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 $\Pr(v \geq P_{mt}) = 1 - G_{mt}(P_{mt})$. 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. For simplicity, we start assuming that every firm, either an incumbent or a potential entrant, has the same cost function. Let q be the amount of output produced by a single firm. The total cost of a firm active in the market is the sum of its variable cost $VC_{mt}(q)$ and its fixed cost FC_{mt}. We specify a quadratic variable cost function:

$$VC_{mt}(q) \quad = \quad \left(\mathbf{X}_{mt}^{MC}\ \boldsymbol{\gamma}_X^{MC} + \varepsilon_{mt}^{MC}\right)\ q + \frac{\gamma_2^{MC}}{2} q^2 \tag{3.4}$$

$\boldsymbol{\gamma}_X^{MC}$ 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

[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.

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:

$$\Pi_{mt}(q_i, \widetilde{Q_i}) = P_{mt}(q_i + \widetilde{Q}_i)\ q_i - VC_{mt}(q_i) - FC_{mt} \tag{3.5}$$

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:

$$P_{mt} + \frac{\partial P_{mt}(q_i + \widetilde{Q_i})}{\partial q_i}\ q_i = MC_{mt}(q_i) \tag{3.6}$$

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:

$$q_{mt}(N) = \frac{A_{mt} - \overline{MC}_{mt}}{B_{mt}\,(N+1) + \gamma_2^{MC}} \tag{3.7}$$

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 (3.7) into this equation, we get the following relationship between price-cost margin and output-per-firm in the Cournot equilibrium:

$$P_{mt} - AVC_{mt} = \frac{\left(B_{mt} + \gamma_2^{MC}/2\right)\ \left(A_{mt} - \overline{MC}_{mt}\right)}{B_{mt}\ \left(N_{mt}+1\right) + \gamma_2^{MC}} = \left(B_{mt} + \gamma_2^{MC}/2\right)\ q_{mt}(N) \tag{3.8}$$

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:

$$\begin{aligned} \Pi^*_{mt}(N) &= (P_{mt} - AVC_{mt})\ q_{mt}(N) - FC_{mt} \\ &= \left(B_{mt} + \gamma_2^{MC}/2\right)\left(\frac{A_{mt} - \overline{MC}_{mt}}{B_{mt}\,(N+1) + \gamma_2^{MC}}\right)^2 - FC_{mt} \end{aligned} \tag{3.9}$$

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

[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.

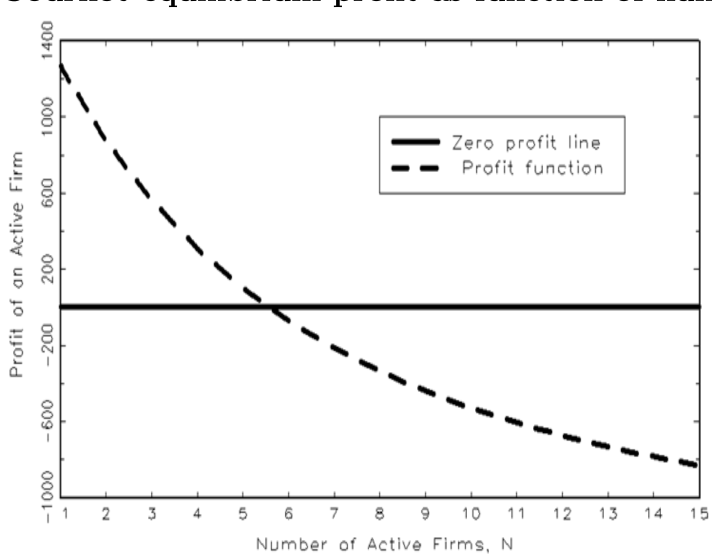
potential entrant decides to enter in the market, it gets negative profits. Additional entry implies negative profits: $\Pi^*_{mt}(N_{mt}+1) < 0$.

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 (3.9), 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:

$$N^*_{mt} \equiv -\left(1 + \frac{\gamma_2^{MC}}{B_{mt}}\right) + \left(A_{mt} - \overline{MC}_{mt}\right)\sqrt{\frac{1 + \gamma_2^{MC}/2B_{mt}}{FC_{mt}\ B_{mt}}} \tag{3.10}$$

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 3.7) and profit (equation 3.9), we can obtain the following expression for the Cournot equilibrium profit: $\Pi^*_{mt}(N) =$

[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.

$(B_{mt} + \gamma_2^{MC}/2)\ q_{mt}(N)^2 - FC_{mt}$. Therefore, the entry equilibrium condition, represented as $\Pi^*_{mt}(N^*_{mt}) = 0$, is equivalent to:

$$\left(\frac{Q_{mt}}{N_{mt}}\right)^2 = \left(\frac{N^*_{mt}}{int(N^*_{mt})}\right)^2 \frac{FC_{mt}}{B_{mt} + \gamma_2^{MC}/2} \tag{3.11}$$

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.

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, $q \equiv Q/N$,

$$\begin{array}{rrcl} \textbf{Demand equation:} & P & = & A - B\ N\ q \\ \textbf{Cournot Equilibrium Condition:} & q & = & \dfrac{A - \overline{MC}}{B\,(N+1) + \gamma_2^{MC}} \\ \textbf{Entry Equilibrium Condition:} & q^2 & = & \dfrac{FC}{B + \gamma_2^{MC}/2} \end{array} \tag{3.12}$$

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

This is a system of simultaneous equations. The system of equations in (3.12) 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 unqiue. To show this, notice that the entry equilibrium condition determines output per firm as a function of the exogenous variables.

$$q = \sqrt{\frac{FC}{B + \gamma_2^{MC}/2}} \tag{3.13}$$

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:

$$N = -\left(1 + \frac{\gamma_2^{MC}}{B}\right) + \left(A - \overline{MC}\right)\sqrt{\frac{1 + \gamma_2^{MC}/2B}{FC\ B}} \tag{3.14}$$

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

$$P \;\; = \;\; \overline{MC} + (\gamma_2^{MC} + B)\sqrt{\frac{FC}{B + \gamma_2^{MC}/2}} \tag{3.15}$$

Equations (3.13), (3.14), and (3.15) 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.

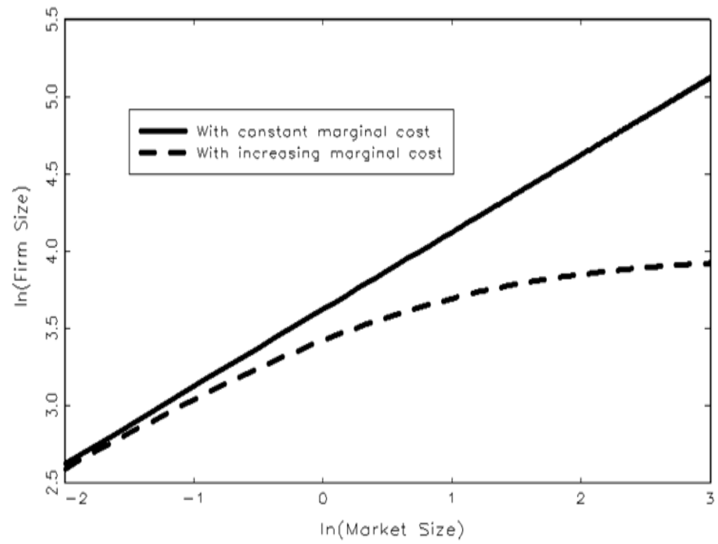
4.1. Empirical evidence from reduced form equations. From an empirical point of view, the reduced form equations establish relationships between exogenous market characteristics, such as market size, and the observable endogenous variables of the model: price, number of firms, and firm size. Can we learn 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 (3.13), implies the following relationship between firm size (or output per firm) q and market size S:

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

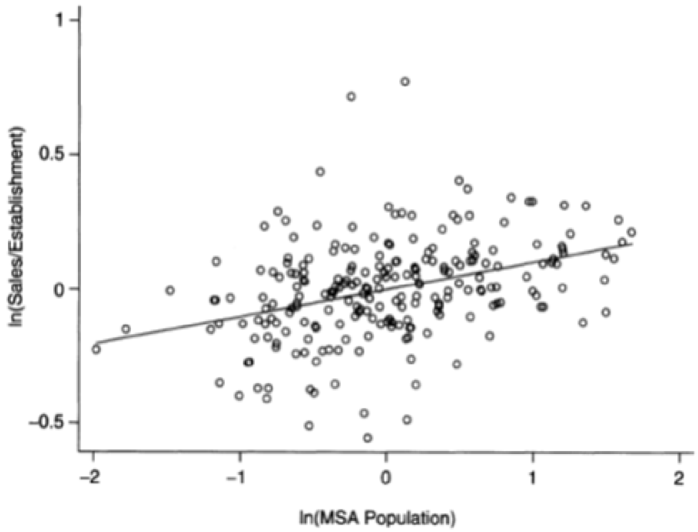
We can distinguish three different possible cases for this relationship. When fixed cost is zero ($FC = 0$) there is not any relationship between firm size and market size. The model becomes one of perfect competition and the equilibrium is characterized by a very large number of firms ($N = \infty$) 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 $\ln(q) = \frac{1}{2}[\ln(\beta_P FC) + \ln(S)]$ 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 nonliear (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).

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

$$\begin{aligned}
\frac{Q_{mt}}{S_{mt}} &= \beta_X \; \mathbf{X}_{mt}^D - \beta_1 \; P_{mt} + \varepsilon_{mt}^D \\
\left(P_{mt} - \frac{1}{\beta_1} \frac{q_{mt}}{S_{mt}} \right) &= \gamma_X^{MC} \; \mathbf{X}_{mt}^{MC} + \gamma_2^{MC} \; q_{mt} + \varepsilon_{mt}^{MC} \qquad (4.2) \\
\frac{q_{mt}^2}{S_{mt}} + \beta_1 \gamma_2^{MC} \; q_{mt} &= \gamma_X^{FC} \; \mathbf{X}_{mt}^{FC} + \varepsilon_{mt}^{FC}
\end{aligned}$$

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

[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).

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

$$\begin{aligned} \frac{Q_{mt}}{S_{mt}} &= \beta_0 + \beta_S\ S_{mt} - \beta_1\ P_{mt} + \varepsilon^D_{mt} \\ \left[P_{mt} - \frac{1}{\beta_1}\frac{q_{mt}}{S_{mt}}\right] &= \gamma_1^{MC} + \gamma_2^{MC}\ q_{mt} + \varepsilon^{MC}_{mt} \\ \left[\frac{1}{\beta_1}\frac{q^2_{mt}}{S_{mt}} + \gamma_2^{MC} q^2_{mt}\right] &= \gamma^{FC} + \varepsilon^{FC}_{mt} \end{aligned} \tag{4.3}$$

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

[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.

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:

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

and

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

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:

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

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,

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

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.

Once the model has been estimated, multiplicity of equilibria generates not trivial issues to make predictions using the estimated model. We will see different approaches to deal with this problem, and discuss their relative advantages and limitations.

4.3. Answering empirical questions: Counterfactual experiments. *** An approach to answer empirical questions using structural models in IO.

4.4. 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. Recommended 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.