'_Q \left(Q_t \ X^D_t \right) \ q_{it}} - 1 = \left[\frac{p_t - MC_{it}}{p_t} \right] \ \left[\frac{1}{q_{it}/Q_t} \right] \ |\eta_t| - 1 \tag{2.11}$$

where η_t is the demand elasticity. Note that $\frac{p_t - MC_{it}}{p_t}$ is the Lerner index and q_{it}/Q_t is the market share of firm i. This equation shows that, given data on quantities, prices, demand and marginal costs, we can identify the firms' beliefs that are consistent with these data and with profit maximization. Let us denote $\left[\frac{p_t - MC_{it}}{p_t} \right] \ \left[\frac{1}{q_{it}/Q_t} \right]$ as the Lerner-index-to-market-share ratio of a firm. If the Lerner-index-to-market-share ratios are close zero, then the estimated values of CV will be close to -1 unless the absolute demand elasticity is large. In contrast, if the Lerner-index-to-market-share ratios are large (i.e., larger than the inverse demand elasticity), then estimated CV values will be greater than zero, and can reject the hypothesis of Cournot competition in favor of collusion.

Under the restriction that all the firms have the same marginal costs and conjectural variations, equation (2.11) that relates the Lerner index with the conjectural variation becomes:

$$\frac{p_t - MC_t}{p_t} = \left[\frac{1 + CV_t}{N_t} \right] \ \frac{1}{|\eta_t|} \tag{2.12}$$

where N_t is the number of firms in the market. This is the equation that we use in the empirical application that we describe at the end of this section. According to this expression, market power, as measured by the Lerner Index, depends on the elasticity of demand (negatively), the number of firms in the market (negatively), and the conjectural variation (positively).

2.2.3. *Estimation without information on marginal costs.* So far, we have considered the estimation of CV parameters when the researcher knows both demand and firms' marginal costs. We now consider the case where the researcher knows the demand, but it does not know firms' marginal costs. Identification of CVs requires also de identification of marginal costs. Under some conditions, we can jointly identify CVs and MCs using the marginal conditions of optimality and the demand.

The researcher observes data $\{p_t, \ q_{it}, \ X_t^D, \ X_t^{MC} : \ i = 1, ...N_t; t = 1, ..., T\}$, where X_t^D are variables affecting consumer demand, e.g., average income, population, and W_t are variables affecting marginal costs, e.g., some input prices. Consider the linear (inverse) demand equation:

$$p_t = \alpha_0 + \alpha_1 \ X_t^D - \alpha_2 \ Q_t + \varepsilon_t^D \tag{2.13}$$

with $\alpha_2 \geq 0$, and ε_t^D is unobservable to the researcher. Consider the marginal cost function:

$$MC_{it} = \beta_0 + \beta_1 \ W_t + \beta_2 \ q_{it} + \varepsilon_{it}^{MC} \tag{2.14}$$

with $\beta_2 \geq 0$, and ε_{it}^{MC} is unobservable to the researcher. Profit maximization implies $p_t + \frac{dP_t}{dQ_t} [1 + CV_{it}] \ q_{it} = MC_{it}$. Since the demand function is linear and $\frac{dP_t}{dQ_t} = -\alpha_2$, we have:

$$p_t = \beta_0 + \beta_1 \ W_t + [\beta_2 + \alpha_2(1 + CV_{it})] \ q_{it} + \varepsilon_{it}^{MC} \tag{2.15}$$

This equation describes the marginal condition for profit maximization. We assume now that $CV_{it} = CV$ for every observation i, t in the data. The structural equations of the model are the demand equation in (2.13) and the equilibrium condition in (2.15).

Using this model and data, can we identify (i.e., estimate consistently) the CV parameter? For the structural model described by equations (2.13) and (2.15), the answer to this question is negative. However, we will see that a simple modification of this model implies separate identification of CV and MC parameters. We first describe the identification problem.

Identification of demand parameters. The estimation of the regression equation for the demand function needs to deal with the well-known simultaneity problem. In equilibrium, output Q_t is correlated with the error term ε_t^D. The model implies a valid instrument to estimate demand. In equilibrium, Q_t depends on the exogenous cost variable W_t. This variable does not enter in the demand equation. If W_t is not correlated with ε_t^D, then this

variable satisfies all the conditions for being a valid instrument. Parameters α_0, α_1, and α_2 are identified using this IV estimator.

Identification of CV and MCs. In the regression equation (2.15) we also need to deal with a simultaneity problem. In equilibrium, output q_{it} is correlated with the error term ε_{it}^{MC}. The model implies a valid instrument to estimate this equation. In equilibrium, q_{it} depends on the exogenous demand shifter X_t^D. Note that X_t^D does not enter in the marginal cost and in the right hand side of the regression equation (2.15). If X_t^D is not correlated with ε_{it}^{MC}, then this variable satisfies all the conditions for being a valid instrument such that the parameters β_0, β_1, and $\gamma \equiv \beta_2 + \alpha_2(1 + CV)$ are identified using this IV estimator.

Note that we can identify the parameter $\gamma \equiv \beta_2 + \alpha_2(1 + CV)$ and the slope of inverse demand function, α_2. However, knowledge of γ and α_2 is not sufficient to identify separately CV and the slope of the MC, β_2. Given estimated values for γ and α_2, equation $\gamma = \beta_2 + \alpha_2(1 + CV)$ implies a linear relationship between CV and β_2 and there are infinite values of these parameters that satisfy this restriction. Even we restrict CV to belong to the values with a clear economic interpretation, such that $CV \in \{-1,\, 0,\, N-1\}$ and β_2 to greater or equal than zero, we do not have identification of these parameters. For instance, suppose that $N = 2$, $\gamma = 2$, and $\alpha_2 = 1$ such that we have the constraint $2 = \beta_2 + (1 + CV)$ or equivalently, $\beta_2 + CV = 1$. This equation is satisfied by any of the following forms of competition and values of $\beta_2 \geq 0$. Perfect competition: $CV = -1$ and $\beta_2 = 2$. Cournot competition: $CV = 0$ and $\beta_2 = 1$. And perfect collusion: $CV = N - 1 = 1$ and $\beta_2 = 0$.

Following Bresnahan (1981), we can provide a graphical representation of this identification problem. Suppose that we have followed the approach described above to estimate consistently demand parameters, marginal cost parameters β_0 and β_1, and the parameter γ. We can define two hypothetical marginal cost functions: the marginal cost under the hypothesis of perfect competition ($CV = -1$ such that $\beta_2 = \gamma$), $MC_c \equiv \beta_0 + \beta_1\, W + \gamma\, q$; and the marginal cost under the hypothesis of monopoly or perfect collusion ($CV = N - 1$ such that $\beta_2 = \gamma - \alpha_2 N$), $MC_m \equiv \beta_0 + \beta_1\, W + (\gamma - \alpha_2 N)\, q$. That is, MC_c and MC_m are the marginal cost functions that rationalize the observed values (p_t, q_{it}) under the hypotheses of perfect competition and monopoly, respectively. Figure 5.1 shows that the observed price and quantity in market 1, say (p_1, q_1), can be rationalized either as the point where the demand function D_1 crosses the marginal cost MC_c, or as the monopoly outcome defined by the marginal revenue MR_1 and the marginal cost MC_m.

Figure 5.1
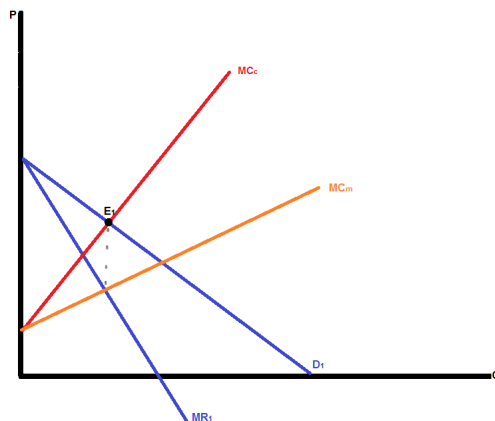
If the model is described by equations (2.13) and (2.15), the observation of prices and quantities (p_t, q_t) from multiple markets does not help to solve this identification problem. Suppose that we keep W_t constant such that the observations (p_t, q_t) for $t = 1, 2, ..., T$ are generated by different values of the demand shifters X_t^D and ε_t^D. This implies parallel vertical shifts in the demand curve and in the corresponding marginal revenue curve, as represented in Figure 5.2. As explained above for observation (p_1, q_1), all the observations $\{p_t, q_t : t = 1, 2, ..., T\}$ can be rationalized either as perfect competive equilibria that come from the intersection of demand curves $\{D_t : t = 1, 2, ..., T\}$ and marginal cost MC_c, or as monopoly outcomes that are determined by the intersection of the marginal revenue curves $\{MR_t : t = 1, 2, ..., T\}$ and the marginal cost MC_m.

Figure 5.2
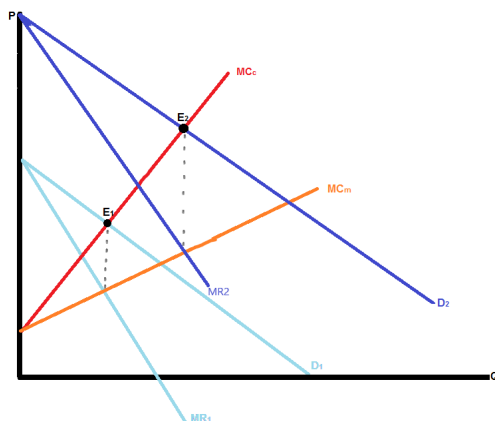
This graphical analysis provides also an intuitive interpretation of a solution to this identification problem. This solution involves generalizing demand function so that changes in exogenous variables do more than just a parallel shift in the demand curve and the marginal revenue. We introduce additional exogenous variables that are capable of **rotating** the demand curve. Consider Figure 5.3. We have two data points as represented by points

E_1 and E_2. Now, point E_2 is associated with a change in the demand curve that consists in a rotation around point E_1. Under perfect competition, this rotation in the demand curve should not have any effect in equilibrium prices and quantities. Therefore, under perfect competition the value of (P, q) in market 2 should be the same as in market 1. Since point E_2 is different to E_1, we can reject the hypothesis of perfect competition. Changes in the slope of the demand have and effect on prices and quantities only if firms market power.

Figure 5.3
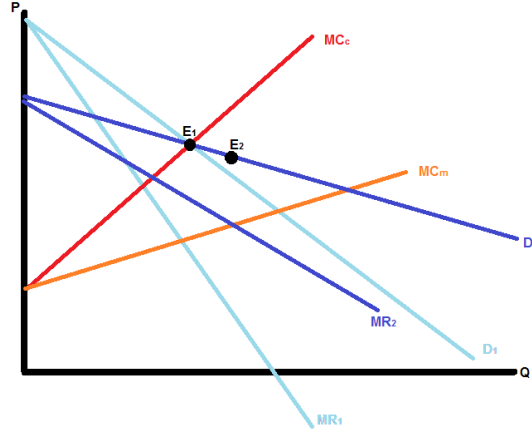
We now present more formally the identification of the model illustrated in Fugure 5.3. Consider now the following demand equation:

$$p_t = \alpha_0 + \alpha_1 \ X_t^D - \alpha_2 \ Q_t - \alpha_3 \left[R_t \ Q_t \right] + \varepsilon_t^D \tag{2.16}$$

R_t is an observable variable that affects the slope of the demand. Some possible candidates for these variables are the price of a substitute or complement product, seasonal variables, or the consumer demographics. The key condition is that the parameter α_3 is different to zero. That is, when R_t varies, there is a rotation in the demand curve, i.e., a change in the slope of the demand curve. Note that this condition is testable. Given this demand model, we have that $\frac{dP_t}{dQ_t} = -\alpha_2 - \alpha_3 \ R_t$, and the marginal condition for profit maximization implies the following regression model:

$$p_t = \beta_0 + \beta_1 \ W_t + \gamma_1 \ q_{it} + \gamma_2 \ (R_t \ q_{it}) + \varepsilon_{it}^{MC} \tag{2.17}$$

with $\gamma_1 \equiv \beta_2 + \alpha_2 \left[1 + CV \right]$ and $\gamma_2 \equiv \alpha_3 \left[1 + CV \right]$.

Equations (2.16) and (2.17) describe the structural model. Using this model and data, we can identify separately CV and MC parameters. Demand parameters can be identified similarly as before, using W_t as an instrument for output. Parameters α_0, α_1, α_2, and α_3 are identified using this IV estimator. The model also implies a valid instrument to estimate the parameters in the equilibrium equation in (2.17). We can instrument q_{it} using

X_t^P. Parameters β_0, β_1, γ_1, and γ_2 are identified. Note that $\gamma_1 = \beta_2 + \alpha_2 \left[1 + CV \right]$ and $\gamma_2 = \alpha_3 \left[1 + CV \right]$ such that that given γ_2 and α_3 we identify CV, and given γ_1, α_2, and CV we identify β_2. The identification of CV is very intuitive: $1 + CV = \gamma_2 / \alpha_3$. It measures the ratio between the sensitivity of price with respect to $(R_t \, q_{it})$ in the equilibrium equation relative to the sensitivty of price with respect to $(R_t \, Q_t)$ in the demand equation.

The sample variation in the slope of the inverse demand plays a key role in the identification of the CV parameter. An increase in the slope means that the demand becomes less price sensitive, more inelastic. For a monopolist, when the demand becomes more inelastic, the optimal price should increase. In general, for a firm with high level of market power (high CV), we should observe an important increase in prices associated with an increase in the slope. On the contrary, if the industry is characterized by very low market power (low CV) the increases in prices should be practically zero. Therefore, the response of prices to an exogenous change in the slope of the demand contains key information for the estimation of θ.

**** HERE ****

2.2.4. *An application: The sugar industry.* Genesove and Mullin (GM) study competition in the US sugar industry during the period 1890-1914. Why this period? The reason is that for this period they can collect high quality information on the value of marginal costs. Two aspects play are important in the collection of information on marginal costs. First, the production technology of refined sugar during this period was very simple and the marginal cost function can be characterized in terms of a simple linear function of the cost of raw sugar, the main intermediate input in the production of refined sugar. Most importantly, during this period there was an important investigation of the industry by the US anti-trust authority. As a result of that investigation, there are multiple reports from expert witnesses that provide estimates about the structure and magnitude of production costs in this industry.

As we describe below, GM use this information on marginal costs to test the validity of the standard conjectural variation approach for estimation of price cost margins and marginal costs. Here I describe briefly the main idea for this approach.

Let $p_t = P(Q_t)$ be the inverse demand function in the industry. Under the conjectural variation approach, and under the assumption that all the firms are identical in their marginal costs and in their conjectural variations, the marginal revenue at period t is:

$$MR_t = p_t - \left[1 + CV_t \right] \frac{Q_t}{N_t} \frac{dP(Q_t)}{dQ_t} \tag{2.18}$$

where $dP(Q_t)/dQ_t$ is the derivative of the inverse demand function. The condition for profit maximization (marginal revenue equals marginal cost) is $p_t - \left[1 + CV_t \right] \frac{Q_t}{N_t} \frac{dP(Q_t)}{dQ_t} = MC_t$,

and it implies the following condition for the Lerner Index:

$$\frac{p_t - MC_t}{p_t} = \left[\frac{1 + CV_t}{N_t}\right] \frac{1}{|\eta_t|} \tag{2.19}$$

Therefore, if we observe prices and can estimate the demand elasticity and the marginal cost, then we have a simple and direct estimate of the conjectural variation. Without information on MCs, the estimation of the CV should be based: (a) on our estimation of demand, and in particular, on exclusion restrictions that permit the identification of demand parameters; and (b) on our estimation of the MC function, on exclusion restrictions that permit the identification of this function. If assumptions (a) or (b) are not correct, our estimation of the CV and therefore of the Lerner Index, will be biased. GM evaluate these assumptions by comparing the estimates of CV using information on MCs and not using that information.

The rest of this section describes the following aspects of this empirical application: (a) The industry; (b) The data; (c) Estimates of demand parameters; and (d) Estimation of CV.

The industry. Homogeneous product industry. Highly concentrated during the sample period, 1890-1914. The industry leader, American Sugar Refining Company (ASRC), had more than 65% of the market share during most of these years.

Production technology. Refined sugar companies buy "raw sugar" from suppliers in national or international markets, transformed it into refined sugar, and sell it to grocers. They sent sugar to grocers in barrels, without any product differentiation. Raw sugar is 96% sucrose and 4% water. Refined sugar is 100% sucrose. The process of transforming raw sugar into refined sugar is called "melting", and it consists of eliminating the 4% of water in raw sugar. Industry experts reported that the industry is a "fixed coefficient" production technology:[2]

$$Q^{refined} = \lambda \, Q^{raw}$$

where $Q^{refined}$ is refined sugar output, Q^{raw} is the input of raw sugar, and $\lambda \in (0,1)$ is a technological parameter. That is, 1 ton of raw sugar generates λ tons units of refined sugar.

Marginal cost function. Given this production technology, the marginal cost function is:

$$MC = c_0 + \frac{1}{\lambda} \, p^{raw}$$

where p^{raw} is the price of the input raw sugar (in dollars per pound), and c_0 is a component of the marginal cost that depends on labor and energy. Industry experts unanimously report that the value of the parameter λ was close to 0.93, and c_0 was around \$0.26 per pound.

---

[2]Actually, the fixed coefficient Leontieff production function is $Q^{refined} = \min\{\lambda \, Q^{raw} \, ; \, f(L,K)\}$ where $f(L,K)$ is a function of labor and capital inputs. However, cost minimization will generally imply that $Q^{refined} = \lambda \, Q^{raw} = f(L,K)$.

Therefore, the marginal cost at period (quarter) t, in dollars per pound of sugar, was:

$$MC_t = 0.26 + 1.075 \; p_t^{raw}$$

The data. Quarterly US data for the period 1890-1914. The dataset contains 97 quarterly observations on industry output, price, price of raw sugar, imports of raw sugar, and a seasonal dummy.

$$\text{Data} = \{ \; Q_t, \; p_t, \; p_t^{raw}, \; IMP_t, \; S_t : t = 1, 2, ..., 97 \}$$

IMP_t represents the imports of raw sugar from Cuba. And S_t is a dummy variable for the Summer season: $S_t = 1$ is observation t is a Summer quarter, and $S_t = 0$ otherwise. The summer was a high demand season for sugar because most the production of canned fruits was concentrated during that season, and the canned fruit industry accounted for an important fraction of the demand of sugar.

Based on this data, we can also obtain a measure of marginal cost as $MC_t = 0.26 + 1.075 \; p_t^{raw}$.

Estimates of demand parameters. GM estimate four different models of demand. The main results are consistent for the four models. Here I concentrate on the linear demand, $Q_t = \beta_t \; (\alpha_t - p_t)$, and the inverse demand equation is:

$$p_t = \alpha_t - \frac{1}{\beta_t} \; Q_t$$

We can refer to β_t to the price sensitivity of demand, which is the inverse of the slope of the demand curve, i.e., higher price sensitivity implies a smaller slope of the demand curve. GM consider the following specification for α_t and β_t:

$$\alpha_t = \alpha_L \; (1 - S_t) + \alpha_H \; S_t + e_t^D$$

$$\beta_t = \beta_L \; (1 - S_t) + \beta_H \; S_t$$

α_L, α_H, β_L, and β_H are parameters. α_L and β_L are the intercept and the slope of the demand during the "Low Season" (when $S_t = 0$). And α_H and β_H are the intercept and the slope of the demand during the "High Season" (when $S_t = 1$). e_t^D is an error term that represents all the other variables that affect demand and that we do not observe. Therefore, we can write the following inverse demand equation:

$$p_t = \alpha_L + (\alpha_H - \alpha_L)S_t + \frac{1}{\beta_L}(-Q_t) + \left(\frac{1}{\beta_H} - \frac{1}{\beta_L} \right) (-S_t Q_t) + e_t^D$$

This is a regression equation where the explanatory variables are a constant term, S_t, $(-Q_t)$, and $(-S_t Q_t)$, and the parameters are α_L, $(\alpha_H - \alpha_L)$, $\frac{1}{\beta_L}$, and $\left(\frac{1}{\beta_H} - \frac{1}{\beta_L} \right)$. From the estimation of these parameters, we can recover α_L, α_H, β_L, and β_H.

As we have discussed before, Q_t is an endogenous regressor in this regression equation. We need to use IV to deal with this endogeneity problem. In principle, it seems that we

could we p_t^{raw} as an instrument. However, GM have a reasonable concern about the validity of this instrument. The demand of raw sugar from the US accounts for a significant fraction of the world demand of raw sugar. Therefore, exogenous shocks in the demand of refined sugar (e_t^D) might generate an increase if the world demand of raw sugar and in p_t^{raw} such that $Cov(e_t^D, p_t^{raw}) \neq 0$. Instead they use imports of raw sugar from Cuba as an instrument: almost 100% of the production of raw sugar in Cuba was exported to US, and the authors claim that variations in Cuban production of raw sugar was driven by supply/weather conditions and not by the demand from US.

These are the parameter estimates.

Demand Estimates		
Parameter	Estimate	Standard Error
α_L	5.81	(1.90)
α_H	7.90	(1.57)
β_L	2.30	(0.48)
β_H	1.36	(0.36)

In the high season the demand shifts upwards and becomes less elastic. The estimated price elasticities of demand in the low and the high season are $|\eta_L| = 2.24$ and $|\eta_H| = 1.04$, respectively. According to this, any model of oligopoly competition where firms have some market power predicts that the price cost margin should increase during the price season due to the lower price sensitivity of demand.

Before we discuss the estimates of the conjectural variation parameter, it is interesting to illustrate the errors that researchers can make if in the absence of information about marginal costs they estimate price cost margins by making an adhoc assumption about the value of CV in the industry. As mentioned above, the industry was highly concentrated during this period. Though there were approximately 6 firms active during most of the sample period, one of the firms accounted for more than two-thirds of total output. Suppose three different researchers of this industry, that we label as researchers M, C, and S. Researcher M considers that the industry was basically a Monopoly/Cartel during this period (in fact, there was anti-trust investigation, so there may be some suspicions of collusive behavior). Therefore, he assumes that $[1 + CV]/N = 1$. Researcher C considers that the industry can be characterized by Cournot competition between the 6 firms, such that $[1 + CV]/N = 1/6$. Finally, researcher S thinks that this industry can be better described by a Stackelberg model with 1 leader and 5 Cournot followers, and therefore $[1 + CV]/N = 1/(2*6 - 1) = 1/11$. What are the respective predictions of these researchers about market power as measured by the Lerner index? The following table presents the researchers' predictions and also the

actual value of the Lerner index based on our information on marginal costs (that we assume is not available for these 3 researchers). Remember that $Lerner = \frac{p-MC}{p} = \left[\frac{1+CV}{N}\right]\frac{1}{|\eta|}$.

Predicted Market Power Based on Different Assumptions on $\frac{1+CV}{N}$				
Assumed $\frac{1+CV}{N}$	Predicted Lerner	Actual Lerner	Predicted Lerner	Actual Lerner
	Low season: $\left[\frac{1+CV}{N\,|\eta_L|}\right]$	Low season: $\frac{p_L-MC}{p_L}$	High season: $\left[\frac{1+CV}{N\,|\eta_H|}\right]$	High season: $\frac{p_H-MC}{p_H}$
Monopoly: $\frac{1+CV}{N}=1$	$\frac{1}{2.24}=44.6\%$	3.8%	$\frac{1}{1.04}=96.1\%$	6.5%
Cournot: $\frac{1+CV}{N}=\frac{1}{6}$	$\frac{1/6}{2.24}=7.4\%$	3.8%	$\frac{1/6}{1.04}=16.0\%$	6.5%
Stackelberg: $\frac{1+CV}{N}=\frac{1}{11}$	$\frac{1/11}{2.24}=4.0\%$	3.8%	$\frac{1/11}{1.04}=8.7\%$	6.5%

This table shows that the researcher M will make a very seriously biased prediction of market power in the industry. Since the elasticity of demand is quite low in this industry, especially during the high season, the assumption of Cartel implies a very high Lerner index, much higher than the actual one. Researcher C also over-estimates the actual Lerner index. The estimates of researcher S are only slightly upward biased.

Consider the judge of an anti-trust case where there is very little reliable information on the actual value of MCs. The picture of industry competition that this judge gets from the three researchers is very different. This judge would be interested in measures of market power in this industry that do not depend on an adhoc assumption about the value of CV.

Estimation of conjectural variation. Suppose that we do not observe the MC and we use the approach described Section 2 to estimate the CV and then the lerner index. The condition marginal revenue equal to marginal cost implies the following equation:

$$p_t = c_0 + c_1\, p_t^{raw} + \theta\, \frac{Q_t}{\beta_t} + e_t^{MC}$$

with $\theta \equiv (1+CV)/N$. We treat c_0 and c_1 (the parameters in the marginal cost function) as parameter to estimate because we do not know that $c_0 = 0.26$ and $c_0 = 1.075$. We interpret e_t^{MC} as an error term in the marginal cost. After the estimation of the demand equation, we have $\hat{\beta}_t = 2.30(1 - S_t) + 1.36\, S_t$. Therefore, we estimate the equation:

$$p_t = c_0 + c_1\, p_t^{raw} + \theta\, \frac{Q_t}{\hat{\beta}_t} + e_t^{MC}$$

Since Q_t is endogeneously determined, it should be correlated with e_t^{MC}. To deal with this endogeneity problem, GM use instrumental variables. Again, the use imports from Cuba as an instrument for Q_t. In principle, they might have considered the seasonal dummy S_t as an instrument, but they were probably concerned that there may be also seasonality in the marginal cost such that e_t^{MC} and S_t might be correlated (e.g., wages of seasonal

workers). The following table presents these IV estimates of c_0, c_1 and θ, their standard errors (in parentheses) and the "true" values of these parameters based on the information on marginal costs.

Estimates of Marginal Costs and θ		
Parameter	Estimate (s.e.)	"True" value$^{(Note)}$
$\frac{1+CV}{N}$	0.038 (0.024)	0.10
c_0	0.466 (0.285)	0.26
c_1	1.052 (0.085)	1.075

The "true" value of $\frac{1+CV}{N}$ using information of MC is obtained using the relationship $\frac{1+CV}{N} = \left(\frac{p-MC}{p}\right)|\eta|$. The estimates of $\frac{1+CV}{N}$, c_0, and c_1, are not too far from their "true" values. This seems a validation of the CV approach for this particular industry. Based on this estimate of $\frac{1+CV}{N}$, the predicted values for the Lerner index in the low season is $\left[\frac{1+CV}{N}\right]\frac{1}{|\eta_L|} = \frac{0.038}{2.24} = 1.7\%$, and the predicted Lerner Index in high season $\left[\frac{1+CV}{N}\right]\frac{1}{|\eta_H|} = \frac{0.038}{1.04} = 3.6\%$. Remember that the true values of the Lerner index using information on marginal costs were 3.8% in the low season and 6.5% in the high season. Therefore, the estimates using the CV approach under-estimate the actual market power in the industry, but by a relatively small magnitude.

3. Differentiated product industry

3.1. Model. Consider an industry with J differentiated products (e.g., automobiles) indexed by $j \in \mathcal{J} = \{1, 2, ..., J\}$. Consumer demand for each of these products can be represented using the demand system:

$$q_j = D_j(\mathbf{p}, \mathbf{x}) \qquad \text{for } j \in \mathcal{J} \tag{3.1}$$

where $\mathbf{p} = (p_1, p_2, ..., p_J)$ is the vector of product prices, and $\mathbf{x} = (x_1, x_2, ..., x_J)$ is a vector of other product attributes. There are F firms in the industry, indexed by $f \in \{1, 2, ..., F\}$. Each firm f owns a subset $\mathcal{J}_f \subset \mathcal{J}$ of the brands. The profit of firm f is:

$$\Pi_f = \sum_{j \in \mathcal{J}_f} p_j \, q_j - C_j(q_j) \tag{3.2}$$

where $C_j(q_j)$ is the cost of producing a quantity q_j of product j. Firms compete in prices.

Bertrand. For the moment, we assume Nash-Bertrand competition: each firm chooses its own prices to maximize profits and takes the prices of other firms as given. The first order

conditions of optimality for profit maximization of firm f are: for any $j \in \mathcal{J}_f$

$$q_j + \sum_{k \in \mathcal{J}_f} [p_k - MC_k] \ \frac{\partial D_k}{\partial p_j} = 0 \qquad (3.3)$$

where MC_j is the marginal cost $C_j'(q_j)$. We can write this system in vector form. For firm f:

$$\mathbf{q}^f + \Delta \mathbf{D}^f \ \left[\mathbf{p}^f - \mathbf{MC}^f\right] = 0 \qquad (3.4)$$

where $\mathbf{q}^f$, $\mathbf{p}^f$, and $\mathbf{MC}^f$ are column vectors with the quantities, prices, and marginal costs, respectively, for every product $j \in \mathcal{J}_f$, and $\Delta \mathbf{D}^f$ is the square martrix with the demand-price derivatives $\frac{\partial D_k}{\partial p_j}$ for every $j, k \in \mathcal{J}_f$. Solving for price-cost margins in this system:

$$\mathbf{p}^f - \mathbf{MC}^f = - \left[\Delta \mathbf{D}^f\right]^{-1} \mathbf{q}^f \qquad (3.5)$$

The RHS of this equation depends only on demand parameters, not costs. Given an estimated demand system, the vector of Price-Cost Margins under Nash-Bertrand competition (and a particular ownership structure of brands),is known to the researcher.

EXAMPLE. Single product firms & Logit model.   For single product firms, the marginal condition of optimality is:

$$p_j - MC_j = - \left[\frac{\partial D_j}{\partial p_j}\right]^{-1} q_j \qquad (3.6)$$

In the logit demand system, we have that:

$$D_j\left(\mathbf{p}, \mathbf{x}\right) = H \ \frac{\exp\left\{x_j'\beta - \alpha \ p_j\right\}}{1 + \sum_{k=1}^{J} \exp\left\{x_k'\beta - \alpha \ p_k\right\}} \qquad (3.7)$$

where H represents market size, and β and α are parameters. This demand system implies that $\frac{\partial D_j}{\partial p_j} = -\alpha \ H \ s_j(1 - s_j)$ where s_j is the market share $s_j \equiv q_j/H$. Therefore, in this model:

$$PCM_j \equiv p_j - MC_j = \frac{1}{\alpha(1 - s_j)} \qquad (3.8)$$

We see that in this model the price-cost margin of a firm declines with the price sensitivity of demand, α, and increases with the own market share, s_j.   ∎

EXAMPLE. Logit model with Multi-product firms. With multiproduct firms we have that, the F.O.C. is $q_j + \sum_{k \in \mathcal{J}_f} PCM_k \ \frac{\partial D_k}{\partial p_j} = 0$. In Logit demand system: $\frac{\partial D_j}{\partial p_j} = -\alpha \ H \ s_j(1 - s_j)$ and for $k \neq j$, $\frac{\partial D_j}{\partial p_k} = \alpha \ H \ s_j \ s_k$. And this implies:

$$PCM_j \ = \ \frac{1}{\alpha} + \sum_{k \in \mathcal{J}_f} PCM_k \ s_k \qquad (3.9)$$

The RHS is firm specific but it does not vary across products within the same firm. This condition implies that all the products owned by a firm have the same price-cost margin. This condition implies that the price-cost margin is:

$$PCM_j = \overline{PCM}_f = \frac{1}{\alpha \left(1 - \sum_{k \in \mathcal{J}_f} s_k \right)} \tag{3.10}$$

For the Logit demand model, a multi-product firm charges the same price-cost margin to all its products. This prediction does not extend to more general/flexible demand systems. Note also that a multi-product firm charges higher prices than a single-product firm:

$$\frac{1}{\alpha \left(1 - \sum_{k \in \mathcal{J}_f} s_k \right)} > \frac{1}{\alpha \left(1 - s_j \right)} \tag{3.11}$$

This prediction is robust and it extends to Bertrand competition when products are substitutes.   ∎

Multiproduct as source of market power. We can write F.O.C. for firm f product j as:

$$
\begin{aligned}
PCM_j &= \left[\frac{-\partial D_j}{\partial p_j} \right]^{-1} q_j \\
\\
&+ \left[\frac{-\partial D_j}{\partial p_j} \right]^{-1} \left[\sum_{k \in J_f;\ k \neq j} PCM_k \frac{\partial D_k}{\partial p_j} \right]
\end{aligned}
\tag{3.12}
$$

With substitutes, $\dfrac{\partial D_k}{\partial p_j} > 0$ for $k \neq j$, and the second term is positive. Selling multiple products contribute to increase the price-cost margin of each of the products.

Collusion and other ownership structures. Suppose that there is collusion between some or all the firms. We can represent a collusive setting as a partition of the set of firms, into a number R of groups or "rinks". Let $\mathcal{F} = \{1, 2, ..., F\}$ be the set of all the firms, and let $\mathcal{R}_1, \mathcal{R}_2, ..., \mathcal{R}_R$ be a partition of the set $\mathcal{F}$ such that $\mathcal{R}_1 \cup \mathcal{R}_2 \cup ... \cup \mathcal{R}_R = \mathcal{F}$ and $\mathcal{R}_r \cap \mathcal{R}_{r'} = \varnothing$ for $r \neq r'$. According to this particition, a firm belongs to one and only one rink. We also use $\mathcal{R}(f)$ to denote the rink the rink to which firm f belongs.

A collusion rink, together with the ownership structure of the products, implies a set of products from all the firms in the rink. Then, we can define:

$$\mathcal{J}_{\mathcal{R}(f)} \equiv \{j : j \in \mathcal{J}_{f'} \text{ for some } f' \in \mathcal{R}(f)\} \equiv \bigcup\nolimits_{f' \in \mathcal{R}(f)} \mathcal{J}_{f'}$$

We also define the dummy variables $\Theta_j^{\mathcal{R}(f)} \equiv 1\{j \in \mathcal{J}_{\mathcal{R}(f)}\}$. No extreme example are no collusion and collusion of all the firms. With no collusion we have that the number of rinks R is equal to the number of firms F, and $\mathcal{J}_{\mathcal{R}(f)} = \mathcal{J}_f$, and $\Theta_j^{\mathcal{R}(f)} = 1\{j \in \mathcal{J}_f\}$. With collusion

of all firms, we have that there is only one rink, $R = 1$, $\mathcal{R}_1 = \mathcal{F}$, $\mathcal{J}_{\mathcal{R}(1)} = \mathcal{J}$, and all the indicators $\Theta_j^{\mathcal{R}}$ are equal to one.

Firm f maximizes its collusion rink profit:

$$\sum_{j=1}^{J} \Theta_j^{\mathcal{R}(f)} \left[p_j \; q_j - C_j(q_j) \right]$$

The F.O.C.s for firm f: for $j \in \mathcal{J}_f$

$$q_j + \sum_{k=1}^{J} [p_k - MC_k] \; \Theta_k^{R(f)} \frac{\partial D_k}{\partial p_j} = 0$$

In vector form, using all the products that belong to the collusion rink $R(f)$

$$\mathbf{q}^{R(f)} + \left[\Delta \mathbf{D}^{R(f)} \right] \left[\mathbf{PCM}^{R(f)} \right] = 0$$

$\Delta \mathbf{D}^{R(f)} =$ matrix of demand-price derivatives $\dfrac{\partial D_k}{\partial p_j}$ for every j, k in the collusion rink of firm f. Such that:

$$\mathbf{PCM}^{R(f)} = -\left[\Delta \mathbf{D}^{R(f)} \right]^{-1} \mathbf{q}^{R(f)}$$

3.2. Estimating MCs based on assumption on form of competition. The researcher has data from J products over T markets, and knows the ownership structure:

$$\text{Data} = \{ p_{jt}, \, q_{jt}, \, x_{jt} : \; j = 1, ..., J; \; t = 1, 2, ..., T \}$$

Suppose that the demand function has been estimated in a fist step, such that there is a consistent estimator of the demand system $D_j \left(\mathbf{p}_t, \mathbf{x}_t \right)$. For every firm f, the research has an estimate of vector $-\left[\Delta \mathbf{D}_t^f \right]^{-1} \mathbf{q}_t^f$ for every firm f. Therefore, under the assumption of Bertrand competition she has consistent estimates of the vectors of MCs:

$$\mathbf{MC}_t^f = \mathbf{p}_t^f + \left[\Delta \mathbf{D}_t^f \right]^{-1} \mathbf{q}_t^f$$

Similarly, given an hypothetical collusion rink $R(f)$ represented by the indicators $\Theta_j^{R(f)}$, the researcher can construct $\left[\Delta \mathbf{D}^{R(f)} \right]^{-1} \mathbf{q}^{R(f)}$ and obtain the estimate of marginal costs:

$$\mathbf{MC}_t^{R(f)} = \mathbf{p}_t^{R(f)} + \left[\Delta \mathbf{D}_t^{R(f)} \right]^{-1} \mathbf{q}_t^{R(f)}$$

Different hypothesis about collusion, or ownership structures of products (e.g., mergers), imply different Price-Cost margins and different estimates of marginal costs. After estimating the realized values of MCs, we can estimate the marginal cost function.

Consider the following cost function:

$$C(q_{jt}) = \frac{1}{\gamma + 1} q_{jt}^{\gamma+1} \; \exp\{x_{jt}'\alpha + \omega_{jt}\}$$

Such that:

$$MC_{jt} = q_{jt}^{\gamma} \exp\{x_{jt}'\alpha + \omega_{jt}\}$$

where ω_{jt} is unobservable to the researcher. The econometric model is:

$$\ln(MC_{jt}) = \gamma \ \ln(q_{jt}) + x'_{jt}\alpha + \omega_{jt}$$

We are interested in the estimation of the parameters α and γ.

Endogeneity: The equilibrium model implies that $E(\ln(q_{tj}) \ \omega_{jt}) \neq 0$. Firms/products with larger ω_{jt} are less efficient in terms of costs (or products are more costly to produce), and this, all else equal, implies a smaller amount of output.

Instrumental variables. Suppose that $E(x_{kt} \ \omega_{jt}) = 0$ for any (k, j). We can use as instruments for $\ln(q_{jt})$ the characteristics of other firms/products.

$$E\left(\begin{bmatrix} x_{jt} \\ \sum_{k \neq j} x_{kt} \end{bmatrix} \left[\ln(MC_{jt}) - \gamma \ \ln(q_{jt}) - x'_{jt}\alpha\right]\right) = \mathbf{0}$$

3.3. Testing the nature of competition. Suppose that the researcher observes the true MC_{jt}. Or more realistically, observes a measure of costs, S_{obs}^{MC}, e.g., the mean value of the MCs of all products and firms in the industry; the mean value of the MC of one particular firm. Given an estimated demand system and an hypothesis about collusion, represented by a matrix of collusion rink dummies $\mathbf{\Theta}^R = \{\Theta_j^{R(f)}\}$, we can obtain the MCs under this hypothesis: $MC_j(\mathbf{\Theta}^R)$.

Let $S^{MC}(\mathbf{\Theta}^R)$ the value of the statistic (e.g., mean value of all MCs) under the hypothesis $\mathbf{\Theta}^R$. We can use $S^{MC}(\mathbf{\Theta}^R)$ and S_{obs}^{MC} to construct a test of the null hypothesis $\mathbf{\Theta}^R$. For instance, if S^{MC} is a vector of sample means, we could use a Chi-square test. This is the approach in Nevo (2001). It is possible to consider $\Theta_j^{R(f)}$ as parameters to estimate, similarly as the conjectural variation parameters in the homogeneous product case. Using the estimated demand, our specification of the MC function, and the F.O.C.s of profit maximization, it is possible to jointly identify $\Theta_j^{R(f)}$ and parameters in MCs. We need similar rotation demand variables as in the homogeneous demand case (Nevo, 1998).

Testing form of competition: Without info on MCs. Instead of estimating $\Theta_j^{R(f)}$ some papers have used non-nested hypothesis tests to test null hypothesis of Collusion against the alternative of Bertrand (or viceversa). The most commonly used non-nested tests procedures are: Cox-Test and Vuong-Test. Davidson & McKinnon provide an intuitive interpretation of these tests: Obtain residuals from the model under H_0; Run regression of the residuals on variable in the model under H_1; Under null, #obs $\times$ R-square of this regression is Chi-square.

3.3.1. *An application: Competition and Collusion in the American Automobile Industry.* Bresnahan (1987) : Descriptive Stats

TABLE I

Year	(1) Auto Production[a]	(2) Real Auto Price-CPI[b]	(3) % Change Auto Price-Cagan[c]	(4) Auto Sales[d]	(5) Auto Quantity Index[e]
1953	6.13	1.01	NA	14.5	86.8
1954	5.51	0.99	NA	13.9	84.9
1955	7.94	0.95	−2.5	18.4	117.2
1956	5.80	0.97	6.3	15.7	97.9
1957	6.12	0.98	6.1	16.2	100.0

Notes: [a] Millions of units over the model year. [Source: *Automotive News*.]
[b] (CPI New automobile component)/CPI. [Source: *Handbook of Labor Statistics*.]
[c] Adjusted for quality change. [See Cagan (1971), especially pp. 232–3.]
[d] Auto output in constant dollars, QIV of previous year through $QIII$ of named year, in billions of 1957 dollars. [Source: *National Income and Product Accounts*.]
[e] (4)/(2), normalized so 1957 = 100.

Bresnahan (1987): Non-nested tests of conduct

TABLE III
COX TEST STATISTICS

Hypotheses	C	N−C	'p'	H
a—1954				
Collusion	—	0.8951	0.9464	−1.934
Nash-Competition	−2.325	—	−0.8878	−2.819
"Products"	−3.978	3.029	—	−1.604
Hedonic	−12.37	−10.94	−13.02	—
b—1955				
Collusion	—	−10.36	−9.884	−13.36
Nash-Competition	−1.594	—	1.260	0.6341
"Products"	−0.7598	−4.379	—	−1.527
Hedonic	−3.353	−8.221	−5.950	—
c—1956				
Collusion	—	1.227	0.8263	1.629
Nash-Competition	−2.426	—	−4.586	0.8314
"Products"	−3.153	0.9951	—	4.731
Hedonic	−5.437	−9.671	−11.58	—

Estimates Demand & MCs: Collusion 1954 & 1956, Collusion 1955

TABLE IV
PARAMETER ESTIMATES 1954–56, <u>MAINTAINED</u> SPECIFICATION

Parameters	1954[a]	1955[b]	1956[a]
Physical Characteristics			
Quality Proxies			
Constant	47.91	48.28	50.87
	(32.8)	(43.2)	(29.4)
Weight #/1000	0.3805	0.5946	0.5694
	(0.332)	(0.145)	(0.374)
Length "/1000	0.1819	0.1461	0.1507
	(0.128)	(0.059)	(0.146)
Horsepower/100	2.665	3.350	3.248
	(0.692)	(0.535)	(0.620)
Cylinders	−0.7387	−0.9375	−0.9639
	(0.205)	(0.115)	(0.186)
Hardtop Dummy	0.9445	0.4531	0.4311
	(0.379)	(0.312)	(0.401)
Demand/Supply			
μ—Marginal Cost	0.1753	0.1747	0.1880
	(0.024)	(0.020)	(0.035)
γ—Lower Endpoint	4.593	3.911	4.441
	(1.49)	(2.08)	(1.46)
V_{max}—Upper Endpoint	$1.92E+7$	$2.41E+7$	$2.83E+7$
	$(8.44E+6)$	$(9.21E+6)$	$(7.98E+6)$
δ—Taste Density	0.4108	0.4024	0.4075
	(0.138)	(0.184)	(0.159)

Estimates Demand & MCs: Bertrand 1954, 1955, 1956. The estimated structural model under the maintained assumption of collusion in years 1954 & 1956 and Bertrand competition

in1955 implies very stable coefficient estimates over the three years. That is, the observed changes in quantity and prices in 1955 can be fully explained by the change in conduct, and not by a change in demand or costs parameters. Instead, the models that impose Collusion over the three years, or Bertrand over the three years imply estimates of structural parameters with strong and implausible changes in demand and costs in year 1955.

3.3.2. *An application: Nevo (2001) on Ready-to-eat cereals.* Ready-to-Eat (RTE) cereal market: highly concentrated; many apparently similar products, and yet price-cost margins (PCM) are high. What are the sources of market power? Product differentiation? Multi-product firms? Collusion? Nevo: (1) estimates a demand system of differentiated products for this industry; (2) recovers PCMs and compare them to rough/aggregate estimates of PCM at the industry level; (3) based on this comparison, tests Bertrand vs (full) Collusion [and rejects collusion]; (4) Under Bertrand, compares estimated PCMs with the counterfactual with single-product firms.

Data. A market is a city-quarter. IRI data on market shares and prices. 65 cities x 20 quarters [Q188-Q492] x 25 brands [total share is 43-62%]. Most of the price variation is cross-brand (88.4%), the remainder is mostly cross-city, and a small amount is cross-quarter. Relatively poor brand characteristics so model includes brand fixed effects.

TABLE I

VOLUME MARKET SHARES

	88Q1	88Q4	89Q4	90Q4	91Q4	92Q4
Kellogg	41.39	39.91	38.49	37.86	37.48	33.70
General Mills	22.04	22.30	23.60	23.82	25.33	26.83
Post	11.80	10.30	9.45	10.96	11.37	11.31
Quaker Oats	9.93	9.00	8.29	7.66	7.00	7.40
Ralston	4.86	6.37	7.65	6.60	5.45	5.18
Nabisco	5.32	6.01	4.46	3.75	2.95	3.11
C3	75.23	72.51	71.54	72.64	74.18	71.84
C6	95.34	93.89	91.94	90.65	89.58	87.53
Private Label	3.33	3.75	4.63	6.29	7.13	7.60

Source: IRI Infoscan Data Base, University of Connecticut, Food Marketing Center.

TABLE III

DETAILED ESTIMATES OF PRODUCTION COSTS

Item	$/lb	% of Mfr Price	% of Retail Price
Manufacturer Price	2.40	100.0	80.0
Manufacturing Cost:	1.02	42.5	34.0
Grain	0.16	6.7	5.3
Other Ingredients	0.20	8.3	6.7
Packaging	0.28	11.7	9.3
Labor	0.15	6.3	5.0
Manufacturing Costs (net of capital costs)[a]	0.23	9.6	7.6
Gross Margin		57.5	46.0
Marketing Expenses:	0.90	37.5	30.0
Advertising	0.31	13.0	10.3
Consumer Promo (mfr coupons)	0.35	14.5	11.7
Trade Promo (retail in-store)	0.24	10.0	8.0
Operating Profits	0.48	20.0	16.0

[a] Capital costs were computed from ASM data.
Source: Cotterill (1996) reporting from estimates in CS First Boston Reports "Kellogg Company," New York, October 25, 1994.

TABLE VI

RESULTS FROM THE FULL MODEL[a]

Variable	Means (β's)	Standard Deviations (σ's)	Interactions with Demographic Variables:			
			Income	Income Sq	Age	Child
Price	−27.198	2.453	315.894	−18.200	—	7.634
	(5.248)	(2.978)	(110.385)	(5.914)		(2.238)
Advertising	0.020	—	—	—	—	—
	(0.005)					
Constant	−3.592[b]	0.330	5.482	—	0.204	—
	(0.138)	(0.609)	(1.504)		(0.341)	
Cal from Fat	1.146[b]	1.624	—	—	—	—
	(0.128)	(2.809)				
Sugar	5.742[b]	1.661	−24.931	—	5.105	—
	(0.581)	(5.866)	(9.167)		(3.418)	
Mushy	−0.565[b]	0.244	1.265	—	0.809	—
	(0.052)	(0.623)	(0.737)		(0.385)	
Fiber	1.627[b]	0.195	—	—	—	−0.110
	(0.263)	(3.541)				(0.0513)
All-family	0.781[b]	0.1330	—	—	—	
	(0.075)	(1.365)				
Kids	1.021[b]	2.031	—	—	—	
	(0.168)	(0.448)				
Adults	1.972[b]	0.247	—	—	—	
	(0.186)	(1.636)				
GMM Objective (degrees of freedom)		5.05 (8)				
MD χ^2		3472.3				
% of Price Coefficients > 0		0.7				

[a] Based on 27,862 observations. Except where noted, parameters are GMM estimates. All regressions include brand and time dummy variables. Asymptotically robust standard errors are given in parentheses.
[b] Estimates from a minimum-distance procedure.

Direct measure of mean value of the price-cost margin in the industry: 31%

TABLE VIII

MEDIAN MARGINS[a]

	Logit (Table V column ix)	Full Model (Table VI)
Single Product Firms	33.6%	35.8%
	(31.8%−35.6%)	(24.4%−46.4%)
Current Ownership of 25 Brands	35.8%	42.2%
	(33.9%−38.0%)	(29.1%−55.8%)
Joint Ownership of 25 Brands	41.9%	72.6%
	(39.7%−44.4%)	(62.2%−97.2%)
Current Ownership of All Brands	37.2%	—
	(35.2%−39.4%)	
Monopoly/Perfect Price Collusion	54.0%	—
	(51.1%−57.3%)	

[a] Margins are defined as $(p − mc)/p$. Presented are medians of the distribution of 27,862 (brand-city-quarter) observations. 95% confidence intervals for these medians are reported in parentheses based on the asymptotic distribution of the estimated demand coefficients. For the Logit model the computation is analytical, while for the full model the computation is based on 1,500 draws from this distribution.

3.4. Conjectural variation model with differentiated products. Consider an industry with a differentiated product. There are two firms in this industry, firm 1 and firm 2.

Each firm produces and sells only one brand of the differentiated product: brand 1 is produced by firm 1, and brand 2 is produced by firm 2. The demand system has the structure of a logit demand model, where consumers choose between three different alternatives: $j = 0$, represents the consumer decision of not purchasing any of the two products; and $j = 1$ and $j = 2$ represent the consumer purchase of product 1 and 2, respectively. The utility of no purchase ($j = 0$) is zero. The utility of purchasing product $j \in \{1, 2\}$ is $\beta\ x_j - \alpha\ p_j + \varepsilon_j$, where the variables and parameters have the interpretation that we have seen in class. Variable x_j is a measure of the quality of product j, e.g., the number of stars of the product according to consumer ratings. Therefore, we have that $\beta > 0$. The random variables ε_1 and ε_2 are independently and identically distributed over consumers with a type I extreme value distribution, i.e., Logit model of demand. Let H be the number of consumers in the market. We define the market shares s_0, s_1, and s_2 such that $s_0 + s_1 + s_2 = 1$ and s_j represents the proportion of consumers choosing alternative j.

The logit model implies that the market share of product 1, s_1, is:

$$s_1 = \frac{\exp\{\beta\ x_1 - \alpha\ p_1\}}{1 + \exp\{\beta\ x_1 - \alpha\ p_1\} + \exp\{\beta\ x_2 - \alpha\ p_2\}}$$

The profit function of firm $j \in \{0, 1\}$ is $\pi_j = p_j\ q_j - c_j\ q_j$, where: q_j is the quantity sold by firm j (i.e., $q_j = H\ s_j$); and c_j is firm $j's$ marginal cost, that is assumed constant, i.e., linear cost function.

Suppose that firms take their qualities x_1 and x_2 as given and compete in prices ala Bertrand. The F.O.C. implies:

$$p_1 - c_1 = \frac{1}{\alpha\ (1 - s_1)}$$

Now, suppose that the researcher is not willing to impose the assumption of Bertrand competition and considers a conjectural variations model. Define the conjecture parameter CV_1 as the belief or conjecture that firm 1 has about how firm 2 will change its price when firm 1 changes marginally its price. That is, CV_1 represents the belief or conjecture of firm 1 about $\frac{\partial p_2}{\partial p_1}$. Similarly, CV_2 represents the belief or conjecture of firm 2 about $\frac{\partial p_2}{\partial p_1}$. Then, the f.o.c. for profit maximization is:

$$q_1 + (p_1 - c_1)\left[\frac{\partial q_1}{\partial p_1} + \frac{\partial q_1}{\partial p_2}CV_1\right] = 0$$

Solving for the price-cost margin, we have that:

$$p_1 - c_1 = \frac{1}{\alpha\ (1 - s_1 - s_2 CV_1)}$$

Suppose that the researcher does not know the magnitude of the marginal costs c_1 and c_2, but she knows that the two firms use the same production technology, the same type of variable inputs, and purchase these inputs in the same markets where they are price

takers. Therefore, the researcher knows that $c_1 = c_2 = c$, though she does not know the magnitude of c. The marginal conditions for profit maximization for the two firms, together with the condition $c_1 = c_2 = c$, imply that price difference between these two firms, $p_1 - p_2$, is a particular function of their markets shares and the conjectural variations. The marginal conditions for firms 1 and 2 are $p_1 - c = \frac{1}{\alpha\,(1-s_1-s_2 CV_1)}$ and $p_2 - c = \frac{1}{\alpha\,(1-s_2-s_1 CV_1)}$, respectively. The difference between these two equations implies:

$$p_1 - p_2 = \frac{1}{\alpha\,(1 - s_1 - s_2 CV_1)} - \frac{1}{\alpha\,(1 - s_2 - s_1 CV_2)}$$

The researcher observes prices $p_1 = \$200$ and $p_2 = \$195$ and market shares $s_1 = 0.5$ and $s_2 = 0.2$. Firm 1 has both a larger price and a larger market share because its product has better quality, i.e., $x_1 > x_2$. Though not really relevant to answer this question, note that in this industry the higher quality product does not imply a larger marginal cost but only a larger fixed cost. The researcher has estimated the demand system and knows that $\alpha = 0.01$. Solving the data into the previous equation, we have:

$$\$200 - \$195 = \frac{100}{1 - 0.5 - 0.2\ CV_1} - \frac{100}{1 - 0.2 - 0.5\ CV_2}$$

This is a condition that the parameters CV_1 and CV_2 should satisfy. Using this equation we can show that the hypothesis of Nash-Bertrand competition (that requires $CV_1 = CV_2 = 0$) implies a prediction about the price difference $p_1 - p_2$ that is substantially larger than the price difference that we observe in the data. The hypothesis of Nash-Bertrand competition, $CV_1 = CV_2 = 0$, implies that the right hand side of the equation in Q15(b) is:

$$\frac{100}{0.5} - \frac{100}{0.8} = 200 - 125 = \$75$$

That is, Nash-Bertrand implies a price difference of \$75 but the price difference in the data is only \$5. The hypothesis of Collusion, $CV_1 = CV_2 = 1$, implies that the right hand side of the equation in Q15(b) is:

$$\frac{100}{0.5 - 0.2} - \frac{100}{0.8 - 0.5} = \$0$$

That is, Collusion implies a price difference of \$0, which is close to the price difference of \$5 that we observe in the data.

4. Exercises

4.1. Exercise 1. Consider an industry with a differentiated product. There are two firms in this industry, firms 1 and 2. Each firm produces and sells only one brand of the differentiated product: brand 1 is produced by firm 1, and brand 2 by firm 2. The demand system is a logit demand model, where consumers choose between three different alternatives: $j = 0$, represents the consumer decision of no purchasing any product; and $j = 1$ and $j = 2$ represent the consumer purchase of product 1 and 2, respectively. The utility of no purchase $(j = 0)$ is zero. The utility of purchasing product $j \in \{1, 2\}$ is $\beta\ x_j - \alpha\ p_j + \varepsilon_j$, where the variables and parameters have the interpretation that we have seen in class. Variable x_j is a measure of the quality of product j, e.g., the number of stars of the product according to consumer ratings. Therefore, we have that $\beta > 0$. The random variables ε_1 and ε_2 are independently and identically distributed over consumers with a type I extreme value distribution, i.e., Logit model of demand. Let H be the number of consumers in the market. Let s_0, s_1, and s_2 be the market shares of the three choice alternatives, such that s_j represents the proportion of consumers choosing alternative j and $s_0 + s_1 + s_2 = 1$.

Question 1.1. Based on this model, write the equation for the market share s_1 as a function of the prices and the qualities x's of all the products.

Question 1.2. Obtain the expression for the derivatives: (a) $\dfrac{\partial s_1}{\partial p_1}$; (b) $\dfrac{\partial s_1}{\partial p_2}$; (c) $\dfrac{\partial s_1}{\partial x_1}$; and (d) $\dfrac{\partial s_1}{\partial x_2}$. Write the expression for these derivatives in terms only of the market shares s_1 and s_2 and the parameters of the model.

The profit function of firm $j \in \{0, 1\}$ is $\pi_j = p_j\ q_j - c_j\ q_j - FC(x_j)$, where: q_j is the quantity sold by firm j (i.e., $q_j = H\ s_j$); c_j is firm $j's$ marginal cost, that is assumed constant, i.e., linear cost function; and $FC(x_j)$ is a fixed cost that depends on the level of quality of the firm.

Question 1.3. Suppose that firms take their qualities x_1 and x_2 as given and compete in prices ala Bertrand.
(a) Obtain the equation that describes the marginal condition of profit maximization of firm 1 in this Bertrand game. Write this equation taking into account the specific form of $\dfrac{\partial s_1}{\partial p_1}$ in the Logit model.
(b) Given this equation, write the expression for the equilibrium price-cost margin $p_1 - c_1$ as a function of s_1 and the demand parameter α.

Now, suppose that the researcher is not willing to impose the assumption of Bertrand competition and considers a conjectural variations model. Define the conjecture parameter CV_1

as the belief or conjecture that firm 1 has about how firm 2 will change its price when firm 1 changes marginally its price. That is, CV_1 represents the belief or conjecture of firm 1 about $\frac{\partial p_2}{\partial p_1}$. Similarly, CV_2 represents the belief or conjecture of firm 2 about $\frac{\partial p_2}{\partial p_1}$.

Question 1.4. Suppose that firm 1 has a conjectural variation CV_1.

(a) Obtain the equation that describes the marginal condition of profit maximization of firm 1 under this conjectural variation. Write this equation taking into account the specific form of $\frac{\partial s_1}{\partial p_1}$ in the Logit model. [Hint: Now, we have that: $\frac{dq_1}{dp_1} = \frac{\partial q_1}{\partial p_1} + \frac{\partial q_1}{\partial p_2}\frac{\partial p_2}{\partial p_1}$, where $\frac{\partial q_1}{\partial p_1}$ and $\frac{\partial q_1}{\partial p_2}$ are the expressions you have derived in Q1.2].

(b) Given this equation, write the expression for the equilibrium price-cost margin $p_1 - c_1$ as a function of the market shares s_1 and s_2, and the parameters α and CV_1.

Question 1.5. Suppose that the researcher does not know the magnitude of the marginal costs c_1 and c_2, but she knows that the two firms use the same production technology, they use the same type of variable inputs, and they purchase these inputs in the same markets where they are price takers. Under these conditions, the researcher knows that $c_1 = c_2 = c$, though she does not know the magnitude of the marginal cost c.

(a) The marginal conditions for profit maximization in Q1.4(b), for the two firms, together with the condition $c_1 = c_2 = c$, imply that price difference between these two firms, $p_1 - p_2$, is a particular function of their markets shares and their conjectural variations. Derive the equation that represents this condition.

(b) The researcher observes prices $p_1 = \$200$ and $p_2 = \$195$ and market shares $s_1 = 0.5$ and $s_2 = 0.2$. Firm 1 has both a larger price and a larger market share because its product has better quality, i.e., $x_1 > x_2$. The researcher has estimated the demand system and knows that $\alpha = 0.01$. Plug in these data into the equation in Q1.5(a) to obtain a condition that the parameters CV_1 and CV_2 should satisfy in this market.

(c) Using the equation in Q1.5(b), show that the hypothesis of Nash-Bertrand competition (that requires $CV_1 = CV_2 = 0$) implies a prediction about the price difference $p_1 - p_2$ that is substantially larger than the price difference that we observe in the data.

(d) Using the equation in Q1.5(b), show that the hypothesis of Collusion (that requires $CV_1 = CV_2 = 1$) implies a prediction about the price difference $p_1 - p_2$ that is much closer to the price difference that we observe in the data.

4.2. Exercise 2. To answer the questions in this part of the problem set 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, i.e., `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, i.e., 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 2.1.

(a) Obtain the OLS-Fixed effects estimator of the Standard logit model. Interpret the results.

(b) Test the null hypothesis that all countries have the same price coefficient.

(c) 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.

Question 2.2.
Consider the equilibrium condition (first order conditions of profit maximization) under the assumption that each product is produced by only one firm.

(a) Write the equation for this equilibrium condition. Write this equilibrium condition as an equation for the Lerner Index, $\dfrac{p_j - MC_j}{p_j}$.

(b) Using the previous equation in Q2.2(a) and the estimated demand in Q2.1, calculate the Lerner index for every car-market-year observation in the data.

(c) Report the mean values of the Lerner Index for each of the counties/markets. Comment the results.

(d) Report the mean values of the Lerner Index for each of the top five car manufacturers (i.e., the five car manufacturers with largest total aggregate sales over these markets and sample period). Comment the results.

Question 2.3.

(a) Using the equilibrium condition and the estimated demand, obtain an estimate of the marginal cost for every car-market-year observation in the data.

(b) Run an OLS-Fixed effects regression where the dependent variable is the estimated value of the marginal cost, and the explanatory variables (regressors) are the product characteristics `hp`, `li`, `wi`, `cy`, `le`, and `he`. Interpret the results.

Bibliography

[1] Anderson, S., A. De Palma and J-F Thisse (1992), Discrete Choice Theory of Product Differentiation, Cambridge, MA: MIT Press.

[2] Armantier, O. and O. Richard (2003): "Exchanges of Cost Information in the Airline Industry," The RAND Journal of Economics, Vol. 34, No. 3, pp. 461-477.

[3] Armantier, O., J.P. Florens, J.F. Richard (2008): "Approximation of Nash equilibria in Bayesian games," Journal of Applied Econometrics, 23(7), 965-981.

[4] Bain, J. (1951): "Relation of Profit Rate to Industry Concentration: American Manufacturing, 1936–1940," The Quarterly Journal of Economics, 65 (3), 293-324.

[5] Bain, J. (1954): "Economies of Scale, Concentration, and the Condition of Entry in Twenty Manufacturing Industries," The American Economic Review , 44 (1), 15-39.

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

[7] Berry, S. and J. Waldfogel (2010): "Quality and Market Size," Journal of Industrial Economics, 58(1), 1-31.

[8] Bresnahan, T. (1980): "Three Essays on the American Automobile Oligopoly," Ph.D. Dissertation, Princeton University.

[9] Bresnahan, T. (1981): "Departures from Marginal-Cost Pricing in the American Automobile Industry: Estimates for 1977-1978," Journal of Econometrics, 17, 201-227.

[10] Bresnahan, T. (1982): "The Oligopoly Solution Concept is Identified," Economics Letters, 10, 87-92.

[11] Bresnahan, T. (1987): "Competition and Collusion in the American Automobile Market: The 1955 Price War," Journal of Industrial Economics, 35, 457-482.

[12] Bresnahan, T. (1989): "Empirical Methods for Industries with Market Power," chapter 17 in Handbook of Industrial Organization, Volume II, Richard Schmalensee and Robert Willig, eds., Amsterdam: Elsevier Science Publishers.

[13] Bresnahan, T. and J. Baker (1992): "Empirical Methods of Identifying and Measuring Market Power," Antitrust Law Journal, Summer 1992.

[14] Ciliberto, F. and J. Williams (2014): "Does multimarket contact facilitate tacit collusion? Inference on conduct parameters in the airline industry," The RAND Journal of Economics, Vol. 45, No. 4, 764-791.

[15] Corts, K. (1999): "Conduct Parameters and the Measurement of Market Power," Journal of Econometrics 88 (2), 227-250.

[16] Ellison, G. (1994): "Theories of Cartel Stability and the Joint Executive Committee," Rand Journal of Economics, 25, 37-57.

[17] Gardete, P. (2016). Competing under asymmetric information: The case of dynamic random access memory manufacturing. Management Science, 62, 3291-3309.

[18] Genesove, D. and W. P. Mullin (1998): Testing static oligopoly models: Conduct and cost in the sugar industry, 1890-1914. The Rand Journal of Economics 29(2), 355–377.

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

[20] Green, E., and R. Porter (1984): "Noncooperative Collusion Under Imperfect Price Information," Econometrica, 52, 87-100.

[21] Graddy, K. 1995. "Testing for Imperfect Competition at the Fulton Fish Market," Rand Journal of Economics 26(Spring): 75-92.

[22] Michel, C. and S. Weiergraeber (2018): "Estimating Industry Conduct in Differentiated Products Markets: The Evolution of Pricing Behavior in the RTE Cereal Industry," manuscript.

[23] Miller, N. and M. Osborne (2013): "Spatial Differentiation and Price Discrimination in the Cement Industry: Evidence from a Structural Model," manuscript. Rotman School of Management. University of Toronto.

[24] Nash, J. (1951): "Non-cooperative games," *Annals of Mathematics*, 286-295.

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

[26] Nevo: Econ Letters

[27] Porter, R. (1983): "Optimal Cartel Trigger Price Strategies," Journal of Economic Theory, 29, 313-338.

[28] Porter, R. (1983): "A Study of Cartel Stability: The Joint Executive Committee, 1880-1886," Bell Journal of Economics, 15, 301-314.

[29] Porter, R. (1985): "On the Incidence and Duration of Price Wars," Journal of Industrial Economics, 33, 415-426.

[30] Porter, R. (1986): "A Note on Tacit Collusion Under Demand Uncertainty," Canadian Journal of Economics, 19, 587-589.

[31] Puller, S. (2009): "Estimation of competitive conduct when firms are efficiently colluding: addressing the Corts critique," Applied Economics Letters, 16, 1497-1500.

[32] Smith, H. (2004), 'Supermarket choice and supermarket competition in market equilibrium', Review of Economic Studies, 71 (1), 235–63.

[33] Vives, X. (2002): "Private Information, Strategic Behavior, and Efficiency in Cournot Markets," The RAND Journal of Economics, 33(3), 361-376.

CHAPTER 5

Empirical Models of Market Entry

A model of entry is a model of market competition where we endogenize the number of firms in the market and, in some cases, the characteristics of the products that firms provide. In the previous chapters, we have taken the number of firms in the market, and the product characteristics of each firm, as exogenously given. In a model of market entry, the entry cost of a firm is a parameter that plays a key role in the determination of the number of firms in the market, in the characteristics of the active firms. Entry costs cannot be identified from the estimation of demand equations or from marginal conditions of optimality for prices and quantities. In the literature of empirical entry models, we use the principle of reveal preference to identify these parameters. If a firm decides to be in the market is because the value of being active in the market is greater than the value of being out. Under some conditions, firms' entry decisions reveal information that allows us to identify firms' entry costs.

An important application of models of entry is the study of firms' decision about the geographic location of their stores. Competition in retail markets is often characterized by the importance of geographic location. The distance from a store to potential customers, wholesalers, and competitors can have substantial effects on demand and costs, and consequently on prices, quantities, profits, and consumer welfare. Brick-and-mortar retailers usually sell their products to consumers who physically visit their stores. Firms need to choose store locations carefully so that they are accessible to many potential customers, who are spatially dispersed. Opening a store in such attractive locations is typically more expensive (for example, higher land prices) and it can be associated with stronger competition. Retailers should consider this trade-off when choosing the best store location. The study of the determinants of when and where to open retail stores is necessary to inform public policy and business debates such as the value of a merger between retail chains, zoning laws, spatial pre-emption, cannibalization between stores of the same chain, or the magnitude of economies of density. Therefore, it is not surprising that models of market entry, store location, and spatial competition have played a fundamental role in the theory of industrial organization at least since the work of Harold Hotelling (1929). However, empirical work on structural estimation of these models has been much more recent and it has followed the seminal work by Bresnahan and Reiss (1990, 1991a).

In a model of entry, the key endogenous variables are firms' decisions to operate or not stores. In some entry models, the set of endogenous variables may also include store geographic locations, prices and quantities, store capacity, product quality, or some product characteristic(s) that provide horizontal product differentiation. Every firm makes these decisions to maximize expected profit. Empirical games of market entry in retail markets share as common features that the payoff of being active in the market depends on market size, entry cost, and the number and characteristics of other active firms. The set of structural parameters of the model varies considerably across models and applications, but it typically includes parameters that represent the entry cost and the strategic interactions between firms (competition effects). Given these common characteristics, there are substantial differences between the models that have been proposed and estimated in empirical applications. The discussion of these different specification assumptions is one of the goals of this chapter.

In empirical applications of games of market entry, structural parameters are estimated using data on firms' entry decisions in a sample of markets. The estimated model is used to answer empirical questions on the nature of competition and the structure of costs in an industry, and to make predictions about the effects of changes in structural parameters or of counterfactual public policies affecting firms' profits, for example, subsidies, taxes, or zoning laws. Entry costs, and their structure in terms of market and firm characteristics, are important primitives of these models because they play a key role in the determination of the number of stores active in a market, their characteristics, and their spatial configuration. These costs cannot be identified from the estimation of demand equations, production functions, or marginal conditions of optimality for prices or quantities. Instead, in a structural entry model, entry costs are identified using the principle of revealed preference: if we observe a firm operating in a market it is because its value in that market is greater than the value of shutting down and putting its assets in alternative uses. Under this principle, firms' entry and exit decisions reveal information about the underlying or latent profit function. Empirical games of market entry can be also useful to identify strategic interactions between firms that occur through variable profits. In empirical applications where a sample variation in prices is very small but there is a substantial variation in entry decisions, an entry model can provide more information about demand substitution between stores and products than the standard approach of using prices and quantities to estimate demand. Furthermore, data on prices and quantities at the store level are typically difficult to obtain, while data on store locations and entry/exit decisions are more commonly available.

1. Some general ideas

1.1. What is a model of market entry? Models of market entry in IO can be characterized in terms of three main features. First, the key endogenous variable is a firm decision to operate or not in a market. Entry in a market should be understood in a broad sense. The standard example is the decision of a firm to enter in an industry by first time, but opening a new store, introducing a new product, adopting a new technology, the release of a new movie, a potential bidder's decision to bid in an auction, etc, are other examples. A second important feature is that there is an **entry cost** associated with being active in the market. And third, the payoff of being active in the market depends on the number (and the characteristics) of other firms active in the market, i.e., the model is a **game**.

Consider a market where there are N firms that potentially may decide to enter in the market. We index firms with $i \in \{1, 2, ..., N\}$. Let $a_i \in \{0, 1\}$ be a binary variable that represents the decision of firm i of being active (enter) in a market: $a_i = 1$ if the firm enters, and $a_i = 0$ if it does not. The profit of not being in the market is zero. The profit of being in the market for firm i is equal to $V_i(n) - F_i$ where $V_i(.)$ is the variable profit, n is the number of firms active in the market, and F_i is the entry cost for firm i. The number of active firms, n, is endogenous and equal to $n = \sum_{i=1}^{N} a_i$. Under Nash assumption, every firm takes as given the decision of the other firms and makes a decision that maximizes its own profit. Therefore, the best response of firm i under Nah equilibrium is:

$$
a_i = \begin{cases} 1 & \text{if} \quad V_i\left(1 + \sum_{j \neq 1} a_j\right) - F_i \geq 0 \\ \\ 0 & \text{if} \quad V_i\left(1 + \sum_{j \neq 1} a_j\right) - F_i < 0 \end{cases}
\tag{1.1}
$$

where $1 + \sum_{j \neq 1} a_j$ represents firm i's Nah-conjecture about the number of active firms.

Example. There are two potential entrants: $N = 2$, with $V_1(n) = V_2(n) = 100 - 20\,n$ and $F_1 = F_2 = F$, such that $V_i(1 + a_j) - F_i = 80 - F - 20\,a_j$. Best responses are:

$$
a_1 = \begin{cases} 1 & \text{if} \quad 80 - F - 20\,a_2 \geq 0 \\ 0 & \text{if} \quad 80 - F - 20\,a_2 < 0 \end{cases} \quad \text{and} \quad a_2 = \begin{cases} 1 & \text{if} \quad 80 - F - 20\,a_1 \geq 0 \\ 0 & \text{if} \quad 80 - F - 20\,a_1 < 0 \end{cases}
$$

The payoff matrix is:

	$a_2 = 0$	$a_2 = 1$
$a_1 = 0$	$(0\,,\,0)$	$(0\,,\,80 - F)$
$a_1 = 1$	$(80 - F\,,\,0)$	$(60 - F\,,\,60 - F)$

We can see that the model has different predictions about market structure depending on the value of the fixed cost. If $F \leq 60$, duopoly, $(a_1, a_2) = (1, 1)$, is the unique Nash equilibrium. If $60 < F \leq 80$, then Monopoly of firm 1 $(a_1, a_2) = (1, 0)$ or Monopoly of firm 2 $(a_1, a_2) = (0, 1)$ are Nash equilibria. If $F > 80$, then no firm in the market $(a_1, a_2) = (0, 0)$ is

the unique Nash equilibrium. The observe actions of the potential entrants reveal information about profits, about fixed costs.

[**The principle of Revealed Preference**] The estimation of structural models of market entry is based on the principle of Revealed Preference. In the context of these models, this principle establishes that if we observe a firm operating in a market it is because its value in that market is greater than the value of shutting down and putting its assets in alternative uses. Under this principle, firms' entry decisions reveal information about the underlying latent firm's profit (or value).

[**Static models**] The first class of models that we study are static. There are many differences between static and dynamic models of market entry. But there is a simple difference that I think it is relevant to point out now. For static models of entry, we should understand "entry" as "being active in the market" and not as a transition from being "out" of the market to being "in" the market. That is, in these static models we ignore the fact that, when choosing whether to be active or not in the market, some firms are already active (incumbents) and other firms not (potential entrants). That is, we ignore that the choice of non-being active in the market means "exit" for some firms and "stay out" for others.

1.2. Why do we estimate models of market entry? As explained in the Introduction to this chapter, the specification and estimation of models of market entry is motivated by the need to endogenize the number of firms in the market, as well as some characteristics that operate at the extensive margin. Endogenizing the number of firms in the market is a key aspect in any model of IO where market structure is treated as endogenous. Once we endogenize the number of firms in the market, we need to identify entry cost parameters, and these parameters cannot be identify from demand equations, production functions, and marginal conditions of optimality for prices and quantities. We identify entry costs from the own entry model. More generally, we can distinguish the following motives for the estimation of models of market entry.

1.   **Identification of entry cost parameters.** Parameters such us fixed production costs, entry costs, or investment costs do not appear in demand or production equations but contribute to the market entry decision. These parameters can be important in the determination of market structure and market power in an industry.

2.   **Efficiency.** The equilibrium entry conditions contain useful information for the identification of structural parameters. Using this information can increase significantly the precision of our estimates. In fact, when the sample variability in prices and quantities is small, the equilibrium entry conditions can have a more important contribution to the identification of demand and cost parameters than demand equations of production functions.

3.   Data on prices and quantities may not be available, or at least, these data are not available at the level we need, e.g., at the level of individual firm, product, and market. Many countries have excellent surveys of manufacturers of retailers with information at the level of specific industry (5 or 6 digits NAICS, SIC) and local markets (census tracts) on the number of establishments and some measure of firm size such as aggregate revenue. Though we observe aggregate revenue at the industry-market level, we do not observe P and Q at that level. Under some assumptions, it is possible to identify structural parameters using these data and the structure of an entry model.

4.   Controlling for endogeneity of firms' entry decisions in the estimation of demand and production functions. The estimation of a demand system or a production function may involve dealing with, in someway or the other, the estimation of a model of market entry. Very often, in the estimation of a demand system or in the estimation of a production function we have to deal with the endogeneity of firms' entry and exit decisions. Olley and Pakes (1996) show that ignoring the endogeneity of a firm's decision to exit from the market (i.e., firm's with smaller values of unobserved productivity are more likely to exit) can generate significant biases in the estimation of production functions. Similarly, some product characteristics (not only price) may be endogenous in the estimation of demand systems. The choice of a product characteristic can be interpreted as an entry decision. For instance, the decision of a coffee shop of having or not wireless internet access. The choice of including or not the product attribute may not be exogenous in the demand system in the sense that it is correlated with unobserved demand factors. That is, we observe more demand in coffee shops with internet wireless not only because consumers like this service but also because the coffee shops that choose to include this service are the ones with higher exogenous demand such that it is profitable for them to pay the fixed cost of including wireless internet. Dealing with this endogenous product attribute requires one to specify and estimate a model of market entry.

1.3. Road map. The type of data used and the assumptions about unobserved firm and market heterogeneity, and about the information of the firms and of the researcher are very important for the estimation of entry models.

[Bresnahan and Reiss] We start with a simple and pioneer model in this literature: the models in Bresnahan and Reiss (JPE, 1991). This paper together with Bresnahan and Reiss (REStud, 1990) were significant contributions to the structural estimation of models of market entry that opened a new literature that has grown significantly during the last 20 years. In that paper, Bresnahan and Reiss show that given a cross-section of "isolated" local markets where we observe the number of firms active, and some exogenous market characteristics, including market size, it is possible to identify fixed costs and the "degree

of competition" or the "nature of competition" in the industry. By "nature of competition" these authors (and after them, this literature) means a measure of how a firm's variable profit declines when the number of competitors in the market increases.

What is most remarkable about Bresnahan and Reiss's result is how with quite limited information (e.g., no information about prices of quantities) the researcher can identify the degree of competition using an entry model.

[**Relaxing the assumption of homogeneous firms**] Bresnahan and Reiss's model is based on some important assumptions. In particular, firms are homogeneous and they have complete information. The assumption of firm homogeneity (both in demand and costs) is strong and can be clearly rejected in many industries. Perhaps more importantly, ignoring firm heterogeneity when present can lead to biased and misleading results about the degree of competition in a industry. Therefore, the first assumption that we relax is the one of homogeneous firms.

As shown originally in the own work of Bresnahan and Reiss (Journal of Econometrics, 1991), relaxing the assumption of firm homogeneity implies two significant econometric challenges. The entry model becomes a system of simultaneous equations with endogenous binary choice variables. Dealing with endogeneity in binary choice system of equations is not a simple econometric problem. In general, IV estimators are not available. Furthermore, the model now has multiple equilibria. Dealing with both endogeneity and multiple equilibria in this class of nonlinear models is an interesting but challenging problem in econometrics.

[**Approaches to deal with endogeneity/multiple equilibria in games of complete information**]. Then, we will go through different approaches that have been used in this literature to deal with the problems of endogeneity and multiple equilibria. I think that it is worthwhile to distinguish two groups of approaches or methods.

The **first group of methods** is characterized by imposing restrictions that imply equilibrium uniqueness for any value of the exogenous variables. Of course, firm homogeneity is a type of assumption that implies equilibrium uniqueness. But there are other assumptions that imply uniqueness even when firms are heterogeneous. For instance, a triangular structure in the strategic interactions between firms (Heckman, Econometrica 1978), or sequential entry decisions (Berry, Econometrica 1993). Given these assumptions, these papers deal with the endogeneity problem by using a maximum likelihood approach.

The **second group of methods** do not impose equilibrium uniqueness. The early work of Jovanovic (Econometrica 1989) and the most recent work by Tamer (2003) were important or influential for this other approach. These authors showed (Jovanovic at a general abstract level, and Tamer in the context of a binary choice game) that identification and multiple equilibria are two very different issues in econometric models. Models with multiple equilibria

can be identified, and we do not need to impose equilibrium uniqueness as a form to get identification. Multiple equilibria can be a computational nuisance in the estimation of these models, but it is not an identification problem. This simple idea has generated a significant and growing literature that deals with computational simple methods to estimate models with multiple equilibria, and more specifically with the estimation of discrete games.

[**Games of incomplete information**] Our next step will be to relax the assumption of complete information by introducing some variables that are private information of each firm. We will see that the identification and estimation of these models can be significantly simpler than in the case of models of complete information.

2. Data

The datasets that have been used in empirical applications of structural models of entry in retail markets consist of a sample of geographic markets with information on firms' entry decisions and consumer socio-economic characteristics over one or several periods of time. In these applications, the number of firms and time periods is typically small such that statistical inference (that is, the construction of sample moments and the application of law of large numbers and central limit theorems) is based on a 'large' number of markets. In most applications, the number of geographic markets is between a few hundred and a few thousand. Within these common features, there is substantial heterogeneity in the type of data that have been used in empirical applications.

In this section, we concentrate on four features of the data that are particularly important because they have substantial implications on the type of model that can be estimated, the empirical questions that we can answer, and the econometric methods to use. These features are: (1) the selection of geographic markets; (2) presence or not of within-market spatial differentiation; (3) information on prices, quantities, or sales at the store level; and (4) information on potential entrants.

2.1. Selection of Geographic Markets. In a seminal paper, Bresnahan and Reiss (1990) use cross-sectional data from 149 small US towns to estimate a model of entry of automobile dealerships. For each town, the dataset contains information on the number of stores in the market, demographic characteristics such as population and income, and input prices such as land prices. The selection of the 149 small towns is based on the following criteria: the town belongs to a county with fewer than 10 000 people; there is no other town with a population of over 1000 people within 25 miles of the central town; and there is no large city within 125 miles. These conditions for the selection of a sample of markets are typically described as the 'isolated small towns' market selection. This approach has been very influential and has been followed in many empirical applications of entry in retail

markets. The main motivation for using this sample selection is in the assumptions of spatial competition in the Bresnahan–Reiss model described in section 2. That model assumes that the location of a store within a market does not have any implication on its profits or in the degree of competition with other stores. This assumption is plausible only in small towns where the possibilities for spatial differentiation are very limited. If this model were estimated using a sample of large cities, we would spuriously find very small competition effects simply because there is negligible or no competition at all between stores located far away of each other within the city. The model also assumes that there is no competition between stores located in different markets. This assumption is plausible only if the market under study is not geographically close to other markets; otherwise the model would ignore relevant competition from stores outside the market.

Although the 'isolated small towns' approach has generated a good number of important applications, it has some limitations. The extrapolation to urban markets of the estimation results obtained in these samples of rural markets is in general not plausible. Focusing on rural areas makes the approach impractical for many interesting retail industries that are predominantly urban. Furthermore, when looking at national retail chains, these rural markets account for a very small fraction of these firms' total profits.

2.2. Within Market Spatial Differentiation. The limitations of the 'isolated small towns' approach have motivated the development of empirical models of entry in retail markets that take into account the spatial locations and differentiation of stores within a city market. The work by Seim (2006) was seminal in this evolution of the literature. In Seim's model, a city is partitioned into many small locations or blocks, for example, census tracts, or a uniform grid of square blocks. A city can be partitioned into dozens, hundreds, or even thousands of these contiguous blocks or locations. In contrast to the 'isolated small towns' approach, these locations are not isolated, and the model allows for competition effects between stores at different locations. The datasets in these applications contain information on the number of stores, consumer demographics, and input prices at the block level. This typically means that the information on store locations should be geocoded, that is, the exact latitude and longitude of each store location. Information on consumer demographics is usually available at a more aggregate geographic level.

The researcher's choice for the size of a block depends on multiple considerations, including the retail industry under study, data availability, specification of the unobservables, and computational cost. In principle, the finer is the grid the more flexible can be the model to measure spatial substitution between stores. The computational cost of estimating the model can increase rapidly with the number of locations. The assumption on the distribution of the unobservables across locations is also important, too. A common approach is to use

a definition of a block/location at which demographic information is available, for example, the set of locations is equal to the set of census tracts within the city.29 While convenient, a drawback of this approach is that some blocks, especially those in the periphery of a city, tend to be very large. These large blocks are often problematic because (1) within-block spatial differentiation seems plausible, and (2) the distance to other blocks becomes highly sensitive to choices of block centroids. In particular, a mere use of geometric centroids in these large blocks can be quite misleading as the spatial distribution of population is often quite skewed. To avoid this problem, Seim (2006) uses population weighted centroids rather than (unweighted) geometric centroids. An alternative approach to avoid this problem is to draw a square grid on the entire city and use each square as a possible location, as in Datta and Sudhir (2013) and Nishida (2015). The value of consumer demographics in a square block is equal to the weighted average of the demographics at the census tracts that overlap with the square. The advantage of this approach is that each submarket has a uniform shape. In practice, implementation of this approach requires the removal of certain squares where entry cost is prohibitive. These areas include those with some particular natural features (for example, lakes, mountains, and wetlands) or where commercial space is prohibited by zoning. For example, Nishida (2015) excludes areas with zero population, and Datta and Sudhir (2013) remove areas that do not have any 'big box' stores as these areas are very likely to be zoned for either residential use or small stores.

So far, all the papers that have estimated this type of model have considered a sample of cities (but not locations within a city) that is still in the spirit of Bresnahan–Reiss isolated small markets approach. For instance, Seim selects US cities with population between 40 000 and 150 000 people, and without other cities with more than 25 000 people within 20 miles. The main reason for this is to avoid the possibility of outside competition at the boundaries of a city. It is interesting that in the current generation of these applications, statistical inference is based on the number of cities and not on the number of locations. A relevant question is whether this model can be estimated consistently using data from a single city with many locations, that is, the estimator is consistent when the number of locations goes to infinity. This type of application can be motivated by the fact that city characteristics that are relevant for these models, such as the appropriate measure of geographic distance, transportation costs, or land use regulations and zoning, can be city specific. Xu (2014) studies an empirical game of market entry for a single city (network) and presents conditions for consistency and asymptotic normality of estimators as the number of locations increases. As far as we know, there are not yet empirical applications following that approach.

2.3. Information on Prices, Quantities, or Sales at the Store Level. Most applications of models of entry in retail markets do not use data on prices and quantities due

to the lack of such data. The most popular alternative is to estimate the structural (or semi-structural) parameters of the model using market entry data only, for example, Bresnahan and Reiss (1990), Mazzeo (2002), Seim (2006), or Jia (2008), among many others. Typically, these studies either do not try to separately identify variable profits from fixed costs, or they do it by assuming that the variable profit is proportional to an observable measure of market size. Data on prices and quantities at store level can substantially help the identification of these models. In particular, it is possible to consider a richer specification of the model that distinguishes between demand, variable cost, and fixed cost parameters, and includes unobservable variables into each of these components of the model.

A sequential estimation approach is quite convenient for the estimation of this type of model. In a first step, data on prices and quantities at the store level can be used to estimate a spatial demand system as in Davis (2006) for movie theatres or Houde (2012) for gas stations. Note that, in contrast to standard applications of demand estimation of differentiated products, the estimation of demand models of this class should deal with the endogeneity of store locations. In other words, in these demand models, not only prices are endogenous but also the set of 'products' or stores available at each location is potentially correlated with unobserved errors in the demand system. In a second step, variable costs can be estimated using firms' best response functions in Bertrand or Cournot model.30 Finally, in a third step, we estimate fixed cost parameters using the entry game and information of firms' entry and store location decisions. It is important to emphasize that the estimation of a demand system of spatial differentiation in the first step provides the structure of spatial competition effects between stores at different locations, such that the researcher does not need to consider other type of semi-reduced form specifications of strategic interactions, as in Seim (2006) among others.

In some applications, price and quantity are not available, but there is information on revenue at the store level (for example, Ellickson and Misra 2012; Aguirregabiria et al. 2013, Suzuki 2013). This information can be used to estimate a (semi reduced form) variable profit function in a first step, and then in a second step the structure of fixed costs is estimated.

2.4. Information on Potential Entrants. An important modelling decision in empirical entry games is to define the set of potential entrants. In most cases, researchers have limited information on the number of potential entrants, let alone their identity. This problem is particularly severe when entrants are mostly independent small stores (for example, mom-and-pop stores). A practical approach is to estimate the model under different numbers of potential entrants and examine how estimates are sensitive to these choices, for example, Seim (2006) and Jia (2008). The problem is less severe when most entrants belong

to national chains (for example, big box stores) because the names of these chains are often obvious and the number is typically small.

****

Most empirical studies have considered cross-sectional data of independent ("isolated") markets. Therefore, these studies try to explain the determinants of cross-sectional differences in market structure. It is important to distinguish three types of data sets. The specification and the identification of the model is different for each of these three types of data.

(1) Only global potential entrants. The same N firms are the potential entrants in every market. We know the identity of these "global" potential entrants. Therefore, we observe the decision of each of these firms in every independent market. We observe market characteristics, and sometimes firm characteristics which may vary or not across markets. The data set is:

$$\{s_m, \ x_{im}, \ a_{im} : m = 1, 2, ..., M; \ i = 1, 2, ..., N\} \tag{2.1}$$

where m is the market index; i is the firm index; s_m is a vector of characteristics of market m such as market size, average consumer income, or other demographic variables; x_{im} is a vector of characteristics of firm i; and a_{im} is the indicator of the event "firm i is active in market m".

Let use x_m to represent the vector of (exogenous or predetermined) market and firms' characteristics: $z_m \equiv \{s_m, z_{im} : i = 1, 2, ..., N\}$. And let a_m be $\{a_{im} : i = 1, 2, ..., N\}$. With this type of data we can nonparametrically identify $\Pr(a_m|x_m)$ and therefore we will be able to identify models with very general forms of firm heterogeneity.

Example 1: Berry (Econometrica, 1992) considers entry in airline markets. A market is a city pair (e.g., Boston-Chicago). The set of markets consists of all the pairs of US cities with airports. Every airline company operating in US is a potential entrant in each of these markets. a_{im} is the indicator of the event "airline i operates in city pair m".

Example 2: Toivanen and Waterson (2000) consider entry in local markets by fast-food restaurants in UK. Potential entrants are Burger King, McDonalds, KFC, Wendys, etc.

(2) Only local potential entrants. We do not know the identity of the potential entrants. In fact, most potential entrants may be local, i.e., they consider entry in only one local market. For this type of data we only observe market characteristics and the number of active firms in the market. The data set is:

$$\{s_m, n_m : m = 1, 2, ..., M\} \tag{2.2}$$

where n_m is the number of firms operating in market m.

With this type of data we can nonparametrically identify $\Pr(n_m|s_m)$. We can identify firm heterogeneity only to a limited extent and based on functional form identification. Notice also that we do not know the number of potential entrants N, and this may vary over markets.

Example 1: Bresnahan and Reiss (REStud, 1990). Car dealers in small towns.

Example 2: Bresnahan and Reiss (JPE, 1991). Restaurants, dentists and other retailer and professional services in small towns.

Example 3: Seim (2003). Video rental stores.

(3) Both global and local potential entrants. This case combines and encompasses the previous two cases. There are N_G firms which are potential entrants in all the markets, and we now the identity of these firms. But there are also other potential entrants which are just local. We observe:

$$\{s_m, n_m, z_{im}, a_{im} : m = 1, 2, ..., M; \ i = 1, 2, ..., N_G\} \tag{2.3}$$

With this data we can nonparametrically identify $\Pr(n_m, a_m|x_m)$. We can allow for heterogeneity between global players in a very general way. Heterogeneity between local players should be much more restrictive.

3. Models

3.1. Static game with single-store firms. We start with the description of a static entry game between single-store firms. Later, we extend this framework to incorporate dynamics and multi-store firms. There are N retail firms that are potential entrants in a market. We index firms by $i \in \{1, 2, \ldots, N\}$. From a geographic point of view, the market is a compact set $\mathbb{C}$ in the Euclidean space $\mathbb{R}^2$, and it contains L locations where firms can operate stores. These locations are exogenously given and they are indexed by $l \in \{1, 2, \ldots, L\}$. Firms play a two-stage game. In the first stage, firms make their entry and store location decisions. Each firm decides whether to be active or not in the market, and if active, the location of its store. We can represent a firm's decision using an L-dimensional vector of binary variables, $a_i \equiv \{a_{il} : l = 1, 2, \ldots, L\}$, where $a_{il} \in \{0, 1\}$ is the indicator of the event 'firm i has a store in location l'. For single-store firms, there is at most one component in the vector a_i that is equal to one while the rest of the binary variables must be zero. In the second stage they compete in prices (or quantities) taking entry decisions as given. The equilibrium in the second stage determines equilibrium prices and quantities at each active store.

The market is populated by consumers. Each consumer is characterized by his preference for the products that firms sell and by his geographical location or home address h that belongs to the set of consumer home addresses $\{1, 2, \ldots, H\}$. The set of consumer home addresses and the set of feasible business locations may be different.5 Following Smith (2004), Davis (2006), or Houde (2012), aggregate consumer demand comes from a discrete choice model of differentiated products where both product characteristics and transportation costs affect demand. For instance, in a spatial logit model, the demand for firm i with a store in location ℓ is:

$$q_{il} = \sum_{h=1}^{H} M(h) \frac{a_{il} \, \exp\{x_i \, \beta - \alpha \, p_{il} - \tau(d_{hl})\}}{1 + \sum_{j=1}^{N} \sum_{l'=1}^{L} a_{jl'} \, \exp\{x_j \, \beta - \alpha \, p_{jl'} - \tau \, d_{hl'}\}}$$

where q_{il} and p_{il} are the quantity sold and the price, respectively, at store (i, l); $M(h)$ represents the mass of consumers living in address h; the term within the square brackets is the market share of store (i,l) among consumers living in address h; x_i is a vector of observable characteristics (other than price) of the product of firm i; and β is the vector of marginal utilities of these characteristics; α is the marginal utility of income; d_{hl} represents the geographic distance between home address h and business location l; and $\tau(d_{hl})$ is an increasing real-valued function that represents consumer transportation costs.

Given this demand system, active stores compete in prices à la Nash–Bertrand to maximize their respective variable profits, $(p_{il} - c_{il}) \, q_{il}$, where c_{il} is the marginal cost of store (i, l), that is exogenously given. The solution of the system of best response functions can be described as a vector of equilibrium prices for each active firm/store. Let $p_i^*(l, a_{-i}, x)$ and $q_i^*(l, a_{-i}, x)$ represent the equilibrium price and quantity for firm i given that this firm has a store at location l , that the rest of the firms' entry/location decisions are represented by the vector $a_{-i} \equiv \{a_j : j \neq i\}$, and that these firms' characteristics are denoted by $x \equiv (x_1, x_2, \ldots, x_N)$. Similarly, we can define the equilibrium (indirect) variable profit,

$$VP_i^*(\ell, a_{-i}, x) = [p(\ell, a_{-i}, x) - c_{il}] \, q_i^*(\ell, a_{-i}, x)$$

Consider now the entry stage of the game. The profit of firm i if it has a store in location ℓ is:

$$\blacksquare_i(\ell, a_{-i}, x) = VP_i^*(\ell, a_{-i}, x) - EC_{i\ell}$$

where $EC_{i\ell}$ represents the entry cost of firm i at location l, that for the moment is exogenously given. The profit of a firm that is not active in the market is normalized to zero, that is, $\blacksquare_i(0, a_{-i}, x) = 0$, where with some abuse of notation, we use $l = 0$ to represent the choice alternative of no entry in any of the L locations.

The description of an equilibrium in this model depends on whether firms have complete or incomplete information about other firms' costs. The empirical literature on entry games

has considered both cases. In the complete information model, a Nash equilibrium is an N-tuple $\{a_i^* : i = 1, 2, \ldots, N\}$ such that for every firm i the following best response condition is satisfied:

$$a_{i\ell}^* = 1\{\pi_i(\ell, a_{-i}^*, x) \geq \pi_i(\ell', a_{-i}^*, x) \text{ for any } \ell' \neq \ell\}$$

where $1\{.\}$ is the indicator function. In equilibrium, each firm is maximizing its own profit given the entry and location decisions of the other firms.

In a game of incomplete information, there is a component of a firm's profit that is private information of the firm. For instance, suppose that the entry cost of firm i is $EC_{i\ell} = ec_{i\ell} + \varepsilon_{i\ell}$, where $ec_{i\ell}$ is public information for all the firms, and $\varepsilon_{i\ell}$ is private information of firm i. These private cost shocks can be correlated across locations for a given firm, but they are independently distributed across firms, that is, $\varepsilon_i \equiv \{\varepsilon_{i\ell} : \ell = 1, 2, \ldots, L\}$ is independently distributed across firms with a distribution function F_i that is continuously differentiable over $\mathbb{R}^L$ and common knowledge to all the firms. A firm's strategy is an L-dimensional mapping $\alpha_i(\varepsilon_i; x) \equiv \{\alpha_{i\ell}(\varepsilon_i; x) : \ell = 1, 2, \ldots, L\}$ where $\alpha_{i\ell}(\varepsilon_i; x)$ is a binary-valued function from the set of possible private information values, R^L and the support of x into $\{0, 1\}$ such that $\alpha_{i\ell}(\varepsilon_i; x) = 1$ means that firm i enters location ℓ when the value of private information is ε_i. A firm has uncertainty about the actual entry decisions of other firms because it does not know the realization of other firms' private information. Therefore, firms maximize expected profits. Let $\pi_i^e(\ell, \alpha_{-i}, x)$ be the expected profit of firm i if it has a store at location ℓ and the other firms follow their respective strategies in α_{-i}^*. By definition, $\pi_i^e(\ell, \alpha_{-i}, x) \equiv E_{\varepsilon_{-i}}[\pi_i(\ell, \alpha_{-i}(\varepsilon_{-i}; x), x)]$, where $E_{\varepsilon_{-i}}$ represents the expectation over the distribution of the private information of firms other than i. A Bayesian Nash equilibrium in this game of incomplete information is an N-tuple of strategy functions $\{\alpha_{-i}^* : i = 1, 2, \ldots, N\}$ such that every firm maximizes its expected profit: for any ε_i,

$$\alpha_{i\ell}^*(\varepsilon_i; x) = 1\{\pi_i^e(\ell, \alpha_{-i}^*, x) \geq \pi_i^e(\ell', \alpha_{-i}^*, x) \text{ for any } \ell' \neq \ell\}$$

In an entry game of incomplete information, firms' strategies (and therefore, a Bayesian Nash equilibrium) can be described also using firms' probabilities of market entry, instead of the strategy functions $\alpha_i(\varepsilon_i; x)$. In sections 2.2.1 and 2.2.4, we present examples of this representation in the context of more specific models.

3.2. Multi-store firms. Multi-store firms, or retail chains, have become prominent in many retail industries such as supermarkets, department stores, apparel, electronics, fast food restaurants, or coffee shops, among others. Cannibalization and economies of scope between stores of the same chain are two important factors in the entry and location decisions of a multi-store firm. The term cannibalization refers to the business stealing effects between stores of the same chain. Economies of scope may appear if some operating costs

are shared between stores of the same retail chain such that these costs are not duplicated when the number of stores in the chain increases. For instance, some advertising, inventory, personnel, or distribution costs can be shared among the stores of the same firm. These economies of scope may become quantitatively more important when store locations are geographically closer to each other. This type of economies of scope is called economies of density. The recent empirical literature on retail chains has emphasized the importance of these economies of density, that is, Holmes (2011), Jia (2008), Ellickson et al. (2013), and Nishida (2015). For instance, the transportation cost associated with the distribution of products from wholesalers to retail stores can be smaller if stores are close to each other. Also, geographic proximity can facilitate sharing inventories and even personnel across stores of the same chain. We now present an extension of the basic framework that accounts for multi-store firms.

A multi-store firm decides its number of stores and their locations. We can represent a firm's entry decision using the L-dimension vector $a_i \equiv \{a_{i\ell} : \ell = 1, 2, \ldots, L\}$, where $a_{i\ell} \in \{0, 1\}$ is still the indicator of the event 'firm i has a store in location ℓ'. In contrast to the case with single-store firms, now the vector a_i can take any value within the choice set $\{0, 1\}^L$. The demand system still can be described using equation (***). The variable profit of a firm is the sum of variable profits over every location where the firm has stores, $\sum_{\ell=1}^{L} a_{i\ell}$ $(p_{i\ell} - c_{i\ell})q_{i\ell}$. Firms compete in prices taking their store locations as given. A retail chain may choose to have a uniform price across all its stores, or to charge a different price at each store. In the Bertrand pricing game with spatial price discrimination (that is, different prices at each store), the best response of firm i can be characterized by the first-order conditions:

$$q_{i\ell} + (p_{i\ell} - c_{i\ell}) \frac{\partial q_{i\ell}}{\partial p_{i\ell}} + \sum_{\ell' \neq \ell}(p_{i\ell'} - c_{i\ell'}) \frac{\partial q_{i\ell'}}{\partial p_{i\ell}} = 0$$

The first two terms represent the standard marginal profit of a single-store firm. The last term represents the effect on the variable profits of all other stores within the firm, and it captures how the pricing decision of the firm internalizes the cannibalization effect among its own stores. A Nash-Bertrand equilibrium is a solution in prices to the system of best response equations in (***). The equilibrium (indirect) variable profit of firm i is:

$$VP_i^*(a_i, a_{-i}; x) = \sum_{\ell=1}^{L}(p_i^*(\ell, a_{-i}; x) - c_{i\ell}) \, q_i^*(\ell, a_{-i}; x)$$

where $p_{i\ell}^*(\ell, a_{-i}; x)$ and $q_i^*(\ell, a_{-i}; x)$ represent Bertrand equilibrium prices and quantities, respectively.

The total profit of the retail chain is equal to total variable profit minus total entry cost: $\pi_i(a_i, a_{-i}; x) = VP_i^*(a_i, a_{-i}; x) - EC_i(a_i)$. The entry costs of a retail chain may depend on

the number of stores (that is, (dis)economies of scale) and on the distance between the stores (for example, economies of density). In section 2.2.5, we provide examples of specifications of entry costs for multi-store retailers.

The description of an equilibrium in this game of entry between retail chains is similar to the game between single-store firms. With complete information, a Nash equilibrium is an N-tuple $\{a_i^* : i = 1, 2, \ldots, N\}$ that satisfies the following best response conditions:

$$\pi_i(a_i^*, a_{-i}^*; x) \geq \pi_i(a_i, a_{-i}^*; x) \ \text{ for any } a_i \neq a_i^*$$

With incomplete information, a Bayesian Nash equilibrium is an N-tuple of strategy functions $\{\alpha_i^*(\varepsilon_i; x) : i = 1, 2, \ldots, N\}$ such that every firm maximizes its expected profit: for any ε_i:

$$\pi_i^e(\alpha_i^*(\varepsilon_i; x), \alpha_{-i}^*, x) \geq \pi_i^e(a_i, \alpha_{-i}^*, x) \ \ \text{ for any } a_i \neq \alpha_i^*(\varepsilon_i; x)$$

3.3. Dynamic game. Opening (or closing) a store is a forward-looking decision with significant non-recoverable entry costs, mainly owing to capital investments which are both firm and location-specific. The sunk cost of setting up new stores, and the dynamic strategic behavior associated with them, are potentially important forces behind the configuration of the spatial market structure that we observe in retail markets. We now present an extension of the previous model that incorporates these dynamic considerations.

Time is discrete and indexed by $t \in \{\ldots, 0, 1, 2, \ldots\}$. At the beginning of period t a firm's network of stores is represented by the vector $a_{it} \equiv \{a_{i\ell t} : \ell = 1, 2, \ldots, L\}$, where $a_{i\ell t}$ is the number of stores that firm $a_{i\ell t}$ operates in location ℓ at period t. For simplicity, we maintain the assumption that a firm can have at most one store in a location, such that $a_{i\ell t} \in \{0, 1\}$. The market structure at period t is represented by the vector $a_t \equiv \{a_{it} : i = 1, 2, \ldots, N\}$ capturing the store network of all firms. Following the structure in the influential work on dynamic games of oligopoly competition by Ericson and Pakes (1995) and Pakes and McGuire (1994), at every period t the model has two stages, similar to the ones described in the static game above. In the second stage, taking the vector of firms' store networks a_t as given, retail chains compete in prices in exactly the same way as in the Bertrand model described in section 2.1.2. The equilibrium in this Bertrand game determines the indirect variable profit function, $VP_i^*(a_t; z_t)$, where z_t is a vector of exogenous state variables in demand and costs. Some components of z_t may be random variables, and their future values may not be known at the current period. In the first stage, every firm decides its network of stores next period, $a_{i,t+1}$, and pays at period t the entry and exit costs associated to opening and closing stores. The period profit of a firm is $\pi_i(a_{i,t+1}, a_t, z_t) = VP_i^*(a_t; z_t) - FC(a_{it}; z_t) - AC_i(a_{i,t+1}, a_{it})$, where FC_i is the fixed cost of operating the network, and AC_i is the cost of adjusting the network from a_{it} to $a_{i,t+1}$, that is, costs of opening and closing stores. A firm chooses its new network $a_{i,t+1}$ to maximize the sum of its discounted expected future profits.

A Markov perfect equilibrium of this dynamic game is an N-tuple of strategy functions $\{\alpha_i^*(a_t, z_t) : i = 1, 2, \ldots, N\}$ such that every firm maximizes its expected intertemporal profit:

$$\alpha_i^*(a_t, z_t) = \arg\max_{a_{i,t+1}} \left[\pi_i(a_{i,t+1}, a_t, z_t) + \blacksquare \, E_t(V_i^{\alpha^*}(a_{i,t+1}, \alpha_{-i}^*(a_t, z_t), z_{t+1})) \right]$$

where $\delta \in (0, 1)$ is the discount factor, and $V_i^{\alpha^*}(a_{it}, a_{-it}, z_t)$ is the value of firm i when firms' networks are equal to a_t, the value of exogenous state variables is z_t, and the other firms follow strategies α_{-i}^*.

3.4. Specification Assumptions. The games of entry in retail markets that have been estimated in empirical applications have imposed different types of restrictions on the framework that we have presented in section 2.1, for example, restrictions on firm and market heterogeneity, firms' information, spatial competition, multi-store firms, dynamics, or the form of the structural functions. The motivations for these restrictions are diverse. Some restrictions are imposed to achieve identification or precise enough estimates of the parameters of interest, given the researcher's limited information on the characteristics of markets and firms. For instance, as we describe in section 3, prices and quantities at the store level are typically not observable to the researcher, and most sample information comes from firms' entry decisions. These limitations in the available data have motivated researchers to use simple specifications for the indirect variable profit function. Other restrictions are imposed for computational convenience in the solution and estimation of the model, for example, to obtain closed form solutions, to guarantee equilibrium uniqueness as it facilitates the estimation of the model, or to reduce the dimensionality of the space of firms' actions or states. In this subsection, we review some important models in this literature and discuss their main identification assumptions. We have organized these models in an approximate chronological order.

3.4.1. *Homogeneous firms.* Work in this field was pioneered by Bresnahan and Reiss. In Bresnahan and Reiss (1991a), they study several retail and professional industries in US, that is, pharmacies, tire dealers, doctors, and dentists. The main purpose of the paper is to estimate the 'nature' or 'degree' of competition for each of the industries: how fast variable profits decline when the number of firms in the market increases. More specifically, the authors are interested in estimating how many entrants are needed to achieve an oligopoly equilibrium equivalent to the competitive equilibrium, that is, hypothesis of contestable markets (Baumol 1982). For each industry, their dataset consists of a cross-section of M small 'isolated markets'. In section 3, we discuss the empirical motivation and implementation of the 'isolated markets' restriction. For the purpose of the model, a key aspect of this restriction is that the M local markets are independent in terms of demand and competition

such that the equilibrium in one market is independent of the one in the other markets. The model also assumes that each market consists of a single location, that is, L = 1, such that spatial competition is not explicitly incorporated in the model. For each local market, the researcher observes the number of active firms (n), a measure of market size (s), and some exogenous market characteristics that may affect demand and/or costs (x). Given this limited information, the researcher needs to restrict firm heterogeneity. Bresnahan and Reiss propose a static game between single-store firms where all the potential entrants in a market are identical and have complete information on demand and costs. The profit of a store is $\pi(n) = s * vp(x,n) - EC(x) - \varepsilon$, where $vp(x,n)$ represents variable profit per capita (per consumer) that depends on the number of active firms n, and $EC(x) + \varepsilon$ is the entry cost, where ε is unobservable to the researcher. The form of competition between active firms is not explicitly modelled. Instead, the authors consider a flexible specification of the variable profit per capita that is strictly decreasing but nonparametric in the number of active stores. Therefore, the specification is consistent with a general model of competition between homogeneous firms, or even between symmetrically differentiated firms.

Given these assumptions, the equilibrium in a local market can be described as a number of firms n^* that satisfies two conditions: (1) every active firm is maximizing profits by being active in the market, that is, $\pi(n^*) \geq 0$; and (2) every inactive firm is maximizing profits by being out of the market, that is, $\pi(n^* + 1) < 0$. That is, every firm is making its best response given the actions of the others. Since the profit function is strictly decreasing in the number of active firms, the equilibrium is unique and it can be represented using the following expression: for any value $n \in \{0, 1, 2, \ldots\}$,

$$\{n^* = n\} \iff \{\pi(n) \geq 0 \text{ and } \pi(n+1) < 0\}$$

$$\iff \{s * vp(x, n+1) - EC(x) < \varepsilon \leq s * vp(x, n) - EC(x)\}$$

Also, this condition implies that the distribution of the equilibrium number of firms given exogenous market characteristics is:

$$\Pr(n^* = n \mid s, x) = F(s * vp(x, n) - EC(x)) - F(s * vp(x, n+1) - EC(x))$$

where $F(.)$ is the CDF of ε. This representation of the equilibrium as an ordered discrete choice model is convenient for estimation.

In the absence of price and quantity data, the separate identification of the variable profit function and the entry cost function is based on the exclusion restrictions that variable profit depends on market size and on the number of active firms while the entry cost does not depend on these variables.

The previous model can be slightly modified to allow for firms' private information. This variant of the original model maintains the property of equilibrium uniqueness and most of the simplicity of the previous model. Suppose that now the entry cost of a firm is $EC(x) + \varepsilon_i$, where ε_i is private information of firm i and it is independently and identically distributed across firms with a CDF $F(.)$. There are N potential entrants in the local market. The presence of private information implies that, when potential entrants make entry decisions, they do not know ex ante the actual number of firms that will be active in the market. Instead, each firm has beliefs about the probability distribution of the number of other firms that are active. We represent these beliefs, for say firm i, using the function $G_i(n) \equiv \Pr(n^*_{-i} = n|s, x)$, where n^*_{-i} represents the number of firms other than i that are active in the market. Then, the expected profit of a firm if active in the market is:

$$\pi^e_i = \left[\sum_{n=0}^{N-1} G_i(n)\ s * vp(x, n+1)\right] - EC(x) - \varepsilon_i$$

The best response of a firm is to be active in the market if and only if its expected profit is positive or zero, that is, $a_i = 1\{\pi^e_i \geq 0\}$. Integrating this best response function over the distribution of the private information ε_i we obtain the best response probability of being active for firm i, that is, $P_i \equiv F(\left[\sum_{n=0}^{N-1} G_i(n)\ s * vp(x, n+1)\right] - EC(x))$. Since all firms are identical, up to their independent private information, it seems reasonable to impose the restriction that in equilibrium they all have the same beliefs and, therefore, the same best response probability of entry. Therefore, in equilibrium, firms' entry decisions $\{a_i\}$ are independent Bernoulli random variables with probability P, and the number of firms active other than i in the market has a Binomial distribution with argument $(N-1, P)$ such that $\Pr\blacksquare(n^*_{-i} = n) = B(n|N-1, P)$. In equilibrium, the beliefs function G(n) should be consistent with firms' best response probability P. Therefore, a Bayesian Nash Equilibrium in this model can be described as a probability of market entry P^* that is the best response probability when firms' beliefs about the distribution of other firms active in the market are $G(n) = B(n\ |N-1, P^*)$. We can represent this equilibrium condition using the following equation:

$$P^* = F\left(\left[\sum_{n=0}^{N-1} B(n|N-1, P^*)\ s * vp(x, n+1)\right] - EC(x)\right)$$

When the variable profit $vp(x, n)$ is a decreasing function in the number of active stores, the right-hand side in equation (14) is also a decreasing function in the probability of entry P, and this implies equilibrium uniqueness. In contrast to the complete information model in Bresnahan and Reiss (1991a), this incomplete information model does not have a closed form solution for the equilibrium distribution of the number of active firms in the market. However, the numerical solution of the fixed point problem in equation (9.14) is computationally very simple, and so are the estimation and comparative statistics using this model.

Given that the only difference between the two models described in section 2.2.1 is in their assumptions about firms' information, it seems reasonable to consider whether these models are observationally different or not. In other words, does the assumption on complete versus incomplete information have implications on the model predictions on competition? Grieco (2014) investigates this question in the context of an empirical application to local grocery markets. In Grieco's model firms are heterogeneous in terms of (common knowledge) observable variables, and this observable heterogeneity plays a key role in his approach to empirically distinguish between firms' public and private information. Note that the comparison of equilibrium conditions in equations (9.12) and (9.14) shows other testable difference between the two models. In the game of incomplete information, the number of potential entrants N has an effect on the whole probability distribution of the number of active firms: a larger number of potential entrants implies a shift to the right in the whole distribution of the number of active firms. In contrast, in the game of complete information, the value of N affects only the probability $Pr(n^* = N|s, x)$ but not the distribution of the number of active firms at values smaller than N. This empirical prediction has relevant economic implications: with incomplete information, the number of potential entrants has a positive effect on competition even in markets where this number is not binding.

3.4.2. *Bresnahan and Reiss (JPE, 1991).* They study several retail and professional industries in US: Doctors; Dentists; Pharmacies; Plumbers; car dealers; etc. For each industry, say car dealers, the dataset consists of a cross-section of M small, "isolated" markets. We index markets by m. For each market m, we observe the number of active firms (N_m), a measure of market size (S_m), and some exogenous market characteristics that may affect demand and/or costs (X_m).

$$\text{Data} = \{ \ N_m, \ S_m, \ X_m : m = 1, 2, ..., M\}$$

There are several empirical questions that they want to answer. First, they want to estimate the "nature" or "degree" of competition for each of the industries: that is, how fast variable profits decline when the number of firms in the market increase. Second, but related to the estimation of the degree of competition, BR are also interested in estimating how many entrants are needed to achieve an equilibrium equivalent to the competitive equilibrium, i.e., hypothesis of contestable markets.

[**Model**] Consider a market m. There is a number N^* of potential entrants in the market. Each firm decides whether to be active or not in the market. Let $\Pi_m(N)$ be the profit of an active firm in market m when there are N active firms. The function $\Pi_m(N)$ is strictly decreasing in N. If N_m is the equilibrium number of firms in market m, then it should satisfy

the following conditions:

$$\Pi_m(N_m) \geq 0 \qquad \text{and} \qquad \Pi_m(N_m + 1) < 0$$

That is, every firm is making his best response given the actions of the others. For active firms, their best response is to be active, and for inactive firms their best response is not to enter in the market.

To complete the model we have to specify the structure of the profit function $\Pi_m(N)$. Total profit is equal to variable profit, $V_m(N)$, minus fixed costs, $F_m(N)$:

$$\Pi_m(N) = V_m(N) - F_m(N)$$

In this model, where we do not observe prices or quantities, the key difference in the specification of variable profit and fixed cost is that variables profits increase with market size (in fact, they are proportional to market size) and fixed costs do not.

The variable profit function of an incumbent firm in market m when there are N active firms is:

$$V_m(N) = S_m \, v_m(N) = S_m \, \left(X_m^D \beta - \alpha(N)\right)$$

where S_m represent market size; $v_m(N)$ is the variable profit per-capita; X_m^D is a vector of market characteristics that may affect the demand of the product, e.g., per capita income, age distribution; β is a vector of parameters; and $\alpha(1)$, $\alpha(2)$, ...$\alpha(N)$ are parameters that capture the degree of competition, such that we expect that $\alpha(1) \leq \alpha(2) \leq \alpha(3) ... \leq \alpha(N)$. Given that there is not firm-heterogeneity in the variable profit function, there is an implicit assumption of homogeneous product or symmetrically differentiated product (e.g., Salop circle city).

The specification fixed cost is:

$$F_m(N) = X_m^C \, \gamma + \delta(N) + \varepsilon_m$$

where X_m^C is a vector of observable market characteristics that may affect the fixed cost, e.g., rental price; and ε_m is a market characteristic that is unobservable to the researchers but observable to the firms; and $\delta(1)$, $\delta(2)$, ...$\delta(N^*)$ are parameters. The dependence of the fixed cost with respect to the number of firms is very unconventional or non-standard in IO. Bresnahan and Reiss allow for this possibility and provide several interpretations. However, the interpretation of the parameters $\delta(1)$, $\delta(2)$, ...$\delta(N^*)$ is not completely clear. In some sense, BR allow the fixed cost to depend on the number firms in the market for robustness reasons. There are several possible interpretations for why fixed costs may depend on the number of firms in the market: (a) entry Deterrence: incumbents create barriers to entry; (b) a shortcut to allow for firm heterogeneity in fixed costs, in the sense that late entrants are less efficient in fixed costs; and (c) actual endogenous fixed costs, for instance rental prices

or other components of the fixed costs, no included in X_m^C, may increase with the number of incumbents (e.g., demand effect on rental prices). For any of these interpretations we expect $\delta(1) \leq \delta(2) \leq \delta(3) \ldots \leq \delta(N^*)$.

Since both $\alpha(N)$ and $\delta(N)$ increase with N, it is clear that the profit function $\Pi_m(N)$ declines with N. Therefore, as we anticipated above, the equilibrium condition for the number of firms in the market can be represented as follows. For $N \in \{0, 1, ..., N^*\}$

$$\{n_m = N\} \Leftrightarrow \{ \ \Pi_m(N) \geq 0 \quad \text{AND} \quad \Pi_m(N+1) < 0 \ \}$$

It is simple to show that the model has a unique equilibrium for any value of the exogenous variables and structural parameters. This is just a direct implication of the strict monotonicity of the profit function $\Pi_m(N)$.

We have a random sample $\{N_m, \ S_m, \ X_m^D, \ X_m^C : m = 1, 2, ..., M\}$ and we want to use this sample to estimate the vector of parameters:

$$\theta = \{\beta, \ \gamma, \ \sigma, \ \alpha(1), ..., \alpha(N^*), \ \delta(1), ..., \delta(N^*)\}$$

The unobserved component of the entry cost, ε_m, is assumed independent of (S_m, X_m^D, X_m^C) and it is i.i.d. over markets with distribution $N(0, \sigma)$. As usual in discrete choice models, σ is not identified. We normalize $\sigma = 1$, which means that we are really identifying the rest of the parameters up to scale. We should keep this in mind for the interpretation of the estimation results.

Given this model and sample, BR estimate θ by (conditional) ML:

$$\hat{\theta} = \arg\max_{\theta} \sum_{m=1}^{M} \log \Pr(N_m \mid \theta, S_m, X_m^D, X_m^C)$$

What is the form of the probabilities $\Pr(N_m | \theta, S_m, X_m, W_m)$ in B&R model? This entry model is equivalent to an **Ordered Probit model** for the number of firms. We can represent the condition $\{\Pi_m(n) \geq 0 \ \text{AND} \ \Pi_m(n+1) < 0\}$ in terms of thresholds for the unobservable variable ε_m.

$$\{N_m = n\} \Leftrightarrow \{T_m(n+1) < \varepsilon_m \leq T_m(n)\}$$

and for any $n \in \{1, 2, ..., N^*\}$ we have that

$$T_m(N) \equiv S_m X_m^D \beta - X_m^C \gamma - \alpha(n) S_m - \delta(n)$$

and $T_m(0) = +\infty$, $T_m(N^* + 1) = -\infty$. This is the structure of an ordered probit model. Therefore, the distribution of the number of firms conditional on the observed exogenous

market characteristics is:

$$\Pr(N_m = n | S_m, X_m^D, X_m^C) \;=\; \Phi\left(T_m(n)\right) - \Phi\left(T_m(n+1)\right)$$

$$\Phi\left(\; S_m X_m^D \beta - X_m^C \gamma - \alpha(n) S_m - \delta(n) \;\right)$$

$$-\;\; \Phi\left(\; S_m X_m^D \beta - X_m^C \gamma - \alpha(n+1) S_m - \delta(n+1) \;\right)$$

This is an **Ordered Probit** model. The model is very simple to estimate. Almost any econometric software package includes a command for the estimation of the ordered probit.

[**Application and Main Results**] **Data:** 202 "isolated local markets". Why isolated local markets? It is very important to include in our definition of market all the firms that are actually competing in the market and not more. Otherwise, we can introduce significant biases in the estimated parameters. If our definition of market is too narrow, such that we do not include all the firms that are actually in a market, we will conclude that there is little entry either because fixed costs are too large or the degree of competition is strong: i.e., we will overestimate the $\alpha's$ or the $\delta's$ or both. If our definition of market is too broad, such that we include firms that are not actually competing in the same market, we will conclude that there is significant entry and to rationalize this wee need fixed costs to be small or to have a low degree of competition between firms. Therefore, we will underestimate the $\alpha's$ or the $\delta's$ or both.

The most common mistake of a broad definition of market is to have a large city as a single market. The common mistake of a narrow definition of market is to have small towns that are close to each other, or close to a large town. To avoid these type of errors, BR construct "isolated local markets". The criteria to select isolated markets in US: (a) at least 20 miles from the nearest town of 1000 people or more; (b) At least 100 miles of cities with 100,000 people or more.

Population sizes between 500 and 75,000 people [see Figure 2 in the]. Industries (16): several retail industries (auto dealers, movie theaters,...) and many professions (doctors, dentists, plumbers, barbers, ...). The model is estimated for each industry separately.

Let $S(N)$ be the minimum market size to sustain N firms in the market. $S(N)$ are called **"entry thresholds"** and they can be obtained (estimated) using the estimated parameters. They do not depend on the normalization $\sigma = 1$. The main empirical results are: (a) For most industries, both $\alpha(N)$ and $\delta(N)$ increase with n. (b) There are very significant cross-industry differences in entry thresholds $S(N)$. (c) For most of the industries, entry thresholds $S(N)/N$ become constant for values of N greater than 4 or 5. Contestable markets?

3.4.3. *Nonparametric identification of Bresnahan-Reiss model [REMOVE ???]* **** REMOVE? IN AN APPENDIX? MERGED WITH THE BASIC MODEL? ADD TO THE PREVIOUS SECTION BUR MUCH REDUCED ******

There are many assumptions in BR model. To distinguish the key identifying assumptions from the accessory assumptions it is useful to present a nonparametric version of their model.

Assumption 1 [Homogeneous firms]: The profit of a firm active in the market depends on exogenous market characteristics, and on the (endogenous) number of firms active in the market. However, every firm active in the market obtains the same profit.

The profit function of an active firm in market m when the number of active firms is n is:

$$\Pi(n, X_m, \varepsilon_m)$$

where X_m is a vector of exogenous market characteristics that are observed by the researcher, and ξ_m is a vector of exogenous market characteristics that are unobserved to the researcher. Note that here we use n to represent an hypothetical value for the number of firms, not the actual realization that we denote by n_m.

The assumption of homogeneous firms is made both because data limitations (i.e., only market level data), and for convenience (i.e., equilibrium uniqueness of no econometric problems associated with the endogeneity of the decisions of "other firms"). However, it also restricts importantly the type of questions that we can analyze with these models. Therefore, this is the first assumption that we will relax in BR model.

Assumption 2: $\Pi(n, X_m, \varepsilon_m)$ is a strictly decreasing function in n.

This is a very weak assumption in a model of competition. In fact, it should hold even if the active firms in the market collude to achieve the monopoly outcome and share the total profits under that outcome.

Under Assumptions 1 and 2, the model establishes that the observed number of firms in the market, n_m, is determined by the following conditions:

$$\Pi(n_m, X_m, \varepsilon_m) \geq 0 \qquad \text{AND} \qquad \Pi(n_m + 1, X_m, \varepsilon_m) < 0$$

Assumption 3: The unobserved ε_m enters additively in the profit function, and its distribution function, F_ε, is independent of (n, X_m).

$$\Pi(n, X_m, \varepsilon_m) = \pi(n, X_m) - \varepsilon_m$$

The additive separability of the unobservable, alone without the independence assumption, is not really an assumption because we can always define ε_m as $\Pi(n, X_m, \varepsilon_m) - \pi(n, X_m)$. Therefore, the most important part of assumption 3 is the independence between ε_m and X_m.

This assumption plays an important role in identification. It can be relaxed to a certain extent. For instance, it is straightforward to allow for ε_m and X_m to be mean dependent. Define the conditional mean function $\mu(X_m) \equiv E(\varepsilon_m | X_m)$. Then, we can define $\pi^*(n, X_m) \equiv$

$\pi(n, X_m) + \mu(X_m)$ and $\varepsilon_m^* \equiv \varepsilon_m - \mu(X_m)$, and therefore $\pi(n, X_m) - \varepsilon_m = \pi^*(n, X_m) - \varepsilon_m^*$. By construction, ε_m^* is mean independent of X_m, and we can prove (see below) the identification of the function $\pi^*(n, X_m)$. The dependence of $\pi^*(n, X_m)$ with respect to n is the same as the dependence of $\pi(n, X_m)$ with respect to n. So, if we are interested only on how the profit function depends on the number of firms (for different values of X) the functions π^* and π are equivalent.

Therefore, without loss of generality we can assume that ε_m is mean independent if X_m and it has zero mean.

According to this model, for any $n \in \{0, 1, ...\}$:

$$\{n_m \;\; = \;\; n\} \Leftrightarrow \{\pi(n, X_m) - \varepsilon_m \geq 0 \quad \text{AND} \quad \pi(n+1, X_m) - \varepsilon_m < 0\}$$

$$\Leftrightarrow \;\; \{\pi(n+1, X_m) < \varepsilon_m \leq \pi(n, X_m)\}$$

Or,

$$\{n_m > n\} \Leftrightarrow \{\varepsilon_m \leq \pi(n+1, X_m)\}$$

Therefore, the predictions of the model can be summarized by the following expression for the conditional distribution of the number of firms: for any value (n, X):

$$\Pr(n_m > n \mid X_m = X) = F_\varepsilon(\pi(n+1, X))$$

Note that the probability $\Pr(n_m > n \mid X_m = X)$ can be identified from the data using a nonparametric estimator such as a frequency estimator (if X includes only discrete variables) or a kernel estimator. For instance, a consistent kernel estimator of $\Pr(n_m > n \mid X_m = X)$ is:

$$\frac{\sum_{m=1}^{M} 1\{n_m > n\} \; K\left(\frac{X_m - X}{b}\right)}{\sum_{m=1}^{M} K\left(\frac{X_m - X}{b}\right)}$$

where $1\{.\}$ is the indicator function, $K(.)$ is a kernel function (e.g., the pdf of the standard normal distribution), and b is a bandwidth.

Given $\Pr(n_m > n \mid X_m = X)$, we want to identify the distribution function F_ε and the profit function $\pi(n, X)$. Without further assumptions we cannot separately identify F_ε and $\pi(n, X)$.

Assumption 4: Let σ^2 be the variance of ε_m. The probability distribution of ε_m/σ is known to the researcher. For instance, ε_m/σ is distributed $N(0, 1)$.

Under Assumption 4, we can identify $\pi(n, X)$ up to scale:

$$\frac{\pi(n, X)}{\sigma} = \Phi^{-1}\left(\Pr(n_m > n - 1 \mid X_m = X)\right)$$

such that $\frac{\pi(1,X)}{\sigma} = \Phi^{-1}\left(\Pr\left(n_m > 0 \mid X_m = X\right)\right)$, $\frac{\pi(2,X)}{\sigma} = \Phi^{-1}\left(\Pr\left(n_m > 1 \mid X_m = X\right)\right)$, and so on, where Φ^{-1} is the inverse of the CDF of the standard normal distribution.

Though the profit function is only identified up to scale, we can identify the percentage reduction in profits when the number of firms increases:

$$
\begin{aligned}
\Delta(n, X) &\equiv \ln\left(\pi(n+1, X)\right) - \ln\left(\pi(n, X)\right) \\
\\
&= \ln\left(\frac{\pi(n+1, X)}{\sigma}\right) - \ln\left(\frac{\pi(n, X)}{\sigma}\right) \\
\\
&= \ln\left(\Phi^{-1}\left(\Pr\left(n_m > n \mid X_m = X\right)\right)\right) - \ln\left(\Phi^{-1}\left(\Pr\left(n_m > n - 1 \mid X_m = X\right)\right)\right)
\end{aligned}
$$

$\Delta(n, X)$ is the percentage reduction in the profit of a firm in a market with characteristics X when we change exogenously the number of competitors from n to $n + 1$.

In fact, $\Delta(n, X)$ is a good measure of the % increase in total welfare associated with an exogenous increase in the number of firms. If $\Delta(n, X)$ is close to zero, then a market with characteristics X and a number of firms n is very close in terms of firms' profits and consumer welfare to a perfectly competitive market. This is an important parameter of interest for instance for a regulator who considers the implications of a policy that tries to encourage more entry, e.g., a policy that reduces entry costs.

Note that the value for the potential number of firms N, does not play any role for the predictions or the identification of the model.

Given a known distribution for ε_m/σ, the assumption of homocedasticity of ε_m is testable. Also, under the assumption of homocedasticity of ε_m/σ, and a variable in X that has a monotonic effect on $\pi(n, X)$ and large range of variation (e.g., market size), then the distribution of ε_m/σ can be identified nonparametrically.

Therefore, the key identifying assumptions of the model are in the assumptions of: (a) homogeneous profits for all firms; (b) free entry conditions; and (c) our definition of local market. We will relax these assumptions.

3.4.4. *Entry with endogenous product choice.* Mazzeo (2002) studies market entry in the motel industry using local markets along US interstate highways.20 A local market is defined as a narrow region around a highway exit. Mazzeo's model maintains most of the assumptions in Bresnahan and Reiss (1991a), such as no spatial competition (that is, L=1), ex ante homogeneous firms, complete information, no multi-store firms, and no dynamics. However, he extends the Bresnahan–Reiss model in an interesting dimension: it introduces endogenous product differentiation. More specifically, firms not only decide whether to enter in a market but they also choose the type of product: low-quality product E (that is, economy hotel), or high-quality product H (that is, upscale hotel).21 Product differentiation makes competition

less intense, and it can increase firms' profits. However, firms have also an incentive to offer the type of product for which demand is stronger.

The profit of an active hotel of type $T \in \{E, H\}$ is:

$$\pi_T(n_E, n_H) = s * v_T(x, n_E, n_H) - EC_T(x) - \varepsilon_T$$

where n_E and n_H represent the number of active hotels with low and high quality, respectively, in the local market. Similarly to the Bresnahan–Reiss model, $v_T(.)$ is the variable profit per capita and $EC_T(x) + \varepsilon_T$ is the entry cost for type T hotels, where ε_T is unobservable to the researcher. Mazzeo solves and estimates his model under two different equilibrium concepts: Stackelberg and what he terms a 'two-stage game'. A computational advantage of the two-stage game is that under the assumptions of the model the equilibrium is unique. In the first stage, the total number of active hotels, $n \equiv n_E + n_H$, is determined in a similar way as in the Bresnahan–Reiss model. Hotels enter the market as long as there is some configuration (n_E, n_H) where both low-quality and high-quality hotels make positive profits. Define the first-stage profit function as:

$$\Pi(n) \equiv \max_{n_E, n_H : n_E + n_H = n} \left\{ \min[\pi_E(n_E, n_H) \ , \ \pi_H(n_E, n_H)] \right\} \blacksquare$$

Then, the equilibrium number of hotels in the first stage is the value n^* that satisfies two conditions: (1) every active firm wants to be in the market, that is, $\Pi(n^*) \geq 0$; and (2) every inactive firm prefers to be out of the market, that is, $\Pi(n^* + 1) < 0$. If the profit functions π_E and π_H are strictly decreasing functions in the number of active firms (n_E, n_H), then $\Pi(n)$ is also a strictly decreasing function, and the equilibrium number of stores in the first stage, n^*, is unique. In the second stage, active hotels choose simultaneously their type or quality level. In this second stage, an equilibrium is a pair (n_E^*, n_H^*) such that every firm chooses the type that maximizes its profit given the choices of the other firms: low quality firms are not better off by switching to high quality, and vice versa,

$$\pi_E(n_E^*, n_H^*) \ \geq \ \pi_H(n_E^* - 1, n_H^* + 1)$$

$$\pi_H(n_E^*, n_E^*) \ \geq \ \pi_E(n_E^* + 1, n_E^* - 1)$$

Mazzeo shows that the equilibrium pair (n_E^*, n_H^*) in this second stage is also unique.

Using these equilibrium conditions, it is possible to obtain a closed form expression for the (quadrangle) region in the space of the unobservables $(\varepsilon_E, \varepsilon_H)$ that generate a particular value of the equilibrium pair (n_E^*, n_H^*). Let $R_\varepsilon(n_E, n_H; s, x)$ be the quadrangle region in $\mathbb{R}^2$ associated with the pair (n_E, n_H) given exogenous market characteristics (s, x) , and let $F(\varepsilon_E, \varepsilon_H)$ be the CDF of the unobservable variables. Then, we have that:

$$\Pr(n_E^* = n_E, n_H^* = n_H | s, x) = \int 1\{(\varepsilon_E, \varepsilon_H) \in R_\varepsilon(n_E, n_H; s, x) \ dF(\varepsilon_E, \varepsilon_H)$$

In the empirical application, Mazzeo finds that hotels have strong incentives to differentiate from their rivals to avoid nose-to-nose competition.

Ellickson and Misra (2008) estimate a game of incomplete information for the US supermarket industry where supermarkets choose the type of 'pricing strategy': 'everyday low price' (EDLP) versus 'high-low' pricing.24 The choice of pricing strategy can be seen as a form of horizontal product differentiation. The authors find evidence of strategic complementarity between supermarkets pricing strategies: firms competing in the same market tend to adopt the same pricing strategy not only because they face the same type of consumers but also because there are positive synergies in the adoption of the same strategy. From an empirical point of view, this result is more controversial than Mazzeo's finding of firms' incentive to differentiate from each other. In particular, the existence of unobservables that are positively correlated across firms but are not fully accounted in the econometric model, may generate a spurious estimate of positive spillovers in the adoption of the same strategy. Vitorino (2012) estimates a game of store entry in shopping centers that allows for incomplete information, positive spillover effects among stores, and also unobserved market heterogeneity for the researcher that is common knowledge to firms. Her empirical results show that, after controlling for unobserved market heterogeneity, firms face business stealing effects but also significant incentives to collocate, and that the relative magnitude of these two effects varies substantially across store types.

3.4.5. *Firm heterogeneity.* The assumption that all potential entrants and incumbents are homogeneous in their variable profits and entry costs is very convenient and facilitates the estimation, but it is also very unrealistic in many applications. A potentially very important factor in the determination of market structure is that firms, potential entrants, are ex-ante heterogeneous. In many applications we want to take into account this heterogeneity. Allowing for firm heterogeneity introduces two important issues in these models: endogenous explanatory variables, and multiple equilibria. We will comment on different approaches that have been used to deal with these issues.

Consider an industry with N potential entrants. For instance, the airline industry. These potential entrants decide whether to be active or not in a market. We observe M different realizations of this entry game. These realizations can be different geographic markets (different routes of or city pairs, e.g., Toronto-New York, Montreal-Washington, etc) or different time periods of time. For the sake of concreteness, we refer to these different realizations of the entry game as "local markets" or "submarkets". We index firms with $i \in \{1, 2, ..., N\}$ and submarkets with $m \in \{1, 2, ..., M\}$.

Let $a_{im} \in \{0,1\}$ be a the binary indicator of the event "firm i is active in market m". For a given market m, the N firms choose simultaneously whether to be active or not in the market. When making his decision, a firm wants to maximize its profit.

Once firms have decided to be active or not in the market, active firms compete in prices or in quantities and firms' profits are realized. For the moment, we do not make it explicit the specific form of competition in this second part of the game, or the structure of demand and variable costs. We take as given an "indirect profit function" that depends on exogenous market and firm characteristics and one the number and the identity of the active firms in the market. This indirect profit function comes from a model of price or quantity competition, but at this point we do not make that model explicit here. Also, we consider that the researcher does not have access to data on firms' prices and quantities such that demand and variable cost parameters in the profit function cannot be estimated from demand, and/or Bertrand/Cournot best response functions.

The (indirect) profit function of an incumbent firm depends on market and firm characteristics affecting demand and costs, and on the entry decisions of the other potential entrants:

$$\Pi_{im} = \begin{cases} \Pi_i\left(x_{im}, \ \varepsilon_{im}, \ a_{-im}\right) & \text{if} \quad a_{im} = 1 \\ \\ 0 & \text{if} \quad a_{im} = 0 \end{cases}$$

where x_{im} and ε_{im} are vectors of exogenous market and firm characteristics, and $a_{-im} \equiv \{a_{jm} : j \neq i\}$. The vector x_{im} is observable to the researcher while ε_{im} is unobserved to the researcher. For the moment we assume that $x_m \equiv \{x_{1m}, x_{2m}, ..., x_{Nm}\}$ and $\varepsilon_m \equiv \{\varepsilon_{1m}, \varepsilon_{2m}, ..., \varepsilon_{Nm}\}$ are common knowledge for all players.

For instance, in the example of the airline industry, the vector x_{im} may include market characteristics such as population and socioeconomic characteristics in the two cities that affect demand, characteristics of the airports such as measures of congestion (that affect costs), and firm characteristics such as the number of other connections that the airline has in the two airports (that affect operating costs due to economies of scale and scope).

The N firms chose simultaneously $\{a_{1m}, a_{2m}, ..., a_{Nm}\}$ and the assumptions of Nash equilibrium hold. A **Nash equilibrium** in this the entry game is an N-tuple $a_m^* = (a_{1m}^*, a_{2m}^*, ..., a_{Nm}^*)$ such that for any player i:

$$a_{im}^* = 1\left\{ \ \Pi_i\left(x_{im}, \ \varepsilon_{im}, \ a_{-im}^*\right) \geq 0 \ \right\}$$

where $1\{.\}$ is the indicator.

Given a dataset with information on $\{a_{im}, x_{im}\}$ for every firm in the M markets, we want to use this model to learn about the structure of the profit function $\Pi_i(.)$. In these applications, we are particularly interested in the effect of other firms' entry decisions on a

firm's profit. For instance, how Southwest entry in the Chicago-Boston submarket affects the profit of American Airlines.

For the sake of concreteness, consider the following specification of the profit function:

$$\Pi_{im} = x_{im}\ \beta_i - \sum_{j \neq i} a_{jm}\ \delta_{ij} + \varepsilon_{im}$$

where x_{im} is a $1 \times K$ vector of observable market and firm characteristics; β_i is a $K \times 1$ vector of parameters; $\delta_i = \{\delta_{ij} : j \neq i\}$ is a $(N-1) \times 1$ vector of parameters, with δ_{ij} being the effect of firm $j's$ entry on firm $i's$ profit; ε_{im} is zero mean random variable that is observable to the players but unobservable to the econometrician.

We assume that ε_{im} is independent of x_m, and it is $i.i.d.$ over m, and independent across i. If x_{im} includes a constant term, then without loss of generality $E(\varepsilon_{im}) = 0$. Define $\sigma_i^2 \equiv Var(\varepsilon_{im})$. Then, we also assume that the probability distribution of $\varepsilon_{im}/\sigma_i$ is known to the researcher. For instance, $\varepsilon_{im}/\sigma_i$ has a standard normal distribution.

The econometric model can be described as system of N simultaneous equations where the endogenous variables are the entry dummy variables:

$$a_{im} = 1\left\{\ x_{im}\ \beta_i - \sum_{j \neq i} a_{jm}\ \delta_{ij} + \varepsilon_{im} \geq 0\ \right\}$$

We want to estimate the vector of parameters $\theta = \left\{\dfrac{\beta_i}{\sigma_i}, \dfrac{\delta_i}{\sigma_i} : i = 1, 2, ..., N\right\}$.

There are two main econometric issues in the estimation of this model: (1) endogenous explanatory variables, a_{jm}; and (2) multiple equilibria.

Endogeneity of other players' actions. In the structural (best response) equation

$$a_{im} = 1\left\{\ x_{im}\ \beta_i - \sum_{j \neq i} a_{jm}\ \delta_{ij} + \varepsilon_{im} \geq 0\ \right\}$$

the actions of the other players, $\{a_{jm} : j \neq i\}$ are endogenous in an econometric sense. That is, a_{jm} is correlated with the unobserved term ε_{im}, and ignoring this correlation can lead to serious biases in our estimates of the parameters β_i and δ_i.

There two sources of endogeneity or correlation between a_{jm} and ε_{im}: simultaneity and common unobservables between ε_{im} and ε_{jm}. It is interesting to distinguish between these two sources of endogeneity because they bias the parameter δ_{ij} in opposite directions.

Simultaneity. An equilibrium of the model is a reduced form equation where we represent the action of each player as a function of only exogenous variables in x_m and ε_m. In this reduced form, a_{jm} depends on ε_{im}. It is possible to show that this dependence is negative: keeping all the other exogenous variables constant if ε_{im} is small enough then $a_{jm} = 0$, and if ε_{im} is large enough then $a_{jm} = 1$. Suppose that our estimator of δ_{ij} ignores this dependence. Then, the negative dependence between a_{jm} and ε_{im} contributes to generate a upward bias in the estimator of δ_{ij}.

That is, we will spuriously over-estimate the negative effect of Southwest on the profit of American Airlines because Southwest tends to enter in those markets where AA has low values of ε_{im}.

Positively correlated unobservables. It is reasonable to expect that ε_{im} and ε_{jm} are positively correlated. This is because both ε_{im} and ε_{jm} contain unobserved market characteristics that affect in a similar way, or at least in the same direction, to all the firms in the same market. Some markets are more profitable than others for every firm, and part of this market heterogeneity is observable to firms but unobservables to us as researchers. The positive correlation between ε_{im} and ε_{jm} generates also a positive dependence between a_{jm} and ε_{im}.

For instance, suppose that $\varepsilon_{im} = \omega_m + u_{im}$, where ω_m represents the common market effect, and u_{im} is independent across firms. Then, keeping x_m and the unobserved u variables constant, if ω_m is small enough then ε_{im} and $a_{jm} = 0$, and if ω_m is large enough then ε_{im} is large and $a_{jm} = 1$. Suppose that our estimator of δ_{ij} ignores this dependence. Then, the negative positive dependence between a_{jm} and ε_{im} contributes to generate a downward bias in the estimator of δ_{ij}. In fact, the estimate of δ_{ij} could have the wrong sign, i.e., being negative instead of positive.

That is, we can spuriously find that American Airlines benefits for the operation of Continental in the same market because we tend to observe that these firms are always active in the same markets. This positive correlation between a_{im} and a_{jm} can be completely driven by the positive correlation between ε_{im} and ε_{jm}.

These two sources of endogeneity generate biases of opposite sign in δ_{ij}. There is evidence from different empirical applications that the biased due to unobserved market effects is much more important than the simultaneity bias. Examples: Collard-Wexler (WP, 2007) US cement industry; Aguirregabiria and Mira (*Econometrica*, 2007) different retail industries in Chile; Aguirregabiria and Ho (WP, 2007) US airline industry; Ellickson and Misra (*Marketing Science*, 2008) US supermarket industry.

How do we deal with this endogeneity problem? The intuition for the identification in this model is similar to the identification using standard Instrumental Variables (IV) and Control Function (CF) approaches.

"IV approach": There are exogenous firm characteristics in x_{jm} that affect the action of firm j but do not have a direct effect on the action of firm i: i.e., observable characteristics with $\beta_j \neq 0$ but $\beta_i = 0$.

"CF approach": There is an observable variable C_{it} that "proxies" or "controls for" the endogenous part of ε_{im} such that if we include C_{it} in the equation for firm i then the new error term in that equation and a_{jm} become independent (conditional on C_{it}).

The method of instrumental variables is the most common approach to deal with endogeneity in linear models. However, IV or GMM cannot be applied to estimate discrete choice models with endogenous variables. Control function approaches: Rivers and Vuong (1988), Vytlacil and Yilditz (2006). These approaches have not been extended yet to deal with models with multiple equilibria or "multiple reduced forms".

An alternative approach is Maximum likelihood: If we derive the probability distribution of the dummy endogenous variables conditional on the exogenous variables (i.e., the reduced form of the model), we can use these probabilities to estimate the model by maximum likelihood.

$$l(\theta) = \sum_{m=1}^{M} \ln \Pr(a_{1m}, a_{2m}, ..., a_{Nm} \mid x_m, \ \theta)$$

This is the approach that has been most commonly used in this literature. However, we will have to deal with the problem of multiple equilibria.

Multiple equilibria. Consider the model with two players and assume that $\delta_1 \geq 0$ and $\delta_2 \geq 0$.

$$a_1 \ = \ 1\{ \ x_1\beta_1 - \delta_1 \ a_2 + \varepsilon_1 \geq 0 \ \}$$

$$a_2 \ = \ 1\{ \ x_2\beta_2 - \delta_2 \ a_1 + \varepsilon_2 \geq 0 \ \}$$

The reduced form of the model is a representation of the endogenous variables (a_1, a_2) only in terms of exogenous variables and parameters. This is the reduced for of this model:

$$\{x_1\beta_1 + \varepsilon_1 < 0\} \ \& \ \{x_2\beta_2 + \varepsilon_2 < 0\} \ \Rightarrow \ (a_1, a_2) = (0, 0)$$

$$\{x_1\beta_1 - \delta_1 + \varepsilon_1 \geq 0\} \& \{x_2\beta_2 - \delta_2 + \varepsilon_2 \geq 0\} \ \Rightarrow \ (a_1, a_2) = (1, 1)$$

$$\{x_1\beta_1 - \delta_1 + \varepsilon_1 < 0\} \ \& \ \{x_2\beta_2 + \varepsilon_2 \geq 0\} \ \Rightarrow \ (a_1, a_2) = (0, 1)$$

$$\{x_1\beta_1 + \varepsilon_1 \geq 0\} \ \& \ \{x_2\beta_2 - \delta_2 + \varepsilon_2 < 0\} \ \Rightarrow \ (a_1, a_2) = (1, 0)$$

The graphical representation in the space $(\varepsilon_1, \varepsilon_2)$ is:

****

FIGURE ON MULTIPLE EQUILIBRIA: 5 regions

(Tamer, RESTUD 2003)

****

Note that when:

$$\{0 \leq x_1\beta_1 + \varepsilon_1 < \delta_1\} \ \text{and} \ \{0 \leq x_2\beta_2 + \varepsilon_2 < \delta_2\}$$

we have two Nash equilibria: $(a_1, a_2) = (0, 1)$ and $(a_1, a_2) = (1, 0)$. For this range of values of $(\varepsilon_1.\varepsilon_2)$, the reduced form (i.e., the equilibrium) is not uniquely determined. Therefore, we can not uniquely determine the probability $\Pr(a_{1m}, a_{2m}|x_m; \theta)$ that we need to estimate the

model by ML. We know $\Pr(1,1|\theta)$, and $\Pr(0,0|\theta)$, but we only have lower and upper bounds for $\Pr(0,1|\theta)$ and $\Pr(1,0|\theta)$.

The problem of indeterminacy of the probabilities of different outcomes becomes even more serious in empirical games with more than 2 players or/and more than two choice alternatives.

There have been different approaches to deal with this problem of multiple equilibria. Some authors have **imposed additional structure in the model to guarantee equilibrium uniqueness** or at least uniqueness of some observable outcome (e.g., number of entrants). A second group of studies do not impose additional structure and use methods such that **moment inequalities** or **pseudo maximum likelihood** to estimate structural parameters. The main motivation of this second group of studies is that identification and multiple equilibria are different problems and we do not need equilibrium uniqueness to identify parameters.

3.4.6. *Entry games with incomplete information.*

Model and basic assumptions. Consider a market with N potential entrants. If firm i does not operate in market m ($a_{im} = 0$), its profit is zero. If the firm is active in the market ($a_{im} = 1$), the profit is:

$$\Pi_{im} = \Pi_i(x_m, a_{-im}) - \varepsilon_{im} \tag{3.1}$$

For instance,

$$\Pi_{im} = x_{im}\ \beta_i - \varepsilon_{im} - \sum_{j \neq i} \delta_{ij}\ a_{jm} \tag{3.2}$$

where β_i and δ_i are parameters. These parameters and the vector $s_m = (s_{1m}, s_{2m}, ..., s_{Nm})$ contain the variables which are common knowledge for all players. Now ε_{im} is private information of firm i. For the moment, we assume that private information variables are independent of s_m, independently distributed over firms with distribution functions $G_i(\varepsilon_{im})$. The distribution function G_i is strictly increasing in $\mathbb{R}$. The information of player i is (s_m, ε_{im}).

A player's strategy depends on the variables in his information set. Let $\alpha \equiv \{\alpha_i(s_m, \varepsilon_{im}) : i = 1, 2, ..., N\}$ be a set of strategy functions, one for each player, such that $\alpha_i : S \times \mathbb{R} \to \{0, 1\}$. The actual payoff/profit Π_{im} is unknown to player i because the private information of the other players is unknown to player i. Players maximize expected profits:

$$\pi_i(s_m, \varepsilon_{im}, \alpha_{-i}) = s_{im}\ \beta_i - \varepsilon_{im} - \sum_{j \neq i} \delta_{ij} \left[\int I\left\{\alpha_j(s_m, \varepsilon_{jm}) = 1\right\} dG_j(\varepsilon_{jm}) \right] \tag{3.3}$$

or:

$$
\begin{aligned}
\pi_i(s_m, \varepsilon_{im}, \alpha_{-i}) &= s_{im}\ \beta_i - \varepsilon_{im} - \sum_{j \neq i} \delta_{ij}\ P_j^\alpha(s_m) \\
&= s_{im}\ \beta_i - \varepsilon_{im} - P_{-i}^\alpha(s_m)'\delta_i
\end{aligned} \tag{3.4}
$$

where $P_j^\alpha(s_m) \equiv \int I\{\alpha_j(s_m, \varepsilon_{jm}) = 1\} dG_j(\varepsilon_{jm})$ is player j's probability of entry if he behaves according to his strategy in α.

Suppose that players other than i play their respective strategies in α. What is player i's best response? Let $b_i(s_m, \varepsilon_{im}, \alpha_{-i})$ be player i's best response function. This function is:

$$
\begin{aligned}
b_i(s_m, \varepsilon_{im}, \alpha_{-i}) &= I\{ \pi_i(s_m, \varepsilon_{im}, \alpha_{-i}) \geq 0 \} \\
&= I\{ \varepsilon_{im} \leq s_{im}\ \beta_i - P_{-i}^\alpha(s_m)' \delta_i \}
\end{aligned}
\tag{3.5}
$$

Associated with the best response function b_i (in the space of strategies), we can define a *best response probability function* in the space of probabilities as:

$$
\begin{aligned}
\Psi_i(s_m, P_{-i}^\alpha) &= \int I\{ b_i(s_m, \varepsilon_{im}, \alpha_{-i}) = 1 \} dG_i(\varepsilon_{im}) \\
&= \int I\{ \varepsilon_{im} \leq s_{im}\ \beta_i - P_{-i}^\alpha(s_m)' \delta_i \} dG_i(\varepsilon_{im})
\end{aligned}
\tag{3.6}
$$

$$
G_i\left(s_{im}\ \beta_i - P_{-i}^\alpha(s_m)' \delta_i \right)
$$

A *Bayesian Nash equilibrium* (BNE) in this model is a set of strategy functions α^* such that, for any player i and any value of (s_m, ε_{im}), we have that:

$$
\alpha_i^*(s_m, \varepsilon_{im}) = b_i(s_m, \varepsilon_{im}, \alpha_{-i}^*)
\tag{3.7}
$$

Associated with the set of strategies α^* we can define a set of choice probability functions $P^* = \{P_i^*(s_m) : i = 1, 2, ..., N\}$ such that $P_i^*(s_m) \equiv \int I\{\alpha_i^*(s_m, \varepsilon_{im}) = 1\} dG_i(\varepsilon_{im})$. Note that these equilibrium choice probabilities are such that, for any player i and any value of s_m:

$$
\begin{aligned}
P_i^*(s_m) &= \Psi_i(s_m, P_{-i}^*) \\
&= G_i\left(s_{im}\ \beta_i - P_{-i}^*(s_m)' \delta_i \right)
\end{aligned}
\tag{3.8}
$$

Therefore, we can define a BNE in terms of strategy functions α^* or in terms of choice probabilities P^*. There is a one-to-one relationship between α^* and P^*. Given α^*, it is clear that there is only one set of choice probabilities P^* defines as $P_i^*(s_m) \equiv \int I\{\alpha_i^*(s_m, \varepsilon_{im}) = 1\} dG_i(\varepsilon_{im})$. And given P^*, there is only one set of strategies α^* that is a BNE and it is consistent with P^*. These strategy functions are:

$$
\alpha_i^*(s_m, \varepsilon_{im}) = I\{ \varepsilon_{im} \leq s_{im}\ \beta_i - P_{-i}^*(s_m)' \delta_i \}
\tag{3.9}
$$

Suppose that the distribution of ε_{im} is known up to some scale parameter σ_i. For instance, suppose that $\varepsilon_{im} \sim iid\ N(0,1)$. Then, we have that equilibrium choice probabilities in market m solve the fixed point mapping in probability space:

$$
P_i^*(s_m) = \Phi\left(s_{im}\ \frac{\beta_i}{\sigma_i} - P_{-i}^\alpha(s_m)' \frac{\delta_i}{\sigma_i} \right)
$$

For notational simplicity we will use β_i and δ_i to represent $\dfrac{\beta_i}{\sigma_i}$ and $\dfrac{\delta_i}{\sigma_i}$, respectively.

We use θ to represent the vector of structural parameters $\{\beta_i, \delta_i : i = 1, 2, ..., N\}$. To emphasize that equilibrium probabilities deped on θ we use $P(s_m, \theta) = \{P_i(s_m, \theta) : i = 1, 2, ..., N\}$ to represent a vector of equilibrium probabilities associated with the exogenous conditions (s_m, θ). In general, there are values of $(s_m; \theta)$ for which the model has multiple equilibria. This is very common in models where players are heterogeneous, but we can find also multiple symmetric equilibria in models with homogeneous players, specially if there is strategic complementarity (i.e., $\delta_i < 0$) as in coordination games.

Data and identification. Suppose that we observe this game played at M independent markets. We observe players' actions and a subset of the common knowledge state variables, $x_{im} \subseteq s_{im}$. That is,

$$Data = \{x_{im}, a_{im} : m = 1, 2, ..., M;\ i = 1, 2, ..., N\} \tag{3.10}$$

The researcher does not observe private information variables. It is important to distinguish two cases:

Case I: No common knowledge unobservables, i.e., $x_{im} = s_{im}$.

Case II: Common knowledge unobservables, i.e., $s_{im} = (x_{im}, \omega_{im})$, where ω_{im} is unobservable.

Case I: No common knowledge unobservables

(A) Data with global players. Suppose that we have a random sample of markets and we observe:

$$\{x_{im}, a_{im} : m = 1, 2, ..., M;\ i = 1, 2, ..., N\} \tag{3.11}$$

Let $P^0 = \{P_i^0(x) : i = 1, 2, .., N; x \in X\}$ be players' entry probabilities in the the population under study. The population is an equilibrium of the model. That is, there is a θ^0 such that, for any i and any $x \in X$:

$$P_i^0(x) = \Phi \left(x_i\ \beta_i^0 - P_{-i}^0(x)'\delta_i^0 \right) \tag{3.12}$$

From our sample, we can nonparametrically identify the population P^0, i.e., $P_i^0(x) = E(a_{im}|x_m = x)$. Given P^0 and the equilibrium conditions in (3.12), can we uniquely identify θ^0? Notice that we can write these equations as:

$$\Phi^{-1} \left(P_i^0(x_m) \right) = x_{im}\ \beta_i^0 - P_{-i}^0(x_m)'\delta_i^0 = Z_{im}\ \theta_i^0$$

Define $Y_{im} \equiv \Phi^{-1} \left(P_i^0(x_m) \right)$; $Z_{im} \equiv (x_{im}, P_{-i}^0(x_m))$; and $\theta_i^0 \equiv (\beta_i^0, \delta_i^0)$. Then,

$$Y_{im} = Z_{im}\ \theta_i^0$$

And we can also write this system as:

$$E(Z_{im}'Y_{im}) = E(Z_{im}'Z_{im})\ \theta_i^0$$

It is clear θ_i^0 is uniquely identified if $E(Z_{im}'Z_{im})$ is a nonsigular matrix. Note that if x_{im} contains variables that variate both over markets and over players then we have exclusion restrictions that imply that $E(Z_{im}'Z_{im})$ is a nonsigular matrix.

(B) Data with only local players. Suppose that we have a random sample of markets and we observe:

$$\{x_m, n_m : m = 1, 2, ..., M\} \tag{3.13}$$

Let $P^0 = \{P^0(x) : x \in X\}$ be the entry probabilities in the the population under study. The population is an equilibrium of the model, and therefore there is a θ^0 such that for any $x \in X$:

$$P^0(x) = \Phi\left(x\ \beta - \delta\ H(P^0[x])\right) \tag{3.14}$$

From our sample, we can nonparametrically identify the population P^0. To see this, notice that: (1) we can identify the distribution for the number of firms: $\Pr(n_m = n | x_m = x)$; (2) the model implies that conditional on $x_m = x$ the number of firms follows a Binomial distribution with arguments N and $P^0(x)$, therefore

$$\Pr(n_m = n | x_m = x) = \binom{n}{N}\ P^0(x)^n\ \left(1 - P^0(x)\right)^{N-n};$$

and (3) given the previous expression, we can obtain the $P^0(x)$ associated with $\Pr(n_m = n | x_m = x)$. Given P^0 and the equilibrium condition $P^0(x) = \Phi\left(x\ \beta - \delta\ H(P^0[x])\right)$, can we uniquely identify θ^0? Notice that we can write these equations as:

$$Y_m = x_m\ \beta^0 - \delta^0\ H(P^0[x_m]) = Z_m\ \theta^0$$

where $Y_m \equiv \Phi^{-1}\left(P^0(x_m)\right)$; $\theta^0 \equiv (\beta^0, \delta^0)$; and $Z_m \equiv (x_m, H(P^0[x_m]))$. And we can also write this system as:

$$E(Z_m'Y_m) = E(Z_m'Z_m)\ \theta^0$$

It is clear θ^0 is uniquely identified if $E(Z_m'Z_m)$ is a nonsingular matrix.

Case II: Common knowledge unobservables

Intuition: conditional on x_m, players actions are still correlated across markets. This is evidence that ....

In applications where we do not observe the identity of the potential entrants, we consider a model without firm heterogeneity:

$$\Pi_{im} = x_m\ \beta - \delta\ h\left(1 + \sum_{j \neq i} a_{jm}\right) + \varepsilon_{im} \tag{3.15}$$

A symmetric Bayesian Nash equilibrium in this model is a probability of entry $P^*(x_m; \theta)$ that solves the fixed point problem:

$$P^*(x_m; \theta) = \Phi\left(x_m\ \beta - \delta\ H(P[x_t, \theta])\right) \tag{3.16}$$

where $H(P)$ is the expected value of $h\left(1 + \sum_{j\neq i} a_j\right)$ conditional on the information of firm i, and under the condition that the other firms behave according to their entry probabilities in P. That is,

$$H(P) = \sum_{a_{-i}} \left(\prod_{j\neq i} P_j^{a_j} \ [1 - P_j]^{1-a_j}\right) \ h\left(1 + \sum_{j\neq i} a_j\right) \tag{3.17}$$

and $\sum_{a_{-i}}$ represents the sum over all the possible actions of firms other than i.

Pseudo ML estimation. The problem is to estimate the vector of structural parameters θ^0 given a random sample $\{x_{im}, a_{im}\}$. Equilibrium probabilities are not uniquely determined for some values of the primitives. However, for any vector of probabilities P, the best response probability functions $\Phi\left(x_{im} \ \beta_i - \sum_{j\neq i} \delta_{ij} \ P_j(x_m)\right)$ are always well-defined. We define a pseudo likelihood function based on best responses to the population probabilities.

$$\begin{aligned}
Q_M(\theta, P^0) &= \sum_{m=1}^{M} \sum_{i=1}^{N} a_{im} \ \ln \Phi\left(x_{im} \ \beta_i - \sum_{j\neq i} \delta_{ij} \ P_j^0(x_m)\right) \\
&+ \ (1 - a_{im}) \ \ln \Phi\left(-x_{im} \ \beta_i + \sum_{j\neq i} \delta_{ij} \ P_j^0(x_m)\right)
\end{aligned} \tag{3.18}$$

It is possible to show that θ^0 uniquely maximizes $Q_\infty(\theta, P^0)$. The PML estimator of θ^0 maximizes $Q_M(\theta, \hat{P}^0)$, where $\hat{P}^0$ is a consistent nonparametric estimator of P^0. This estimator is consistent and asymptotically normal. Iterating in this procedure can provide efficiency gains both in finite samples and asymptotically (Aguirregabiria, Economics Letters, 2004).

3.4.7. *Entry and spatial competition.* Based on Seim (RAND, 2006)

How do market power and profits of a retail firm depend on the location of its store(s) relative to the location of competitors? How important is spatial differentiation to explain market power? These are important questions in the study of competition in retail markets. Seim (2006) studies these questions in the context of the video rental industry. Seim's work is the first study that endogenizes store locations and introduces spatial competition in a game of market entry. Her model has important similarities with the static game with single-store firms and incomplete information that we have presented above in section 2.1.1. The main difference is that Seim's model does not include an explicit model of spatial consumer demand and price competition. Instead, she considers a 'semi-structural' specification of a store's profit that captures the idea that the profit of a store declines when competing stores get closer in geographic space. The specification seems consistent with the idea that consumers face transportation costs and therefore spatial differentiation between stores can increase profits.

From a geographical point of view, a market in this model is a compact set in the two-dimension Euclidean space. There are L locations in the market where firms can operate stores. These locations are exogenously given and they could be chosen as the set grid points where the grid can be as fine as we want. We index locations by ℓ that belongs to the set $\{1, 2, ..., L\}$.

There are N potential entrants in the market. Each firm takes two decisions: (1) whether to be active or not in the market; and (2) if it decides to be active, it chooses the location of its store. Note that Seim does not model multi-store firms. Aguirregabiria and Vicentini (2007) present an extension of Seim's model with multi-store firms, endogenous consumer behavior, and dynamics.

Let a_i represent the decisions of firm i, such that $a_i \in \{0, 1, ..., L\}$ and $a_i = 0$ represents "no entry", and $a_i = \ell > 0$ represents entry in location ℓ.

The profit of not being active in the market is normalized to zero. Let $\Pi_{i\ell}$ be the profit of firm i if it has a store in location ℓ. These profits depend on the store location decisions of the other firms. In particular, $\Pi_{i\ell}$ declines with the number of other stores "close to" location ℓ.

Of course, the specific meaning of being close to location ℓ is key for the implications of this model. This should depend on how consumers perceive as close substitutes stores in different locations. In principle, if we have data on quantities and prices for the different stores active in this city, we could estimate a demand system that would provide a measures of consumers' transportation costs and of the degree of substitution in demand between stores at different locations. That is what Jackie Wang did in his job market paper for the banking industry (Wang, 2010). However, for this industry we do not have information on prices and quantities at the store level, and even if we had, stores location decisions may contain useful (and even better) information to identify the degree of competition between stores at different locations.

Seim's specification of the profit function is "semi-structural" in the sense that it does not model explicitly consumer behavior,but it is consistent with the idea that consumers face transportation costs and therefore spatial differentiation (ceteris paribus) can increase profits.

****
FIGURE on definition of local markets
(Seim, RAND 2006)
****

For every location ℓ in the city, Seim defines B rings around the location. A first ring of radius d_1 (say half a mile); a second ring of radius $d_2 > d_1$ (say one mile), and so on. The profit of a store depends on the number of other stores located within each of the B rings. We expect that closer stores should have stronger negative effects on profits. The profit function of an active store at location ℓ is:

$$\Pi_{i\ell} = x_\ell \, \beta + \sum_{b=1}^{B} \gamma_b \, N_{b\ell} + \xi_\ell + \varepsilon_{i\ell}$$

where β, γ_1, γ_2, ..., and γ_B are parameters; x_ℓ is a vector of observable exogenous character-istics that affect profits in location ℓ; $N_{b\ell}$ is the number of stores in ring b around location ℓ excluding ; ξ_ℓ represents exogenous characteristics of location ℓ that are unobserved to the researcher but common and observable to firms; and $\varepsilon_{i\ell}$ is component of the profit of firm i in location ℓ that is private information to this firm. For the no entry choice, $\Pi_{i0} = \varepsilon_{i0}$.

ASSUMPTION: Let $\varepsilon_i = \{\varepsilon_{i\ell} : \ell = 0, 1, ..., L\}$ be the vector with the private information variables of firm i at every possible location. ε_i is i.i.d. over firms and locations with a extreme value type 1 distribution.

The information of firm i is (x, ξ, ε_i), where x and ξ represent the vectors with x_ℓ and ξ_ℓ, respectively, at every location in the city. Firm i does not know the $\varepsilon's$ of other firms. Therefore, $N_{b\ell}$ is unknown to a firm. Firms only know the probability distribution of $N_{b\ell}$. Therefore, firms maximize expected profits. The expected profit of firm i is:

$$\Pi_{i\ell}^e = x_\ell\ \beta + \sum_{b=1}^{B} \gamma_b\ N_{b\ell}^e + \xi_\ell + \varepsilon_{i\ell}$$

where $N_{b\ell}^e$ represents $E(N_{b\ell}|x,\xi)$.

A firm's strategy depends on the variables in his information set. Let $\alpha_i(x, \xi, \varepsilon_i)$ be a strategy function for firm i such that $\alpha_i : X \times \mathbb{R}^2 \to \{0, 1, ..., L\}$. Given expectations $N_{b\ell}^e$, the best response strategy of player i is:

$$\alpha_i(x, \xi, \varepsilon_i) = \arg\max_{\ell \in \{0,1,...,L\}} \left\{ x_\ell\ \beta + \sum_{b=1}^{B} \gamma_b\ N_{b\ell}^e + \xi_\ell + \varepsilon_{i\ell} \right\}$$

Or similarly, $\alpha_i(x, \xi, \varepsilon_i) = \ell$ if and only if $x_\ell\ \beta + \sum_{b=1}^{B} \gamma_b\ N_{b\ell}^e + \xi_\ell + \varepsilon_{i\ell}$ is greater that $x_{\ell'}\ \beta + \sum_{b=1}^{B} \gamma_b\ N_{b\ell'}^e + \xi_{\ell'} + \varepsilon_{i\ell'}$ for any other location ℓ.

From the point of view of other firms that do not know the private information of firm i but know the strategy function $\alpha_i(x, \xi, \varepsilon_i)$, the strategy of firm i can be described as a probability distribution: $P_i \equiv \{P_{i\ell} : \ell = 0, 1, ..., L\}$ where $P_{i\ell}$ is the probability that firm i chooses location ℓ when following his strategy $\alpha_i(x, \xi, \varepsilon_i)$. That is,

$$P_{i\ell} \equiv \int 1\{\alpha_i(x, \xi, \varepsilon_i) = \ell\}\ dF(\varepsilon_i)$$

where $F(\varepsilon_i)$ is the CDF of ε_i. By construction, $\sum_{\ell=0}^{L} P_{i\ell} = 1$.

Given expectations $N_{b\ell}^e$, we can also represent the best response strategy of firm i as a choice probability. A best response probability $P_{i\ell}$ is:

$$P_{i\ell} = \int 1 \left[\ell = \arg\max_{\ell'} \left\{ x_{\ell'}\ \beta + \sum_{b=1}^{B} \gamma_b\ N_{b\ell'}^e + \xi_{\ell'} + \varepsilon_{i\ell'} \right\} \right]\ dF(\varepsilon_i)$$

And given the extreme value assumption on ε_i:

$$P_{i\ell} = \frac{\exp\left\{x_\ell\beta + \sum_{b=1}^{B}\gamma_b\ N_{b\ell}^e + \xi_\ell\right\}}{1 + \exp\left\{x_{\ell'}\beta + \sum_{b=1}^{B}\gamma_b\ N_{b\ell'}^e + \xi_{\ell'}\right\}}$$

In this application, there is not information on firm exogenous characteristics, and Seim assumes that the equilibrium is symmetric: $\alpha_i(x, \xi, \varepsilon_i) = \alpha(x, \xi, \varepsilon_i)$ and $P_{i\ell} = P_\ell$ for every firm i.

The expected number of firms in ring b around location ℓ, $N_{b\ell}^e$, is determined by the vector of entry probabilities $P \equiv \{P_{\ell'} : \ell' = 1, 2, ..., L\}$. That is:

$$N_{b\ell}^e = \sum_{\ell'=1}^{L} 1\{\ell' \text{ belongs to ring } b \text{ around } \ell\}\ P_{\ell'}\ N$$

To emphasize this dependence we use the notation $N_{b\ell}^e(P)$.

Therefore, we can define a (symmetric) equilibrium in this game as a vector of probabilities $P \equiv \{P_\ell : \ell = 1, 2, ..., L\}$ that solve the following system of equilibrium conditions: for every $\ell = 1, 2, ..., L$:

$$P_\ell = \frac{\exp\left\{x_\ell\beta + \sum_{b=1}^{B}\gamma_b\ N_{b\ell}^e(P) + \xi_\ell\right\}}{1 + \exp\left\{x_{\ell'}\beta + \sum_{b=1}^{B}\gamma_b\ N_{b\ell'}^e(P) + \xi_{\ell'}\right\}}$$

By Brower's Theorem an equilibrium exist. The equilibrium may not be unique. Seim shows that if the γ parameters are not large and they decline fast enough with b, then the equilibrium is unique.

Let $\theta = \{N, \beta, \gamma_1, \gamma_2, ..., \gamma_B\}$ be the vector of parameters of the model. These parameters can be estimated even if we have data only from one city. Suppose that the data set is $\{x_\ell, n_\ell : \ell = 1, 2, ..., L\}$ for L different locations in a city, where L is large, and n_ℓ represents the number of stores in location ℓ. We want to use these data to estimate θ. I describe the estimation with data from only one city. Later, we will see that the extension to data from more than one city is trivial.

Let x be the vector$\{x_\ell : \ell = 1, 2, ..., L\}$. All the analysis is conditional on x, that is a description of the "landscape" of observable socioeconomic characteristics in the city. Given x, we can think in $\{n_\ell : \ell = 1, 2, ..., L\}$ as **one realization of a spatial stochastic process**. In terms of the econometric analysis, this has similarities with time series econometrics in the sense that a time series is a single realization from a stochastic process. Despite having just one realization of a stochastic process, we can estimate consistently the parameters of that process as long as we make some stationarity assumptions.

This is the model considered by Seim (2006): there is city unobserved heterogeneity (her dataset includes multiple cities) but within a city there is not unobserved location heterogeneity.

Conditional on x, spatial correlation/dependence in the unobservable variables ξ_ℓ can generate dependence between the number of firms at different locations $\{n_\ell\}$. We start with the simpler case where there is not the unobserved location heterogeneity: i.e., $\xi_\ell = 0$ for every location ℓ.

Without unobserved location heterogeneity, and conditional on x, the variables n_ℓ are independently distributed, and n_ℓ is a random draw from Binomial random variable with arguments $(N, P_\ell(x, \theta))$, where $P_\ell(x, \theta)$ are the equilibrium probabilities defined above where now I explicitly include (x, θ) as arguments.

$$n_\ell \sim i.i.d. \text{ over } \ell \text{ Binomial}(N, P_\ell(x, \theta))$$

Therefore,

$$\Pr\left(n_1, n_2, ..., n_L \mid x, \theta\right) = \prod_{\ell=1}^{L} \Pr\left(n_\ell \mid x, \theta\right)$$

$$= \prod_{\ell=1}^{L} \frac{N!}{n_\ell(N - n_\ell)!} P_\ell(x, \theta)^{n_\ell} (1 - P_\ell(x, \theta))^{N-n_\ell}$$

The log-likelihood function is:

$$l(\theta) = \sum_{\ell=1}^{L} \ln\left(\frac{N!}{(N - n_\ell)!}\right) + n_\ell \ln P_\ell(x, \theta) + (N - n_\ell)\ln(1 - P_\ell(x, \theta))$$

And the maximum likelihood estimator, $\hat{\theta}$, is the value of θ that maximizes this likelihood. Later, I will present and describe in detail several algorithms to obtain this MLE. The part of this estimation that is computationally more demanding is that the probabilities are the solution of a fixed point/equilibrium problem.

The parameters of the model, including the number of potential entrants N, are identified. Partly, the identification comes form functional form assumptions. However, there also exclusion restrictions that can provide identification even if some of these assumptions are relaxed. In particular, for the identification of β and γ_b, the model implies that $N_{b\ell}^e$ depends on socioeconomic characteristics at locations other than ℓ (i.e., $x_{\ell'}$ for $\ell' \neq \ell$). Therefore, $N_{b\ell}^e$ has sample variability that is independent of x_ℓ and this implies that the effects of x_ℓ and $N_{b\ell}^e$ on a firm's profit can be identified even if we relax the linearity assumption.

Haiqing Xu's job market paper (2010) (titled "Parametric and Semiparametric Structural Estimation of Hotelling-type Discrete Choice Games in A Single Market with An Increasing Number of Players") studies the asymptotics of this type of estimator. His model is a bit different to Seim's model because players and locations are the same thing.

Now, let's consider the model where $\xi_\ell \neq 0$. A simple (but restrictive approach) is to assume that there is a number R of "regions" or districts in the city, where the number of regions R is small relative to the number of locations L, such that all the unobserved heterogeneity is between regions but there is no unobserved heterogeneity within regions. Under this assumption, we can control for unobserved heterogeneity by including region dummies. In fact, this case is equivalent to the previous case without unobserved location heterogeneity with the only difference is that the vector of observables x_ℓ now includes region dummies.

A more interesting case is when the unobserved heterogeneity is at the location level. We assume that $\xi = \{\xi_\ell : \ell = 1, 2, ..., L\}$ is independent of x and it is a random draw from a spatial stochastic process. The simplest process is when ξ_ℓ is $i.i.d.$ with a known distribution, say $N(0, \sigma_\xi^2)$ where the zero mean is without loss of generality. However, we can allow for spatial dependence in this unobservable. For instance, we may consider a Spatial autorregressive process (SAR):

$$\xi_\ell = \rho \, \bar{\xi}_\ell^C + u_\ell$$

where u_ℓ is $i.i.d.$ $N(0, \sigma_u^2)$, ρ is a parameter, and $\bar{\xi}_\ell^C$ is the mean value of ξ at the C locations closest to location ℓ, excluding location ℓ itself. To obtain, a random draw of the vector ξ from this stochastic process it is convenient to write the process in vector form:

$$\xi = \rho \, \mathbf{W}^C \, \xi + u$$

where ξ and u are $L \times 1$ vectors, and $\mathbf{W}^C$ is a $L \times L$ weighting matrix such that every row, say row ℓ, has values $1/C$ at positions that correspond to locations close to location ℓ, and zeroes otherwise. Then, we can write $\xi = (I - \rho \, \mathbf{W}^C)^{-1}u$. First, we take independent draws from $N(0, \sigma_u^2)$ to generate the vector u, and then we pre-multiple that vector by $(I - \rho \, \mathbf{W}^C)^{-1}$ to obtain ξ.

Note that now the vector of structural parameters includes the parameters in the stochastic process of ξ, i.e., σ_u and ρ.

Now, conditional on x AND ξ, the variables n_ℓ are independently distributed, and n_ℓ is a random draw from Binomial random variable with arguments $(N, P_\ell(x, \xi, \theta))$, where $P_\ell(x, \xi, \theta)$ are the equilibrium probabilities. Importantly, for different values of ξ we have

different equilibrium probabilities. Then,

$$
\begin{aligned}
\Pr\left(n_1, n_2, ..., n_L \mid x, \theta\right) &= \int \Pr\left(n_1, n_2, ..., n_L \mid x, \xi, \theta\right)\ dG(\xi) \\[2mm]
&= \int \left[\prod_{\ell=1}^{L} \Pr\left(n_\ell \mid x, \xi, \theta\right)\right]\ dG(\xi) \\[2mm]
&= \prod_{\ell=1}^{L} \frac{N!}{n_\ell(N - n_\ell)!} \\
&\quad \int \left[\prod_{\ell=1}^{L} P_\ell(x, \xi, \theta)^{n_\ell}(1 - P_\ell(x, \xi, \theta))^{N-n_\ell}\right] dG(\xi)
\end{aligned}
$$

And the log-likelihood function is:

$$
\begin{aligned}
l(\theta) &= \sum_{\ell=1}^{L} \ln\left(\frac{N!}{(N - n_\ell)!}\right) \\[2mm]
&\quad + \ln\left(\int \left[\prod_{\ell=1}^{L} P_\ell(x, \xi, \theta)^{n_\ell}(1 - P_\ell(x, \xi, \theta))^{N-n_\ell}\right] dG(\xi)\right)
\end{aligned}
$$

And the maximum likelihood estimator is defined as usual.

In their empirical study on competition between big-box discount stores in US (that is, Kmart, Target and Walmart), Zhu and Singh (2009) extend Seim's entry model by introducing firm heterogeneity. The model allows competition effects to be asymmetric across three different chains. The model can incorporate a situation where, for example, the impact on the profit of Target of a Walmart store 10 miles away is stronger than the impact of a Kmart store located 5 miles away. The specification of the profit function of a store of chain i at location l is:

$$
\pi_{i\ell} = x_\ell\ \beta_i + \sum_{j \neq i} \sum_{b=1}^{B} \gamma_{bij}\ n_{b\ell j} + \xi_\ell + \varepsilon_{i\ell}
$$

where $n_{b\ell j}$ represents the number of stores that chain j has within the $b - ring$ around location ℓ. Despite the paper studies competition between retail chains, it still makes similar simplifying assumptions as in Seim's model that ignores important aspects of competition between retail chains. In particular, the model ignores economies of density, and firms' concerns on cannibalization between stores of the same chain. It assumes that the entry decisions of a retail chain are made independently at each location. Under these assumptions, the equilibrium of the model can be described as a vector of $N * L$ entry probabilities, one for each firm and location, that solves the following fixed point problem:

$$
P_{i\ell} = \frac{\exp\left\{x_\ell \beta_i + \sum_{j \neq i} \sum_{b=1}^{B} \gamma_{bij}\ N\left[\sum_{\ell'=1}^{L} D_{\ell\ell'}^b\ P_{j\ell'}\right] + \xi_\ell\right\}}{1 + \sum_{\ell'=1}^{L} \exp\left\{x_{\ell'} \beta_i + \sum_{j \neq i} \sum_{b=1}^{B} \gamma_{bij}\ N\left[\sum_{\ell''=1}^{L} D_{\ell'\ell''}^b\ P_{j\ell''}\right] + \xi_{\ell'}\right\}}
$$

The authors find substantial heterogeneity in the competition effects between these three big-box discount chains, and in the pattern of how these effects decline with distance. For instance, Walmart's supercenters have a very substantial impact even at large distance.

Datta and Sudhir (2013) estimate an entry model of grocery stores that endogenizes both location and product type decisions. Their main interests are the consequence of zoning on market structure. Zoning often reduces firms' ability to avoid competition by locating remotely each other. Theory suggests that in such a market firms have a stronger incentive to differentiate their products. Their estimation results support this theoretical prediction. The authors also investigate different impacts of various types of zoning ('centralized zoning', 'neighborhood zoning', and 'outskirt zoning') on equilibrium market structure.25

3.4.8. *Multi-store firms.* As we have mentioned above, economies of density and cannibalization are potentially important factors in store location decisions of retail chains. A realistic model of competition between retail chains should incorporate this type of spillover effects. Taking into account these effects requires a model of competition between multi-store firms similar to the one in section 2.1.2. The model takes into account the joint determination of a firm's entry decisions at different locations. A firm's entry decision is represented by the L-dimension vector $a_i \equiv \{a_{i\ell} : \ell = 1, 2, \ldots, L\}$, with $a_{i\ell} \in \{0,1\}$, such that the set of possible actions contains $2\hat{\ }L$ elements. For instance, Jia (2008) studies competition between two chains (Walmart and Kmart) over 2065 locations (US counties). The number of possible decisions of a retail chain is $2\hat{\ }2065$, which is larger than $10\hat{\ }621$. It is obvious that, without further restrictions, computing firms' best responses is intractable.

Jia (2008) proposes and estimates a game of entry between Kmart and Walmart over more than 2000 locations (counties). Her model imposes restrictions on the specification of firms' profits that imply the supermodularity of the game and facilitate substantially the computation of an equilibrium. Suppose that we index the two firms as i and j. The profit function of a firm, say i, is $\Pi_i = VP_i(a_i, a_j) - EC_i(a_i)$, where $VP_i(a_i, a_j)$ is the variable profit function such that:

$$VP_i(a_i, a_j) = \sum_{\ell=1}^{L} a_{i\ell} \left[x_\ell \, \beta_i + \gamma_{ij} \, a_{j\ell} \right]$$

x_ℓ is a vector of market/location characteristics. γ_{ij} is a parameter that represents the effect on the profit of firm i of competition from a store of chain j. $EC_i(a_i)$ is the entry cost function such that:

$$EC_i(a_i) = \sum_{\ell=1}^{L} a_{i\ell} \left[\theta_{i\ell}^{EC} - \frac{\theta^{ED}}{2} \sum_{\ell'=1}^{L} \frac{a_{i\ell'}}{d_{\ell\ell'}} \right]$$

$\theta_{i\ell}^{EC}$ is the entry cost that firm i would have in location l in the absence of economies of density (that is, if it were a single-store firm); θ^{ED} is a parameter that represents the magnitude of the economies of density and is assumed to be positive; and $d_{\ell\ell'}$ is the distance between locations ℓ and ℓ'. Jia further assumes that the entry cost $\theta_{i\ell}^{EC}$ consists of three parts: $\theta_{i\ell}^{EC} = \theta_i^{EC} + (1 - \rho) \, \xi_\ell + \varepsilon_{i\ell}$, where θ_i^{EC} is chain-fixed effects, ρ is a scale parameter, ξ_ℓ is a

location random effect, and $\varepsilon_{i\ell}$ is a firm-location error term. Both ξ_ℓ and $\varepsilon_{i\ell}$ are i.i.d. draws from the standard normal distribution and known to all the players when making decisions. To capture economies of density, the presence of the stores of the same firm at other locations is weighted by the inverse of the distance between locations, $1/d_{\ell\ell'}$. This term is multiplied by one-half to avoid double counting in the total entry cost of the retail chain.

The specification of the profit function in equations (9.29) and (9.30) imposes some important restrictions. Under this specification, locations are interdependent only through economies of density. In particular, there are no cannibalization effects between stores of the same chain at different locations. Similarly, there is no spatial competition between stores of different chains at different locations. In particular, this specification ignores the spatial competition effects between Kmart, Target, and Walmart that Zhu and Singh (2009) find in their study. The specification also rules out cost savings that do not depend on store density such as lower wholesale prices owing to strong bargaining power of chain stores. The main motivation for these restrictions is to have a supermodular game that facilitates very substantially the computation of an equilibrium, even when the model has a large number of locations.

In a Nash equilibrium of this model, the entry decisions of a firm, say i, should satisfy the following L optimality conditions:

$$a_{i\ell} = 1\left\{ x_\ell\,\beta_i + \gamma_{ij}\,a_{j\ell} - \theta_{i\ell}^{EC} + \frac{\theta^{ED}}{2}\sum_{\ell'=1}^{L}\frac{a_{i\ell'}}{d_{\ell\ell'}} \geq 0 \right\}$$

These conditions can be interpreted as the best response of firm i in location l given the other firm's entry decisions, and given also firm i's entry decisions at locations other than l. We can write this system of conditions in a vector form as $a_i = br_i(a_i, a_j)$. Given a_j, a fixed point of the mapping $br_i(., a_j)$ is a (full) best response of firm i to the choice a_j by firm j. With $\theta^{ED} > 0$ (that is, economies of density), it is clear from equation (9.31) that the mapping br_i is increasing in a_i. By Topkis's theorem, this increasing property implies that: (1) the mapping has at least one fixed point solution; (2) if it has multiple fixed points they are ordered from the lowest to the largest; and (3) the smallest (largest) fixed point can be obtained by successive iterations in the mapping br_i using as starting value $a_i = 0$ ($a_i = 1$). Given these properties, Jia shows that the following algorithm provides the Nash equilibrium that is most profitable for firm i:

Step [i]: Given the lowest possible value for $a_j = 0$, that is, $a_i = (0, 0, \dots 0)$, we apply successive iterations with respect to a_i in the fixed point mapping $br_i(., a_j = 0)$ starting at $a_i = (1, 1, \dots 1)$. These iterations converge to the largest best response of firm i, that we denote by $a_i^{(1)} = BR_i^{(High)}(0)$.

Step [j]: Given $a_i^{(1)}$, we apply successive iterations with respect to a_j in the fixed point mapping $br_j(., a_i^{(1)})$ starting at $a_j = 0$. These iterations converge to the lowest best response of firm j, that we denote by $a_j^{(1)} = BR_j^{(Low)}(a_i^{(1)})$.

We keep iterating in (Step [i]) and (Step [j]) until convergence.

At any iteration, say k, given $a_j^{(k-1)}$ we first apply (Step [i]) to obtain $a_i^{(k)} = BR_i^{(High)}(a_j^{(k-1)})$, and then we apply (Step [j]) to obtain $a_j^{(k)} = BR_j^{(Low)}(a_i^{(k)})$. The supermodularity of the game assures the convergence of this process and the resulting fixed point is the Nash equilibrium that most favors firm i. Jia combines this solution algorithm with a simulation of unobservables to estimate the parameters of the model using the method of simulated moments (MSM).

In his empirical study of convenience stores in Okinawa Island of Japan, Nishida (2015) extends Jia's model in two directions. First, a firm is allowed to open multiple stores (up to four) in the same location. Second, the model explicitly incorporates some form of spatial competition: a store's revenue is affected not only by other stores in the same location but also by those in adjacent locations.

Although the approach used in these two studies is elegant and useful, its use in other applications is somewhat limited. First, supermodularity requires that the own network effect on profits is monotonic, that is, the effect of *** is either always positive ($\blacksquare^{ED} > 0$) or always negative ($\blacksquare^{ED} < 0$). This condition rules out situations where the net effect of cannibalization and economies of density varies across markets. Second, the number of (strategic) players must be equal to two. For a game to be supermodular, players' strategies must be strategic complements. In a model of market entry, players' strategies are strategic substitutes. However, when the number of players is equal to two, any game of strategic substitutes can be transformed into one of strategic complements by changing the order of strategies of one player (for example, use zero for entry and one for no entry). This trick no longer works when we have more than two players.26

Ellickson et al. (2013, hereafter EHT) propose an alternative estimation strategy and apply it to data of US discount store chains. Their estimation method is based on a set of inequalities that arise from the best response condition of a Nash equilibrium. Taking its opponents' decisions as given, a chain's profit associated with its observed entry decision must be larger than the profit of any alternative entry decision. Ellickson et al. (2013) consider particular deviations that relocate one of the observed stores to another location. Let a_i^* be the observed vector of entry decisions of firm i, and suppose that in this observed vector the firm has a store in location ℓ but not in location ℓ'. Consider the alternative (hypothetical) choice $a_i^{\ell \to \ell'}$ that is equal to a_i^* except that the store in location ℓ is closed and relocated to location ℓ'. Revealed preference implies that $\pi_i(a_i^*) \geq \pi_i(a_i^{\ell \to \ell'})$. Ellickson

et al. (2013) further simplify this inequality by assuming that there are no economies of scope or density (for example, $\theta^{ED} = 0$), and that there are no firm-location-specific factors unobservable to the researcher, that is, $\varepsilon_{i\ell} = 0$. Under these two assumptions, the inequality above can be written as the profit difference between two locations:

$$[x_\ell - x_{\ell'}]\beta_i + \sum_{j \neq i} \gamma_{ij} \left[a_{j\ell}^* - a_{j\ell'}^* \right] + [\xi_\ell - \xi_{\ell'}] \geq 0$$

Now, consider another chain, say k, that has an observed choice a_k^* with a store in location ℓ' but not in location ℓ. For this chain, we consider the opposite (hypothetical) relocation decision that for firm i above: the store in location ℓ' is closed and a new store is open in location ℓ. For this chain, revealed preference implies that

$$[x_{\ell'} - x_\ell]\beta_k + \sum_{j \neq k} \gamma_{kj} \left[a_{j\ell}^* - a_{j\ell'}^* \right] + [\xi_{\ell'} - \xi_\ell] \geq 0$$

Summing up the inequalities for firms i and k, we generate an inequality that is free from location fixed effects ξ.

$$[x_{\ell'} - x_\ell] \; [\beta_i - \beta_k] + \sum_{j \neq i} \gamma_{ij} \left[a_{j\ell}^* - a_{j\ell'}^* \right] + \sum_{j \neq k} \gamma_{kj} \left[a_{j\ell}^* - a_{j\ell'}^* \right] \geq 0$$

Ellickson et al. (2013) construct a number of inequalities of this type and obtain estimates of the parameters of the model by using a smooth maximum score estimator (Manski 1975; Horowitz, 1992; Fox, 2010).

Unlike the lattice theory approach of Jia and Nishida, the approach applied by EHT can accommodate more than two players, allows the researcher to be agnostic about equilibrium selections, and is robust to the presence of unobserved market heterogeneity. Their model, however, rules out any explicit interdependence between stores in different locations, including spatial competition, cannibalization and economies of density. Although incorporating such inter-locational interdependencies does not seem to cause any fundamental estimation issue, doing so can be difficult in practice as it considerably increases the amount of computation. Another possible downside of this approach is the restriction it imposes on unobservables. The only type of structural errors that this model includes are the variables ξ_ℓ that are common for all firms. Therefore, to accommodate observations that are incompatible with inequalities in (9.33) above, the model requires non-structural errors, which may be interpreted as firms' optimization errors.

3.4.9. *Dynamics with single-store firms.* When the entry cost is partially sunk, firms' entry decisions depend on their incumbency status, and dynamic models become more relevant. The role of sunk entry costs in shaping market structure in an oligopoly industry was first empirically studied by Bresnahan and Reiss (1993). They estimate a two-period model using panel data of the number of dentists. Following recent developments in the

econometrics of dynamic games of oligopoly competition,27 several studies have estimated dynamic games of market entry-exit in different retail industries.

Aguirregabiria and Mira (2007) estimate dynamic games of market entry and exit for five different retail industries: restaurants, bookstores, gas stations, shoe shops, and fish shops. They use annual data from a census of Chilean firms created for tax purposes by the Chilean Internal Revenue Service during the period 1994–99. The estimated models show significant differences in fixed costs, entry costs, and competition effects across the five industries, and these three parameters provide a precise description of the observed differences in market structure and entry-exit rates between the five industries. Fixed operating costs are a very important component of total profits of a store in the five industries, and they range between 59 percent (in restaurants) to 85 percent (in bookstores) of the variable profit of a monopolist in a median market. Sunk entry costs are also significant in the five industries, and they range between 31 percent (in shoe shops) and 58 percent (in gas stations) of a monopolist variable profit in a median market. The estimates of the parameter that measures competition effect show that restaurants are the retailers with the smallest competition effects, that might explained by a higher degree of horizontal product differentiation in this industry.

Suzuki (2013) examines the consequence of tight land use regulation on market structure of hotels through its impacts on entry costs and fixed costs. He estimates a dynamic game of entry-exit of mid-scale hotels in Texas that incorporates detailed measures of land use regulation into cost functions of hotels. The estimated model shows that imposing stringent regulation increases costs considerably and has substantial effects on market structure and hotel profits. Consumers also incur a substantial part of the costs of regulation in the form of higher prices.

Dunne et al. (2013) estimate a dynamic game of entry and exit in the retail industries of dentists and chiropractors in the US, and use the estimated model to evaluate the effects on market structure of subsidies for entry in small geographic markets, that is, markets that were designated by the government as Health Professional Shortage Areas (HPSA). The authors compare the effects of this subsidy with those of a counterfactual subsidy on fixed costs, and they find that subsidies on entry costs are cheaper, or more effective for the same present value of the subsidy.

Yang (2014) extends the standard dynamic game of market entry-exit in a retail market by incorporating information spillovers from incumbent firms to potential entrants.28 In his model, a potential entrant does not know a market-specific component in the level of profitability of a market (for example, a component of demand or operating costs). Firms learn about this profitability only when they actually enter that market. In this context, observing incumbents stay in this market is a positive signal for potential entrants about the

quality of this market. Potential entrants use these signals to update their beliefs about the profitability of the market (that is, Bayesian updating). These information spillovers from incumbents may contribute to explain why we observe retail clusters in some geographic markets. Yang estimates his model using data from the fast food restaurant industry in Canada, which goes back to the initial conditions of this industry in Canada. He finds significant evidence supporting the hypothesis that learning from incumbents induces retailers to herd into markets where others have previously done well in, and to avoid markets where others have previously failed in.

3.4.10. *Dynamics and spatial competition between multi-store firms.* A structural empirical analysis of economies of density, cannibalization, or spatial entry deterrence in retail chains requires the specification and estimation of models that incorporate dynamics, multi-store firms, and spatial competition. Some recent papers present contributions on this research topic.

Holmes (2011) studies the temporal and spatial pattern of store expansion by Walmart during the period 1971–2005. He proposes and estimates a dynamic model of entry and store location by a multi-store firm similar to the one that we have described in section 2.1.3 above. The model incorporates economies of density and cannibalization between Walmart stores, though it does not model explicitly competition from other retailers or chains (for example, Kmart or Target), and therefore it abstracts from dynamic strategic considerations such as spatial entry deterrence. The model also abstracts from price variation and assumes that Walmart sets constant prices across all stores and over time. However, Holmes takes into account three different types of stores and plants in Walmart retail network: regular stores that sell only general merchandise; supercenters, that sell both general merchandise and food; and distribution centers, which are the warehouses in the network, and that have also two different types, that is, general and food distribution centers. The distinction between these types of stores and warehouses is particularly important to explain the evolution of Walmart retail network over time and space. In the model, every year Walmart decides the number and the geographic location of new regular stores, supercenters, and general and food distribution centers. Economies of density are channeled through the benefits of stores being close to distribution centers. The structural parameters of the model are estimated using the Moment Inequalities estimation method in Pakes et al. (forthcoming). More specifically, moment inequalities are constructed by comparing the present value of profits from Walmart's actual expansion decision with the present value from counterfactual expansion decisions which are slight deviations from the observed ones. Holmes finds that Walmart obtains large savings in distribution costs by having a dense store network.

Igami and Yang (forthcoming) study the trade-off between cannibalization and spatial pre-emption in the fast-food restaurant industry, for example, McDonalds, Burger King, and so on. Consider a chain store that has already opened its first store in a local market. Opening an additional store increases this chain's current and future variable profits by, first, attracting more consumers and, second, preventing its rivals' future entries (pre-emption). However, the magnitude of this increase could be marginal when the new store steals customers from its existing store (cannibalization). Whether opening a new store economically makes sense or not depends on the size of the entry cost. Igami and Yang estimate a dynamic structural model and find the quantitative importance of preemptive motives. However, they do not model explicitly spatial competition, or allow for multiple geographic locations within their broad definition of geographic market.

Schiraldi et al. (2013) study store location and spatial competition between UK supermarket chains. They propose and estimate a dynamic game similar to the one in Aguirregabiria and Vicentini (forthcoming) that we have described in section 2.1.3. A novel and interesting aspect of this application is that the authors incorporate the regulator's decision to approve or reject supermarkets' applications for opening a new store in a specific location. The estimation of the model exploits a very rich dataset from the U.K. supermarket industry on exact locations and dates of store openings/closings, applications for store opening, approval/rejection decisions by the regulator, as well as rich data of consumer choices and consumer locations. The estimated model is used to evaluate the welfare effects of factual and counterfactual decision rules by the regulator.

4. Estimation

The estimation of games of entry and spatial competition in retail markets should deal with some common issues in the econometrics of games and dynamic structural models. Here we do not try to present a detailed discussion of this econometric literature. Instead, we provide a brief description of the main issues, with an emphasis on aspects that are particularly relevant for empirical applications in retail industries.31

4.1. Multiple Equilibria. Entry models with heterogeneous firms often generate more than one equilibria for a given set of parameters. Multiple equilibria pose challenges to the researcher for two main reasons. First, standard maximum likelihood estimation no longer works because the likelihood of certain outcomes is not well defined without knowing the equilibrium selection mechanism. Second, without further assumptions, some predictions or counterfactual experiments using the estimated model are subject to an identification problem. These predictions depend on the type of equilibrium that is selected in a hypothetical scenario not included in the data.

Several approaches have been proposed to estimate an entry game with multiple equilibria. Which method works the best depends on assumptions imposed in the model, especially its information structure. In a game of complete information, there are at least four approaches. The simplest approach is to impose some particular equilibrium selection rule beforehand and estimate the model parameters under this rule. For instance, Jia (2008) estimates the model of competition between big-box chains using the equilibrium that is most preferable to K-mart. She also estimates the same model under alternative equilibrium selection rules to check for the robustness of some of her results. The second approach is to construct a likelihood function for some endogenous outcomes of the game that are common across all the equilibria. Bresnahan and Reiss (1991b) estimate their model by exploiting the fact that, in their model, the total number of entrants is unique in all the equilibria.

A third approach is to make use of inequalities that are robust to multiple equilibria. One example is the profit inequality approach of EHT, which we described in section 2.2.5. Another example is the method of moment inequality estimators proposed by Ciliberto and Tamer (2009). They characterize the lower and upper bounds of the probability of a certain outcome that are robust to any equilibrium selection rule. Estimation of structural parameters relies on the set of probability inequalities constructed from these bounds. In the first step, the researcher nonparametrically estimates the probabilities of equilibrium outcomes conditional on observables. The second step is to find a set of structural parameters such that the resulting probability inequalities are most consistent with the data. The application of Ciliberto and Tamer's approach to a spatial entry model may not be straightforward. In models of this class, the number of possible outcomes (that is, market structures) is often very large. For example, consider a local market consisting of ten sub-blocks. When two chains decide whether they enter into each of these sub-blocks, the total number of possible market structures is 1024 ($=2^{10}$). Such a large number of possible outcomes makes it difficult to implement this approach for two reasons. The first stage estimate is likely to be very imprecise even when a sample size is reasonably large. The second stage estimation can be computationally intensive because one needs to check, for a given set of parameters, whether each possible outcome meets the equilibrium conditions or not.

A fourth approach proposed by Bajari et al. (2010b) consists in the specification of a flexible equilibrium selection mechanism and in the joint estimation of the parameters in this mechanism and the structural parameters in firms' profit functions. Together with standard exclusion restrictions for the identification of games, the key specification and identification assumption in this paper is that the equilibrium selection function depends only on firms' profits.

In empirical games of incomplete information, the standard way to deal with multiple equilibria is to use a two-step estimation method (Aguirregabiria and Mira 2007; Bajari et al. 2010a).32 In the first step, the researcher estimates the probabilities of firms' entry conditional on market observables (called policy functions) in a nonparametric way, for example, a sieves estimator. The second step is to find a set of structural parameters that are most consistent with the observed data and these estimated policy functions. A key assumption for the consistency of this approach is that, in the data, two markets with the same observable characteristics do not select different types of equilibria, that is, same equilibria conditional on observables. Without this assumption, the recovered policy function in the first stage would be a weighted sum of firms' policies under different equilibria, making the second-stage estimates inconsistent. Several authors have recently proposed extensions of this method to allow for multiplicity of equilibria in the data for markets with the same observable characteristics.

4.1.1. *Identification and multiple equilibria.* Tamer (2003) showed that all the parameters of the previous entry model with $N = 2$ is (point) identified under standard exclusion restrictions, and that multiple equilibria do not play any role in this identification result. Tamer's result can be extended to any number N of players, as long as we have the appropriate exclusion restrictions.

More generally, equilibrium uniqueness is neither a necessary nor a sufficient condition for the identification of a model (Jovanovic, 1989). To see this, consider a model with vector of structural parameters $\theta \in \Theta$, and define the mapping $C(\theta)$ from the set of parameters Θ to the set of measurable predictions of the model. For instance, $C(\theta)$ may contain the probability distribution of players actions conditional on exogenous variables $\Pr(a_1, a_2, ..., a_N | x, \theta)$.

Multiple equilibria implies that the mapping $C(.)$ is a correspondence. A model is not point-identified if at the observed data (say $P^0 = \Pr(a_1, a_2, ..., a_N | x, \theta)$ for any vector of actions and $x's$) the inverse mapping C^{-1} is a correspondence. In general, C being a function (i.e., equilibrium uniqueness) is neither a necessary nor a sufficient condition for C^{-1} being a function (i.e., for point identification).

****

FIGURE ON MULTIPLE EQUILIBRIA AND IDENTIFICATION

Mapping and inverse mapping. four cases.

(Cooper, 2002)

****

To illustrate the identification of a game with multiple equilibria, we start with a simple binary choice game with identical players and where the equilibrium probability P is implicitly defined as the solution of the condition $P = \Phi\left(-1.8 + \theta\, P\right)$, where θ is a structural parameter, and $\Phi\left(.\right)$ is the CDF of the standard normal. Suppose that the true value θ_0 is 3.5. It is possible to verify that the set of equilibria associated with θ_0 is $C(\theta_0) = \{ P^{(A)}(\theta_0) = 0.054,\ P^{(B)}(\theta_0) = 0.551,\ \text{and}\ P^{(C)}(\theta_0) = 0.924\}$. The game has been played M times and we observe players' actions for each realization of the game $\{a_{im} : i, m\}$. Let P_0 be the population probability $\Pr(a_{im} = 1)$. Without further assumptions the probability P_0 can be estimated consistently from the data. For instance, a simple frequency estimator $\hat{P}_0 = (NM)^{-1} \sum_{i,m} a_{im}$ is a consistent estimator of P_0. Without further assumption, we do not know the relationship between population probability P_0 and the equilibrium probabilities in $C(\theta_0)$. If all the sample observations come from the same equilibrium, then P_0 should be one of the points in $C(\theta_0)$. However, if the observations come from different equilibria in $C(\theta_0)$, then P_0 is a mixture of the elements in $C(\theta_0)$. To obtain identification, we can assume that every observation in the sample comes from the same equilibrium. Under this condition, since P_0 is an equilibrium associated with θ_0, we know that $P_0 = \Phi\left(-1.8 + \theta_0\, P_0\right)$. Given that $\Phi(.)$ is an invertible function, we have that $\theta_0 = (\Phi^{-1}\left(P_0\right) + 1.8)/P_0$. Provided that P_0 is not zero, it is clear that θ_0 is point identified regardless the existence of multiple equilibria in the model.

4.2. Unobserved Market Heterogeneity. Some market characteristics affecting firms' profits may not be observable to the researcher. For example, consider local attractions that spur the demand for hotels in a particular geographic location. Observing and controlling for all the relevant attractions are often impossible to the researcher. This demand effect implies that markets with such attractions should have more hotels than those without such attractions but with equivalent observable characteristics. Therefore, without accounting for this type of unobservables, researchers may wrongly conclude that competition boosts profits, or underestimate the negative effect of competition on profits.

Unobserved market heterogeneity usually appears as an additive term (ω_ℓ) in the firm's profit function $(\pi_{i\ell})$ where ω_ℓ is a random effect from a distribution known up to some parameters.34 The most common assumption (for example, Seim 2006; Zhu and Singh 2009; Datta and Sudhir 2013) is that these unobservables are common across locations in the same local market (that is, $\omega_\ell = \omega$ for all ℓ). Under this assumption the magnitude of unobserved market heterogeneity matters whether the firm enters some location in this market but not which location. Orhun (2013) relaxes this assumption by allowing unobserved heterogeneity to vary across locations in the same market.

In a game of complete information, accommodating unobserved market heterogeneity does not require a fundamental change in the estimation process. In a game of incomplete information, however, unobserved market heterogeneity introduces an additional challenge. Consistency of the two-step method requires that the initial nonparametric estimator of firms' entry probabilities in the first step should account for the presence of unobserved market heterogeneity. A possible solution is to use a finite mixture model. In this model, every market's ω_ℓ is drawn from a distribution with finite support. Aguirregabiria and Mira (2007) show how to accommodate such market-specific unobservables into their nested pseudo likelihood (NPL) algorithm. Arcidiacono and Miller (2011) propose an expectation-maximization (EM) algorithm in a more general environment. An alternative way to deal with this problem is to use panel data with a reasonably long time horizon. In that way, we can incorporate market fixed effects as parameters to be estimated. This approach is popular when estimating a dynamic game (for example, Ryan 2012; Suzuki 2013).35 A necessary condition to implement this approach is that every market at least observes some entries during the sample period.36 Dropping markets with no entries from the sample may generate a selection bias.

4.3. Computation. The number of geographic locations, L, introduces two dimensionality problems in the computation of firms' best responses in games of entry with spatial competition. First, in a static game, a multi-store firm's set of possible actions includes all the possible spatial configurations of its store network. The number of alternatives in this set is equal to 2^L, and this number is extremely large even with modest values of L, such as a few hundred geographic locations. Without further assumptions, the computation of best responses becomes impractical. This is an important computational issue that has deterred some authors to account for multi-store retailers in their spatial competition models, for example, Seim (2006), or Zhu and Singh (2009), among many others. As we have described in section 2.2.5, two approaches that have been applied to deal with this issue are (1) to impose restrictions that guarantee supermodularity of the game (that is, only two players, no cannibalization effects), and (2) to avoid the exact computation of best responses and use instead inequality restrictions implied by these best responses.

Looking at the firms' decision problem as a sequential or dynamic problem helps also to deal with the dimensionality in the space of possible choices. In a given period of time (for example, year, quarter, or month), we typically observe that a retail chain makes small changes in its network of stores, that is, it opens a few new stores, or closes a few existing stores. Imposing these small changes as a restriction on the model implies a very dramatic reduction in the dimension of the action space such that the computation of best responses becomes practical, at least in a 'myopic' version of the sequential decision problem.

However, to fully take into account the sequential or dynamic nature of a firm's decision problem, we also need to acknowledge that firms are forward looking. In the firm's dynamic programming problem, the set of possible states is equal to all the possible spatial configurations of a store network, and it has 2^L elements. Therefore, by going from a static model to a dynamic forward-looking model, we have just 'moved' the dimensionality problem from the action space into the state space. Recent papers propose different approaches to deal with this dimensionality problem in the state space. Arcidiacono et al. (2013) present a continuous-time dynamic game of spatial competition in a retail industry and propose an estimation method of this model. The continuous-time assumption eliminates the curse of dimensionality associated to integration over the state space. Aguirregabiria and Vicentini (forthcoming) propose a method of spatial interpolation that exploits the information provided by the (indirect) variable profit function.

5. Further topics

Spillovers between different retail sectors. Existing applications of games of entry and spatial competition in retail markets concentrate on a single retail industry. However, there are also interesting spillover effects between different retail industries. Some of these spillovers are positive, for example, good restaurants can make a certain neighborhood more attractive for shopping. There are also negative spillovers effects through land prices, that is, retail sectors with high value per unit of space (for example, jewelry stores) are willing to pay higher land prices than supermarkets that have low markups and are intensive in the use of land. The consideration and measurement of these spillover effects are interesting in themselves, and they can help to explain the turnover and reallocation of industries in different parts of a city. Relatedly, endogenizing land prices would also open the possibility of using these models for the evaluation of specific public policies at the city level.

Richer datasets with store level information on prices, quantities, inventories. The identification and estimation of competition effects based mainly on data of store locations have been the rule more than the exception in this literature. This approach typically requires strong restrictions in the specification of demand and variable costs. The increasing availability of datasets with rich information on prices and quantities at product and store level should create a new generation of empirical games of entry and spatial competition that relax these restrictions. Also, data on store characteristics such as product assortments or inventories will enable the introduction of these important decisions as endogenous variables in empirical models of competition between retail stores.

Measuring spatial pre-emption. So far, all the empirical approaches to measure the effects of spatial pre-emption are based on the comparison of firms' actual entry with firms' behavior

in a counterfactual scenario characterized by a change in either (1) a structural parameter (for example, a store exit value), or (2) firms' beliefs (for example, a firm believes that other firms' entry decisions do not respond to this firm's entry behavior). These approaches suffer the serious limitation that they do not capture only the effect of pre-emption and are contaminated by other effects. The development of new approaches to measure the pure effect of pre-emption would be a methodological contribution with relevant implications in this literature.

Geography. Every local market is different in its shape and its road network. These differences may have important impacts on the resulting market structure. For example, the center of a local market may be a quite attractive location for retailers when all highways go through there. However, it may not be the case anymore when highways encircle the city center (for example, Beltway in Washington DC). These differences may affect retailers' location choices and the degree of competition in an equilibrium. The development of empirical models of competition in retail markets that incorporate, in a systematic way, these idiosyncratic geographic features will be an important contribution in this literature.

Bibliography

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

[2] Ackerberg, D., and Gowrisankaran, G. (2006): "Quantifying Equilibrium Network Externalities in the ACH Banking Industry", RAND Journal of Economics, 37, 738-761.

[3] Aguirregabiria, V., R. Clark and H. Wang (2013), 'Diversification of geographic risk in retail bank networks: evidence from bank expansion after the Riegle-Neal Act', working paper, University of Toronto.

[4] Aguirregabiria, V. and P. Mira (2013), 'Identification of games of incomplete information with multiple equilibria and common unobserved heterogeneity', working paper, University of Toronto.

[5] Bajari, P., H. Hong, and S. Ryan (2010): "Identification and Estimation of A Discrete Game of Complete Information," Econometrica, 78(5), 1529-68.

[6] Bajari, P., H. Hong, and d. Nekipelov (2010): "Econometrics for Game Theory," in Advances in Economics and Econometrics: Theory and Applications, Tenth World Congress (Econometric Society Monographs).

[7] Bajari, P., H. Hong, J. Krainer, and D. Nekipelov (2011): "Estimating Static Models of Strategic Interactions," Journal of Business and Economic Statistics, forthcoming.

[8] Berry, S. (1990): "Airport Presence as Product Differentiation," American Economic Review, Papers and Proceedings , 80, 394-399.

[9] Berry, S. (1992): "Estimation of a Model of Entry in the Airline Industry," Econometrica , 60, 889-917.

[10] Berry, S. and P. Reiss (2007): "Empirical Models of Entry and Market Structure," Chapter 29 in Handbook of Industrial Organization, vol. 3, Mark Armstrong and Robert Porter, eds. North-Holland Press.

[11] Berry, S. and E. Tamer (2007): "Identification in Models of Oligopoly Entry," in Advances in Economics and Econometrics: Theory and Applications, Ninth World Congress, vol. 2, R. Blundell, W.K. Newey and T. Persson, eds., Cambridge Univ. Press.

[12] Berry, S. and J. Waldfogel (2010): "Quality and Market Size," Journal of Industrial Economics, 58(1), 1-31.

[13] Bresnahan, T. and P. Reiss (1987): "Do Entry Conditions Vary across Markets? Brookings Papers on Economic Activity: Special Issue on Microeconomics, no. 3, Martin Baily and Clifford Winston, eds., 833-871.

[14] Bresnahan, T. and P. Reiss (1990): "Entry into Monopoly Markets," Review of Economic Studies, 57, 531-553.

[15] Bresnahan, T. and P. Reiss (1991): "Econometric Models of Discrete Games," Journal of Econometrics, 48, 57-81.

[16] Bresnahan, T. and P. Reiss (1991): "Entry and Competition in Concentrated Markets," Journal of Political Economy, 95, 977-1009.

[17] Bresnahan, T. and P. Reiss (1994): "Measuring the Importance of Sunk Costs," Annales D'Économie et de Statistique, 31, 183-217.

[18] Campbell, J. and H. Hopenhayn: (2005): "Market size matters," Journal of Industrial Economics ...

[19] Ciliberto, F. and E. Tamer (2009), 'Market structure and multiple equilibria in airline markets', Econometrica, 77 (6), 1791–828.

[20] Datta, S. and K. Sudhir (2013), 'Does reducing spatial differentiation increase product differentiation? Effects of zoning on retail entry and format variety', Quantitative Marketing and Economics, 11 (1), 83–116.

[21] De Paula, A. (2013), 'Econometric analysis of games with multiple equilibria,' Annual Reviews, 5 (1), 107–31.

[22] De Paula, A. and X. Tang (2012), 'Inference of signs of interaction effects in simultaneous games with incomplete information', Econometrica, 80 (1), 143–72.

[23] Draganska, M., M. Mazzeo and K. Seim. (2009), 'Beyond plain vanilla: modeling joint product assortment and pricing decisions', Quantitative Marketing & Economics, 7 (2), 105–46.

[24] Ellickson, P. (2007), 'Does Sutton apply to supermarkets?' RAND Journal of Economics, 38 (1), 43–59.

[25] Ellickson, P., S. Houghton and C. Timmins (2013), 'Estimating network economies in retail chains: a revealed preference approach', RAND Journal of Economics, 44 (2), 169–93.

[26] Ellickson, P. and S. Misra (2008), 'Supermarket pricing strategies', Marketing Science, 27 (5), 811–28.

[27] Ellickson, P. and S. Misra (2012), 'Enriching interactions: incorporating outcome data into static discrete games', Quantitative Marketing and Economics, 10 (1), 1–26.

[28] Gowrisankaran, G. and J. Krainer (2011), 'Entry and pricing in a differentiated products industry: evidence from the ATM market', RAND Journal of Economics, 42 (1), 1–22.

[29] Grieco, P. (2014), 'Discrete games with flexible information structures: an application to local grocery markets', RAND Journal of Economics, 45 (2), 303–40.

[30] Hotelling, H. (1929), 'Stability in competition', Economic Journal, 39 (153), 41–57.

[31] Jia, P. (2008), 'What happens when Wal-Mart comes to town: an empirical analysis of the discount retailing industry', Econometrica, 76 (6), 1263–316.

[32] Matsa, D. (2011), 'Competition and product quality in the supermarket industry', Quarterly Journal of Economics, 126 (3), 1539–91.

[33] Mazzeo, M.J. (2002), 'Product choice and oligopoly market structure', RAND Journal of Economics, 33 (2), 221–42.

[34] Nishida, M. (2015), 'Estimating a model of strategic network choice: the convenience-store industry in Okinawa', Marketing Science, 34 (1), 20-38.

[35] Orhun, A. Y. (2013), 'Spatial differentiation in the supermarket industry: the role of common information', Quantitative Marketing and Economics, 11 (1), 3–37.

[36] Pakes, A., J. Porter, K. Ho and J. Ishii (forthcoming), 'Moment inequalities and their application', Econometrica.

[37] Seim, K. (2006): "An Empirical Model of Firm Entry with Endogenous Product-Type Choices," RAND Journal of Economics 37(3).

[38] Sutton, J. (1991): "Sunk costs and market structure: Price competition, advertising, and the evolution of concentration." MIT press. Cambridge, MA, USA.

[39] Sweeting, A. (2006): "Coordination, Differentiation and the Timing of Radio Commercials, Journal of Economics and Management Strategy, 15(4), 909-942.

[40] Sweeting, A. (2009): "The Strategic Timing of Radio Commercials: An Empirical Analysis Using Multiple Equilibria," RAND Journal of Economics, 40(4), 710-742.

[41] Sweeting, A. (2010): "The Effects of Horizontal Mergers on Product Positioning: Evidence from the Music Radio Industry, RAND Journal of Economics, 41(2), 372-397.

[42] Toivanen, O. and M. Waterson (2005): "Market Structure and Entry: Where's the Beef?," RAND Journal of Economics, 36(3), 680-699.

[43] Vitorino, M. (2012), 'Empirical entry games with complementarities: an application to the shopping center industry', Journal of Marketing Research, 49 (2), 175–91.

[44] Watson, R. (2009), 'Product variety and competition in the retail market for eyeglasses', Journal of Industrial Economics, 57 (2), 217–51.

[45] Walrath, M. (2010): "Religion as an Industry: Estimating a Strategic Entry Model for Churches," manuscript. University of St. Thomas.

[46] Zhu T. and V. Singh (2009), 'Spatial competition with endogenous location choices: an application to discount retailing', Quantitative Marketing and Economics, 7 (1), 1–35.

CHAPTER 6

Dynamic Structural Models of Industrial Organization

1. Introduction

Dynamics in demand and/or supply can be important aspects of competition in oligopoly markets. In many markets demand is dynamic in the sense that (a) consumers current decisions affect their future utility, and (b) consumers' current decisions depend on expectations about the evolution of future prices (states). Some sources of dynamics in demand are consumer switching costs, habit formation, brand loyalty, learning, and storable or durable products. On the supply side, most firm investment decisions have implications on future profits. Some examples are market entry, investment in capacity, inventories, or equipment, or choice of product characteristics. Firms' production decisions have also dynamic implications if there is learning by doing. Similarly, the existence of menu costs, or other forms of price adjustment costs, imply that pricing decisions have dynamic effects.

Identifying the factors governing the dynamics is important to understanding competition and the evolution of market structure, and for the evaluation of public policy. To identify and understand these factors, we specify and estimate dynamic structural models of demand and supply in oligopoly industries. A dynamic structural model is a model of individual behavior where agents are forward looking and maximize expected intertemporal payoffs. The parameters are structural in the sense that they describe preferences and technological and institutional constraints. Under the *principle of revealed preference*, these parameters are estimated using longitudinal micro data on individuals' choices and outcomes over time.

I start with some examples and a brief discussion of applications of dynamic structural models of Industrial Organization. These examples illustrate why taking into account forward-looking behavior and dynamics in demand and supply is important for the empirical analysis of competition in oligopoly industries.

1.1. Example 1: Demand of storable goods. For a storable product, purchases in a given period (week, month) are not equal to consumption. When the price is low, consumers have incentives to buy a large amount to store the product and consume it in the future. When the price is high, or the household has a large inventory of the product, consumers do not buy an consume from his inventory. Dynamics arise because consumers' past purchases and consumption decisions impact their current inventory and therefore the

benefits of purchasing today. Furthermore, consumers expectations about future prices also impact the perceived trade-offs of buying today versus in the future.

What are the implications of ignoring consumer dynamic behavior when we estimate the demand of differentiated storable products? An important implication is that we can get serious biases in the estimates of price demand elasticities. In particular, we can interpret a short-run intertemporal substitution as a long-run substitution between brands (or stores). To illustrate this issue, it is useful to consider an specific example. The following figure presents weekly times series data of prices and sales of canned tuna in a supermarket store. The time series of prices is characterized by "High-Low" pricing, what is quite common in many supermarkets. The price fluctuates between a high regular price and a low promotion price. The promotion price is infrequent and last only few days, after which the price returns to its "regular" level. Sales of this storable product respond to this type of dynamics in prices. As we can see in figure 6.1, most sales are concentrated at the very few days with low prices. Apparently, the short-run response of sales to these temporary price reductions is very large: the typical discount of a sales promotion is between 10% and 20%, and the increase in sales are around 300%.

Figure 6.1: Price promotions and sales of a storable good

In a static demand model, this type of respond would suggest that the price elasticity of demand of the product is very large. In particular, with these data the estimation of a static demand model provides estimates of own-price elasticities greater than 8. The static model interprets the large response of sales to a price reduction in terms of consumers substitution between brands (and to some extend between supermarkets too). Based on this estimates of demand elasticities, our model of competition would imply that price-cost margins are very small and firms (both supermarkets and brand manufacturers) have very little market power. A large degree of substitution between brands implies that product differentiation is small and market power is low.

This interpretation that ignores dynamics in consumer purchasing decision can be seriously wrong. Most of the short-run response of sales to a temporary price reduction is not substitution between brands or stores but intertemporal substitution in households' purchases. The temporary price reduction induces consumers to buy for storage today and to buy less in the future. The long-run substitution effect is much smaller, and it is this long-run effect what is relevant to measure firms' market power.

In order to distinguish between short-run and long-run responses to price changes, we have to specify and estimate a dynamic model of demand of differentiated products. In this

type of models consumers are forward looking and take into account their expectations about future prices as well as storage costs.

1.2. Example 2: Demand of a new durable product. Melnikov (2000), Esteban and Shum (*RAND*, 2007), Carranza (2006), Gowrisankaran and Rysman (2009).

The price of new durable products typically declines over time during the months after the introduction of the product. Figure 6.2 illustrates this point for the case of *****. Different factors may explain this price decline, e.g., intertemporal price discrimination, increasing competition, exogenous cost decline, or endogenous cost decline due to learning by doing. As in the case of the "high-low" pricing of storable goods, explaining this pricing dynamics also requires one to take into account dynamics in supply. For the moment, we concentrate here in the demand. If consumers are forward looking, they expect the price will be lower in the future and this generates an incentive to wait and buying the good in the future.

<p align="center">Figure 6.2: Price decline of new durable products</p>

A static model that ignores dynamics in demand of durable goods can introduce two different type of biases in the estimates of the distribution consumers willingness to pay and therefore of demand. The first source of bias comes from the failure to recognize that each period the potential market size is changing. Each period the demand curve is changing because some high willingness-to-pay consumers have already bought the product and left the market. A second source of bias comes from ignoring consumer forward-looking behavior. In the static model, consumers willingness-to-pay can is contaminated by consumers' willingness to wait because the expectation of future lower prices.

To illustrate the first source of bias, consider a market with an initial mass of 100 consumers and a uniform distribution of willingness to pay over the the unit interval. To concentrate on the first source of bias, consider that consumers are myopic and buy the product if the price is below their willingness to pay. Once consumers buy the product they are out of the market forever. Time is discrete and indexed by $t \in \{1, 2, ...\}$. Every period t, the aggregate demand is $Q_t = H_t \Pr(v_t \geq P_t) = H_t [1 - F_t(P_t)]$, where Q_t and P_t are quantity and price, respectively, H_t is the mass of consumers who remain in market at period t, and F_t is the distribution function of willingness to pay for consumers who remain in the market at period t. Suppose that we observe a sequence of prices equal to $P_1 = 0.9$, $P_2 = 0.8$, $P_3 = 0.7$, etc. Given this price sequence, it is easy to show that the demand curve at period $t = 1$ is $Q_1 = 100(1 - P_1)$, at period $t = 2$ the demand is $Q_2 = 90(\frac{0.9 - P_2}{0.9}) = 100(0.9 - P_2)$, at period $t = 3$ it is $Q_3 = 80(\frac{0.8 - P_3}{0.8}) = 100(0.8 - P_3)$, and so on. Therefore, the sequence of quantities is constant over time: $Q_1 = Q_2 = Q_3 = ... = 10$. A static demand model lead the

researcher to conclude that consumers are not sensitive to price, since the same quantity is sold as prices decline. The estimate of the price elasticity would be zero. This example but it illustrates how ignoring dynamics in demand of durable goods can lead to serious biases in the estimates of the price sensitivity of demand.

1.3. Example 3: Product repositioning in differentiated product markets. A common assumption in many static (and dynamic) demand models of differentiated products is that product characteristics, other than prices, are exogenous. However, in many industries, product characteristics are very important strategic variables.

Ignoring the endogeneity of product characteristics has several implications. First, it can biases in the estimated demand parameters. A dynamic game that acknowledges the endogeneity of some product characteristics and exploits the dynamic structure of the model to generate valid moment conditions can deal with this problem.

A second important limitation of a static model of firm behavior is that it cannot recover the costs of repositioning product characteristics. As a result, the static model cannot address important empirical questions such as the effect of a merger on product repositioning. That is, the evaluation of the effects of a merger using a static model should assume that the product characteristics (other than prices) of the new merging firm would remain the same as before the merger. This is at odds both with the predictions of theoretical models and with informal empirical evidence. Theoretical models of horizontal mergers show that product repositioning is a potentially very important source of value for a merging firm, and informal empirical evidence shows that soon after a merger firms implement significant changes in their product portfolio.

Sweeting (2007) and Aguirregabiria and Ho (2009) are two examples of empirical applications that endogenize product attributes using a dynamic game of competition in a differentiated products industry. Sweeting estimates a dynamic game of oligopoly competition in the US commercial radio industry. The model endogenizes the choice of radio stations format (genre), and estimates product repositioning costs. Aguirregabiria and Ho (2009) propose and estimate a dynamic game of airline network competition where the number of direct connections that an airline has in an airport is an endogenous product characteristic.

1.4. Example 4: Dynamics of market structure. Ryan (2006) and Kasahara (JBES, 2010) provide excellent examples of how ignoring supply-side dynamics and firms' forward looking behavior can lead to misleading results.

Ryan (2006) studies the effects of the 1990 Amendments to the Clean Air Act on the US cement industry. This environmental regulation added new categories of regulated emissions, and introduced the requirement of an environmental certification that cement plants have to

pass before starting their operation. Ryan estimates a dynamic game of competition where the sources of dynamics are sunk entry costs and adjustment costs associated with changes in installed capacity. The estimated model shows that the new regulation had negligible effects on variable production costs but it increased significantly the sunk cost of opening a new cement plant. A static analysis, that ignores the effects of the policy on firms' entry-exit decisions, would conclude that the regulation had negligible effects on firms profits and consumer welfare. In contrast, the dynamic analysis shows that the increase in sunk-entry costs caused a reduction in the number of plants that in turn implied higher markups and a decline in consumer welfare.

Kasahara (2010) proposes and estimates a dynamic model of firm investment in equipment and it uses the model to evaluate the effect of an important increase in import tariffs in Chile during the 1980s. The increase in tariffs had a substantial effect of the price of imported equipment and it may have a significant effect on firms' investment. An important feature of this policy is that the government announced that it was a temporary increase and that tariffs would go back to their original levels after few years. Kasahara shows that the temporary aspect of this policy exacerbated its negative effects on firm investment. Given that firms anticipated the future decline in import tariffs and the price of capital, a significant fraction of firms decided not invest and waiting until the reduction of tariffs. This waiting and inaction would not appear if the policy change were perceived as permanent. Kasahara shows that the Chilean economy would have recovered faster from the economic crisis of 1982-83 if the increase in tariffs would have been perceived as permanent.

1.5. Example 5: Dynamics of prices in a retail market. The significant cross-sectional dispersion of prices is a well-known stylized fact in retail markets. Retailing firms selling the same product, and operating in the same (narrowly defined) geographic market and at the same period of time, do charge prices that differ by significant amounts, e.g., 10% price differentials or even larger. This empirical evidence has been well established for gas stations and supermarkets, among other retail industries. Interestingly, the price differentials between firms, and the ranking of firms in terms prices, have very low persistence over time. A gas station that charges a price 5% below the average in a given week may be charging a price 5% above the average the next week. Using a more graphical description we can say that a firm's price follows a cyclical pattern, and the price cycles of the different firms in the market are not synchronized. Understanding price dispersion and the dynamics of price dispersion is very important to understand not only competition and market power but also for the construction of price indexes.

Different explanations have been suggested to explain this empirical evidence. Some explanations have to do with dynamic pricing behavior or "state dependence" in prices.

For instance, an explanation is based on the relationship between firm inventory and optimal price. In many retail industries with storable products, we observe that firms' orders to suppliers are infrequent. For instance, for products such as laundry detergent, a supermarket ordering frequency can be lower than one order per month. A simple and plausible explanation of this infrequency is that there are fixed or lump-sum costs of making an order that do not depend on the size of the order, or at least they do not increase proportionally with the size of the order. Then, inventories follow a so called (S,s) cycle: the increase by a large amount up to a maximum when a place is order and then they decline slowly up a minimum value where a new order is placed. Given this dynamics of inventories, it is simple to show that optimal price of the firm should also follow a cycle. The price drops to a minimum when a new order is placed and then increases over time up to a maximum just before the next order when the price drops again. Aguirregabiria (*REStud*, 1999) shows this joint pattern of prices and inventories for many products in a supermarket chain. I show that this type of inventory-depedence price dynamics can explain more than 20% of the time series variability of prices in the data.

CHAPTER 7

Single-Agent Models of Firm Investment

0.6. Firm Investment at the Extensive Margin. Some important firm investment decisions are discrete or at the *extensive margin*, e.g., market entry and exit, machine replacement, or adoption of a new technology. Starting with the seminal work by Pakes (1986) on patent renewal and Rust (1987) on machine replacement, models and methods for dynamic discrete choice structural models have been applied to study these investment decisions.[1]

Let $a_{it} \in \mathcal{A} = \{0, 1, ..., J\}$ be the discrete variable that represents the investment decision of firm i at period t. The profit function is:

$$\Pi_{it} = p_{it} \ Y \left(a_{it}, k_{it}, \omega_{it}; \theta_y \right) - C \left(a_{it}, r_{it}; \theta_c \right) + \varepsilon_{it}(a_{it}) \tag{0.1}$$

p_{it} represents output price. The term $y_{it} = Y \left(a_{it}, k_{it}, \omega_{it}; \theta_y \right)$ is a production function that depends on investment, a_{it}, predetermined installed capital, k_{it}, productivity, ω_{it}, and the structural parameters θ_y. The term $C \left(a_{it}, r_{it}; \theta_c \right)$ captures the cost of investment, where r_{it} is the price of the new capital and θ_c is a vector of structural parameters. The vector of variables $\varepsilon_{it} = \{\varepsilon_{it}(a) : a \in \mathcal{A}\}$ represents a component of the investment cost that is unobservable to the researcher. These unobservables have mean zero and typically they are assumed *i.i.d.* across plants and over time. The capital stock k_{it} depreciates exogenously and increases when new investments are made according to the standard transition rule, $k_{it+1} = (1 - \lambda) \left(k_{it} + a_{it} \right)$. The state variables ω_{it} and r_{it} follow exogenous first order Markov processes.

Every period t, the manager observes the state variables $x_{it} \equiv (k_{it}, p_{it}, r_{it})$, ω_{it}, and ε_{it} and decides its investment in order to maximize expected and discounted profits $\mathbb{E}_t \left(\sum_{j=0}^{\infty} \delta^j \ \Pi_{i,t+j} \right)$, where $\delta \in (0, 1)$ is the discount factor. The optimal decision rule for investment is:

$$a_{it} = \arg \max_{a \in \mathcal{A}} \left\{ \ v(a, x_{it}, \omega_{it}) + \varepsilon_{it}(a) \ \right\}, \tag{0.2}$$

where $v(a, x_{it}, \omega_{it})$ is the *conditional choice value function*, that is the unique solution to a Bellman equation. Given the distribution of the unobservables $\{\omega_{it}, \varepsilon_{it}\}$, the observable exogenous state variables, and the vector of structural parameters $\theta = (\theta_y, \theta_c, \delta)$, this model implies a probability for the observed path of output and investment of a firm, $\{y_{it}, a_{it} : t = 1, 2, ..., T\}$. A standard approach to estimate the parameters of this model

---

[1]In this section, we review applications that abstract from dynamic oligopoly competition or assume explicitly that firms operate in either competitive or monopolistic markets.

is by (Conditional) Maximum Likelihood. Rust (1987) proposed the Nested Fixed Point algorithm (NFXP) for the computation of this estimator. Hotz and Miller (1993) propose a two-step *Conditional Choice Probabilities* (CCP) estimator that avoids computing a solution of the dynamic programming problem. Aguirregabiria and Mira (2002) propose the Nested Pseudo Likelihood algorithm (NPL) that is a recursive extension of the CCP method that returns the maximum likelihood estimates at a substantially lower computing time than NFXP.

Das (1992) studies the decision to operate, hold idle, or retire a kiln by plants in the U.S. cement industry. Kennet (1994) analyzes airlines' replacement decisions of aircraft engines and identifies significant changes in the decision rule after the deregulation of the US airline industry in 1978. Roberts and Tybout (1997) investigate why the decision to export by Colombian manufacturing plants is very persistent over time. The authors disentangle the effects of sunk costs, prior exporting experience, and serially correlated unobserved heterogeneity. Kasahara (2009) studies the effect of import tariffs on capital investment decisions by Chilean manufacturing plants. He shows that the temporary feature of a tariff increase in the mid-1980s exacerbated firms' zero-investment response. Rota (2004) and Aguirregabiria and Alonso-Borrego (2014) estimate dynamic discrete choice models of labor demand and use them to measure the magnitude of labor adjustment costs for permanent works in Italy and Spain, respectively, and the effects of labor market reforms.

Holmes (2011) studies the geographic expansion by Wal-Mart stores during the period 1971-2005. He estimates a dynamic model of entry and store location by a multi-store firm that incorporates economies of density and cannibalization between Wal-Mart stores, though it abstracts from competition of other retail chains and therefore from preemptive motives. In the model, every year Wal-Mart decides the number and the geographic location of new regular stores, supercenters, and general and food distribution centers. Holmes finds that Wal-Mart obtains large savings in distribution costs by having a dense store network.

1. Model and Assumptions

To present some common features of dynamic structural models, we start with a simple model of firm investment that we can represent as a machine replacement model.

Suppose that we have panel data of N plants operating in the same industry with information on output, investment, and capital stock over T periods of time.

$$\text{Data} = \{\ Y_{it},\ I_{it},\ K_{it} : i = 1, 2, ..., N \text{ and } t = 1, 2, ..., T\ \}$$

Suppose that the investment data is characterized by infrequent and lumpy investments. That is, I_{it} contains a large proportion of zeroes (no investment), and when investment is positive the investment-to-capital ratio I_{it}/K_{it} is quite large. For instance, for some industries

and samples we can find that the proportion of zeroes is above 60% (even with annual data!) and the average investment-to-capital ratio conditional on positive investment is above 50%.

A possible explanation for this type of dynamics in firms' investment is that there are significant indivisibilities in the purchases of new capital, or/and fixed or lump-sum costs associated with purchasing and installing new capital. Machine replacement models are models of investment that emphasize the existence of these indivisibilities and lump-sum costs of investment.

This type of investment models have been applied before in papers by Rust (Ectca, 1987), Das (REStud, 1991), Kennet (RAND, 1994), Rust and Rothwell (JAE, 1995), Cooper, Halti-wanger and Power (AER, 1999), Cooper and Haltiwanger (REStud 2006), and Kasahara (JBES, 2010), among others. In Rust (1987) the firm is a bus company (in Madison, Wisconsin), a plant is a bus, and a machine is a bus engine. Das (1991) considers cement firms and a plant is a cement kiln. In Kennet (1994) studies airline companies and the machine is an aircraft engine. Rust and Rothwell (1995) consider nuclear power plants. Cooper, Halti-wanger and Power (1999), Cooper and Haltiwanger (2006), and Kasahara (2010) consider manufacturing firms and investment in equipment in general.

We index plants by i and time by t. A plant's profit function is:

$$\Pi_{it} = Y_{it} - C_t \ I_{it} - RC_{it}$$

Y_{it} is the revenue of market value of the output produced by plant i at period t. I_{it} is the amount of investment at period t. C_t is the price of new capital. And RC_{it} represents investment costs other than the cost of purchasing the new capital, i.e., costs of replacing the old equipment (machine) by the new equipment.

Let K_{it} be the capital stock of plant i at the beginning of period t. As usual, capital depreciates exogenously and it increases when new investments are made. This transition rule of the capital stock is:

$$K_{it+1} = (1 - \delta) \ (K_{it} + I_{it})$$

Following the key feature in models of machine replacement, we assume that there is an indivisibility in the investment decision. In the standard machine replacement model, the firm decides between zero investment ($I_{it} = 0$) or the replacement of the old capital by a "new machine" that implies a fixed amount of capital K^*. Therefore,

$$I_{it} \in \left\{ 0 \ , \ K^* - K_{it} \ \right\}$$

Therefore,

$$K_{it+1} = \begin{cases} (1 - \delta) \ K_{it} & \text{if} \ \ I_{it} = 0 \\[2mm] (1 - \delta) \ K^* & \text{if} \ \ I_{it} > 0 \end{cases}$$

or

$$K_{it+1} = (1 - \delta) \; [(1 - a_{it}) \; K_{it} + a_{it} \; K^*]$$

where a_{it} is the indicator of positive investment, i.e., $a_{it} \equiv 1\{I_{it} > 0\}$.

This implies that the possible values of the capital stock are $(1 - \delta)K^*$, $(1 - \delta)^2 K^*$, etc. Let T_{it} be the number of periods since the last machine replacement, i.e., time duration since the last time that investment was positive. There is a one-to-one relationship between capital K_{it} and the time duration T_{it}:

$$K_{it} = (1 - \delta)^{T_{it}} \; K^*$$

or in logarithms, $k_{it} = k^* - d \; T_{it}$, where $k^* \equiv \log K^*$ and $d \equiv -\log(1 - \delta) > 0$.

These assumptions on the values of investment and capital seem natural in applications where the investment decision is actually a machine replacement decision, as in the papers by Rust (1987), Das (1991), Kennet (1994), or Rust and Rothwell (1995), among others. However, this framework may be restrictive when we look at less specific investment decisions, such as investment in equipment as in the papers by Cooper, Haltiwanger and Power (1999), Cooper and Haltiwanger (2006), and Kasahara (2010). In these other papers, investment in the data is very lumpy, which is a prediction of a model of machine replacement, but firms in the sample have very different sizes (average over long periods of time) and their capital stocks in those periods with positive investment are very different. These papers consider that investment is either zero or a constant proportion of the installed capital, i.e., $I_{it} \in \{0 \; , \; q \; K_{it}\}$ where q is a constant, e.g., $q = 25\%$. Here I maintained the most standard assumption of machine replacement models.

The production function (actually, revenue function) is:

$$Y_{it} = \exp\left\{ \alpha_0 + \eta_i^Y \right\} \; [(1 - a_{it}) \; K_{it} + a_{it} \; K^*]^{\alpha_1}$$

where α_0 and α_1 are parameters, and η_i^Y captures productivity differences between firms that are time-invariant. The specification of the replacement cost function is:

$$RC_{it} = a_{it} \; (\; r(K_{it}) + \eta_i^C + \varepsilon_{it} \;)$$

$r(K)$ is a function that is increasing in K, and η_i^C and ε_{it} are zero mean random variables that captures firm heterogeneity in replacement costs. Therefore, the profit function is:

$$\Pi_{it} = \begin{cases} \exp\left\{ \alpha_0 + \eta_i^Y \right\} \; K_{it}^{\alpha_1} & \text{if } \; a_{it} = 0 \\[2ex] \exp\left\{ \alpha_0 + \eta_i^Y \right\} \; K^{*\alpha_1} - C_t \; I^* - r(K_{it}) - \eta_i^C - \varepsilon_{it} & \text{if } \; a_{it} = 1 \end{cases}$$

Every period t, the firm observes the state variables K_{it}, C_t, and ε_{it} and then it decides its investment in order to maximize its expected value:

$$E_t \left(\sum_{j=0}^{\infty} \beta^j \; \Pi_{i,t+j} \right)$$

where $\beta \in (0,1)$ is the discount factor. The main trade-off in this machine replacement decision is simple. On the one hand, the productivity/efficiency of a machine declines over time and therefore the firm prefers younger machines. However, using younger machines requires frequent replacement and replacing a machine is costly.

The firm has uncertainty about future realizations of C_t and ε_{it}. To complete the model we have to specify the stochastic processes of these variables. We assume that C_t follows a Markov process with transition probability $f_C(C_{t+1}|C_t)$. For the shock in replacement costs ε_{it} we consider that it is i.i.d. with a logistic distribution with dispersion parameter σ_ε. The individual effects (η_i^Y, η_i^C) have a finite mixture distribution, i.e., (η_i^Y, η_i^C) is a pair of random variables from a distribution with discrete and finite support F_η.

Let $S_{it} = (K_{it}, C_t, \varepsilon_{it})$ be the vector of state variables in the decision problem of a plant and let $V_i(S_{it})$ be the value function. This value function is the solution to the **Bellman equation**:

$$V_i(S_{it}) = \max_{a_{it} \in \{0,1\}} \left\{ \Pi_i(a_{it}, S_{it}) + \beta \int V_i(S_{it+1})\, f_S(S_{it+1}|a_{it}, S_{it})\, dS_{it+1} \right\}$$

where $f_S(S_{it+1}|a_{it}, S_{it})$ is the (conditional choice) transition probability of the state variables:

$$f_S(S_{it+1}|a_{it}, S_{it}) = 1\{K_{it+1} = (1 - \delta)\ [(1 - a_{it})\ K_{it} + a_{it}\ K^*]\}\ f_C(C_{t+1}|C_t)\ f_\varepsilon(\varepsilon_{it})$$

where $1\{.\}$ is the indicator function, and f_ε is the density function of ε_{it}.

We can also represent the Bellman equation as:

$$V_i(S_{it}) = \max\{ v_i(0; K_{it}, C_t)\ ;\ v_i(1; K_{it}, C_t) - \varepsilon_{it}\ \}$$

where $v_i(0; K_{it}, C_t)$ and $v_i(1; K_{it}, C_t)$ are the choice-specific value functions:

$$v_i(0; K_{it}, C_t) \equiv \exp\{\alpha_0 + \eta_i^Y\}\ K_{it}^{\alpha_1} + \beta \int V_i((1 - \delta)K_{it}, C_{t+1}, \varepsilon_{it+1})\ f_C(C_{t+1}|C_t)\ df_\varepsilon(\varepsilon_{it})$$

$$v_i(1; K_{it}, C_t) \equiv \begin{array}{l} \exp\{\alpha_0 + \eta_i^Y\}\ K_{it}^{\alpha_1} - C_t\ I^* - r(K_{it}) - \eta_i^C \\ + \beta \int V_i((1 - \delta)K^*, C_{t+1}, \varepsilon_{it+1})\ f_C(C_{t+1}|C_t)\ df_\varepsilon(\varepsilon_{it}) \end{array}$$

2. Solving the dynamic programming (DP) problem

For given values of structural parameters and functions, $\{\alpha_0,\ \alpha_1,\ r(.),\ f_C(.),\ \sigma_\varepsilon\}$, and of the individual effects η_i^Y and η_i^C, we can solve the DP problem of firm i by simply using successive approximations to the value function, i.e., iterations in the Bellman equation.

In models where some of the state variables are not serially correlated, it is computationally very convenient (and also convenient for the estimation of the model) to define versions of the value function and the Bellman equation that are integrated over the non-serially

correlated variables. In our model, ε is not serially correlated state variables. The *integrated value function* of firm i is:

$$\bar{V}_i(K_{it}, C_t) \equiv \int V_i(K_{it}, C_t, \varepsilon_{it}) \, df_\varepsilon(\varepsilon_{it})$$

And the integrated Bellman equation is:

$$\bar{V}_i(K_{it}, C_t) = \int \max \left\{ \, v_i(0; K_{it}, C_t) \, ; \, v_i(1; K_{it}, C_t) - \varepsilon_{it} \, \right\} df_\varepsilon(\varepsilon_{it})$$

The main advantage of using the integrated value function is that it has a lower dimensionality than the original value function.

Given the extreme value distribution of ε_{it}, the integrated Bellman equation is:

$$\bar{V}_i(K_{it}, C_t) = \sigma_\varepsilon \, \ln \left[\exp \left\{ \frac{v_i(0; K_{it}, C_t)}{\sigma_\varepsilon} \right\} + \exp \left\{ \frac{v_i(1; K_{it}, C_t)}{\sigma_\varepsilon} \right\} \right]$$

where

$$v_i(0; K_{it}, C_t) \equiv \exp \left\{ \alpha_0 + \eta_i^Y \right\} \, K_{it}^{\alpha_1} + \beta \int \bar{V}_i((1-\delta)K_{it}, C_{t+1}) \, f_C(C_{t+1}|C_t)$$

$$v_i(1; K_{it}, C_t) \equiv \exp \left\{ \alpha_0 + \eta_i^Y \right\} \, K_{it}^{\alpha_1} - C_t \, I^* - r(K_{it}) - \eta_i^C + \beta \int \bar{V}_i((1-\delta)K^*, C_{t+1}) \, f_C(C_{t+1}|C_t)$$

The optimal decision rule of this dynamic programming (DP) problem is:

$$a_{it} = 1 \left\{ \, \varepsilon_{it} \leq v_i(1; K_{it}, C_t) - v_i(0; K_{it}, C_t) \, \right\}$$

Suppose that the price of new capital, C_t, has a discrete a finite range of variation: $C_t \in \{c^1, c^2, ..., c^L\}$. Then, the value function $\bar{V}_i$ can be represented as a $M \times 1$ vector in the Euclidean space, where $M = T * L$ and the T is the number of possible values for the capital stock. Let $\mathbf{V}_i$ be that vector. The integrated Bellman equation in matrix form is:

$$\mathbf{V}_i = \sigma_\varepsilon \ln \left(\exp \left\{ \frac{\mathbf{\Pi}_i(0) + \beta \, \mathbf{F}(0) \, \mathbf{V}_i}{\sigma_\varepsilon} \right\} + \exp \left\{ \frac{\mathbf{\Pi}_i(1) + \beta \, \mathbf{F}(1) \, \mathbf{V}_i}{\sigma_\varepsilon} \right\} \right)$$

where $\mathbf{\Pi}_i(0)$ and $\mathbf{\Pi}_i(1)$ are the $M \times 1$ vectors of one-period profits when $a_{it} = 0$ and $a_{it} = 1$, respectively. $\mathbf{F}(0)$ and $\mathbf{F}(0)$ are $M \times M$ transition probability matrices of (K_{it}, C_t) conditional on $a_{it} = 0$ and $a_{it} = 1$, respectively.

Given this equation, the vector $\mathbf{V}_i$ can be obtained by using value function iterations in the Bellman equation. Let $\mathbf{V}_i^0$ be an arbitrary initial value for the vector $\mathbf{V}_i$. For instance, $\mathbf{V}_i^0$ could be a $M \times 1$ vector of zeroes. Then, at iteration $k = 1, 2, ...$ we obtain:

$$\mathbf{V}_i^k = \sigma_\varepsilon \ln \left(\exp \left\{ \frac{\mathbf{\Pi}_i(0) + \beta \, \mathbf{F}(0) \, \mathbf{V}_i^{k-1}}{\sigma_\varepsilon} \right\} + \exp \left\{ \frac{\mathbf{\Pi}_i(1) + \beta \, \mathbf{F}(1) \, \mathbf{V}_i^{k-1}}{\sigma_\varepsilon} \right\} \right)$$

Since the (integrated) Bellman equation is a contraction mapping, this algorithm always converges (regardless the initial $\mathbf{V}_i^0$) and it converges to the unique fixed point. Exact convergence requires infinite iterations. Therefore, we stop the algorithm when the distance

(e.g., Euclidean distance) between $\mathbf{V}_i^k$ and $\mathbf{V}_i^{k-1}$ is smaller than some small constant, e.g., 10^{-6}.

An alternative algorithm to solve the DP problem is the **Policy Iteration algorithm**. Define the *Conditional Choice Probability (CCP) function* $P_i(K_{it}, C_t)$ as:

$$P_i(K_{it}, C_t) \quad \equiv \quad \Pr\left(\varepsilon_{it} \leq v_i(1; K_{it}, C_t) - v_i(0; K_{it}, C_t) \right)$$

$$= \quad \frac{\exp\left\{ \dfrac{v_i(1; K_{it}, C_t) - v_i(0; K_{it}, C_t)}{\sigma_\varepsilon} \right\}}{1 + \exp\left\{ \dfrac{v_i(1; K_{it}, C_t) - v_i(0; K_{it}, C_t)}{\sigma_\varepsilon} \right\}}$$

Given that (K_{it}, C_t) are discrete variables, we can describe the CCP function $P_i(.)$ as a $M \times 1$ vector of probabilities $\mathbf{P}_i$. The expression for the CCP in vector form is:

$$\mathbf{P}_i \quad = \quad \frac{\exp\left\{ \dfrac{\mathbf{\Pi}_i(1) - \mathbf{\Pi}_i(0) + \beta\left[\mathbf{F}(1) - \mathbf{F}(0)\right]\mathbf{V}_i}{\sigma_\varepsilon} \right\}}{1 + \exp\left\{ \dfrac{\mathbf{\Pi}_i(1) - \mathbf{\Pi}_i(0) + \beta\left[\mathbf{F}(1) - \mathbf{F}(0)\right]\mathbf{V}_i}{\sigma_\varepsilon} \right\}}$$

Suppose that the firm behaves according to the probs in $\mathbf{P}_i$. Let $\mathbf{V}_i^{\mathbf{P}}$ the vector of values if the firm behaves according to $\mathbf{P}$. That is $\mathbf{V}_i^{\mathbf{P}}$ is the expected discounted sum of current and future profits if the firm behaves according to $\mathbf{P}_i$. Ignoring for the moment the expected future $\varepsilon's$, we have that:

$$\mathbf{V}_i^{\mathbf{P}} = (1 - \mathbf{P}_i) * \left[\mathbf{\Pi}_i(0) + \beta\,\mathbf{F}(0)\mathbf{V}_i^{\mathbf{P}}\right] + \mathbf{P}_i * \left[\mathbf{\Pi}_i(1) + \beta\,\mathbf{F}(1)\mathbf{V}_i^{\mathbf{P}}\right]$$

And solving for $\mathbf{V}_i^{\mathbf{P}}$:

$$\mathbf{V}_i^{\mathbf{P}} = \left(I - \beta\,\mathbf{F}_i^{\mathbf{P}}\right)^{-1}\left((1 - \mathbf{P}_i) * \mathbf{\Pi}_i(0) + \mathbf{P}_i * \mathbf{\Pi}_i(1)\right)$$

where $\mathbf{F}_i^{\mathbf{P}} = (1 - \mathbf{P}_i) * \mathbf{F}(0) + \mathbf{P}_i * \mathbf{F}(1)$.

Taking into account this expression for $\mathbf{V}_i^{\mathbf{P}}$, we have that the optimal CCP $\mathbf{P}_i$ is such that:

$$\mathbf{P}_i \quad = \quad \frac{\exp\left\{ \dfrac{\tilde{\mathbf{\Pi}}_i + \beta\,\tilde{\mathbf{F}}\left(I - \beta\,\mathbf{F}_i^{\mathbf{P}}\right)^{-1}\left((1 - \mathbf{P}_i) * \mathbf{\Pi}_i(0) + \mathbf{P}_i * \mathbf{\Pi}_i(1)\right)}{\sigma_\varepsilon} \right\}}{1 + \exp\left\{ \dfrac{\tilde{\mathbf{\Pi}}_i + \beta\,\tilde{\mathbf{F}}\left(I - \beta\,\mathbf{F}_i^{\mathbf{P}}\right)^{-1}\left((1 - \mathbf{P}_i) * \mathbf{\Pi}_i(0) + \mathbf{P}_i * \mathbf{\Pi}_i(1)\right)}{\sigma_\varepsilon} \right\}}$$

where $\tilde{\mathbf{\Pi}}_i \equiv \mathbf{\Pi}_i(1) - \mathbf{\Pi}_i(0)$, and $\tilde{\mathbf{F}} \equiv \mathbf{F}(1) - \mathbf{F}(0)$. This equation defines a fixed point mapping in $\mathbf{P}_i$. This fixed point mapping is called the Policy Iteration mapping. This is also a contraction mapping. Optimal $\mathbf{P}_i$ is its unique fixed point.

Therefore we compute $\mathbf{P}_i$ by iterating in this mapping. Let $\mathbf{P}_i^0$ be an arbitrary initial value for the vector $\mathbf{P}_i$. For instance, $\mathbf{P}_i^0$ could be a $M \times 1$ vector of zeroes. Then, at each iteration $k = 1, 2, ...$ we do "two things":

Valuation step:

$$\mathbf{V}_i^k = \left(I - \beta \ \mathbf{F}^{\mathbf{P}_i^{k-1}} \right)^{-1} \left((1 - \mathbf{P}_i^{k-1}) * \mathbf{\Pi}_i(0) + \mathbf{P}_i^{k-1} * \mathbf{\Pi}_i(1) \right)$$

Policy step:

$$\mathbf{P}_i^k = \frac{\exp \left\{ \dfrac{\tilde{\mathbf{\Pi}}_i + \beta \ \tilde{\mathbf{F}} \ \mathbf{V}_i^k}{\sigma_\varepsilon} \right\}}{1 + \exp \left\{ \dfrac{\tilde{\mathbf{\Pi}}_i + \beta \ \tilde{\mathbf{F}} \ \mathbf{V}_i^k}{\sigma_\varepsilon} \right\}}$$

Policy iterations are more costly than Value function iterations (especially because the matrix inversion in the valuation step). However, the policy iteration algorithm requires a much lower number of iterations, especially with β is close to one. Rust (1987, 1994) proposes an hybrid algorithm: start with a few value function iterations and then switch to policy iterations.

3. Estimation

The primitives of the model are: (a) The parameters in the production function; (b) the replacement costs function $r(.)$; (c) the probability distribution of firm heterogeneity $F_\eta(.)$; (d) the dispersion parameter σ_ε; and (e) the discount factor β. Let θ represent the vector of structural parameters. We are interested in the estimation of θ.

Here I describe the Maximum Likelihood estimation of these parameters. Conditional on the observe history of price of capital and on the initial condition for the capital stock, we have that:

$$\Pr\left(Data \mid C, K_{i1}, \theta\right) \ = \ \prod_{i=1}^{N} \Pr\left(a_{i1}, Y_{i1}, ..., \ a_{iT}, Y_{iT} \mid C, K_{i1}, \theta\right)$$

The probability $\Pr\left(a_{i1}, Y_{i1}, ..., \ a_{iT}, Y_{iT} \mid C, K_{i1}, \theta\right)$ is the contribution of firm i to the likelihood function. Conditional on the individual heterogeneity, $\eta_i \equiv (\eta_i^Y, \eta_i^C)$, we have that:

$$\Pr\left(a_{i1}, Y_{i1}, ..., \ a_{iT}, Y_{iT} \mid C, K_{i1}, \eta_i, \theta\right) \ = \ \prod_{t=1}^{T} \Pr\left(a_{it}, Y_{it} \mid C_t, K_{it}, \eta_i, \theta\right)$$

$$= \ \prod_{t=1}^{T} \Pr\left(Y_{it} \mid a_{it}, C_t, K_{it}, \eta_i, \theta\right) \Pr\left(a_{it} \mid C_t, K_{it}, \eta_i, \theta\right)$$

where $\Pr\left(a_{it} \mid C_t, K_{it}, \eta_i, \theta\right)$ is the CCP function:

$$\Pr\left(a_{it} \mid C_t, K_{it}, \eta_i, \theta\right) \ = \ P_i\left(K_{it}, C_t, \theta\right)^{a_{it}} \ \left[1 - P_i\left(K_{it}, C_t, \theta\right)\right]^{1 - a_{it}}$$

and $\Pr\left(Y_{it} \mid a_{it}, C_t, K_{it}, \eta_i, \theta\right)$ comes from the production function, $Y_{it} = \exp\left\{\alpha_0 + \eta_i^Y\right\}$ $\left[(1 - a_{it}) \, K_{it} + a_{it} \, K^*\right]^{\alpha_1}$. In logarithms, the production function is:

$$\ln Y_{it} = \alpha_0 + \alpha_1 \, (1 - a_{it}) \ln K_{it} + \kappa \, a_{it} + \eta_i^Y + e_{it}$$

where κ is a parameter that represents $\alpha_1 \ln K^*$, and e_{it} is a measurement error in output, that we assume i.i.d. $N(0, \sigma_e^2)$ and independent of ε_{it}. Therefore,

$$\Pr\left(Y_{it} \mid a_{it}, C_t, K_{it}, \eta_i, \theta\right) \;=\; \phi\left(\frac{\ln Y_{it} - \alpha_0 - \alpha_1 \, (1 - a_{it}) \ln K_{it} - \kappa \, a_{it} - \eta_i^Y}{\sigma_e}\right)$$

where $\phi\left(.\right)$ is the PDF of the standard normal distribution.

Putting all these pieces together, we have that the log-likelihood function of the model is $\ell(\theta) = \sum_{i=1}^{N} \ln \mathcal{L}_i(\theta)$ where $\mathcal{L}_i(\theta) \equiv \Pr\left(a_{i1}, Y_{i1}, ..., \; a_{iT}, Y_{iT} \mid C, \, K_{i1}, \theta\right)$ and:

$$\mathcal{L}_i(\theta) = \sum_{\eta \in \Omega} F_\eta(\eta) \left[\begin{array}{c} \displaystyle\prod_{t=1}^{T} \phi\left(\dfrac{\ln Y_{it} - \alpha_0 - \alpha_1 \, (1 - a_{it}) \ln K_{it} - \kappa \, a_{it} - \eta^Y}{\sigma_e}\right) \\[2em] P_i\left(K_{it}, C_t, \eta, \theta\right)^{a_{it}} \; \left[1 - P_i\left(K_{it}, C_t, \eta, \theta\right)\right]^{1 - a_{it}} \end{array} \right]$$

Given this likelihood, we can estimate by Maximum Likelihood (ML)

The NFXP algorithm is a gradient iterative search method to obtain the MLE of the structural parameters.

This algorithm nests a BHHH method (outer algorithm), that searches for a root of the likelihood equations, with a value function or policy iteration method (inner algorithm), that solves the DP problem for each trial value of the structural parameters. The algorithm is initialized with an arbitrary vector $\hat{\theta}_0$.

A BHHH iteration is defined as:

$$\hat{\theta}_{k+1} = \hat{\theta}_k + \left(\sum_{i=1}^{N} \nabla l_i(\hat{\theta}_k) \nabla l_i(\hat{\theta}_k)'\right)^{-1} \left(\sum_{i=1}^{N} \nabla l_i(\hat{\theta}_k)\right)$$

where $\nabla l_i(\theta)$ is the gradient in θ of the log-likelihood function for individual i. In a partial likelihood context, the score $\nabla l_i(\theta)$ is:

$$\nabla l_i(\theta) = \sum_{t=1}^{T_i} \nabla \log P(a_{it} | x_{it}, \theta)$$

To obtain this score we have to solve the DP problem.

In our machine replacement model:

$$l(\theta) \;=\; \sum_{i=1}^{N} \sum_{t=1}^{T_i} a_{it} \, \log P(x_{it}, \theta) + (1 - a_{it}) \, \log(1 - P(x_{it}, \theta))$$

with:

$$\mathbf{P}(\theta) \;=\; F_{\tilde{\varepsilon}}\left(\begin{array}{c} [\theta_{Y0} + \theta_{Y1}\mathbf{X} + \beta \, \mathbf{F}_x(0)\mathbf{V}(\theta)] \\ -[\theta_{Y0} - \theta_{R0} - \theta_{Y1}\mathbf{X} + \beta \, \mathbf{F}_x(1)\mathbf{V}(\theta)] \end{array}\right)$$

The NFXP algorithm works as follows. At each iteration we can distinguish three main tasks or steps.

Step 1: Inner iteration: DP solution. Given $\hat{\boldsymbol{\theta}}_0$, we obtain the vector $\bar{\mathbf{V}}(\hat{\boldsymbol{\theta}}_0)$ by using either successive iterations or policy iterations.

Step 2: Construction of scores. Then, given $\hat{\boldsymbol{\theta}}_0$ and $\bar{\mathbf{V}}(\hat{\boldsymbol{\theta}}_0)$ we construct the choice probabilities

$$\mathbf{P}(\hat{\boldsymbol{\theta}}_0) \;=\; F_{\tilde{\varepsilon}}\left(\begin{array}{c} \left[\theta_{Y0} + \theta_{Y1}\mathbf{X} + \beta\ \mathbf{F}_x(0)\mathbf{V}(\hat{\boldsymbol{\theta}}_0)\right] \\ -\left[\theta_{Y0} - \theta_{R0} - \theta_{Y1}\mathbf{X} + \beta\ \mathbf{F}_x(1)\mathbf{V}(\hat{\boldsymbol{\theta}}_0)\right] \end{array} \right)$$

the Jacobian $\dfrac{\partial \bar{\mathbf{V}}(\hat{\boldsymbol{\theta}}_0)'}{\partial \theta}$ and the scores $\nabla l_i(\hat{\boldsymbol{\theta}}_0)$

Step 3: BHHH iteration. We we use the scores $\nabla l_i(\hat{\boldsymbol{\theta}}_0)$ to make a new BHHH iteration to obtain $\hat{\boldsymbol{\theta}}_1$.

$$\hat{\boldsymbol{\theta}}_1 = \hat{\boldsymbol{\theta}}_0 + \left(\sum_{i=1}^{N} \nabla l_i(\hat{\boldsymbol{\theta}}_0)\nabla l_i(\hat{\boldsymbol{\theta}}_0)' \right) \left(\sum_{i=1}^{N} \nabla l_i(\hat{\boldsymbol{\theta}}_0) \right)$$

Then, we replace $\hat{\boldsymbol{\theta}}_0$ by $\hat{\boldsymbol{\theta}}_1$ and go back to step 1.

* We repeat stesp 1 to 3 until convergence: i.e., until the distance between $\hat{\boldsymbol{\theta}}_1$ and $\hat{\boldsymbol{\theta}}_0$ is smaller than a pre-specified convergence constant.

The main advantages of the NFXP algorithm are its conceptual simplicity and, more importantly, that it provides the MLE which is the most efficient estimator asymptotically under the assumptions of the model.

The main limitation of this algorithm is its computational cost. In particular, the DP problem should be solved for each trial value of the structural parameters.

4. Patent Renewal Models

• **What is the value of a patent? How to measure it?**

• The valuation of patents is very important for: merger & acquisition decisions; using patents as collateral for loans; value of innovations; value of patent protection.

• Very few patents are traded, and there is substantial selection. We cannot use an "hedonic" approach.

• The number of citations of a patent is a very imperfect measure of patent value.

• Multiple patents are used in the production of multiple products, and in generating new patents. A "production function approach" seems also unfeasible.

4.1. Pakes (1986). • Pakes (1986) proposes using information on patent renewal fees together with a *Reveal Preference approach* to estimate the value of a patent.

• Every year, a patent holder should pay a renewal fee to keep her patent.

• If the patent holder decides to renew, it is because her expected value of holding the patent is greater than the renewal fee (that is publicly known).

• Therefore, observed decisions on patent renewal / non renewal contain information on the value of a patent.

Model: Basic Framework

• Consider a patent holder who has to decide whether to renew her patent or not. We index patents by i.

• This decision should be taken at ages $t = 1, 2, ..., T$ where $T < \infty$ is the regulated term of a patent (e.g., 20 years in US, Europe, or Canada).

• Patent regulation also establishes a sequence of **Maintenance Fees** $\{c_t : t = 1, 2, ..., T\}$. This sequence of renewal fees is deterministic such that a patent owner knows with certainty future renewal fees.

• The schedule $\{c_t : t = 1, 2, ..., T\}$ is typically increasing in patent age t and it may god from a few hundred dollars to a few thousand dollars.

• A patent generates a sequence of profits $\{\pi_{it} : t = 1, 2, ..., T\}$.

• At age t, a patent holder knows current profit π_{it} but has uncertainty about future profits $\pi_{i,t+1}$, $\pi_{i,t+2}$, ...

• The evolution of profits depends on the following factors:

(1) the initial "quality" of the idea/patent;

(2) innovations (new patents) which are substitutes of the patent and therefore, depreciate its value or even make it obsolete;

(3) innovations (new patents) which are complements of the patent and therefore, increase its value.

Stochastic process of patent profits

• Pakes proposes the following stochastic process for profits, that tries to capture the three forces mentioned above.

• A patent profit at the first period is a random draw from a log-normal distribution with parameters μ_1 and σ_1:

$$\ln(\pi_{i1}) \sim N(\mu_1, \sigma_1^2)$$

• After the first year, profit evolves according to the following formula:

$$\pi_{i,t+1} = \tau_{i,t+1} \ \max\left\{ \ \delta \ \pi_{it} \ ; \ \xi_{i,t+1}\right\}$$

- $\delta \in (0,1)$ is the depreciation rate. In the absence of unexpected shocks, the value of the patent depreciates according to the rule: $\pi_{i,t+1} = \delta \, \pi_{it}$.

- $\tau_{i,t+1} \in \{0,1\}$ is a binary variable that represents that the patent becomes obsolete (i.e., zero value) due to competing innovations. The probability of this event is a decreasing function of profit at previous year:

$$\Pr(\tau_{i,t+1} = 0 \mid \pi_{it}, t) = \exp\{-\lambda \, \pi_{it}\}$$

- The largest is the profit of the patent at age t, the smallest is the probability that it becomes obsolete.

- Variable $\xi_{i,t+1}$ represents innovations which are complements of the patent and increase its profitability.

- $\xi_{i,t+1}$ has an exponential distribution with mean γ and standard deviation $\phi^t \sigma$:

$$p(\xi_{i,t+1} \mid \pi_{it}, t) = \frac{1}{\phi^t \sigma} \exp\left\{-\frac{\gamma + \xi_{i,t+1}}{\phi^t \sigma}\right\}$$

- If $\phi < 1$, the variance of $\xi_{i,t+1}$ declines over time (and the $E(\max\{\,x\,;\,\xi_{i,t+1}\})$ value declines as well).

- If $\phi > 1$, the variance of $\xi_{i,t+1}$ increases over time (and the $E(\max\{\,x\,;\,\xi_{i,t+1}\})$ value increases as well).

- Under this specification, profits $\{\pi_{it}\}$ follow a non-homogeneous Markov process with initial density $\pi_{i1} \sim \ln N(\mu_1, \sigma_1^2)$, and transition density function:

$$f_\varepsilon\left(\pi_{it+1} \mid \pi_{it}, t\right) = \begin{cases} \exp\{-\lambda \, \pi_{it}\} & \text{if } \pi_{it+1} = 0 \\[2mm] \Pr\left(\xi_{it+1} < \delta\pi_{it} \mid \pi_{it}, t\right) & \text{if } \pi_{it+1} = \delta\pi_{it} \\[2mm] \dfrac{1}{\phi^t \sigma} \exp\left\{-\dfrac{\gamma + \pi_{it+1}}{\phi^t \sigma}\right\} & \text{if } \pi_{it+1} > \delta\pi_{it} \end{cases}$$

- The vector of structural parameters is $\theta = (\lambda, \delta, \gamma, \phi, \sigma, \mu_1, \sigma_1)$.

   Model: Dynamic Decision Model

- $V_t(\pi)$ is the value of an active patent of age t and current profit π.

- Let $a_{it} \in \{0,1\}$ be the decision variable that represents the event "the patent owner decides to renew the patent at age t".

- The value function is implicitly defined by the Bellman equation:

$$V_t(\pi_{it}) = \max\left\{\,0\,;\,\pi_{it} - c_t + \beta \int V_{t+1}(\pi_{i,t+1})\, f_\varepsilon(d\pi_{i,t+1} \mid \pi_{it}, t)\,\right\}$$

with $V_t(\pi_{it}) = 0$ for any $t \geq T+1$.

- The value of not renewal ($a_{it} = 0$) is zero. The value of renewal ($a_{it} = 1$) is the current profit $\pi_{it} - c_t$ plus the expected and discounted future value.

Model: Solution (Backwards induction)

• We can use backwards induction to solve for the sequence of value functions $\{V_t\}$ and optimal decision rules $\{\alpha_t\}$:

• Starting at age $t = T$, for any profit π:

$$V_T(\pi) = \max\{ 0 \; ; \; \pi - c_T\}$$

and

$$\alpha_T(\pi) = 1\{ \pi - c_T \geq 0 \}$$

• Then, for age $t < T$, and for any profit π:

$$V_t(\pi) = \max\left\{ 0 \; ; \; \pi - c_t + \beta \int V_{t+1}(\pi') \; f_\varepsilon(d\pi'|\pi, t) \right\}$$

and

$$\alpha_t(\pi) = 1\left\{ \pi - c_t + \beta \int V_{t+1}(\pi') \; f_\varepsilon(d\pi'|\pi, t) \geq 0 \right\}$$

Solution - A useful result

• Given the form of $f_\varepsilon(\pi'|\pi, t)$, the future and discounted expected value, $\beta \int V_{t+1}(\pi') \; f_\varepsilon(d\pi'|\pi, t)$, is increasing in current π.

• This implies that the solution of the DP problem can be described as a **sequence of threshold values for profits** $\{\pi_t^* : t = 1, 2, ..., T\}$ such that the optimal decision rule is:

$$\alpha_t(\pi) = 1\{ \pi \geq \pi_t^* \}$$

• π_t^* is the level of current profits that leaves the owner indifferent between renewing the patent or not: $V_t(\pi_t^*) = 0$.

• These threshold values are obtained using backwards induction:
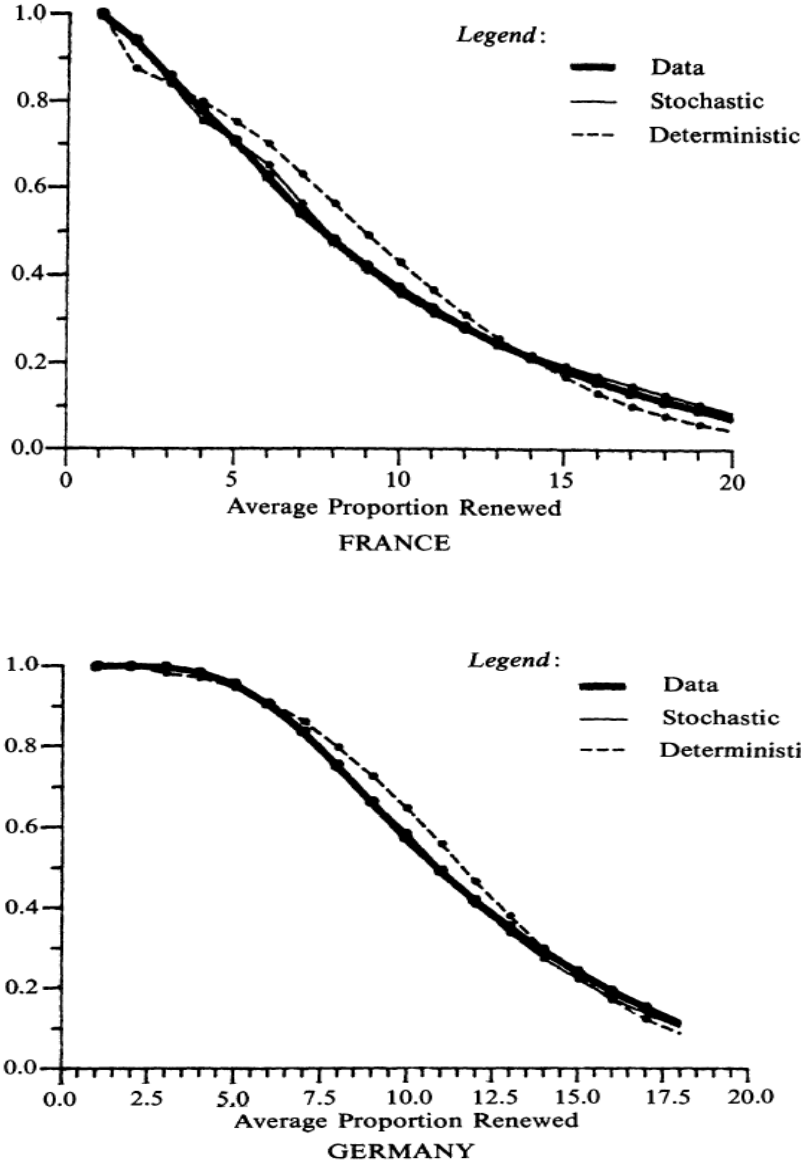
• At period $t = T$:

$$\pi_T^* = c_T$$

• At period $t < T$, π_t^* is the unique solution to the equation:

$$\pi_t^* - c_t + E\left(\sum_{s=t+1}^{T} \beta^{s-t} \; \max\{ 0 \; ; \; \pi_{t+1} - \pi_{t+1}^* \} \mid \pi_t = \pi_t^* \right) = 0$$

• Solving for a sequence of threshold values is much simpler that solving for a sequence of value functions.

Data

• Sample of N patents with complete (uncensored) durations $\{d_i : i = 1, 2, ...N\}$, where $d_i \in \{1, 2, ..., T+1\}$ is patent i's duration or age at its last renewal period.

FRANCE

GERMANY

- The information in this sample can be summarized by the empirical distribution of $\{d_i\}$:

$$\widehat{p}(t) = \frac{1}{N} \sum_{i=1}^{N} 1\{d_i = t\}$$

Estimation: Likelihood

- The log-likelihood function of this model and data is:

$$
\begin{aligned}
l(\theta) &= \sum_{i=1}^{N} \sum_{t=1}^{T+1} 1\{d_i = t\} \ \ln \Pr(d_i = t | \theta) \\
&= N \sum_{t=1}^{T+1} \widehat{p}(t) \ \ln P(t | \theta)
\end{aligned}
$$

where:

$$P(t|\theta) \;=\; \Pr\left(\pi_s \geq \pi_s^* \text{ for } s \leq t-1, \text{and } \pi_t < \pi_t^* \mid \theta\right)$$

$$= \int_{\pi_1^*}^{\infty} \cdots \int_{\pi_{t-1}^*}^{\infty} \int_0^{\pi_t^*} dF(\pi_1, ..., \pi_{t-1}, \pi_t)$$

• Computing $P(t|\theta)$ involves solving an integral of dimension t. For t greater than 4 or 5, it is computationally very costly to obtain the exact value of these probabilities. Instead, we approximate these probabilities using Monte Carlo simulation.

   Estimation: Simulation of Probabilities

• For a given value of θ, let $\{\pi_t^{sim}(\theta) : t = 1, 2, ..., T\}$ be a simulated history of profits for patent i.

• Suppose that, for a given value of θ, we simulate R **independent** profit histories. Let $\{\pi_{rt}^{sim}(\theta) : t = 1, 2, ..., T; r = 1, 2, ..., R\}$ be these histories.

• Then, we can approximate the probability $P(t|\theta)$ using the following simulator:

$$\tilde{P}_R(t|\theta) = \frac{1}{R} \sum_{r=1}^{R} 1\{\pi_{rs}^{sim}(\theta) \geq \pi_s^* \text{ for } s \leq t-1, \text{and } \pi_{rt}^{sim} < \pi_t^*\}$$

• $\tilde{P}_R(t|\theta)$ is a *raw frequency simulator*. It has the following properties (Note that these are properties of a simulator, not of an estimator. $\tilde{P}_R(t|\theta)$ does not depend on the data).

   (1) Unbiased: $E\left(\tilde{P}_R(t|\theta)\right) = P(t|\theta)$

   (2) $Var(\tilde{P}_R(t|\theta)) = P(t|\theta)(1 - P(t|\theta))/R$

   (3) Consistent as $R \to \infty$.

• It is possible to obtain better simulators (with lower variance) by using importance-sampling simulation. This is relevant because the bias and variance of simulated-based estimators depend on the variance (and bias) of the simulator.

• Furthermore, when $P(t|\theta)$ is small, the simulator $\tilde{P}_R(t|\theta)$ can be zero even when R is large, and this creates problems for ML estimation.

• A simple solution to this problem is to consider the following simulator which is based on the raw-frequency simulated probabilities $\tilde{P}_R(1|\theta)$, $\tilde{P}_R(2|\theta)$, ... $\tilde{P}_R(T+1|\theta)$:

$$P_R^*(t|\theta) = \frac{\exp\left\{\dfrac{\tilde{P}_R(t|\theta)}{\eta}\right\}}{\sum_{s=1}^{T+1} \exp\left\{\dfrac{\tilde{P}_R(s|\theta)}{\eta}\right\}}$$

where $\eta > 0$ is a smoothing parameter.

• The simulator P_R^* is biased. However, if $\eta \to 0$ as $R \to \infty$, then P_R^* is consistent, it has lower variance than $\tilde{P}_R$, and it is always strictly positive.

Simulation-Based Estimation

• The estimator of θ (Simulated Method of Moments estimator) is the value that solves the system of T equations: for $t = 1, 2, ...T$:

$$\frac{1}{N} \sum_{i=1}^{N} \left[1\{d_i = t\} - \tilde{P}_{R,i}(t|\theta) \right] = 0$$

where the subindex i in the simulator $\tilde{P}_{R,i}(t|\theta)$ indicates that for each patent i in the sample we draw R independent histories and compute independent simulators.

• **Effect of simulation error.** Note that $\tilde{P}_{R,i}(t|\theta)$ is unbiased such that $\tilde{P}_{R,i}(t|\theta) = P(t|\theta) + e_i(t, \theta)$, where $e_i(t, \theta)$ is the simulation error. Since the simulation errors are independent random draws:

$$\frac{1}{N} \sum_{i=1}^{N} e_i(t, \theta) \to_p 0 \quad \text{and} \quad \frac{1}{\sqrt{N}} \sum_{i=1}^{N} e_i(t, \theta) \to_d N(0, V_R)$$

The estimator is consistent an asymptotically normal for any R. The variance of the estimator declines with R.

Identification

• Since there are only 20 different values for the renewal fees $\{c_t\}$ we can at most identify 20 different points in the probability distribution of patent values.

• The estimated distribution at other points is the result of interpolation or extrapolation based on the functional form assumptions on the stochastic process for profits.

• It is important to note that the identification of the distribution of patent values is NOT up to scale but in dollar values.

• For a given patent of with age t, all what we can say is that: if $a_{it} = 0$, then $V_{it} < V(\pi_t^*)$; and if $a_{it} = 1$, then $V_{it} \geq V(\pi_t^*)$.

Empirical Questions

• The estimated model can be used to address important empirical questions.

• **Valuation of the stock of patents.** Pakes uses the estimated model to obtain the value of the stock of patents in a country.

• According to the estimated model, the value of the stock of patents in 1963 was $315 million in France, $385 million in UK, and $511 in Germany.

• Combining these figures with data on R&D investments in these countries, Pakes calculates rates of return of 15.6%, 11.0% and 13.8%, which look like quite reasonable.

Empirical Questions

• **Factual policies.** The estimated model shows that a very important part of the observed between-country differences in patent renewal can be explained by differences in policy parameters (i.e., renewal fees and maximum length).

• **Counterfactual policy experiments.** The estimated model can be used to evaluate the effects of policy changes (in renewal fees and/or in maximum length) which are not observed in the data.

4.2. Lanjow (1999). Estimates the private value of patent protection for four technology areas—computers, textiles, combustion engines, and pharmaceuticals - using new patent data for West Germany, 1953-1988. The model takes into account that patentees must pay not only renewal fees to keep their patents but also legal expenses to enforce them. The dynamic structural model takes into account the potential need to prosecute infringement. Results show that the aggregate value of protection generated per year is on the order of 10% of related R&D expenditure.

4.3. Trade of patents: Serrano (2018). The sale of patents is an incentive to invest in R&D, especially for small firms. This market can generate social gains by reallocating patent rights from innovators to firms that may be more effective in using, commercializing, or enforcing these rights. There are also potential social costs, if the acquiring firms can exercise more market power. Serrano (IER, 2018) investigates the value of trading patents by estimating a structural model that includes renewal and trading decisions.

Data. Panel of patents granted to U.S small firms (no more than 500 employees) in the period 1988-1997 (15% of patents granted to firms). In the U.S. patent system, the patent holder needs to pay renewal fees to maintain the patent only at ages 5, 9, and 13 years. Fee increases with age: $c_{13} > c_9 > c_5$. Serrano (2000) constructs the dataset with renewals and transfers/sales. Working sample: 54,840 patents from 10 granting cohorts (1988 to 1997) followed from granting period until 2001 or not renewal.

Renewal and trading frequencies. Probability that a patent is traded (between renewal dates): - higher if previously untraded. - decreases with age. Probability of patent expiration (at renewal dates) - lower for previously traded. - increase over time.

Renewal and trading frequencies

A-1: Percentage of Active Small Business Patents Traded and

Age	All	Not Previously Traded	Previously Traded (Years since last trade)	
			Any Year	One year
A. Probability that an active patent is traded				
2	2.99	2.85	7.47	7.47
7	2.81	2.46	4.79	6.63
11	2.51	2.13	3.77	2.55
B. Probability that an active patents is allowed to expire				
5	17.2	17.7	12.7	6.2
9	25.6	26.6	21.4	11.6
13	25.5	26.6	22.5	14.1

Model: Key features. The transfer/sale of a patent involves a transaction cost. This transaction cost creates a selection effect: patents with higher per period returns are more likely to be traded. This selection effect explains the observed pattern that previously traded patents are: - more likely to be traded; - less likely to expire.

Returns. At age t, a patent has: - an **internal return** for the current patent owner, x_t; - a potential **external return** for the best alternative user, y_t. There is an "improvement factor", g_t^e, that relates external and internal returns: $y_t = g_t^e x_t$, g_t^e is i.i.d. with a truncated (at zero) exponential distribution: $\gamma^e \equiv \Pr(g_t^e = 0)$, and σ^e is the mean of the exponential. Initial (internal) returns: $\log(x_1) \sim N(\mu, \sigma_R^2)$. Next period returns:

$$x_{t+1} = \begin{cases} g_t^i \, x_t & \text{if not traded at age } t \\ g_t^i \, y_t & \text{if traded at age } t \end{cases}$$

g_t^i is a random variable with a truncated (at zero) exponential distribution: $\gamma^i \equiv \Pr(g_t^i = 0)$, and σ_t^i is the mean of this exponential, and $\sigma_t^i = \phi^t \sigma_0^i$, with $\phi \in (0,1)$. This implies that x_{t+1} follows a first order Markov process. Remember that there is a lump-sum transaction cost, τ. It is assumed that is paid by the buyer.

Model: Renewal and Sale decisions. Let $V_t(x_t, y_t)$ be the value of a patent with age t, current internal and external returns x_t and y_t, resp.

$$V_t(x_t, y_t) = \max \left\{ 0, \; V_t^K(x_t, y_t), \; V_t^S(x_t, y_t) \right\}$$
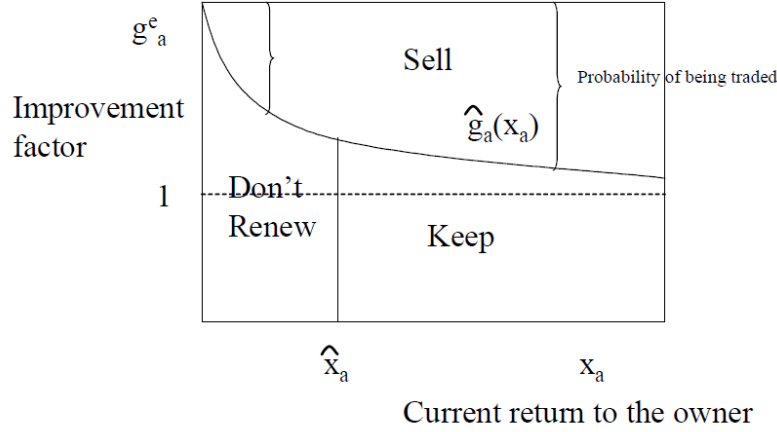
$V_t^K(x_t, y_t) =$ value of keeping; $V_t^S(x_t, y_t) =$ value of selling. And for $t \le T = 17$:

$$V_t^K(x_t, y_t) \;\;=\;\; x_t - c_t + \beta \, \mathbb{E}\left[V_{t+1}(x_{t+1}, y_{t+1}) \mid x_t, y_t, a_t = K\right]$$

$$V_t^S(x_t, y_t) \;\;=\;\; x_t - c_t - \tau + \beta \, \mathbb{E}\left[V_{t+1}(x_{t+1}, y_{t+1}) \mid x_t, y_t, a_t = S\right]$$

with $V_{T+1}^K = V_{T+1}^S = 0$.

Figure 1: Optimal choices of a Patent Holder

Model: Optimal decision rule. Lemma 1: $V_t(x_t, y_t)$ is weakly increasing in x_t and y_t, and weakly decreasing in t. Proposition 1. There are two threshold values: $x_t^*(\theta)$ that depends on age and structural parameters, and $g_t^*(x, \theta)$, that depends on age, internal return, and parameters, such that the optimal decision rule a_t is:

$$
a_t = \begin{cases}
S & \text{if } g_t^e \geq g_t^*(x_t, \theta) \\[2mm]
K & \text{if } g_t^e < g_t^*(x_t, \theta) \text{ and } x_t \geq x_t^*(\theta) \\[2mm]
0 & \text{if } g_t^e < g_t^*(x_t, \theta) \text{ and } x_t < x_t^*(\theta)
\end{cases}
$$

Identification and Estimation. Method: Simulated method of moments. Moments describing the history of trading and renewal decisions of patent owners. (1) probability that an active patent is traded at different ages conditional on having been previously traded, and conditional on not having been previously traded. (2) probability that an active patent is allowed to expire at different renewal dates conditional on having been previously traded, and conditional on not having been previously traded. A total of 186 moments.

Parameter estimates. Transaction cost: \$5,850, about one-third of the average return at age 1 (8% of the average value at age 1). On average, internal growth of returns is greater than external.

Parameter estimates

Evaluating the value of the market for patents. The possibility of trading patents has two types of the effects on the value of the pool of patents: - a direct causal effect due to the reallocation to an owner with higher returns; - a selection effect, through the renewal decisions (renewal decision is different with and without the possibility of trading).

Table 1: Parameter Estimates

Description (Parameter)	Estimate[a]
A. Patent initial returns	
Mean parameter of the Lognormal Initial Distribution (μ)	8.4179 ($4.0*10^{-2}$)
Std. Deviation parameter of the Lognormal Initial Distribution (σ_R)	1.6911 ($1.2*10^{-2}$)
B. Internal growth of returns	
Depreciation factor (δ)	0.8917 ($5.2*10^{-3}$)
Not obsolescence (γ^i)	0.9673 ($6.5*10^{-3}$)
Internal Growth of Returns (σ^i)	0.4450 ($3.6*10^{-3}$)
Upside opportunities (ϕ)	0.5941 ($6.1*10^{-3}$)
C. Market for patents and transaction costs	
Transaction cost (τ)	5,850.1 (50.49)
Mean External Growth of Returns (σ^e)	0.3745 ($2.9*10^{-3}$)
Proportion of unsuccessful transfers (γ^e)	0.0385 ($4*10^{-4}$)
Random transfers (ε)	0.0059 ($4*10^{-4}$)
Size of sample	54,840
Simulations in the estimation	164,520
MSE[b]	$3.784*10^{-4}$

Serrano measures these two sources of value. Evaluating the value of the market for patents (1) Total effect on the value of patents: - 50% of the total value of patents. - Only 23% of patents are sold, but the value of a traded patent is 3 times higher than untraded patent ($173,668 vs. $54,960). (2) Direct gains from trade (from reallocation) - accounts for 10% of the total value of the traded patents. - The distribution of the gains from trade is very skewed.

Counterfactual: Reducing transaction cost Lowering transaction cost by 50% (from $5,850 to $2,925). It raises the proportion of patents traded by 6 percentage points: from 23.1% to 29.6%. It boosts the gains from trade (reallocation) by an additional 8.7%. It increases the total value of the patent market by 3%.

5. Dynamic pricing

Retail firms selling the same product and operating in the same narrowly defined geographic market can charge prices that differ by significant amounts. Cross-sectional dispersion of prices has been well established in different retail markets such as gas stations or supermarkets, among others. Recent empirical papers show that temporary sales account for approximately half of all price changes of retail products in US (see Hosken and Reiffen, 2004, Nakamura and Steinsson, 2008, or Midrigan, 2011). Sales promotions can also account for a substantial part of cross-sectional price dispersion. Therefore, understanding the determinants of temporary sales is important to understand price stickiness, price dispersion, and firms' market power and competition.

Varian (1980) presents a model of price competition in an homogeneous product market with two types of consumers according to their information about prices. Informed customers always buy in the store with the lowest price. Uninformed consumers choose a store at

random and buy there as long as the price of the store is not above their reservation price. The model does not have an equilibrium in pure strategies. In mixed strategies, there is a unique symmetric equilibrium characterized by a U-shape density function on the price interval between the marginal cost and the reservation price of uninformed consumers. According to this equilibrium, the price charged by a store changes randomly over time between a "low" and a "high" price.

Though Varian's model can explain some important empirical features in the cross-section and time series of prices in retail markets, it cannot explain the time dependence of sales promotions that have been reported in many empirical studies (e.g., Slade, 1998, Aguirregabiria, 1999, or Pesendorfer, 2002, among others). The probability of a sales promotion increases with the duration since the last sale. Several studies have proposed and estimated dynamic structural models of retail pricing that can explain price dispersion, sales promotions and their state dependence. These studies provide also estimates of the magnitude and structure of firms' price adjustments costs.

Slade (1998) proposes a model where the demand of a product in a store depends on a stock of goodwill that accumulates over time when the store charges low prices, and erodes when the price is high. The model incorporates also menu costs of changing prices. The optimal pricing policy consists of a cycle between a low price (or sales promotion) and a high price. Slade estimates this model using weekly scanner data of prices and quantities of saltine crackers in four supermarket chains. The estimated model fits well the join dynamics of prices and quantities. Her estimates of the cost of adjusting prices are approximately 4% of revenue.

Aguirregabiria (1999) studies the relationship between inventories and prices in supermarkets. The cost of placing orders to wholesalers has a fixed component. Retailers have also menu costs of changing prices, face substantial demand uncertainty, and have stockouts. Aguirregabiria proposes a model of price and inventory decisions that incorporates these features. In the optimal decision rule of this model, inventories follow an (S,s) cycle, and prices have a "high-low" cyclical pattern. When a new order is placed, the probability of stockout declines, expected demand becomes more elastic, and the optimal price drops to a minimum. When inventories decline between two orders, the probability of stockout increases, expected sales become more inelastic, and the optimal price eventually increases and stays high until the next order. Aguirregabiria estimates this model using data on inventories, prices, and sales from the warehouse of a supermarket chain. The estimated model fits well the joint cyclical pattern of prices and inventories in the data and can explain temporary sales. The estimated values for the fixed ordering cost and the menu cost are 3.1% and 0.7% of monthly

revenue, respectively. According to the estimated model, almost 50% of sales promotions are associated to the dynamics of inventories.

Pesendorfer (2002) studies the dynamics of consumer demand as a factor explaining sales promotions of storable products. He proposes a model of demand of a storable product with two types of consumers: store-loyal consumers and shoppers. The equilibrium of the model predicts that the probability that a store has a sale increases with the duration since the last sale both in that store and in other stores. The implied pattern of prices path consists of an extended period of high prices followed by a short period of low prices. He tests the predictions of the model using supermarket scanner data for ketchup products. The effects of the duration variables are significant and have the predicted sign. Though this evidence suggests that demand accumulation could be important in the decision to conduct a sale, it is also consistent with models in Slade (1998) and Aguirregabiria (1999). As far as we know, there is not an empirical study that has tried to disentangle the relative contribution of consumer inventories, firm inventories, and goodwill to explain temporary sales promotions.

Kano (2013) makes an interesting point on the estimation menu costs in oligopoly markets. Dynamic price competition in oligopoly markets implies a positive strategic interaction between the prices of different firms. This strategic interaction may be an importance source of price inertia even when menu costs are small. If a firm experiences an idiosyncratic increase in its marginal cost, it may prefer not to increase its price if the competitor maintains its price constant. A model of monopolistic competition that ignores strategic interactions among firms may spuriously overestimate menu costs. Kano estimates a dynamic pricing model that accounts for these strategic interactions and finds that they account for an substantial part of price rigidity.

5.1. Aguirregabiria (1999). The significant cross-sectional dispersion of prices is a well-known stylized fact in retail markets. Retailing firms selling the same product, and operating in the same (narrowly defined) geographic market and at the same period of time, do charge prices that differ by significant amounts, e.g., 10% price differentials or even larger. This empirical evidence has been well established for gas stations and supermarkets, among other retail industries. Interestingly, the price differentials between firms, and the ranking of firms in terms prices, have very low persistence over time. A gas station that charges a price 5% below the average in a given week may be charging a price 5% above the average the next week. Using a more graphical description we can say that a firm's price follows a cyclical pattern, and the price cycles of the different firms in the market are not synchronized. Understanding price dispersion and the dynamics of price dispersion is very important to understand not only competition and market power but also for the construction of price indexes.

Different explanations have been suggested to explain this empirical evidence. Some explanations have to do with dynamic pricing behavior or "state dependence" in prices.

For instance, an explanation is based on the relationship between firm inventory and optimal price. In many retail industries with storable products, we observe that firms' orders to suppliers are infrequent. For instance, for products such as laundry detergent, a supermarket ordering frequency can be lower than one order per month. A simple and plausible explanation of this infrequency is that there are fixed or lump-sum costs of making an order that do not depend on the size of the order, or at least they do not increase proportionally with the size of the order. Then, inventories follow a so called (S,s) cycle: the increase by a large amount up to a maximum when a place is order and then they decline slowly up a minimum value where a new order is placed. Given this dynamics of inventories, it is simple to show that optimal price of the firm should also follow a cycle. The price drops to a minimum when a new order is placed and then increases over time up to a maximum just before the next order when the price drops again. Aguirregabiria (*REStud*, 1999) shows this joint pattern of prices and inventories for many products in a supermarket chain. I show that this type of inventory-depedence price dynamics can explain more than 20% of the time series variability of prices in the data.

Temporary sales and inventories. Recent empirical papers show that temporary sales account for approximately half of all price changes of retail products in US: Hosken and Reiffen (RAND, 2004); Nakamura and Steinsson (QJE, 2008); Midrigan (Econometrica, 2011). Understanding the determinants of temporary sales is important to understand price stickiness and price dispersion, and it has important implications on the effects of monetary policy. It has also important implications in the study of firms' market power and competition. Different empirical models of sales promotions: Slade (1998) [Endogenous consumer loyalty], Aguirregabiria (1999) [Inventories], Pesendorfer (2002) [Intertemporal price discrimination], and Kano (2013).

This paper studies how retail inventories, and in particular (S,s) inventory behavior, can explain both price dispersion and sales promotions in retail markets. Three factors are key for the explanation provided in this paper:

(1) Fixed (lump-sum) ordering costs, that generates (S,s) inventory behavior.

(2) Demand uncertainty.

(3) Sticky prices (Menu costs) that, together with demand uncertainty, creates a positive probability of excess demand (stockout).

Model: Basic framework

Consider a retail firm selling a product. We index products by i. Every period (month) t the firm decides the retail price and the quantity of the product to order to manufacturers/wholesalers. **Monthly sales** are the minimum of supply an demand:

$$y_{it} = \min\left\{ d_{it} \; ; \; s_{it} + q_{it} \right\}$$

y_{it} = sales in physical units; d_{it} = demand; s_{it} = inventories at the beginning of month t; q_{it} = orders (and deliveries) during month t.

Demand and Expected sales. The firm has uncertainty about **current demand**:

$$d_{it} = d_{it}^e \; \exp\left(\xi_{it}\right)$$

d_{it}^e = expected demand; ξ_{it} = zero mean demand shock unknown to the firm at t. Therefore, **expected sales** are:

$$y_{it}^e = \int \min\left\{d_{it}^e \; \exp\left(\xi\right) \; ; \; s_{it} + q_{it} \right\} dF_\xi(\xi)$$

Assume monopolistic competition. **Expected Demand** depends on the own price, p_{it}, and a demand shock ω_{it}. The functional form is isoelastic:

$$d_{it}^e = \exp\left\{\gamma_0 - \gamma_1 \; \ln(p_{it}) + \omega_{it} \right\}$$

where γ_0 and $\gamma_1 > 0$ are parameters.

Price elasticity of expected sales. **Demand uncertainty** has important implications for the relationship between prices and inventories. The price elasticity of expected sales is a function of the **supply-to-expected-demand ratio** $(s_{it} + q_{it})/d_{it}^e$:

$$\eta_{y^e|p} \equiv \frac{-\partial y^e}{\partial p} \frac{p}{y^e} = -\left[\int I\left\{d^e \exp\left(\xi\right) \; ; \; s+q \right\} dF_\xi(\xi)\right] \frac{\partial d^e}{\partial p} \frac{p}{y^e}$$

$$= \gamma_1 \; F_\xi\left(\log\left[\frac{s+q}{d^e}\right]\right) \frac{d^e}{y^e}$$

And we have that:

$$\eta_{y^e|p} \longrightarrow \begin{cases} \gamma_1 & \text{as} \quad (s+q)/d^e \longrightarrow \infty \\ \\ 0 & \text{as} \quad (s+q)/d^e \longrightarrow 0 \end{cases}$$

Price elasticity of expected sales

$$\eta_{y^e|p} = \gamma_1 \; F_\xi\left(\log\left[\frac{s+q}{d^e}\right]\right) \frac{d^e}{y^e}$$

[FIGURE: $\eta_{y^e|p}$ increasing in $\dfrac{s+q}{d^e}$, with asymptote at γ_1]

When the supply-to-expected-demand ratio is large, the probability of stockout is very small and $y^e \simeq d^e$, so the elasticity of expected sales is just the elasticity of demand. However, when the supply-to-expected-demand ratio is small, the probability of stockout is large and the elasticity of expected sales can be much lower than the elasticity of demand.

Markup and inventories (myopic case). This has potentially important implications for the optimal price of an oligopolistic firm. To give some intuition, consider the pricing decision of the monopolistic firm without forward-looking behavior. That optimal price is:

$$\frac{p - c}{p} = \frac{1}{\eta_{y^e|p}}$$

$$\text{OR}$$

$$\frac{p - c}{c} = \frac{1}{\eta_{y^e|p} - 1}$$

Variability over time in the supply-to-expected-demand ratio can generate significant fluctuations in price-cost margins. It can also explain temporary sales promotions. That can be the case under (S, s) inventory behavior. Evolution of inventories and price without menu cost; and Evolution of inventories and price with menu cost

****

Figure: Cyclical pattern of inventories and prices

****

Empirical Application

The paper investigates this hypothesis using a data from a supermarket chain, with rich information on prices, sales, inventories, orders, and wholesale prices for many different products. Reduced form estimations present evidence that supports the hypothesis:

(1) Prices depend negatively and very significantly on the level of inventories.

(2) Inventories of many products follow (S,s) cycles.

(3) Price cost margins increase at the beginning of an (S,s) cycle, and decline monotonically during the cycle.

I estimate the parameters in the profit function (demand parameters, ordering costs, inventory holding costs) and use the estimated model to analyze how much of price variation and temporary sales promotions can be explained by firm inventories.

Profit function. **Expected current profits** are equal to expected revenue, minus ordering costs, inventory holding costs and price adjustment costs:

$$\pi_{it} = p_{it}\, y_{it}^e - OC_{it} - IC_{it} - PAC_{it}$$

OC_{it} = ordering costs; IC_{it} = inventory holding costs; PAC_{it} = price adjustment (menu) costs. **Ordering costs:**

$$OC_{it} = \begin{cases} 0 & \text{if } q_{it} = 0 \\ F_{oc} + \varepsilon_{it}^{oc} - c_{it}\, q_{it} & \text{if } q_{it} > 0 \end{cases}$$

$F_{oc} =$ fixed (lump-sum) ordering cost. Parameter; $\varepsilon_{it}^{oc} =$ zero mean shock in the fixed ordering cost; $c_{it} =$ wholesale price. **Inventory holding costs:**

$$IC_{it} = \alpha \; s_{it}$$

Menu costs:

$$PAC_{it} = \begin{cases} 0 & \text{if} \quad p_{it} = p_{i,t-1} \\[2mm] F_{mc}^{(+)} + \varepsilon_{it}^{mc(+)} & \text{if} \quad p_{it} > p_{i,t-1} \\[2mm] F_{mc}^{(-)} + \varepsilon_{it}^{mc(-)} & \text{if} \quad p_{it} < p_{i,t-1} \end{cases}$$

$F_{mc}^{(+)}$ and $F_{mc}^{(-)}$ are price adjustment cost parameters; $\varepsilon_{it}^{mc(+)}$ and $\varepsilon_{it}^{mc(-)}$ are zero mean shocks in menu costs

State variables. The state variables of this DP problem are:

$$\left\{ \underbrace{s_{it}, \; c_{it}, \; p_{i,t-1}, \; \omega_{it},}_{x_{it}} \; \underbrace{\varepsilon_{it}^{oc}, \; \varepsilon_{it}^{mc(+)}, \; \varepsilon_{it}^{mc(+)}}_{\varepsilon_{it}} \right\}$$

The decision variables are q_{it} and $\Delta p_{it} \equiv p_{it} - p_{i,t-1}$. We use a_{it} to denote $(q_{it}, \Delta p_{it})$. Let $V(x_{it}, \varepsilon_{it})$ be the value of the firm associated with product i. This value function solves the Bellman equation:

$$V(x_{it}, \varepsilon_{it}) = \max_{a_{it}} \left\{ \begin{array}{c} \pi(a_{it}, x_{it}, \varepsilon_{it}) \\[2mm] + \beta \; \int V(x_{i,t+1}, \varepsilon_{i,t+1}) \; dF(x_{i,t+1}, \varepsilon_{i,t+1} | a_{it}, x_{it}, \varepsilon_{it}) \end{array} \right\}$$

Discrete Decision variables. Most of the variability of q_{it} and Δp_{it} in the data is discrete. For simplicity, we assume that these variables have a discrete support.

$$q_{it} \quad \in \quad \{0 \; , \; \kappa_i\}$$

$$\Delta p_{it} \quad \in \quad \{0 \; , \; \delta_i^{(+)}, \; \delta_i^{(-)}\}$$

where $\kappa_i > 0$, $\delta_i^{(+)} > 0$, and $\delta_i^{(-)} < 0$ are parameters. Therefore, the set of choice alternatives at every period t is:

$$a_{it} \in A = \left\{ \; (0,0), \; (0,\delta_i^{(+)}), \; (0,\delta_i^{(-)}), \; (\kappa_i,0), \; (\kappa_i,\delta_i^{(+)}), \; (\kappa_i,\delta_i^{(-)}) \; \right\}$$

The transition rules for the state variables are:

$$\begin{array}{rcc} s_{i,t+1} & = & s_{it} + q_{it} - y_{it} \\ p_{it} & = & p_{i,t-1} + \Delta p_{it} \\ c_{i,t+1} & \sim & AR(1) \\ \omega_{i,t+1} & \sim & AR(1) \\ \varepsilon_{it} & \sim & i.i.d. \end{array}$$

(Integrated) Bellman Equation. The components of ε_{it} are independently and extreme value distributed with dispersion parameter σ_ε. Therefore, as in Rust (1987), the integrated

value function $\bar{V}(x_{it})$ is the unique fixed point of the integrated Bellman equation:

$$\bar{V}(x_{it}) = \sigma_\varepsilon \ln \left(\sum_{a \in A} \exp \left\{ \frac{v(a, x_{it})}{\sigma_\varepsilon} \right\} \right)$$

where:

$$v(a, x_{it}) = \bar{\pi}(a, x_{it}) + \beta \sum_{x_{i,t+1}} \bar{V}(x_{i,t+1}) \ f_x(x_{i,t+1}|a, x_{it})$$

Discrete choice profit function

- $\bar{\pi}(a, x_{it})$ is the part of current profit which does not depend on ε_{it}:

$$\bar{\pi}(a, x_{it}) = \begin{cases} R_{it}(0, 0) - \alpha \ s_{it} & \text{if} \quad a = (0, 0) \\[2mm] R_{it}(0, \delta_i^{(+)}) - \alpha \ s_{it} - F_{mc}^{(+)} & \text{if} \quad a = (0, \delta_i^{(+)}) \\[2mm] R_{it}(0, \delta_i^{(-)}) - \alpha \ s_{it} - F_{mc}^{(-)} & \text{if} \quad a = (0, \delta_i^{(-)}) \\[2mm] R_{it}(\kappa_i, 0) - \alpha \ s_{it} - F_{oc} - c_{it}\kappa_i & \text{if} \quad a = (\kappa_i, 0) \\[2mm] R_{it}(\kappa_i, \delta_i^{(+)}) - \alpha \ s_{it} - F_{oc} - c_{it}\kappa_i - F_{mc}^{(+)} & \text{if} \quad a = (\kappa_i, \delta_i^{(+)}) \\[2mm] R_{it}(\kappa_i, \delta_i^{(-)}) - \alpha \ s_{it} - F_{oc} - c_{it}\kappa_i - F_{mc}^{(-)} & \text{if} \quad a = (\kappa_i, \delta_i^{(-)}) \end{cases}$$

where $R_{it}(.,.)$ is the expected revenue function.

Some predictions of the model. Fixed ordering cost F_{oc} generate infrequent orders: $(\mathbf{S}, \mathbf{s})$ **inventory policy**. (S, s) inventory behavior, together demand uncertainty (i.e., optimal prices depend on the supply-to-expected demand ratio) generate a cyclical pattern in the price elasticity of sales. Prices decline significantly when an order is placed (sales promotion). This price decline and the consequently inventory reduction generate a price increase. Then, as inventories decline between two orders, prices tend to increase.

Data. Data from the central warehouse of a supermarket chain in the Basque Country (Spain). Monthly data: period January 1990 to May 1992. Estimation of Structural Parameters. Counterfactual Experiments

Bibliography

[1] Ackerberg, D. (2009): "A New Use of Importance Sampling to Reduce Computational Burden in Simulation Estimation" Quantitative Marketing and Economics, 7, 343-376.

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

[3] Adda & Cooper

[4] Aguirregabiria, V. (1999): "The dynamics of markups and inventories in retailing firms," The Review of Economic Studies, 66, 275-308.

[5] Aguirregabiria, V. and P. Mira (2002): "Swapping the nested fixed point algorithm: A class of estimators for discrete Markov decision models," Econometrica 70, 1519-1543.

[6] Aguirregabiria, V. and P. Mira (2010), 'Dynamic discrete choice structural models: a survey', Journal of Econometrics, 156 (1), 38–67.

[7] Aguirregabiria, V. and A. Nevo (2013), 'Recent developments in empirical IO: dynamic demand and dynamic games', in D. Acemoglu, M. Arellano and E. Deckel (eds), Advances in Economics and Econometrics. Tenth World Congress. Vol. 3, Econometrics, Cambridge: Cambridge University Press, pp. 53–122.

[8] Aguirregabiria, V. and J. Suzuki (2014), 'Identification and counterfactuals in dynamic models of market entry and exit', Quantitative Marketing and Economics, 12 (3), 267–304.

[9] Arcidiacono, P. and R. Miller (2011), 'Conditional choice probability estimation of dynamic discrete choice models with unobserved heterogeneity', Econometrica, 79 (6), 1823–67.

[10] Arcidiacono, P., P. Bayer, J. Blevins and P. Ellickson (2013), 'Estimation of dynamic discrete choice models in continuous time with an application to retail competition', working paper, Duke University.

[11] Das, M. , 1992, A Micro-econometric Model of Capital Utilization and Retirement: The Case of the Cement Industry. Review of Economic Studies 59, 277-297.

[12] De Pinto, A., and G. Nelson, 2008, Land Use Change with Spatially Explicit Data: A Dynamic Approach. Environmental and Resource Economics. Forthcoming.

[13] Geweke, J., 1996, Monte Carlo simulation and numerical integration,in H. Amman, D. Kendrick, and J. Rust (eds.) Handbook of Computational Economics, chapter 15, pages 731-800. North-Holland. Amsterdam.

[14] Geweke, J. and M. Keane (2000): "Bayesian Inference for Dynamic Discrete Choice Models without the Need for Dynamic Programming," In book Simulation Based Inference and Econometrics: Methods and Applications, Mariano, Schuermann and Weeks (eds.), Cambridge University Press, 100-131.

[15] Hajivassiliou, V. and P. Ruud, 1994, Classical Estimation Methods for LDV Models Using Simulation, in The Handbook of Econometrics, Volume 4, D. McFadden and R. Engle (eds.). North-Holland: Amsterdam.

[16] Heckman, J. (1981): "The incidental parameters problem and the problem of initial conditions in estimating a discrete time - discrete data stochastic process," in C. Manski and D. McFadden (eds.), *Structural Analysis of Discrete Data with Econometric Applications*. MIT Press.

[17] Heckman, J. and S. Navarro, 2007, Dynamic Discrete Choice and Dynamic Treatment Effects. Journal of Econometrics 136, 341-396.

[18] Holmes, T., 2008, The Diffusion of Wal-Mart and Economies of Density. NBER Working Paper No. 13783.

[19] Hong, H. and M. Shum (2010): "Pairwise-Difference Estimation of a Dynamic Optimization Model," Review of Economic Studies, 77, 273-304.

[20] Hotz, J., and R.A. Miller (1993): "Conditional choice probabilities and the estimation of dynamic models," *Review of Economic Studies,* 60, 497-529.

[21] Hotz, J., R.A. Miller, S. Sanders, and J. Smith (1994): "A simulation estimator for dynamic models of discrete choice," *Review of Economic Studies,* 61, 265-89.

[22] Imai, S., N. Jain, and A. Ching (2009): "Bayesian Estimation of Dynamic Discrete Choice Models," Econometrica, 77(6), 1865-1899.

[23] Kasahara, H. (2009): "Temporary Increases in Tariffs and Investment: The Chilean Case," Journal of Business and Economic Statistics, 27(1), 113-127.

[24] Kasahara, H. and K. Shumitsu (2008a): "Pseudo-likelihood Estimation and Bootstrap Inference for Structural Discrete Markov Decision Models," Journal of Econometrics, 146(1), 2008, 92-106.

[25] Kasahara, H. and K. Shumitsu (2008b): "Nonparametric Identification of Finite Mixture Models of Dynamic Discrete Choices," Econometrica, 77(1), 2009, 135-175.

[26] Kasahara, H. and K. Shumitsu (2009): "Sequential Estimation of Structural Models with a Fixed Point Constraint," manuscript. University of Western Ontario.

[27] Kennet, M., 1993, Did Deregulation Affect Aircraft Engine Maintenance? An Empirical Policy Analysis. RAND Journal of Economics 24, 542-558.

[28] Kennet, M., 1994, A Structural Model of Aircraft Engine Maintenance. Journal of Applied Econometrics 9, 351-368.

[29] Kryukov, Y. (2010): "Dynamic R&D and the Effectiveness of Policy Intervention in the Pharmaceutical Industry," manuscript. Tepper School of Business. Carnegie Mellon University.

[30] Lanjow, J. (1998): "Patent Protection in the Shadow of Infringement: Simulation Estimations of Patent Value," Review of Economic Studies, 65, 671-710.

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

[32] Magnac, T. and D. Thesmar (2002): "Identifying Dynamic Discrete Decision Processes," Econometrica, 70(2), 801-816.

[33] Manski, C., 1993, Dynamic choice in social settings: Learning from the experience of others. Journal of Econometrics 58, 121-36.

[34] Miller, R., 1997, Estimating Models of Dynamic Optimization with Microeconomic Data. In H. Pesaran and P. Smidth (eds.) Handbook of Applied Econometrics: Microeconomics. Blackwell.

[35] Norets, A. (2009): "Inference in Dynamic Discrete Choice Models with Serially Correlated Unobserved State Variables," Econometrica, 77, 1665–1682.

[36] Pakes, A. (1986): "Patents as Options: Some Estimates of the Value of Holding European Patent Stocks," Econometrica, 54, 755-784.

[37] Pakes, A. (1994): "Dynamic structural models, problems and prospects," in C. Sims (ed.) Advances in Econometrics. Sixth World Congress, Cambridge University Press.

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

[39] Pesendorfer, M. (2002): "Retail Sales: A Study of Pricing Behavior in Supermarkets," Journal of Business, 75(1), pages 33-66.

[40] Rota, P., 2004, Estimating Labor Demand with Fixed Costs. International Economic Review 45, 25-48.

[41] Rust, J., 1987, Optimal replacement of GMC bus engines: An empirical model of Harold Zurcher. Econometrica 55, 999-1033.

[42] Rust, J., 1994a, Structural estimation of Markov decision processes. In R. E. Engle and McFadden (eds.) Handbook of Econometrics Volume 4, North-Holland. Amsterdam.

[43] Rust, J., 1994b, Estimation of dynamic structural models, problems and prospects: discrete decision processes, in C. Sims (ed.) Advances in Econometrics. Sixth World Congress, Cambridge University Press.

[44] Rust, J. and G. Rothwell, 1995, Optimal Response to a Shift in Regulatory Regime: The Case of the US Nuclear Power Industry. Journal of Applied Econometrics 10, S75-S118.

[45] Schankerman, M. (1998). How valuable is patent protection? Estimates by technology field. the RAND Journal of Economics, 77-107.

[46] Serrano, C. (2018): "Estimating the Gains from Trade in the Market for Patent Rights," International Economic Review, forthcoming.

[47] Slade, M. (1998): "Optimal Pricing with Costly Adjustment: Evidence from Retail Grocery Stores," Review of Economic Studies, 65, 87-108.

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

[49] Sturm, R., 1991, A Structural Economic Model of Operating Cycle Management in European Nuclear Power Plants. Manuscript, RAND Corporation.

[50] Wolpin, K., 1996, Public-Policy Uses of Discrete-Choice Dynamic Programming Models. American Economic Review 86, 427-32.

[51] Wolpin, K., 2007, Model Validation and Model Comparison Ex Ante Policy Evaluation, Structural Estimation, and Model Selection. American Economic Review 97, 48–52.

[52] Wooldridge, J. (2009): "On Estimating Firm-Level Production Functions Using Proxy Variables to Control for Unobservables," Economics Letters, 104, 112-114.

Structural Models of Dynamic Demand of Differentiated Products

1. Introduction

Consumers can stockpile a storable good when prices are low and use the stock for future consumption. This stockpiling behavior can introduce significant differences between short-run and long-run responses of demand to price changes. Also, the response of demand to a price change depends on consumers' expectations/beliefs about how permanent the price change is. For instance, if a price reduction is perceived by consumers as very transitory (e.g., a sales promotion), then a significant proportion of consumers may choose to increase purchases today, stockpile the product and reduce their purchases during future periods when the price will be higher. If the price reduction is perceived as permanent, this intertemporal substitution of consumer purchases will be much lower or even zero.

Ignoring consumers' stockpiling and forward-looking behavior can introduce serious biases in estimated own- and cross- price demand elasticities. These biases can be particularly serious when the time series of prices is characterized by "High-Low" pricing. The price fluctuates between a (high) regular price and a (low) promotion price. The promotion price is infrequent and last only few days, after which the price returns to its "regular" level. Most sales are concentrated in the very few days of promotion prices.

****

Figure. Hi-Lo pricing
Pesendorfer (*Journal of Business*, 2002)
****

Static demand models assume that all the substitution is either between brands or product expansion. They rule out intertemporal substitution. This can imply serious biases in the estimated demand elasticities. With High-Low pricing, we expect the static model to over-estimate the own-price elasticity. The bias in the estimated elasticities implies also a biased in the estimated Price Cost Margins (PCM). We expect PCMs to be underestimated. These biases have serious implications on policy analysis, such as merger analysis and antitrust cases.

Here we discuss two papers that have estimated dynamic structural models of demand of differentiated products using consumer level data (scanner data): Hendel and Nevo (Econometrica, 2006) and Erdem, Keane and Imai (QME, 2003). These papers extend microeconometric discrete choice models of product differentiation to a dynamic setting, and contains useful methodological contributions. Their empirical results show that ignoring the dynamics of demand can lead to serious biases. Also the papers illustrate how the use of **micro level data on household choices** (in contrast to only aggregate data on market shares) is key for credible identification of the dynamics of differentiated product demand.

2. Data and descriptive evidence

We assume that the researcher has access to consumer level data. Such data is widely available from several data collection companies and recently researchers in several countries have been able to gain access to such data for academic use. The data include the history of shopping behavior of a consumer over a period of one to three years. The researcher knows whether a store was visited, if a store was visited then which one, and what product (brand and size) was purchased and at what price. From the view point of the model, the key information that is not observed is consumer inventory and consumption decisions.

Hendel and Nevo use consumer-level scanner data from Dominicks, a supermarket chain that operates in the Chicago area. The dataset comes from 9 supermarket stores and it set covers the period June 1991 to June 1993. Purchases and price information is available in real (continuous) time but for the analysis in the paper it is aggregated at weekly frequency.

The dataset has two components: store-level and household-level data. **Store level data:** For each detailed product (brand–size) in each store in each week we observe the (average) price charged, (aggregate) quantity sold, and promotional activities. **Household level data:** For a sample of households, we observe the purchases of households at the 9 supermarket stores: supermarket visits and total expenditure in each visit; purchases (units and value) of detailed products (brand-size) in 24 different product categories (e.g., laundry detergent, milk, etc). The paper studies demand of laundry detergent products.

Table I in the paper presents summary statistics on household demographics, purchases, and store visits.

Table II in the paper presents the market shares of the main brands of laundry detergent in the data. The market is significantly concentrated, especially the market for Powder laundry detergent where the concentration ratios are $CR1 = 40\%$, $CR2 = 55\%$, and $CR3 = 65\%$. For most brands, the proportion of sales under a promotion price is important. However, this proportion varies importantly between brands, showing that different brands have different patterns of prices.

Descriptive evidence.   H&N present descriptive evidence which is consistent with household inventory holding. See also Hendel and Nevo (RAND, 2006). Though household purchase histories are observable, household inventories and consumption are unobservable. Therefore, empirical evidence on the importance of household inventory holding is indirect.

(a) Time duration since previous sale promotion has a positive effect on the aggregate quantity purchased.

(b) Indirect measures of storage costs (e.g., house size) are negatively correlated with households' propensity to buy on sale.

3. Model

3.1. Basic Assumptions. Consider a differentiated product, laundry detergent, with J different brands. Every week a household has some level of inventories of the product (that may be zero) and chooses (a) how much to consume from its inventory; and (b) how much to purchase (if any) of the product, and the brand to purchase.

An important simplifying assumption in Hendel-Nevo model is that consumers care about brand choice when they purchase the product, but not when they consume or store it. I explain below the computational advantages of this assumption. Of course, the assumption imposes some restrictions on the intertemporal substitution between brands, and I will discuss this point too. Erdem, Imai, and Keane (2003) do not impose that restriction.

The subindex t represents time, the subindex j represents a brand, and the subindex h represents a consumer or household. A household current utility function is:

$$u_h(c_{ht}, v_{ht}) - C_h(i_{h,t+1}) + m_{ht}$$

$u_h(c_{ht}, v_{ht})$ is the utility from consumption of the storable product, with c_{ht} being consumption and v_{ht} is a shock in the utility of consumption:

$$u_h(c_{ht}, v_{ht}) = \gamma_h \ \ln\left(c_{ht} + v_{ht}\right)$$

$C_h(i_{h,t+1})$ is the inventory holding cost, where $i_{h,t+1}$ is the level of inventory at the end of period t, after consumption and new purchases:

$$C_h(i_{h,t+1}) = \delta_{1h} \ i_{h,t+1} + \delta_{2h} \ i^2_{h,t+1}$$

m_{ht} is the indirect utility function from consumption of the composite good (outside good) plus the utility from brand choice (i.e., the utility function in a static discrete model of differentiated product):

$$m_{ht} = \sum_{j=1}^{J} \sum_{x=0}^{X} d_{hjxt} \left(\beta_h \ a_{jxt} - \alpha_h \ p_{jxt} + \xi_{jxt} + \varepsilon_{hjxt}\right)$$

$j \in \{1, 2, ..., J\}$ is the brand index. $x \in \{0, 1, 2, ..., X\}$ is the index of quantity choice, where the maximum possible size is X units. In this application $X = 4$. Brands with different sizes are standardized such that the same measurement unit is used in x. The variable $d_{hjxt} \in \{0, 1\}$ is a binary indicator for the event "household purchases x units of brand j at week t". p_{jxt} is the price of x units of brand j at period t. Note that the models allows for nonlinear pricing, i.e., for some brands and weeks p_{jxt} and $x * p_{j1t}$ can take different values. This is potentially important because the price data shows significant degree of nonlinear pricing. a_{jxt} is a vector of product characteristics other than price that is observable to the researcher. In this application, the most important variables in a_{jxt} are those that represent store-level advertising, e.g., display of the product in the store, etc. The variable ξ_{jxt} is a random variable that is unobservable to the researcher and that represents all the product characteristics which are known to consumers but not in the set of observable variables in the data.

α_h and β_h represent the marginal utility of income and the marginal utility of product attributes in a_{jxt}, respectively. As it is well-known in the empirical literature of demand of differentiated products, it is important to allow for heterogeneity in these marginal utilities in order to have demand systems with flexible and realistic own and cross elasticities or substitution patterns. Allowing for this heterogeneity is much simpler with consumer level data on product choices than with aggregate level data on product market shares. In particular, micro level datasets can include information on a rich set of household socioeconomic characteristics such as income, family size, age, education, gender, occupation, house-type, etc, that can be included as observable variables that determine the marginal utilities α_h and β_h. That is the approach in Hendel and Nevo's paper.

Finally, ε_{hjxt} is a consumer idiosyncratic shock that is independently and identically distributed over (h, j, x, t) with an extreme value type 1 distribution. This is the typical logit error that is included in most discrete models of demand of differentiated products. Note that while ε_{hjxt} vary over individuals, ξ_{jxt} do not.

Let $\mathbf{p}_t$ be the vector of product characteristics, observable or unobservable, for all the brands and sizes at period t:

$$\mathbf{p}_t \equiv \left\{ p_{jxt}, \, a_{jxt}, \, \xi_{jxt} : j = 1, 2, ..., J \text{ and } x = 1, 2, ..., X \right\}$$

Every week t, the household knows his level of inventories, i_{ht}, observes product attributes $\mathbf{p}_t$, and its idiosyncratic shocks in preferences, v_{ht} and ε_{ht}. Given this information, the household decides his consumption of the storable product, c_{ht}, and how much to purchase and which product, $d_{ht} = \{d_{hjxt}\}$. The household makes this decision to maximize his expected and

discounted stream of current and future utilities,

$$E_t \left(\sum_{s=0}^{\infty} \delta^s \left[u_h(c_{ht+s}, v_{ht+s}) - C_h(i_{h,t+s+1}) + m_{ht+s} \right] \right)$$

where δ is the discount factor.

The vector of state variables of this DP problem is $\{i_{ht}, v_{ht}, \varepsilon_{ht}, \mathbf{p}_t\}$. The decision variables are c_{ht} and d_{ht}. To complete the model we need to make some assumptions on the stochastic processes of the state variables. The idiosyncratic shocks v_{ht} and ε_{ht} are assumed iid over time. The vector of product attributes $\mathbf{p}_t$ follows a Markov processes. Finally, consumer inventories i_{ht} has the obvious transition rule:

$$i_{h,t+1} = i_{h,t+1} - c_{ht} + \left(\sum_{j=1}^{J} \sum_{x=0}^{X} d_{hjxt} \ x \right)$$

where $\sum_{j=1}^{J} \sum_{x=0}^{X} d_{hjxt} \ x$ represents the units of the product purchased by household h at period t.

Let $V_h(\mathbf{s}_{ht})$ be the value function of a household, where $\mathbf{s}_{ht}$ is the vector of state variables $(i_{ht}, v_{ht}, \varepsilon_{ht}, \mathbf{p}_t)$. A household decision problem can be represented using the Bellman equation:

$$V_h(\mathbf{s}_{ht}) = \max_{\{c_{ht}, d_{ht}\}} \left[u_h(c_{ht}, v_{ht}) - C_h(i_{h,t+1}) + m_{ht} + \delta \ E\left(V_h(\mathbf{s}_{ht+1}) \mid \mathbf{s}_{ht}, \ c_{ht}, \ d_{ht} \right) \right]$$

where the expectation $E\left(. \mid \mathbf{s}_{ht}, \ c_{ht}, \ d_{ht}\right)$ is over the distribution of $\mathbf{s}_{ht+1}$ conditional on $(\mathbf{s}_{ht}, c_{ht}, d_{ht})$. The solution of this DP problem implies optimal decision rules for consumption and purchasing decisions: $c_{ht} = c_h^*(\mathbf{s}_{ht})$ and $d_{ht} = d_h^*(\mathbf{s}_{ht})$ where $c_h^*(.)$ and $d_h^*(.)$ are the decision rules. Note that they are household specific because there is time-invariant household heterogeneity in the marginal utility of product attributes $(\alpha_h$ and $\beta_h)$, in the utility of consumption of the storable good u_h, and in inventory holding costs, C_h.

The optimal decision rules $c_h^*(.)$ and $d_h^*(.)$ depend also on the structural parameters of the model: the parameters in the utility function, and in the transition probabilities of the state variables. In principle, we could use the equations $c_{ht} = c_h^*(\mathbf{s}_{ht})$ and $d_{ht} = d_h^*(\mathbf{s}_{ht})$ and our data on (some) decision and state variables to estimate the parameters of the model. To apply this revealed preference approach, there are three main issues we have to deal with.

First, the dimension of the state space of $\mathbf{s}_{ht}$ is extremely large. In most applications of demand of differentiated products, there are dozens (or even more than a hundred) products. Therefore, the vector of product attributes $\mathbf{p}_t$ contains more than a hundred continuous state variables. Solving a DP problem with this state space, or even approximating the solution with enough accuracy using Monte Carlo simulation methods, is computationally very demanding even with the most sophisticated computer equipment. We will see how Hendel and Nevo propose and implement a method to reduce the dimension of the state space. The method is based on some assumptions that we discuss below.

Second, though we have good data on households purchasing histories, information on households' consumption and inventories of storable goods is very rare. In this application, consumption and inventories, c_{ht} and i_{ht}, are unobservable to the researchers. Not observing inventories is particularly challenging. This is the key state variable in a dynamic demand model of demand of a storable good. We will discuss below the approach used by Hendel and Nevo to deal with this issue, and also the approach used by Erdem, Imai, and Keane (2003).

And third, as usual in the estimation of a model of demand, we should deal with the endogeneity of prices. Of course, this problem is not specific of a dynamic demand model. However, dealing with this problem may not be independent of the other issues mentioned above.

3.2. Reducing the dimension of the state space. Given that the state variables $(v_{ht}, \varepsilon_{ht})$ are independently distributed over time, it is convenient to reduce the dimension of this DP problem by using a value function that is integrated over these iid random variables. The integrated value function is defined as:

$$\bar{V}_h(i_{ht}, \mathbf{p}_t) \equiv \int V_h(\mathbf{s}_{ht}) \; dF_\varepsilon(\varepsilon_{ht}) \; dF_v(v_{ht})$$

where F_ε and F_v are the CDFs of ε_{ht} and v_{ht}, respectively. Associated with this integrated value function there is an integrated Bellman equation. Given the distributional assumptions on the shocks ε_{ht} and v_{ht}, the integrated Bellman equation is:

$$\bar{V}_h(i_{ht}, \mathbf{p}_t) = \max_{c_{ht}, d_{ht}} \int \ln \left(\sum_{j=1}^{J} \exp \left\{ \begin{array}{l} u_h(c_h, v_{ht}) - C_i(i_{ht+1}) + m_{ht} \\[6pt] + \delta \; \mathbb{E} \left[\bar{V}_h(i_{ht+1}, \mathbf{p}_{t+1}) \mid i_{ht}, \mathbf{p}_t, c_{ht}, d_{ht} \right] \end{array} \right\} \right) dF_v(v_{ht}).$$

This Bellman equation is also a contraction mapping in the value function. The main computational cost in the computation of the functions $\bar{V}_h$ comes from the dimension of the vector of product attributes $\mathbf{p}_t$. We now explore ways to reduce this cost.

First, note that the assumption that there is only one inventory, the aggregate inventory of all the products, and not one inventory for each brand, $\{i_{hjt}\}$, has already reduced importantly the dimension of the state space. This assumption not only reduces the state space but, as we see below, it also allows us to modify the dynamic problem, which can significantly aid in the estimation of the model.

Taken literally, this assumption implies that there is no differentiation in consumption: the product is homogenous in use. Note, that through ξ_{jxt} and ε_{ijxt} the model allows differentiation in purchase, as is standard in the IO literature. It is well known that this differentiation is needed to explain purchasing behavior. This seemingly creates a tension in the model: products are differentiated at purchase but not in consumption. Before explaining

how this tension is resolved we note that the tension is not only in the model but potentially in reality as well. Many products seem to be highly differentiated at the time of purchase but its hard to imagine that they are differentiated in consumption. For example, households tend to be extremely loyal to the laundry detergent brand they purchase – a typical household buys only 2-3 brands of detergent over a very long horizon – yet its hard to imagine that the usage and consumption are very different for different brands.

A possible interpretation of the model that is consistent with product differentiation in consumption is that the variables ξ_{jxt} not only captures instantaneous utility at period t but also the discounted value of consuming the x units of brand j. This is a valid interpretation if brand-specific utility in consumption is additive such that it does not affect the marginal utility of consumption.

This assumption has some implications that simplify importantly the structure of the model. It implies that the optimal consumption does not depend on which brand is purchased, only on the size. And relatedly, it implies that the brand choice can be treated as a static decision problem.

We can distinguish two components in the choice d_{ht}: the quantity choice, x_{ht}, and the brand choice j_{ht}. Given $x_{ht} = x$, the optimal brand choice is:

$$j_{ht} = \arg\max_{j \in \{1,2,\dots,J\}} \left\{ \beta_h\, a_{jxt} - \alpha_h\, p_{jxt} + \xi_{jxt} + \varepsilon_{hjxt} \right\}$$

Then, given our assumption about the distribution of ε_{hjxt}, the component m_{ht} of the utility function can be written as $m_{ht} = \sum_{x=0}^{X} \omega_h(x, \mathbf{p}_t) + e_{ht}$ where $\omega_{ht}(x, \mathbf{p}_t)$ is the inclusive value:

$$\omega_h(x, \mathbf{p}_t) \equiv E\left(\max_{j \in \{1,2,\dots,J\}} \left\{ \beta_h\, a_{jxt} - \alpha_h\, p_{jxt} + \xi_{jxt} + \varepsilon_{hjxt} \right\} \mid x_{ht} = x,\ \mathbf{p}_t \right)$$

$$= \ln\left(\sum_{j=1}^{J} \exp\left\{ \beta_h\, a_{jxt} - \alpha_h\, p_{jxt} + \xi_{jxt} \right\} \right)$$

and e_{ht} does not depend on size x (or on inventories and consumption), and therefore we can ignore this variable for the dynamic decisions on size and consumption.

Therefore, the dynamic decision problem becomes:

$$\bar{V}_h(i_{ht}, \mathbf{p}_t) = \max_{c_{ht}, x_{ht}} \int \left\{ u_h(c_{ht}, v_{ht}) - C_i(i_{ht+1}) + \omega_h(x, \mathbf{p}_t) + \delta\, \mathbb{E}\left[\bar{V}_h(i_{ht+1}, \mathbf{p}_{t+1}) \mid i_{ht+1}, \mathbf{p}_t \right] \right\} dF_v(v_{ht})$$

In words, the problem can now be seen as a choice between sizes, each with a utility given by the size-specific inclusive value (and extreme value shock). The dimension of the state space is still large and includes all product attributes, because we need these attributes to compute the evolution of the inclusive value. However, in combination with additional assumptions the modified problem is easier to estimate.

Note also, that expression that describes the optimal brand choice, $j_{ht} = \arg\max_{j \in \{1,2,...,J\}}$ $\{\beta_h \, a_{jxt} - \alpha_h \, p_{jxt} + \xi_{jxt} + \varepsilon_{hjxt}\}$ is a "standard" multinomial logit model with the caveat that prices are endogenous explanatory variables because they depend on the unobserved attributes in ξ_{jxt}. We describe below how to deal with this endogeneity problem. With household level data, dealing with the endogeneity of prices is much simpler than with aggregate data on market shares. More specifically, we do not need to use Monte Carlo simulation techniques, or an iterative algorithm to compute the "average utilities" $\{\delta_{jxt}\}$.

To reduce the dimension of the state space, Hendel and Nevo (2006) introduce the following assumption. Let $\boldsymbol{\omega}_h(\mathbf{p}_t)$ be the vector with the inclusive values for every possible size $\{\omega_h(x, \mathbf{p}_t) : x = 1, 2, ..., X\}$.

Assumption: The vector $\boldsymbol{\omega}_h(\mathbf{p}_t)$ is a sufficient statistic of the information in $\mathbf{p}_t$ that is useful to predict $\boldsymbol{\omega}_h(\mathbf{p}_{t+1})$:

$$\Pr(\boldsymbol{\omega}_h(\mathbf{p}_{t+1}) \mid \mathbf{p}_t) = \Pr(\boldsymbol{\omega}_h(\mathbf{p}_{t+1}) \mid \boldsymbol{\omega}_h(\mathbf{p}_t))$$

In words, the vector $\boldsymbol{\omega}_h(\mathbf{p}_t)$ contains all the relevant information in $\mathbf{p}_t$ to obtain the probability distribution of $\boldsymbol{\omega}_h(\mathbf{p}_{t+1})$ conditional on $\mathbf{p}_t$. Instead of all the prices and attributes, we only need a single index for each size. Two vectors of prices that yield the same (vector of) current inclusive values imply the same distribution of future inclusive values. This assumption is violated if individual prices have predictive power above and beyond the predictive power of $\boldsymbol{\omega}_h(\mathbf{p}_t)$.

The inclusive values can be estimated outside the dynamic demand model. Therefore, the assumption can be tested and somewhat relaxed by including additional statistics of prices in the state space. Note, that $\boldsymbol{\omega}_h(\mathbf{p}_t)$ is consumer specific: different consumers value a given set of products differently and therefore this assumption does not further restrict the distribution of heterogeneity.

Given this assumption, the integrated value function is $\bar{V}_h(i_{ht}, \boldsymbol{\omega}_{ht})$ that includes only $X + 1$ variables, instead of $3 * J * X + 1$ state variables.

4. Estimation

4.1. Estimation of brand choice. Let j_{ht} represent the brand choice of household h at period t. Under the assumption that there is product differentiation in purchasing but not in consumption or in the cost of inventory holding, a household brand choice is a static decision problem. Given $x_{ht} = x$, with $x > 0$, the optimal brand choice is:

$$j_{ht} = \arg\max_{j \in \{1,2,...,J\}} \left\{ \beta_h \, a_{jxt} - \alpha_h \, p_{jxt} + \xi_{jxt} + \varepsilon_{hjxt} \right\}$$

The estimation of demand models of differentiated products, either static or dynamic, should deal with two important issues. First, the endogeneity of prices. The model implies that

p_{jxt} depends on observed and unobserved products attributes, and therefore p_{jxt} and ξ_{jxt} are not independently distributed. The second issue, is that the model should allow for rich heterogeneity in consumers marginal utilities of product attributes, β_h and α_h. Using consumer-level data (instead of aggregate market share data) facilities significantly the econometric solution of these issues.

Consumer-level scanner datasets contain rich information on household socioeconomic characteristics. Let z_h be a vector of observable socioeconomic characteristics that have a potential effect on demand, e.g., income, family size, age distribution of children and adults, education, occupation, type of housing, etc. We assume that β_h and α_h depend on this vector of household characteristics:

$$\beta_h = \beta_0 + (z_h - \bar{z})\sigma_\beta$$

$$\alpha_h = \alpha_0 + (z_h - \bar{z})\sigma_\alpha$$

β_0 and α_0 are scalar parameters that represent the marginal utility of advertising and income, respectively, for the average household in the sample. $\bar{z}$ is the vector of household attributes of the average household in the sample. And σ_β and σ_α are $K \times 1$ vectors of parameters that represent the effect of household attributes on marginal utilities. Therefore, the utility of purchasing can be written as:

$$\left[\beta_0 + (z_h - \bar{z})\sigma_\beta\right] a_{jxt} - \left[\alpha_0 + (z_h - \bar{z})\sigma_\alpha\right] p_{jxt} + \xi_{jxt} + \varepsilon_{hjxt}$$

$$= \left[\beta_0\, a_{jxt} - \alpha_0\, p_{jxt} + \xi_{jxt}\right] + (z_h - \bar{z}) \left[a_{jxt}\, \sigma_\beta - p_{jxt}\, \sigma_\alpha\right] + \varepsilon_{hjxt}$$

$$= \delta_{jxt} + (z_h - \bar{z})\, \sigma_{jxt} + \varepsilon_{hjxt}$$

where $\delta_{jxt} \equiv \beta_0\, a_{jxt} - \alpha_0\, p_{jxt} + \xi_{jxt}$, and $\sigma_{jxt} \equiv a_{jxt}\, \sigma_\beta - p_{jxt}\, \sigma_\alpha$. δ_{jxt} is a scalar that represents the utility of product (j, x, t) for the average household in the sample. σ_{jxt} is a vector and each element in this vector represents the effect of a household attribute on the utility of product (j, x, t).

In fact, it is possible to allow also for interactions between the observable household attributes and the unobservable product attributes, to have a term $\lambda_h \xi_{jxt}$ where $\lambda_h = 1 + (z_h - \bar{z})\sigma_\lambda$. With this more general specification, we still have that $\delta_{jxt} \equiv \beta_0\, a_{jxt} - \alpha_0\, p_{jxt} + \xi_{jxt}$, but now $\sigma_{jxt} \equiv a_{jxt}\, \sigma_\beta - p_{jxt}\, \sigma_\alpha + \xi_{jxt}\sigma_\lambda$.

4.1.1. *Dummy-Variables Maximum Likelihood + IV estimator.* Given this representation of the brand choice model, the probability that a household with attributes z_h purchases brand j at period t given that he buys x units of the product is:

$$P_{hjxt} = \frac{\exp\left\{\delta_{jxt} + (z_h - \bar{z})\, \sigma_{jxt}\right\}}{\sum_{k=1}^{J} \exp\left\{\delta_{kxt} + (z_h - \bar{z})\, \sigma_{kxt}\right\}}$$

Given a sample with a large number of households, we can estimate δ_{jxt} and σ_{jxt} for every (j, x, t) in a multinomial logit model with probabilities $\{P_{hjxt}\}$. For instance, we can estimate these "incidental parameters" δ_{jxt} and σ_{jxt} separately for every value of (x, t). For $(t = 1, x = 1)$ we select the subsample of households in sample who purchase $x = 1$ unit of the product at week $t = 1$. Using this subsample, we estimate the vector of $J(K + 1)$ parameters $\{\delta_{j11}, \sigma_{j11} : j = 1, 2, ..., J\}$ by maximizing the multinomial log-likelihood function:

$$\sum_{h=1}^{H} 1\{x_{h1} = 1\} \sum_{j=1}^{J} 1\{j_{h1} = j\} \ln P_{hj11}$$

We can proceed in the same way to estimate all the parameters $\{\delta_{jxt}, \sigma_{jxt}\}$.

This estimator is consistent as H goes to infinity for fixed T, X, and J. For a given (finite) sample, there are some requirements on the number of observations in order to be able to estimate the incidental parameters. For every value of (x, t), the number of incidental parameters to estimate is $J(K + 1)$, and the number of observations is equal to the number of households who purchase x units at week t, i.e., $H(x, t) = \sum_{h=1}^{H} 1\{x_{ht} = x\}$. We need that $H(x, t) > J(K + 1)$. For instance, with $J = 25$ products and $K = 4$ household attributes, we need $H(x, t) > 125$ for every week t and every size x. We may need a very large number of households H in the sample in order to satisfy these conditions. An assumption that may eliminate this problem is that the utility from brand choice is proportional to quantity: $x(\beta_h \ a_{jt} - \alpha_h \ p_{jt} + \xi_{jt} + \varepsilon_{hjt})$. Under this assumption, we have that for every week t, the number of incidental parameters to estimate is $J(K + 1)$, but the number of observations is now equal to the number of households who purchase any quantity $x > 0$ at week t, i.e., $H(t) = \sum_{h=1}^{H} 1\{x_{ht} > 0\}$. We need that $H(t) > J(K + 1)$ which is a much weaker condition.

Given estimates of the incidental parameters, $\{\hat{\delta}_{jxt}, \hat{\sigma}_{jxt}\}$, now we can estimate the structural parameters β_0, α_0, σ_β, and σ_α using an IV (or GMM) method. For the estimation of β_0 and α_0, we have that:

$$\hat{\delta}_{jxt} = \beta_0 \ a_{jxt} - \alpha_0 \ p_{jxt} + \xi_{jxt} + e_{jxt}$$

where e_{jxt} represents the estimation error $(\hat{\delta}_{jxt} - \delta_{jxt})$. This is a linear regression where the regressor p_{jxt} is endogenous. We can estimate this equation by IV using the so-called "BLP instruments", i.e., the characteristics other than price of products other than j, $\{a_{kxt} : k \neq j\}$. Of course, there are other approaches to deal with the endogeneity of prices in this equation. For instance, we could consider the following Error-Component structure in the endogenous part of the error term: $\xi_{jxt} = \xi_{jx}^{(1)} + \xi_{jxt}^{(2)}$ where $\xi_{jxt}^{(2)}$ is assumed not serially correlated. Then, we can control for $\xi_{jx}^{(1)}$ using product-size dummies, and use lagged values of prices and other product attributes to deal with the endogeneity of prices that comes from the correlation with the transitory shock $\xi_{jxt}^{(2)}$.

For the estimation of σ_β, and σ_α, we have the system of equations:

$$\hat{\sigma}_{jxt} = a_{jxt}\sigma_\beta - p_{jxt}\,\sigma_\alpha + \xi_{jxt}\sigma_\lambda + e_{jxt}$$

where e_{jxt} represents the estimation error $(\hat{\sigma}_{jxt} - \sigma_{jxt})$. We have one equation for each household attribute. We can estimate each of these equations using the same IV procedure as for the estimation of β_0 and α_0.

Once we have estimated $(\beta_0, \alpha_0, \sigma_\beta, \sigma_\alpha)$, we can also obtain estimates of ξ_{jxt} as residuals from the estimated equation. We can get also consistent estimates of the marginal utilities β_h and α_h as:

$$\hat{\beta}_h = \hat{\beta}_0 + (z_h - \bar{z})\hat{\sigma}_\beta$$

$$\hat{\alpha}_h = \hat{\alpha}_0 + (z_h - \bar{z})\hat{\sigma}_\alpha$$

Finally, we can get estimates of the inclusive values:

$$\hat{\omega}_{hxt} = \ln\left(\sum_{j=1}^{J}\exp\left\{\hat{\beta}_h\,a_{jxt} - \hat{\alpha}_h\,p_{jxt} + \hat{\xi}_{jxt}\right\}\right)$$

4.1.2. *Control function approach.* The previous approach, though simple, has the limitation that we need to have, for every week in the sample, a large enough number of households making positive purchases. That requirement is not needed for identification of the parameters. It is only needed for the implementation of the simple two-step dummy variables approach to deal with the endogeneity of prices.

When our sample does not satisfy that requirement, there is other simple method that we can use. This method is a control function approach that is in the spirit of the methods proposed by Rivers and Vuong (Journal of Econometrics, 1988), Blundell and Powell (REStud, 2004), and in the specific context of demand of differentiated products, Petrin and Train (Journal of Marketing Research, 2010).

If firms choose prices to maximize profits, we expect that prices depend on the own product characteristics and also on the characteristics of competing products: $p_{jxt} = f_{jxt}(a_t, \xi_t)$, where $a_t = \{a_{jxt} : \text{for any } j, x\}$, and $\xi_t = \{\xi_{jxt} : \text{for any } j, x\}$. Define the conditional mean function:

$$g^p_{jx}(a_t) \equiv E(p_{jxt} \mid a_t) = E(f_{jxt}(a_t, \xi_t) \mid a_t)$$

Then, we can write the regression equation:

$$p_{jxt} = g^p_{jx}(a_t) + e_{jxt}$$

where the error term e_{jxt} is by construction mean independent of a_t.

The first step of the control function method consists in the estimation of the conditional mean functions $g^p_{jx}(.)$ for every brand and size (j, x). Though we have a relatively large number of weeks in our dataset (more than 100 weeks in most scanner datasets), the number

of variables in the vector a_t is $J * X$, that is a pretty large number. Therefore, we need to impose some restrictions on how the exogenous product characteristics in a_t affect prices. For instance, we may assume that,

$$g_{jx}^p(a_t) = g_{jx}^p \left(a_{jxt}, \ \bar{a}_{j(-x)t}, \ \bar{a}_{(-j)xt}, \ \bar{a}_{(-jx)t} \right)$$

where $\bar{a}_{j(-x)t}$ is the sample mean of variable a at period t for all the products of brand j but with different size than x; $\bar{a}_{(-j)xt}$ is the sample mean for all the products with size x but with brand different than j; and $\bar{a}_{(-jx)t}$ is the sample mean for all the products with size different than x and brand different than j. Of course, we can consider more flexible specifications but still with a number of regressors much smaller than $J * X$.

The second step of the method is based on a decomposition of the error term ξ_{jxt} in two components: an endogenous that is a deterministic function of the error terms in the first step, $e_t \equiv \{e_{jxt} :$ for any j and $x\}$, and an "exogenous" component that is independent of the price p_{jxt} once we have controlled for e_{jxt}. Define the conditional mean function:

$$g_{jx}^\xi(e_t) \equiv E(\xi_{jxt} \mid e_t)$$

Then, we can write ξ_{jxt} as the sum of two components, $\xi_{jxt} = g_{jx}^\xi(e_t) + v_{jxt}$. By construction, the error term v_{jxt} is mean independent of e_t. But then, v_{jxt} is mean independent of all the product prices because prices depend only on the exogenous product characteristics a_t (that by assumption are independent of ξ_{jxt}) and on the "residuals" e_t (that by construction are mean independent of v_{jxt}). Then, we can write the utility of product (j, x) as:

$$\beta_h \ a_{jxt} - \alpha_h \ p_{jxt} + g_{jx}^\xi(e_t) + (v_{jxt} + \varepsilon_{hjxt})$$

The term $g_{jx}^\xi(e_t)$ is the control function.

Under the assumption that $(v_{jxt} + \varepsilon_{hjxt})$ is iid extreme value type 1 distributed, we have that the brand choice probabilities conditional on $x_{ht} = x$ are:

$$P_{hjxt} = \frac{\exp\left\{\beta_0 a_{jxt} - \alpha_0 p_{jxt} + a_{jxt}(z_h - \bar{z})\sigma_\beta - p_{jxt}(z_h - \bar{z})\sigma_\alpha + g_{jx}^\xi(e_t)\right\}}{\sum_{k=1}^J \exp\left\{\beta_0 a_{kxt} - \alpha_0 p_{kxt} + a_{kxt}(z_h - \bar{z})\sigma_\beta - p_{kxt}(z_h - \bar{z})\sigma_\alpha + g_{kx}^\xi(e_t)\right\}}$$

where the control functions $\{g_{jx}^\xi(e_t)\}$ consists of a brand dummies and polynomial in the residual variables $\{e_{jxt} : j = 1, 2, ..., J\}$. Then, we can estimate $(\beta_0, \alpha_0, \sigma_\beta, \sigma_\alpha)$ and the parameters of the control function by using Maximum Likelihood in this multinomial logit model. The log-likelihood function is:

$$\ell(\theta) = \sum_{h=1}^H \sum_{t=1}^T \sum_{x=1}^X \sum_{j=1}^J 1\{x_{ht} = x \ , \ j_{ht} = j\} \ \ln P_{hjxt}$$

As in the previous method, once we have estimated these parameters, we can construct consistent estimates of the inclusive values ω_{hxt}.

4.2. Estimation of quantity choice. As mentioned above, the lack of data on house-hold inventories is a challenging econometric problem because this is a key state variable in a dynamic demand model of demand of a storable good. Also, this is not a "standard" unobservable variable in the sense that it follows a stochastic process that is endogenous. That is, not only inventories affect purchasing decision, but also purchasing decisions affect the evolution of inventories.

The approach used by Erdem, Imai, and Keane (2003) to deal with this problem is to assume that household inventories is a (deterministic) function of "number of weeks (duration) since last purchase", T_{ht}, and the quantity purchased in the last purchase, x_{ht}^{last}:

$$i_{ht} = f_h(x_{ht}^{last}, T_{ht})$$

In general, this assumption holds under two conditions: (1) consumption is deterministic; and (2) when a new purchase is made, the existing inventory at the beginning of the week is consumed or scrapped. For instance, suppose that these conditions hold and that the level of consumption is constant $c_{ht} = c_h$. Then,

$$i_{ht+1} = \max\left\{ 0 \; ; \; x_{ht}^{last} - c_h \, T_{ht} \right\}$$

The constant consumption can be replace by a consumption rate that depends on the level of inventories. For instance, $c_{ht} = \lambda_h i_{ht}$. Then:

$$i_{ht+1} = \max\left\{ 0 \; ; \; (1 - \lambda_h)^{T_{ht}} \, x_{ht}^{last} \right\}$$

Using this approach, the state variable i_{ht} should be replaced by the state variables (x_{ht}^{last}, T_{ht}), but the rest of the features of the model remain the same. The parameters c_h or λ_h can be estimated together with the rest of parameters of the structural model. Also, we may not need to solve for the optimal consumption decision.

There is no doubt that using observable variables to measure inventories is very useful for the estimation of the model and for identification. It also provides a more intuitive interpretation of the identification of the model.

The individual level data provide the probability of purchase conditional on current prices, and past purchases of the consumer (amounts purchased and duration from previous purchases): $\Pr(x_{ht}|x_{ht}^{last}, T_{ht}, \mathbf{p}_t)$. Suppose that we see that this probability is not a function of past behavior (x_{ht}^{last}, T_{ht}), we would then conclude that dynamics are not relevant and that consumers are purchasing for immediate consumption and not for inventory. On the other hand, if we observe that the purchase probability is a function of past behavior, and we assume that preferences are stationary then we conclude that there is dynamic behavior.

Regarding the identification of storage costs, consider the following example. Suppose we observe two consumers who face the same price process and purchase the same amount over

a relatively long period. However, one of them purchases more frequently than the other. This variation leads us to conclude that this consumer has higher storage costs. Therefore, the storage costs are identified from the average duration between purchases.

Hendel and Nevo use a different approach, though the identification of their model is based on the same intuition.

5. Empirical Results

To Be Completed

6. Dynamic Demand of Differentiated Durable Products

- Gowrisankaran and Rysman (2009). TBW

CHAPTER 9

Empirical Dynamic Games of Oligopoly Competition

1. Introduction

The last three lectures of the course deal with methods and applications of empirical dynamic games of oligopoly competition. More generally, some of the methods that I will describe can be applied to estimate dynamic games in other applied fields such as political economy (e.g., competition between political parties), or international economics (e.g., ratification of international treaties), among others.

Dynamic games are powerful tools for the analysis of phenomena characterized by **dynamic strategic interactions** between multiple agents. By *dynamic strategic interactions* we mean that:

> **(a)** players' current decisions affect their own and other players' payoffs in the future (i.e., *multi-agent dynamics*);

> **(b)** players' decisions are forward looking in the sense that they take into account the implications on their own and on their rivals' future behavior and how this behavior affects future payoffs (i.e., *strategic behavior*).

Typical sources of dynamic strategic interactions are decisions that are partially irreversible (costly to reverse) or that involve sunk costs. Some examples in the context of firm oligopoly competition: **(1)** entry-exit in markets; **(2)** introduction of a new product; timing of the release of a new movie; **(3)** reposition of product characteristics; **(4)** investment in capacity, or equipment, or R&D, or quality, or goodwill, or advertising; **(5)** pricing of a durable good; pricing when demand is characterized by consumer switching costs; **(6)** production when there is learning-by-doing.

Taking into account dynamic strategic interactions may change substantially our interpretation of some economic phenomena or the implications of some public policies. We have already discussed some examples from recent applications in IO: **(1)** Short-run and long-run responses to changes in industry regulations (Ryan, 2006); **(2)** Product repositioning in differentiated product markets (Sweeting, 2007); **(3)** Dynamic aspects of network competition (Aguirregabiria and Ho, 2008).

Road Map: 1.   Structure of empirical dynamic games; 2.   Identification; 3. Estimation; 4.   Dealing with unobserved heterogeneity; 5.   Empirical Applications; 5.1. Dynamic effects of industry regulation (Ryan, 2006);   5.2. Product repositioning in differentiated product markets (Sweeting, 2007);   5.3. Dynamic aspects of network competition (Aguirregabiria and Ho, 2008).

2. Dynamic version of Bresnahan-Reiss model

Based on Bresnahan and Reiss (AES, 1994)

Complete information; homogeneous firms

2.1. Motivation. Suppose that we have panel data of M markets over T periods of time.

$$\text{Data} = \{ \ n_{mt}, \ X_{mt} : m = 1, 2, ..., M; \ t = 1, 2, ..., T \ \}$$

In these data, we observe how in market the number of firms grow or decline. Suppose that we do not know the gross changes in the number of firms, i.e., we do not observe the number of new entrants, en_{mt}, and number of exits, ex_{mt}. We only observe the net change $n_{mt} - n_{mt-1} = en_{mt} - ex_{mt}$.

To explain the observed variation, across markets and over time, in the number of firms, we could propose and estimate the BR static model that we have considered so far. The only difference is that now we have multiple realizations of the game both because the game is played at different locations and because it is played at different periods of time.

However, the static BR model imposes a strong and unrealistic restriction on this type of panel data. According to the static model, the number of firms at previous period, n_{mt-1}, does not play any role in the determination of the current number of firms n_{mt}. This is because the model considers that the profit of an active firm is the same regardless it was active at previous period or not. That is, the model assumes that either there are not entry costs, or that entry costs are paid every period the firm is active such that both new entrants and incumbents should pay these costs. Of course, this assumption is very unrealistic for most industries.

Bresnahan and Reiss (AES, 1994) propose and estimate a dynamic extension of their static model of entry. This dynamic model distinguishes between incumbents and potential entrants and takes into account the existence of sunk entry costs. The model is simple but interesting and useful because its own simplicity. We could call it a "semi-structural" model. It is structural in the sense that is fully consistent with dynamic game of entry-exit in an oligopoly industry. But it is only "semi" in the sense that it does not model explicitly how the future expected value function of an incumbent firm depends on the sunk-cost. Ignoring this relationship has clear computational advantages in the estimation of the model, that is

very simple. It has also limitations in terms of the type of the counterfactuals and empirical questions that can be studied using this model.

2.2. Model. Let n_t be the number of firms active in the market at period t. n_t belongs to the set $\{0, 1, ..., N\}$ where N is a large but finite number. Let $V(n_t, X_t) - \varepsilon_t$ be the value function of an active firm in a market with exogenous characteristics (X_t, ε_t) and number of firms n_t. The additive error term ε_t can be interpreted as an iid shock in the fixed cost of being active in the market. The function $V(n, X)$ is strictly decreasing in n.

This value function does not include the cost of entry. Let EC be the entry cost that a new entrant should pay to be active in the market at period t. And let SV be the scrapping value of a firm that decides to exit from the market. For the moment, we consider that EC and SV are constant parameters but I will discuss later how this assumption can be relaxed.

An important and obvious condition is that $SV \leq EC$. That is, firms cannot make profits by constantly entering and exiting in a market. It is an obvious arbitrage condition. The parameter $EC - SV$ is called the sunk entry cost, i.e., it is the part of the entry cost that is sunk and cannot be recovered upon exit. For instance, administrative costs, costs of market research, and in general any investment in capital that is firm specific and therefore will not have market value when the firm exits the market.

The values or payoffs of incumbents and potential entrants are: Incumbent that decides to stay: $V(n_t, X_t) - \varepsilon_t$; Incumbent that exits: SV; New entrant: $V(n_t, X_t) - \varepsilon_t - EC$; Potential entrant stays out: 0.

Now, I describe the entry-exit equilibrium conditions that determine the equilibrium number of firms n_t as a function of (X_t, ε_t).

Regime 1: Exit. Suppose that $n_{t-1} > 0$ and $V(n_{t-1}, X_t) - \varepsilon_t < SV$. That is, at the beginning of period t, the values of the exogenous variables X_t and ε_t are realized, the incumbent firms at previous period find out that the value of being active in the market is smaller than the scrapping value of the firm. Therefore, these firms want to exit.

It should be clear that under this regime there is not entry. Since $SV \leq EC$, we have that $V(n_{t-1}, X_t) - \varepsilon_t < EC$ and therefore $V(n_{t-1} + 1, X_t) - \varepsilon_t < EC$. The value for a new entrant is smaller than the entry cost and therefore there is not entry.

Therefore, incumbent firms will start quitting the market up to the point when: either (a) there are no more firms in the market, i.e., $n_t = 0$; or (b) there are still firms in the market and the value of an active firm is greater or equal the scrapping value. The equilibrium

number of firms in this regime is given by the conditions:

$$
\begin{cases}
n_t = 0 & \text{if } V(1, X_t) - \varepsilon_t < SV \\
\text{OR} \\
n_t = n > 0 & \text{if } \{V(n, X_t) - \varepsilon_t \geq SV\} \quad \text{AND} \quad \{V(n+1, X_t) - \varepsilon_t < SV\}
\end{cases}
$$

The condition $\{V(n_t, X_t) - \varepsilon_t \geq SV\}$ says that an active firm in the market does not want to exit. Condition $\{V(n_t + 1, X_t) - \varepsilon_t < SV\}$ establishes that if there were any number of firms in the market greater than n_t, firms would prefer to exit.

Summarizing, **Regime 1 [Exit]** is described by the following condition on exogenous variables $\{n_{t-1} > 0\}$ and $\{\varepsilon_t > V(n_{t-1}, X_t) - SV\}$, and this condition implies that:

$$n_t < n_{t-1}$$

and n_t is determined by

$$
\begin{cases}
n_t = 0 & \text{if } V(1, X_t) - \varepsilon_t < SV \\
\text{OR} \\
n_t = n > 0 & \text{if } V(n+1, X_t) - SV < \varepsilon_t \leq V(n, X_t) - SV
\end{cases}
$$

Regime 2: Entry. Suppose that $n_{t-1} < N$ and $V(n_{t-1} + 1, X_t) - \varepsilon_t \geq EC$. At the beginning of period t, potential entrants realize that the value of being active in the market is greater than the entry cost. Therefore, potential entrants want to enter in the market.

It should be clear that under this regime there is not exit. Since $SV \leq EC$ and $V(n_{t-1} + 1, X_t) < V(n_{t-1}, X_t)$, we have that the condition $\{V(n_{t-1} + 1, X_t) - \varepsilon_t \geq EC\}$ implies that $\{V(n_{t-1}, X_t) - \varepsilon_t > SV\}$. The value of an incumbent is greater than the scrapping value and therefore there is not exit.

Therefore, new firms will start entering the market up to the point when: either (a) there are no more potential entrants to enter in the market, i.e., $n_t = N$; or (b) there are still potential entrants that may enter the market but the value of an active firm goes down to a level such that there are not more incentives for additional entry. The equilibrium number of firms in this regime is given by the conditions:

$$
\begin{cases}
n_t = N & \text{if } V(N, X_t) - \varepsilon_t \geq EC \\
\text{OR} \\
n_t = n < N & \text{if } \{V(n, X_t) - \varepsilon_t \geq EC\} \quad \text{AND} \quad \{V(n+1, X_t) - \varepsilon_t < EC\}
\end{cases}
$$

Condition $\{V(n_t, X_t) - \varepsilon_t \geq EC\}$ says that the last firm that entered the market wanted to enter. Condition $\{V(n_t + 1, X_t) - \varepsilon_t < EC\}$ establishes that one more firm in the market would not get enough value to cover the entry cost.

Summarizing, **Regime 2 [Entry]** is described by the following condition on exogenous variables $\{n_{t-1} < N\}$ and $\{\varepsilon_t \leq V(n_{t-1} + 1, X_t) - EC\}$, and this condition implies that:

$$n_t > n_{t-1}$$

and n_t is determined by

$$
\begin{cases}
n_t = N & \text{if } V(N, X_t) - \varepsilon_t \geq EC \\
\text{OR} \\
n_t = n < N & \text{if } V(n+1, X_t) - EC < \varepsilon_t \leq V(n, X_t) - EC
\end{cases}
$$

Regime 3: Inaction. The third possible regime is given by the complementary condi-
tions to those that define regimes 1 and 2. Under these conditions, incumbent firms do not
want to exit and potential entrants do not want to enter.

$$
\{n_t = n_{t-1}\} \text{ iff }
\begin{cases}
\{n_{t-1} = 0\} \text{ AND } \{V(1, X_t) - \varepsilon_t < EC\} \\
\text{OR} \\
\{n_{t-1} = N\} \text{ AND } \{V(N, X_t) - \varepsilon_t \geq SV\} \\
\text{OR} \\
\{0 < n_{t-1} < N\} \text{ AND } \{V(n_{t-1}+1, X_t) - \varepsilon_t < EC\} \text{ AND } \{V(n_{t-1}, X_t) - \varepsilon_t \geq SV\}
\end{cases}
$$

Putting the three regimes together, we can obtain the probability distribution of the
endogenous n_t conditional on (n_{t-1}, X_t). Assume that ε_t is i.i.d. and independent of X_t with
CDF F_ε. Then:

$$
\Pr(n_t = n \mid n_{t-1}, X_t) =
\begin{cases}
F_\varepsilon\left(\frac{V(n, X_t) - SV}{\sigma}\right) - F_\varepsilon\left(\frac{V(n+1, X_t) - SV}{\sigma}\right) & \text{if } n < n_{t-1} \\[2ex]
F_\varepsilon\left(\frac{V(n_{t-1}, X_t) - SV}{\sigma}\right) - F_\varepsilon\left(\frac{V(n_{t-1}+1, X_t) - EC}{\sigma}\right) & \text{if } n = n_{t-1} \\[2ex]
F_\varepsilon\left(\frac{V(n, X_t) - EC}{\sigma}\right) - F_\varepsilon\left(\frac{V(n+1, X_t) - EC}{\sigma}\right) & \text{if } n > n_{t-1}
\end{cases}
$$

It is interesting to compare this probability distribution of the number of firms with the
one from the static BR model. In the static BR model:

$$
\Pr(n_t = n \mid n_{t-1}, X_t) = F_\varepsilon\left(\frac{V(n, X_t)}{\sigma}\right) - F_\varepsilon\left(\frac{V(n+1, X_t)}{\sigma}\right)
$$

This is exactly the distribution that we get in the dynamic model when $EC = SV$. Note
that $EC = SV$ the sunk cost $EC - SV$ is zero and firms' entry-exit decisions are static.

When $EC > SV$ (positive sunk cost), the dynamic model delivers different predictions
than the static model. There are two main differences. First, number of firms is more
persistence over time, i.e., there is "structural state dependence" in the number of firms.

$$
\Pr(n_t = n_{t-1} \mid n_{t-1}, X_t) =
\begin{cases}
F_\varepsilon\left(\frac{V(n_{t-1}, X_t) - SV}{\sigma}\right) - F_\varepsilon\left(\frac{V(n_{t-1}+1, X_t) - EC}{\sigma}\right) & \text{if } EC > SV \\[2ex]
F_\varepsilon\left(\frac{V(n_{t-1}, X_t)}{\sigma}\right) - F_\varepsilon\left(\frac{V(n_{t-1}+1, X_t)}{\sigma}\right) & \text{if } EC = SV
\end{cases}
$$

In the static model, all the persistent in the number of firms is because this variable
is indivisible, it is an integer. In the dynamic model, sunk entry costs introduce more
persistence. A purely transitory shock (in X_t or in ε_t) that increases the number of firms at
some period t will have a persistent effect for several periods in the future.

Second, there the number of firms responses asymmetrically to positive and negative shocks. Given $EC > SV$, it is possible to show that the upward response is less elastic than the downward response.

2.3. Identification. It is interesting to explore the identification of the model. With this model and data, we cannot identify nonparametrically the distribution of ε_t. So we make a parametric assumption on this distribution. For instance, we assume that ε_t has a $N(0, \sigma^2)$ distribution.

Define $P_{entry}(n_{t-1},\, X_t)$ and $P_{exit}(n_{t-1},\, X_t)$ as the probabilities of positive (entry) and negative (exit) changes in the number of firms, respectively. That is, $P_{entry}(n_{t-1},\, X_t) \equiv \Pr\left(n_t > n_{t-1} \mid n_{t-1},\, X_t\right)$ and $P_{exit}(n_{t-1},\, X_t) \equiv \Pr\left(n_t < n_{t-1} \mid n_{t-1},\, X_t\right)$. These probability functions are nonparametrically identified from our panel data on $\{n_t,\, X_t\}$.

The model predicts the following structure for the probabilities of entry and exit

$$P_{entry}(n_{t-1},\, X_t) \;=\; \Pr\left(V\left(n_{t-1}+1, X_t\right) - \varepsilon_t > EC \mid X_t\right) =$$

$$= \;\; \Phi\left(\frac{V\left(n_{t-1}+1, X_t\right) - EC}{\sigma}\right)$$

and:

$$P_{exit}(n_{t-1},\, X_t) \;=\; \Pr\left(V\left(n_{t-1}, X_t\right) - \varepsilon_t < SV \mid X_t\right)$$

$$= \;\; 1 - \Phi\left(\frac{V\left(n_{t-1}, X_t\right) - SV}{\sigma}\right)$$

Using these expressions, it is simple to obtain that, for any $(n_{t-1},\, X_t)$:

$$\frac{EC - SV}{\sigma} = \Phi^{-1}\left(1 - P_{exit}(n_{t-1},\, X_t)\right) - \Phi^{-1}\left(P_{entry}(n_{t-1}-1,\, X_t)\right)$$

where Φ^{-1} is the inverse function of the CDF of ε_t.

Therefore, even with a nonparametric specification of the value function $V(n, X)$, we can identify the sunk cost up to scale. Note that this expression provides a clear intuition about the source of identification of this parameter. The magnitude of this parameter is identified by "a distance" between the probability of entry of potential entrants and the probability of staying of incumbents $(1 - P_{exit})$. In a model without sunk costs, both probabilities should be the same. In a model with sunk costs, the probability of staying in the market should be greater than the probability of entry.

Since we do not know the value of σ, the value of the parameter $\frac{EC-SV}{\sigma}$ is not meaningful from an economic point of view. However, based on the identification of $\frac{EC-SV}{\sigma}$ and the identification up to scale of the value function $V(n, X)$, that we show below, it is possible

to get an economically meaningful estimate of the importance of sunk cost. Suppose that $V(n, X)/\sigma$ is identified. Then, we can identify the ratio:

$$\frac{EC - SV}{V(n, X)} = \frac{\frac{EC-SV}{\sigma}}{\frac{V(n,X)}{\sigma}}$$

For instance, we have $\frac{EC-SV}{V(1,X)}$ which is the percentage of the sunk cost over the value of a monopoly in a market with characteristics X.

Following the same argument as for the identification of the constant parameter $\frac{EC-SV}{\sigma}$, we can show the identification of a sunk cost that depends nonparametrically on the state variables (n_{t-1}, X_t). That is, we can identify a sunk cost function $\frac{EC(n_{t-1},X_t)-SV(n_{t-1},X_t)}{\sigma}$. This has economic interest. In particular, the dependence of the sunk cost with respect to the number of incumbents n_{t-1} is evidence of endogenous sunk costs (see John Sutton's book titled *"Sunk Costs and Market Structure,"* MIT Press, 1991). Therefore, we can test nonparametrically for the existence of endogenous sunk costs by testing the dependence of the estimated function $\frac{EC(n_{t-1},X_t)-SV(n_{t-1},X_t)}{\sigma}$ with respect to n_{t-1}.

We can also use the probabilities of entry and exit to identify the value function $V(n, X)$. The model implies that:

$$\Phi^{-1}\left(P_{entry}(n_{t-1} - 1, X_t)\right) = \frac{V(n_{t-1},X_t)-EC}{\sigma}$$

$$\Phi^{-1}\left(1 - P_{exit}(n_{t-1}, X_t)\right) = \frac{V(n_{t-1},X_t)-SV}{\sigma}$$

The left-hand-side of these equations is identified from the data. From these expressions, it should be clear that we cannot identify EC/σ separately from a constant term in the value function (a fixed cost), and we cannot identify SV/σ separately from a constant term in the value function.

Let $-FC$ be the constant term or fixed cost in the value function. More formally, define the parameter FC as the expected value:

$$FC \equiv -E\left(V(n_{t-1}, X_t)\right)$$

Also define the function $V^*(n_{t-1}, X_t)$ as the deviation of the value function with respect to its mean:

$$V^*(n_{t-1}, X_t) \equiv V(n_{t-1}, X_t) - -E\left(V(n_{t-1}, X_t)\right)$$

$$= V(n_{t-1}, X_t) + FC$$

Also, define also $EC^* \equiv EC + FC$, and $SV^* \equiv SV + FC$ such that, by definition, $V(n_{t-1}, X_t) - EC = V^*(n_{t-1}, X_t) - EC^*$, and $V(n_{t-1}, X_t) - SV = V^*(n_{t-1}, X_t) - SV^*$.

Then, $\frac{EC^*}{\sigma}$, $\frac{SV^*}{\sigma}$, and $V^*(n_{t-1}, X_t)/\sigma$ are identified nonparametrically from the following expressions:

$$\frac{EC^*}{\sigma} = E\left(\Phi^{-1}\left(P_{entry}(n_{t-1} - 1, X_t)\right)\right)$$

$$\frac{SV^*}{\sigma} = E\left(\Phi^{-1}\left(1 - P_{exit}(n_{t-1}, X_t)\right)\right)$$

And

$$\frac{V^*(n_{t-1}, X_t)}{\sigma} = \Phi^{-1}\left(P_{entry}(n_{t-1} + 1, X_t)\right) - E\left(\Phi^{-1}\left(P_{entry}(n_{t-1} + 1, X_t)\right)\right)$$

$$\text{and}$$

$$\frac{V^*(n_{t-1}, X_t)}{\sigma} = \Phi^{-1}\left(1 - P_{exit}(n_{t-1}, X_t)\right) - E\left(\Phi^{-1}\left(1 - P_{exit}(n_{t-1}, X_t)\right)\right)$$

In fact, we can see that the function $V^*(.,.)$ is over identified: it can be identified either from the probability of entry or from the probability of exit. This provides over-identification restrictions that can be used to test the restrictions or assumptions of the model.

Again, one of the main limitations of this model is the assumption of homogeneous firms. In fact, as an implication of that assumption, the model predicts that there should not be simultaneous entry and exit. This prediction is clearly rejected in many panel datasets on industry dynamics.

2.4. Estimation of the model. Given a parametric assumption about the distribution of ε_t, and a parametric specification of the value function $V(n, X)$, we can estimate the model by conditional maximum likelihood. For instance, suppose that ε_t is i.i.d. across markets and over time with a distribution $N(0, \sigma^2)$, and the value function is linear in parameters:

$$V(n_t, X_t) = g(n_t, X_t)'\beta - FC$$

where $g(.,.)$ is a vector of known functions, and β is a vector of unknown parameters.

Let θ be the vectors of parameters to estimate:

$$\theta = \{\ \beta/\sigma,\ EC^*/\sigma,\ SV/\sigma\}$$

Then, we can estimate θ using the conditional maximum likelihood estimator:

$$\hat{\theta} = \arg\max_{\theta} \sum_{m=1}^{M} \sum_{t=1}^{T} \sum_{n=0}^{N} 1\{n_{mt} = n\}\ \log \Pr(n \mid n_{mt-1},\ X_{mt}; \theta)$$

where:

$$
\Pr(n_t = n \mid n_{t-1},\ X_t) = \begin{cases} \Phi\left(g(n, X_t)'\dfrac{\beta}{\sigma} - \dfrac{SV}{\sigma}\right) - \Phi\left(g(n+1, X_t)'\dfrac{\beta}{\sigma} - \dfrac{SV}{\sigma}\right) & \text{if } n < n_{t-1} \\[3ex] \Phi\left(g(n, X_t)'\dfrac{\beta}{\sigma} - \dfrac{SV}{\sigma}\right) - \Phi\left(g(n+1, X_t)'\dfrac{\beta}{\sigma} - \dfrac{EC}{\sigma}\right) & \text{if } n = n_{t-1} \\[3ex] \Phi\left(g(n, X_t)'\dfrac{\beta}{\sigma} - \dfrac{EC}{\sigma}\right) - \Phi\left(g(n+1, X_t)'\dfrac{\beta}{\sigma} - \dfrac{EC}{\sigma}\right) & \text{if } n > n_{t-1} \end{cases}
$$

Based on the previous identification results, we can also construct a simple least squares estimator of θ. Let $\hat{P}_{mt}^{entry}$ and $\hat{P}_{mt}^{exit}$ be nonparametric Kernel estimates of $P_{entry}(n_{mt-1} + 1, X_{mt})$ and $P_{exit}(n_{mt-1}, X_{mt})$, respectively. The model implies that:

$$
\Phi^{-1}\left(\hat{P}_{mt}^{entry}\right) = \left(-\frac{EC^*}{\sigma}\right) + g(n_{mt-1}, X_{mt})'\frac{\beta}{\sigma} + e_{mt}^{entry}
$$

$$
\Phi^{-1}\left(1 - \hat{P}_{mt}^{exit}\right) = \left(-\frac{SV^*}{\sigma}\right) + g(n_{mt-1}, X_{mt})'\frac{\beta}{\sigma} + e_{mt}^{exit}
$$

where e_{mt}^{entry} and e_{mt}^{exit} are error terms that come from the estimation error in $\hat{P}_{mt}^{entry}$ and $\hat{P}_{mt}^{exit}$. We can put together these regression equations in a single regression as:

$$
Y_{dmt} = D_{dmt}\left(-\frac{EC^*}{\sigma}\right) + (1 - D_{dmt})\left(-\frac{SV^*}{\sigma}\right) + g(n_{mt-1}, X_{mt})'\frac{\beta}{\sigma} + e_{mt}
$$

where $Y_{dmt} \equiv D_{dmt}\ \Phi^{-1}\left(\hat{P}_{mt}^{entry}\right) + (1 - D_{dmt})\Phi^{-1}\left(1 - \hat{P}_{mt}^{exit}\right)$; the subindex d represents the "regime", $d \in \{entry,\ exit\}$, and D_{dmt} is a dummy variable that is equal to one when $d = entry$ and it is equal to zero when $d = exit$.

OLS estimation of this linear regression equation provides a consistent estimator of θ. This estimator is not efficient but we can easily obtain an asymptotically efficient estimator by making one Newton-Raphson iteration in the maximization of the likelihood function.

2.5. Structural model and counterfactual experiments. This dynamic model is fully consistent with a dynamic game of entry-exit. However, the value function $V(n, X)$ is not a primitive or a structural function. It implicitly depends on the one-period profit function, on the entry cost EC, on the scrapping value SV, and on the equilibrium of the model (i.e., on equilibrium firms' strategies).

The model and the empirical approach that we have described above does not make explicit the relationship between the primitives of the model and the value function, or how this value function depends on the equilibrium transition probability of the number of firms, $P^*(n_{t+1}|n_t, X_t)$. This "semi-structural" approach has clear advantages in terms of computational and conceptual simplicity. However, it has also its limitations. We discuss here its advantages and limitations.

Similar approaches have been proposed and applied for the estimation of dynamic models of occupational choice by Geweke and Keane (2001) and Florian Hoffmann (2009). This type of approach is different to other methods that have been proposed and applied to the estimation of dynamic structural models and that also try to reduce the computational cost in estimation, such as Hotz and Miller (1993 and 1994) and Aguirregabiria and Mira (2002 and 2007). We will see these other approaches later in the course.

To understand the advantages and limitations of Bresnahan and Reiss "semi-structural" model of industry dynamics it is useful to relate the value function $V(n_t, X_t)$ with the actual primitives of the model. Let $\pi(n_t, X_t, \varepsilon_t)$ be the profit function of an incumbent firm that stays in the market is $\pi(n_t, X_t) - \varepsilon_t$. Therefore:

$$V(n_t, X_t) = E\left(\sum_{j=0}^{\infty} \delta^j \; [(1 - Exit_{t+j}) \; \pi(n_{t+1}, X_{t+j}, \varepsilon_{t+j}) + Exit_{t+j}SV] \mid n_t, X_t\right)$$

where δ is the discount factor; and $Exit_{t+j}$ is a binary variable that indicates if the firm exits from the market at period $t+j$ (i.e., $Exit_{t+j} = 1$) or stays in the market (i.e., $Exit_{t+j} = 0$). The expectation is taken over all future paths of the state variables $\{n_{t+1}, X_{t+j}, \varepsilon_{t+j}\}$. In particular, this expectation depends on the stochastic process that follows the number of firms in equilibrium and that is governed by the transition probability $\Pr(n_{t+1}|n_t, X_t)$.

The transition probability $\Pr(n_{t+1}|n_t, X_t)$ is determined in equilibrium and it depends on all the structural parameters of the model. More specifically, this transition probability can be obtained as the solution of a fixed point problem. Solving this fixed point problem is computationally demanding. The "semi-structural" approach avoids this computational cost by ignoring the relationship between the value function $V(n_t, X_t)$ and the structural parameters of the model. This can provide huge computational advantages, especially when the dimension of the state space of (n_t, X_t) is large and/or when the dynamic game may have multiple equilibria.

These significant computational gains come with a cost. The range of predictions and counterfactual experiments that we can make using the estimated "semi-structural" model is very limited. In particular, we cannot make predictions about how the equilibrium transition $\Pr(n_{t+1}|n_t, X_t)$ (or the equilibrium steady-state distribution of n_t) changes when we perturb one the parameters in θ.

There are two types of problems in this model to implement these predictions of counterfactual experiments. First, the parameters β are not structural such that we cannot change one of this parameters and assume that the rest will stay constant. Or in other words, we do not know what that type of experiment means.

Second, though EC^* and SV^* are structural parameters, the parameters β in the value function should depend on EC^* and SV^*, but we do not know the form of that relationship.

We cannot assume that EC^* or SV^* and β remains constant. Or that type of experiment does not have a clear interpretation or economic interest.

For instance, suppose that we want to predict how a 20% increase in the entry cost would affect the transition dynamic and the steady state distribution of the number of firms. If $\lambda_0 = EC^*/\sigma$ is our estimate of the value of the parameter in the sample, then its counterfactual value is $\lambda_1 = 1.2\lambda_0$. However, we also know that the value function should change. In particular, the value of an incumbent firm increases when the entry costs increase.

The "semi-structural" model ignores that the value function V will change as the result of the change in the entry cost. Therefore, it predicts that entry will decline, and that the exit/stay behavior of incumbent firms will not be affected because V and SV have not changed.

There are two errors in the prediction of the "semi-structural" model. First, it overestimates the decline in the amount of entry because it does not take into account that being an incumbent in the market now has more value. And second, it ignores that, for the same reason, exit of incumbent firms will also decline.

Putting these two errors together we have that this counterfactual experiment using the "semi-structural" model can lead to a serious under-estimate of the number of firms in the counterfactual scenario.

Later in the course we will study other methods for the estimation of structural models of industry dynamics that avoid the computational cost of solving for the equilibrium of the game but that do not have the important limitations, in terms of counterfactual experiments, of the semi-structural model here.

Nevertheless, it is difficult to overemphasize the computational advantages of Bresnahan-Reiss empirical model of industry dynamics. It is a useful model to obtain a first cut of the data, and to answer empirical questions that do not require the implementation of counterfactual experiments, e.g., testing for endogenous sunk costs, or measuring the magnitude of sunk costs relative to the value of an incumbent firm.

3. The structure of dynamic games of oligopoly competition

3.1. Basic Framework and Assumptions. Time is discrete and indexed by t. The game is played by N firms that we index by i. Following the standard structure in the Ericson-Pakes (1995) framework, firms compete in two different dimensions: a static dimension and a dynamic dimension. We denote the dynamic dimension as the "investment decision". Let a_{it} be the variable that represents the investment decision of firm i at period t. This investment decision can be an entry/exit decision, a choice of capacity, investment in equipment, R&D, product quality, other product characteristics, etc. Every period, given

their capital stocks that can affect demand and/or production costs, firms compete in prices or quantities in a static Cournot or Bertand model. Let p_{it} be the static decision variables (e.g., price) of firm i at period t.

For simplicity and concreteness, I start presenting a simple dynamic game of market entry-exit where every period incumbent firms compete a la Bertrand. In this entry-exit model, the dynamic investment decision a_{it} is a binary indicator of the event "firm i is active in the market at period t". The action is taken to maximize the expected and discounted flow of profits in the market, $E_t \left(\sum_{r=0}^{\infty} \delta^r \ \Pi_{it+r} \right)$ where $\delta \in (0,1)$ is the discount factor, and Π_{it} is firm i's profit at period t. The profits of firm i at time t are given by

$$\Pi_{it} = VP_{it} - FC_{it} - EC_{it}$$

where VP_{it} represents variable profits, FC_{it} is the fixed cost of operating, and EC_{it} is a one time entry cost. We now describe these different components of the profit function.

(a) Variable Profit Function. The variable profit VP_{it} is an "indirect" variable profit function that comes from the equilibrium of a static Bertrand game with differentiated product. Consider the simplest version of this type of model. Suppose that all firms have the same marginal cost, c, and product differentiation is symmetric. Consumer utility of buying product i is $u_{it} = \nu - \alpha p_{it} + \varepsilon_{it}$, where ν and α are parameters, and ε_{it} is a consumer-specific i.i.d. extreme value type 1 random variable. Under these conditions, the equilibrium variable profit of an active firm depends only on the number of firms active in the market.

$$VP_{it} = (p_{it} - c)q_{it}$$

where p_{it} and q_{it} represent the price and the quantity sold by firm i at period t, respectively. According this model, the quantity is:

$$q_{it} = H_t \ \frac{a_{it} \ \exp\{v - \alpha \ p_{it}\}}{1 + \sum_{j=1}^{N} a_{jt} \ \exp\{v - \alpha \ p_{jt}\}} = H_t \ s_{it}$$

where H_t is the number of consumers in the market (market size) and s_{it} is the market share of firm i. Under the Nash-Bertrand assumption the first order conditions for profit maximization are:

$$q_{it} + (p_{it} - c) \ (-\alpha) \ q_{it} \ (1 - s_{it}) = 0$$

or

$$p_{it} = c + \frac{1}{\alpha \ (1 - s_{it})}$$

Since all firms are identical, we consider a symmetric equilibrium, $p_t^* = p_{it}^*$ for every firm i. Therefore, $s_{it} = a_{it} \ s_t^*$, and:

$$s_t^* = \frac{\exp\{v - \alpha \ p_t^*\}}{1 + n_t \ \exp\{v - \alpha \ p_t^*\}}$$

where $n_t \equiv \sum_{j=1}^{N} a_{jt}$ is the number of active firms at period t. Then, it is simple to show that the equilibrium price p_t^* is implicitly defined as the solution to the following fixed point problem:

$$p_t^* = \left(c + \frac{1}{\alpha}\right) + \frac{1}{\alpha}\left(\frac{\exp\{v - \alpha\ p_t^*\}}{1 + (n_t - 1)\ \exp\{v - \alpha\ p_t^*\}}\right)$$

It is simple to show that an equilibrium always exists. The equilibrium price depends on the number of firms active in the market, but in this model it does not depend on market size: $p_t^* = p^*(n_t)$. Similarly, the equilibrium market share s_t^* is a function of the number of active firms: $s_t^* = s^*(n_t)$. Therefore, the indirect or equilibrium variable profit of an active firm is:

$$VP_{it} = a_{it}\ H_t\ (p^*(n_t) - c)\ s^*(n_t)$$

$$= a_{it}\ H_t\ \theta^{VP}(n_t)$$

where $\theta^{VP}(.)$ is a function that represents variable profits per capita.

For most of the analysis below, I will consider that the researcher does not have access to information on prices and quantities. Therefore, we will treat $\{\theta^{VP}(1), \theta^{VP}(2), ..., \theta^{VP}(N)\}$ as parameters to estimate from the structural dynamic game.

Of course, we can extend the previous approach to incorporate richer form of product differentiation. In fact, product differentiation can be endogenous. Suppose that the quality parameter v in the utility function can take A possible values: $v(1) < v(2) < ... < v(A)$. And suppose that the investment decision a_{it} combines an entry/exit decision with a "quality" choice decision. That is, $a_{it} \in \{0, 1, ..., A\}$ where $a_{it} = 0$ represents that firm i is not active in the market, and $a_{it} = a > 0$ implies that firm i is active in the market with a product of quality a. It is straightforward to show that, in this model, the equilibrium variable profit of an active firm is:

$$VP_{it} = \sum_{a=1}^{A} 1\{a_{it} = a\}\ H_t\ \theta^{VP}(a, n_t^{(1)}, n_t^{(2)}, ..., n_t^{(A)})$$

where $\theta^{VP}(.)$ is the variable profit per capita that now depends on the firm's own quality, and one the number of competitors at each possible level of quality.

(b) Fixed Cost. The fixed cost is paid every period that the firm is active in the market, and it has the following structure:

$$FC_{it} = a_{it}\ \left(\theta_i^{FC} + \varepsilon_{it}\right)$$

θ_i^{FC} is a parameter that represents the mean value of the fixed operating cost of firm i. ε_{it} is a zero-mean shock that is private information of firm i. There are two main reasons why we incorporate these private information shocks in the model. First, as shown in Doraszelski

and Satterthwaite (2007) it is way to guarantee that the dynamic game has at least one equilibrium in pure strategies. And second, they are convenient econometric errors. If private information shocks are independent over time and over players, and unobserved to the researcher, they can 'explain' players heterogeneous behavior without generating endogeneity problems.

We will see later that the assumption that these private information shocks are the only unobservables for the researcher can be too restrictive. We will study how to incorporate richer forms of unobserved heterogeneity.

For the model with endogenous quality choice, we can generalize the structure of fixed costs:

$$FC_{it} = \sum_{a=1}^{A} 1\{a_{it} = a\} \ \left(\theta_i^{FC}(a) + \varepsilon_{it}(a)\right)$$

where now the mean value of the fixed cost, $\theta_i^{FC}(a)$, and the private information shock, $\varepsilon_{it}(a)$, depend on the level quality.

(c) Entry Cost and Repositioning costs. The entry cost is paid only if the firm was not active in the market at previous period:

$$EC_{it} = a_{it} \ (1 - x_{it}) \ \theta_i^{EC}$$

where x_{it} is a binary indicator that is equal to 1 if firm i was active in the market in period $t-1$, i.e., $x_{it} \equiv a_{i,t-1}$, and θ_i^{EC} is a parameter that represents the entry cost of firm i. For the model with endogenous quality, we can also generalize this entry cost to incorporate also costs of adjusting the level of quality, or repositioning product characteristics. For instance,

$$
\begin{aligned}
EC_{it} \ &= \ 1\{x_{it} = 0\} \left(\sum_{a=1}^{A} 1\{a_{it} = a\} \ \theta_i^{EC}(a)\right) \\
&+ \ 1\{x_{it} > 0\} \left(\theta_i^{AC(+)} 1\{a_{it} > x_{it}\} + \theta_i^{AC(-)} 1\{a_{it} < x_{it}\}\right)
\end{aligned}
$$

Now, $x_{it} = a_{i,t-1}$ also represents the firm's quality at previous period. $\theta_i^{EC}(a)$ is the cost of entry with quality a, and $\theta_i^{AC(+)}$ and $\theta_i^{AC(-)}$ represents the costs of increasing and reducing quality, respectively, once the firm is active.

The payoff relevant state variables of this model are: (1) market size H_t; (2) the incumbent status (or quality levels) of firms at previous period $\{x_{it} : i = 1, 2, ..., N\}$; and (3) the private information shocks $\{\varepsilon_{it} : i = 1, 2, ..., N\}$. The specification of the model is completed with the transition rules of these state variables. Market size follows an exogenous Markov process with transition probability function $F_H(H_{t+1}|H_t)$. The transition of the incumbent status is trivial: $x_{it+1} = a_{it}$. Finally, the private information shock ε_{it} is i.i.d. over time and independent across firms with CDF G_i.

Note that in this example, I consider that firms' dynamic decisions are made at the beginning of period t and they are effective during the same period. An alternative timing that has been considered in some applications is that there is a one-period time-to-build, i.e., the decision is made at period t, and entry costs are paid at period t, but the firm is not active in the market until period $t + 1$. The latter is in fact the timing of decisions in Ericson and Pakes (1995). All the results below can be generalized in a straightforward way to that case, and we will see empirical applications with that timing assumption.

3.2. Markov Perfect Equilibrium. Most of the recent literature in IO studying industry dynamics focuses on studying a Markov Perfect Equilibrium (MPE), as defined by Maskin and Tirole (Econometrica, 1988). The key assumption in this solution concept is that players' strategies are functions of only payoff-relevant state variables. We use the vector $\mathbf{x}_t$ to represent all the common knowledge state variables at period t, i.e., $\mathbf{x}_t \equiv (H_t, x_{1t}, x_{2t}, ..., x_{Nt})$. In this model, the payoff-relevant state variables for firm i are $(\mathbf{x}_t, \varepsilon_{it})$.

Note that if private information shocks are serially correlated, the history of previous decisions contains useful information to predict the value of a player's private information, and it should be part of the set of payoff relevant state variables. Therefore, the assumption that private information is independently distributed over time has implications for the set of payoff-relevant state variables.

Let $\boldsymbol{\alpha} = \{\alpha_i(\mathbf{x}_t, \varepsilon_{it}) : i \in \{1, 2, ..., N\}\}$ be a set of strategy functions, one for each firm. A MPE is a set of strategy functions $\boldsymbol{\alpha}^*$ such that every firm is maximizing its value given the strategies of the other players. For given strategies of the other firms, the decision problem of a firm is a single-agent dynamic programming (DP) problem. Let $V_i^{\boldsymbol{\alpha}}(\mathbf{x}_t, \varepsilon_{it})$ be the value function of this DP problem. This value function is the unique solution to the Bellman equation:

$$V_i^{\boldsymbol{\alpha}}(\mathbf{x}_t, \varepsilon_{it}) = \max_{a_{it}} \left\{ \Pi_i^{\boldsymbol{\alpha}}(a_{it}, \mathbf{x}_t) - \varepsilon_{it}(a_{it}) + \delta \int V_i^{\boldsymbol{\alpha}}(\mathbf{x}_{t+1}, \varepsilon_{it+1}) \, dG_i(\varepsilon_{it+1}) \, F_i^{\boldsymbol{\alpha}}(\mathbf{x}_{t+1}|a_{it}, \mathbf{x}_t) \right\}$$

$$(3.1)$$

where $\Pi_i^{\boldsymbol{\alpha}}(a_{it}, \mathbf{x}_t)$ and $F_i^{\boldsymbol{\alpha}}(\mathbf{x}_{t+1}|a_{it}, \mathbf{x}_t)$ are the expected one-period profit and the expected transition of the state variables, respectively, for firm i given the strategies of the other firms. For the simple entry/exit game, the expected one-period profit $\Pi_i^{\boldsymbol{\alpha}}(a_{it}, \mathbf{x}_t)$ is:

$$\Pi_i^{\boldsymbol{\alpha}}(a_{it}, \mathbf{x}_t) = a_{it} \left[H_t \sum_{n=0}^{N-1} \Pr\left(\sum_{j \neq i} \alpha_j(\mathbf{x}_t, \varepsilon_{jt}) = n \mid \mathbf{x}_t \right) \theta^{VP}(n+1) - \theta_i^{FC} - (1 - x_{it})\theta_i^{EC} \right]$$

And the expected transition of the state variables is:

$$F_i^{\boldsymbol{\alpha}}(\mathbf{x}_{t+1}|a_{it}, \mathbf{x}_t) \;\; = \;\; 1\{x_{it+1} = a_{it}\} \left[\prod_{j \neq i} \Pr\left(x_{j,t+1} = \alpha_j(\mathbf{x}_t, \varepsilon_{jt}) \mid \mathbf{x}_t \right) \right] F_H(H_{t+1} \mid H_t)$$

A player's best response function gives his optimal strategy if the other players behave, now and in the future, according to their respective strategies. In this model, the best response function of player i is:

$$\alpha_i^*(\mathbf{x}_t, \varepsilon_{it}) = \arg\max_{a_{it}} \left\{ v_i^{\boldsymbol{\alpha}}(a_{it}, \mathbf{x}_t) - \varepsilon_{it}(a_{it}) \right\}$$

where $v_i^{\boldsymbol{\alpha}}(a_{it}, \mathbf{x}_t)$ is the conditional choice value function that represents the value of firm i if: (1) the firm chooses alternative a_{it} today and then behaves optimally forever in the future; and (2) the other firms behave according to their strategies in $\boldsymbol{\alpha}$. By definition,

$$v_i^{\boldsymbol{\alpha}}(a_{it}, \mathbf{x}_t) \equiv \Pi_i^{\boldsymbol{\alpha}}(a_{it}, \mathbf{x}_t) + \delta \int V_i^{\boldsymbol{\alpha}}(\mathbf{x}_{t+1}, \varepsilon_{it+1}) \ dG_i(\varepsilon_{it+1}) \ F_i^{\boldsymbol{\alpha}}(\mathbf{x}_{t+1}|a_{it}, \mathbf{x}_t)$$

A Markov perfect equilibrium (MPE) in this game is a set of strategy functions $\boldsymbol{\alpha}^*$ such that for any player i and for any $(\mathbf{x}_t, \varepsilon_{it})$ we have that:

$$\alpha_i^*(\mathbf{x}_t, \varepsilon_{it}) = \arg\max_{a_{it}} \left\{ v_i^{\boldsymbol{\alpha}^*}(a_{it}, \mathbf{x}_t) - \varepsilon_{it}(a_{it}) \right\}$$

3.3. Conditional Choice Probabilities. Given a strategy function $\alpha_i(\mathbf{x}_t, \varepsilon_{it})$, we can define the corresponding *Conditional Choice Probability (CCP)* function as :

$$P_i(a|\mathbf{x}) \equiv \Pr\left(\alpha_i(\mathbf{x}_t, \varepsilon_{it}) = a \mid \mathbf{x}_t = \mathbf{x}\right)$$

$$= \int 1\{\alpha_i(\mathbf{x}_t, \varepsilon_{it}) = a\} \ dG_i(\varepsilon_{it})$$

Since choice probabilities are integrated over the continuous variables in ε_{it}, they are lower dimensional objects than the strategies $\boldsymbol{\alpha}$. For instance, when both a_{it} and $\mathbf{x}_t$ are discrete, CCPs can be described as vectors in a finite dimensional Euclidean space. In our entry-exit model, $P_i(1|\mathbf{x}_t)$ is the probability that firm i is active in the market given the state $\mathbf{x}_t$. Under standard regularity conditions, it is possible to show that there is a one-to-one relationship between strategy functions $\alpha_i(\mathbf{x}_t, \varepsilon_{it})$ and CCP functions $P_i(a|\mathbf{x}_t)$. From now on, we use CCPs to represent players' strategies, and use the terms 'strategy' and 'CCP' as interchangeable. We also use $\Pi_i^{\mathbf{P}}$ and $F_i^{\mathbf{P}}$ instead of $\Pi_i^{\boldsymbol{\alpha}}$ and $F_i^{\boldsymbol{\alpha}}$ to represent the expected profit function and the transition probability function, respectively.

Based on the concept of CCP, we can represent the equilibrium mapping and a MPE in way that is particularly useful for the econometric analysis. This representation has two main features:

(1) a MPE is a vector of CCPs;

(2) a player's best response is an optimal response not only to the other players' strategies but also to his own strategy in the future.

A MPE is a vector of CCPs, $\mathbf{P} \equiv \{P_i(a|\mathbf{x}) :$ for any $(i, a, x)\}$, such that for every firm and any state $\mathbf{x}$ the following equilibrium condition is satisfied:

$$P_i(a|\mathbf{x}) = \Pr\left(a = \arg\max_{a_i} \left\{v_i^{\mathbf{P}}(a_i, \mathbf{x}) - \varepsilon_i(a_i)\right\} \mid \mathbf{x}\right)$$

The right hand side of this equation is a best response probability function. $v_i^{\mathbf{P}}(a_i, \mathbf{x})$ is a conditional choice probability function, but it has a slightly different definition that before. Now, $v_i^{\mathbf{P}}(a_i, \mathbf{x})$ represents the value of firm i if: (1) the firm chooses alternative a_i today; and (2) **all the firms**, including firm i, behave according to their respective CCPs in $\mathbf{P}$. The Representation Lemma in Aguirregabiria and Mira (2007) shows that every MPE in this dynamic game can be represented using this mapping. In fact, this is result is a particular application of the so called "one-period deviation principle".

The form of this equilibrium mapping depends on the distribution of ε_i. For instance, in the entry/exit model, if ε_i is $N(0, \sigma_\varepsilon^2)$:

$$P_i(1|\mathbf{x}) = \Phi\left(\frac{v_i^{\mathbf{P}}(1, \mathbf{x}) - v_i^{\mathbf{P}}(0, \mathbf{x})}{\sigma_\varepsilon}\right)$$

In the model with endogenous quality choice, if $\varepsilon_i(a)$'s are extreme value type 1 distributed:

$$P_i(a|\mathbf{x}) = \frac{\exp\left\{\frac{v_i^{\mathbf{P}}(a,\mathbf{x})}{\sigma_\varepsilon}\right\}}{\sum_{a'=0}^{A} \exp\left\{\frac{v_i^{\mathbf{P}}(a',\mathbf{x})}{\sigma_\varepsilon}\right\}}$$

3.4. Computing $v_i^{\mathbf{P}}$ for arbitrary P. Now, I describe how to obtain the conditional choice value functions $v_i^{\mathbf{P}}$. Since $v_i^{\mathbf{P}}$ is not based on the optimal behavior of firm i in the future, but just in an arbitrary behavior described by $P_i(.|.)$, calculating $v_i^{\mathbf{P}}$ does not require solving a DP problem, and it only implies a valuation exercise.

By definition:

$$v_i^{\mathbf{P}}(a_i, \mathbf{x}) = \Pi_i^{\mathbf{P}}(a_i, \mathbf{x}) + \delta \sum_{\mathbf{x}'} V_i^{\mathbf{P}}(\mathbf{x}') \ F_i^{\mathbf{P}}(\mathbf{x}'|a_i, \mathbf{x})$$

$\Pi_i^{\mathbf{P}}(a_i, \mathbf{x})$ is the expected current profit. In the entry/exit example:

$$\Pi_i^{\mathbf{P}}(a_i, \mathbf{x}) = a_i\left[H \sum_{n=0}^{N-1} \Pr\left(n_{-i} = n \mid \mathbf{x}, \mathbf{P}\right) \theta^{VP}(n+1) - \theta_i^{FC} - (1 - x_i)\theta_i^{EC}\right]$$

$$= a_i\left[\mathbf{z}_i^{\mathbf{P}}(\mathbf{x}) \ \boldsymbol{\theta}_i\right]$$

where $\boldsymbol{\theta}_i$ is the vector of parameters:

$$\boldsymbol{\theta}_i = \left(\theta^{VP}(1), \theta^{VP}(2), ...,\theta^{VP}(N), \theta_i^{FC}, \theta_i^{EC}\right)'$$

and $\mathbf{z}_i^{\mathbf{P}}(\mathbf{x})$ is the vector that depends only on the state $\mathbf{x}$ and on the CCPs at state $\mathbf{x}$, but not on structural parameters.

$$\mathbf{z}_i^{\mathbf{P}}(\mathbf{x}) = \left(H \Pr\left(n_{-i} = 1 | \mathbf{x}, \mathbf{P}\right), \; ..., \; H \Pr\left(n_{-i} = N - 1 | \mathbf{x}, \mathbf{P}\right), \; -1, \; -(1 - x_i) \right)$$

For the dynamic game with endogenous quality choice, we can also represent the expected current profit $\Pi_i^{\mathbf{P}}(a_i, \mathbf{x})$ as:

$$\Pi_i^{\mathbf{P}}(a_i, \mathbf{x}) = \mathbf{z}_i^{\mathbf{P}}(a_i, \mathbf{x}) \, \boldsymbol{\theta}_i$$

The value function $V_i^{\mathbf{P}}(.)$ represents the value of firm i if all the firms, including firm i, behave according to their CCPs in $\mathbf{P}$. We can obtain $V_i^{\mathbf{P}}$ as the unique solution of the recursive expression:

$$V_i^{\mathbf{P}}(\mathbf{x}) = \sum_{a_i=0}^{A} P_i(a_i | \mathbf{x}) \left[\mathbf{z}_i^{\mathbf{P}}(a_i, \mathbf{x}) \boldsymbol{\theta}_i + \delta \sum_{\mathbf{x}'} V_i^{\mathbf{P}}(\mathbf{x}') \, F^{\mathbf{P}}(\mathbf{x}' | a_i, \mathbf{x}) \right]$$

When the space $\mathcal{X}$ is discrete and finite, we can obtain $V_i^{\mathbf{P}}$ as the solution of a system of linear equations of dimension $|\mathcal{X}|$. In vector form:

$$\mathbf{V}_i^{\mathbf{P}} = \left[\sum_{a_i=0}^{A} P_i(a_i) * \mathbf{z}_i^{\mathbf{P}}(a_i) \right] \boldsymbol{\theta}_i + \delta \left[\sum_{a_i=0}^{A} P_i(a_i) * \mathbf{F}_i^{\mathbf{P}}(a_i) \right] \mathbf{V}_i^{\mathbf{P}}$$

$$= \; \bar{\mathbf{z}}_i^{\mathbf{P}} \, \boldsymbol{\theta}_i + \delta \, \bar{\mathbf{F}}^{\mathbf{P}} \mathbf{V}_i^{\mathbf{P}}$$

where $\bar{\mathbf{z}}_i^{\mathbf{P}} = \sum_{a_i=0}^{A} P_i(a_i) * \mathbf{z}_i^{\mathbf{P}}(a_i)$, and $\bar{\mathbf{F}}^{\mathbf{P}} = \sum_{a_i=0}^{A} P_i(a_i) * \mathbf{F}_i^{\mathbf{P}}(a_i)$. Then, solving for $\mathbf{V}_i^{\mathbf{P}}$, we have:

$$\mathbf{V}_i^{\mathbf{P}} = \left(\mathbf{I} - \delta \, \bar{\mathbf{F}}^{\mathbf{P}}\right)^{-1} \bar{\mathbf{z}}_i^{\mathbf{P}} \, \boldsymbol{\theta}_i$$

$$= \; \mathbf{W}_i^{\mathbf{P}} \, \boldsymbol{\theta}_i$$

where $\mathbf{W}_i^{\mathbf{P}} = \left(\mathbf{I} - \delta \, \bar{\mathbf{F}}^{\mathbf{P}}\right)^{-1} \bar{\mathbf{z}}_i^{\mathbf{P}}$ is a matrix that only depends on CCPs and transition probabilities but not on $\boldsymbol{\theta}$.

Solving these expressions into the formula for the conditional choice value function, we have that:

$$v_i^{\mathbf{P}}(a_i, \mathbf{x}) = \tilde{z}_i^{\mathbf{P}}(a_i, \mathbf{x}) \, \boldsymbol{\theta}_i$$

where:

$$\tilde{z}_i^{\mathbf{P}}(a_i, \mathbf{x}) = z_i^{\mathbf{P}}(a_i, \mathbf{x}) + \delta \sum_{\mathbf{x}'} F_i^{\mathbf{P}}(\mathbf{x}' | a_i, \mathbf{x}) \, \mathbf{W}_i^{\mathbf{P}}$$

Finally, the equilibrium or best response mapping in the space of CCPs becomes:

$$P_i(a | \mathbf{x}) = \Pr\left(a = \arg\max_{a_i} \left\{ \tilde{z}_i^{\mathbf{P}}(a_i, \mathbf{x}) \, \boldsymbol{\theta}_i - \varepsilon_i(a_i) \right\} \; | \; \mathbf{x} \right)$$

For the entry/exit model with $\varepsilon_i \sim N(0, \sigma_\varepsilon^2)$:

$$P_i(1|\mathbf{x}) = \Phi\left(\left[\tilde{z}_i^{\mathbf{P}}(1, \mathbf{x}) - \tilde{z}_i^{\mathbf{P}}(0, \mathbf{x})\right]\frac{\boldsymbol{\theta}_i}{\sigma_\varepsilon}\right)$$

In the model with endogenous quality choice with $\varepsilon_i(a)$'s extreme value type 1 distributed:

$$P_i(a|\mathbf{x}) = \frac{\exp\left\{\tilde{z}_i^{\mathbf{P}}(a, \mathbf{x})\dfrac{\boldsymbol{\theta}_i}{\sigma_\varepsilon}\right\}}{\sum_{a'=0}^{A}\exp\left\{\tilde{z}_i^{\mathbf{P}}(a', \mathbf{x})\dfrac{\boldsymbol{\theta}_i}{\sigma_\varepsilon}\right\}}$$

Identification

First, let's summarize the structure of the dynamic game of oligopoly competition.

Let θ be the vector of structural parameters of the model, where $\theta = \{\theta_i : i = 1, 2, ..., N\}$ and θ_i includes the vector of parameters in the variable profit, fixed cost, and entry cost of firm i: for instance, in the entry-exit example, $\theta_i = (\theta^{VP}(1), \theta^{VP}(2), ..., \theta^{VP}(N), \theta_i^{FC}, \theta_i^{EC})'$. Let $\mathbf{P}(\theta) = \{P_i(a|\mathbf{x},\theta) : \text{for any } (i, a, \mathbf{x})\}$ be a MPE of the model associated with θ. $\mathbf{P}(\theta)$ is a solution to the following equilibrium mapping: for any $(i, a_i, \mathbf{x})$:

$$P_i(a_i|\mathbf{x},\theta) = \frac{\exp\left\{\tilde{z}_i^{\mathbf{P}}(a_i, \mathbf{x})\dfrac{\boldsymbol{\theta}_i}{\sigma_\varepsilon}\right\}}{\sum_{a'=0}^{A}\exp\left\{\tilde{z}_i^{\mathbf{P}}(a', \mathbf{x})\dfrac{\boldsymbol{\theta}_i}{\sigma_\varepsilon}\right\}}$$

where the vector of values $\tilde{z}_i^{\mathbf{P}}(a, \mathbf{x})$ are

$$\tilde{z}_i^{\mathbf{P}}(a_i, \mathbf{x}) = z_i^{\mathbf{P}}(a_i, \mathbf{x}) + \delta\sum_{\mathbf{x}'}F_i^{\mathbf{P}}(\mathbf{x}'|a_i, \mathbf{x})\,\mathbf{W}_i^{\mathbf{P}}$$

and $\mathbf{W}_i^{\mathbf{P}} = \mathbf{W}_i^{\mathbf{P}} = \left(\mathbf{I} - \delta\,\bar{\mathbf{F}}^{\mathbf{P}}\right)^{-1}\bar{\mathbf{z}}_i^{\mathbf{P}}$, and $z_i^{\mathbf{P}}(a_i, \mathbf{x})$ is a vector with the different components of the current expected profit. For instance, in the entry-exit example:

$$\mathbf{z}_i^{\mathbf{P}}(0, \mathbf{x}) = (0, 0, 0, ...0)$$

$$\mathbf{z}_i^{\mathbf{P}}(1, \mathbf{x}) = (H\Pr(n_{-i} = 1|\mathbf{x},\mathbf{P}), ..., H\Pr(n_{-i} = N - 1|\mathbf{x},\mathbf{P}), -1, -(1 - x_i))$$

That is, $\tilde{z}_i^{\mathbf{P}}(a_i, \mathbf{x})$ represents the expected present value of the different components of the current profit of firm i if he chooses alternative a_i today, and then all the firms, including firm i, behave in the future according to their CCPs in the vector $\mathbf{P}$.

In general, I will use the function $\Psi_i(a_i, \mathbf{x}; \mathbf{P}, \boldsymbol{\theta})$ to represent the best response or equilibrium function that in our example is $\dfrac{\exp\left\{\tilde{z}_i^{\mathbf{P}}(a_i,\mathbf{x})\dfrac{\boldsymbol{\theta}_i}{\sigma_\varepsilon}\right\}}{\sum_{a'=0}^{A}\exp\left\{\tilde{z}_i^{\mathbf{P}}(a',\mathbf{x})\dfrac{\boldsymbol{\theta}_i}{\sigma_\varepsilon}\right\}}$. Then, we can represent in a compact form a MPE as:

$$\mathbf{P} = \Psi(\mathbf{P}, \boldsymbol{\theta})$$

where $\Psi(\mathbf{P}, \boldsymbol{\theta}) = \{\Psi_i(a_i, \mathbf{x}; \mathbf{P}, \boldsymbol{\theta}) : \text{for any } (i, a, \mathbf{x})\}$.

Our first goal is to use data on firms' investment decisions $\{a_{it}\}$ and state variables $\{x_{it}\}$ to estimate the parameters θ.

Our second goal is to use the estimated model to perform counterfactual analysis/experiments that will help us to understand competition in this industry and to evaluate the effects of public policies or/and changes in structural parameters.

Data. In most applications of dynamic games in empirical IO the researcher observes a random sample of M markets, indexed by m, over T periods of time, where the observed variables consists of players' actions and state variables. In the standard application in IO, the values of N and T are small, but M is large. Two aspects of the data deserve some comments. For the moment, we consider that the industry and the data are such that: (a) each firm is observed making decisions in every of the M markets; and (b) the researcher knows all the payoff relevant market characteristics that are common knowledge to the firms. We describe condition (a) as a data set with *global players*. For instance, this is the case in a retail industry characterized by competition between large retail chains which are potential entrants in any of the local markets that constitute the industry. With this type of data we can allow for rich firm heterogeneity that is fixed across markets and time by estimating firm-specific structural parameters, $\boldsymbol{\theta}_i$. This 'fixed-effect' approach to deal with firm heterogeneity is not feasible in data sets where most of the competitors can be characterized as *local players*, i.e., firms specialized in operating in a few markets. Condition (b) rules out the existence of unobserved market heterogeneity. Though it is a convenient assumption, it is also unrealistic for most applications in empirical IO. Later I present estimation methods that relax conditions (a) and (b) and deal with unobserved market and firm heterogeneity.

Suppose that we have a random sample of M local markets, indexed by m, over T periods of time, where we observe:

$$Data = \{\boldsymbol{a}_{mt}, \ \mathbf{x}_{mt} : m = 1, 2, ..., M; \ t = 1, 2, ..., T\}$$

We want to use these data to estimate the model parameters in the population that has generated this data: $\boldsymbol{\theta}^0 = \{\boldsymbol{\theta}_i^0 : i \in I\}$.

Identification. A significant part of this literature has considered the following identification assumptions.

Assumption (ID 1): Single equilibrium in the data. Every observation in the sample comes from the same Markov Perfect Equilibrium, i.e., for any observation (m, t), $\mathbf{P}_{mt}^0 = \mathbf{P}^0$.

Assumption (ID 2): No unobserved common-knowledge variables. The only unobservables for the econometrician are the private information shocks ε_{imt} and the structural parameters $\boldsymbol{\theta}$.

Comments on these assumptions: .... The assumption of no unobserved common knowledge variables (e.g., no unobserved market heterogeneity) is particularly strong.

It is possible to relax these assumptions. We will see later identification and estimation when we relax assumption ID 2. The following is a standard regularity condition.

Assumption (ID 3): For some benchmark choice alternative, say $a_i = 0$, define $Z_{imt} \equiv \tilde{z}_i^{\mathbf{P}^0}(a_{imt}, \mathbf{x}_{mt}) - \tilde{z}_i^{\mathbf{P}^0}(0, \mathbf{x}_{mt})$. Then, $E(Z'_{imt}Z_{imt})$ is a non-singular matrix.

Under assumptions ID-1 to ID-3, the proof of identification is straightforward. First, under assumptions ID-1 and ID-2, the equilibrium that has generated the data, $\mathbf{P}^0$, can be estimated consistently and nonparametrically from the data. For any $(i, a_i, \mathbf{x})$:

$$P_i^0(a_i|\mathbf{x}) = \Pr(a_{imt} = a_i \mid \mathbf{x}_{mt} = \mathbf{x})$$

For instance, we can estimate consistently $P_i^0(a_i|\mathbf{x})$ using the following simple kernel estimator:

$$P_i^0(a_i|\mathbf{x}) = \frac{\sum_{m,t} 1\{a_{imt} = a_i\} \, K\left(\dfrac{\mathbf{x}_{mt} - \mathbf{x}}{b_n}\right)}{\sum_{m,t} K\left(\dfrac{\mathbf{x}_{mt} - \mathbf{x}}{b_n}\right)}$$

Second, given that $\mathbf{P}^0$ is identified, we can identify also the expected present values $\tilde{z}_i^{\mathbf{P}^0}(a_i, \mathbf{x})$ at the "true" equilibrium in the population. Third, we know that $\mathbf{P}^0$ is an equilibrium associated to $\boldsymbol{\theta}^0$. Therefore, the following equilibrium conditions should hold: for any $(i, a_i, \mathbf{x})$,

$$P_i^0(a_i|\mathbf{x}) = \frac{\exp\left\{\tilde{z}_i^{\mathbf{P}^0}(a_i, \mathbf{x})\dfrac{\boldsymbol{\theta}_i^0}{\sigma_\varepsilon^0}\right\}}{\sum_{a'=0}^{A} \exp\left\{\tilde{z}_i^{\mathbf{P}^0}(a', \mathbf{x})\dfrac{\boldsymbol{\theta}_i^0}{\sigma_\varepsilon^0}\right\}}$$

It is straightforward to show that under Assumption ID-3, these equilibrium conditions identify $\dfrac{\boldsymbol{\theta}_i^0}{\sigma_\varepsilon^0}$. For instance, in this logit example, we have that for $(i, a_i, \mathbf{x})$,

$$\ln\left(\frac{P_i^0(a_i|\mathbf{x})}{P_i^0(0|\mathbf{x})}\right) = \left[\tilde{z}_i^{\mathbf{P}^0}(a_i, \mathbf{x}) - \tilde{z}_i^{\mathbf{P}^0}(0, \mathbf{x})\right]\frac{\boldsymbol{\theta}_i^0}{\sigma_\varepsilon^0}$$

Define $Y_{imt} \equiv \ln\left(\frac{P_i^0(a_{imt}|\mathbf{x}_{mt})}{P_i^0(0|\mathbf{x}_{mt})}\right)$ and $Z_{imt} \equiv \tilde{z}_i^{\mathbf{P}^0}(a_{imt}, \mathbf{x}_{mt}) - \tilde{z}_i^{\mathbf{P}^0}(0, \mathbf{x}_{mt})$. Then,

$$Y_{imt} = Z_{imt}\,\frac{\boldsymbol{\theta}_i^0}{\sigma_\varepsilon^0}$$

And we can also write this system as, $E(Z'_{imt}Y_{imt}) = E(Z'_{imt}Z_{imt}) \dfrac{\boldsymbol{\theta}_i^0}{\sigma_\varepsilon^0}$. Under assumption ID-3:

$$\frac{\boldsymbol{\theta}_i^0}{\sigma_\varepsilon^0} = E(Z'_{imt}Z_{imt})^{-1}E(Z'_{imt}Y_{imt})$$

and $\dfrac{\boldsymbol{\theta}_i^0}{\sigma_\varepsilon^0}$ is identified.

Note that under the single-equilibrium-in-the-data assumption, the multiplicity of equilibria in the model does not play any role in the identification of the structural parameters. The single-equilibrium-in-the-data assumption is a sufficient for identification but it is not necessary. Sweeting (2009), Aguirregabiria and Mira (2009), and Paula and Tang (2010) present conditions for the point-identification of games of incomplete information when there are multiple equilibria in the data.

Estimation

The use of a 'extended' or 'pseudo' likelihood (or alternatively GMM criterion) function plays an important role in the different estimation methods. For arbitrary values of the vector of structural parameters $\boldsymbol{\theta}$ and firms' strategies $\mathbf{P}$, we define the following likelihood function of observed players' actions $\{a_{imt}\}$ conditional on observed state variables $\{\mathbf{x}_{mt}\}$:

$$Q(\boldsymbol{\theta}, \mathbf{P}) = \sum_{i,m,t} \sum_{a_i=0}^{A} 1\{a_{imt} = a_i\} \ \ln \Psi_i(a_i, \mathbf{x}_{mt}; \mathbf{P}, \boldsymbol{\theta})$$

We call $Q(\boldsymbol{\theta}, \mathbf{P})$ a 'Pseudo' Likelihood function because players' CCPs in $\mathbf{P}$ are arbitrary and do not represent the equilibrium probabilities associated with $\boldsymbol{\theta}$ implied by the model.

An important implication of using arbitrary CCPs, instead of equilibrium CCPs, is that likelihood Q is a function and not a correspondence. To compute this pseudo likelihood, a useful construct is the representation of equilibrium in terms of CCPs, which I presented above.

We could also consider a Pseudo GMM Criterion function:

$$Q(\boldsymbol{\theta}, \mathbf{P}) = -r(\boldsymbol{\theta}, \mathbf{P})' \ \boldsymbol{\Omega} \ r(\boldsymbol{\theta}, \mathbf{P})$$

where $\boldsymbol{\Omega}$ is the weighting matrix and $r(\boldsymbol{\theta}, \mathbf{P})$ is the vector of moment conditions:

$$r(\boldsymbol{\theta}, \mathbf{P}) = \frac{1}{MT} \sum_{m,t} \left[h(x_{mt}) \otimes \begin{pmatrix} 1\{a_{imt} = a_i\} - \Psi_i(a_i, \mathbf{x}_{mt}; \mathbf{P}, \boldsymbol{\theta}) \\ ... \\ \text{for any } (i, a_i) \end{pmatrix} \right]$$

and $h(x_{mt})$ is a vector of functions of x_{mt} (instruments).

Full Maximum Likelihood. The dynamic game imposes the restriction that the strategies in $\mathbf{P}$ should be in equilibrium. The ML estimator is defined as the pair $(\hat{\boldsymbol{\theta}}_{MLE}, \hat{\mathbf{P}}_{MLE})$ that maximizes the pseudo likelihood subject to the constraint that the strategies in $\hat{\mathbf{P}}_{MLE}$ are equilibrium strategies associated with $\hat{\boldsymbol{\theta}}_{MLE}$. That is,

$$(\hat{\boldsymbol{\theta}}_{MLE}, \hat{\mathbf{P}}_{MLE}) \quad = \quad \arg\max_{(\boldsymbol{\theta}, \mathbf{P})} \ Q(\boldsymbol{\theta}, \mathbf{P})$$

$$\text{s.t. } P_i(a_i|\mathbf{x}) = \Psi_i(a_i, \mathbf{x}; \mathbf{P}, \boldsymbol{\theta}) \text{ for any } (i, a_i, \mathbf{x})$$

This is a constrained ML estimator that satisfies the standard regularity conditions for consistency, asymptotic normality and efficiency of ML estimation.

The numerical solution of the constrained optimization problem that defines these estimators requires one to search over an extremely large dimensional space. In the empirical applications of dynamic oligopoly games, the vector of probabilities $\mathbf{P}$ includes thousands or millions of elements. Searching for an optimum in that kind of space is computationally demanding. Su and Judd (2008) have proposed to use a MPEC algorithm, which is a general purpose algorithm for the numerical solution of constrained optimization problems. However, even using the most sophisticated algorithm such as MPEC, the optimization with respect to $(\mathbf{P}, \boldsymbol{\theta})$ can be extremely demanding when $\mathbf{P}$ has a high dimension.

Two-step methods. Let $\mathbf{P}^0$ be the vector with the population values of the probabilities $P_i^0(a_i|\mathbf{x}) \equiv \Pr(a_{imt} = a_i | \mathbf{x}_{mt} = \mathbf{x})$. Under the assumptions of "no unobserved common knowledge variables" and "single equilibrium in the data", the CCPs in $\mathbf{P}^0$ represent also firms' strategies in the only equilibrium that is played in the data. These probabilities can be estimated consistently using standard nonparametric methods. Let $\hat{\mathbf{P}}^0$ be a consistent nonparametric estimator of $\mathbf{P}^0$. Given $\hat{\mathbf{P}}^0$, we can construct a consistent estimator of $\tilde{z}_i^{\mathbf{P}^0}(a_i, \mathbf{x})$. Then, the two-step estimator of $\boldsymbol{\theta}^0$ is defined as:

$$\hat{\boldsymbol{\theta}}_{2S} = \arg\max_{\boldsymbol{\theta}} Q(\boldsymbol{\theta}, \hat{\mathbf{P}}^0)$$

After the computation of the expected present values $\tilde{z}_i^{\mathbf{P}^0}(a_i, \mathbf{x})$, this second step of the procedure is computationally very simple. It consists just in the estimation of a standard discrete choice model, e.g., a binary probit/logit in our entry-exit example, or a conditional logit in our example with quality choice. Under standard regularity conditions, this two-step estimator is root-M consistent and asymptotically normal.

This idea was originally exploited, for estimation of single agent problems, by Hotz and Miller (1993) and Hotz, Miller, Sanders and Smith (1994). It was expanded to the estimation of dynamic games by Aguirregabiria and Mira (2007), Bajari, Benkard and Levin (2007), Pakes, Ostrovsky and Berry (2007), and Pesendorfer and Schmidt-Dengler (2008).

The main advantage of these two-step estimators is their computational simplicity. The first step is a simple nonparametric regression, and the second step is the estimation of a standard discrete choice model with a criterion function that in most applications is globally concave (e.g., such as the likelihood of a standard probit model in our entry-exit example). The main computational burden comes from the calculation of the present values $W_i^{\hat{\mathbf{P}}}(\mathbf{x})$. Though the computation of these present values may be subject to a curse of dimensionality, the cost of obtaining a two-step estimator is several orders of magnitude smaller than solving (just once) for an equilibrium of the dynamic game. In most applications, this makes the difference between being able to estimate the model or not.

However, these two-step estimators have some important limitations. A first limitation is the restrictions imposed by the assumption of no unobserved common knowledge variables. Ignoring persistent unobservables, if present, can generate important biases in the estimation of structural parameters. We deal with this issue later.

A second problem is finite sample bias. The finite sample bias of the two-step estimator of $\boldsymbol{\theta}^0$ depends very importantly on the properties of the first-step estimator of $\mathbf{P}^0$. In particular, it depends on the rate of convergence and on the variance and bias of $\hat{\mathbf{P}}^0$. It is well-known that there is a **curse of dimensionality in the nonparametric estimation** of a regression function such as $\mathbf{P}^0$. The rate of convergence of the estimator (and its asymptotic variance) declines (increase) with the number of explanatory variables in the regression. The initial nonparametric estimator can be very imprecise in the samples available in actual applications, and this can generate serious finite sample biases in the two-step estimator of structural parameters.

In dynamic games with heterogeneous players, the number of observable state variables is proportional to the number of players and therefore the so called *curse of dimensionality in nonparametric estimation* (and the associated bias of the two-step estimator) can be particularly serious. For instance, in our dynamic game of product quality choice, the vector of state variables contains the qualities of the N firms.

The source of this bias is well understood in two-step methods: $\hat{\mathbf{P}}$ enters nonlinearly in the sample moment conditions that define the estimator, and the expected value of a nonlinear function of $\hat{\mathbf{P}}$ is not equal to that function evaluated at the expected value of $\hat{\mathbf{P}}$. The larger the variance or the bias of $\hat{\mathbf{P}}$, the larger the bias of the two-step estimator of $\boldsymbol{\theta}_0$. To see this, note that the PML or GMM estimators in the second step are based on moment conditions at the true $\mathbf{P}^0$:

$$E\left(\ h(x_{mt})\ [1\{a_{imt} = a_i\} - \Psi_i(a_i, \mathbf{x}; \mathbf{P^0}, \boldsymbol{\theta})]\ \right)\ =\ 0$$

The same moment conditions evaluated at $\hat{\mathbf{P}}^0$ do not hold because of the estimation error:

$$E\left(\ h(x_{mt})\ \left[1\{a_{imt} = a_i\} - \Psi_i(a_i, \mathbf{x}; \hat{\mathbf{P}}^{\mathbf{0}}, \boldsymbol{\theta}_0)\right]\ \right)\ \neq\ 0$$

This generates a finite sample bias. The best response function $\Psi_i(a_i, \mathbf{x}; \hat{\mathbf{P}}^{\mathbf{0}}, \boldsymbol{\theta}_0)$ is a nonlinear function of the random vector $\hat{\mathbf{P}}_0$, and the expected value of a nonlinear function is not equal to the function evaluated at the expected value. The largest the finite sample bias or the variance of $\hat{\mathbf{P}}^{\mathbf{0}}$, the largest the bias of the two-step estimation of $\boldsymbol{\theta}_0$.

Recursive K-step estimators. To deal with finite sample bias, Aguirregabiria and Mira (2002, 2007) consider a recursive K-step extension. Given the two-step estimator $\hat{\boldsymbol{\theta}}_{2S}$ and the initial nonparametric estimator of CCPs, $\hat{\mathbf{P}}^0$, we can construct a new estimator of CCPs, $\hat{\mathbf{P}}^1$, such that, for any $(i, a_i, \mathbf{x})$:

$$\hat{P}_i^1(a_i|\mathbf{x}) = \Psi_i(a_i, \mathbf{x}; \hat{\mathbf{P}}^{\mathbf{0}}, \hat{\boldsymbol{\theta}}_{2S})$$

or in our example:

$$\hat{P}_i^1(a_i|\mathbf{x}) = \frac{\exp\left\{\tilde{z}_i^{\hat{\mathbf{P}}^{\mathbf{0}}}(a_i, \mathbf{x})\ \hat{\boldsymbol{\theta}}_{i,2S}\right\}}{\sum_{a'=0}^{A} \exp\left\{\tilde{z}_i^{\hat{\mathbf{P}}^{\mathbf{0}}}(a', \mathbf{x})\ \hat{\boldsymbol{\theta}}_{i,2S}\right\}}$$

This new estimator of CCPs exploits the parametric structure of the model, and the structure of best response functions. It seems intuitive that this new estimator of CCPs has better statistical properties than the initial nonparametric estimator, i.e., smaller asymptotic variance, and smaller finite sample bias and variance. As we explain below, this intuition is correct as long as the equilibrium that generated the data is (Lyapunov) stable.

Under this condition, it seems natural to obtain a new two-step estimator by replacing $\hat{\mathbf{P}}^0$ with $\hat{\mathbf{P}}^1$ as the estimator of CCPs. Then, we can obtain the new estimator:

$$\hat{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta}} Q(\boldsymbol{\theta}, \hat{\mathbf{P}}^1)$$

The same argument can be applied recursively to generate a sequence of $K - step$ estimators. Given an initial consistent nonparametric estimator $\hat{\mathbf{P}}^0$, the sequence of estimators $\{\hat{\boldsymbol{\theta}}^K, \hat{\mathbf{P}}^K : K \geq 1\}$ is defined as:

$$\hat{\boldsymbol{\theta}}^K = \arg\max_{\boldsymbol{\theta}} Q(\boldsymbol{\theta}, \hat{\mathbf{P}}^{K-1})$$

and

$$\hat{\mathbf{P}}^K = \Psi(\hat{\mathbf{P}}^{K-1}, \hat{\boldsymbol{\theta}}^K)$$

Monte Carlo experiments in Aguirregabiria and Mira (2002, 2007), and Kasahara and Shimotsu (2008a, 2009) show that iterating in the NPL mapping can reduce significantly the finite sample bias of the two-step estimator. The Monte Carlo experiments in Pesendorfer and Schmidt-Dengler (2008) present a different, more mixed, picture. While for some of their experiments NPL iteration reduces the bias, in other experiments the bias remains

constant or even increases. A closer look at the Monte Carlo experiments in Pesendorfer and Schmidt-Dengler shows that the NPL iterations provide poor results in those cases where the equilibrium that generates the data is not (Lyapunov) stable. As we explain below, this is not a coincidence. It turns out that the computational and statistical properties of the sequence of K-step estimators depend critically on the stability of the NPL mapping around the equilibrium in the data.

Convergence properties of recursive K-step estimators. To study the properties of these K-step estimators, it is convenient to represent the sequence $\{\hat{\mathbf{P}}^K : K \geq 1\}$ as the result of iterating in a fixed point mapping. For arbitrary $\mathbf{P}$, define the mapping:

$$\varphi(\mathbf{P}) \equiv \Psi(\mathbf{P},\ \hat{\boldsymbol{\theta}}(\mathbf{P}))$$

where $\hat{\boldsymbol{\theta}}(\mathbf{P}) \equiv \arg\max_{\boldsymbol{\theta}} Q(\boldsymbol{\theta}, \mathbf{P})$. The mapping $\varphi(\mathbf{P})$ is called the Nested Pseudo Likelihood (NPL) mapping. The sequence of estimators $\{\hat{\mathbf{P}}^K : K \geq 1\}$ can be obtained by successive iterations in the mapping φ starting with the nonparametric estimator $\hat{\mathbf{P}}^0$, i.e., for $K \geq 1$, $\hat{\mathbf{P}}^K = \varphi(\hat{\mathbf{P}}^{K-1})$.

Lyapunov stability. Let $\mathbf{P}^*$ be a fixed point of the NPL mapping such that $\mathbf{P}^* = \varphi(\mathbf{P}^*)$. We say that the mapping φ is Lyapunov-stable around the fixed point $\mathbf{P}^*$ if there is a neighborhood of $\mathbf{P}^*$, $\mathcal{N}$, such that successive iterations in the mapping φ starting at $\mathbf{P} \in \mathcal{N}$ converge to $\mathbf{P}^*$. A necessary and sufficient condition for Lyapunov stability is that the *spectral radius* of the Jacobian matrix $\partial\varphi(\mathbf{P}^*)/\partial\mathbf{P}'$ is smaller than one. The neighboring set $\mathcal{N}$ is denoted the dominion of attraction of the fixed point $\mathbf{P}^*$. The spectral radius of a matrix is the maximum absolute eigenvalue. If the mapping φ is twice continuously differentiable, then the spectral radius is a continuous function of $\mathbf{P}$. Therefore, if φ is Lyapunov stable at $\mathbf{P}^*$, for any $\mathbf{P}$ in the dominion of attraction of $\mathbf{P}^*$ we have that the spectral radius of $\partial\varphi(\mathbf{P})/\partial\mathbf{P}'$ is also smaller than one. Similarly, if $\mathbf{P}^*$ is an equilibrium of the mapping $\Psi(.,\boldsymbol{\theta})$, we say that this mapping is Lyapunov stable around $\mathbf{P}^*$ if and only if the *spectral radius* of the Jacobian matrix $\partial\Psi(\mathbf{P}^*,\boldsymbol{\theta})/\partial\mathbf{P}'$ is smaller than one.

There is a relationship between the stability of the NPL mapping and of the equilibrium mapping $\Psi(.,\boldsymbol{\theta}^0)$ around $\mathbf{P}^0$ (i.e., the equilibrium that generates the data). The Jacobian matrices of the NPL and equilibrium mapping are related by the following expression (see Kasahara and Shimotsu, 2009):

$$\frac{\partial\varphi(\mathbf{P}^0)}{\partial\mathbf{P}'} = M(\mathbf{P}^0)\ \frac{\partial\Psi(\mathbf{P}^0,\boldsymbol{\theta}^0)}{\partial\mathbf{P}'}$$

where $M(\mathbf{P}^0)$ is an idempotent projection matrix $I - \Psi_\theta(\Psi_\theta'\ diag\{\mathbf{P}^0\}^{-1}\ \Psi_\theta)^{-1}\Psi_\theta'\ diag\{\mathbf{P}^0\}^{-1}$, where $\Psi_\theta \equiv \partial\Psi(\mathbf{P}^0,\boldsymbol{\theta}^0)/\partial\theta'$. In single-agent dynamic programming models, the Jacobian matrix $\partial\Psi(\mathbf{P}^0,\boldsymbol{\theta}^0)/\partial\mathbf{P}'$ is zero (i.e., zero Jacobian matrix property, Aguirregabiria and

Mira, 2002). Therefore, for that class of models $\partial \varphi \left(\mathbf{P}^0 \right) / \partial \mathbf{P}' = 0$ and the NPL mapping is Lyapunov stable around $\mathbf{P}^0$. In dynamic games, $\partial \Psi \left(\mathbf{P}^0, \boldsymbol{\theta}^0 \right) / \partial \mathbf{P}'$ is not zero. However, given that $M(\mathbf{P}^0)$ is an idempotent matrix, it is possible to show that the spectral radius of $\partial \varphi \left(\mathbf{P}^0 \right) / \partial \mathbf{P}'$ is not larger than the spectral radius of $\partial \Psi(\mathbf{P}^0, \boldsymbol{\theta}^0) / \partial \mathbf{P}'$. Therefore, Lyapunov stability of $\mathbf{P}^0$ in the equilibrium mapping implies stability of the NPL mapping.

Convergence of NPL iterations. Suppose that the true equilibrium in the population, $\mathbf{P}^0$, is Lyapunov stable with respect to the NPL mapping. This implies that with probability approaching one, as M goes to infinity, the (sample) NPL mapping is stable around a consistent nonparametric estimator of $\mathbf{P}^0$. Therefore, the sequence of K-step estimators converges to a limit $\hat{\mathbf{P}}^0_{\text{lim}}$ that is a fixed point of the NPL mapping, i.e., $\hat{\mathbf{P}}^0_{\text{lim}} = \varphi(\hat{\mathbf{P}}^0_{\text{lim}})$. It is possible to show that this limit $\hat{\mathbf{P}}^0_{\text{lim}}$ is a consistent estimator of $\mathbf{P}^0$ (see Kasahara and Shimotsu, 2009). Therefore, under Lyapunov stability of the NPL mapping, if we start with a consistent estimator of $\mathbf{P}^0$ and iterate in the NPL mapping, we converge to a consistent estimator that is an equilibrium of the model. It is possible to show that this estimator is asymptotically more efficient than the two-step estimator (Aguirregabiria and Mira, 2007).

Pesendorfer and Schmidt-Dengler (2010) present an example where the sequence of K-step estimators converges to a limit estimator that is not consistent. As implied by the results presented above, the equilibrium that generates the data in their example is not Lyapunov stable. The concept of Lyapunov stability of the best response mapping at an equilibrium means that if we marginally perturb players' strategies, and then allow players to best respond to the new strategies, then we will converge to the original equilibrium. To us this seems like a plausible equilibrium selection criterion. Ultimately, whether an unstable equilibrium is interesting depends on the application and the researchers taste. Nevertheless, at the end of this section we present simple modified versions of the NPL method that can deal with data generated from an equilibrium that is not stable.

Reduction of finite sample bias. Kasahara and Shimotsu (2008a, 2009) derive a second order approximation to the bias of the K-step estimators. They show that the key component in this bias is the distance between the first step and the second step estimators of $\mathbf{P}^0$, i.e., $\left\| \varphi \left(\hat{\mathbf{P}}^0 \right) - \hat{\mathbf{P}}^0 \right\|$. An estimator that reduces this distance is an estimator with lower finite sample bias. Therefore, based on our discussion in point (b) above, the sequence of K-step estimators are decreasing in their finite sample bias if and only if the NPL mapping is Lyapunov stable around $\mathbf{P}^0$.

The Monte Carlo experiments in Pesendorfer and Schmidt-Dengler (2008) illustrate this point. They implement experiments using different DGPs: in some of them the data is generated from a stable equilibrium, and in others the data come from a non-stable equilibrium. It is simple to verify (see Aguirregabiria and Mira, 2010) that the experiments where NPL

iterations do not reduce the finite sample bias are those where the equilibrium that generates the data is not (Lyapunov) stable.

Modified NPL algorithms. Note that Lyapunov stability can be tested after obtaining the first NPL iteration. Once we have obtained the two-step estimator, we can calculate the Jacobian matrix $\partial\varphi(\hat{\mathbf{P}}^0)/\partial\mathbf{P}'$ and its eigenvalues, and then check whether Lyapunov stability holds at $\hat{\mathbf{P}}^0$.

If the applied researcher considers that his data may have been generated by an equilibrium that is not stable, then it will be worthwhile to compute this Jacobian matrix and its eigenvalues. If Lyapunov stability holds at $\hat{\mathbf{P}}^0$, then we know that NPL iterations reduce the bias of the estimator and converge to a consistent estimator.

When the condition does not hold, then the solution to this problem is not simple. Though the researcher might choose to use the two-step estimator, the non-stability of the equilibrium has also important negative implications on the properties of this simple estimator. Non-stability of the NPL mapping at $\mathbf{P}^0$ implies that the asymptotic variance of the two-step estimator of $\mathbf{P}^0$ is larger then asymptotic variance of the nonparametric reduced form estimator. To see this, note that the two-step estimator of CCPs is $\hat{\mathbf{P}}^1 = \varphi(\hat{\mathbf{P}}^0)$, and applying the delta method we have that $Var(\hat{\mathbf{P}}^1) = [\partial\varphi(\mathbf{P}^0)/\partial\mathbf{P}'] \, Var(\hat{\mathbf{P}}^0) \, [\partial\varphi(\mathbf{P}^0)/\partial\mathbf{P}']'$. If the spectral radius of $\partial\varphi(\mathbf{P}^0)/\partial\mathbf{P}'$ is greater than 1, then $Var(\hat{\mathbf{P}}^1) > Var(\hat{\mathbf{P}}^0)$. This is a puzzling result because the estimator $\hat{\mathbf{P}}^0$ is nonparametric while the estimator $\hat{\mathbf{P}}^1$ exploits most of the structure of the model. Therefore, the non-stability of the equilibrium that generates the data is an issue for this general class of two-step or sequential estimators.

In this context, Kasahara and Shimotsu (2009) propose alternative recursive estimators based on fixed-point mappings other than the NPL that, by construction, are stable. Iterating in these alternative mappings is significantly more costly than iterating in the NPL mapping, but these iterations guarantee reduction of the finite sample bias and convergence to a consistent estimator.

Aguirregabiria and Mira (2010) propose two modified versions of the NPL algorithm that are simple to implement and that always converge to a consistent estimator with better properties than two-step estimators. A *first modified-NPL-algorithm* applies to dynamic games. The first NPL iteration is standard but in every successive iteration best response mappings are used to update guesses of each player's own future behavior without updating beliefs about the strategies of the other players. This algorithm always converges to a consistent estimator even if the equilibrium generating the data is not stable and it reduces monotonically the asymptotic variance and the finite sample bias of the two-step estimator.

The *second modified-NPL-algorithm* applies to static games and it consists in the application of the standard NPL algorithm both to the best response mapping and to the inverse

of this mapping. If the equilibrium that generates the data is unstable in the best response mapping, it should be stable in the inverse mapping. Therefore, the NPL applied to the inverse mapping should converge to the consistent estimator and should have a largest value of the pseudo likelihood that the estimator that we converge to when applying the NPL algorithm to the best response mapping. Aguirregabiria and Mira illustrate the performance of these estimators using the examples in Pesendorfer and Schmidt-Dengler (2008 and 2010).

Estimation using Moment Inequalities. Bajari, Benkard, and Levin (2007) proposed a two-step estimator in the spirit of the ones described before but with two important differences:

(a) they used moment inequalities (instead of moment equalities);

(b) they do not calculate exactly the present value $\mathbf{W}_i^{\mathbf{P}}(\mathbf{x}_t)$ but they approximate them using Monte Carlo simulation.

(a) and (b) are two different ideas than can be applied separately. In my opinion, these two ideas have different merits and therefore I will discuss them separately.

Estimation using Moment Inequalities. Remember that $V_i^{\mathbf{P}}(\mathbf{x}_t)$ is the value of player i at state $\mathbf{x}_t$ when all the players behave according to their strategies in $\mathbf{P}$. In a model where the one-period payoff function is multiplicatively separable in the structural parameters, we have that

$$V_i^{\mathbf{P}}(\mathbf{x}_t) \;\; = \;\; W_i^{\mathbf{P}}(\mathbf{x}_t) \; \boldsymbol{\theta}_i$$

and the matrix of present values $\mathbf{W}_i^{\mathbf{P}} \equiv \{W_i^{\mathbf{P}}(\mathbf{x}_t) : \mathbf{x}_t \in X\}$ can be obtained exactly as:

$$\mathbf{W}_i^{\mathbf{P}} \equiv \left(\mathbf{I} - \beta \; \mathbf{F}_i^{\mathbf{P}}\right)^{-1} \bar{\mathbf{z}}_i^{\mathbf{P}}$$

For notational simplicity, I'll use $W_{it}^{\mathbf{P}}$ to represent $W_i^{\mathbf{P}}(\mathbf{x}_t)$.

Let's split the vector of choice probabilities $\mathbf{P}$ into the sub-vectors $\mathbf{P}_i$ and $\mathbf{P}_{-i}$,

$$\mathbf{P} \equiv (\mathbf{P}_i \; , \; \mathbf{P}_{-i})$$

where $\mathbf{P}_i$ are the probabilities associated to player i and $\mathbf{P}_{-i}$ contains the probabilities of players other than i. $\mathbf{P}^0$ is an equilibrium associated to $\boldsymbol{\theta}^0$. Therefore, $\mathbf{P}_i^0$ is firm i's best response to $\mathbf{P}_{-i}^0$. Therefore, for any $\mathbf{P}_i \neq \mathbf{P}_i^0$ the following inequality should hold:

$$W_{it}^{\left(\mathbf{P}_i^0, \mathbf{P}_{-i}^0\right)} \; \boldsymbol{\theta}_i^0 \geq W_{it}^{\left(\mathbf{P}_i, \mathbf{P}_{-i}^0\right)} \; \boldsymbol{\theta}_i^0$$

We can define an estimator of $\boldsymbol{\theta}^0$ based on these (moment) inequalities. There are infinite alternative policies $\mathbf{P}_i$, and therefore there are infinite moment inequalities. For estimation, we should select a finite set of alternative policies. This is a very important decision for this

class of estimators (more below). Let H be a **(finite) set of alternative policies** for each player. Define the following criterion function:

$$R\left(\boldsymbol{\theta}, \mathbf{P}^0\right) \equiv \sum_{i,m,t} \sum_{\mathbf{P} \in H} \left(\min\left\{ 0 \; ; \; \left[W_{imt}^{(\mathbf{P}_i^0, \mathbf{P}_{-i}^0)} - W_{imt}^{(\mathbf{P}_i, \mathbf{P}_{-i}^0)} \right] \boldsymbol{\theta}_i \right\} \right)^2$$

This criterion function penalizes departures from the inequalities. Then, given an initial NP estimator of $\mathbf{P}^0$, say $\hat{\mathbf{P}}^0$, we can define the following estimator of $\boldsymbol{\theta}^0$ based on moment inequalities (MI):

$$\hat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}} \; R\left(\boldsymbol{\theta}, \hat{\mathbf{P}}^0\right)$$

There are several relevant comments to make on this MI estimator: **(1)** Computational properties (relative to two-step ME estimators); **(2)** Point identification / Set identification; **(3)** How to choose the set of alternative policies?; **(4)** Statistical properties; **(5)** Continuous decision variables.

Computational Properties. **The two-step MI estimator is more computationally costly than a two-step ME estimator.** There at least three factors than contribute to this larger cost.

> **(a)** In both types of estimators, the main cost comes from calculating the present values $\mathbf{W}_i^{\mathbf{P}}$. In a 2-step ME estimator this evaluation is done once. In the MI estimator this is done as many times as alternative policies in the set H;
>
> **(b)** The ME criterion functions $Q\left(\boldsymbol{\theta}, \hat{\mathbf{P}}\right)$ is typically globally concave in $\boldsymbol{\theta}$, but $R\left(\boldsymbol{\theta}, \hat{\mathbf{P}}\right)$ is not;
>
> **(c)** Set estimation versus point estimation. The MI estimator needs an algorithm for set optimization.

MI Estimator: Point / Set identification. This estimator is based on exactly the same assumptions as the 2-step moment equalities (ME) estimator. We have seen that $\boldsymbol{\theta}^0$ **is point identified** by the moment equalities of the ME estimators (e.g., by the pseudo likelihood equations). Therefore, if the set H of alternative policies is large enough, then $\boldsymbol{\theta}^0$ should be point identified as the unique minimizer of $R\left(\boldsymbol{\theta}, \mathbf{P}^0\right)$. However, it is very costly to consider a set H with many alternative policies. For the type of H sets which are considered in practice, minimizing $R\left(\boldsymbol{\theta}, \mathbf{P}^0\right)$ does uniquely identifies $\boldsymbol{\theta}^0$. Therefore, $\boldsymbol{\theta}^0$ **is set identified.**

How to choose the set of alternative policies? The choice of the alternative policies in the set H plays a key role in the statistical properties (e.g., precision, bias) of this estimator. However, there is a clear rule on how to select these policies.

Statistical properties of MI estimator (relative to ME). **The MI estimator is not more 'robust' than the ME estimator**. Both estimators are based on exactly the same model

and assumptions. Set identification. **Asymptotically, the MI estimator is less efficient than the ME estimator.** The efficient 2-step Moment Equalities (ME) estimator has lower asymptotic variance than the MI estimator, even as the set H becomes very large.

Continuous decision variables. BBL show that, when combined with simulation techniques to approximate the values $\{W_{it}^{\mathbf{P}}\}$, the MI approach can be easily applied to the **estimation of dynamic games with continuous decision variables.** In fact, the BBL estimator of a model with continuous decision variable is basically the same as with a discrete decision variable. The ME estimator of models with continuous decision variable may be more complicated.

A different approach to construct inequalities in dynamic games. In a MPE a player equilibrium strategy is his best response not only within the class of Markov strategies but also within the class of non Markov strategies: e.g., strategies that vary over time. Maskin and Tirole: if all the other players use Markov strategies, a player does not have any gain from using non Markov strategies.

Suppose that to construct the inequalities $W_{it}^{\left(\mathbf{P}_i^0, \mathbf{P}_{-i}^0\right)} \boldsymbol{\theta}_i^0 \geq W_{it}^{\left(\mathbf{P}_i, \mathbf{P}_{-i}^0\right)} \boldsymbol{\theta}_i^0$ we use alternative strategies which are non-Markov. In a MPE a player equilibrium strategy is his best response not only within the class of Markov strategies but also within the class of non Markov strategies: e.g., strategies that vary over time. Maskin and Tirole: if all the other players use Markov strategies, a player does not have any gain from using non Markov strategies. More specifically, suppose that the alternative strategy of player i is

$$\mathbf{P}_i = \{P_{it}(\mathbf{x}_t) : t = 1, 2, 3, ...; \ \mathbf{x}_t \in X\}$$

with the following features.

(a) Two-periods deviation: $P_{it}(.) \neq P_i^0(.)$, $P_{it+1}(.) \neq P_i^0(.)$, but $P_{it+s}(.) = P_i^0(.)$ for any $s \geq 2$;

(b) $P_{it+1}(.)$ is constructed in such a way that it compensates the effects of the perturbation $P_{it}(.)$ on the distribution of $\mathbf{x}_{t+2}$ conditional on $\mathbf{x}_t$, i.e.,

$$F_x^{\mathbf{P}^0(2)}(\mathbf{x}_{t+2} \mid \mathbf{x}_t) = F_x^{\mathbf{P}^0(2)}(\mathbf{x}_{t+2} \mid \mathbf{x}_t)$$

Given this type of alternative policies, we have that the value differences

$$W_{it}^{\left(\mathbf{P}_i^0, \mathbf{P}_{-i}^0\right)} \boldsymbol{\theta}_i^0 \geq W_{it}^{\left(\mathbf{P}_i, \mathbf{P}_{-i}^0\right)} \boldsymbol{\theta}_i^0$$

only depends on differences between expected payoffs at periods t and $t + 1$. We do not have to use simulation, invert huge matrices, etc, and we can consider thousands (or even millions) of alternative policies.

Dealing with Unobserved Heterogeneity. So far, we have maintained the assumption that the only unobservables for the researcher are the private information shocks that are i.i.d. over firms, markets, and time. In most applications in IO, this assumption is not realistic and it can be easily rejected by the data. Markets and firms are heterogenous in terms of characteristics that are payoff-relevant for firms but unobserved to the researcher. Not accounting for this heterogeneity may generate significant biases in parameter estimates and in our understanding of competition in the industry.

For instance, in the empirical applications in Aguirregabiria and Mira (2007) and Collard-Wexler (2006), the estimation of a model without unobserved market heterogeneity implies estimates of strategic interaction between firms (i.e., competition effects) that are close to zero or even have the opposite sign to the one expected under competition. In both applications, including unobserved heterogeneity in the models results in estimates that show significant and strong competition effects.

Aguirregabiria and Mira (2007) and Arcidiacono and Miller (2008) have proposed methods for the estimation of dynamic games that allow for persistent unobserved heterogeneity in players or markets. Here we concentrate on the case of permanent unobserved market heterogeneity in the profit function.

$$\Pi_{imt} \;=\; z_i^{\mathbf{P}}(a_i, \mathbf{x}_{mt}) \, \theta_{ii}^{EC} - \sigma_{\xi i} \, \xi_m - \varepsilon_{imt}$$

$\sigma_{\xi i}$ is a parameter, and ξ_m is a time-invariant 'random effect' that is common knowledge to the players but unobserved to the researcher.

The distribution of this random effect has the following properties: (A.1) it has a discrete and finite support $\{\xi^1, \xi^2, \ldots, \xi^L\}$, each value in the support of ξ represents a 'market type', and we index market types by $\ell \in \{1, 2, ..., L\}$; (A.2) it is i.i.d. over markets with probability mass function $\lambda_\ell \equiv \Pr(\xi_m = \xi^\ell)$; and (A.3) it does not enter into the transition probability of the observed state variables, i.e., $\Pr(\mathbf{x}_{mt+1} \mid \mathbf{x}_{mt}, \mathbf{a}_{mt}, \xi_m) = F_x(\mathbf{x}_{mt+1} \mid \mathbf{x}_{mt}, \mathbf{a}_{mt})$. Without loss of generality, ξ_m has mean zero and unit variance because the mean and the variance of ξ_m are incorporated in the parameters θ_i^{FC} and $\sigma_{\xi i}$, respectively. Also, without loss of generality, the researcher knows the points of support $\{\xi^\ell : \ell = 1, 2, \ldots, L\}$ though the probability mass function $\{\lambda_\ell\}$ is unknown.

Assumption (A.1) is common when dealing with permanent unobserved heterogeneity in dynamic structural models. The discrete support of the unobservable implies that the contribution of a market to the likelihood (or pseudo likelihood) function is a finite mixture of likelihoods under the different possible best responses that we would have for each possible market type. With continuous support we would have an infinite mixture of best responses and this could complicate significantly the computation of the likelihood. Nevertheless, as

we illustrate before, using a pseudo likelihood approach and a convenient parametric specification of the distribution of ξ_m simplifies this computation such that we can consider many values in the support of this unobserved variable at a low computational cost. Assumption (A.2) is also standard when dealing with unobserved heterogeneity. Unobserved spatial correlation across markets does not generate inconsistency of the estimators that we present here because the likelihood equations that define the estimators are still valid moment conditions under spatial correlation. Incorporating spatial correlation in the model, if present in the data, would improve the efficiency of the estimator but at a significant computational cost. Assumption (A.3) can be relaxed, and in fact the method by Arcidiacono-Miller deals with unobserved heterogeneity both in payoffs and transition probabilities.

Each market type ℓ has its own equilibrium mapping (with a different level of profits given ξ^ℓ) and its own equilibrium. Let $\mathbf{P}_\ell$ be a vector of strategies (CCPs) in market-type ℓ: $\mathbf{P}_\ell \equiv \{P_{i\ell}(\mathbf{x}_t) : i = 1, 2, ..., N; \mathbf{x}_t \in \mathcal{X}\}$. The introduction of unobserved market heterogeneity also implies that we can relax the assumption of only 'a single equilibrium in the data' to allow for different market types to have different equilibria. It is straightforward to extend the description of an equilibrium mapping in CCPs to this model. A vector of CCPs $\mathbf{P}_\ell$ is a MPE for market type ℓ if and only if for every firm i and every state $\mathbf{x}_t$ we have that: $P_{i\ell}(\mathbf{x}_t) = \Phi\left(\tilde{\mathbf{z}}_i^{\mathbf{P}_\ell}(\mathbf{x}_t, \xi^\ell)\,\boldsymbol{\theta}_i + \tilde{e}_i^{\mathbf{P}_\ell}(\mathbf{x}_t, \xi^\ell)\right)$, where now the vector of structural parameters $\boldsymbol{\theta}_i$ is $\{\theta_{i,0}^{VP}, ..., \theta_{i,N-1}^{VP}, \theta_i^{FC}, \theta_i^{EC}, \sigma_{\xi i}\}$ that includes $\sigma_{\xi i}$, and the vector $\tilde{\mathbf{z}}_i^{\mathbf{P}_\ell}(\mathbf{x}_t, \xi^\ell)$ has a similar definition as before with the only difference that it has one more component associated with $-\xi^\ell$. Since the points of support $\{\xi^\ell : \ell = 1, 2, \ldots, L\}$ are known to the researcher, he can construct the equilibrium mapping for each market type.

Let λ be the vector of parameters in the probability mass function of ξ, i.e., $\lambda \equiv \{\lambda_\ell : \ell = 1, 2, ..., L\}$, and let $\mathbf{P}$ be the set of CCPs for every market type, $\{\mathbf{P}_\ell : \ell = 1, 2, ..., L\}$. The (conditional) pseudo log likelihood function of this model is $Q(\boldsymbol{\theta}, \lambda, \mathbf{P}) = \sum_{m=1}^M \log \Pr(\mathbf{a}_{m1}, \mathbf{a}_{m2}, ..., \mathbf{a}_{mT} \mid \mathbf{x}_{m1}, \mathbf{x}_{m2}, ..., \mathbf{x}_{mT}; \boldsymbol{\theta}, \lambda, \mathbf{P})$. We can write this function as $\sum_{m=1}^M \log q_m(\boldsymbol{\theta}, \lambda, \mathbf{P})$, where $q_m(\boldsymbol{\theta}, \lambda, \mathbf{P})$ is the contribution of market m to the pseudo likelihood:

$$q_m(\boldsymbol{\theta}, \boldsymbol{\lambda}, \mathbf{P}) \;=\; \sum_{\ell=1}^L \lambda_{\ell|\mathbf{x}_{m1}} \left[\prod_{i,t} \Phi\left(\tilde{\mathbf{z}}_{im\ell t}^{\mathbf{P}_\ell}\, \boldsymbol{\theta}_i + \tilde{e}_{im\ell t}^{\mathbf{P}_\ell} \right)^{a_{imt}} \Phi\left(-\tilde{\mathbf{z}}_{im\ell t}^{\mathbf{P}_\ell}\, \boldsymbol{\theta}_i - \tilde{e}_{im\ell t}^{\mathbf{P}_\ell} \right)^{1-a_{imt}} \right]$$

where $\tilde{\mathbf{z}}_{im\ell t}^{\mathbf{P}_\ell} \equiv \tilde{\mathbf{z}}_i^{\mathbf{P}_\ell}(\mathbf{x}_{mt}, \xi^\ell)$, $\tilde{e}_{im\ell t}^{\mathbf{P}_\ell} \equiv \tilde{e}_i^{\mathbf{P}_\ell}(\mathbf{x}_{mt}, \xi^\ell)$, and $\lambda_{\ell|\mathbf{x}}$ is the conditional probability $\Pr(\xi_m = \xi^\ell|\mathbf{x}_{m1} = \mathbf{x})$. The conditional probability distribution $\lambda_{\ell|x}$ is different from the unconditional distribution λ_ℓ. In particular, ξ_m is not independent of the predetermined endogenous state variables that represent market structure. For instance, we expect a negative correlation between the indicators of incumbent status, s_{imt}, and the unobserved component of the fixed cost ξ_m, i.e., markets where it is more costly to operate tend to have a smaller number of incumbent firms. This is the so called *initial conditions problem* (Heckman, 1981).

In short panels (for T relatively small), not taking into account this dependence between ξ_m and $\mathbf{x}_{m1}$ can generate significant biases, similar to the biases associated to ignoring the existence of unobserved market heterogeneity. There are different ways to deal with the initial conditions problem in dynamic models (Heckman, 1981). One possible approach is to derive the joint distribution of $\mathbf{x}_{m1}$ and ξ_m implied by the equilibrium of the model. That is the approach proposed and applied in Aguirregabiria and Mira (2007) and Collard-Wexler (2006). Let $\mathbf{p}^{\mathbf{P}_\ell} \equiv \{p^{\mathbf{P}_\ell}(\mathbf{x}_t) : \mathbf{x}_t \in \mathcal{X}\}$ be the ergodic or steady-state distribution of $\mathbf{x}_t$ induced by the equilibrium $\mathbf{P}_\ell$ and the transition F_x. This stationary distribution can be simply obtained as the solution to the following system of linear equations: for every value $\mathbf{x}_t \in \mathcal{X}$, $p^{\mathbf{P}_\ell}(\mathbf{x}_t) = \sum_{\mathbf{x}_{t-1} \in \mathcal{X}} p^{\mathbf{P}_\ell}(\mathbf{x}_{t-1}) \, F_x^{\mathbf{P}_\ell}(\mathbf{x}_t \mid \mathbf{x}_{t-1})$, or in vector form, $\mathbf{p}^{\mathbf{P}_\ell} = \mathbf{F}_x^{\mathbf{P}_\ell} \, \mathbf{p}^{\mathbf{P}_\ell}$ subject to $\mathbf{p}^{\mathbf{P}_\ell\prime}\mathbf{1} = 1$. Given the ergodic distributions for the L market types, we can apply Bayes' rule to obtain:

$$\lambda_{\ell|\mathbf{x}_{m1}} = \frac{\lambda_\ell \, p^{\mathbf{P}_\ell}(\mathbf{x}_{m1})}{\displaystyle\sum_{\ell'=1}^{L} \lambda_{\ell'} \, p^{\mathbf{P}_{\ell'}}(\mathbf{x}_{m1})} \tag{3.2}$$

Note that given the CCPs $\{\mathbf{P}_\ell\}$, this conditional distribution does not depend on parameters in the vector $\boldsymbol{\theta}$, only on the distribution λ. Given this expression for the probabilities $\{\lambda_{\ell|\mathbf{x}_{m1}}\}$, we have that the pseudo likelihood in (??) only depends on the structural parameters $\boldsymbol{\theta}$ and $\boldsymbol{\lambda}$ and the incidental parameters $\mathbf{P}$.

For the estimators that we discuss here, we maximize $Q(\boldsymbol{\theta}, \lambda, \mathbf{P})$ with respect to $(\boldsymbol{\theta}, \lambda)$ for given $\mathbf{P}$. Therefore, the ergodic distributions $p^{\mathbf{P}_\ell}$ are fixed during this optimization. This implies a significant reduction in the computational cost associated with the initial conditions problem. Nevertheless, in the literature of finite mixture models, it is well known that optimization of the likelihood function with respect to the mixture probabilities λ is a complicated task because the problem is plagued with many local maxima and minima. To deal with this problem, Aguirregabiria and Mira (2007) introduce an additional parametric assumption on the distribution of ξ_m that simplifies significantly the maximization of $Q(\boldsymbol{\theta}, \lambda, \mathbf{P})$ for fixed $\mathbf{P}$. They assume that the probability distribution of unobserved market heterogeneity is such that the only unknown parameters for the researcher are the mean and the variance which are included in θ_i^{FC} and $\sigma_{\xi i}$, respectively. Therefore, they assume that the distribution of ξ_m (i.e., the points of support and the probabilities λ_ℓ) are known to the researcher. For instance, we may assume that ξ_m has a discretized standard normal distribution with an arbitrary number of points of support L. Under this assumption, the pseudo likelihood function is maximized only with respect to $\boldsymbol{\theta}$ for given $\mathbf{P}$. Avoiding optimization with respect to λ simplifies importantly the computation of the different estimators that we describe below.

NPL estimator. As defined above, the NPL mapping $\varphi(.)$ is the composition of the equilibrium mapping and the mapping that provides the maximand in $\boldsymbol{\theta}$ to $Q(\boldsymbol{\theta}, \mathbf{P})$ for given $\mathbf{P}$. That is, $\varphi(\mathbf{P}) \equiv \Psi(\hat{\theta}(\mathbf{P}), \mathbf{P})$ where $\hat{\theta}(\mathbf{P}) \equiv \arg\max_{\boldsymbol{\theta}} Q(\boldsymbol{\theta}, \mathbf{P})$. By definition, an NPL fixed point is a pair $(\hat{\boldsymbol{\theta}}, \hat{\mathbf{P}})$ that satisfies two conditions: (a) $\hat{\boldsymbol{\theta}}$ maximizes $Q(\boldsymbol{\theta}, \hat{\mathbf{P}})$; and (b) $\hat{\mathbf{P}}$ is an equilibrium associated to $\hat{\boldsymbol{\theta}}$. The NPL estimator is defined as the NPL fixed point with the maximum value of the likelihood function. The NPL estimator is consistent under standard regularity conditions (Aguirregabiria and Mira, 2007, Proposition 2).

When the equilibrium that generates the data is Lyapunov stable, we can compute the NPL estimator using a procedure that iterates in the NPL mapping, as described in section 3.2 to obtain the sequence of K-step estimators (i.e., NPL algorithm). The main difference is that now we have to calculate the steady-state distributions $\mathbf{p}(\mathbf{P}_\ell)$ to deal with the initial conditions problem. However, the pseudo likelihood approach also reduces significantly the cost of dealing with the initial conditions problem. This NPL algorithm proceeds as follows. We start with L arbitrary vectors of players' choice probabilities, one for each market type: $\{\hat{\mathbf{P}}_\ell^0 : \ell = 1, 2, ..., L\}$. Then, we perform the following steps. Step 1: For every market type we obtain the steady-state distributions and the probabilities $\{\lambda_{\ell|x_{m1}}\}$. Step 2: We obtain a pseudo maximum likelihood estimator of $\boldsymbol{\theta}$ as $\hat{\boldsymbol{\theta}}^1 = \arg\max_{\boldsymbol{\theta}} Q(\boldsymbol{\theta}, \hat{\mathbf{P}}^0\})$. Step 3: Update the vector of players' choice probabilities using the best response probability mapping. That is, for market type ℓ, firm i and state $\mathbf{x}$, $\hat{P}_{i\ell}^1(\mathbf{x}) = \Phi(\tilde{\mathbf{z}}_i^{\hat{\mathbf{P}}_\ell^0}(\mathbf{x}, \xi^\ell)\hat{\boldsymbol{\theta}}_i^1 + \tilde{e}_i^{\hat{\mathbf{P}}_\ell^0}(\mathbf{x}, \xi^\ell))$. If, for every type ℓ, $||\hat{\mathbf{P}}_\ell^1 - \hat{\mathbf{P}}_\ell^0||$ is smaller than a predetermined small constant, then stop the iterative procedure and keep $\hat{\boldsymbol{\theta}}^1$ as a candidate estimator. Otherwise, repeat steps 1 to 4 using $\hat{\mathbf{P}}^1$ instead of $\hat{\mathbf{P}}^0$.

The NPL algorithm, upon convergence, finds an NPL fixed point. To guarantee consistency, the researcher needs to start the NPL algorithm from different CCP's in case there are multiple NPL fixed points. This situation is similar to using a gradient algorithm, designed to find a local root, in order to obtain an estimator which is defined as a global root. Of course, this global search aspect of the method makes it significantly more costly than the application of the NPL algorithm in models without unobserved heterogeneity. This is the additional computational cost that we have to pay for dealing with unobserved heterogeneity. Note, however, that this global search can be parallelized in a computer with multiple processors.

Arcidiacono and Miller (2008) extend this approach in several interesting and useful ways. First, they consider a more general form of unobserved heterogeneity that may enter both in the payoff function and in the transition of the state variables. Second, to deal with the complexity in the optimization of the likelihood function with respect to the distribution of the finite mixture, they combine the NPL method with an EM algorithm. Third, they

show that for a class of dynamic decision models, that includes but it is not limited to optimal stopping problems, the computation of the inclusive values $\tilde{\mathbf{z}}^{\mathbf{P}_\ell}_{im\ell t}$ and $\tilde{e}^{\mathbf{P}_\ell}_{im\ell t}$ is simple and it is not subject to a 'curse of dimensionality', i.e., the cost of computing these value for given $\mathbf{P}_\ell$ does not increase exponentially with the dimension of the state space. Together, these results provide a relatively simple approach to estimate dynamic games with unobserved heterogeneity of finite mixture type. Note that Lyapunov stability of each equilibrium type that generates the data is a necessary condition for the NPL and the Arcidiacono-Miller algorithms to converge to a consistent estimator.

Kasahara and Shimotsu (2008). The estimators of finite mixture models in Aguirregabiria and Mira (2007) and Arcidiacono and Miller (2008) consider that the researcher cannot obtain consistent nonparametric estimates of market-type CCPs $\{\mathbf{P}^0_\ell\}$. Kasahara and Shimotsu (2008b), based on previous work by Hall and Zhou (2003), have derived sufficient conditions for the nonparametric identification of market-type CCPs $\{\mathbf{P}^0_\ell\}$ and the probability distribution of market types, $\{\lambda^0_\ell\}$. Given the nonparametric identification of market-type CCPs, it is possible to estimate structural parameters using a two-step approach similar to the one described above. However, this two-step estimator has three limitations that do not appear in two-step estimators without unobserved market heterogeneity. First, the conditions for nonparametric identification of $\mathbf{P}^0$ may not hold. Second, the nonparametric estimator in the first step is a complex estimator from a computational point of view. In particular, it requires the minimization of a sample criterion function with respect to the large dimensional object $\mathbf{P}$. This is in fact the type of computational problem that we wanted to avoid by using two-step methods instead of standard ML or GMM. Finally, the finite sample bias of the two-step estimator can be significantly more severe when $\mathbf{P}^0$ incorporates unobserved heterogeneity and we estimate it nonparametrically.

4. Reducing the State Space

Although two-step and sequential methods are computationally much cheaper than full solution-estimation methods, they are still impractical for applications where the dimension of the state space is large. The cost of computing exactly the matrix of present values $\mathbf{W}^{\mathbf{P}}_{z,i}$ increases cubically with the dimension of the state space. In the context of dynamic games, the dimension of the state space increases exponentially with the number of heterogeneous players. Therefore, the cost of computing the matrix of present values may become intractable even for a relatively small number of players.

A simple approach to deal with this curse of dimensionality is to assume that players are homogeneous and the equilibrium is symmetric. For instance, in our dynamic game of market entry-exit, when firms are heterogeneous, the dimension of the state space is $|H| * 2^N$, where

$|H|$ is the number of values in the support of market size H_t. To reduce the dimensionality of the state space, we need to assume that: (a) only the number of competitors (and not their identities) affects the profit of a firm; (b) firms are homogeneous in their profit function; and (c) the selected equilibrium is symmetric. Under these conditions, the payoff relevant state variables for a firm i are $\{H_t, s_{it}, n_{t-1}\}$ where s_{it} is its own incumbent status, and n_{t-1} is the total number of active firms at period $t-1$. The dimension of the state space is $|H|*2*(N+1)$ that increases only linearly with the number of players.[1] It is clear that the assumption of homogeneous firms and symmetric equilibrium can reduce substantially the dimension of the state space, and it can be useful in some empirical applications. Nevertheless, there are many applications where this assumption is too strong. For instance, in applications where firms produce differentiated products.

To deal with this issue, Hotz, Miller, Sanders and Smith (1994) proposed an estimator that uses Monte Carlo simulation techniques to approximate the values $\mathbf{W}_{z,i}^{\mathbf{P}}$. Bajari, Benkard, and Levin (2007) have extended this method to dynamic games and to models with continuous decision variables. This approach has proved useful in some applications. Nevertheless, it is important to be aware that in those applications with large state spaces, simulation error can be sizeable and it can induce biases in the estimation of the structural parameters. In those cases, it is worthwhile to reduce the dimension of the state space by making additional structural assumptions. That is the general idea in the inclusive-value approach that we have discussed in section 2 and that can be extended to the estimation of dynamic games. Different versions of this idea have been proposed and applied by Nevo and Rossi (2008), Maceria (2007), Rossi (2009), and Aguirregabiria and Ho (2009).

To present the main ideas, we consider here a dynamic game of quality competition in the spirit of Pakes and McGuire (1994), the differentiated product version of Ericson-Pakes model. There are N firms in the market, that we index by i, and B brands or differentiated products, that we index by b. The set of brands sold by firm i is $\mathcal{B}_i \subset \{1, 2, ..., B\}$. Demand is given by a model similar to that of Section 2.1: consumers choose one of the B products offered in the market, or the outside good. The utility that consumer h obtains from purchasing product b at time t is $U_{hbt} = x_{bt} - \alpha\ p_{bt} + u_{hbt}$, where x_{bt} is the quality of the product, p_{bt} is the price, α is a parameter, and u_{hbt} represents consumer specific taste for product b. These idiosyncratic errors are identically and independently distributed over (h, b, t) with type I extreme value distribution. If the consumer decides not to purchase any of the goods, she chooses the outside option that has a mean utility normalized to zero. Therefore, the aggregate demand for product b is $q_{bt} = H_t \exp\{x_{bt} - \alpha\ p_{bt}\} [1 + \sum_{b'=1}^{B} \exp\{x_{b't} - \alpha p_{b't}\}]^{-1}$, where H_t represents market size at period t. The market

---

[1]This is a particular example of the 'exchangeability assumption' proposed by Pakes and McGuire (2001).

structure of the industry at time t is characterized by the vector $\mathbf{x}_t = (H_t, x_{1t}, x_{2t}, ..., x_{Bt})$. Every period, firms take as given current market structure and decide simultaneously their current prices and their investment in quality improvement. The one-period profit of firm i can be written as

$$\Pi_{it} = \sum_{b \in \mathcal{B}_i} (p_{bt} - mc_b)\, q_{bt} - FC_b - (c_b + \varepsilon_{bt})\, a_{bt} \qquad (4.1)$$

where $a_{bt} \in \{0, 1\}$ is the binary variable that represents the decision to invest in quality improvement of product b; mc_b, FC_b, and c_b are structural parameters that represent marginal cost, fixed operating cost, and quality investment cost for product b, respectively; and ε_{bt} is an iid private information shock in the investment cost. Product quality evolves according to a transition probability $f_x(x_{bt+1}|a_{bt}, x_{bt})$. For instance, in Pakes-McGuire model, $x_{bt+1} = x_{bt} - \zeta_t + a_{bt}\, v_{bt}$ where ζ_t and v_{bt} are two independent and non-negative random variables that are independently and identically distributed over (b, t).

In this model, price competition is static. The Nash-Bertrand equilibrium determines prices and quantities as functions of market structure $\mathbf{x}_t$, i.e., $p_b^*(\mathbf{x}_t)$ and $q_b^*(\mathbf{x}_t)$. Firms' quality choices are the result of a dynamic game. The one-period profit function of firm i in this dynamic game is $\Pi_i(\mathbf{a}_{it}, \mathbf{x}_t) = \sum_{b \in \mathcal{B}_i} (p_i^*(\mathbf{x}_t) - mc_b)\, q_b^*(\mathbf{x}_t) - FC_b - (c_b + \varepsilon_{bt})\, a_{bt}$, where $\mathbf{a}_{it} \equiv \{a_{bt} : b \in \mathcal{B}_i\}$. This dynamic game of quality competition has the same structure as the game that we have described in Section 3.1 and it can be solved and estimated using the same methods. However, the dimension of the state space increases exponentially with the number of products and the solution and estimation of the model becomes impractical even when B is not too large.

Define the *cost adjusted inclusive value* of firm i at period t as $\omega_{it} \equiv \log[\sum_{b \in \mathcal{B}_i} exp\{x_{bt} - \alpha\, mc_b\}]$. This value is closely related to the inclusive value that we have discussed in Section 2. It can be interpreted as the net quality level, or a value added of sort, that the firm is able to produce in the market. Under the assumptions of the model, the variable profit of firm i in the Nash-Bertrand equilibrium can be written as a function of the vector of inclusive values $\boldsymbol{\omega}_t \equiv (\omega_{1t}, \omega_{2t}, ..., \omega_{Nt}) \in \Omega$, i.e., , $\sum_{b \in \mathcal{B}_i} (p_i^*(\mathbf{x}_t) - mc_b)\, q_b^*(\mathbf{x}_t) = vp_i(\boldsymbol{\omega}_t)$. Therefore, the one-period profit Π_{it} is a function $\tilde{\Pi}_i(\mathbf{a}_{it}, \boldsymbol{\omega}_t)$. The following assumption is similar to Assumption A2 made in Section 2 and it establishes that given vector $\boldsymbol{\omega}_t$ the rest of the information contained in the in $\mathbf{x}_t$ is redundant for the prediction of future values of $\boldsymbol{\omega}$.

Assumption: The transition probability of the vector of inclusive values $\boldsymbol{\omega}_t$ from the point of view a firm (i.e., conditional on a firm's choice) is such that $\Pr(\boldsymbol{\omega}_{t+1} \mid \mathbf{a}_{it}, \mathbf{x}_t) = \Pr(\boldsymbol{\omega}_{t+1} \mid \mathbf{a}_{it}, \boldsymbol{\omega}_t)$.

Under these assumptions, $\boldsymbol{\omega}_t$ is the vector of payoff relevant state variables in the dynamic game. The dimension of the space Ω increases exponentially with the number of firms but not with the number of brands. Therefore, the dimension of Ω can be much smaller than the dimension of the original state space of $\mathbf{x}_t$ in applications where the number of brands is large relative to the number of firms.

Of course, the assumption of sufficiency of $\boldsymbol{\omega}_t$ in the prediction of next period $\boldsymbol{\omega}_{t+1}$ is not trivial. In order to justify it we can put quite strong restrictions on the stochastic process of quality levels. Alternatively, it can be interpreted in terms of limited information, and/or bounded rationality. For instance, a possible way to justify this assumption is that firms face the same type of computational burdens that we do. Limiting the information that they use in their strategies reduces a firm's computational cost of calculating a best response.

Note that the dimension of the space of $\boldsymbol{\omega}_t$ still increases exponentially with the number of firms. To deal with this curse of dimensionality, Aguirregabiria and Ho (2009) consider a stronger inclusive value / sufficiency assumption. Let vp_{it} the variable profit of firm i at period t. Assumption: $\Pr(\omega_{it+1}, vp_{it+1} \mid \mathbf{a}_{it}, \mathbf{x}_t) = \Pr(\omega_{it+1}, vp_{it+1} \mid \mathbf{a}_{it}, \omega_{it}, vp_{it})$. Under this assumption, the vector of payoff relevant state variables in the decision problem of firm i is (ω_{it}, vp_{it}) and the dimension of the space of (ω_{it}, vp_{it}) does not increase with the number of firms.

5. Counterfactual experiments with multiple equilibria

One of the attractive features of structural models is that they can be used to predict the effects of new counterfactual policies. This is a challenging exercise in a model with multiple equilibria. Under the assumption that our data has been generated by a single equilibrium, we can use the data to identify which of the multiple equilibria is the one that we observe. However, even under that assumption, we still do not know which equilibrium will be selected when the values of the structural parameters are different to the ones that we have estimated from the data. For some models, a possible approach to deal with this issue is to calculate all of the equilibria in the counterfactual scenario and then draw conclusions that are robust to whatever equilibrium is selected. However, this approach is of limited applicability in dynamic games of oligopoly competition because the different equilibria typically provide contradictory predictions for the effects we want to measure.

Here we describe a simple homotopy method that has been proposed in Aguirregabiria (2009) and applied in Aguirregabiria and Ho (2009). Under the assumption that the equilibrium selection mechanism, which is unknown to the researcher, is a smooth function of the structural parameters, we show how to obtain a Taylor approximation to the counterfactual

equilibrium. Despite the equilibrium selection function is unknown, a Taylor approximation of that function, evaluated at the estimated equilibrium, depends on objects that the researcher knows.

Let $\Psi(\boldsymbol{\theta}, \mathbf{P})$ be the equilibrium mapping such that an equilibrium associated with $\boldsymbol{\theta}$ can be represented as a fixed point $\mathbf{P} = \Psi(\boldsymbol{\theta}, \mathbf{P})$. Suppose that there is an equilibrium selection mechanism in the population under study, but we do not know that mechanism. Let $\boldsymbol{\pi}(\boldsymbol{\theta})$ be the selected equilibrium given $\boldsymbol{\theta}$. The approach here is quite agnostic with respect to this equilibrium selection mechanism: it only assumes that there is such a mechanism, and that it is a smooth function of $\boldsymbol{\theta}$. Since we do not know the mechanism, we do not know the form of the mapping $\boldsymbol{\pi}(\boldsymbol{\theta})$ for every possible $\boldsymbol{\theta}$. However, we know that the equilibrium in the population, $\mathbf{P}^0$, and the vector of the structural parameters in the population, $\boldsymbol{\theta}^0$, belong to the graph of that mapping, i.e., $\mathbf{P}^0 = \boldsymbol{\pi}(\boldsymbol{\theta}^0)$.

Let $\boldsymbol{\theta}^*$ be the vector of parameters under the counterfactual experiment that we want to analyze. We want to know the counterfactual equilibrium $\boldsymbol{\pi}(\boldsymbol{\theta}^*)$ and compare it to the factual equilibrium $\boldsymbol{\pi}(\boldsymbol{\theta}^0)$. Suppose that Ψ is twice continuously differentiable in θ and P. The following is the key assumption to implement the homotopy method that we describe here.

Assumption: The equilibrium selection mechanism is such that $\pi(.)$ is a continuous differentiable function within a convex subset of Θ that includes $\boldsymbol{\theta}^0$ and $\boldsymbol{\theta}^*$.

That is, the equilibrium selection mechanism does not "jump" between the possible equilibria when we move over the parameter space from $\boldsymbol{\theta}^0$ to $\boldsymbol{\theta}^*$. This seems a reasonable condition when the researcher is interested in evaluating the effects of a change in the structural parameters but "keeping constant" the same equilibrium type as the one that generates the data.

Under these conditions, we can make a Taylor approximation to $\boldsymbol{\pi}(\boldsymbol{\theta}^*)$ around $\boldsymbol{\theta}^0$ to obtain:

$$\boldsymbol{\pi}(\boldsymbol{\theta}^*) \;=\; \boldsymbol{\pi}\left(\boldsymbol{\theta}^0\right) + \frac{\partial \boldsymbol{\pi}\left(\boldsymbol{\theta}^0\right)}{\partial \boldsymbol{\theta}'}\left(\boldsymbol{\theta}^* - \boldsymbol{\theta}^0\right) + O\left(\left\|\boldsymbol{\theta}^* - \boldsymbol{\theta}^0\right\|^2\right) \tag{5.1}$$

We know that $\boldsymbol{\pi}\left(\boldsymbol{\theta}^0\right) = \mathbf{P}^0$. Furthermore, by the implicit function theorem, $\partial \boldsymbol{\pi}\left(\boldsymbol{\theta}^0\right)/\partial \boldsymbol{\theta}'$ $= \partial\Psi(\boldsymbol{\theta}^0, \mathbf{P}^0)/\partial \boldsymbol{\theta}' + \partial\Psi(\boldsymbol{\theta}^0, \mathbf{P}^0)/\partial \mathbf{P}' \; \partial \boldsymbol{\pi}\left(\boldsymbol{\theta}^0\right)/\partial \boldsymbol{\theta}'$. If $\mathbf{P}^0$ is not a singular equilibrium then $I - \partial\Psi(\boldsymbol{\theta}^0, \mathbf{P}^0)/\partial \mathbf{P}'$ is not a singular matrix and $\partial \boldsymbol{\pi}\left(\boldsymbol{\theta}^0\right)/\partial \boldsymbol{\theta}' = (I - \partial\Psi(\boldsymbol{\theta}^0, \mathbf{P}^0)/\partial \mathbf{P}')^{-1}$ $\partial\Psi(\boldsymbol{\theta}^0, \mathbf{P}^0)/\partial \boldsymbol{\theta}'$. Solving this expression into the Taylor approximation, we have the following approximation to the counterfactual equilibrium:

$$\hat{\mathbf{P}}^* = \hat{\mathbf{P}}^0 + \left(I - \frac{\partial\Psi(\hat{\boldsymbol{\theta}}^0, \hat{\mathbf{P}}^0)}{\partial \mathbf{P}'}\right)^{-1} \frac{\partial\Psi(\hat{\boldsymbol{\theta}}^0, \hat{\mathbf{P}}^0)}{\partial \boldsymbol{\theta}'}\left(\boldsymbol{\theta}^* - \hat{\boldsymbol{\theta}}^0\right) \tag{5.2}$$

where $(\hat{\boldsymbol{\theta}}^0, \hat{\mathbf{P}}^0)$ represents our consistent estimator of $(\boldsymbol{\theta}^0, \mathbf{P}^0)$. It is clear that $\hat{\mathbf{P}}^*$ can be computed given the data and $\boldsymbol{\theta}^*$. Under our assumptions, $\hat{\mathbf{P}}^*$ is a consistent estimator of the linear approximation to $\boldsymbol{\pi}(\boldsymbol{\theta}^*)$.

As in any Taylor approximation, the order of magnitude of the error depends on the distance between the value of the structural parameters in the factual and counterfactual scenarios. Therefore, this approach can be inaccurate when the counterfactual experiment implies a large change in some of the parameters. For these cases, we can combine the Taylor approximation with iterations in the equilibrium mapping. Suppose that $\mathbf{P}^*$ is a (Lyapunov) stable equilibrium. And suppose that the Taylor approximation $\hat{\mathbf{P}}^*$ belongs to the dominion of attraction of $\mathbf{P}^*$. Then, by iterating in the equilibrium mapping $\Psi(\boldsymbol{\theta}^*, .)$ starting at $\hat{\mathbf{P}}^*$ we will obtain the counterfactual equilibrium $\mathbf{P}^*$. Note that this approach is substantially different to iterating in the equilibrium mapping $\Psi(\boldsymbol{\theta}^*, .)$ starting with the equilibrium in the data $\hat{\mathbf{P}}^0$. This approach will return the counterfactual equilibrium $\mathbf{P}^*$ *if only if* $\hat{\mathbf{P}}^0$ belongs to the dominion of attraction of $\mathbf{P}^*$. This condition is stronger than the one establishing that the Taylor approximation $\hat{\mathbf{P}}^*$ belongs to the domination of attraction of $\mathbf{P}^*$.

Empirical Application: Environmental Regulation in the Cement Industry

Ryan studies the effects in the US cement industry of the 1990 Amendments to Air Clean Act. I have talked about this paper before in the course, and problem set #1 was inspired on this empirical application. In the problem set, we considered a static model of Cournot competition and market entry with homogeneous firms. Ryan's model extends that simple framework to dynamic game of oligopoly competition where firms compete in quantities but they also make investment decisions in capacity and in market entry/exit, and they are heterogeneous in their different costs, i.e.,marginal costs, fixed costs, capacity investment costs, and sunk entry costs.

Here, I will comment the following points of the paper. (a) Motivation and Empirical Questions; (b) The US Cement Industry; (c) The Regulation (Policy Change); (d) Empirical Strategy; (e) Data; (f) Model; (g) Estimation and Results.

5.1. Motivation and Empirical Questions. Most previous studies that measure the welfare effects of environmental regulation (ER) have ignored dynamic effects of these policies.

ER has potentially important effects on firms' entry and investment decisions, and, in turn, these can have important welfare effects.

This paper estimates a dynamic game of entry/exit and investment in the US Portland cement industry.

The estimated model is used to evaluate the welfare effects of the 1990 Amendments to the Clean Air Act (CAA).

5.2. The US Cement Industry. For the purpose of this paper, the most important features of the US cement industry are: (1) Indivisibilities in capacity investment, and economies of scale; (2) Highly polluting and energy intensive industry; and (3) Local competition, and highly concentrated local markets

Indivisibilities in capacity investment, and economies of scale. Portland cement is the binding material in concrete, which is a primary construction material. It is produced by first pulverizing limestone and then heating it at very high temperatures in a rotating kiln furnace. These kilns are the main piece of equipment. Plants can have one or more kilns (indivisibilities). Marginal cost increases rapidly when a kiln is close to full capacity.

Highly polluting and energy intensive industry. The industry generates a large amount of pollutants by-products. High energy requirements and pollution make the cement industry an important target of environmental policies.

Local competition, and highly concentrated local markets. Cement is a commodity difficult to store and transport, as it gradually absorbs water out of the air rendering it useless. This is the main reason why the industry is spatially segregated into regional markets. These regional markets are very concentrated.

5.3. The Regulation (Policy Change). In 1990, the Amendments to the Clean Air Act (CAA) added new categories of regulated emissions. Also, cement plants were required to undergo an environmental certification process. It has been the most important new environmental regulation affecting this industry in the last three decades. This regulation may have increased sunk costs, fixed operating costs or even investment costs in this industry.

5.4. Empirical Strategy. Previous evaluations of these policies have ignored effects on entry/exit and on firms' investment. They have found that the regulation contributed to reduce marginal costs and therefore prices. Positive effects on consumer welfare and total welfare. Ignoring effects on entry/exit and on firms' investment could imply an overestimate of these positive effects.

Specify a model of the cement industry, where oligopolists make optimal decisions over entry, exit, production, and investment given the strategies of their competitors. Estimate the model for the cement industry using a 20 year panel and allowing the structural parameters to differ before and after the 1990 regulation. Changes in cost parameters are attributed to the new regulation. The MPEs before and after the regulation are computed and they are used for welfare comparisons.

Comments on this empirical approach and its potential limitations: (a) anticipation of the policy; (b) technological change; (c) learning about the new policy.

5.5. Data. Period: 1980 to 1999 (20 years); 27 regional markets. Index local markets by m, plants by i and years by t.

$$Data = \{S_{mt}, \ W_{mt}, \ P_{mt}, \ n_{mt}, \ q_{imt}, \ i_{imt}, \ s_{imt}\}$$

S_{mt} = Market size; W_{mt} = Input prices (electricity prices, coal prices, natural gas prices, and manufacturing wages); P_{mt} = Output price; n_{mt} = Number of cement plants; q_{imt} = Quantity produced by plant i; s_{imt} = Capacity of plant i (number and capacity of kilns); i_{imt} = Investment in capacity by plant i.

5.6. Model. Regional homogenous-goods market. Firms compete in quantities in a static equilibrium, but they are subject to capacity constraints. Capacity is the most important strategic variable. Firms invest in future capacity and this decision is partly irreversible (and therefore dynamic). Incumbent firms also make optimal decisions over whether to exit.

Inverse demand curve (iso-elastic):

$$\log P_{mt} = \alpha_{mt} + \frac{1}{\epsilon} \log Q_{mt}$$

Production costs:

$$C(q_{imt}) \quad = \quad (MCOST + \omega_{imt}) \, q_{imt}$$

$$+ CAPCOST * I\left\{\frac{q_{imt}}{s_{imt}} > BINDING\right\} \left(\frac{q_{imt}}{s_{imt}} - BINDING\right)^2$$

s_{imt} = installed capacity; q_{imt}/s_{imt} = degree of capacity utilization; ω_{imt} = private information shock; $MCOST$, $CAPCOST$ and $BINDING$ are parameters.

Investment costs

$$IC_{imt} \quad = \quad I\{i_{imt} > 0\} \left(ADJPOS + INVMCPOS * i_{imt} + INVMCPOS2 * i_{imt}^2\right)$$

$$+ I\{i_{imt} < 0\} \left(ADJNEG + INVMCNEG * i_{imt} + INVMCNEG2 * i_{imt}^2\right)$$

Entry costs

$$EC_{imt} = I\{s_{imt} = 0 \text{ and } i_{imt} > 0\} \ \left(SUNK + \varepsilon_{imt}^{EC}\right)$$

In equilibrium, investment is a function:

$$i_{imt} = i(\alpha_{mt}, W_{mt}, s_{imt}, s_{-imt})$$

Similarly, entry and exit probabilities depend on $(\alpha_{mt}, W_{mt}, s_{imt}, s_{-imt}, \varepsilon_{imt}^{EC})$.

5.7. Estimation and Results. Estimation of demand curve. Includes local market region fixed effects (estimated with 19 observations per market). Instruments: local variation in input prices. The market specific demand shocks, A_{mt}, are estimated as residuals in this equation.

Estimation of variable production costs. From the Cournot equilibrium conditions. Firm specific cost shocks, ω_{imt}, are estimated as residuals in this equation.

Estimation of investment functions. Assumption:

$$i_{imt} = i(\alpha_{mt}, W_{mt}, s_{imt}, s_{-imt}) = i\left(\alpha_{mt}, W_{mt}, s_{imt}, \sum_{j\neq i} s_{jmt}\right)$$

6. Product repositioning in differentiated product markets

(Sweeting, 2007) To Be Completed

7. Dynamic Game of Airlines Network Competition

7.1. Motivation and Empirical Questions. An airline network is a description of the city-pairs (or airport pairs) that the airline connects with non-stop flights. The first goal of this paper is to develop a **dynamic game of network competition between airlines**, a model that can be estimated using publicly available data.

The model endogenizes airlines' networks, and the dynamics of these networks. Prices and quantities for each airline-route are also endogenous in the model. It extends previous work by Hendricks et al (1995, 1999) on airline networks, and previous literature on structural models of the airline industry: Berry (1990 and 1992), Berry, Carnall and Spiller (2006), Ciliberto and Tamer (2009).

The second of the paper is to apply this model to study empirically the contribution of demand, cost, and strategic factors to explain why most companies in the US airline industry operate using **hub-and-spoke networks**. The model incorporates **different hypotheses** that have been suggested in the literature to explain **hub-and-spoke networks**. We estimate the model and use counterfactual experiments to obtain the contribution of demand, costs and strategic factors.

Hub-and-Spoke Networks

Hypotheses that have been suggested in the literature to explain airlines' adoption of hub-spoke networks:

- **Demand:** Travellers may be willing to pay for the services associated with an airline's scale of operation in an airport.

- **Costs:** Economies of scale at the plane level (marginal costs); Economies of scope at the airport level (fixed costs and entry costs); Contracts with airports (fixed costs and entry costs).

- **Strategic:** Entry deterrence (Hendricks, Piccione and Tan, 1997).

The paper has several contributions to the literature on empirical dynamic games of oligopoly competition: (1) first application of dynamic network competition; (2) first paper to study empirically the strategic entry-deterrence aspect of hub-and-spoke networks; (3) first paper to apply the inclusive-values approach to a dynamic game; and (4) it proposes and implements a new method to make counterfactual experiments in dynamic games.

7.2. Model: Dynamic Game of Network Competition. N airlines and C cities, exogenously given. In our application, $N = 22$ and $C = 55$.

City-Pairs and Routes. Given the C cities, there are $M \equiv C(C-1)/2$ **non-directional city-pairs** (or markets). For each city-pair, an airline decides whether to operate non-stop flights. A **route** (or path) is a **directional round-trip between 2 cities**. A route may or may not have stops. A route-airline is a product, and there is a demand for each route-airline product. Airlines choose prices for each route they provide.

Networks. We index city-pairs by m, airlines by i, and time (quarters) by t. $x_{imt} \in \{0,1\}$ is a binary indicator for the event "airline i operates non-stop flights in city-pair m". $\boldsymbol{x}_{it} \equiv \{x_{imt} : m = 1,2,...,M\}$ is the network of airline i at period t. The network $\boldsymbol{x}_{it}$ describes all the routes (products) that the airline provides, and whether they are non-stop or stop routes. The industry network is $\boldsymbol{x}_t \equiv \{\boldsymbol{x}_{it} : i = 1,2,...,N\}$.

Airlines' Decisions. An airline network $\boldsymbol{x}_{it}$ determines the set of routes (products) that the airline provides, that we denote by $L(\boldsymbol{x}_{it})$. Every period, active airlines in a route compete in prices. Price competition determines variable profits for each airline. Every period (quarter), each airline decides also its network for next period. There is *time-to-build*. We represent this decision as $\boldsymbol{a}_{it} \equiv \{a_{imt} : m = 1,2,...,M\}$, though $a_{imt} \equiv x_{imt+1}$.

Profit Function. The airline's total profit function is:

$$\Pi_{it} = \sum_{r \in L(\boldsymbol{x}_{it})} (p_{irt} - c_{irt})q_{irt}$$

$$- \sum_{m=1}^{M} a_{imt} \left(FC_{imt} + (1 - x_{imt}) \ EC_{imt} \right)$$

$(p_{irt} - c_{irt})q_{irt}$ is the variable profit in route r. FC_{imt} and EC_{imt} are fixed cost and entry cost in city-pair m.

Network effects in demand and costs. An important feature of the model is that demand, variable costs, fixed costs, and entry costs depend on the scale of operation (number of connections) of the airline in the origin and destination airports of the city-pair. Let HUB_{imt} be the "hub size" of airline i in market m at period t as measured by the total number of connections to other cities that airline i has in the origin and destination cities of market m at the beginning of period t. This is the most important endogeneous state variable of this model. It is endogenous because, though HUB_{imt} does not depend on the entry-exit decision of the airline in market m, a_{imt-1}, it does depend on the airline's entry-exit decisions in any other market that has common cities with market m, $\{a_{im't-1}$ for $m' \neq m$ and markets m' and m have common cities$\}$.

This implies that markets are interconnected through these hub-size effects. Entry-exit in a market has implications of profits in other markets. An equilibrium of this model is an equilibrium for the whole airline industry and not only for a single city-pair.

Dynamic Game / Strategy Functions. Airlines maximize intertemporal profits, are forward-looking, and take into account the implications of their entry-exit decisions on future profits and on the expected future reaction of competitors. Airlines' strategies depend only on payoff-relevant state variables, i.e., Markov perfect equilibrium assumption. An airline's payoff-relevant information at quarter t is $\{\boldsymbol{x}_t, \boldsymbol{z}_t, \boldsymbol{\varepsilon}_{it}\}$. Let $\boldsymbol{\sigma} \equiv \{\sigma_i(\boldsymbol{x}_t, \boldsymbol{z}_t, \boldsymbol{\varepsilon}_{it}) : i = 1, 2, ..., N\}$ be a set of strategy functions, one for each airline. A MPE is a set of strategy functions such that each airline's strategy maximizes the value of the airline for each possible state and taking as given other airlines' strategies.

7.3. Data. Airline Origin and Destination Survey (DB1B) collected by the Office of Airline Information of the BTS. Period 2004-Q1 to 2004-Q4. $C = 55$ largest metropolitan areas. $N = 22$ airlines. City Pairs: $M = (55 * 54)/2 = 1,485$.

Airlines: Passengers and Markets

Airline (Code)	# Passengers (in thousands)	# City-Pairs in 2004-Q4 (maximum = 1,485)
1. Southwest (WN)	25,026	373
2. American (AA)[3]	20,064	233
3. United (UA)[4]	15,851	199
4. Delta (DL)[5]	14,402	198
5. Continental (CO)[6]	10,084	142
6. Northwest (NW)[7]	9,517	183
7. US Airways (US)	7,515	150
8. America West (HP)[8]	6,745	113
9. Alaska (AS)	3,886	32
10. ATA (TZ)	2,608	33
11. JetBlue (B6)	2,458	22

Airlines, their Hubs, and Hub-Spoke Ratios

Airline (Code)	1st largest hub	Hub-Spoke Ratio (%) One Hub	2nd largest hub	Hub-Spoke Ratio (%) Two Hubs
Southwest	Las Vegas (35)	9.3	Phoenix (33)	18.2
American	Dallas (52)	22.3	Chicago (46)	42.0
United	Chicago (50)	25.1	Denver (41)	45.7
Delta	Atlanta (53)	26.7	Cincinnati (42)	48.0
Continental	Houston (52)	36.6	New York (45)	68.3
Northwest	Minneapolis (47)	25.6	Detroit (43)	49.2
US Airways	Charlotte (35)	23.3	Philadelphia (33)	45.3
America West	Phoenix (40)	35.4	Las Vegas (28)	60.2
Alaska	Seattle (18)	56.2	Portland (10)	87.5
ATA	Chicago (16)	48.4	Indianapolis (6)	66.6
JetBlue	New York (13)	59.0	Long Beach (4)	77.3

Cumulative Hub-and-Spoke Ratios

Distribution of Markets by Number of Incumbents	
Markets with 0 airlines	35.44%
Markets with 1 airline	29.06%
Markets with 2 airlines	17.44%
Markets with 3 airlines	9.84%
Markets with 4 or more airlines	8.22%

Number of Monopoly Markets by Airline	
Southwest	157
Northwest	69
Delta	56
American	28
Continental	24
United	17

Entry and Exit	
	All Quarters
Distribution of Markets by Number of New Entrants	
Markets with 0 Entrants	84.66%
Markets with 1 Entrant	13.37%
Markets with 2 Entrants	1.69%
Markets with 3 Entrants	0.27%
Distribution of Markets by Number of Exits	
Markets with 0 Exits	86.51%
Markets with 1 Exit	11.82%
Markets with 2 Exits	1.35%
Markets with more 3 or 4 Exits	0.32%

7.4. Specification and Estimation of Demand. Demand. Let $d \in \{0, 1\}$ be the indicator of "direct" or non-stop flight. Let q_{irdt} be the number of tickets sold by airline i for route r, type of flight d, at quarter t. For a given route r and quarter t, the quantities $\{q_{irdt} :$ for every airline i and $d = 0, 1\}$ come from a system of demand of differentiated product. More specifically, we consider Nested Logit of demand. For notational simplicity, I omit here the subindexes (r, t), but the demand system refers to a specific route and quarter.

Let H be the number of travelers in the route. Each traveler in the route demands only one trip (per quarter) and chooses which product to purchase. The indirect utility of a traveler who purchases product (i,d) is $U_{id} = b_{id} - p_{id} + v_{id}$, where p_{id} is the price of product (i,d), b_{id} is the "quality" or willingness to pay for the product of the average consumer in the market, and v_{id} is a consumer-specific component that captures consumer heterogeneity in preferences. Product quality b_{ird} depends on exogenous characteristics of the airline and the route, and on the endogenous "hub-size" of the airline in the origin and destination airports.

$$b_{id} = \alpha_1 \; d + \alpha_2 \; HUB_i + \alpha_3 \; DIST + \xi_i^{(1)} + \xi^{(2)} + \xi_{id}^{(3)}$$

α_1 to α_3 are parameters. $DIST$ is the flown distance between the origin and destination cities of the route. $\xi_i^{(1)}$ is an airline fixed-effect that captures between-airlines differences in quality which are constant over time and across markets. $\xi^{(2)}$ represents the interaction of (origin and destination) city dummies and time dummies. These terms account for demand shocks, such as seasonal effects, which vary across cities and over time. $\xi_{id}^{(3)}$ is a demand shock that is airline and route specific. The variable HUB_i represents the "hub size" airline i in the origin and destination airports of the route r.

In the Nested Logit, we have that $v_{id} = \sigma_1 \; v_i^{(1)} + \sigma_2 \; v_{id}^{(2)}$, where $v_i^{(1)}$ and $v_{id}^{(2)}$ are independent Type I extreme value random variables, and σ_1 and σ_2 are parameters that measure the dispersion of these variables, with $\sigma_1 \geq \sigma_2$. A property of the nested logit model is that the demand system can be represented using the following closed-form demand equations:

$$\ln(s_{id}) - \ln(s_0) = \frac{b_{id} - p_{id}}{\sigma_1} + \left(1 - \frac{\sigma_2}{\sigma_1}\right) \ln(s_{id}^*) \tag{7.1}$$

where s_0 is the share of the outside alternative in route r, i.e., $s_{0r} \equiv 1 - \sum_{i=1}^N (s_{ir0} + s_{ir1})$, and s_{id}^* is the market share of product (i,d) within the products of airline i in this route, i.e., $s_{id}^* \equiv s_{id}/(s_{i0} + s_{i1})$.

Therefore, we have the following demand regression equation:

$$\ln(s_{irdt}) - \ln(s_{0rdt}) = W_{irdt} \; \alpha + \left(\frac{-1}{\sigma_1}\right) p_{irdt} + \left(1 - \frac{\sigma_2}{\sigma_1}\right) \ln(s_{irdt}^*) + \xi_{irdt}^{(3)} \tag{7.2}$$

The regressors in vector W_{irdt} are: dummy for nonstop-flight, hub-size, distance, airline dummies, origin-city dummies × time dummies, and destination-city dummies × time dummies.

Issues: Is HUB_{irt} correlated with $\xi_{irdt}^{(3)}$? Are the BLP instruments (HUB size of competing airlines in route r at period t) valid in this equation, i.e., are they correlated with $\xi_{irdt}^{(3)}$?

ASSUMPTION D1: *Idiosyncratic demand shocks* $\{\xi_{irdt}^{(3)}\}$ *are not serially correlated over time.*

ASSUMPTION D2: The idiosyncratic demand shock $\{\xi_{irdt}^{(3)}\}$ is private information of the corresponding airline. Furthermore, the demand shocks of two different airlines at two different routes are independently distributed.

Under assumption *D1*, the hub-size variable is not correlated with $\xi_{irdt}^{(3)}$ because HUB_{irt} is predetermined. Under assumption D2, HUB sizes of competing airlines in route r at period t are not correlated with $\xi_{irdt}^{(3)}$ and they are valid instruments for price p_{irdt}. Note that both assumptions D1 and D2 are testable. We can use the residuals of $\xi_{irdt}^{(3)}$ to test for no serial correlation (assumption D1) and no spatial correlation (assumption D2) in the residuals.

Table 7 presents estimates of the demand system.

Table 7
Demand Estimation[1]
Data: 85,497 observations. 2004-Q1 to 2004-Q4

	OLS		IV	
FARE (in \$100) $\left(-\frac{1}{\sigma_1}\right)$	-0.329	(0.085)	-1.366	(0.110)
ln(s*) $\left(1 - \frac{\sigma_2}{\sigma_1}\right)$	0.488	(0.093)	0.634	(0.115)
NON-STOP DUMMY	1.217	(0.058)	2.080	(0.084)
HUBSIZE-ORIGIN (in million people)	0.032	(0.005)	0.027	(0.006)
HUBSIZE-DESTINATION (in million people)	0.041	(0.005)	0.036	(0.006)
DISTANCE	0.098	(0.011)	0.228	(0.017)
σ_1 **(in \$100)**	3.039	(0.785)	0.732	(0.059)
σ_2 **(in \$100)**	1.557	(0.460)	0.268	(0.034)
Test of Residuals Serial Correlation				
m1$\sim N(0,1)$ **(p-value)**	0.303	(0.762)	0.510	(0.610)

(1) All the estimations include airline dummies, origin-airport dummies $\times$ time dummies,
and destination-airport dummies $\times$ time dummies. Stadard errors in parentheses.

The most important result is that the effect of hub-size on demand is statistically significant but very small: on average consumers are willing to pay approx. \$2 for an additional connection of the airline at the origin or destination airports ($\$2 \simeq \$100 * (0.027/1.366)$).

7.5. Specification and Estimation of Marginal Cost. Static Bertrand competition between airlines active in a route imply:

$$p_{irdt} - c_{irdt} = \frac{\sigma_1}{1 - \bar{s}_{irt}}$$

where $\bar{s}_{irt} = (e_{ir0t} + e_{ir1t})^{\sigma_2/\sigma_1}[1 + \sum_{j=1}^{N}(e_{jr0t} + e_{jr1t})^{\sigma_2/\sigma_1}]^{-1}$, $e_{irdt} \equiv \exp\{(b_{irdt} - p_{irdt})/\sigma_2\}$. Then, given the estimated demand parameters we can obtain estimates of the marginal costs c_{irdt}.

We are interested in estimation the effect of "hub-size" on marginal costs. We estimated the following model for marginal costs:

$$c_{irdt} = W_{irdt}\ \delta + \omega_{irdt}$$

where the regressors in vector W_{irdt} are: dummy for nonstop-flight, hub-size, distance, airline dummies, origin-city dummies $\times$ time dummies, and destination-city dummies $\times$ time dummies.

Again, under the assumption the error term ω_{irdt} is not serially correlated, hub-size is an exogenous regressor and we can estimate the equation for marginal costs using OLS.

Table 8
Marginal Cost Estimation[1]
Data: 85,497 observations. 2004-Q1 to 2004-Q4
Dep. Variable: Marginal Cost in $100

	Estimate (Std. Error)
NON-STOP DUMMY	0.006 (0.010)
HUBSIZE-ORIGIN (in million people)	-0.023 (0.009)
HUBSIZE-DESTINATION (in million people)	-0.016 (0.009)
DISTANCE	5.355 (0.015)
Test of Residuals Serial Correlation	
m1$\sim N(0,1)$ (p-value)	0.761 (0.446)

(1) All the estimations include airline dummies, origin-airport dummies $\times$ time dummies, and destination-airport dummies $\times$ time dummies.

Again, the most important result from this estimation is that the effect of hub-size on marginal cost is statistically significant but very small: on average an additional connection of the airline at the origin or destination airports implies a reduction in marginal cost between $1.6 and $2.3.

7.6. Simplifying assumptions for solution and estimation of dynamic game of network competition. The next step is the estimation of the effects if hub-size on fixed operating costs and sunk entry-costs. We consider the following structure in these costs.

$$FC_{imt} \;=\; \gamma_1^{FC} + \gamma_2^{FC}\,HUB_{imt} + \gamma_3^{FC}DIST_m + \gamma_{4i}^{FC} + \gamma_{5c}^{FC} + \varepsilon_{imt}^{FC}$$

$$EC_{imt} \;=\; \eta_1^{EC} + \eta_2^{EC}\,HUB_{imt} + \eta_3^{EC}DIST_m + \eta_{4i}^{EC} + \eta_{5c}^{EC}$$

where γ_{4i}^{FC} and η_{4i}^{EC} are airline fixed effects, and γ_{5c}^{FC} and η_{5c}^{EC} are city (origin and destination) fixed effects. ε_{imt}^{FC} is a private information shock. The parameters in these functions are estimated using data on airlines entry-exit decisions and the dynamic game.

However, this dynamic game has really a large dimension. Given the number of cities and airlines in our empirical analysis, the number of possible industry networks is $|X| = 2^{NM} \simeq 10^{10,000}$ (much larger than all the estimates of the number of atoms in the observable universe, around 10^{100}). We should make simplifying assumptions.

We consider **two types of simplifying assumptions** that reduce the dimension of the dynamic game and make its solution and estimation manageable.

1. An **airline's choice of network is decentralized** in terms of the separate decisions of local managers.

2. The state variables of the model can be aggregated in a vector of **inclusive-values** that belongs to a space with a much smaller dimension than the original state space.

(1) **Decentralizing the Airline's Choice of Network.** Each airline has M local managers, one for each city-pair. A local manager decides whether to operate or not non-stop flights in his local-market: i.e., he chooses a_{imt}. The private information shock ε_{imt}^{FC} is private information of the manager (i, m).

IMPORTANT: A local manager is not only concern about profits in her own route. She internalizes the effects of her own entry-exit decision in many other routes. This is very important to allow for entry deterrence effects of hub-and-spoke networks.

ASSUMPTION: Let R_{imt} be the sum of airline i's variable profits over all the routes that include city-pair m as a segment. *Local managers maximize the expected and discounted value of*

$$\Pi_{imt} \equiv R_{imt} - a_{imt}\left(FC_{imt} + (1 - x_{imt})EC_{imt}\right).$$

(2) **Inclusive-Values.** Decentralization of the decision simplifies the computation of players' best responses, but the state space of the decision problem of a local manager is still huge. Notice that the profit of a local manager depends only on the state variables:

$$\mathbf{x}_{imt}^* \equiv (x_{imt},\; R_{imt},\; HUB_{imt})$$

ASSUMPTION: The *vector* $\mathbf{x}^*_{imt}$ *follows a controlled first-order Markov Process:*

$$\Pr\left(\mathbf{x}^*_{im,t+1} \mid \mathbf{x}^*_{imt}, a_{imt}, \mathbf{x}_t, \mathbf{z}_t\right) = \Pr\left(\mathbf{x}^*_{im,t+1} \mid \mathbf{x}^*_{imt}, a_{imt}\right)$$

A MPE of this game can be describe as a vector of probability functions, one for each local-manager:

$$P_{im}(\mathbf{x}^*_{imt}) : i = 1, 2, ..., N; \ m = 1, 2, ..., M$$

$P_{im}(\mathbf{x}^*_{imt})$ is the probability that local-manager (i, m) decides to be active in city-pair m given the state $\mathbf{x}^*_{imt}$. An equilibrium exits. The model typically has multiple equilibria.

7.7. Estimation of dynamic game of network competition. We use the Nested Pseudo Likelihood (NPL) method.

Table 9
Estimation of Dynamic Game of Entry-Exit[1]

Data: 1,485 markets $\times$ 22 airlines $\times$ 3 quarters = 98,010 observations

	Estimate (Std. Error) (in thousand $)
Fixed Costs (quarterly):[2]	
$\gamma_1^{FC} + \gamma_2^{FC}$**mean hub-size** $+\gamma_3^{FC}$**mean distance** (average fixed cost)	119.15 (5.233)
γ_2^{FC} **(hub-size in # cities connected)**	-1.02 (0.185)
γ_3^{FC} **(distance, in thousand miles)**	4.04 (0.317)
Entry Costs:	
$\eta_1^{EC} + \eta_2^{EC}$**mean hub-size** $+\eta_2^{EC}$**mean distance** (average entry cost)	249.56 (6.504)
η_2^{EC} **(hub-size in # cities connected)**	-9.26 (0.140)
η_3^{EC} **(distance, in thousand miles)**	0.08 (0.068)
σ_ε	8.402 (1.385)
β	0.99 (not estimated)
Pseudo R-square	0.231

(1) All the estimations include airline dummies, and city dummies.

(2) Mean hub size = 25.7 million people. Mean distance (nonstop flights) = 1996 miles

• Goodness of fit:

Table 10
Comparison of Predicted and Actual Statistics of Market Structure
1,485 city-pairs (markets). Period 2004-Q1 to 2004-Q4

		Actual (Average All Quarters)	Predicted (Average All Quarters)
	Herfindahl Index (median)	5338	4955
Distribution of Markets	Markets with 0 airlines	35.4%	29.3%
by Number of Incumbents	" " 1 airline	29.1%	32.2%
	" " 2 airlines	17.4%	24.2%
	" " 3 airlines	9.8%	8.0%
	" " $\geq$4 airlines	8.2%	6.2%
Number (%) of Monopoly	Southwest	151 (43.4%)	149 (38.8%)
Markets for top 6 Airlines	Northwest	66 (18.9%)	81 (21.1%)
	Delta	57 (16.4%)	75 (19.5%)
	American	31 (8.9%)	28 (7.3%)
	Continental	27 (7.7%)	27 (7.0%)
	United	16 (4.6%)	24 (6.2%)
Distribution of Markets	Markets with 0 Entrants	84.7%	81.9%
by Number of New Entrants	" " 1 Entrant	13.4%	16.3%
	" " 2 Entrants	1.7%	1.6%
	" " $\geq$3 Entrants	0.3%	0.0%
Distribution of Markets	Markets with 0 Exits	86.5%	82.9%
by Number of Exits	" " 1 Exit	11.8%	14.6%
	" " 2 Exits	1.4%	1.4%
	" " $\geq$3 Exits	0.3%	0.0%

7.8. Counterfactual Experiments. To deal with multiple equilibria or equilibrium selection in the counterfactual experiment, we use the homotopy method that I described in lecture #11.

Table 11
Counterfactual Experiments
Hub-and-Spoke Ratios when Some Structural Parameters Become Zero

Carrier	Observed	Method 1: Taylor Approximation			
		Experiment 1 No hub-size effects in variable profits	Experiment 2 No hub-size effects in fixed costs	Experiment 3 No hub-size effects in entry costs	Experiment 4 No complementarity across markets
Southwest	18.2	17.3	15.6	8.9	16.0
American	42.0	39.1	36.5	17.6	29.8
United	45.7	42.5	39.3	17.8	32.0
Delta	48.0	43.7	34.0	18.7	25.0
Continental	68.3	62.1	58.0	27.3	43.0
Northwest	49.2	44.3	36.9	18.7	26.6
US Airways	45.3	41.7	39.0	18.1	34.4

Carrier	Observed	Method II: Policy Iterations Starting from Taylor Approx.			
		Experiment 1 No hub-size effects in variable profits	Experiment 2 No hub-size effects in fixed costs	Experiment 3 No hub-size effects in entry costs	Experiment 4 No complementarity across markets
Southwest	18.2	16.9	14.4	8.3	16.5
American	42.0	37.6	34.2	16.6	24.5
United	45.7	40.5	37.3	15.7	30.3
Delta	48.0	41.1	32.4	17.9	22.1
Continental	68.3	60.2	57.4	26.0	42.8
Northwest	49.2	40.8	35.0	17.2	23.2
US Airways	45.3	39.7	37.1	16.4	35.2

Experiment 1: Counterfactual model: $\alpha_2 = \alpha_3 = \delta_2 = \delta_3 = 0$

Experiment 2: Counterfactual model: $\gamma_2^{FC} = 0$

Experiment 3: Counterfactual model: $\eta_2^{EC} = 0$

Experiment 4: Counterfactual model: Variable profit of local manager in city-pair AB includes only variable profits from non-stop routes AB and BA.

Main results:

-Hub-size effects on **demand, variable costs and fixed operating costs** are significant but can **explain very little of the propensity to hub-spoke networks**.

- **Hub-size effects on Sunk Entry Costs are large**. This is the most important factor to explain hub-spoke networks.

- **Strategic factors: hub-spoke network as a strategy to deter entry** is the second most important factor for some of the largest carriers (Northwest and Delta).

8. Dynamic strategic behavior in firms' innovation

8.1. Competition and Innovation: static analysis. Competition and Innovation. Long lasting debate on the effect of competition on innovation (e.g., Schumpeter, Arrow). Apparently, there are contradictory results between a good number of theory papers showing that "competition" has a negative effect on innovation (Dasgupta & Stiglitz, 1980: Spence, 1984), and a good number of reduced-form empirical papers showing a positive relationship between measures of competition and measures of innovation (Porter, 1990; Geroski, 1990; Blundell, Griffith and Van Reenen 1999). Vives (JIND, 2008) presents a systematic theoretical analysis of this problem that tries to explain the apparent disparaty between existing theoretical and empirical results.

Competition and Innovation: Vives (2008) considers:

[1] Different sources of exogenous increase in competition. (i) reduction in entry cost; (ii) increase in market size; (iii) increase in degree of product substitutability.

[2] Different types of innovation. (i) process or cost-reduction innovation; (ii) product innovation / new products.

[3] Different models of competition and specifications. (i) Bertrand; (ii) Cournot

[4] Specification of demand. linear, CES, expontetial, logit, nested logit.

Vives shows that [1] the form of increase in competition; and [2] the type of innovation are key to detemine a positive or a negative relatioship betwween competition and innovation. However, the results are very robust: [3] the form of competition (Bertrand or Cournot) and [4] the specification of the demand system.

Model. Static model with symmetric firms, endogenous entry. Profit of firm i: $\pi_j = [p_j - c(z_j)] \; s \; d(p_j, p_{-j}, n; \alpha) - z_j - F$, s = market size; n = number of firm; $d(p_j, p_{-j}, n; \alpha)$ = demand per-consumer; α = degree of substitutability; $c(z_j)$ = marginal cost (constant); z_i = expenditure in cost reduction; $c' < 0$ and $c'' > 0$; F = entry cost.

Equilibrium. Nash equilibrium for simultaneous choice of (p_j, z_j). Symmetric equilibrium. There is endogenous entry. Marginal condition w.r.t cos-reduction R&D (z) is: $-c'(z) \; s \; d(p, n; \alpha) - 1 = 0$. Since $c'' > 0$, this implies $z = g(s \; d(p, n; \alpha))$, where $g(.)$ is an increasing

function. The incentive to invest in cost reduction increases with output per firm, $q \equiv s\, d(p,n;\alpha)$.

Any exogenous change in competition (say in α, s, or F) has three effects on output per firm and therefore on investment in cost-reduction R&D.

$$\frac{dz}{d\alpha} = g'(q) \left[\frac{\partial\left[s\, d(p,n;\alpha)\right]}{\partial\alpha} + \frac{\partial\left[s\, d(p,n;\alpha)\right]}{\partial p}\frac{\partial p}{\partial\alpha} + \frac{\partial\left[s\, d(p,n;\alpha)\right]}{\partial n}\frac{\partial n}{\partial\alpha} \right]$$

$\dfrac{\partial\left[s\, d(p,n;\alpha)\right]}{\partial\alpha}$ is the **direct demand effect,** $\dfrac{\partial\left[s\, d(p,n;\alpha)\right]}{\partial p}\dfrac{\partial p}{\partial\alpha}$ is the **price pressure effect.** $\dfrac{\partial\left[s\, d(p,n;\alpha)\right]}{\partial n}\dfrac{\partial n}{\partial\alpha}$ is the **number of entrants effect.** The effects of different changes in competition on cost-reduction R&D can be explained in terms of these three effects.

Summary of comparative statics. **(i) Increase in market size.** - Increases per-firm expenditures in cost-reduction; - Effect on product innovation (# varieties) can be either positive or negative. **(ii) Reduction in cost of market entry.** - Reduces per-firm expenditures in cost-reduction; - Increases number of firms and varieties. **(iii) Increase in degree of product substitution.** - Increases per-firm expenditures in cost-reduction; - # varieties may increase or decline.

Some limitations in this analysis. The previous analysis is static, without uncertainty, with symmetric and single product firms. Therefore, the following factors that relate competition and innovation are absent from the analysis. (1) **Preemptive motives.** (2) **Cannibalization of own products**. (3) **Increasing uncertainty** in returns to R&D due competition (asymmetric info). To study these factors, we need dynamic games with uncertainty, and asymmetric multi-product firms.

8.2. Creative destruction: incentives to innovate of incumbents and new entrants. Innovation and creative destruction (Igami, 2017). Innovation, the creation of new products and technologies, necessarily implies the "destruction" of existing products, technologies, and firms. In other words, the survival of existing products / technologies / firms is at the cost of preempting the birth of new ones. The speed (and the effectiveness) of the innovation process in an industry depends crucially on the dynamic strategic interactions between "old" and "new" products/technologies. Igami (JPE, 2017) studies these interactions in the context of the Hard-Disk-Drive (HDD) industry during 1981-1998.
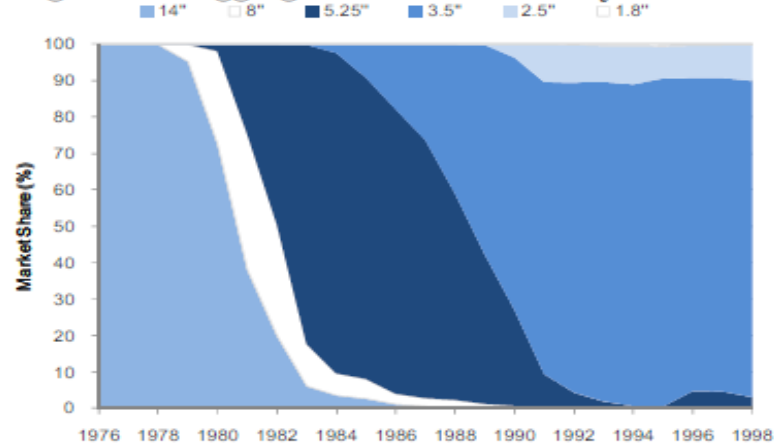
HDD: Different generations of products

Figure 2: Shifting Generations of Technology
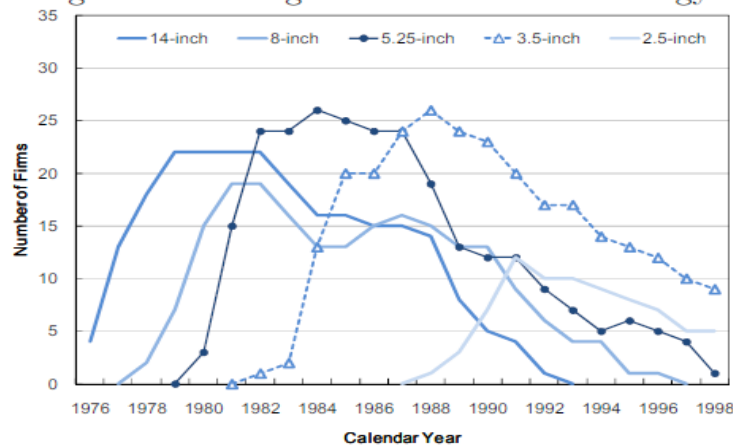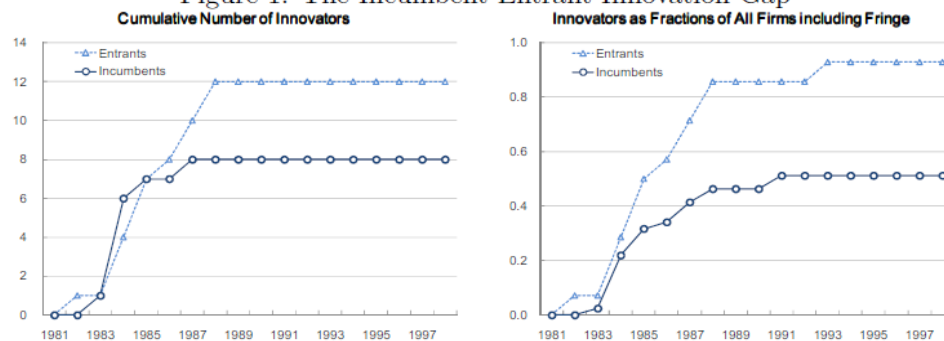
HDD: Different generations of products

Figure 12: Aggregate Market Share by Diameter

Adoption new tech: Incumbents vs. New Entrants


Figure 1: The Incumbent-Entrant Innovation Gap

Adoption new tech: Incumbents vs. New Entrants. Igami focuses on the transition from 5.25 to 3.5 inch products. He consider three main factors that contribute to the relative propensity to innovate of incumbents and potential entrants. **Cannibalization.** For incumbents, the introduction of a new product reduces the demand for their pre-existing

products. **Preemption.** Early adoption by incumbents can deter entry and competition from potential new entrants. **Differences in entry/innovation costs.** It can play either way. Incumbents have knowledge capital and **economies of scope**, but they also have **organizational inertia**.
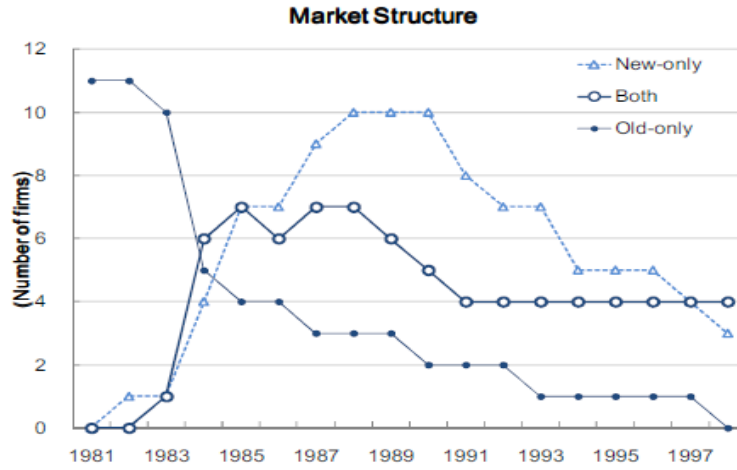
Data. Market shares New/Old products
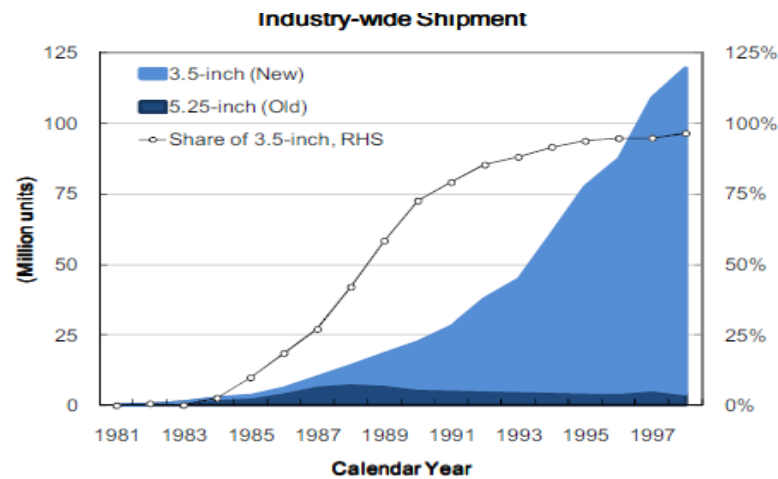
Average Prices: New/Old products
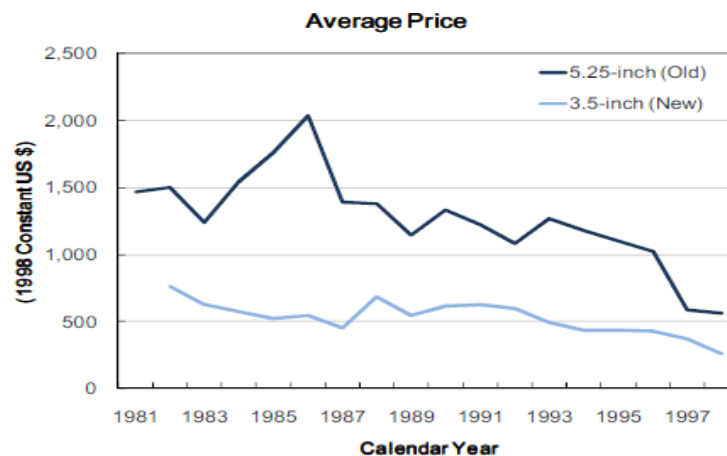
Average Quality: New/Old products

Average Quality (Information Storage Capacity)
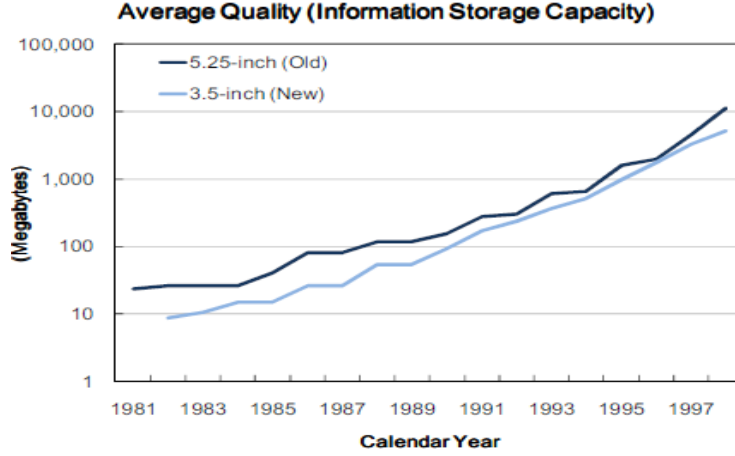
Market Structure: New/Old products

Market Structure

Model. Market structure at period t is described by four type of firms according to the products they produce: $s_t = \{N_t^{old}, N_t^{both}, N_t^{new}, N_t^{pe}\}$

● Initialy, $N_0^{both} = N_0^{new} = 0$. Timing within a period t:

1 Incumbents compete (a la Cournot) $\rightarrow$ Period profits $\pi_t(s_{it}, s_{-it})$

2. The N_t^{old} firms draw private info shocks and simultaneously choose $a_{it}^{old} \in \{exit, stay, innovate\}$

3. The N_t^{both} observe a_t^{old}, draw private info shocks, and simultaneously choose $a_{it}^{both} \in \{exit, stay\}$

4. The N_t^{new} observe a_t^{old}, a_t^{both}, draw private info shocks, and simultaneously choose $a_{it}^{new} \in \{exit, stay\}$

5. The N_t^{pe} observe a_t^{old}, a_t^{both}, a_t^{new}, draw private info shocks, and simultaneously choose $a_{it}^{pe} \in \{entry, noentry\}$.

Given these choices, next period market structure is obtained, s_{t+1}, and demand and cost variables evolve exogenously. Why imposing this order of move? This Assumption,

together with: - Finite horizon T; Homogeneous firms (up to the i.i.d. private info shocks) withing each type, implies that there is a **unique Markov Perfect equilibrium**. This is very convenient for estimation (Igami uses a standard/Rust Nested Fixed Point Algorithm for estimation) and especially for counterfactuals.

Demand. Simple logit model of demand. A product is defined as a pair {technology, quality}, where technology $\in \{old, new\}$ and *quality* represents different storage sizes. There is no differentiation across firms (perhaps true, but assumption comes from data limitations).

Estimation:

$$\ln \left(\frac{s_j}{s_k} \right) = \alpha_1 \left[p_j - p_k \right] + \alpha_2 \left[1_j^{new} - 1_k^{new} \right] + \alpha_3 \left[x_j - x_k \right] + \xi_j - \xi_k$$

Data: multiple periods and regions. IVs: Hausman-Nevo. Prices in other regions.

Estimates of Demand

Market definition:	Broad		Narrow	
Estimation method:	OLS	IV	OLS	IV
	(1)	(2)	(3)	(4)
Price ($000)	-1.66^{***}	-2.99^{***}	$-.93^{**}$	-3.28^{***}
	(.45)	(.55)	(.46)	(.63)
Diameter = 3.5-inch	$.84^{*}$	.75	1.75^{***}	$.91^{**}$
	(.46)	(.45)	(.31)	(.38)
Log Capacity (MB)	.18	$.87^{***}$	.04	1.20^{***}
	(.33)	(.27)	(.26)	(.31)
Year dummies	Yes	Yes	Yes	Yes
Region/user dummies	–	–	Yes	Yes
Adjusted R^2	.43	.33	.50	.28
Number of obs.	176	176	405	405
Partial R^2 for Price	–	.32	–	.16
P-value	–	.00	–	.00

Evolution of unobserved Quality (epsi)

Estimated Unobserved Quality (ξ)
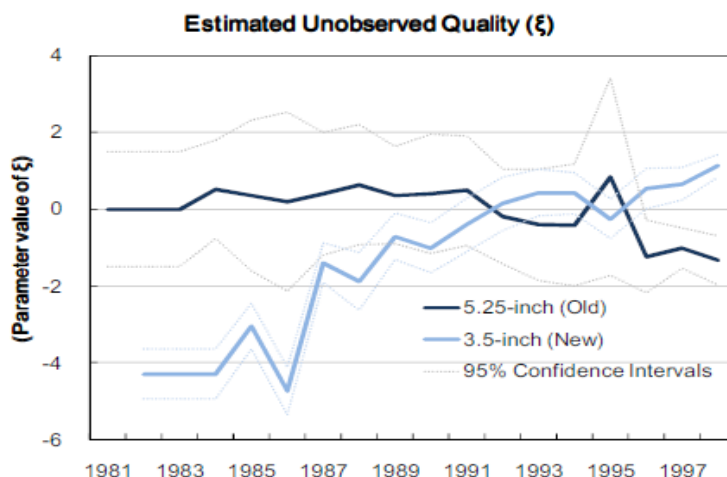
Evolution of Marginal Costs

Estimated Marginal Cost
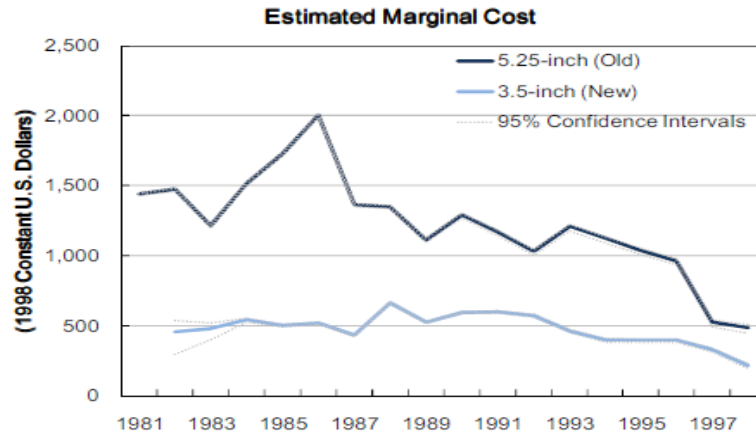
Evolution of Period Profits [keeping market structure]
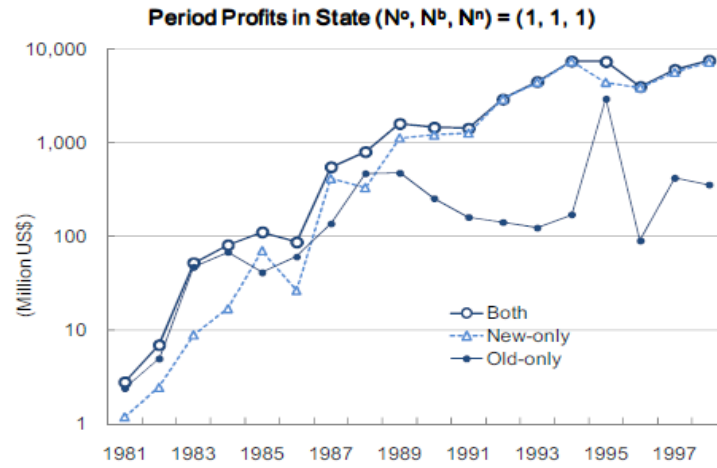
Estimates of Dynamic Parameters

Table 4: Estimates of the Dynamic Parameters

| ($ Billion)	Maximum Likelihood Estimates		
	(1)	(2)	(3)
Assumed order of moves:	Old-Both-New-PE	PE-New-Both-Old	PE-Old-Both-New
Fixed cost of operation (ϕ)	0.1474	0.1472	0.1451
	$\begin{bmatrix} -0.02, & 0.33 \end{bmatrix}$	$\begin{bmatrix} -0.02, & 0.33 \end{bmatrix}$	$\begin{bmatrix} -0.03, & 0.33 \end{bmatrix}$
Incumbents' sunk cost $\left(\kappa^{inc}\right)$	1.2439	1.2370	1.2483
	$\begin{bmatrix} 0.51, & 2.11 \end{bmatrix}$	$\begin{bmatrix} 0.50, & 2.10 \end{bmatrix}$	$\begin{bmatrix} 0.51, & 2.11 \end{bmatrix}$
Entrants' sunk cost $\left(\kappa^{ent}\right)$	2.2538	2.2724	2.2911
	$\begin{bmatrix} 1.74, & 2.85 \end{bmatrix}$	$\begin{bmatrix} 1.76, & 2.87 \end{bmatrix}$	$\begin{bmatrix} 1.78, & 2.89 \end{bmatrix}$
Log likelihood	-112.80	-112.97	-113.46

Estimates of Dynamic Parameters

Different estimates depending on the order of move within a period. Cost for innovation is smaller for incumbents than for new entrants ($\kappa^{inc} < \kappa^{pe}$). Organizational inertia does not seem an important factor. Magnitude of entry costs are comparable to the annual R&D budget of specialized HDD manufacturers, e.g., Seagate Tech: between $\$0.6B - \$1.6B$.
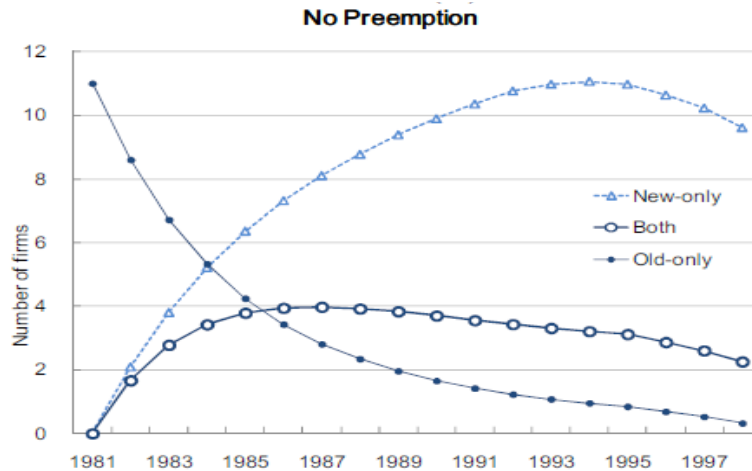
Estimated Model: Goodness of fit

Figure 5: Fit of Market Structure Dynamics
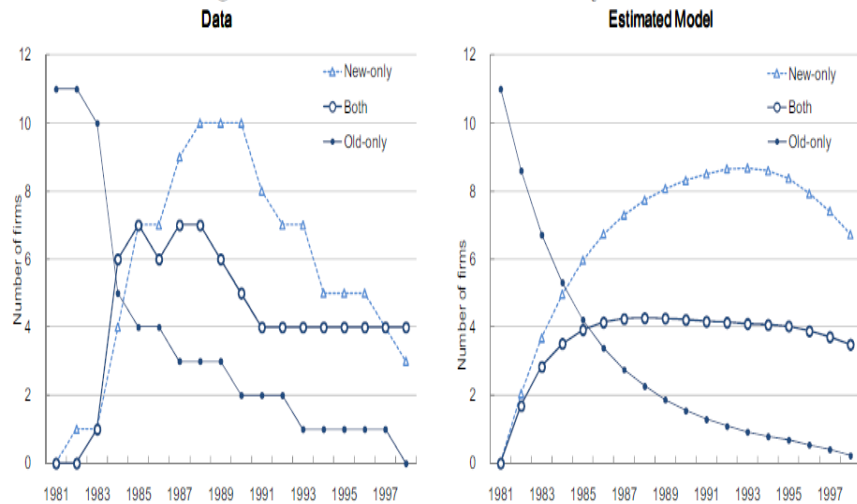
Counterfactual: Removing Cannibalization
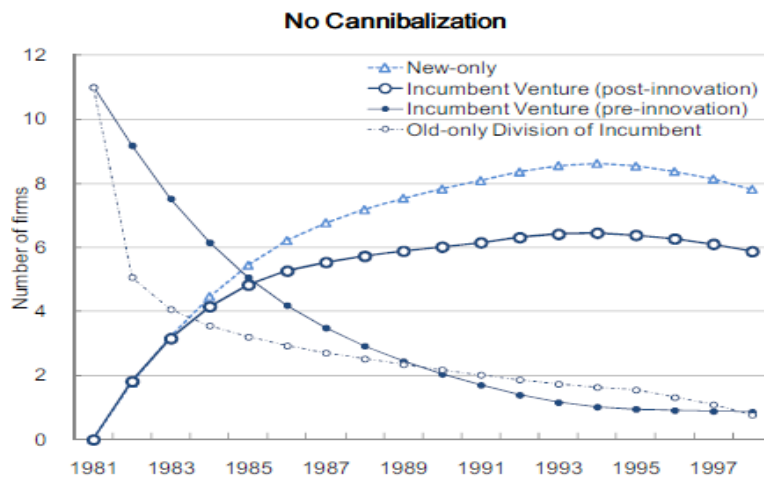
Counterfactual: Removing Preemption

No Preemption

8.3. Competition and innovation in the CPU industry: Intel and AMD. Study competition between Intel and AMD in the PC microprocessor industry. Incorporates durability of the product as a potentially important factor. Two forces drive innovation: - competition between firms for the technological frontier; - since PCs have little physical depreciacion, firms have the incentive to innovate to generate a tenological depreciation of consumers' installed PCs that encourages them to upgrade [most of the demand during the period >89% was upgrading]. Duopolists face both forces, whereas a monopolist faces only the latter (but in a stronger way).

The PC microprocessor industry. Very important to the economy: - Computer equipment manufacturing industry generated 25% of U.S. productivity growth from 1960 to 2007. - Innovations in microprocessors are directly measured via improved performance on benchmark tasks. Most important: CPU speed. Interesting also from the point of view of antitrust: - In 2004: several antitrust lawsuits claiming Intel's anticompetitive practices, e.g., rewarding PC manufacturers that exclusively use Intel microprocessors. - Intel foreclosures AMD to access some consumers.- Intel settled these claims in 2009 with a $1.25 billion payment to AMD.

Market is essentially a duopoly, with AMD and Intel selling 95% CPUs. Firmst have high R&D intensities, R&D/Revenue (1993-2004): - AMD 20% ; and Intel 11%. Innovation is rapid: new products are released nearly every quarter. CPU performance (speed) doubles every 7 quarters, i.e., Moore'e law. AMD and Intel extensively cross-license each other's technologies, i.e., positive spillovers.

As microprocessors are durable, replacement drives and important part of demand. The importance of replacement is partly exogenous (new consumers arriving to the marker), and partly endogenous: speed of improvements in frontier microprocessors that encourages consumers to upgrade. In 2004, 82% of PC purchases were replacements. After an upgrade

(a) Frontier CPU Log–Quality
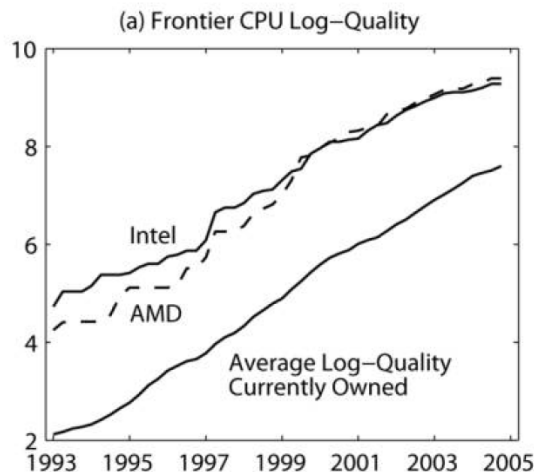
boom, prices and sales fall as replacement demand drops. Firms must continue to innovate to rebuild replacement demand.
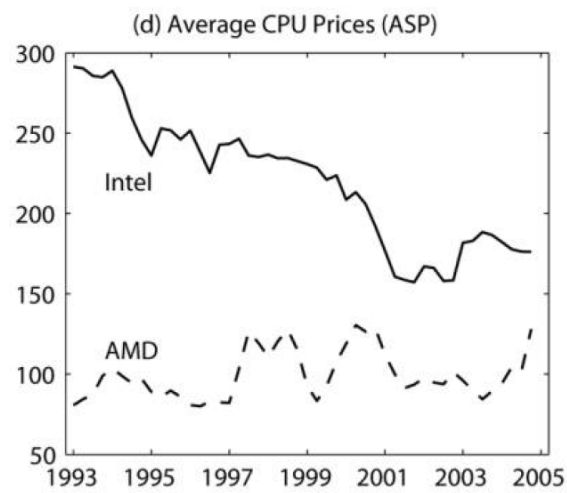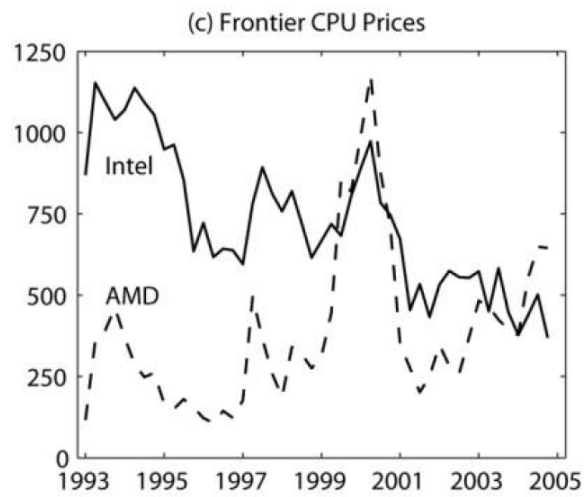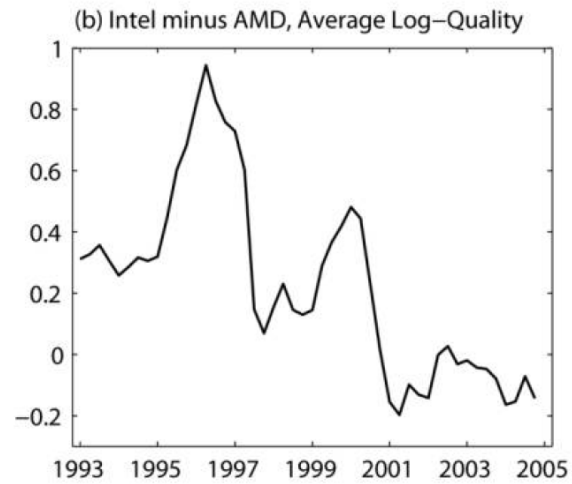
Data. Proprietary data from a market research firm specializing in the microprocessor industry. Quarterly data from Q1-1993 to Q4-2004 (48 quarters). Information on: shipments in physical units for each type of CPU; manufacturers' average selling prices (ASP); **production costs**; CPU characteristics (speed). All prices and costs are converted to base year 2000 dollars. Quarterly R&D investment levels, obtained from firms' annual reports.
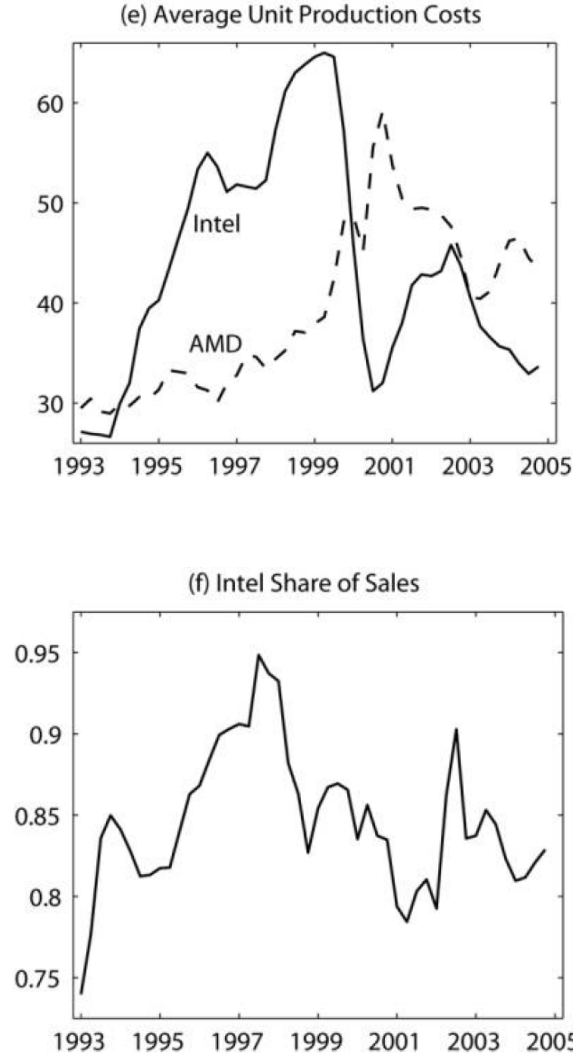
Moore's Law. Intel cofounder Gordon Moore predicted in 1965 that the number of transistors ina CPU (and therefore the CPU speed) would double every 2 years. Following figure shows "Moore's law" over the 48 quarters in the data. Quality is measured using processor speed. Quarterly % change in CPU speed is 10.2% for Intel and 11% for AMD.

Moore's Law (Frontier CPU speed)

Differential log-quality between Intel and AMD. Intel's initial quality advantage is moderate in 1993–94. Then, it becomes large in 1995-96 when Intel releases the Pentium. AMD's responded in 1997 introduccing the K6 processor that narrows the gap. But parity is not achieved until the mid-2000 when AMD released the Athlon.

Model: General features. Dynamic model of an oligopoly with differentiated and durable products. Each firm j sells a single product and invests in R&D to improve its quality. If investments are successful, quality improves next quarter by a fixed proportion δ; otherwise it is unchanged: log quality $q_{jt} \in \{0,\ \delta,\ 2\delta,\ 3\delta,\ \dots\}$. Consumers: a key feature of demand for durable goods is that the value of the no-purchase option is endogenous, determined by last purchase. The distribution of currently owned products by consumerts is represented by the vector Δ_t. Δ_t affects current consumer demand. [Details]

(b) Intel minus AMD, Average Log–Quality

(c) Frontier CPU Prices

(d) Average CPU Prices (ASP)

(e) Average Unit Production Costs

(f) Intel Share of Sales

Firms and consumers are forward looking. A consumer's i state space consists of $(q_{it}^*, q_t, \Delta_t)$: - q_{it}^* = the quality of her currently owned product q_t^*; - q_t = vector of firms' current qualities q_t;   - Δ_t = distribution of qualities of consumers currently owned products. Δ_t is part of the consumers' state space because it affects expectations on future prices. State space for firms is (q_t, Δ_t). Given these state variables firms simultaneously choose prices p_{jt} and investment x_{jt}.

Consumer Demand. Authors: *"We restrict firms to selling only one product because the computational burden of allowing multiproduct firms is prohibitive"*. Consumers own no more than one microprocessor at a time. Utility for a consumer i from firm j's new product with quality q_{jt} is given by: $u_{ijt} = \gamma\, q_{jt} - \alpha\, p_{jt} + \xi_j + \varepsilon_{ijt}$. Utility from the no-purchase option is: $u_{i0t} = \gamma\, q_{it}^* + \varepsilon_{i0t}$. A consumer maximizes her intertemporal utility given her beliefs about the evolution of future qualities and prices given (q_t, Δ_t).

Market shares for consuerms currently owning q^* are:

$$s_{jt}(q^*) = \frac{\exp\{v_j(q_t, \Delta_t, q^*)\}}{\sum_{k=0}^{J} \exp\{v_k(q_t, \Delta_t, q^*)\}}$$

Using Δ_t to integrate over the distribution of q^* yields the market share of product j.

$$s_{jt}(q^*) = \sum_{q^*} s_{jt}(q^*)\, \Delta_t(q^*)$$

Transition rule of Δ_t. By definition, next period Δ_{t+1} is determined by a known closed-form function of Δ_t, q_t, and s_t.

$$\Delta_{t+1} = F_\Delta(\Delta_t, q_t, s_t)$$

The period profit function is:

$$\pi_j(p_t, q_t, \Delta_t) = M\ s_j(p_t, q_t, \Delta_t)\ [p_{jt} - mc_j(q_{jt})]$$

The specification of the marginal cost is:

$$mc_j(q_{jt}) = \lambda_{0j} - \lambda_1(q_t^{\max} - q_{jt})$$

Marginal costs are smaller for non-frontier firms. Parameter λ_1 captures an spillover effect from the innovation of other firms.

Model: Firms. Innovation process. Relationship between investment in R&D (x_{jt}) and log-quality improvement ($\Delta q_{jt+1} = q_{jt+1} - q_{jt}$). Log-Quality improvement can take two values, 0 or δ. The probability that $\Delta q_{jt+1} = \delta$ is (Pakes & McGure, 1994):

$$\chi_j(x_{jt}, q_{jt}) = \frac{a_j(q_{jt})\ x_{jt}}{1 + a_j(q_{jt})\ x_{jt}}$$

$a_j(q_{jt})$ is the "investment efficiency" function. It is a decreasing function, to capture the idea that of an increasing ed difficulty of advancing the frontier relative to catching up.

Let $W_j(q_t, \Delta_t)$ be the value function. The Bellman equation is:

$$W_j(q_t, \Delta_t) = \max_{x_{jt}, p_{jt}} \left[\pi_j(p_t, q_t, \Delta_t) - x_{jt} + \beta\, \mathbb{E}_t\left[W_j(q_{t+1}, \Delta_{t+1})\right]\ \right]$$

The decision variables are continuous, and the best response function should satisfy the F.O.C.

$$\frac{\partial \pi_{jt}}{\partial p_{jt}} + \beta \frac{\partial \mathbb{E}_t\left[W_{j,t+1}\right]}{\partial p_{jt}} = 0$$

$$\frac{\partial \pi_{jt}}{\partial x_{jt}} - 1 + \beta \frac{\partial \mathbb{E}_t\left[W_{j,t+1}\right]}{\partial x_{jt}} = 0$$

Markov Perfect Equilibrium. (1) firms' and consumers' equilibrium strategies depend only on current payoff relevant state variables (q_t, Δ_t). (2) consumers have rational expectations about firms' policy functions. (3) each firm has rational expectations about competitors' policy functions and about the evolution of the ownership distribution.

TABLE 1
EMPIRICAL AND SIMULATED MOMENTS

Moment	Actual	Actual Standard Error	Fitted
Intel price equation:			
Average Intel price	219.7	5.9	206.2
$q_{Intel,t} - q_{AMD,t}$	47.4	17.6	27.3
$q_{Intel,t} - \bar{\Delta}_t$	94.4	31.6	43.0
AMD price equation:			
Average AMD price	100.4	2.3	122.9
$q_{Intel,t} - q_{AMD,t}$	−8.7	11.5	−22.3
$q_{AMD,t} - \bar{\Delta}_t$	16.6	15.4	5.9
Intel share equation:			
Constant	.834	.007	.846
$q_{Intel,t} - q_{AMD,t}$	.055	.013	.092
Potential upgrade gains:			
Mean $(\bar{q}_t - \hat{\Delta}_t)$	1.146	.056	1.100
Mean innovation rates:			
Intel	.557	.047	.597
AMD	.610	.079	.602
Relative qualities:			
Mean $q_{Intel,t} - q_{AMD,t}$	1.257	.239	1.352
Mean $\mathcal{I}(q_{Intel,t} \geq q_{AMD,t})$	.833	.054	.929
Mean R&D/revenue:			
Intel	.114	.004	.101
AMD	.203	.009	.223

Estimation. Marginal cost parameters (λ_0, λ_1) are estimated in a first step because the dataset includes data on marginal costs. The rest of the structural parameters, $\theta = (\gamma, \alpha, \xi_{intel}, \xi_{amd}, a_{0,intel}, a_{0,amd}, a_1)$. Demand: $\gamma, \alpha, \xi_{intel}, \xi_{amd}$; Investment innovation effiiency: $a_{0,intel}, a_{0,amd}, a_1$. θ is estimated using Indirect Inference or Simulated Method of Moments (SMM).

Moments to match: Mean of innovation rates $q_{j,t+1} - q_{jt}$ for each firm. Mean R&D intensities $x_{jt}/revenue_{jt}$ for each firm. Mean of differential quality $q_{intel,t} - q_{amd,t}$, and share of quarters with $q_{intel,t} \geq q_{amd,t}$. Mean of gap $q_t^{max} - \bar{\Delta}_t$. Average prices, and OLS estimated coefficients of the regressions of p_{jt} on $q_{intel,t}$, $q_{amd,t}$, and average $\bar{\Delta}_t$. OLS estimated coefficients of the regression of $s_{intel,t}$ on $q_{intel,t} - q_{amd,t}$.

Empirical and predicted moments

Parameter estimates

Demand: Dividing γ by α: consumers are willing to pay \$21 for enjoying during 1 quarter a $\delta = 20\%$ increase in log quality. Dividing $\xi_{intel} - \xi_{amd}$ by α: consumers are willing to pay \$194 for Intel over AMD. The model needs this strong brand effect to explain the fact that AMD's share never rises above 22 percent in the period during which AMD had a faster product. Intel and AMD's innovation efficiencies are estimated to be .0010 and .0019, respectively, as needed for AMD to occasionally be the technology leader while investing much less.

Counterfactuals

From current duopoly (1) to Intel Monopoly (3) Innovation rate increases from 0.599 to 0.624. Mean quality upgrade increases 261% to 410%. Investment in R&D: increases by $1.2B$ per quarter: more than doubles. Price increases in \$102 (70%). Consumer surplus declines

TABLE 2
PARAMETER ESTIMATES

Parameter	Estimate	Standard Error
Price, α	.0131	.0017
Quality, γ	.2764	.0298
Intel fixed effect, ξ_{Intel}	$-.6281$	.0231
AMD fixed effect, ξ_{AMD}	-3.1700	.0790
Intel innovation, $a_{0,\text{Intel}}$	.0010	.0002
AMD innovation, $a_{0,\text{AMD}}$	.0019	.0002
Spillover, a_1	3.9373	.1453
Stage 1 marginal cost equation:		
Constant, λ_0	44.5133	1.1113
$\max(0, q_{\text{competitor},t} - q_{\text{own},t})$, λ_1	-19.6669	4.1591

TABLE 3
INDUSTRY OUTCOMES UNDER VARIOUS SCENARIOS

	AMD-INTEL DUOPOLY (1)	SYMMETRIC DUOPOLY (2)	MONOPOLY (3)	NO SPILLOVER DUOPOLY (4)	MYOPIC PRICING AMD-Intel (5)	MYOPIC PRICING Monopoly (6)
Industry profits ($ billions)	408	400	567	382	318	322
Consumer surplus (CS)	2,978	3,012	2,857	3,068	2,800	2,762
CS as share of monopoly CS	1.042	1.054	1.000	1.074	.980	.967
Social surplus (SS)	3,386	3,412	3,424	3,450	3,118	3,084
SS as share of planner SS	.929	.906	.940	.916	.828	.819
Margins, $(p-mc)/mc$	3.434	2.424	5.672	3.478	2.176	2.216
Price	194.17	146.73	296.98	157.63	140.06	143.16
Frontier innovation rate	.599	.501	.624	.438	.447	.438
Industry investment ($ millions)	830	652	1,672	486	456	787
Mean quality upgrade (%)	261	148	410	187	175	181
Intel or leader share	.164	.135	.143	.160	.203	.211
AMD or laggard share	.024	.125		.091	.016	

in $121M$ (4.2%). Industry profits increase in $159M$. Social surplus increases in $38M$ (less than 1%)

From current duopoly (1) to symmetruic duopoly (2) Innovation rate declines from 0.599 to 0.501. Mean quality declines from 261% to 148%. Investment in R&D: declines by $178M$ per quarter. Price declines in $48 (24%). Consumer surplus increases in $34M$ (1.2%). Industry profits decline in $8M$. Social surplus increases in $26M$ (less than 1%)

From current scenario (1) to myopic pricing. It reduces prices, increases CS, and reduces firms' profits. Innovation rates and investment in R&D decline dramatically. Why? The higher induce firms to innovate more rapidly. Prices are higher with dynamic pricing because firms want to preserve future demand.
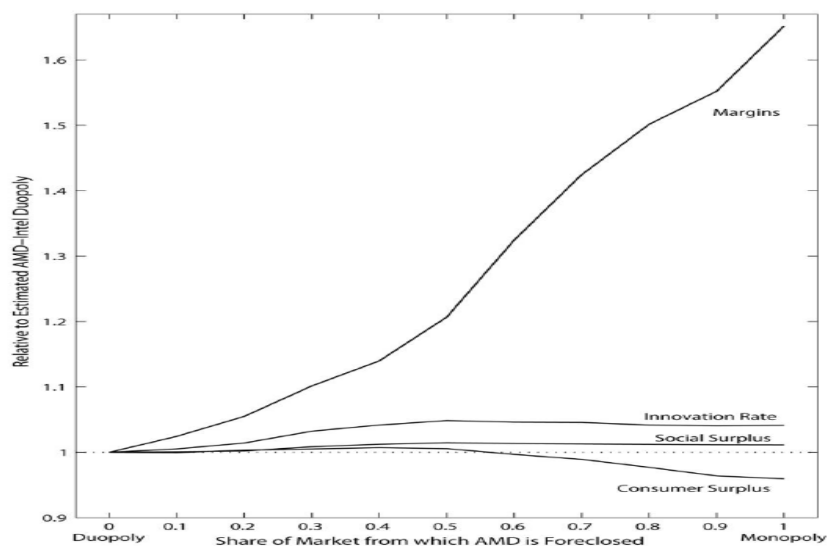
Fig. 6.—Foreclosing AMD from the market

The finding that innovation by a monopoly exceeds that of a duopoly reflects two features of the model: the monopoly must innovate to induce consumers to upgrade; the monopoly is able to extract much of the potential surplus from these upgrades because of its substantial pricing power. If there were a steady flow of new consumers into the market, such that most demand were not replacement, the monopoly would reduce innovation below that of the duopoly.

Counterfactuals: Foreclosure. In 2009, Intel paid AMD $1.25 billion to settle claims that Intel's anticompetitive practices foreclosed AMD from many consumers. To study the effect of such practices on innovation, prices, and welfare, the authors perform a series of counterfactual simulations in which they vary the portion of the market to which Intel has exclusive access. Let ζ be the proportion of foreclosure market. Intel market share becomes: $s_j^* = \zeta \, \widehat{s}_j + (1 - \zeta) \, s_j$, where s_j is the market share when AMD is competing, and $\widehat{s}_j$ is the market share when Intel competes only with the outside alternative.

Counterfactuals: Foreclosure

Margins monotonically rise steeply. Innovation exhibits an inverted U with a peak at $\zeta = 0.5$. Consumer surplus is actually higher when AMD is barred from a portion of the market, peaking at 40% foreclosure. This finding highlights the importance of accounting for innovation in antitrust policy: the decrease in consumer surplus from higher prices can be more than offset by the compounding effects of higher innovation rates.

Counterfactuals: Product substitutability

Innovation in the monopoly exhibits an inverted U as substitutability increases. Innovation in the duopoly increases as substitutability increases until Var( ) becomes too small for firms with similar qualities to coexist. - Beyond this "shakeout" threshold, the laggard
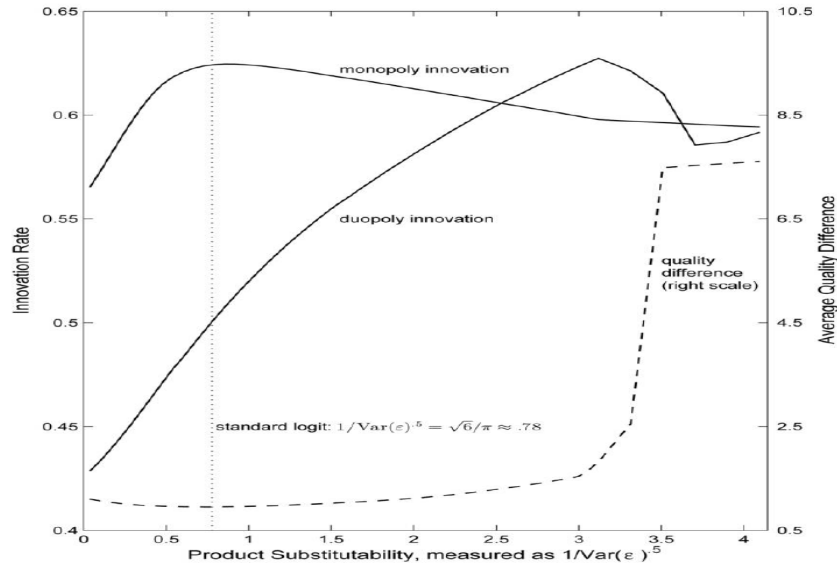
eventually concedes the market as evidenced by the sharp increase in the quality difference. Duopoly innovation is higher than monopoly innovation when substitutability is near the shakeout threshold.

Summary of results. The rate of innovation in product quality would be 4.2% higher if Intel were a monopolist, consistent with Schumpeter.

Without AMD, higher margins spur Intel to innovate faster to generate upgrade sales. As in Coase's (1972) conjecture, product durability can limit welfare losses from market power. This result, however, depends on the degree of competition from past sales. If first-time purchasers were to arrive sufficiently faster than we observe, innovation in an Intel monopoly would be lower, not higher, since upgrade sales would be less important.

Bibliography

[1] Aguirregabiria, V. (2011): "A Method for Implementing Counterfactual Experiments in Models with Multiple Equilibria," Economics Letters, forthcoming.

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

[3] Aguirregabiria, V. and P. Mira, 2007, Sequential estimation of dynamic discrete games. Econometrica 75, 1–53.

[4] Aguirregabiria, V. and P. Mira (2010), 'Dynamic discrete choice structural models: a survey', Journal of Econometrics, 156 (1), 38–67.

[5] Aguirregabiria, V., P. Mira, and H. Roman, 2007a, An estimable dynamic model of entry, exit and growth in oligopoly retail markets. American Economic Review 97, 449-454.

[6] Aguirregabiria, V., P. Mira, and H. Roman, 2007b, Inter-industry heterogeneity in market structure and dynamic oligopoly structural models. Manuscript. The University of Toronto.

[7] Aguirregabiria, V. and A. Nevo (2013), 'Recent developments in empirical IO: dynamic demand and dynamic games', in D. Acemoglu, M. Arellano and E. Deckel (eds), Advances in Economics and Econometrics. Tenth World Congress. Vol. 3, Econometrics, Cambridge: Cambridge University Press, pp. 53–122.

[8] Aguirregabiria, V. and J. Suzuki (2014), 'Identification and counterfactuals in dynamic models of market entry and exit', Quantitative Marketing and Economics, 12 (3), 267–304.

[9] Aguirregabiria, V. and G. Vicentini (forthcoming), 'Dynamic spatial competition between multi-store firms', Journal of Industrial Economics.

[10] Arcidiacono, P. and R. Miller, 2008, CCP Estimation of Dynamic Discrete Choice Models with Unobserved Heterogeneity. Manuscript. Duke University.

[11] Bajari, P., L. Benkard and J. Levin (2007): "Estimating dynamic models of imperfect competition," Econometrica, 75, 1331-1370.

[12] Bajari, P. and H. Hong, 2005, Semiparametric Estimation of a Dynamic Game of Incomplete Information. Manuscript. Duke University.

[13] Benkard, L. (2000): "Learning and Forgetting: The Dynamics of Aircraft Production," American Economic Review, September 2000.

[14] Benkard, L. (2004): "A Dynamic Analysis of the Market for Wide-Bodied Commercial Aircraft," Review of Economic Studies, July 2004.

[15] Benkard, L., G. Weintraub and B. Van Roy (2008): "Markov Perfect Industry Dynamics with Many Firms," Econometrica, November 2008.

[16] Benkard, L., G. Weintraub and B. Van Roy (2010): "Computational Methods for Oblivious Equilibrium," Operations Research, July-August 2010.

[17] Beresteanu, A. and P. Ellickson, 2005, The Dynamics of Retail Oligopoly. Manuscript. Department of Economics. Duke University.

[18] Bond, S. and J. Van Reenen (2007): "Microeconometric Models of Investment and Employment," in J. Heckman and E. Leamer (editors) *Handbook of Econometrics, Vol. 6A*. North Holland. Amsterdam.

[19] Bresnahan, T. and P. Reiss (1994): "Measuring the Importance of Sunk Costs," Annales D'Économie et de Statistique, 31, 183-217.

[20] Collard-Wexler, A. (2006): "Productivity Dispersion and Plant Selection in the Ready-Mix Concrete Industry," Manuscript. New York University.

[21] Collard-Wexler, A. (2006): "Demand Fluctuations in the Ready-Mix Concrete Industry," Manuscript. New York University.

[22] Doraszelski, U., and Satterthwaite, M. (2010): "Computable Markov-Perfect Industry Dynamics," *RAND Journal of Economics*, 41(2), 215–243.

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

[24] Dunne, T., S. Klimek, M. Roberts, and Y. Xu, 2006, Entry and Exit in Geographic Markets. Manuscript. Department of Economics. Pennsylvania State University.

[25] Einav, L. (2010): "Not All Rivals Look Alike: Estimating an Equilibrium Model of the Release Date Timing Game," Economic Inquiry 48(2), 369-390.

[26] Ericson, R. and A. Pakes (1995): "Markov Perfect Industry Dynamics: A Framework for Empirical Work," Review of Economic Studies, 62, 53-82.

[27] Esteban, S. & Shum, M. (2007). "Durable goods oligopoly with secondary markets: The case of automobiles," RAND Journal of Economics, 38, 332-354.

[28] Fershtman, C. and A. Pakes (2000): "A Dynamic Game with Collusion and Price Wars," RAND Journal of Economics," 31, 207-36.

[29] Goettler, R. and B. Gordon (2009): "Does AMD spur Intel to innovate more?" manuscript. University of Chicago Booth School of Business.

[30] Hashmi, A. and J. Van Biesebroeck (2010): "Market Structure and Innovation: A Dynamic Analysis of the Global Automobile Industry," NBER Working Paper 15959.

[31] Heckman, J., and B. Singer, 1984, A method for minimizing the impact of distributional assumptions in economic models for duration data. Econometrica 52, 271-320.

[32] Heckman, J. and S. Navarro, 2007, Dynamic Discrete Choice and Dynamic Treatment Effects. Journal of Econometrics 136, 341-396.

[33] Hendricks, K., M. Piccione and G. Tan (1995): "The Economics of Hubs: The Case of Monopoly," (with M. Piccione and G. Tan), Review of Economic Studies, 62, 83–99.

[34] Hendricks, K., M. Piccione and G. Tan (1997): "Entry and Exit in Hub-Spoke Networks," The RAND Journal of Economics, 28, 291–303.

[35] Hendricks, K., M. Piccione and G. Tan (1999): "Equilibria in Networks," Econometrica,

[36] Jofre-Bonet, M. and M. Pesendorfer, 2003, Estimation of a Dynamic Auction Game. Econometrica 71, 1443–1489.

[37] Kano, K., 2006, Menu Costs, Strategic Interactions and Retail Price Movements. Manuscript. Queen's University.

[38] Kryukov, Y. (2010): "Dynamic R&D and the Effectiveness of Policy Intervention in the Pharmaceutical Industry," manuscript. Tepper School of Business. Carnegie Mellon University.

[39] Macieira, J. (2007): "Extending the frontier: A structural model of investment and technological competition in the supercomputer industry," manuscript. Virginia Tech.

[40] Miller, N. and M. Osborne (2013): "Spatial Differentiation and Price Discrimination in the Cement Industry: Evidence from a Structural Model," manuscript. Rotman School of Management. University of Toronto.

[41] Nevo, A. and Rossi, F. (2008): "An approach for extending dynamic models to settings with multi-product firms," Economics Letters, 100(1), 49-52.

[42] Pakes, A. and R. Ericson (1998): "Empirical Implications of Alternative Models of Firm Dynamics," Journal of Economic Theory, 79, 1-46.

[43] Pakes, A. and P. McGuire (2001): "Stochastic Algorithms, Symmetric Markov Perfect Equilibria, and the 'Curse' of Dimensionality", Econometrica, 69, 1261-81.

[44] Pakes, A. and P. McGuire (1994): "Computing Markov Perfect Nash Equilibrium: Numerical Implications of a Dynamic Differentiated Product Model," RAND Journal of Economics, 25, 555-589.

[45] Pakes, A., M. Ostrovsky, and S. Berry (2007), "Simple Estimators for the Parameters of Discrete Dynamic Games, with Entry/Exit Examples", RAND Journal of Economics, 38(2), 373-399.

[46] Pesendorfer, M. and M. Jofre-Bonet (2003): "Estimation of a Dynamic Auction Game," Econometrica, 71(5), 1443-1489.

[47] Pesendorfer, M. and P. Schmidt-Dengler (2008): "Asymptotic Least Squares Estimators for Dynamic Games," Review of Economic Studies, 75(3), pages 901-928.

[48] Pesendorfer, M. and P. Schmidt-Dengler (2010): "Sequential Estimation of Dynamic Discrete Games: A Comment," Econometrica, 78(2), pages 833-842.

[49] Rust, J., 1994b, Estimation of dynamic structural models, problems and prospects: discrete decision processes, in C. Sims (ed.) Advances in Econometrics. Sixth World Congress, Cambridge University Press.

[50] Ryan, S., 2004, The Costs of Environmental Regulation in a Concentrated Industry. Manuscript, MIT Department of Economics.

[51] Schmidt-Dengler, P. (2006): "The Timing of New Technology Adoption: The Case of MRI," manuscript.

[52] Sutton, J. (1991): "Sunk costs and market structure: Price competition, advertising, and the evolution of concentration." MIT press. Cambridge, MA, USA.

[53] Suzuki, J. (2010): "Land Use Regulation as a Barrier to Entry: Evidence from the Texas Lodging Industry," manuscript. Department of Economics. University of Toronto.

[54] Sweeting, A. (2007): "Dynamic Product Repositioning in Differentiated Product Markets: The Case of Format Switching in the Commercial Radio Industry," NBER Working Paper #13522.

[55] Toivanen, O. and M. Waterson (2011): "Retail Chain Expansion: The Early Years of McDonalds in Great Britain," CEPR Discussion Papers 8534, C.E.P.R. Discussion Papers.

[56] Walrath, M. (2010): "Religion as an Industry: Estimating a Strategic Entry Model for Churches," manuscript. University of St. Thomas.

[57] AW, B., ROBERTS, M. and XU, D. (2011), "R&D Investment, Exporting, and Productivity Dynamics", American Economic Review, 101, 1312-1344.

[58] ERICSON, R., and A. PAKES (1995): "Markov-Perfect Industry Dynamics: A Framework for Empirical Work," Review of Economic Studies, 62, 53-82.

[59] GOETLER, R. and B. GORDON (2011): "Does AMD spur Intel to innovate more?" Journal of Political Economy, 119(6), 1141-1200.

[60] HASHMI, A. and J. VAN BIESEBROECK (2016): "The relationship between market structure and innovation in industry equilibrium: a case study of the global automobile industry," Review of Economics and Statistics, 98(1), 192-208.

[61] IGAMI, M (2017): "Estimating the innovator's dilemma: Structural analysis of creative destruction in the hard disk drive industry, 1981–1998," Journal of Political Economy, 125(3), 798-847.

[62] PETERS, B., M. ROBERTSs, V. VUONG, and H FRYGES (2017): "Estimating dynamic R&D choice: an analysis of costs and long-run benefits," The RAND Journal of Economics, 48(2), 409-437.

[63] VIVES, X. (2008): "Innovation and competitive pressure," Journal of Industrial Economics, 56(3): 419-469.

[64] XU, D. (2018): "A Structural Empirical Model of R&D, Firm Heterogeneity, and Industry Evolution," Journal of Industrial Economics, forthcoming.

CHAPTER 10

Empirical Models of Auctions

Auctions are common mechanisms for selling goods and services such as agricultural products (e.g., fish, livestock, wine), natural resources (e.g., timber, oil and gas drilling rights), government contracts (procurement auctions), money in interbank markets, treasury bonds, electricity, or art work. More recently, internet auctions (e.g., eBay) have become a popular way of selling a diverse range of products.

Auctions can be modelled as games of incomplete information. A seller (or a buyer, in the case of a procurement auction) is interested in selling an object. The seller faces a number of potential buyers or bidders, and she does not know their valuations of the object. A bidder knows his own valuation of the object but not other bidders' values. Each bidder submits a bid to maximize his expected payoff. The rules of the auction determines who gets the object and the price he should pay. These rules (e.g., first price sealed bids, second price), and the conditions on bidders' information and on the correlation between their valuations (e.g., independent private values, common values) are important features that determine the predictions of the model.

Consider the auction of a single object with N bidders indexed by $i \in \{1, 2, ..., N\}$. Bidder i's valuation for the object is $u_i = U(v_i, c)$ where $U(.,.)$ is an increasing function in both arguments; v_i represents a bidder's private signal; and c is a common value that affects the valuations of all the bidders. In an auction, the value of the vector $(v_1, v_2, ..., v_n, c)$ is a random draw from the joint cumulative distribution function $\mathbf{F}(v_1, v_2, ..., v_n, c)$ that is continuously differentiable and has compact support $[\underline{v}, \overline{v}]^2 \times [\underline{c}, \overline{c}]$. Each bidder knows her own private value v_i and the functions U and $\mathbf{F}$, but she does not know the other bidders' private values. Depending on the model, she may or may not know the common component c. The game is said to be symmetric if bidders are identical ex ante, i.e., if the distribution $\mathbf{F}$ is exchangeable in its first N arguments.

Each bidder decides her bid, b_i, to maximize her expect payoff. Most of the empirical literature has focused on first-price auctions: the winner is the highest bidder (provided it is higher than the seller's reservation price) and she pays his bid. Under this rule, the expected payoff is:

$$\pi_i^e(b_i) = \mathbb{E}\left(1\{b_i > b_j \; \forall j \neq i\} \; [U(v_i, c) - b_i]\right) \tag{0.1}$$

where $1\{.\}$ is the indicator function. This literature assumes that bids come from a Bayesian Nash equilibrium (BNE). This BNE is described as a vector of N strategy functions $\{s_i(v_i) : i = 1, 2, ..., N\}$ such that each bidder's strategy maximizes her expected payoff taking as given the strategy functions of the other bidders:

$$s_i(v_i) = \arg\max_{b_i} \ \mathbb{E}\left(1\left\{b_i > s_j(v_j) \ \forall j \neq i\right\} \ [U(v_i, c) - b_i]\right) \tag{0.2}$$

where the expectation is taken over the joint distribution of $\{v_j : j \neq i\}$ (and c, if this is not common knowledge). This BNE can be described as the solution of a system of differential equations.

Most empirical applications of structural auction models have focused on the *Independent Private Values* (IPV) model. This model assumes that valuations depend only on private information signals, $U(v_i, c) = v_i$, and they are independently and identically distributed, i.e., $\mathbf{F}(v_1, v_2, ..., v_n) = \prod_{i=1}^{N} F(v_i)$. It also imposes the restriction that the data comes from a symmetric BNE: $s_i(v_i) = s(v_i)$ for every bidder i. A BNE of the IPV model can be described as a strategy function $s(.)$ that solves the differential equation:

$$s(v_i) = v_i - \frac{F(v_i) \ s'(v_i)}{(N-1) \ f(v_i)} \tag{0.3}$$

subject to the boundary condition boundary $s(\underline{v}) = \underline{v}$, and where f is the density function of the distribution F. This differential equation has a unique solution that has following closed-form expression:

$$b_i = s(v_i) = v_i - \frac{1}{\left[F(v_i)\right]^{N-1}} \int_{\underline{v}}^{v_i} \left[F(u)\right]^{N-1} du \tag{0.4}$$

Auction data is widely available. In many countries, procurement auction data must be publicly available by law. Empirical researchers have used these data to answer different empirical questions such as detecting collusion among bidders, testing different auction models, or designing auction rules that maximize seller's revenue or total welfare.

The first empirical papers on auctions focused on testing important predictions of the model, without estimating the structural parameters (Hendricks and Porter, 1988; Hendricks, Porter, and Wilson, 1994; Porter, 1995). The papers by Paarsch (1992, 1997) and Laffont, Ossard, and Vuong (1995) present the first structural estimations of auction models.

In the structural estimation of auction models, the researcher has some information on bids and uses this information and the equilibrium conditions to estimate the distribution of bidders' valuations. Auction data may come in different forms, and this has important implications for the identification and estimation of the model. In an ideal situation, the researcher has a random sample of T independent auctions (indexed by t) of the same type of object from the same population of bidders, and she observes the bids of each of the N_t bidders at every auction t in the sample. Such ideal situations are quite uncommon. For

instance, often the researcher observes only the winning bid. It is also common that there is heterogeneity across the T auctions (e.g., different environments, or non identical objects) such that it is not plausible to assume that the same distribution of bidders' valuations, $F(.)$, applies to the T auctions. In that case, it is useful to have observable auction characteristics, X_t, such that the researcher may assume that two auctions with the same observable characteristics have the same distributions of valuations: $F_t(v|X_t) = F(v|X_t)$ for every auction t. In general, an auction dataset can be described as:

$$Data = \left\{ b_t^{(n)}, X_t : n = 1, ..., \overline{N}_t; t = 1, 2, ..., T \right\} \tag{0.5}$$

where $b_t^{(1)}$ is the largest bid, $b_t^{(2)}$ is the second largest, and so on; and $\overline{N}_t$ is the number of bids the researcher observes in auction t. When the dataset includes only information on winning bids, we have that $\overline{N}_t = 1$ for any auction t.

Tree planting procurement auctions in British Columbia. Paarsch (1992) studies first price sealed-bid auctions of tree planting contracts operated by the Forest Service (government agency) in the province of British Columbia, Canada. The object of an auction is described by the number and type of trees to plant and the location. The bidding variable is the price per tree, and the winner of the auction is the firm with the lowest price. The dataset consists of 144 auctions in the same forest region between 1985 and 1988 with information on all the bids. Paarsch estimates structurally independent private value models and common value models under different parametric specifications of the distribution of firms' costs. All the specifications of private value models are rejected. The estimated common value models are consistent with observed bidders' behavior. More specifically, there is evidence consistent with bidders' concern for the *winner's curse* and with bid functions that increase with the number of bidders.

The first empirical applications on structural auction models consider parametric specifications of the distribution of valuations (Paarsch, 1992, 1997; Laffont, Ossard, and Vuong, 1995; Baldwin, Marshall, and Richard, 1997). However, the more recent literature has focused on the nonparametric identification and estimation of this distribution. Guerre, Perrigne, and Vuong (2000) obtained an important identification result and estimation method in this literature. They show that equation (0.4), that characterizes the equilibrium of the model, implies that a bidder's valuation is a known function of his bid and the distribution of observed bids. Let $G(b)$ and $g(b)$ be the distribution and the density function of bids, respectively, implied by the equilibrium of the model. Since the equilibrium bidding strategy, $s(v_i)$, is strictly increasing, we have that $v_i = s^{-1}(b_i)$ and $G(b_i) = F(s^{-1}(b_i))$, and this implies that $g(b_i) = f(v_i)/s'(v_i)$. Solving these expressions into the differential equation

(0.3), we get:

$$v_i = \xi\left(b_i, G\right) = b_i + \frac{G(b_i)}{(N-1)\ g(b_i)} \tag{0.6}$$

Based on this equation, the distribution of valuations can be estimated from the data using a two-step procedure. Suppose for the moment that the data consists of a random sample of independent and identical auctions with information on all bids. Then, the distribution function G can be consistently estimated at any value $b \in [\underline{b}, \overline{b}]$ using the estimator $\widehat{G}(b) = (NT)^{-1} \sum_{t=1}^{T} \sum_{n=1}^{N} 1\{b_t^{(n)} \leq b\}$, and the density function can be estimated using the kernel method, $\widehat{g}(b) = (NT\ h_g)^{-1} \sum_{t=1}^{T} \sum_{n=1}^{N} K[(b_t^{(n)} - b)/h_g]$, where h_g is the bandwidth parameter. In a second-step, we can use equation (0.6) to construct the estimated pseudo-values $\widehat{v}_t^{(n)} = \xi\left(b_t^{(n)}, \widehat{G}\right)$ and use them to obtain a kernel estimator of the density of values: for any $v \in [\underline{v}, \overline{v}]$, $\widehat{f}(v) = (NT\ h_f)^{-1} \sum_{t=1}^{T} \sum_{n=1}^{N} K[(\widehat{v}_t^{(n)} - v)/h_f]$. GPV show consistency and asymptotic normality of this estimator and its speed of convergence. The also show that the estimator can be easily generalized to datasets where only the winning bid, $b_t^{(1)}$, is observed.

Athey and Haile (2002) provide a comprehensive study on the nonparametric identification of auction models. For IPV models, they show that the asymmetric IPV model is identified from data of winning bids if the identity of the winner is observed. When the distribution of values depends on observable auction characteristics, $F(v|X_t)$, they show that this distribution is identified from data of winning bids, both in the symmetric and the asymmetric IPV model. They also provide conditions for the identification of the affiliated private value model and the common values model.

In some applications, especially in procurement auctions, there may be substantial heterogeneity across auctions after controlling for the observable characteristics. Not controlling for this heterogeneity can generate important biases in the estimated distributions of valuations. Krasnokutskaya (2011) and Asker (2010) propose and estimate auction models of IPV with unobserved auction heterogeneity.

In Krasnokutskaya's model, bidders' valuations have a multiplicative structure: $u_{it} = v_{it} * c_t$, where v_{it} is private information of bidder i at auction t, and c_t is common knowledge to all the bidders in auction t. She provides sufficient conditions for the nonparametric identification of the distribution of the two components, and proposes an estimation method.[1] Krasnokutskaya applies her method to data from Michigan highway procurement auctions. She finds that, after conditioning on observable auction characteristics (e.g., number of bidders and project size), private information explains only 34% of the sample variation in winning bids. The remaining sample variation comes from unobserved heterogeneity from the point of view of the researcher. Estimates of the model that ignore this unobserved

---

[1] A key (and very intuitive) identification condition is that the researcher observes multiple bids for each auction. Data with only winning bids is not sufficient.

heterogeneity provide substantial biases in the average and the variance of firms' costs, and underestimate firms' mark-ups.

Asker (2010) considers a similar model where bidders' valuations have a multiplicative structure between IPVs and common knowledge auction heterogeneity. He applies this model to estimate the damages and efficiency costs of a "bidding ring" (cartel) in the US market for collectible stamps. Like Krasnokutskaya, he finds that accounting for unobserved auction heterogeneity has an important impact on the estimated model and its economic implications. The model without unobserved heterogeneity over-estimates the cartel's damages to the seller by more than 100%, and under-estimates the efficiency loss from the cartel by almost 50%.

The recent literature on structural auction models has extended the standard model in different important directions. Bajari and Hortacsu (2003), Li and Zheng (2009), Athey, Levin, and Seira (2011), Marmer, Shneyerov, and Xu (2013), and Gentry and Li (2014) study endogenous entry of bidders (and sellers). Jofre-Bonet and Pesendorfer (2003) estimate a dynamic structural model of procurement auctions where firms have capacity constraints and are forward-looking. Groeger (2014) estimates a dynamic model of entry in procurement auctions where firms have sunk entry costs. Lu and Perrigne (2008), Guerre, Perrigne, and Vuong (2009), Campo et al. (2011) incorporate bidders' risk aversion, provide conditions for nonparametric identification, and propose estimation methods. Finally, Lewis and Bajari (2011) and Takahashi (2014) study procurement auctions where the winner is determined by a scoring rule that weights both the price and the quality in a firm's bid.

Bibliography

[1] Asker, J. (2010): "A Study of the Internal Organization of a Bidding Cartel", American Economic Review, 100(3), 724-762, 2010.

[2] Athey, S., and P. Haile (2002) "Identification of Standard Auction Models," Econometrica, 70(6), 2107-2140.

[3] Bajari, P. and A. Hortacsu (2003): "The Winner's Curse, Reserve Prices and Endogenous Entry: Empirical Insights from eBay Auctions," Rand Journal of Economics, 3(2), 329-355.

[4] Bajari, P. and A. Hortacsu (2003): "Economic Insights from Internet Auctions," Journal of Economic Literature, 42, 457-486.

[5] Bajari, P. and S. Tadelis (2001): "Incentives Versus Transaction Costs: A Theory of Procurement Contracts," Rand Journal of Economics, 32(3), 287-307.

[6] Chernozhukov, V., Hong, H., and E. Tamer, 2007, Estimation and Confidence Regions for Parameter Sets in Econometric Models. Econometrica 75, 1243-1284.

[7] Green, E., and R. Porter (1984): "Noncooperative Collusion Under Imperfect Price Information," Econometrica, 52, 87-100.

[8] Guerre, E., I. Perrigne, and Q. Vuong (2000): "Optimal Nonparametric Estimation of First-Price Auctions," Econometrica, 68, 525-574.

[9] Hendricks, K., and H. Paarsch (1995): "A Survey of Recent Empirical Work Concerning Auctions," Canadian Journal of Economics, 403–426.

[10] Hendricks, K., and R. Porter (1988): "An Empirical Study of an Auction with Asymmetric Information," American Economic Review, 78 (5), 865–883.

[11] Hendricks, K., and R. Porter (1996): "Determinants of the Timing and Incidence of Exploratory Drilling on Offshore Wildcat Tracts," American Economic Review, 86 (3), 388–407.

[12] Hendricks, K., and R. Porter (2007): "A Survey of Empirical Work in Auctions," Handbook of Industrial Organization, Vol. III, edited by R. Porter and M. Armstrong, Vickers, Amsterdam: North-Holland.

[13] Hendricks, K., R. Porter and G. Tan (2009): "Bidding Rings and the Winner's Curse: The Case of Offshore Federal Oil and Gas Lease Auctions," The RAND Journal of Economics.

[14] Jofre-Bonet, M. and M. Pesendorfer, 2003, Estimation of a Dynamic Auction Game. Econometrica 71, 1443–1489.

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

[16] Pesendorfer, M. and M. Jofre-Bonet (2003): "Estimation of a Dynamic Auction Game," Econometrica, 71(5), 1443-1489.

[17] Vuong, Q., 1989, Likelihood Ratio Test for Model Selection and Non-Nested Hypotheses. Econometrica 57, 307–333.

APPENDIX A

Appendix 1

1. Random Utility Models

Consider a discrete choice Random Utility Model (RUM) where the optimal choice, a^*, is defined as:

$$a^* = \arg\max_{a \in A} \; \{u_a + \varepsilon_a\}$$

$A = \{1, 2, ..., J\}$ is the set of feasible choice alternative. $u = (u_1, u_2, ..., u_J)$ is the vector with the deterministic or constant components of the utility. $\varepsilon = (\varepsilon_1, \varepsilon_2, ..., \varepsilon_J)$ is the vector with the stochastic or random component of the utility. The vector ε has a joint CDF $G(.)$ that is continuous and strictly increasing with respect to the Lebesgue measure in the Euclidean space.

This note derives closed form expressions for the distribution of the maximum utility (i.e., $\max_{a \in A} \{u_a + \varepsilon_a\}$), the expected maximum utility (i.e., $E(\max_{a \in A} \{u_a + \varepsilon_a\} | u)$), and the choice probabilities (i.e., $\Pr(a^* = a | u)$) under three different assumptions on the distribution of the vector ε: (1) iid Extreme Value distribution (MNL model); (2) nested Extreme Value distribution (NL model); and (3) Ordered Generalized Extreme Value distribution (OGEV model).

The following definitions and properties are used in the note.

Definition: A random variable X has a Double Exponential or Extreme Value distribution with location parameter μ and dispersion parameter σ if its CDF is:

$$G(X) = \exp\left\{ -\exp\left(-\left[\frac{X - \mu}{\sigma}\right] \right) \right\}$$

Definition: Maximum utility. Let v^* be the random variable that represents the maximum utility, i.e., $v^* \equiv \max_{a \in A} \{u_a + \varepsilon_a\}$. This maximum utility is a random variable because it depends on the vector of random variables ε.

Definition: Social Surplus function (McFadden). The social surplus function $S(u)$ is the expected value of the maximum utility conditional on the vector of constants u, i.e., $S(u) \equiv E(\max_{a \in A} \{u_a + \varepsilon_a\} | u)$.

Definition: Conditional choice probabilities (CCPs). The conditional choice probability $P(a)$ is the probability that alternative a is optimal choice, i.e., $P(a|u) \equiv Pr(a^* = a | u)$.

Definition: Conditional choice expected utilities (CCEU). The conditional choice expected utility $e(a,u)$ is the expected value of $u_a + \varepsilon_a$ conditional on the vector u and on alternative a being the optimal choice, i.e., $e(a,u) \equiv E(u_a + \varepsilon_a | u, a^* = a)$.

Williams-Daly-Zachary (WDZ) Theorem. Let $S(u)$ be the function that represents the expected maximum utility conditional on the vector of constants u, i.e., $S(u) \equiv E(\max_{a \in A} \{u_a + \varepsilon_a\} | u)$. Then, the conditional choice probabilities (CCPs), $Pr(a^* = a|u)$, can be obtained as the partial derivatives of the function $S(u)$, i.e.,

$$Pr(a^* = a|u) = \frac{\partial S(u)}{\partial u_a}$$

Proof: (reference here). By definition of $S(u)$, $\dfrac{\partial S(u)}{\partial u_a} = \dfrac{\partial}{\partial u_a} \displaystyle\int \max_{j \in A} \{u_j + \varepsilon_j\} \, dG(\varepsilon)$. Given the assumptions on the CDF of ε, we have that $\dfrac{\partial}{\partial u_a} \displaystyle\int \max_{j \in A} \{u_j + \varepsilon_j\} \, dG(\varepsilon) = \displaystyle\int \dfrac{\partial}{\partial u_a} \max_{j \in A} \{u_j + \varepsilon_j\} \, dG(\varepsilon)$. Therefore,

$$
\begin{aligned}
\frac{\partial S(u)}{\partial u_a} &= \int \frac{\partial}{\partial u_a} \max_{j \in A} \{u_j + \varepsilon_j\} \, dG(\varepsilon) \\[2mm]
&= \int 1\{u_a + \varepsilon_a \geq u_j + \varepsilon_j \text{ for any } j \in A\} dG(\varepsilon) \\[2mm]
&= Pr(a^* = a|u)
\end{aligned}
$$

Theorem. For any distribution of ε, any value of the vector u, and any choice alternative a, the conditional choice expected utility $e(a,u)$ is equal to the social surplus function $S(u)$, i.e., $e(a,u) = S(u)$ for any (a,u). Furthermore, this implies that $E(\varepsilon_a \mid u, a^* = a) = S(u) - u_a$.

Proof: (reference here). By definition, $e(a,u) = u_a + E(\varepsilon_a \mid u, a^* = a)$. Taking into account that the random variable v^* represents maximum utility, we have that the event $\{a^* = a\}$ is equivalent to the event $\{v^* = u_a + \varepsilon_a\}$. Therefore,

$$
\begin{aligned}
e(a,u) &= u_a + E(\varepsilon_a \mid u, \ v^* = u_a + \varepsilon_a) \\[2mm]
&= u_a + E(v^* - u_a \mid u) \\[2mm]
&= E(v^*|u) = S(u)
\end{aligned}
$$

Inversion Theorem. *** *Representation of $S(u)$ in terms of CCPs and utilities only. How to do it in general?* ****

2. Multinomial logit (MNL)

Suppose that the random variables in the vector ε are independent and identically distributed with double exponential distribution with zero location and dispersion σ. That is, for every alternative a, the CDF of ε_a is $G(\varepsilon_a) = \exp\left\{-\exp\left(-\frac{\varepsilon_a}{\sigma}\right)\right\}$.

(a) Distribution of the Maximum Utility

Let v^* be the random variable that represents the maximum utility, i.e., $v^* \equiv \max_{a \in A}\{u_a + \varepsilon_a\}$. This maximum utility is a random variable because it depends on the vector of random variables ε. By definition, the cumulative probability distribution of v^* is:

$$H(v) \equiv \Pr(v^* \leq v) \;=\; \prod_{a \in A} \Pr(\varepsilon_a \leq v - u_a)$$

$$= \prod_{a \in A} \exp\left\{-\exp\left(-\frac{v - u_a}{\sigma}\right)\right\}$$

$$= \exp\left\{-\exp\left(-\frac{v}{\sigma}\right) U\right\}$$

where $U \equiv \sum_{a \in A} \exp\left(\frac{u_a}{\sigma}\right)$. We can also write $H(v) = \exp\left\{-\exp\left(-\frac{v - \sigma \ln U}{\sigma}\right)\right\}$. This expression shows that the maximum utility v^* is a double exponential random variable with dispersion parameter σ and location parameter $\sigma \ln U$. Therefore, the maximum of a vector of i.i.d. double exponential random variables is also a double exponential random variable. This is the reason why this family of random variables is also called "extreme value". The density function of v^* is:

$$h(v) \equiv H'(v) \;=\; H(v) \frac{U}{\sigma} \exp\left(-\frac{v}{\sigma}\right)$$

(b) Expected maximum utility

By definition, $S(u) = E(v^*|u)$. Therefore,

$$S(u) \;=\; \int_{-\infty}^{+\infty} v^* \, h(v^*) \, dv^* \;=\; \int_{-\infty}^{+\infty} v^* \, \exp\left\{-\exp\left(-\frac{v^*}{\sigma}\right) U\right\} \frac{U}{\sigma} \exp\left(-\frac{v^*}{\sigma}\right) dv^*$$

We apply the change in variable: $z = \exp(-v^*/\sigma)$, such that $v^* = -\sigma \ln(z)$, and $dv^* = -\sigma(dz/z)$. Then,

$$S(u) \;=\; \int_{+\infty}^{0} -\sigma \ln(z) \, \exp\{-z\,U\} \frac{U}{\sigma} z \left(-\sigma \frac{dz}{z}\right)$$

$$= \; -\sigma U \int_{0}^{+\infty} \ln(z) \, \exp\{-z\,U\} \, dz$$

Using Laplace transformation we have that $\int_0^{+\infty} \ln(z) \exp\{-z\,U\}\,dz = \dfrac{\ln(U) + \gamma}{U}$, where γ is Euler's constant. Therefore, the expected maximum utility is:

$$S \;=\; \sigma U \left(\frac{\ln(U) + \gamma}{U} \right) \;=\; \sigma\,(\,\ln(U) + \gamma\,)$$

(c) Choice probabilities

By Williams-Daly-Zachary (WDZ) theorem, the optimal choice probabilities can be obtained by differentiating the surplus function. Therefore, for the MNL model,

$$P(a|u) \;=\; \sigma\,\frac{\partial \ln(U)}{\partial u_a} \;=\; \sigma\,\frac{\partial U}{\partial u_a}\,\frac{1}{U}$$

$$= \; \exp\left(\frac{u_a}{\sigma}\right)\frac{1}{U} \;=\; \frac{\exp\left(u_a/\sigma\right)}{\sum_{j \in A} \exp\left(u_j/\sigma\right)}$$

(d) Conditional choice expected utilities

As shown in general, $e(a, u) = S(u)$. This implies that $E(\varepsilon_a \mid u, a^* = a) = S(u) - u_a$. For the case of the i.i.d. double exponential ε we have that:

$$E(\varepsilon_a|u, a^* = a) = \sigma\,(\,\ln(U) + \gamma\,) - u_a$$

(e) Function relating $E(\varepsilon_a \mid u, a^* = a)$ and CCPs.

In some applications we are interested in the function that relates the expected value $E(\varepsilon_a|u, a^* = a)$ with conditional choice probabilities $\{P(j|u) : j = 1, 2, ..., J\}$. From the expression for $P(a|u)$ in the MNL model, we have that $\ln P(a|u) = u_a/\sigma - \ln U$, and therefore $\ln(U) = u_a/\sigma - \ln P_a$. Solving this expression in the previous formula for the expectation $E(\varepsilon_a|u, a^* = a)$ we get:

$$E(\varepsilon_a|u, a^* = a) = \sigma\,(u_a/\sigma - \ln P(a|u) + \gamma) - u_a = \sigma\,(\gamma - \ln P(a|u))$$

3. Nested logit (NL)

Suppose that the random variables in the vector ε have the following joint CDF:

$$G(\varepsilon) = \exp\left\{ -\sum_{r=1}^{R}\left[\sum_{a \in A_r} \exp\left(-\frac{\varepsilon_a}{\sigma_r}\right)\right]^{\frac{\sigma_r}{\delta}} \right\}$$

where $\{A_1, A_2, ..., A_R\}$ is a partition of A, and δ, σ_1, σ_2, ..., σ_R are positive parameters, with $\delta \leq 1$.

(a) Distribution of the Maximum Utility

$$H(v) \equiv \Pr(v^* \leq v) \;=\; \Pr(\varepsilon_a \leq v - u_a : for\ any\ a \in A)$$

$$= \exp\left\{-\sum_{r=1}^{R}\left[\sum_{a\in A_r}\exp\left(-\frac{v-u_a}{\sigma_r}\right)\right]^{\frac{\sigma_r}{\delta}}\right\}$$

$$= \exp\left\{-\exp\left(-\frac{v}{\delta}\right)\sum_{r=1}^{R}\left[\sum_{a\in A_r}\exp\left(\frac{u_a}{\sigma_r}\right)\right]^{\frac{\sigma_r}{\delta}}\right\}$$

$$= \exp\left\{-\exp\left(-\frac{v}{\delta}\right)U\right\}$$

where:

$$U \equiv \sum_{r=1}^{R}\left[\sum_{a\in A_r}\exp\left(\frac{u_a}{\sigma_r}\right)\right]^{\frac{\sigma_r}{\delta}} = \sum_{r=1}^{R}U_r^{1/\delta}$$

where $U_r \equiv \left[\sum_{a\in A_r}\exp\left(\frac{u_a}{\sigma_r}\right)\right]^{\sigma_r}$. The density function of v^* is:

$$h(v) \equiv H'(v) \;=\; H(v)\,\frac{U}{\delta}\,\exp\left(-\frac{v}{\delta}\right)$$

(b) Expected maximum utility

By definition, $S(u) = E(v^*)$. Therefore,

$$S(u) \;=\; \int_{-\infty}^{+\infty} v^*\,h(v^*)\,dv^* \;=\; \int_{-\infty}^{+\infty} v^*\,\exp\left\{-\exp\left(-\frac{v^*}{\delta}\right)U\right\}\frac{U}{\delta}\exp\left(-\frac{v^*}{\delta}\right)\,dv^*$$

Let's apply the following change in variable: $z = \exp(-v^*/\delta)$, such that $v^* = -\delta\ln(z)$, and $dv^* = -\delta(dz/z)$. Then,

$$S \;=\; \int_{+\infty}^{0} -\delta\ln(z)\,\exp\{-z\,U\}\,\frac{U}{\delta}\,z\left(-\delta\frac{dz}{z}\right) \;=\; -\delta U\int_{+\infty}^{0}\ln(z)\,\exp\{-z\,U\}\,dz$$

And using Laplace transformation:

$$S \;=\; \delta U\left(\frac{\ln(U)+\gamma}{U}\right) \;=\; \delta\left(\,\ln(U)+\gamma\,\right)$$

where γ is the Euler's constant.

(c) Choice probabilities

By Williams-Daly-Zachary (WDZ) theorem, choice probabilities can be obtained differentiating the surplus function. For the NL model:

$$P(a|u) = \delta\, \frac{\partial \ln(U)}{\partial u_a} = \delta \frac{\partial U}{\partial u_a} \frac{1}{U} =$$

$$= \delta \frac{\sigma_{ra}}{\delta} \left[\sum_{j \in A_{ra}} \exp\left(\frac{u_j}{\sigma_{ra}} \right) \right]^{\frac{\sigma_{ra}}{\delta} - 1} \frac{1}{\sigma_{ra}} \exp\left(\frac{u_a}{\sigma_{ra}} \right) \frac{1}{U}$$

$$= \frac{\exp\left(u_a/\sigma_{ra} \right)}{\sum_{j \in A_{ra}} \exp\left(u_j/\sigma_{ra} \right)} \; \frac{\left[\sum_{j \in A_{ra}} \exp\left(u_j/\sigma_{ra} \right) \right]^{\frac{\sigma_{ra}}{\delta}}}{\sum_{r=1}^{R} \left[\sum_{j \in A_r} \exp\left(u_j/\sigma_r \right) \right]^{\frac{\sigma_r}{\delta}}}$$

The first term is $q(a|r_a)$ (i.e., probability of choosing a given that we are in group A_{ra}), and the second term is $Q(r_a)$ (i.e., probability of selecting the group A_{ra}).

(d) Conditional choice expected utilities

As shown in general, $e(a, u) = S(u)$. This implies that $E(\varepsilon_a \mid u, a^* = a) = S(u) - u_a$. Given that for the NL model $S(u) = \delta \left(\ln U + \gamma \right)$ we have that:

$$E(\varepsilon_a|u, a^* = a) = \delta\gamma + \delta \ln U - u_a$$

(e) Function relating $E(\varepsilon_a \mid u, a^* = a)$ and CCPs.

To write $E(\varepsilon_a|u, a^* = a)$ in terms of choice probabilities, note that from the definition of $q(a|r_a)$ and $Q(r_a)$, we have that:

$$\ln q(a|r_a) = \frac{u_a - \ln U_{ra}}{\sigma_{ra}} \quad \Rightarrow \quad \ln U_{ra} = u_a - \sigma_{ra} \ln q(a|r_a)$$

and

$$\ln Q(r_a) = \frac{\ln U_{ra}}{\delta} - \ln U \quad \Rightarrow \quad \ln U = \frac{\ln U_{ra}}{\delta} - \ln Q(r_a)$$

Combining these expressions, we have that:

$$\ln U = \frac{u_a - \sigma_{ra} \ln q(a|r_a)}{\delta} - \ln Q(r_a)$$

Therefore,

$$e_a = \delta\gamma + \delta \left(\frac{u_a - \sigma_{ra} \ln q(a|r_a)}{\delta} - \ln Q(r_a) \right) - u_a$$

$$= \delta\gamma - \sigma_{ra} \; \ln q(a|r_a) - \delta \; \ln Q(r_a)$$

4. Ordered GEV (OGEV)

Suppose that the random variables in the vector ε have the following joint CDF:

$$G(\varepsilon) = \exp\left\{-\sum_{r=1}^{J+M}\left[\sum_{a\in B_r} W_{r-a}\exp\left(-\frac{\varepsilon_a}{\sigma_r}\right)\right]^{\frac{\sigma_r}{\delta}}\right\}$$

where:

(1) M is a positive integer;

(2) $\{B_1, B_2, ..., B_{J+M}\}$ are $J+M$ subsets of A, with the following definition:

$$B_r = \{a \in A : r - M \le a \le r\}$$

For instance, if $A = \{1, 2, 3, 4, 5\}$ and $M = 2$, then $B_1 = \{1\}$, $B_2 = \{1, 2\}$, $B_3 = \{1, 2, 3\}$, $B_4 = \{2, 3, 4\}$, $B_5 = \{3, 4, 5\}$, $B_6 = \{4, 5\}$, and $B_7 = \{5\}$.

(3) δ, and σ_1, σ_2, ..., σ_{J+M} are positive parameters, with $\delta \le 1$;

(4) W_0, W_1, ..., W_M are constants (weights) such that: $W_m \ge 0$, and $\sum_{m=0}^{M} W_m = 1$.

(a) Distribution of the Maximum Utility

$$
\begin{aligned}
H(v) \equiv \Pr(v^* \le v) \;&=\; \Pr(\varepsilon_a \le v - u_a : \text{ for any } a \in A) \\[2mm]
&=\; \exp\left\{-\sum_{r=1}^{J+M}\left[\sum_{a\in B_r} W_{r-a}\exp\left(-\frac{v - u_a}{\sigma_r}\right)\right]^{\frac{\sigma_r}{\delta}}\right\} \\[2mm]
&=\; \exp\left\{-\exp\left(-\frac{v}{\delta}\right)\sum_{r=1}^{J+M}\left[\sum_{a\in B_r} W_{r-a}\exp\left(\frac{u_a}{\sigma_r}\right)\right]^{\frac{\sigma_r}{\delta}}\right\} \\[2mm]
&=\; \exp\left\{-\exp\left(-\frac{v}{\delta}\right)U\right\}
\end{aligned}
$$

where:

$$U \equiv \sum_{r=1}^{J+M}\left[\sum_{a\in B_r} W_{r-a}\exp\left(\frac{u_a}{\sigma_r}\right)\right]^{\frac{\sigma_r}{\delta}} = \sum_{r=1}^{J+M} U_r^{1/\delta}$$

where $U_r \equiv \left[\sum_{a\in B_r} W_{r-a}\exp\left(\frac{u_a}{\sigma_r}\right)\right]^{\sigma_r}$. The density function of v^* is:

$$h(v) \equiv H'(v) \;=\; H(v)\,\frac{U}{\delta}\,\exp\left(-\frac{v}{\delta}\right)$$

(b) Expected maximum utility

By definition, $S(u) = E(v^*|u)$. Therefore,

$$S(u) \;=\; \int_{-\infty}^{+\infty} v^* \, h(v^*) \, dv^* \;=\; \int_{-\infty}^{+\infty} v^* \, \exp\left\{-\exp\left(-\frac{v^*}{\delta}\right) U\right\} \frac{U}{\delta} \, \exp\left(-\frac{v^*}{\delta}\right) dv^*$$

Let's apply the following change in variable: $z = \exp(-v^*/\delta)$, such that $v^* = -\delta \ln(z)$, and $dv^* = -\delta(dz/z)$. Then,

$$S \;=\; \int_{+\infty}^{0} -\delta \ln(z) \, \exp\left\{-z \, U\right\} \frac{U}{\delta} \, z \, \left(-\delta\frac{dz}{z}\right) \;=\; -\delta U \int_{0}^{+\infty} \ln(z) \, \exp\left\{-z \, U\right\} \, dz$$

And using Laplace transformation:

$$S \;=\; \delta U \left(\frac{\ln U + \gamma}{U}\right) \;=\; \delta\left(\ln U + \gamma\right) = \delta\gamma + \delta \, \ln\left[\sum_{r=1}^{J+M} \left[\sum_{a\in B_r} W_{r-a} \exp\left(\frac{u_a}{\sigma_r}\right)\right]^{\frac{\sigma_r}{\delta}}\right]$$

where γ is the Euler's constant.

(c) Choice probabilities

By Williams-Daly-Zachary (WDZ) theorem, choice probabilities can be obtained differentiating the surplus function.

$$P(a|u) = \frac{1}{U} \sum_{r=a}^{a+M} \left[\sum_{j\in B_r} W_{r-j} \exp\left(\frac{u_j}{\sigma_r}\right)\right]^{\frac{\sigma_r}{\delta}-1} W_{r-a} \exp\left(\frac{u_a}{\sigma_r}\right) = \sum_{r=a}^{a+M} q(a|r) \, Q(r)$$

where:

$$q(a|r) \;=\; \frac{W_{r-a} \, \exp\left(u_a/\sigma_r\right)}{\displaystyle\sum_{j\in B_r} W_{r-j} \exp\left(u_j/\sigma_r\right)} = \frac{\exp\left(u_a/\sigma_r\right)}{\exp(\ln U_r/\sigma_r)}$$

$$Q(r) \;=\; \frac{\exp\left(\ln U_r/\delta\right)}{\displaystyle\sum_{j=1}^{J+M} \exp\left(\ln U_j/\delta\right)} = \frac{\exp\left(\ln U_r/\delta\right)}{U}$$

(d) Conditional choice expected utilities

As shown in general, $e(a,u) = S(u)$. This implies that $E(\varepsilon_a \mid u, a^* = a) = S(u) - u_a$. Given that for the OGEV model $S(u) = \delta\left(\ln U + \gamma\right)$ we have that:

$$E(\varepsilon_a|u, a^* = a) = \delta\gamma + \delta \ln U - u_a$$

(e) Function relating $E(\varepsilon_a \mid u, a^* = a)$ and CCPs.

To write $E(\varepsilon_a|u, a^* = a)$ in terms of choice probabilities, note that from the definition of $q(a|r)$ and $Q(r)$, we have that: ....

TBC

Appendix 2. Problems

1. Problem set #1

Context. At the end of year 2002, the federal government of the Republic of Greenishtan introduced a new environmental regulation on the cement industry, one of the major polluting industries. The most important features of this regulation is that new plants, in order to operate in the industry, should pass an environmental test and should install new equipment that contributes to reduce pollutant emissions. Industry experts consider that this new law increased the fixed cost of operating in this industry. However, these experts disagree in the magnitude of the effect. There is also disagreement with respect to whether the new law affected variable costs, competition, prices, and output. You have been hired by the Ministry of Industry as an independent researcher to study and to evaluate the effects of this policy on output, prices, firms' profits, and consumer welfare.

Data. To perform your evaluation, you have a panel dataset with annual information on the industry for the period 1998-2007. The Stata datafile `eco2901_problemset_01_2011.dta` contains panel data from 200 local markets (census tracts) over 10 years (1998-2007) for the cement industry in the Republic of Greenishtan. The local markets in this dataset have been selected following criteria similar to the ones in Bresnahan and Reiss (1991). This is the list of variables in the dataset:

Variable name		Description
`market`	:	Code of local market
`year`	:	Year
`pop`	:	Population of local market
`income`	:	Per capita income in local market
`output`	:	Annual output produced in the local market
`price`	:	Price of cement in local market
`pinput`	:	Price index of intermediate inputs in local market
`nplant`	:	Number of cement plants in local market at current year

Model. To answer our empirical questions, we consider a model in the spirit of the model by Bresnahan and Reiss that we have seen in class. The main difference with respect to

that model is that we specify the demand function and the cost function in the industry and make it explicit the relationship between these primitives and the profit of a plant.

Demand of cement in market m at period t. We assume that cement is an homogeneous product and consider the following inverse demand function:

$$\ln P_{mt} = \alpha_0^D + \alpha_1^D \ln POP_{mt} + \alpha_2^D \ln INC_{mt} - \alpha_3^D \ln Q_{mt} + \varepsilon_{mt}^D$$

where $\alpha^{D\prime}s$ are demand parameters, Q_{mt} represents output, POP_{mt} is population, INC_{mt} is per capita income, P_{mt} is price, and ε_{mt}^D is a component of the demand that is unobserved to the researcher.

Production costs. Let q_{mt} be the amount of output of a cement plant in market m and period t. The production cost function is $C_{mt}(q_{mt}) = FC_{mt} + MC_{mt} \, q_{mt}$, where FC_{mt} and MC_{mt} are the fixed cost function and the marginal cost, respectively. We consider the following specification for FC_{mt} and MC_{mt}:

$$FC_{mt} \;=\; \exp\left\{ X_{mt} \, \alpha^{FC} + \varepsilon_{mt}^{FC} \right\}$$

$$MC_{mt} \;=\; \exp\left\{ X_{mt} \, \alpha^{MC} + \varepsilon_{mt}^{MC} \right\}$$

where X_{mt} is the vector $(1, \ln POP_{mt}, \ln INC_{mt}, \ln PINPUT_{mt})$, where $PINPUT_{mt}$ is the index price of inputs (energy and limestone); α^{FC} and α^{MC} are vectors of parameters; and ε_{mt}^{FC} and ε_{mt}^{MC} are components of the fixed cost and the marginal cost, respectively, that are unobserved to the researcher. The main reason why we consider an exponential function in the specification of FC_{mt} and MC_{mt} is to impose the natural restriction that costs should be always positive.

Entry costs and scrapping value. For simplicity, we consider a static model and therefore we assume that there are not sunk entry costs.

Unobservables. Let ε_{mt} be the vector of unobservables $\varepsilon_{mt} \equiv (\varepsilon_{mt}^D, \varepsilon_{mt}^{MC}, \varepsilon_{mt}^{FC})$. We allow for serial correlation in these unobservables. In particular, we assume that each of these unobservables follows an AR(1) process. For $j \in \{D, MC, FC\}$:

$$\varepsilon_{mt}^j = \rho^j \, \varepsilon_{mt-1}^j + u_{mt}^j$$

where $\rho^j \in [0,1)$ is the autorregressive parameter, and the vector $u_{mt} = (u_{mt}^D, u_{mt}^{MC}, u_{mt}^{FC})$ is i.i.d. over markets and over time with a joint normal distribution with zero means and variance-covariance matrix Ω.

Question 1 [20 points]. **(a)** Propose an estimator of the demand parameters and explain the assumptions under which the estimator is consistent. **(b)** Obtain estimates and standard

errors. **(c)** Test the null hypothesis of "no structural break" in demand parameters after year 2002.

Question 2 [20 points]. **(a)** Describe how to use the Cournot equilibrium conditions to estimate the parameters in the marginal cost function. Explain the assumptions under which the estimator is consistent. **(b)** Obtain estimates and standard errors. **(c)** Test the null hypothesis of "no structural break" in the variable cost parameters after year 2003.

Question 3 [30 points]. Assume that $\rho^{FC} = 0$. **(a)** Describe how to estimate the parameters in the fixed cost function. Show that these costs are identified in dollar amounts (i.e., not only up to scale). Explain the assumptions under which the estimator is consistent. How does the estimation of fixed costs change if $\rho^{FC} \neq 0$? Explain. **(b)** Obtain estimates and standard errors. **(c)** Test the null hypothesis of "no structural break" in the fixed cost parameters after year 2003.

Question 4 [30 points]. Now, we use our estimates to evaluate the effects of the policy change. Suppose that we attribute to the new policy the estimated change in the parameters of the cost function, but not the estimated change in the demand parameters.

   (a) [10 points] Given the estimated parameters "after 2002", calculate the equilibrium values of the endogenous variables $\{P_{m,2003}, Q_{m,2003}, N_{m,2003}\}$ for every local market in 2003, i.e., for every value of the exogenous variables $(X_{m,2003}, \varepsilon_{m,2003})$. Obtain also firms' profits, consumer welfare, and total welfare.

   (b) [10 points] Now, consider the counterfactual scenario where demand parameters are the ones "after 2002" but cost parameters are the ones "before 2003". For this scenario, calculate the "counterfactual" equilibrium values of the endogenous variables $\{P^*_{m,2003}, Q^*_{m,2003}, N^*_{m,2003}\}$ for every local market in 2003. Also obtain the counterfactual values for firms' profits, consumer welfare, and total welfare.

   (c) [10 points] Obtain the effects of the policy one the number of firms, output, prices, firms' profits, consumer welfare, and total welfare. Comment the results. Present two-way graphs of these effects with the logarithm of population in the horizontal axis and the estimated on a certain endogenous variable in the vertical axis. Comment the results. What are the most important effects of this policy?

2. Problem set #2

   The Stata datafile `eco2901_problemset_01_chiledata_2010.dta` contains a panel dataset of 167 local markets in Chile with annual information over the years 1994 to 1999 and for

five retail industries: Restaurants ('Restaurantes,' product code 63111); Gas stations ('Gaso-lineras,' product code 62531); Bookstores ('Librerias,' product code 62547); Shoe Shops ('Calzado,' product code 62411); and Fish shops ('Pescaderias,' product code 62141). The 167 "isolated" local markets in this dataset have been selected following criteria similar to the ones in Bresnahan and Reiss (1991). This is the list of variables in the dataset with a brief description of each variable:

`comuna_code`	:	Coder of local market
`comuna_name`	:	Name of local market
`year`	:	Year
`procode`	:	Code of product/industry
`proname`	:	Name of product/industry
`pop`	:	Population of local market (in # people)
`areakm2`	:	Area of local market (in square Km)
`expc`	:	Annual expenditure per capita in all retail products in the local market
`nfirm`	:	Number of firms in local market and industry at current year
`nfirm_1`	:	Number of firms in local market and industry at previous year
`entries`	:	Number of new entrants in local market and industry during current year
`exits`	:	Number of exiting firms in local market and industry during current year

Consider the following static entry model in the spirit of Bresnahan and Reiss (JPE, 1991, hereinafter *BR-91*). The profit of an active firm in market m at year t is:

$$\Pi_{mt} = S_{mt} \ v(n_{mt}) - F_{mt}$$

where S_{mt} is a measure of market size; n_{mt} is the number of firms active in the market; $v(.)$ is the *variable profit per capita* and it is a decreasing function; and F_{mt} represents fixed operating costs in market m at period t. The function $v(.)$ is nonparametrically specific. The specification of market size is:

$$S_{mt} = POP_{mt} \ \exp\left\{\beta_0^S + \beta_1^S \ expc_{mt} + \varepsilon_{mt}^S\right\}$$

where POP_{mt} is the population in the local market; $expc_{mt}$ is per capita sales in all retail industries operating in the local market; β_0^S and β_1^S are parameters; and ε_{mt}^S is an unobservable component of market size. The specification of the fixed cost is:

$$F_{mt} = \exp\left\{\beta^F + \varepsilon_{mt}^F\right\}$$

where β^F is a parameter, and ε_{mt}^F in an unobservable component of the fixed cost. Define the unobservable $\varepsilon_{mt} \equiv \varepsilon_{mt}^S - \varepsilon_{mt}^F$. And let $X_{mt} \equiv (\ln POP_{mt}, expc_{mt})$ be the vector with the observable characteristics of the local market. We assume that ε_{mt} is independent of X_{mt} and iid over (m, t) $N(0, \sigma^2)$.

Question 1. [10 points] Show that the model implies the following probability distribution for the equilibrium number of firms: let $n_{\max}$ be the maximum value of n_{mt}, then for any

$n \in \{0, 1, ..., n_{\max}\}$:

$$\Pr\left(n_{mt} = n \mid X_{mt}\right) = \Pr\left(cut(n) \leq X_{mt}\left[\begin{array}{c}\frac{1}{\sigma} \\ \frac{\beta_1^S}{\sigma}\end{array}\right] + \frac{\varepsilon_{mt}}{\sigma} \leq cut(n+1)\right)$$

$$= \Phi\left(cut(n+1) - X_{mt}\left[\begin{array}{c}\frac{1}{\sigma} \\ \frac{\beta_1^S}{\sigma}\end{array}\right]\right) - \Phi\left(cut(n) - X_{mt}\left[\begin{array}{c}\frac{1}{\sigma} \\ \frac{\beta_1^S}{\sigma}\end{array}\right]\right)$$

where $cut(0)$, $cut(1)$, $cut(2)$, ... are parameters such that for $n \in \{1, 2, ..., n_{\max}\}$, $cut(n) \equiv (\beta^F - \beta_0^S - \ln v(n))/\sigma$, and $cut(0) \equiv -\infty$, and $cut(n_{\max} + 1) \equiv -\infty$.

Question 2. **[20 points]** Given the Ordered Probit structure of the model, estimate the vector of parameters $\{1/\sigma, \beta_1^S/\sigma, cut(1), cut(2), ..., cut(n_{\max})\}$ for each of the five industries separately. Given these estimates, obtain estimates of the parameters $\dfrac{v(n+1)}{v(n)}$ for $n \in \{1, 2, ..., n_{\max}\}$. Present a figure of the estimated function $\dfrac{v(n+1)}{v(n)}$ for each of the five industries. Interpret the results. Based on these results, what can we say about *the nature of competition* in each of these industries?

Question 3. [20 points] Repeat the same exercise as in Question 3 but using the following specification of the unobservable ε_{mt}:

$$\varepsilon_{mt} = \gamma_t + \delta_m + u_{mt}$$

where γ_t are time effects that can be captured by using time-dummies; δ_m are fixed market effects that can be captured by using market-dummies; and u_{mt} is independent of X_{mt} and iid over (m, t) $N(0, \sigma^2)$. Comment the results.

────────────────────────────────

Now, consider the following static entry model of incomplete information. There are N_{mt} potential entrants in market m at period t. The profit of an active firm in market m at year t is:

$$\Pi_{imt} = S_{mt} \, v(n_{mt}) - F_{imt}$$

Market size, S_{mt}, has the same specification as in Question 2. The firm-specific fixed cost, F_{mt}, has the following specification:

$$F_{imt} = \exp\left\{\beta^F + \varepsilon_{mt}^F + \xi_{imt}\right\}$$

The random variables ε_{mt}^S, ε_{mt}^F, and ξ_{imt} are unobservable to the researcher. From the point of view of the firms in the market, the variables ε_{mt}^S and ε_{mt}^F are common knowledge, while ξ_{imt} is private information of firm i. We assume that ξ_{imt} is independent of X_{mt} and iid over (m, t) $N(0, \sigma_\xi^2)$.

The number of potential entrants, N_{mt}, is assumed to be proportional to population: $N_{mt} = \lambda\, POP_{mt}$, where the parameter λ is industry specific.

Question 4. [5 points] Consider the following estimator of the number of potential entrants:

$$\hat{N}_{mt} = \text{integer} \left\{ \max_{\text{over all}\{m',t'\}} \left[\frac{entrants_{m't'} + incumbents_{m't'}}{POP_{m't'}} \right] POP_{mt} \right\}$$

where $entrants_{m't'}$ and $incumbents_{m't'}$ are the number of new entrants and the number of incumbents, respectively, in market m' at period t'. Show that $\hat{N}_{mt}$ is a consistent estimator of $N_{mt} = \lambda\, POP_{mt}$.

Question 5. [15 points] Let $P(X_{mt}, \varepsilon_{mt})$ be the equilibrium probability of entry given the common knowledge variables $(X_{mt}, \varepsilon_{mt})$. And let $G(n|X_{mt}, \varepsilon_{mt})$ be the distribution of the number of active firms in equilibrium conditional on $(X_{mt}, \varepsilon_{mt})$ and given that one of the firms is active with probability one. (i) Obtain the expression of the probability distribution $G(n|X_{mt}, \varepsilon_{mt})$ in terms of the probability of entry $P(X_{mt}, \varepsilon_{mt})$. (ii) Derive the expression for the expected profit of an active firm in terms of the probability of entry. (iii) Obtain the expression of the equilibrium mapping that defines implicitly the equilibrium probability of entry $P(X_{mt}, \varepsilon_{mt})$.

NOTE: *For Questions 6 and 7, consider the following approximation to the function* $\ln E(v(n_{mt})$ $| X_{mt}, \varepsilon_{mt}, 1 \text{ sure})$:

$$\ln E(v(n_{mt})|X_{mt}, \varepsilon_{mt}, 1\text{sure}) \simeq \ln v(1) + \sum_{n=1}^{N_{mt}} G(n|X_{mt}, \varepsilon_{mt}) \left[\frac{v(n) - v(1)}{v(1)} \right]$$

This is a first order Taylor approximation to $\ln E(v(n_{mt})|X_{mt}, \varepsilon_{mt}, 1\text{sure})$ *around the values* $v(1) = v(2) = ... = v(N)$*, i.e., no competition effects. The main advantage of using this approximation for estimation is that it is linear in the parameters* $\left[\frac{v(n)-v(1)}{v(1)} \right]$*.*

Question 6. [20 points] Suppose that $\varepsilon_{mt} \equiv \varepsilon_{mt}^S - \varepsilon_{mt}^F$ is just an aggregate time effect, $\varepsilon_{mt} = \gamma_t$. Use a two-step pseudo maximum likelihood method to estimate the vector of parameters:

$$\theta \equiv \left\{ \frac{1}{\sigma_\xi},\ \frac{\beta_1^S}{\sigma_\xi},\ \frac{\ln v(1) + \beta_0^S - \beta^F}{\sigma_\xi},\ \frac{v(n) - v(1)}{\sigma_\xi\, v(1)} : n = 2, 3, ... \right\}$$

for each of the five industries separately. Given these estimates, obtain estimates of the parameters $\dfrac{v(n+1)}{v(n)}$ for $n \in \{1, 2, ..., n_{\max}\}$. Present a figure of the estimated function $\dfrac{v(n+1)}{v(n)}$ for each of the five industries. Interpret the results. Based on these results, what

can we say about *the nature of competition* in each of these industries? Compare these results to those from the estimation of the *BR-91* models in Questions 2 and 3.

Question 7. **[10 points]** Repeat the same exercise as in Question 7 but using the following specification of the unobservable ε_{mt}:

$$\varepsilon_{mt} = \gamma_t + \delta_m$$

where γ_t are time effects that can be captured by using time-dummies; and δ_m are fixed market effects that can be captured by using market-dummies. Comment the results. Compare these results to those in Questions 2, 3, and 6.

3. Problem set #3

This problem set describes a dynamic game of entry/exit in an oligopoly market. To answer the questions below, you have to write computer code (e.g., GAUSS, MATLAB) for the solution, simulation and estimation of the model. Please, submit the program code together with your answers.

Consider the UK fast food industry during the period 1991-1995, as analyzed by Toivanen and Waterson (RAND, 2005). During this period, the industry was dominated by two large retail chains: McDonalds (MD) and Burger King (BK). The industry can be divided into isolated/independent local markets. Toivanen and Waterson consider local districts as the definition of local market (of which there are almost 500 in UK). At each local market these retail chains decide whether to have an outlet or not.

We index firms by $i \in \{MD, BK\}$ and time (years) by t. The current profit of firm i in a local market is equal to variable profits, VP_{it}, minus fixed costs of operating an outlet, FC_{it}, and minus the entry cost of setting up an outlet by first time, EC_{it}. Variable profits are $VP_{it} = (p_{it} - c_i)q_{it}$, where p_{it} represents the price, c_i is firm i's marginal cost (i.e., the marginal cost of an average meal in chain i), and q_{it} is the quantity sold (i.e., total number of meals served in the outlet at year t). The demand for an outlet of firm i in the local market is:

$$q_{it} = \frac{S_t \, \exp\left\{w_i - \alpha \, p_{it}\right\}}{1 + \exp\left\{w_i - \alpha \, p_{it}\right\} + a_{jt} \, \exp\left\{w_j - \alpha \, p_{jt}\right\}}$$

S_t represents the size of the local market at period t (i.e., total number of restaurant meals over the year). w_i and w_j are the average willingness to pay for products i and j, respectively. α is a parameter. And a_{jt} is the indicator of the event "firm j is active in the local market at period t". Every period t, the active firms compete in prices. There is not dynamics in consumers demand or in variable costs, and therefore price competition is static. Fixed costs

and entry costs have the following form:

$$FC_{it} = FC_i + \varepsilon_{it}$$

$$EC_{it} = (1 - a_{i,t-1})\ EC_i$$

The fixed cost is paid every year that the firm is active in the market. The entry cost, or setup cost, is paid only if the firm was not active at previous year (if $a_{i,t-1} = 0$). Both fixed costs and entry costs are firm-specific. The entry cost is time invariant. ε_{it} represents a firm-idiosyncratic shock in firm i's fixed cost that is iid over firms and over time with a distribution $N(0, \sigma^2)$. We also assume that ε_{it} is private information of firm i. If a firm is not active in the market, its profit is zero. For notational simplicity I "normalize" the variance of ε_{it} to be 1, though it should be understood that the structural parameters in the profit function are identified up to scale.

———

QUESTION 1. [5 POINTS] Consider the static model of price competition. Show that equilibrium price-cost margins, $p_{it} - c_i$, and equilibrium market shares, q_{it}/S_t, do not depend on market size S_t. Therefore, we can write the equilibrium variable profit function as:

$$VP_{it} = (1 - a_{jt})\ S_t\ \theta_i^M + a_{jt}\ S_t\ \theta_i^D$$

where θ_i^M and θ_i^D represent the equilibrium variable profits per-capita (per-meal) when firm i is a monopolist and when it is a duopolist, respectively.

———

The payoff-relevant information of firm i at period t is $\{x_t, \varepsilon_{it}\}$ where $x_t \equiv \{S_t, a_{1,t-1}, a_{2,t-1}\}$. Let $P_j(x_t)$ represents firm $i's$ belief about the probability that firm j will be active in the market given state x_t. Given this belief, the expected profit of firm i at period t is:

$$\pi_{it}^P = (1 - P_j(x_t))\ S_t\ \theta_i^M + P_j(x_t)\ S_t\ \theta_i^D - FC_i - (1 - a_{i,t-1})\ EC_i - \varepsilon_{it}$$

$$= Z_{it}^P\ \theta_i - \varepsilon_{it}$$

where $Z_{it}^P \equiv ((1 - P_j(x_t))\ S_t, P_j(x_t)S_t, -1, -(1 - a_{i,t-1}))$ and $\theta_i \equiv \left(\theta_i^M, \theta_i^D, FC_i, EC_i\right)'$.

For the rest of this problem set, we consider the following values for the profit parameters:

$$\theta_{MD}^M = 1.5\ \ ;\ \ \theta_{MD}^D = 0.7\ \ ;\ \ FC_{MD} = 6\ \ ;\ \ EC_{MD} = 6$$

$$\theta_{BK}^M = 1.2\ \ ;\ \ \theta_{BK}^D = 0.3\ \ ;\ \ FC_{BK} = 4\ \ ;\ \ EC_{BK} = 4$$

MD's product has higher quality (even after adjusting for marginal costs) than BK's. This implies that MD has higher variable profits than BK, either under monopoly or under duopoly. However, MD has also higher costs of setting up and operating an outlet.

Market size S_t follows a discrete Markov process with support $\{4, 5, 6, 7, 8, 9\}$ and transition probability matrix:

$$F_S = \begin{bmatrix} 0.9 & 0.1 & 0.0 & 0.0 & 0.0 & 0.0 \\ 0.1 & 0.8 & 0.1 & 0.0 & 0.0 & 0.0 \\ 0.0 & 0.1 & 0.8 & 0.1 & 0.0 & 0.0 \\ 0.0 & 0.0 & 0.1 & 0.8 & 0.1 & 0.0 \\ 0.0 & 0.0 & 0.0 & 0.1 & 0.8 & 0.1 \\ 0.0 & 0.0 & 0.0 & 0.0 & 0.1 & 0.9 \end{bmatrix}$$

A.   STATIC (MYOPIC) ENTRY-EXIT GAME

We first consider a static (not forward-looking) version of the entry-exit game. A Bayesian Nash Equilibrium (BNE) in this game can be described as a pair of probabilities, $\{P_{MD}(x_t), P_{BK}(x_t)\}$ solving the following system of equations:

$$P_{MD}(x_t) = \Phi\left(Z_{MDt}^P \, \theta_{MD}\right)$$

$$P_{BK}(x_t) = \Phi\left(Z_{BKt}^P \, \theta_{BK}\right)$$

where $\Phi(.)$ is the CDF of the standard normal.

QUESTION 2. [10 POINTS] For every possible value of the state x_t (i.e., 24 values) obtain all the BNE of the static entry game.

Hint: Define the functions $f_{MD}(P) \equiv \Phi\left(Z_{MDt}^P \, \theta_{MD}\right)$ and $f_{BK}(P) \equiv \Phi\left(Z_{BKt}^P \, \theta_{BK}\right)$. Define also the function $g(P) \equiv P - f_{MD}(f_{BK}(P))$. A BNE is zero of the function $g(P)$. You can search for all the zeroes of $g(P)$ in different ways, but in this case the simpler method is to consider a discrete grid for P in the interval $[0, 1]$, e.g., uniform grid with 101 points.

For some values of the state vector x_t, the static model has multiple equilibria. To answer Questions 3 to 5, assume that, in the population under study, the "equilibrium selection mechanism" always selects the equilibrium with the higher probability that MD is active in the market.

Let X be the set of possible values of x_t. And let $\mathbf{P}^0 \equiv \{P_{MD}^0(x), P_{BK}^0(x) : x \in X\}$ be the equilibrium probabilities in the population. Given $\mathbf{P}^0$ and the transition probability matrix for market size, F_S. We can obtain the steady-state distribution of x_t. Let $p^*(x_t)$ be

the steady-state distribution. By definition, for any $x_{t+1} \in X$:

$$
\begin{aligned}
p^*(x_{t+1}) &= \sum_{x_t \in X} p^*(x_t) \ \Pr\left(x_{t+1}|x_t\right) \\
&= \sum_{x_t \in X} p^*(x_t) \ F_S(S_{t+1}|S_t) \\
&\quad [P^0_{MD}(x_t)]^{a_{MDt+1}} [1 - P^0_{MD}(x_t)]^{1-a_{MDt+1}} [P^0_{BK}(x_t)]^{a_{BKt+1}} [1 - P^0_{BK}(x_t)]^{1-a_{BKt+1}}
\end{aligned}
$$

QUESTION 3. [10 POINTS] Compute the steady-state distribution of x_t in the population.

QUESTION 4. [20 POINTS] Using the values of P^0, F_S and p^* obtained above, simulate a data set $\{x_{mt} : t = 0, 1, ..., T; m = 1, 2, ..., M\}$ for $M = 500$ local markets and $T+1 = 6$ years with the following features: (1) local markets are independent; and (2) the initial states x_{m0} are random draws from the steady-state distribution p^*. Present a table with the mean values of the state variables in x_t and with the sample frequencies for the following events: (1) MD is a monopolist; (2) BK is a monopolist; (3) duopoly; (4) MD is active given that (conditional) he was a monopolist at the beginning of the year (the same for BK); (5) MD is active given that BK was a monopolist at the beginning of the year (the same for BK); (6) MD is active given that there was a duopoly at the beginning of the year (the same for BK); and (7) MD is active given that there were no firms active at the beginning of the year (the same for BK).

QUESTION 5. [20 POINTS] Use the simulated data in Question 4 to estimate the structural parameters of the model. Implement the following estimators: (1) two-step PML using a frequency estimator of $\mathbf{P}^0$ in the first step; (2) two-step PML using random draws from a U(0,1) for $\mathbf{P}^0$ in the first step; (3) 20-step PML using a frequency estimator of $\mathbf{P}^0$ in the first step; (4) 20-step PML using random draws from a U(0,1) for P^0 in the first step; and (5) NPL estimator based on 10 NPL fixed points (i.e., 10 different initial $P's$). Comment the results.

QUESTION 6. [30 POINTS] Suppose that the researcher knows that local markets are heterogeneous in their market size, but he does not observed market

size S_{mt}. Suppose that the researcher assumes that market size is constant over time but it varies across markets, and it has a uniform distribution with discrete support $\{4, 5, 6, 7, 8, 9\}$. Obtain the NPL estimator under this assumption (use 20 NPL fixed points). Comment the results.

_____

_____

QUESTION 7. [30 POINTS] Use the previous model (both the true model and the model estimated in Question 5) to evaluate the effects of a value added tax. The value added tax is paid by the retailer and it is such that the parameters θ_i^M and θ_i^D are reduced by 10%. Obtain the effects of this tax on average firms' profits, and on the probability distribution of market structure.

_____

B.   DYNAMIC ENTRY-EXIT GAME

Now, consider the dynamic (forward-looking) version of the entry-exit game. A Markov Perfect Equilibrium (MPE) in this game can be described as a vector of probabilities $\mathbf{P} \equiv \{P_i(x_t) : i \in \{MD, BK\}, x_t \in X\}$ such that, for every (i, x_t):

$$P_i(x_t) = \Phi\left(\tilde{Z}_{it}^P \, \theta_{MD} + \tilde{e}_{it}^P\right)$$

where $\tilde{Z}_{it}^P$ and $\tilde{e}_{it}^P$ are defined in the class notes.

_____

QUESTION 8. [20 POINTS] Obtain the MPE that we obtain when we iterate in the equilibrium mapping starting with an initial $\mathbf{P} = 0$. Find other MPEs.

_____

_____

QUESTION 9. [10 POINTS] Compute the steady-state distribution of x_t in the population.

_____

_____

QUESTION 10. [20 POINTS] The same as in Question 4 but using the dynamic game and the MPE in Question 8.

_____

_____

QUESTION 11. [20 POINTS] The same as in Question 5 but using the dynamic game and the MPE in Question 8.

———

———

QUESTION 12. [30 POINTS] The same as in Question 6 but using the dynamic game and the MPE in Question 8.

———

———

QUESTION 13. [30 POINTS] The same as in Question 7 but using the dynamic game and the MPE in Question 8.

———

4. Problem set #4

QUESTION 1 (25 POINTS): This question deals with the paper by Hendel and Nevo (*Econometrica,* 2006).

>   **(a)** Explain the implications on estimated elasticities and market power of ignoring (when present) consumer forward-looking behavior and dynamics in the demand of differentiated storable products. Discuss how the biases depend on the stochastic process of prices (e.g., Hi-Lo pricing versus a more stable price).

>   **(b)** Describe the main issues in the estimation of Hendel-Nevo model. Discuss the assumptions that they make to deal with these issues.

QUESTION 2 (25 POINTS): The geographic definition of a local market is an important modelling decision in empirical models of market entry.

>   **(a)** Explain the implications on the empirical predictions of these model of using a definition of local that is too broad or too narrow.

>   **(b)** Explain the approach in Seim (2006). Discuss its advantages and limitations.

QUESTION 3 (50 POINTS): There is a significant number of empirical applications of static and dynamic models of entry in local markets which find the following empirical regularity: after conditioning on observable market characteristics (e.g., population, income, age) there is a positive correlation between the entry decisions of potential entrants. Three main hypotheses have been proposed to explain this evidence: (1) spillover effects in consumer

traffic; (2) information externalities (see Caplin and Leahy [*Economic Journal,* 1998] and Toivanen and Waterson [*RAND*, 2005]); and (3) market characteristics which are observable for the firms but unobservable to the researcher.

(a) Explain how these hypotheses can explain the empirical evidence.

(b) Discuss why it is important to distinguish between these hypothesis. Do they have different policy implications?

(c) Consider the data and the empirical application in Toivanen and Waterson (*RAND*, 2005). Explain how it is possible to identify empirically the contribution of the three hypotheses.

(d) Consider the dynamic game of entry-exit in the Problem Set of this course. Explain how to extend this model to incorporate information externalities as in Caplin and Leahy (1998). Discuss identification issues.

5. Problem set #5

Consider a market with N firms who can potentially operate in it. We index firms by $i \in \{1, 2, ..., N\}$. Firms produce and sell a differentiated product. There are S consumers and each consumer buys at most one unit (per period) of this differentiated product. A consumer (indirect) utility of buying firm $i's$ product is:

$$U_i = w_i - p_i + \varepsilon_i$$

w_i is the "quality" of product i which is valued in the same way by every consumer. p_i is the price. And $\{\varepsilon_1, \varepsilon_2, ..., \varepsilon_N\}$ are consumer specific preferences which are i.i.d. with a type 1 extreme value distribution with dispersion parameter α. The utility of not buying any of these products is normalized to zero. For simplicity, we consider that there are only two levels of quality, high and low: $w_i \in \{w_L, w_H\}$, with $w_L < w_H$. Firms choose endogenously their qualities and prices. They also decide whether to operate in the market or not. Let n_L and n_H be the number of active firms with low and high quality products, respectively. Then, the demand for an active firm with quality w_i and price p_i is:

$$q_i = \frac{S \, \exp\left\{\dfrac{w_i - p_i}{\alpha}\right\}}{1 + n_L \, \exp\left\{\dfrac{w_L - p_L}{\alpha}\right\} + n_H \, \exp\left\{\dfrac{w_H - p_H}{\alpha}\right\}}$$

where we have imposed the (symmetric) equilibrium restriction that firms with the same quality charge the same price. Inactive firms get zero profit. The profit of an active firm is:

$$\Pi_i = (p_i - c(w_i)) \, q_i - F(w_i)$$

where $c(w_i)$ and $F(w_i)$ are the (constant) marginal cost and the fixed cost of producing a product with quality w_i. Each firm decides: (1) whether or not to operate in the market; (2) the quality of its product; and (3) its price. The game that firms play is a sequential game with the following two steps:

> Step 1: Firms make entry and quality decisions. This determines n_L and n_H.
> Step 2: Given (n_L, n_H), firms compete in prices a la Bertrand.

We start describing the Bertrand equilibrium at step 2 of the game.

QUESTION 1. [10 POINTS] Show that the best response functions of the Bertrand game in step 2 have the following form.

$$
p_L = c_L + \alpha \left[1 - \frac{\exp\left\{ \dfrac{w_L - p_L}{\alpha} \right\}}{1 + n_L \, \exp\left\{ \dfrac{w_L - p_L}{\alpha} \right\} + n_H \, \exp\left\{ \dfrac{w_H - p_H}{\alpha} \right\}} \right]^{-1}
$$

$$
p_H = c_H + \alpha \left[1 - \frac{\exp\left\{ \dfrac{w_H - p_H}{\alpha} \right\}}{1 + n_L \, \exp\left\{ \dfrac{w_L - p_L}{\alpha} \right\} + n_H \, \exp\left\{ \dfrac{w_H - p_H}{\alpha} \right\}} \right]^{-1}
$$

ANSWER:

Note that equilibrium prices depend on (n_L, n_H).

QUESTION 2. [30 POINTS] Write a computer program that computes equilibrium prices in this Bertrand game. For given values of the structural parameters (e.g., $\alpha = 1$, $w_L = 2$, $w_H = 4$, $c_L = 1$, $c_H = 2$) calculate equilibrium prices for every possible combination of (n_L, n_H) given that $N = 4$. Present the results in a table.

n_L	n_H	p_L	p_H
1	0	?	?
0	1	?	?
1	1	?	?
2	0	?	?
...	...	...	...

Now, consider the game at step 1. It is useful to define the indirect variable profit function that results from the Bertrand equilibrium in step 2 of the game. Let $\Pi_L(n_L, n_H)$ and

$\Pi_H(n_L, n_H)$ be this indirect variable profit, i.e., $\Pi_L(n_L, n_H) = (p_L - c_L)q_L$ and $\Pi_H(n_L, n_H) = (p_H - c_H)q_H$, where prices and quantities are equilibrium ones.

QUESTION 3. [10 POINTS] Show that: $\Pi_L(n_L, n_H) = \alpha S q_L/(S - q_L)$ and $\Pi_H(n_L, n_H) = \alpha S q_H/(S - q_H)$.

Let $n_{L(-i)}$ and $n_{H(-i)}$ be the number of low and high quality firms, respectively, excluding firm i. Let's use $w_i = \varnothing$ to represent no entry. And let $b(n_{L(-i)}, n_{H(-i)})$ be the best response mapping of a firm at step 1 of the game.

QUESTION 4. [10 POINTS] Show that the best response function $b(n_{L(-i)}, n_{H(-i)})$
can be described as follows:

$$b(n_{L(-i)}, n_{H(-i)}) = \begin{cases} \varnothing & if & \left[\begin{array}{c} \{\Pi_L(n_{L(-i)} + 1, n_{H(-i)}) - F_L < 0\} \\ and \ \{\Pi_H(n_{L(-i)}, n_{H(-i)} + 1) - F_H < 0\} \end{array} \right] \\[3ex] w_L & if & \left[\begin{array}{c} \{\Pi_L(n_{L(-i)} + 1, n_{H(-i)}) - F_L \geq 0\} \\ and \ \{\Pi_L(n_{L(-i)} + 1, n_{H(-i)}) - F_L > \Pi_H(n_{L(-i)}, n_{H(-i)} + 1) - F_H\} \end{array} \right] \\[3ex] w_H & if & \left[\begin{array}{c} \{\Pi_H(n_{L(-i)}, n_{H(-i)} + 1) - F_H \geq 0\} \\ and \ \{\Pi_L(n_{L(-i)} + 1, n_{H(-i)}) - F_L \leq \Pi_H(n_{L(-i)}, n_{H(-i)} + 1) - F_H\} \end{array} \right] \end{cases}$$

Now, suppose that a component of the fixed cost is private information of the firm: i.e., $F_i(w_L) = F_L + \xi_{iL}$ and $F_i(w_H) = F_H + \xi_{iH}$, where F_L and F_H are parameters and ξ_{iL} and ξ_{iH} are private information variables which are iid extreme value distributed across firms. In this Bayesian game a firm's strategy is a function of his own private information $\xi_i \equiv (\xi_{iL}, \xi_{iH})$ and of the common knowledge variables (i.e., parameters of the model and market size S). Let $\omega(\xi_i, S)$ be a firm's strategy function. A firm's strategy can be also described in terms of two probabilities: $P_L(S)$ and $P_H(S)$, such that:

$$P_L(S) \equiv \int I\{\omega(\xi_i, S) = w_L\} \, dF_\xi(\xi_i)$$

$$P_H(S) \equiv \int I\{\omega(\xi_i, S) = w_H\} \, dF_\xi(\xi_i)$$

where $I\{.\}$ is the indicator function and $F_\xi(.)$ is the CDF of ξ_i.

QUESTION 5. [20 POINTS] Show that a Bayesian Nash Equilibrium (BNE) in this game is a pair (P_L, P_H) that is a solution to the following fixed problem:

$$P_L = \frac{\exp\left\{\Pi_L^e(P_L, P_H) - F_L\right\}}{1 + \exp\left\{\Pi_L^e(P_L, P_H) - F_L\right\} + \exp\left\{\Pi_H^e(P_L, P_H) - F_H\right\}}$$

$$p_H = \frac{\exp\left\{\Pi_H^e(P_L, P_H) - F_H\right\}}{1 + \exp\left\{\Pi_L^e(P_L, P_H) - F_L\right\} + \exp\left\{\Pi_H^e(P_L, P_H) - F_H\right\}}$$

with:

$$\Pi_L^e(P_L, P_H) = \sum_{n_{L(-i)}, n_{H(-i)}} \Pi_L\big(n_{L(-i)} + 1, n_{H(-i)}\big)\, T(n_{L(-i)}, n_{H(-i)}|N - 1, P_L, P_H)$$

$$\Pi_H^e(P_L, P_H) = \sum_{n_{L(-i)}, n_{H(-i)}} \Pi_H\big(n_{L(-i)}, n_{H(-i)} + 1\big)\, T(n_{L(-i)}, n_{H(-i)}|N - 1, P_L, P_H)$$

where $T(x, y|n, p_1, p_2)$ is the **PDF of a trinomial distribution with parameters** (n, p_1, p_2).

QUESTION 6. [50 POINTS] Write a computer program that computes the BNE in this entry/quality game. Consider $N = 4$. For given values of the structural parameters, calculate the equilibrium probabilities $(P_L(S), P_H(S))$ for a grid of points for market size S. Present a graph for $(P_L(S), P_H(S))$ (on the vertical axis) on S (in the horizontal axis). Does the proportion of high quality firms depend on market size?

QUESTION 7. [30 POINTS] Define the function $\lambda(S) \equiv P_H(S)/P_L(S)$ that represents the average ratio between high and low quality firms in the market. Repeat the same exercise as in Question 1.6. but for three different values of the ratio F_H/F_L. Present a graph of $\lambda(S)$ on S for the three values of F_H/F_L. Comment the results.

QUESTION 8. [50 POINTS] A regulator is considering a policy to encourage the production of high quality products. The policy would provide a subsidy of 20% of the additional fixed cost of producing a high quality product. That is, the new fixed cost of producing a high quality product would be $F_H^* = F_H - 0.20 * (F_H - F_L)$. Given a parametrization of the model, obtain the equilibrium before and after the policy and calculate the effect of the policy on: (1) prices; (2) quantities; (3) firms' profits; (4) average consumers' surplus; and (5) total surplus.

Suppose that the researcher observes a random sample of M isolated markets, indexed by m, where these N firms compete. More specifically, the researcher observes:

$$Data = \{S_m,\ n_{Hm},\ n_{Lm},\ q_{Hm},\ q_{Lm},\ p_{Hm},\ p_{Lm} : m = 1, 2, ..., M\}$$

For instance, consider data from the hotel industry in a region where "high quality" is defined as four stars or more (low quality as three stars or less). We incorporate two sources of market heterogeneity in the econometric model (i.e., unobservables for the researcher).

(A) Consumers' average valuations: $w_{Lm} = w_L + \eta_m^{(wL)}$ and $w_{Hm} = w_H + \eta_m^{(wH)}$, where $\eta_m^{(wL)}$ and $\eta_m^{(wH)}$ are zero mean random variables.

(B) Marginal costs: $c_{Lm} = c_L + \eta_m^{(cL)}$ and $c_{Hm} = c_H + \eta_m^{(cH)}$, where $\eta_m^{(cL)}$ and $\eta_m^{(cH)}$ are zero mean random variables.

We assume that the vector of unobservables $\eta_m \equiv \{\eta_m^{(wL)}, \eta_m^{(wH)}, \eta_m^{(cL)}, \eta_m^{(cH)}\}$ is iid over markets and independent of market size S_m. We also assume that these variables are common knowledge. We want to use these data to estimate the structural parameters $\theta = \{\alpha, w_j, c_j, F_j : j = L, H\}$.

_____

QUESTION 9. [30 POINTS] Show that the econometric model can be described in terms of three sets of equations.

(1) Demand equations: For $j \in \{L, H\}$ let s_{jm} be the market share q_{jm}/S_m. Then:

$$\ln\left(\frac{s_{jm}}{1 - s_{Lm} - s_{Hm}}\right) = \frac{w_j}{\alpha} - \frac{1}{\alpha} p_{jm} + \frac{\eta_m^{(wj)}}{\alpha} \qquad \text{if } n_{jm} > 0$$

(2) Price equations: For $j \in \{L, H\}$:

$$p_{jm} = c_j + \alpha\left(\frac{1}{1 - s_{jm}}\right) + \eta_m^{(cj)} \qquad \text{if } n_{jm} > 0$$

(3) Entry/Quality choice: Suppose that from the estimation of (1) and (2) we can obtain consistent estimates of η_m as residuals. After that estimation, we can treat η_m as "observable" (though we should account for estimation error). Then,

$$\Pr(n_{Lm}, n_{Hm}|S_m, \eta_m) = T(n_{Lm}, n_{Hm}|N, P_L(S_m, \eta_m), P_H(S_m, \eta_m))$$

where $P_L(S_m, \eta_m), P_H(S_m, \eta_m)$ are equilibrium probabilities in market m.

_____

QUESTION 10. [30 POINTS] Discuss in detail the econometric issues in the estimation of the parameters $\{w_L, w_H, \alpha\}$ from the demand equations: for $j \in$

$\{L, H\}$:

$$\ln\left(\frac{s_{jm}}{1 - s_{Lm} - s_{Hm}}\right) = \frac{w_j}{\alpha} - \frac{1}{\alpha}\, p_{jm} + \frac{\eta_m^{(wj)}}{\alpha} \qquad \text{if } n_{jm} > 0$$

Propose and describe in detail a method that provides consistent estimates of $\{w_L, w_H, \alpha\}$.

QUESTION 11. [30 POINTS] Suppose for the moment that α has not been estimated from the demand equations. Discuss in detail the econometric issues in the estimation of the parameters $\{c_L, c_H, \alpha\}$ from the pricing equations: for $j \in \{L, H\}$:

$$p_{jm} = c_j + \alpha\left(\frac{1}{1 - s_{jm}}\right) + \eta_m^{(cj)} \qquad \text{if } n_{jm} > 0$$

Propose and describe in detail a method that provides consistent estimates of $\{c_L, c_H, \alpha\}$. What if α has been estimated in a first step from the demand equations? Which are the advantages of a joint estimation of demand and supply equations?

QUESTION 12. [50 POINTS] For simplicity, suppose that the parameters $\{w_L, w_H, c_L, c_H, \alpha\}$ are known and that η_m is observable (i.e., we ignore estimation error from the first step estimation). We want to estimate the fixed costs F_L and F_H using information on firms' entry/quality choices. Discuss in detail the econometric issues in the estimation of these parameters. Propose and describe in detail a method that provides consistent estimates of $\{F_L, F_H\}$.

QUESTION 13. [50 POINTS] Suppose that you incorporate a third source of market heterogeneity in the model:

> **(C) Fixed costs: $F_{Lm} = F_L + \eta_m^{(FL)}$ and $F_{Hm} = F_H + \eta_m^{(FH)}$, where $\eta_m^{(FL)}$ and $\eta_m^{(FH)}$ are zero mean random variables, and they are common knowledge to the players.**

Explain which are the additional econometric issues in the estimation of $\{F_L, F_H\}$ when we have these additional unobservables. Propose and describe in detail a method that provides consistent estimates of $\{F_L, F_H\}$ and the distribution of $\{\eta_m^{(FL)}, \eta_m^{(FH)}\}$.

QUESTION 14. [50 POINTS] Consider the econometric model without $\{\eta_m^{(FL)}, \eta_m^{(FH)}\}$. Suppose that S_m is log normally distributed and $\eta_m \equiv \{\eta_m^{(wL)}, \eta_m^{(wH)}, \eta_m^{(cL)}, \eta_m^{(cH)}\}$ has a normal distribution with zero means. Generate a random sample of $\{S_m, \eta_m\}$

with sample size of $M = 500$ markets. **Given a parametrization of the model, for every value** $\{S_m, \eta_m\}$ **in the sample, solve the model and obtain the endogenous variables** $\{n_{Hm}, \ n_{Lm}, \ q_{Hm}, \ q_{Lm}, \ p_{Hm}, \ p_{Lm}\}$. **Present a table with the summary statistics of these variables: e.g., mean, median, standard deviation, minimum, maximum.**

_____

QUESTION 15. [50 POINTS] Write a computer program that implements the method for the estimation of the demand that you proposed in Question 10. Apply this method to the data simulated in Question 14. Present and comment the results.

_____

QUESTION 16. [50 POINTS] Write a computer program that implements the method for the estimation of the pricing equations that you proposed in Question 11. Apply this method to the data simulated in Question 14. Present and comment the results.

_____

QUESTION 17. [100 POINTS] Write a computer program that implements the method for the estimation of the entry/quality choice game that you proposed in Question 12. Apply this method to the data simulated in Question 14. Present and comment the results.

_____

QUESTION 18. [50 POINTS] Use the estimated model to evaluate the policy in question 8. Present a table that compares the average (across markets) "actual" and estimated effects of the policy on: (1) prices; (2) quantities; (3) firms' profits; (4) average consumers' surplus; and (5) total surplus.

6. Problem set #6

In the paper _"The Interpretation of Instrumental Variables Estimators in Simultaneous Equations Models with an Application to the Demand for Fish,"_ (REStud, 2000), Angrist, Graddy and Imbens consider the following random coefficients model of supply and demand for an homogeneous product:

$$\text{Inverse Demand:} \quad p_t \ = \ x_t \, \beta^D - \left(\alpha^D + \upsilon_t^D\right) \ q_t + \varepsilon_t^D$$

$$\text{Inverse Supply:} \quad p_t \ = \ x_t \, \beta^S + \left(\alpha^S + \upsilon_t^S\right) \ q_t + \varepsilon_t^S$$

where p_t is logarithm of price; q_t is the logarithm of the quantity sold; and ε_t^D, ε_t^S, v_t^D and v_t^S are unobservables which have zero mean conditional on x_t. The variables v_t^D and v_t^S account for random shocks in the price elasticities of demand and supply. Suppose that the researcher has a sample $\{p_t, q_t, x_t : t = 1, 2, ..., n\}$ and is interested in the estimation of the demand parameters β^D and α^D.

(1) Explain why instrumental variables (or 2SLS) provides inconsistent estimates of the parameters β^D and α^D.

(2) Descrine an estimation method that provides consistent estimates of β^D and α^D.

(3)

7. Problem set #7

Mitsubishi entered the Canadian automobile market in September 2002. You can consider this to be an exogenous change. Subsequently, the firm had to decide in which local markets to open dealerships. This, you should consider to be endogenous choices.

(1) How could you use this type of variation to estimate a model of entry like Bresnahan & Reiss (1988, 1990, 1991)? What variation in the data will be useful to identify which underlying economic parameters? How would you learn about or control for the competitiveness of market operation?

(2) (It is not necessary to derive any equations, although you can if it helps your exposition.).

(3) Could you use the same data to estimate an entry model like Berry (1992)? How?

(4) How would you use data for this industry to estimate the lower bound on concentration in the sense of Sutton?

(5) Give an example of an economic question that you would be able to address with this type of variation over time —entry by a new firm— that the previous authors were unable to address using only cross sectional data.

(6)

8. Problem set #8

In the paper *"The valuation of new goods under perfect and imperfect competition,"* Jerry Hausman estimates a demand system for ready-to eat cereals using panel data on quantities and prices for multiple markets (cities), brands and quarters. The demand system is (Deaton-Muellbauer demand system):

$$w_{jmt} = \alpha_j^0 + \alpha_m^1 + \alpha_t^2 + \sum_{k=1}^{J} \beta_{jk} \ \ln(p_{kmt}) + \gamma_j \ \ln(x_{mt}) + \varepsilon_{jmt}$$

where: j, m and t are the product, market (city) and quarter subindexes, respectively; x_{mt} represents exogenous market characteristics such as population and average income. There are not observable cost shifters. The terms α_j^0, α_m^1 and α_t^2 represent product, market and time effects, respectively, which are captured using dummies. As instruments for prices, Hausman uses average prices is nearby markets. More specifically, the instrument for price p_{jmt} is z_{jmt} which is defined as:

$$z_{jmt} = \frac{1}{\#(R_m)} \sum_{\substack{m' \neq m \\ m' \in R_m}} p_{jm't}$$

where R_m is the set of markets nearby market m, and, $\#(R_m)$ is the number of elements in that set.

(1) Explain under which economic assumptions, on supply or price equations, these instruments are valid.

(2) Describe how Deaton-Muellbauer demand system can be used to calculate the value of a new product.

(3) Comment the limitations of this approach as a method to evaluate the effects of new product on consumers' welfare and firms' profits.

(4) Explain how the empirical literature on demand models in characteristics space deals with some of the limitations that you have mentioned in question (c).

(5)

9. Problem set #9

Consider Berry-Levinshon-Pakes (BLP) model for the demand of a differentiated product. The (indirect) utility of buying product j for consumer i is:

$$U_{ij} = (\beta_1 + \omega_{1i})x_{1j} + \ldots + (\beta_K + \omega_{Ki})x_{Kj} - \alpha\, p_j + \xi_j + \varepsilon_{ij}$$

where α, β_1, ..., and β_K are parameters; $\omega_i \equiv (\omega_{1i}, \omega_{2i}, ..., \omega_{Ki})$ is a vector of normal random variables (with zero mean); and $\varepsilon_i \equiv (\varepsilon_{i1}, \varepsilon_{i2}, ..., \varepsilon_{iJ})$ is a vector of independent extreme value random variables.

(1) Describe in detail BLP estimation method.

(2) Explain why it is important to allow for consumer heterogeneity in the marginal utility with respect to product characteristics.

(3) A key identifying assumption in BLP method is that unobserved product characteristics, ξ_j, are not correlated with observed product characteristics other than price, $(x_{1j}, x_{2j}, ..., x_{Kj})$. Comment on this assumption.

(4) Suppose that there is only one observable product characteristic, x_j, that we can interpret as a measure of product quality. Let x_j^* is the "true" quality of product j, which is unobservable to the researcher. That is, $x_j = x_j^* + e_j$ where e_j is measurement error which is assumed independent of x_j^*. According to this model, the unobservable ξ_j is equal to $-\beta e_j$. Show that the type of instrumental variables proposed by BLP can still be valid in this model with measurement error in quality.

(5)

10. Problem set #10

Consider an oligopoly industry in which competition takes place at the level of local markets. For concreteness, suppose that there are only two firms in the industry: firm 1 and firm 2. There are M local markets, where M is a large number. Consider the following adaptation to this industry of the simultaneous equations model in *Olley and Pakes (Econometrica, 1996)*.

Production Function: $y_{imt} = \alpha_{Li}\, l_{imt} + \alpha_{Ki}\, k_{imt} + \omega_{imt} + e_{imt}$

Investment Function: $i_{imt} = f_i\left(k_{1mt}, k_{2mt}, \omega_{1mt}, \omega_{2mt}, r_{mt}\right)$

Stay-in-the-market decision: $\chi_{imt} = I\{\omega_{imt} \geq \omega_i^*\left(k_{1mt}, k_{2mt}, r_{mt}\right)\}$

where: i is the firm subindexl; m is the local-market subindex; t is the time subindex; r_{mt} represents input prices in market m at period t; and all the other variables and parameters have the same interpretation as in Olley-Pakes. Following Olley-Pakes we assume that labor is a perfectly flexible input and that new investment is not productivity until next period (i.e., time-to-build). We are interested in the estimation of the production function parameters $\{\alpha_{L1}, \alpha_{K1}, \alpha_{L2}, \alpha_{K2}\}$.

(1) Explain why a direct application of Olley-Pakes method to this model will not provide consistent estimates of the parameters of interest.

(2) Describe how Olley-Pakes method can be adapted/extended to this industry and data to obtain a consistent estimator of $\{\alpha_{L1}, \alpha_{K1}, \alpha_{L2}, \alpha_{K2}\}$.

(3) Suppose that the average productivity of labor is larger in markets where both firms are active (relative to markets where only one of the two firms is active). Mention different hypotheses that might explain this evidence. Explain how one can use the estimated model to measure the contribution of each of these hypothesis to the observed differential in the average productivity of labor.

(4)

11. Problem set #11

Consider the following description of a hotel industry. There are N firms/hotel chains in the industry. These firms compete in independent local markets (cities). We index hotel chains by $i \in \{1, 2, ..., N\}$ and local markets by $m \in \{1, 2, ..., N\}$. The product that hotels sell is vertically differentiated. For simplicity, we consider that there are only two levels of quality, high (H) and low (L). At each local market, each firm decides whether or not to operate in the market, the quality of its product, and its price. The game that hotel chains play is a sequential game with the following two steps. Step 1: firms make entry and quality decisions. This step determines the number of low and high quality hotels in the market: n_m^L and n_m^H respectively. Step 2: Given (n_m^L, n_m^H), firms compete in prices a la Bertrand. Associated to the Bertrand equilibrium we can define the (indirect) variable profit functions $V_L(n_m^L, n_m^H, S_m)$ and $V_H(n_m^L, n_m^H, S_m)$: i.e., $V_L(n_m^L, n_m^H, S_m)$ $(V_H(n_m^L, n_m^H, S_m))$ is the variable profit of a low (high) quality hotel in a market with size S_m, with n_m^L low quality hotels and with n_m^H high quality hotels. Total operating costs are: $\Pi_{Lim} = V_L(n_m^L, n_m^H, S_m) - F_L - \varepsilon_{Lim}$ and $\Pi_{Him} = V_H(n_m^L, n_m^H, S_m) - F_H - \varepsilon_{Him}$, where F_L and F_H are the fixed costs for low and high quality firms, respectively, and ε_{Lim} and ε_{Him} are private information shocks which are iid extreme value distributed across firms and markets. A firm's strategy can be described in terms of two probabilities: the probability of being active with low quality, P_L, and the probability of being active and high quality, P_H.

(1) Show that a Bayesian Nash Equilibrium (BNE) in this game is a pair (P_L, P_H) that is a solution to the following fixed problem:

$$P_L = \frac{\exp\{V_L^e(P_L, P_H) - F_L\}}{1 + \exp\{V_L^e(P_L, P_H) - F_L\} + \exp\{V_H^e(P_L, P_H) - F_H\}}$$

$$P_H = \frac{\exp\{V_H^e(P_L, P_H) - F_H\}}{1 + \exp\{V_L^e(P_L, P_H) - F_L\} + \exp\{V_H^e(P_L, P_H) - F_H\}}$$

with:

$$V_L^e(P_L, P_H) = \sum_{n_{L(-i)}, n_{H(-i)}} V_L(n_{L(-i)} + 1, n_{H(-i)}) \, T(n_{L(-i)}, n_{H(-i)}|N - 1, P_L, P_H)$$

$$V_H^e(P_L, P_H) = \sum_{n_{L(-i)}, n_{H(-i)}} V_H(n_{L(-i)}, n_{H(-i)} + 1) \, T(n_{L(-i)}, n_{H(-i)}|N - 1, P_L, P_H)$$

where $T(x, y|n, p_1, p_2)$ is the PDF of a trinomial distribution with parameters (n, p_1, p_2).

(2) Suppose that the indirect profit functions $V_L(n_L, n_H, S)$ and $V_H(n_L, n_H, S)$ are known, i.e., they have been estimated using price an quantity data). The researcher observes the sample $\{n_{Hm}, n_{Lm}, S_m : m = 1, 2, ..., M\}$. We want to estimate the fixed costs F_L and F_H using information on firms' entry/quality choices. Discuss in

detail the econometric issues in the estimation of these parameters. Propose and describe in detail a method that provides consistent estimates of $\{F_L, F_H\}$.

(3) Suppose that you incorporate unobserved market heterogeneity in fixed costs: $F_{Lm} = F_L + \eta_m^L$ and $F_{Hm} = F_H + \eta_m^H$, where η_m^L and η_m^H are zero mean random variables, and they are common knowledge to the players. Explain which are the additional econometric issues in the estimation of $\{F_L, F_H\}$ when we have these additional unobservables. Propose and describe in detail a method that provides consistent estimates of $\{F_L, F_H\}$ and the distribution of $\{\eta_m^L, \eta_m^H\}$.

(4)

12. Problem set #12

Consider an extension of Rust's machine replacement model (Rust, 1987) that incorporates asymmetric information in the market of machines. A firm produces at several independent plants (indexed by i) that operate independently. Each plant has a machine. The cost of operation and maintenance of a machine increases with the age of the machine. Let x_{it}.be the age of the machine at plant i and at period t. There are two types of machines according to their maintenance costs: low and high maintenance costs. When the firm's manager decides to buy a machine, he does not observe its type. However, the manager learns this type just after one year of operation. The maintenance cost is: $c_i\ x_{it} + \varepsilon_{it}(0)$ where $c_i \in \{\theta_L, \theta_H\}$ is a parameter and $\varepsilon_{it}(0)$ is a component of the maintenance cost that is unobserved for the researcher. There is a cost of replacing an old machine by a new one. This replacement cost is: $RC + \varepsilon_{it}(1)$ where RC is a parameter, and $\varepsilon_{it}(1)$ is a component of the maintenance cost that is unobserved for the researcher. The firm has decide when to replace a machine in order to minimize the present value of the sume of maintenance and replacement costs. Suppose that the researcher has a random sample of machines.