1. Introduction

1.1 Empirical Industrial Organization

Industrial Organization (IO) deals with the behavior of firms in markets. We are interested in understanding how firms interact strategically, and how their actions affect market outcomes. IO economists are particularly interested in three aspects related to 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, their market shares, and the products they sell. 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 atomistic 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 above those needed to remunerate all the inputs at market prices. A *firm's strategy* is a description of the firms' actions (for instance, pricing, production and market entry decisions) contingent on the state of demand and cost conditions. We say that a firm behaves strategically if it takes into account that its actions affect other firms' profits and behavior.

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, as well as 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 a firm's profit maximizing strategy, given its beliefs about the behavior of other firms, and in determining equilibrium strategies: the set of all firms' strategies which are consistent with profit maximization and rational beliefs about each others' behavior. 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 today's economy. Many interesting markets are characterized by a relatively small number of firms who behave strategically and take into account how their decisions affect market prices and other firms' profits. 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 (for instance, 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 these types of questions.

Example 1: Estimating the demand for a new product. A company considers launching a new product, for instance, 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 (that is, 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: Evaluating the effects of a policy change. 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 a higher 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: Explain the persistence of market power. 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 anti-competitive behavior are potential 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.

1.2 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 power and a measure of market structure that captures the degree of concentration of output in a few firms (*market concentration*). In these studies, the typical measure of market power was the *Lerner Index (LI)* which is defined as price minus marginal cost divided by price, $LI \equiv (P - MC)/P$. And a common measure of market concentration is the *Herfindahl-Hirschman Index (HHI)*, defined as the sum of the squares of the market shares of the firms in the market: $HHI = \sum_{i=1}^{N} (q_i/Q)^2$, where q_i is firm i's output, and Q represents total industry output. Given a sample of N industries (indexed by n) with information on the Lerner and the Herfindahl-Hirschman indexes for each indistry, these studies related the two indexes using a linear regression model as follows,

$$LI_n = \beta_0 + \beta_1 \; HHI_n + \varepsilon_n \tag{1.1}$$

This linear regression model was 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, that is, the OLS estimate of β_1 was statistically greater than zero. One of the main purposes of these empirical studies was to identify a relationship between market concentration and market structure that could be applied to most industries. Furthermore, the estimated regression function was a causal relationship. That is, the parameter β_1 is interpreted as the increase in the Lerner Index of a one-unit increase in market concentration as measured by the HHI. This interpretation does not take into account that both market power (LI) and concentration (HHI) are endogenous variables which are jointly determined in equilibrium and affected by the same exogenous variables, and some of these variables (ε) are unobservable to the researcher.

In the 1980s, the seminal work of Bresnahan (1981, 1982, 1987), Porter, (1983), Schmalensee (1989), Sutton (1991), among others, configured the basis for the so called *New Empirical IO*. These authors pointed out 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 also vary greatly across industries. In reality, the parameters of these linear regression models are heterogeneous across industries (that is, we have β_{1n} instead of β_1) but they are estimated as constants in this previous literature. 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. Overlooking the correlation between the explanatory variable (market structure) and the error term (unobserved heterogeneity in industry fundamentals) in this regression model implies a 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* emphasize the need to study competition by looking at each industry separately using richer data at a more disaggregate level and combining these data with games 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 now 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. This leads to the question: what are the main sources of sample variability in empirical studies in modern 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 (for instance, 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 identifying and estimating 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. There are several approaches we can take. 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.

1.3 Structural models 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 provides us with useful tools for understanding competition, business strategies, and the evolution of an industry. They also help us 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, the following section presents a simple empirical model of oligopoly competition. Though simple, this model incorporates some important features related to modelling and econometric issues such as specification, endogeneity, identification, estimation, and policy experiments. We will study these issues in detail throughout this book. This example is inspired by Ryan (2012), and the model below can be seen as a simplified version of the model in that paper.

1.3.1 Empirical question

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 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* (that is, before the new policy is actually implemented) or *ex-post* (that is, after the implementation of the policy change).

1.3.2 Model

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 of 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 a 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 the spatial location of cement plants a potentially important dimension for competition.

In this simple example, we ignore spatial competition, that we will analyze in chapters 4 and 5. 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 costs of cement (per dollar value of the product) are very high. This explains why the industry is very local. Cement plants are located in proximity to the point of demand (that is, construction cites 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 become relevant for our empirical question: installed capacity is a dynamic decision that depends on the plant's capacity investments and on depreciation.

1.3.3 Data

The specification of the model depends importantly on the **data** that is available for the researcher. The level of aggregation of the data (for instance, 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 consider when she specifies the model. Model features that are important to explain firm-level data might be quite irrelevant, or they may be under-identified, 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 (for instance, 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.

$$Data = \{\ P_{mt}\ ,\ Q_{mt}\ ,\ N_{mt},\ \mathbf{X}_{mt} : m = 1,2,...,M;\ t = 1,2,...,T\ \} \tag{1.2}$$

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 form of the 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.

1.3.4 Components of the model

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

[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.

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.

1.3.5 Endogeneity and identification

A key econometric issue in the estimation of parameters in our econometric models is the endogeneity of the explanatory variables. For instance, prices and quantities that appear in a demand equation are jointly determined in the equilibrium of the model and they both depend on the exogenous variables affecting demand and costs. Some of these exogenous variables are unobservable to the researcher and are part of the error terms in our econometric models. Therefore, these error terms are correlated with some of the explanatory variables in the econometric model. For instance, the error term in the demand equation is correlated with the explanatory variable price. Ignoring this correlation can imply serious biases in the estimation of the parameters of the model and in the conclusions of the research. Dealing with this endogeneity problem is a fundamental element in EIO and in econometrics in general.

1.3.6 Demand equation

In this simple model we assume cement is a 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.

$$Q_{mt} = S_{mt} \ \left(\mathbf{X}_{mt}^{D} \ \beta_X - \beta_P \ P_{mt} + \varepsilon_{mt}^{D} \right) \tag{1.3}$$

β_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 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] A possible interpretation of this demand equation is that $\mathbf{X}_{mt}^{D} \ \beta_X - \beta_P \ P_{mt} + \varepsilon_{mt}^{D}$ is the downward sloping demand curve of a *representative consumer* in market m at period t. According to this interpretation, $\mathbf{X}_{mt}^{D}$ $\beta_X + \varepsilon_{mt}^{D}$ is the willingness to pay of this representative consumer for the first unit that she 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}. [4]

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

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

[4] 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

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

$$P_{mt} = A_{mt} - B_{mt}\, Q_{mt} \tag{1.4}$$

where the intercept A_{mt} is $\left(\mathbf{X}_{mt}^{D}\, \beta_X + \varepsilon_{mt}^{D}\right)/\beta_P$, 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).[5]

1.3.7 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, suppose that the production function has the Cobb-Douglas form $q = L^{\alpha_L}\, K^{\alpha_K}\, \exp\{\varepsilon\}$ where L and K are the amounts of labor and capital inputs, respectively, α_L and α_K are parameters, and ε represents total factor productivity which is unobservable to the researcher. We can take a logarithm transformation of this production function to have the linear in parameters regression model, $\ln q = \alpha_L \ln L + \alpha_K \ln K + \varepsilon$. In chapter 3, we present methods for the estimation of the parameters in this production function. Suppose that labor is a variable input and capital is a fixed input. The variable cost function $VC(q)$ is the minimum variable cost (in this case, labor cost) to produce an amount of output q. For this production function, we have that:[6]

$$VC(q) \quad = \quad p_L \left[\frac{q}{\exp\{\varepsilon\} K^{\alpha_K}}\right]^{1/\alpha_L} \tag{1.5}$$

and

$$FC \quad = \quad p_K\, K \tag{1.6}$$

where p_L and p_K represent the price of labor and capital, respectively.

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* .

---

that the distribution function G_{mt} is uniform with support $[(A_{mt}-1)/\beta_P\, , A_{mt}/\beta_P]$ and $A_{mt} \equiv \mathbf{X}_{mt}^{D}\, \beta_X + \varepsilon_{mt}^{D}$. Then, the aggregate market demand has the form in equation (1.3).

[5] 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.

[6] Since capital is fixed, the production function implies a one-to-one relationship between output and labor. That is, to produce q units of output (given fixed K), the firm needs $L = \left[\frac{q}{\exp\{\varepsilon\} K^{\alpha_K}}\right]^{1/\alpha_L}$ units of labor. Therefore, if p_L is the price of labor, we have that $VC(q) = p_L\, L = p_L \left[\frac{q}{\exp\{\varepsilon\} K^{\alpha_K}}\right]^{1/\alpha_L}$.

1.3.8 Revealed Preference

Under the assumption that agents make decisions to maximize a utility or payoff, observed agents' choices reveal information to us about their payoff functions. In this case, a firm's choice of output reveals information about its marginal costs, and its decision to be active in the market or not reveals information about its fixed costs.

We start by 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:

$$VC_{mt}(q) \quad = \quad \left(\mathbf{X}_{mt}^{MC}\ \gamma_x^{MC} + \varepsilon_{mt}^{MC}\right)\ q + \frac{\gamma_q^{MC}}{2} q^2 \tag{1.7}$$

γ_x^{MC} and γ_q^{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 $MC_{mt}(q) = \overline{MC}_{mt} + \gamma_q^{MC} q$, where $\overline{MC}_{mt} \equiv \mathbf{X}_{mt}^{MC}\ \gamma_x^{MC} + \varepsilon_{mt}^{MC}$ represents the exogenous part of the marginal cost – that is, the part of the marginal cost that does not depend on the amount of output. Since $q \geq 0$, we have that $\overline{MC}_{mt}$ is the minimum possible value for the marginal cost. The component $\gamma_q^{MC} q$ captures how the marginal cost increases with output.

The fixed cost is associated with inputs that are used in a fixed amount, regardless the level of output. These inputs can be land, the physical plant, or some equipment. This fixed cost is specified as $FC_{mt} = \mathbf{X}_{mt}^{FC}\ \gamma^{FC} + \varepsilon_{mt}^{FC}$, where γ^{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. 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.

1.3.9 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 à la Cournot. The assumption of Cournot competition is far from 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:

$$\Pi_{mt}(q_i, \widetilde{Q}_i) = P_{mt}(q_i + \widetilde{Q}_i)\ q_i - VC_{mt}(q_i) - FC_{mt} \tag{1.8}$$

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 her own output q_i to maximize her 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. That is, a firm's best response is a function. This best response output is characterized by the following condition of optimality which establishes that marginal revenue equals marginal cost:

$$P_{mt} + P'_{mt}(q_i + \widetilde{Q}_i)\ q_i = MC_{mt}(q_i) \tag{1.9}$$

where $P'_{mt}(Q)$ is the derivative of the inverse demand function.

Given the linear demand function $P_{mt} = A_{mt} - B_{mt} Q$, the derivative $P'_{mt}(Q) = -B_{mt}$, 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:

$$q_{mt}(N) = \frac{A_{mt} - \overline{MC}_{mt}}{B_{mt}\,(N+1) + \gamma_q^{MC}} \tag{1.10}$$

This equation shows that, keeping the number of active firms fixed, output per firm increases with the intercept in the demand curve (A_{mt}), declines with marginal cost and the slope of the demand curve (B_{mt}), and it does not depend on fixed cost. The later 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. However, 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 the output per firm.

Price over average variable cost is $P_{mt} - AVC_{mt} = [A_{mt} - B_{mt}\ N_{mt}\ q_{mt}(N)] - [\overline{MC}_{mt} + \gamma_q^{MC}/2\ q_{mt}(N)] = [A_{mt} - \overline{MC}_{mt}] - [B_{mt}\ N_{mt} + \gamma_q^{MC}/2]\ q_{mt}(N)$. Plugging expression (1.10) into this equation, we get the following relationship between price-cost margin and output-per-firm in the Cournot equilibrium:

$$P_{mt} - AVC_{mt} = \frac{\left(B_{mt} + \gamma_q^{MC}/2\right)\ \left(A_{mt} - \overline{MC}_{mt}\right)}{B_{mt}\ \ (N_{mt}+1) + \gamma_2^{MC}} = \left(B_{mt} + \gamma_q^{MC}/2\right)\ q_{mt}(N) \tag{1.11}$$

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:

$$\begin{aligned} \Pi^*_{mt}(N) &= (P_{mt} - AVC_{mt})\ q_{mt}(N) - FC_{mt} \\ &= (B_{mt} + \gamma_q^{MC}/2)\left(\frac{A_{mt} - \overline{MC}_{mt}}{B_{mt}\,(N+1) + \gamma_q^{MC}}\right)^2 - FC_{mt} \end{aligned} \tag{1.12}$$

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.

1.3.10 Market entry

Now, we specify a model for how the number of active firms in a local market is determined in equilibrium. Remember that the profit of a firm that is not active in the industry is zero.[7] 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 (1.12), 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 $\Pi^*_{mt}(N+1) < 0$.[8] In the example of Figure 1.1, the equilibrium number is 5 firms. In this particular model, 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:

$$N^*_{mt} \equiv -\left(1 + \frac{\gamma_q^{MC}}{B_{mt}}\right) + \left(A_{mt} - \overline{MC}_{mt}\right)\sqrt{\frac{1 + \gamma_q^{MC}/2B_{mt}}{FC_{mt}\, B_{mt}}} \tag{1.13}$$

The equilibrium number of firms is the largest integer that is smaller or equal to N^*_{mt}. We represent this relationship as $N_{mt} = int(N^*_{mt})$ where int is the integer function, that is, the 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.

We can combine the equilibrium output per firm in equation 1.10 and the profit function in equation 1.12), to obtain the following expression for the Cournot equilibrium profit: $\Pi^*_{mt}(N) = (B_{mt} + \gamma_q^{MC}/2)\; q_{mt}(N)^2 - FC_{mt}$. This provides the following expression for the entry equilibrium condition – $\Pi^*_{mt}(N^*_{mt}) = 0$, that is particularly useful for the estimation of the model:[9]

$$\left(\frac{Q_{mt}}{N_{mt}}\right)^2 = \frac{FC_{mt}}{B_{mt} + \gamma_q^{MC}/2} \tag{1.14}$$

This equation 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

[7] 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.

[8] 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.

[9] To derive this equation, we consider that the ratio $N^*_{mt}/int(N^*_{mt})$ is approximately equal to one.

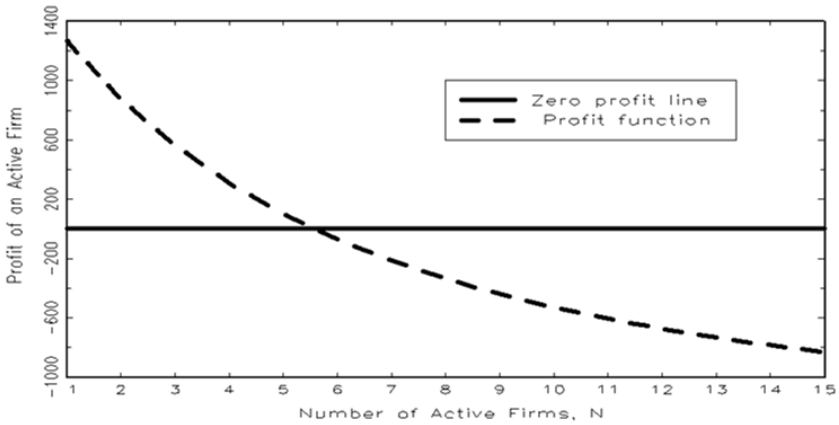

Figure 1.1: Cournot equilibrium profit as function of number of firms

with the degree of increasing marginal costs. Industries with large fixed costs, inelastic demand curves, and rapidly increasing marginal costs, have larger firms and a smaller 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 a large number of very small firms. It is clear that exogenous demand and cost are key in determining the industry market structure and market power.

1.3.11 Structural 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 expressions for the equilibrium values which are linear in parameters, and this simplifies our analysis of model identification and estimation. 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, $q \equiv Q/N$,

$$\begin{aligned} \textbf{Demand equation:} \quad P &= A - B\,N\,q \\ \textbf{Cournot Equilibrium Condition:} \quad q &= \frac{A - \overline{MC}}{B\,(N+1) + \gamma_q^{MC}} \\ \textbf{Entry Equilibrium Condition:} \quad q^2 &= \frac{FC}{B + \gamma_q^{MC}/2} \end{aligned} \tag{1.15}$$

This is a system of simultaneous equations. The system of equations in (1.15) is denoted as the *structural equations* of the model. Given a value of the exogenous variables, $\mathbf{X}$ and $\varepsilon \equiv (\varepsilon^D, \varepsilon^{MC}, \varepsilon^{FC})$, and of the structural parameters, $\theta \equiv \{\beta_x, \beta_p, \gamma_x^{MC}, \gamma_q^{MC}, \gamma^{FC}\}$, an *equilibrium* of the model is a 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:

$$q = \sqrt{\frac{FC}{B + \gamma_q^{MC}/2}} \tag{1.16}$$

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:

$$N = -\left(1 + \frac{\gamma_q^{MC}}{B}\right) + \left(\frac{A - \overline{MC}}{B}\right)\sqrt{\frac{B + \gamma_q^{MC}/2}{FC}} \tag{1.17}$$

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

$$P = \overline{MC} + (\gamma_q^{MC} + B)\sqrt{\frac{FC}{B + \gamma_q^{MC}/2}} \tag{1.18}$$

Equations (1.16), (1.17), and (1.18) 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, because 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.

1.4 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.

1.4.1 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, providing answers to some other questions requires the estimation of the structural model. For instance, the estimation of the structural model is helpful to answer our policy question.

Relationship between market size and firm size

The reduced form equation for output-per-firm in (1.16), implies the following relationship between firm size (or output per firm) q and market size S, given that $B = 1/\beta_P S$:

$$\ln(q) = \frac{1}{2}\left[\ln(\beta_p FC) + \ln(S) - \ln\left(1 + \frac{\beta_p\, \gamma_q^{MC}\, S}{2}\right)\right] \tag{1.19}$$

We can distinguish three different cases for this relationship. When fixed cost is zero $(FC = 0)$ there is no 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 larger demand have larger firms. We can distinguish between two different cases when the fixed cost is strictly positive. When the marginal cost is constant $(\gamma_q^{MC} = 0)$, 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 $(\gamma_q^{MC} > 0)$, the limit of firm size when market size goes to infinity is equal to $\sqrt{2FC/\gamma_q^{MC}}$, and this constant represents the maximum size of a firm in the industry. The value $\sqrt{2FC/\gamma_q^{MC}}$ is the level of output-per-firm that minimizes the Average Total Cost, and it is denoted 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 $FC = 1{,}400$, $\beta_p = 1$, $\gamma_q^{MC} = 0$ and $\gamma_q^{MC} = 1$.

Equation (1.19) 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 homogeneous product industry, the representation of the scatter-plot 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}

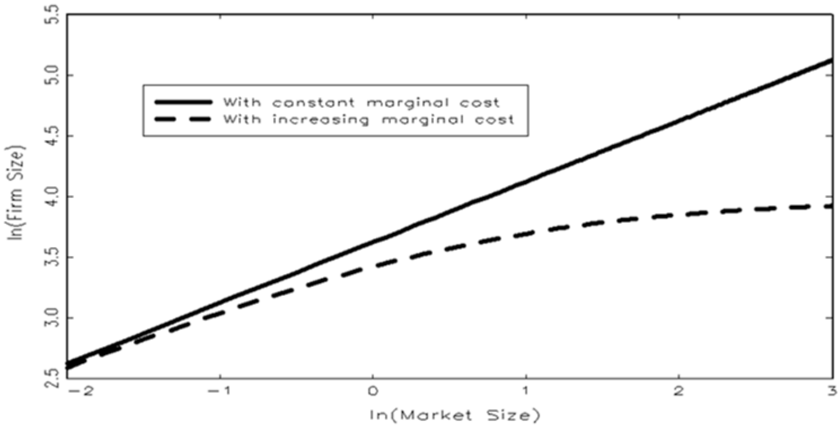

Figure 1.2: Relationship between firm size and market size

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 scatter-plot and the estimated regression line for the logarithm of firm size on the logarithm of market size in the *Women's Clothing* retail industry. In this example, the relationship in logarithms is linear, which is consistent with $FC > 0$ and $\gamma_q^{MC} = 0$ for this industry. In logarithms, for small γ_q^{MC}, we have that $\ln(q_{mt}) = \alpha_0 + \alpha_1 \ln(S_{mt}) + \alpha_q S_{mt} + u_{mt}$, where $\alpha_1 \equiv 1/2$, and $\alpha_2 \equiv -\beta_p \gamma_q^{MC}/2$. Therefore, testing the null hypothesis $\alpha_2 = 0$ is equivalent to testing for non-convexity in the variable cost, that is, $\gamma_q^{MC} = 0$. 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.

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.

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_q^{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

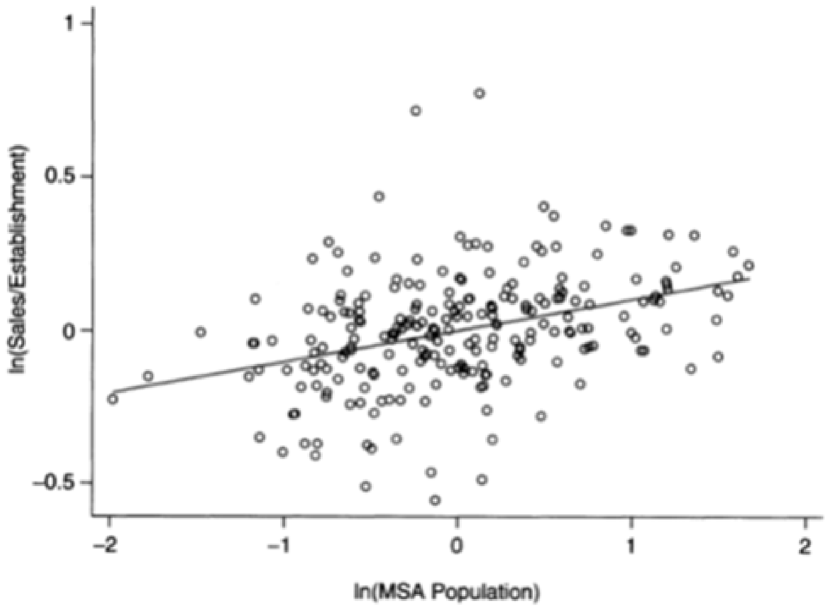

Figure 1.3: Market size matters (Campbell and Hopenhayn, 2005)

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 $P = \overline{MC} + \sqrt{FC/\beta_p S^*}$. 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 $\overline{MC} + \sqrt{2\gamma_q^{MC} FC}$, which is the marginal cost when output-per-firm is at the *Minimum Efficient Scale* (MES).

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.

Policy Question and Reduced Form Equations

Recall our initial objective: evaluating the effects of a policy which generates an increase in the fixed cost and a reduction in the marginal cost on firms in the cement industry. What do the reduced 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 (1.19) shows that an increase in the fixed cost FC and a reduction in the marginal cost parameter γ_q^{MC} imply a larger firm size. Therefore, the model predicts that the new policy will transform the industry into 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 has a negative effect on consumer welfare. A reduction in the marginal cost affects all the endogenous variables in the opposite direction. The net effects are ambiguous, and 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 between two cases or scenarios in terms of the information for the researcher about the policy change. In the first case, which 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 requires weaker assumptions than identification of structural parameters.

Many empirical questions in IO deal with predicting the effects of changes that have not yet occurred. For instance, when an industry regulator makes a recommendation on whether to approve a merger between two companies or not, she has to predict the effects of a merger that has not yet taken place. Similarly, a company that decides whether or not to introduce 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 first consider the case where the regulator has not yet implemented the new environmental regulation and wants to predict 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 the problem of identification of the structural parameters.[10]

1.4.2 Estimation of structural model

The researcher wants to use the available data to estimate the vector of structural parameters $\theta = \{\beta_x, \beta_p, \gamma_x^{MC}, \gamma_q^{MC}, \gamma^{FC}\}$. Given an estimate of the true θ, we can use

[10] 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 Marschak (1953), and it has been exploited recently in different studies. See also Chetty (2009) and Aguirregabiria (2010).

our model to evaluate/predict the effects of a hypothetical change in the cost parameters γ_x^{MC}, γ_q^{MC}, and γ^{FC} implied by the policy. For simplicity, we start by considering a version of the model without measurement error in the observable measure of market size, that is, $\exp\{\varepsilon_{mt}^S\} = 1$ for every market and period (m,t).

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

$$\begin{aligned}
\frac{Q_{mt}}{S_{mt}} &= \beta_x\, \mathbf{X}_{mt}^D - \beta_p\, P_{mt} + \varepsilon_{mt}^D \\
\left(P_{mt} - \frac{1}{\beta_p}\frac{q_{mt}}{S_{mt}}\right) &= \gamma_x^{MC}\, \mathbf{X}_{mt}^{MC} + \gamma_q^{MC}\, q_{mt} + \varepsilon_{mt}^{MC} \\
q_{mt}^2 \left(\frac{1}{\beta_p S_{mt}} + \frac{\gamma_q^{MC}}{2}\right) &= \gamma^{FC}\, \mathbf{X}_{mt}^{FC} + \varepsilon_{mt}^{FC}
\end{aligned} \tag{1.20}$$

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

We say the parameters of the model are identified if there is a feasible estimator of θ that is *consistent* in a statistical or econometric sense.[11]

To prove the vector of parameters is identified, a standard approach is using the moment restrictions implied by the model to show that we can **uniquely** determine the value of θ as a function of moments of the observable variables. For instance, in a classical linear regression model $Y = \beta_0 + \beta_1\, X + \varepsilon$ under the assumption of no correlation between the error term and the regressor, we have that $\mathbb{E}(\varepsilon) = 0$ and $\mathbb{E}(X\ \varepsilon)$ and these conditions imply that $\beta_1 = cov(X,Y)/var(X)$ and $\beta_0 = \mathbb{E}(Y) -\ [cov(X,Y)/var(X)]\ \mathbb{E}(X)$. These expressions show that the parameters β_0 and β_1 are 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.

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. In the presence of endogeneous regressors, OLS estimation produces biased and inconsistent parameter estimates. In the second equation, the left-hand-side is the price minus the

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

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

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

price-cost-margin and this should be equal to the marginal cost on 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 on 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 on the 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 on the right-hand-side of the 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, that is, $\mathbb{E}(P_{mt}\ \varepsilon_{mt}^{D}) \neq 0$; and (2) in the Cournot equilibrium equation, output per firm is an endogenous regressor, that is, $\mathbb{E}(q_{mt}\ \varepsilon_{mt}^{MC}) \neq 0$.

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.

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 is independent of the unobserved demand shock ε_{mt}^{D}. The experiment consists of a firm subsidy per unit of output produced and sold in the market. In market-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. For instance, it is determined as a random draw from some distribution. We also need 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 and with market size $\mathbb{E}(\tau_{mt}\ S_{mt}) = 0$, but it is correlated with price. That is,

$$\mathbb{E}(\tau_{mt}\varepsilon_{mt}^{D}) = 0,\ \ \mathbb{E}(\tau_{mt}S_{mt}) = 0,\ \text{ but }\ \mathbb{E}(\tau_{mt}P_{mt}) \neq 0 \tag{1.21}$$

These conditions imply that we can use the amount of subsidy, τ_{mt}, as an instrument for P_{mt} in the demand equation, to identify all the parameters in the demand equation. More precisely, the moment conditions

$$\mathbb{E}(\varepsilon_{mt}^{D}) = 0, \;\; \mathbb{E}(S_{mt}\varepsilon_{mt}^{D}) = 0, \;\; \text{and } \mathbb{E}(\tau_{mt}\varepsilon_{mt}^{D}) = 0 \tag{1.22}$$

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

$$\mathbb{E}(\varepsilon_{mt}^{MC}) = 0, \;\; \mathbb{E}(S_{mt}\varepsilon_{mt}^{MC}) = 0, \;\; \text{and } \;\mathbb{E}(\tau_{mt}\varepsilon_{mt}^{MC}) = 0 \tag{1.23}$$

to identify variable cost parameters in the Cournot equation, and the moment conditions

$$\mathbb{E}(\varepsilon_{mt}^{FC}) = 0, \;\; \mathbb{E}(S_{mt}\varepsilon_{mt}^{FC}) = 0, \;\; \text{and } \mathbb{E}(\tau_{mt}\varepsilon_{mt}^{FC}) = 0 \tag{1.24}$$

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 involved in the experiment may change the way they compete during the time the experiment is implemented. For instance, they may decide to agree not to change their levels of output such that the subsidy will not pass through to the price and they will keep the subsidy as a pure transfer. However, as long as the subsidy has some effect on price (that is, there is at least a partial pass-through of the subsidy to price), this concern does not affect the identification of the demand parameters. A key aspect in this experimental design is to ensure consumers are not aware of this experiment such that they do not change their demand behaviour. In contrast, if some consumers were aware of the temporary nature of this experiment, they may decide to buy excess cement for inventory. If that is the case, the experiment will affect demand, and the estimates of the demand parameters based on this randomized experiment will be biased.

Exclusion restrictions – Instrumental Variables

The method of instrumental variables is the most common approach to deal with endogeneity in econometrics, and in empirical micro fields in particular. 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 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 they affect marginal cost, they should be correlated with price, but they do not enter in the demand equation. What is not so obvious is whether these variables are uncorrelated with the unobserved demand shock. If the demand for 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.

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 shock". 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, that is, 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 on the dummy variable I_{mt}. For I_{mt} to be a valid instrument for price, the key identification assumption required is that demand was unaffected by the natural event. Under this assumption, the moment condition $\mathbb{E}(I_{mt}\ \varepsilon^D_{mt}) = 0$, together with the conditions $\mathbb{E}(\varepsilon^D_{mt}) = 0$ and $\mathbb{E}(S_{mt}\ \varepsilon^D_{mt}) = 0$, identify the demand parameters.

The condition 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 have taken 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 $\mathbb{E}(I_{mt}\ \varepsilon^D_{mt}) = 0$. To try to deal with this issue, most applications exploiting identification from 'natural experiments' assume a particular structure for the unobserved error:

$$\varepsilon^D_{mt} = \omega^D_m + \delta^D_t + u^D_{mt}, \tag{1.25}$$

We can control for ω^D_m using market dummies, and for δ_t using time dummies. The 'natural experiment' dummy I_{mt} can be correlated with ω^D_m and/or with δ^D_t. The identification assumption is that I_{mt} is not correlated with the shock u^D_{mt}.

Restrictions on unobservables

Suppose that the unobservables in the demand and in the marginal cost have the covariance structure:

$$\begin{array}{rcl} \varepsilon^D_{mt} & = & \omega^D_m + \delta^D_t + u^D_{mt}, \\ \varepsilon^{MC}_{mt} & = & \omega^{MC}_m + \delta^{MC}_t + u^{MC}_{mt} \end{array} \tag{1.26}$$

These components of the variance specification of the unobservables, together with restrictions on the serial or/and the spatial correlation of the demand shocks u^D_{mt}, 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 (that is, Arellano-Bond Instruments; Arellano and Bond, 1991), or on the spatial correlation (that is, Hausman-Nevo Instruments; Hausman, 1996, and Nevo, 2001).

Arellano-Bond instruments. Suppose that the shock u^D_{mt} is not serially correlated over time. That is, all the time persistence in unobserved demand comes from the time-invariant effect ω^D_m, and from the common industry shocks δ^D_t, but the idiosyncratic demand shock u^D_{mt} is not persistent over time. Under these conditions, in the demand equation in first-differences, $\Delta Q_{mt}/S_{mt} = \beta_S\ \Delta S_{mt} - \beta_1\ \Delta P_{mt} + \Delta\delta^D_t + \Delta u^D_{mt}$, the lagged endogenous variables $\{P_{mt-2}, Q_{mt-2}, N_{mt-2}\}$ are not correlated with the error Δu^D_{mt}, and they 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 than 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 a 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 cost 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 identify 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 restrictions imply the moment conditions:

$$\mathbb{E}\left(\left[q_{mt}^2\left(\frac{1}{\beta_p S_{mt}}+\frac{\gamma_2^{MC}}{2}\right)-\gamma^{FC}\,\mathbf{X}_{mt}^{FC}\right]\left[\frac{Q_{mt}}{S_{mt}}-\beta_x\,\mathbf{X}_{mt}^{D}-\beta_p\,P_{mt}\right]\right)=0 \tag{1.27}$$

and

$$\mathbb{E}\left(\left[q_{mt}^2\left(\frac{1}{\beta_p S_{mt}}+\frac{\gamma_2^{MC}}{2}\right)-\gamma^{FC}\,\mathbf{X}_{mt}^{FC}\right]\left[P_{mt}-\frac{1}{\beta_p}\frac{q_{mt}}{S_{mt}}-\gamma_1^{MC}\,\mathbf{X}_{mt}^{MC}-\gamma_2^{MC}\,q_{mt}\right]\right)=0 \tag{1.28}$$

These moment restrictions, together with 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.

Multiple equilibria and Identification

Multiplicity of equilibria is a 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, that is, linear demand; (b) homogeneous firms, that is, homogeneous product and costs; and (c) no dynamics. Once we relax any of these assumptions, multiple equilibria becomes 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 be, 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 is a decreasing function of the number of firms in the local market. For instance, the supply of equipment (fixed input) increases with the number of firms in the market, and the price of this fixed input declines. 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:

$$\left(\frac{Q_{mt}}{N_{mt}}\right)^2 = \frac{\gamma^{FC} - \delta\ N_{mt} + \varepsilon_{mt}^{FC}}{B_{mt} + \gamma_q^{MC}/2} \tag{1.29}$$

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 generate 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,

$$q_{mt}^2 \left(\frac{1}{\beta_p S_{mt}} + \frac{\gamma_q^{MC}}{2}\right) \quad = \quad \gamma^{FC}\ \mathbf{X}_{mt}^{FC} - \delta\ N_{mt} + \varepsilon_{mt}^{FC} \tag{1.30}$$

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 a known variable to the researcher. In the right hand side, we now have 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, multiple equilibria may even 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 no such 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 defined 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 book.

1.4.3 Extensions

The rest of the book deals with empirical models of market structure that relax some of these assumptions. (a) *Product differentiation* and more general forms of demand (see chapter 2 on demand estimation). (b) *Heterogeneity in firms’ costs*: exploiting information on firms’ inputs to identify richer cost structures(see chapter 3, on production function estimation). (c) *Relaxing the assumption of Cournot competition*, and identification of the "nature of competition" from the data, for instance, collusion (see chapter 4 on models of price and quantity competition). (d) *Heterogeneity of entry costs* in oligopoly games of entry (see chapter 5 on static games of entry). (e) *Spatial differentiation* and plant spatial location. (see chapter 5 on games of spatial competition). (f) Competition in quality and other product characteristics (see chapter 5 on games of quality competition). (g) *Investment in capacity* and physical capital (see chapters 6 and 7 on dynamic structural models of firm investment decisions). (h) *Consumers intertemporal substitution and dynamic demand* of storable and durable products (see chapter 8 on dynamic demand). (i) Dynamic strategic interactions in firms’ investment and innovation decisions (see chapter 9 dynamic games]. (j) *Mergers* (see chapter 5 on conduct parameters and chapter 9 on dynamic games). (k) *Firm networks*, chains, and competition between networks (see chapter 9 on dynamic games). (l) Firms’ competition in auctions (see chapter 10 on auctions).

1.5 Summary

In this chapter, we have described Empirical Industrial Organization as a discipline that deals with the combination of data, models, and econometric methods to answer empirical questions related to the behavior of firms in markets. The answers to empirical questions IO are typically based on the estimation of structural models of competition. These models have four key components: demand, costs, price or quantity competition, and market entry. The identification and estimation of the structural parameters in these models are based on the principle of revealed preference. Endogeneity is an important issue in the estimation of the model parameters. We have described different approaches to deal with endogeneity problems, from randomized control trials and natural experiments, to instrumental variables, and restrictions on the structure of the unobserved variables. Multiplicity of equilibria is also a common feature in some empirical games.

1.6 Exercises

1.6.1 Exercise 1

Write a computer program in your favorite mathematical software (for instance, R, Gauss, Matlab, Stata, Julia, Python, etc) that implements the following tasks.

(a) 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 $\{1, 2, ..., 100\}$.

(b) For each market type/size, obtain equilibrium values of the endogenous variables including 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.

(c) Now, consider a policy change that increases fixed cost and reduces marginal cost. Obtain two-way graphs of each variable against market size representing the curves both before and after the policy change.

1.6.2 Exercise 2

Write a computer program in your favorite mathematical software that implements the following tasks.

(a) 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.

(b) 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 $\{P_{mt}, Q_{mt}, N_{mt}, S_{mt}\}$

(c) Use these data to estimate the model by OLS, and also try some of the identification approaches to identify the parameters of the model.

1.6.3 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 computer program that implements the following tasks.

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

(b) Consider the following comparative statics (counterfactual) exercises and obtain the empirical distribution (histogram) for the distribution of prices under each of the following changes: (i) eliminate heterogeneity in market size: set all market sizes equal to the one in the median market; (ii) eliminate market heterogeneity in demand shocks: set all demand shocks equal to zero; (iii) eliminate all the market heterogeneity in marginal costs; and (iv) remove all the market heterogeneity in fixed costs. Generate figures of each of these counterfactual distributions together with the factual distribution.