# 11. Entry games and free entry equilibria*

*Michele Polo*

## 1 INTRODUCTION

What are the elements that may explain why certain industries are populated by a large number of firms, each covering a small fraction of total output, whereas other markets are dominated by few large firms that supply a relevant fraction of customers? These questions have been at the core of the topics studied in industrial organization (IO) from the very beginning.[1] These research topics have been approached in the early phases of industrial economics mostly from an empirical perspective[2] within the structure–conduct–performance paradigm, while the theoretical foundations of endogenous market structures have been explored more rigorously in the game-theoretic framework of the new IO literature. The analytical framework that has been developed looks at market entry and exit as the process that endogenously determines the number and characteristics of active firms in the long run. In this setting, then, other research questions emerge. How do these market structures change in reaction to a variation in some key parameters? Are we able to identify a set of robust comparative statics properties in oligopoly markets, despite the rich variety of models in the IO literature? And finally, on the normative side, does entry into the market, a key component of the competitive process, lead to a welfare-maximizing outcome, or might the number and characteristics of firms exceed or fall short of the level of efficiency?

This chapter deals with the theories of market equilibria when the number and characteristics of active firms are endogenously determined through the process of entry. More precisely, we shall review the literature on entry games and free entry equilibria in a multi-stage game framework. A large number of potential entrants decide first whether to enter or not; once all the firms have undertaken their entry decisions, the active firms compete according to some oligopoly game. The chapter is organized as follows. In Section 2 we present the general analytical framework. In Section 3 we analyze a wide range of symmetric oligopoly models to identify the relationship between the number of firms and the market equilibria: we start with homogeneous products and competition in strategic substitutes (Section 3.1), moving then to differentiated products and competition in strategic complements (Section 3.2), next offering a general explanation of the comparative statics properties (Section 3.3) and concluding with cartels (Section 3.4). We then move to free entry equilibria and the determinants of the maximum number of firms (Section 4). Finally, we consider symmetric entry games under a normative perspective (Section 5), looking at the comparison between the free entry and the welfare-maximizing number of firms. In Section 6 we review asymmetric

* I thank Simon Anderson, Emilio Calvano, Chiara Fumagalli and Michele Grillo for very useful discussions and suggestions. Usual disclaimers apply.

1 See Bain (1956) and Scherer (1980).

2 See Schmalensee (1989) for a comprehensive survey of the empirical literature.

free entry equilibria that exploit the aggregative nature of most oligopoly models. We then present the case of endogenous sunk cost and persistent concentration (Section 7) and the case of frictionless entry and contestable markets (Section 8). Concluding remarks follow.

## 2 ENTRY GAMES

There are several ways to model the entry process and market interaction among active firms. The various set-ups allow us to highlight different issues, focussing on distinct effects that interact in the overall market dynamics. We can draw a key distinction between the environments in which the entry decisions precede the market strategies, and those where some firms undertake entry decisions after observing their competitors' market strategies.

In the former case, the *market strategies* of individual firms cannot be chosen with the purpose of affecting the entry decisions of any firm, since entry already occurred, although the features of the *market equilibria* that result from the aggregate process of entry affect the early decision to enter the market. In this perspective, multi-stage games represent a suitable formal framework. There is a large group of $m$ potential entrants $j \in I_m$ that choose whether to enter, incurring a fixed set-up cost $F > 0$, or not; then, once they have taken their decision and the set of $n \leq m$ entrants $i \in I_n$ is common knowledge, the active firms play a market game. This set-up is usually adopted to study long-run free entry equilibria, in which a set of exogenous variables referring to the primitives of technology and preferences explains the long-run market structure.

Alternatively, in a second class of strategic environments, a subset of early entrants (incumbents) commit to observable market strategies before the other firms (entrants) decide whether to enter or not. The incumbents' initial strategy, then, may affect the entry decisions of the latecomers, explaining why this set-up is widely used to study strategic entry deterrence and foreclosure. In this environment, the market structure is explained by foreclosure strategies, based on a rich set of strategic tools, rather than by market fundamentals.

The two set-ups are useful to explore different and complementary issues and they are characterized by a different time horizon. Sequential entry with incumbents and entrants is a more realistic representation of short-run market dynamics, since entry is typically an ongoing process where already established and new firms interact. The possibility of foreclosure, then, is an empirically relevant issue that characterizes the evolution of markets. At the same time, multi-stage entry games allow us to move away from these short-run phenomena and focus on the underlying features of preferences and technology as long-run drivers of market evolution. By shifting attention to this complementary perspective we can identify fundamental forces that, despite the frictions that in the short run may slow down the process and foreclose the market, push towards a more or less concentrated market. Since in this chapter the focus is on long-run market structures rather than foreclosure, we will consider several and different specifications of multi-stage entry games.

A second relevant feature recurring across models is the assumption of symmetric firms. Supply-side symmetry is a natural assumption in a long-run perspective, since we may think that any barrier to adopting best practice technologies, such as patent protection or private know-how, tends to vanish in the long run. Demand-side symmetry, consistent with homogeneous products or horizontal product differentiation and different varieties, is a convenient assumption when we want to analyze the number of entrants and the distribution of market shares.[3]

The different models considered in the following sections make use of the symmetry assumption at different levels, either by applying it to the whole population of potential entrants, or to a subset of them identified as marginal entrants, while allowing for asymmetries across major market players. We shall see that the symmetry assumption is also at the core of the analysis of potential competition and contestable markets.

## 3 SYMMETRIC OLIGOPOLY MARKETS

We start our analysis of entry games by considering the (second-stage) market games where $n$ firms are active, having decided to enter in the first stage. In this section we consider symmetric market games where all the $n$ firms share the same (best-practice) technology and no one has an advantage on the demand side, e.g. a higher-quality product. In this setting, when firms adopt the same strategies $a_i = a$, $i \in I_n$, then they obtain the same level of profits. A symmetric environment greatly simplifies the analysis of free entry equilibria, since the equilibrium profits, as well as the equilibrium strategies, consumers' surplus and welfare, all depend on a vector $x$ of parameters related to the properties of costs (technology) and demand (preferences), and on the number of firms $n$: $\Pi_i(a_i^*, a_{-1}^*) = \Pi^*(n; x)$. Market equilibria, once the entry process has been completed, therefore can be analyzed simply in terms of the number of firms $n$. The individual equilibrium profits $\Pi^*(n; x)$ are therefore the object that potential entrants consider when, at the initial stage of the game, they choose whether to enter or not, given their expectation of the number of firms that will enter.

Oligopoly theory offers a very rich set of models that describe market interaction among $n$ competitors, ranging from homogeneous to differentiated products and distinguishing competition in strategic substitutes or complements. In all these environments, moreover, demand and cost functions can be specified differently. Finally, beyond static, possibly multi-stage games, the literature on tacit collusion adds to the toolkit for the analysis of cartels. A general theory of free entry equilibria has to encompass all these classes of models, admitting a variety of business strategies, modes of strategic interaction and features of demand and costs. In this perspective, then, the key point is whether there exist some regularities across different models in the relationship between the number of (symmetric) active firms $n$ and the equilibrium profits they obtain $\Pi^*(n; x)$. A first, relevant result that we are going to present in the following sections, is that, despite the significant differences in oligopoly equilibria across models, we can establish under very general conditions a negative relationship between the equilibrium profits and the number of firms.

We organize the discussion by considering three different cases: homogeneous products and strategic substitutes, differentiated products and strategic complements, and repeated games.

[3] As will be clear in the following sections, this approach does not prevent us from also considering environments where, for instance, firms offer goods of different quality, which are therefore attractive to consumers in different ways. What we maintain is that, even in these cases, there is a further dimension of (horizontal) product differentiation such that for each level of quality several firms may further differentiate their products by variety. In this case, symmetry is preserved at each layer of quality.

### 3.1 Homogeneous Products and Strategic Substitutes

Our first look at symmetric oligopoly equilibria refers to a market with $n$ firms producing a homogeneous product and competing in strategic substitutes, usually associated with the Cournot model. Since the pioneering work of Cournot (1838) a large number of contributions have explored the conditions for the existence of and characteristics of the equilibria when $n$ firms compete in quantities. McManus (1962, 1964) and Roberts and Sonnenschein (1976), independently proved the existence of a symmetric equilibrium in symmetric Cournot games with convex costs. Novshek (1985) showed that an $n$-oligopoly has a Nash equilibrium if each firm's marginal revenue is decreasing in the other firms' aggregate output. A step forward in proving the existence of Cournot equilibria under general conditions is in Vives (1990), who showed in the duopoly case the relationship between the assumptions of the previous literature and the submodularity of Cournot games. Supermodular games and the techniques of monotone comparative statics,[4] have proved to be extremely useful tools to explore the properties of Cournot oligopolies and to identify the general conditions under which the comparative statics of equilibria can be analyzed. We summarize here the main results following this approach as in Amir and Lambson (2000).

Consider an oligopoly with $n$ firms offering a homogeneous product and producing with the same cost function $C(q_i)$ and incurring no capacity constraint over the relevant output range. Market inverse demand $P(Q)$ is a continuous and differentiable function of total output $Q = \sum_{i=1}^{n} q_i$. The profit function of firm $i$, then, is:

$$\Pi_i(q_i, \mathbf{Q}_{-i}) = P(Q)q_i - C(q_i)$$

where $\mathbf{Q}_{-i} = \{q_j\}_{j \neq i}$ is the vector of outputs of the other firms. In this traditional specification, each firm maximizes its profits by choosing a level of output for given strategies of the other firms, $\mathbf{Q}_{-i}$. It is well recognized that under standard assumptions, firm $i$'s best reply $\widehat{q}_i(\mathbf{Q}_{-i}) = \arg\max_{q_i} \Pi_i(q_i, \mathbf{Q}_{-i})$ is downward sloping in the other firms' output, implying a submodular game and competition in strategic substitutes.

Let us define

$$\Delta(q_i, Q) := -P'(Q) + C''(q_i). \tag{11.1}$$

Then, Amir and Lambson (2000) prove that if $\Delta(q_i, Q) > 0$ on the relevant range of outputs and the inverse demand function is log-concave, there exists a unique and symmetric equilibrium, with individual output $q^*(n)$ nonincreasing in $n$ and total output $Q^*(n)$ (market price $P(Q^*(n))$) nondecreasing (nonincreasing) in $n$.[5] This condition holds, for instance, in the set-up adopted in the works of McManus (1962, 1964), Roberts and Sonnenschein (1976) and Novshek (1985) quoted above and is consistent with the framework proposed in Vives (1999).

---

[4] See Milgrom and Roberts (1990, 1994) and Milgrom and Shannon (1994).

[5] Amir and Lambson (2000) prove (Theorem 2.2) a more general result that does not require log-concavity of the inverse demand function and that allows for multiplicity of Cournot equilibria. In this case the comparative statics properties with respect to $n$ of total equilibrium output and the equilibrium output of $n-1$ firms are preserved by considering the values of the extremal equilibria. We focus in the text on uniqueness to ease the exposition.

To illustrate this result with an example let us consider the linear Cournot model: market demand is $Q = S * \left[\alpha - \beta p\right]$, where $S$ measures market size, e.g. the number of consumers. Then, the inverse demand is $P(\frac{Q}{S}) = a - b\frac{Q}{S}$ where $a = \frac{\alpha}{\beta}$, $b = \frac{1}{\beta}$ and $Q$ is total supply. Firms produce at constant marginal cost $c \in (0, a)$ and compete in quantities. Then, each firm selects its optimal output by solving $q_i^* = \arg\max_{q_i} \left(P(\frac{Q}{S}) - c\right) q_i$. The symmetric equilibrium quantity $q^*(n)$ satisfies for all firms the first-order conditions:

$$\left(P\left(\frac{nq^*}{S}\right) - c\right) - P'\frac{q^*}{S} = 0, \tag{11.2}$$

Substituting and solving for the symmetric equilibrium we get:

$$q^*(n) = S\frac{a-c}{b(n+1)}, \quad p^*(n) = \frac{a+nc}{n+1} \geq c, \quad \Pi^*(n) = \frac{S}{b}\left(\frac{a-c}{n+1}\right)^2. \tag{11.3}$$

When the number of firms increases, therefore, the individual quantity decreases, whereas total output increases. Consequently, the market clearing price falls and tends to the marginal cost when the number of firms increases indefinitely. Finally, the equilibrium profits, gross of the fixed entry costs, decrease in $n$ and tends to zero at the limit, due to the combined quantity and price effects.

This pattern characterizes the so-called *Cournotian paradigm*, a representation of the market equilibrium that depends on the number of firms and that moves from the monopoly to the perfectly competitive equilibrium as $n$ increases from 1 to infinity. Perfect competition, in this setting, corresponds to the limiting case when each firm supplies an infinitesimal amount of output in a market populated by an infinite number of negligible firms.

This structural view of perfect competition can be easily derived from the first-order conditions that guarantee a profit-maximizing solution for any number of firms. Equation (11.2), indeed, implies that the market clearing price tends to the marginal cost when the last term vanishes. There are two possible explanations why $P'\frac{q^*}{S} \to 0$. One argues that when firms are small with respect to the market, they follow a *price-taking behavior*; that is, they expect the market price not to react to any change in their individual output. This case corresponds to assuming $P' = 0$ in a perfectly competitive market. The other explanation, which is consistent with the structuralist view of the Cournotian paradigm, instead focusses on the fact that it is the individual quantity that vanishes as $n$ becomes indefinitely large, whereas $P' < 0$ even at the limit. In this latter case, indeed, $\lim_{n\to\infty} q^*(n) = 0$, as evident from (11.3).

It is interesting to notice that the last term in (11.2) also represents the negative externality that characterizes strategic interaction in a Cournot game, i.e. $\frac{\partial \Pi_i}{\partial q_j} = P'\frac{q^*}{S}$. In other words, with Cournot competition each firm affects the rivals' profits when it increases its quantity since it makes the price fall and reduces the revenues that the competitors obtain from their production. The level of individual production, therefore, multiplicatively affects this externality, which vanishes when each firm produces a negligible output. Then, a perfectly competitive market in a Cournotian perspective is also characterized at the limit by vanishing externalities across firms. This result confirms the idea that in a perfectly competitive market no externality occurs, a feature that is driven by the same effect ($\lim_{n\to\infty} q^*(n) = 0$) that explains why the competitive price tends to the marginal cost.

Finally, market size $S$ increases individual and total quantities as well as the equilibrium profits.

### 3.2 Differentiated Products and Strategic Complements

A different class of oligopoly models moves into the realm of differentiated products and assumes that firms compete in prices, a framework that entails strategic complementarities. In the product differentiation literature, moreover, we can assume that either differentiation does not break the intrinsic symmetry of firms' market positions, or alternatively that product differentiation introduces a competitive advantage for some firms with respect to the others. The former case recalls the idea of (horizontal) differentiation by variety, where products differ in terms of characteristics, each one being more suited to a specific subset of customers. The latter, instead, captures the idea of (vertical) differentiation in quality. Given our focus on symmetric equilibria, in this section we shall consider several approaches to differentiation by variety. We shall consider entry and differentiation by quality in Section 7.

There are three main ways to model the demand side when products are (horizontally) differentiated: the representative consumer approach characterized by preference for variety; the discrete choice model where the external observer is able to reconstruct consumers' behavior up to a random component related to unobservable individual characteristics; and the address approach that assumes heterogeneous consumers with inelastic demand.[6]

Let $q_i = S * D_i(p_i, \mathbf{p}_{-i})$ be the demand for product $i \in I_n$, where $S$ measures the size of the market and $\mathbf{p}_{-i}$ is the vector of prices other than $p_i$. Let us further assume $D_i(.)$ is continuous and differentiable and $C_i(D_i(.)) = cD_i(p_i, \mathbf{p}_{-i})$. Finally, let us assume that each firm offers only one variety.[7] Each firm solves the following problem: $\max_{p_i}(p_i - c)D_i(p_i, \mathbf{p}_{-i})$. Under standard assumptions on the strategy space being compact and convex, and the profit function being quasi-concave, the following equation identifies the necessary and sufficient conditions for a maximum:

$$\frac{p_i^* - c}{p_i^*} = \frac{D_i\left(p_i^*, \mathbf{p}_{-i}\right)}{p_i^* \frac{\partial D_i}{\partial p_i}} = \frac{1}{\varepsilon_i} \tag{11.4}$$

where $\varepsilon_i$ is the price elasticity of demand for product $i$. In a symmetric equilibrium $p_i^* = p^*(n)$, $i \in I_n$, and

$$\varepsilon^*(n) = \frac{p^*(n)\frac{\partial D_i}{\partial p_i}}{D_i\left(p^*(n), \mathbf{p}^*(\mathbf{n})\right)}. \tag{11.5}$$

[6] For a detailed analysis of these three approaches and the relationships among them see Anderson, De Palma and Thisse (1992). On the representative consumer models see, for instance, the constant elasticity of substitution (CES) representation adopted in Spence (1976) and Dixit and Stiglitz (1977) and the linear representation in Shubik and Levitan (1980) and Singh and Vives (1984). On the interpretations of random utility models, we find two approaches: Manski (1977) assumes that utility is deterministic but it is not perfectly observed by the other agents, with a random term capturing the unobserved component; Quandt (1956) instead assumes the individual behavior to be intrinsically probabilistic. Finally, the address model approach was first proposed in Hotelling (1929). See also Salop (1979) and D'Aspremont et al. (1979).

[7] As we shall discuss in Section 4, assuming single-product firms makes the analysis of the maximum number of varieties and that of firms equivalent. With multi-product firms, instead, the maximum number of varieties will be larger than the number of active firms in a free entry equilibrium.

Hence, the pattern of equilibrium prices $p^*(n)$ when the number of firms increases depends inversely on the corresponding pattern of $\varepsilon^*(n)$. If $\lim_{n\to\infty} \varepsilon^*(n) = \infty$, then at the limit the price converges to the marginal cost, and we replicate the perfectly competitive equilibrium already found in the case of Cournot competition. When, instead, $\lim_{n\to\infty} \varepsilon^*(n) = \overline{\varepsilon}$ with $\overline{\varepsilon}$ finite, a positive mark-up persists at the limit, a pattern associated to Chamberlinian monopolist competition.[8] As we shall see, the limiting properties of the different approaches to product differentiation are consistent with either of the two alternatives.

Let us consider first the case of convergence to competitive equilibria. Generalizing the duopoly linear model originally proposed by Singh and Vives (1984) and further developed in Häckner (2000), the utility function of the representative consumer is quasi-linear according to the expression:

$$U(q_1,\ldots,q_n;I) = \alpha \sum_{i=1}^{n} q_i - \frac{1}{2}\left(\sum_{i=1}^{n} q_i^2 + 2\gamma \sum_{j\neq i} q_i q_j\right) + O \tag{11.6}$$

where $\gamma \in [0,1)$ measures product substitutability and $O$ is the money spent on outside goods. The demand system, then, is:

$$D_i(p_i,\mathbf{p}_{-i}) = S * \frac{\alpha(1-\gamma) + \gamma \sum_{j\neq i} p_j - \left[\gamma(n-2)+1\right] p_i}{(1-\gamma)\left[\gamma(n-1)+1\right]} \tag{11.7}$$

where $S$ measures the size of the market, i.e. the number of representative consumers. Notice that in a symmetric price configuration $p_i = p$ for $i \in I_n$, firm $i$'s demand

$$D_i(p,\mathbf{p}) = S * \frac{\alpha - p}{\left[\gamma(n-1)+1\right]}$$

decreases in the number of firms, since consumers spread their purchases over a larger set of varieties. The demand elasticity in a symmetric price equilibrium is:

$$\varepsilon^*(n) = \frac{\left[\gamma(n-2)+1\right] p^*(n)}{(\alpha - p^*(n))\,(1-\gamma)}. \tag{11.8}$$

Hence, $\lim \varepsilon^*(n) = \infty$ being $p^*(n) < \alpha$. Indeed, the equilibrium price

$$p^*(n) = \frac{\alpha(1-\gamma) + c\left[\gamma(n-2)+1\right]}{\gamma(n-3)+2} \tag{11.9}$$

8 See Vives (1999), pp. 160–64 for a detailed discussion.

tends to the marginal cost when $n \to \infty$. Moreover, the equilibrium quantity and profits

$$q^*(n) = S * \frac{(\alpha - c)\left[\gamma(n-2)+1\right]}{\left[\gamma(n-1)+1\right]\left[\gamma(n-3)+2\right]} \tag{11.10}$$

and

$$\Pi^*(n) = S * \frac{(\alpha - c)^2(1-\gamma)\left[\gamma(n-2)+1\right]}{\left[\gamma(n-1)+1\right]\left[\gamma(n-3)+2\right]^2} \tag{11.11}$$

are decreasing in the number of firms $n$.

A similar pattern can be obtained within the address models of product differentiation. Following Salop (1979) we can extend the original linear Hotelling duopoly to encompass $n$ active firms by considering a circular market of length 1 where $S$ consumers are uniformly distributed according to their individual preferred version $t$. Firms $i \in I_n$ produce at constant marginal cost $c$ and sell at price $p_i$ horizontally differentiated varieties that are evenly distributed at $x_i = i/n$ around the circle. Finally, a consumer of type $t$ purchasing variety $i$ has a net utility $u^* - p_i - (x_i - t)^2/\gamma$. We also use in this class of address models parameter $\gamma$ to positively affect product substitutability. When $\gamma$ is large the utility mostly depends on the price and the consumers are ready to switch to a more convenient, although more distant, variety. The demand system, in this setting, is given by:

$$D_i(p_i, p_{i-1}, p_{i+1}) = S\left[\frac{1}{n} - n\gamma p_i + \frac{n\gamma}{2}\left(p_{i+1} + p_{i-1}\right)\right] \tag{11.12}$$

and displays localized competition between neighboring varieties, a notable feature of the address approach. The demand elasticity in a symmetric equilibrium is

$$\varepsilon^*(n) = \gamma n^2 p^*(n) \tag{11.13}$$

and $\lim_{n\to\infty} \varepsilon^*(n) = \infty$, implying convergence to the marginal cost. Notice also that, for given $n$, the elasticity is increasing in the substitutability parameter $\gamma$.

The symmetric equilibrium price, quantity and profits, indeed, are given by:

$$p^*(n) = c + \frac{1}{\gamma n^2}, \quad q^*(n) = \frac{S}{n} \quad \Pi^*(n) = \frac{S}{\gamma n^3}. \tag{11.14}$$

Comparing the symmetric equilibria in the Singh and Vives (1984) and in the Salop (1979) models of product differentiation with those obtained in the Cournot linear model we find significantly similar comparative statics properties, with price and individual quantity falling in the number of firms and the price approaching the marginal cost as the number of firms tends to infinity. Indeed, the driving effect we highlighted in Cournot, based on vanishing individual quantities still applies. In the Salop model, however, an additional interesting effect is at work. When $n$ increases indefinitely the market is completely covered with (locally) almost identical varieties. Localized competition between adjacent varieties reproduces a Bertrand environment, leading to marginal cost pricing. This latter effect corresponds to an increasingly intense price competition between closer and closer variety. In other words, in the localized competition model of product differentiation an increase in $n$ produces at the same time a vanishing quantity externality and an increasing price externality, both pushing towards convergence to a competitive outcome.

We can now turn to the case of monopolistic competition, when positive mark-ups are associated with a market populated by a very large (i.e. infinite) number of infinitesimal firms. We illustrate this case referring to the multinomial logit model, thereby also covering the discrete choice approach to product differentiation. Let the utility of a consumer be described by a deterministic component $U(p_i) = \alpha - p_i$ and an additive random independent and identically distributed (i.i.d.) component $\eta_i$ that is distributed according to the double exponential distribution $F(x) = \exp - \left[\exp -(\gamma x + \epsilon\right]$ where $\epsilon$ is Euler's constant and $\gamma$ a positive constant that negatively affects the variance. Then, the resulting probability of choosing product $i$ given the vector of prices $(p_1, \ldots, p_n)$ is

$$P_i(p_i, \mathbf{p}_{-i}) = \frac{\exp(-\gamma p_i)}{\sum_{j=1}^{n} \exp(-\gamma p_j(\mu)}. \tag{11.15}$$

Then firm $i$'s expected profits are:

$$\Pi_i(p_i, \mathbf{p}_{-i}) = S * (p_i - c) P_i(p_i, \mathbf{p}_{-i}).$$

We can observe that $\frac{\partial P_i}{\partial p_i} = \gamma P_i(1 - P_i)$ and that, therefore, parameter $\gamma$, once again, captures product substitutability. Moreover, in a symmetric equilibrium $P_i(p, \mathbf{p}){=}\frac{1}{n}$. Then, the elasticity of demand is

$$\varepsilon^*(n) = \frac{\gamma(n-1)p^*(n)}{n}, \tag{11.16}$$

with $\lim_{n\to\infty} \varepsilon^*(n) = \gamma p^*(n)$ finite.[9] Hence, the firms obtain a positive mark-up when $n$ tends to infinity. The equilibrium price, quantity and profits are:

$$p^*(n) = c + \frac{n}{\gamma\,(n-1)}, \quad q^*(n) = \frac{S}{n}, \quad \Pi^*(n) = \frac{S}{\gamma\,(n-1)}. \tag{11.17}$$

9 Parameter $\gamma$, as in the previous models, positively affects price elasticity for given $n$.

The multinomial logit model[10] presents a different pattern of price adjustment, with the equilibrium price decreasing in the number of firms and converging to a mark-up $1/\gamma$ when $n \rightarrow \infty$. Despite the positive mark-up, the firm's profits vanish at the limit, since the individual output becomes negligible, as it is in a monopolistic competition environment. We can also notice that the basic channel of interaction across firms vanishes as well at the limit: $\frac{\partial P_i}{\partial p_j} = \gamma P_i P_j = \frac{\gamma}{n^2}$. Hence, the "competitive" component of monopolistic competition is associated with vanishing externalities, as already observed when discussing the Cournot model.

To sum up, the different models of product differentiation display similar comparative static properties with respect to the number of firms, with the equilibrium price, quantity and profits falling in $n$. The main difference rests on the convergence of the equilibrium prices to the marginal cost, as in a perfectly competitive market, or instead to a positive mark-up over costs that characterizes monopolistic competition. Moreover, the size of the market, in all cases, pushes up profits.

The results of the product differentiation literature provide an additional insight related to the intensity of price competition and its effect on $n$-firms market equilibria. In the three models, with a little abuse of notation, we have represented product substitutability through parameter $\gamma$, with the price elasticity increasing and the price and profits falling in $\gamma$.

### 3.3 Explaining the Comparative Statics in a Unified Framework

In the previous sections we have shown that the market equilibria, described by prices and quantities, share similar comparative statics properties across a wide range of different oligopoly models and features of preferences and technology. This raises a natural question of whether this common pattern may be accounted for through a unified explanation. The theory of monotone comparative statics developed by Milgrom and Roberts (1990, 1994) and Milgrom and Shannon (1994) offers an enlightening perspective. Their approach allows the development of new tools with which to study how equilibria change in reaction to a variation in the parameters and constraints of the maximization problem, moving beyond the tradition approach based on the implicit function theorem.[11] Quoting Amir (2003, p. 2), "if in a maximization problem, the objective reflects a complementarity between an endogenous variable and an exogenous parameter, in the sense that having more of one increases the marginal return to having more of the other, then the optimal value of the former will be increasing in the latter. In the case of multiple endogenous variables, then all of them must also

[10] A similar result is obtained, within the representative consumer approach, assuming Cobb-Douglas preferences between a numéraire good $q_0$ and a set of differentiated products $q_i$ with CES preferences:

$$U(q_o, q_{1,\dots}, q_n) = q_0^{1-\beta} \widetilde{q}^{\beta} \quad \text{with } \gamma \in (0, 1)$$

and

$$\widetilde{q} = \left( \sum_{i=1}^{n} q_i^{\frac{\sigma-1}{\sigma}} \right)^{\frac{\sigma}{\sigma-1}}.$$

See Spence (1976), Dixit and Stiglitz (1977) and Anderson et al. (1992) pp. 226–9.

[11] Importantly, the new tools help to deal with the comparative statics of multiple equilibria, by studying how extremal equilibria move in reaction to a change in exogenous variables. For the purpose of our discussion, however, we shall focus on the case of unique equilibria.

be complements so as to guarantee that their increases are mutually reinforcing". The former property corresponds to increasing differences (between the endogenous and the exogenous variables, and more in general between two variables), whereas the latter qualifies the function to be maximized as supermodular.[12]

When a game is supermodular and characterized by increasing differences, an increase in the strategy of one player increases the marginal payoff of the strategy of the other players, inducing them to adjust their optimal choice upwards. This case, therefore, corresponds to upward-sloping reaction functions or, in the classification of Bulow, Geanakoplos and Klemperer (1985), strategic complementarity. Moreover, increasing differences between the endogenous variables and the exogenous variable implies that an increase in the exogenous variable increases the marginal payoff of the strategy of the players, with an upwards shift in the best-reply functions.

Increasing differences then can be easily turned into decreasing differences by reverting the sign of the adjustment or defining a new exogenous variable that is the negative of the original one. In this case, an increase in the exogenous variable induces a contraction in the endogenous one.[13]

We can borrow from the theory of monotone comparative statics two conditions, described in the statements of Theorem 5 and 6 in Milgrom and Shannon (1994) that, in our setting, fit the problem. The exogenous variable[14] is the number of firms $n$ whereas the endogenous variables are, depending on the model specification, the quantities $q_i$ or prices $p_i$. Then, we require the profit functions to be supermodular in the strategic variables and to display decreasing differences in the number of firms. Since we consider continuous and differentiable functions, the two conditions correspond to $\frac{\partial \Pi_i}{\partial a_i \partial a_j} > 0$ and $\frac{\partial \Pi_i}{\partial a_i \partial n} < 0$ for $i, j = I_n$, $i \neq j$, where $a_i$ describes firm $i$'s strategy, i.e. quantity or price. Moreover, in order to focus on the comparative statics, we take for granted that an equilibrium exists and is unique, by assuming that the profit function is strictly quasi-concave in the choice variable and that the best-reply slope meets the contraction mapping requirement.

Starting with the Cournot case, a first problem arises since in the traditional description competition is in strategic substitutes, and the game is submodular rather than supermodular.[15] A way out of this problem borrows from an early intuition in Novshek (1985) and is developed in Amir and Lambson (2000). Indeed, a notable property of the Cournot model is that the profits can be expressed as a function of own output $q_i$ and of the aggregate level of output of the other $n-1$ firms $Q_{-i} = \sum_{j \neq i} q_j$, i.e.

$$\Pi_i(q_i, Q_{-i}) = P\,(q_i + Q_{-i})\, q_i - C(q_i).$$

[12] See Vives (1999), Chapter 2. When the payoff functions are smooth and the strategy space of each firm and the exogenous parameters space are one-dimensional, supermodularity and increasing differences boil down to the condition that the second cross-partials between each firm's strategic variable and the other firm's strategic variable and with the exogenous parameter are positive.

[13] Increasing differences is a cardinal property and can be replaced by the ordinal Spence-Mirlees single-crossing property considered in Milgrom and Shannon (1994). When this property holds, if an increase in the choice variable is profitable when the exogenous variable is low it is still profitable when the exogenous variable is high, although it is not required, as in the case of increasing differences, that the profitability is higher in the latter case.

[14] Here for convenience we measure the number of firms $n$ as a continuous variable defined on the positive reals.

[15] While in a Cournot duopoly this issue is easily adjusted by describing one of the strategies as $-q$, transforming the setting into a supermodular game, with $n > 2$ firms this trick cannot be applied anymore.

Moreover, we can equivalently describe firm $i$'s strategy, rather than refer to the choice of its own output $q_i$, as the selection of a certain level of total output $Q$ for given output $Q_{-i}$ supplied by the competitors. In this alternative formulation

$$\widehat{\Pi}_i(Q, Q_{-i}) = P(Q)(Q - Q_{-i}) - C((Q - Q_{-i})). \tag{11.18}$$

Then,

$$\frac{\partial^2 \widehat{\Pi}_i}{\partial Q \partial Q_{-i}} = C''(Q - Q_{-i}) - P'(Q) = \Delta, \tag{11.19}$$

which corresponds to (11.1). Then, the condition $\Delta > 0$ implies the supermodularity of the modified Cournot game. Decreasing differences can be easily established by noting that when the other $n-1$ firms choose the same output $q$ then $Q_{-i} = (n-1)q$. Then, substituting in the first-order conditions for the choice of $Q$ in the modified Cournot problem we have:

$$\frac{\partial \widehat{\Pi}_i}{\partial Q} = P'(Q)(Q - (n-1)q) + P(Q) - C'(Q - (n-1)q). \tag{11.20}$$

Hence,

$$\frac{\partial^2 \widehat{\Pi}_i}{\partial Q \partial n} = q\Delta > 0 \tag{11.21}$$

when the game is supermodular. We conclude that the equilibrium total output $Q^*(n)$ is increasing in the number of firms. In a symmetric equilibrium $Q^*_{-i}(n) = \frac{n-1}{n} Q^*(n)$, and therefore the output of the firms other than $i$ is increasing in $n$ as well, since both terms $\frac{n-1}{n}$ and $Q^*(n)$ are positive and increasing in $n$. Moreover, since firm $i$'s best reply in the original Cournot problem is downward sloping and $Q^*_{-i}(n)$ is increasing in $n$, the individual output $q_i^*(n)$ is decreasing in the number of firms. Finally, since demand is bounded, when $n \to \infty$ we must have $Q^*(n) = nq^*(n)$ finite and therefore $\lim_{n\to\infty} q^*(n) = 0$. Then, given the first-order conditions of the original Cournot problem, $p^*(n) \to C'(q^*(n))$.

Our discussion offers a clear insight into the advantages of the techniques of monotone comparative statics. A single and general condition, $\Delta = C''(q_i) - P'(Q) > 0$, generates supermodularity of the modified Cournot problem and $Q^*(n)$ and $Q^*_{-i}(n)$ increasing in the number of firms, while the comparative statics on individual output $q_i^*(n)$ and the limiting competitive result on the price derive from the first-order conditions of the original Cournot problem. Interestingly, the condition $\Delta > 0$ includes elements of demand and costs, and both jointly define the relevant condition. This extends with respect to previous contributions that explored the properties of Cournot equilibria by making specific assumptions on costs or demand.[16]

[16] See Amir and Lambson (2000) for a general analysis of equilibria in Cournot games.

Turning to the models of product differentiation and price competition, in an $n$-firm oligopoly each one solves $\max_{p_i} p_i D_i(p_i, \mathbf{p}_{-i}; n) - C(D_i(.))$ where we emphasize that, differently from the homogeneous product case, the number of substitute products $n$ may directly enter into the expression of the demand for product $i$. Moreover, notice that in our symmetric environment we assume that all firms have the same cost structure, i.e. $C_i(D_i(.)) = C(D_i(.))$.

If

$$\frac{\partial^2 \Pi_i}{\partial p_i \partial p_j} = \frac{\partial D_i}{\partial p_j} + (p_i - C')\frac{\partial^2 D_i}{\partial p_i \partial p_j} - C'' \frac{\partial D_i}{\partial p_j}\frac{\partial D_i}{\partial p_i} > 0, \tag{11.22}$$

for any $i, j = I_n$, $i \neq j$, the game is in strategic complements, that is the condition for supermodularity is met. Then, the equilibrium prices fall in the number of firms if

$$\frac{\partial^2 \Pi_i}{\partial p_i \partial n} = \frac{\partial D_i}{\partial n} + (p_i - C')\frac{\partial^2 D_i}{\partial p_i \partial n} - C'' \left(\frac{\partial D_i}{\partial p_i}\right)^2 < 0.$$

Substituting the first-order conditions $p_i - C' = -\frac{D_i}{\partial D_i / \partial p_i}$ and rearranging we get:

$$\frac{\partial^2 \Pi_i}{\partial p_i \partial n} = \frac{\partial D_i}{\partial n} + \frac{p_i}{\varepsilon_p}\frac{\partial^2 D_i}{\partial p_i \partial n} - C'' \left(\frac{\partial D_i}{\partial p_i}\right)^2. \tag{11.23}$$

Differentiating the elasticity of demand with respect to $n$, we obtain:

$$\frac{\partial \varepsilon_p}{\partial n} = -\frac{\varepsilon_p}{D_i}\left[\frac{\partial D_i}{\partial n} + \frac{p_i}{\varepsilon_p}\frac{\partial^2 D_i}{\partial p_i \partial n}\right].$$

Hence, we can rewrite (11.23) as

$$\frac{\partial^2 \Pi_i}{\partial p_i \partial n} = -\frac{D_i}{\varepsilon_p}\frac{\partial \varepsilon_p}{\partial n} - C'' \left(\frac{\partial D_i}{\partial p_i}\right)^2. \tag{11.24}$$

Then, if (11.22) holds and (11.24)$< 0$ for all $i \in I_n$, the symmetric equilibrium prices fall in the number of firms. We can notice that the conditions (11.22) and (11.24) display a combination of demand and cost elements, a feature already noticed in the Cournot model. For instance, if the marginal costs are not decreasing and the demand elasticity is increasing in the number of firms, then the conditions are met.

Turning to our three examples of differentiated products models referred to in the different approaches, we have directly derived the equilibrium prices and observed that they fall in the number of firms. It is easy to check that the two conditions (11.22) and (11.24) are satisfied in our examples. Indeed, we assumed in the examples linear costs, i.e. $C'' = 0$. Moreover, it can be easily verified that when the other $n - 1$ firms set thesame price $p$, the elasticity of demand is increasing in $n$. Hence, the game features supermodularity and increasing differences and the prices fall in $n$.

### 3.4 Collusive Equilibria

We conclude our review of $n$-firms oligopolies by considering the case of collusive equilibria. We refer to the infinite horizon repeated game approach pioneered by Friedman (1971) and further developed in Fudenberg and Maskin (1986). Since we are considering symmetric oligopolies, we assume that the basic market interaction can be represented in each period $t = 1, .., T$ by a symmetric and stationary constituent game $\Gamma^t = \left\{I_n, a_i^t \in A, \pi_i^t = \pi(\mathbf{a}^t)\right\}$, where $I_n$ is the set of $n$ firms, $\mathbf{a}^t = (a_i^t, \mathbf{a}_{-i}^t)$ is the vector of actions chosen by firm $i$ and the other $n-1$ firms at time $t$, $A$ is the set of feasible actions and $\pi_i^t = \pi(\mathbf{a}^t)$ the per-period payoff. We further assume that $\Gamma^t$ has a unique symmetric Nash equilibrium $\widehat{\mathbf{a}} = (\widehat{a}, .., \widehat{a})$ that is Pareto dominated by other market configurations $A^{n*} = \left\{(a_i^*, \mathbf{a}_{-i}^*) \in A^n \left| \pi(a_i^*, \mathbf{a}_{-i}^*) \geq \pi(\widehat{\mathbf{a}}) \; \forall i \in I_n\right.\right\}$. Let $\overline{\mathbf{a}}^*$ be the maximal collusive symmetric configuration. The firms maximize the discounted sum of profits $V_0 = \sum_{t=0}^{T} \delta^t \pi_i^t$, where $\delta = 1/(1+r)$ is the discount factor. Each firm observes the other firms' actions with a one-period lag. The set of observed actions at time $t$, the history of the game, then, is $H^t = \left\{\mathbf{a}^0, .., \mathbf{a}^{t-1}\right\}$.

In what follows we concentrate on symmetric collusive equilibria, in the spirit of the overall section. Let $a^C$ be firm $i$'s collusive action, $\mathbf{a}^C \in A^{n*}$ be the vector of collusive actions, and $\pi^C = \pi(\mathbf{a}^C)$ the corresponding individual profits. Notice[17] that $\mathbf{a}^C \in \left[\overline{a}^*, \widehat{\mathbf{a}}\right]$; that is, the collusive symmetric allocation is in between the Nash equilibrium and the maximal collusive allocation. Further, define $a^P = \widehat{a}$ firm $i$'s action during the punishment phase, corresponding to the symmetric Nash equilibrium action in the constituent game, and $\pi^P = \pi(\widehat{\mathbf{a}})$ the punishment individual profits. Finally, let $a^D = \arg\max_{a_i} \pi(a_i, \mathbf{a}_{-i}^C)$ be firm $i$'s optimal deviation when the other firms stick to the collusive action, yielding $\pi^D = \pi(a^D, \mathbf{a}_{-i}^C)$. Our previous discussion implies that $\pi^P \leq \pi^C \leq \pi^D$ with strict inequalities if $a^C < \widehat{a}$. We focus on closed-loop grim-trigger strategies:

$$\sigma_i{}^* = \begin{cases} a_i^t = a^C & \text{for } t = 0 \\ a_i^t = a^C & \text{for } t > 0 \text{ and } H^t = \left\{\mathbf{a}^C, .., \mathbf{a}^C\right\} \\ a_i^t = a^P & \text{for } t > 0 \text{ and } H^t \neq \left\{\mathbf{a}^C, .., \mathbf{a}^C\right\} \end{cases}.$$

When $T = \infty$ (infinite horizon), given the strategy followed by the other firms and the stationarity of the repeated game each firm chooses to collude if the following incentive compatibility constraint holds:

$$V^C = \frac{\pi^C}{1-\delta} \geq V^D = \pi^D + \frac{\delta}{1-\delta}\pi^P.$$

[17] We implicitly assume in this notation that $\widehat{a} > a^*$, as is the case if the action corresponds to an output level. If, instead, the action corresponds to a price, the boundaries of the interval should be inverted.

Then, a well-known result (Folk theorem) states that any allocation $\mathbf{a}^* \in A^{n*}$ can be implemented as a subgame perfect equilibrium in the game repeated indefinitely ($T = \infty$) if the following condition holds[18] for all firms $i \in I_n$:

$$\delta \geq \delta^* = \frac{\pi^D - \pi^C}{\pi^D - \pi^P}. \tag{11.25}$$

We can now address the key issue of whether the price(s), quantities and profits change, and in which direction, when the number of firms increases. To answer these questions we can consider two examples of market interaction when firms offer homogeneous products, characterizing the constituent game $\Gamma^t$ as a price-setting Bertrand game or a quantity-setting Cournot game. Let $\Pi^C = n\pi^C$ be the total profits of the cartel. Then, in a Bertrand setting $\pi^C = \Pi^C/n$, $\pi^D = \Pi^C$ and $\pi^P = 0$. Then, the condition (11.25) boils down to

$$\delta \geq \delta^*(n) = \frac{n-1}{n}$$

that is increasing in $n$. In other words, if the basic market interaction takes the form of Bertrand competition with homogeneous products, the incentive compatibility constraint becomes tighter the larger the number of firms. The economic intuition is pretty simple: a cartel with more members distributes the overall profits $\Pi^C$ among a larger number of participants, making the individual profits fall. Deviation and punishment profits, in this setting, are instead unaffected by the number of cartel members, making the condition for cartel sustainability harder to meet. We can further observe that the incentive compatibility constraint does not depend on the specific (symmetric) collusive allocation $\mathbf{a}^C$ the cartelists agree upon, since the gains from deviations are always proportional to the collusive profits. Then, a focal outcome would be to mimic the monopoly price $p^m$. Our prediction, then, is that the market price will be $p^m$ if the number of firms is $n \leq \frac{1}{1-\delta}$, falling to the Nash equilibrium price $p = c$ thereafter. To sum up, individual profits are strictly decreasing and the market price is weakly decreasing in the number of firms.

Turning to the Cournot model, we can identify a further element in the comparative statics. Indeed, in a Cournot setting the profits in the different states vary nonproportionally in the collusive allocation $Q^C$ the firms choose to implement. More precisely, the incentive compatibility constraint becomes tighter when the firms coordinate on an allocation, summarized by total output $Q^C$; that is, closer to the monopoly output $Q^m$. Hence, in a Cournot setting the critical discount factor $\delta^*(Q^C, n)$ is decreasing in the collusive output $Q^C$, whereas it continues to be increasing in the number of firms $n$.[19] The most collusive sustainable output in a symmetric cartel, $\overline{Q}^C$, then, is (weakly) increasing in the number of firms: if we start from $\overline{Q}^C = Q^m$, we can find a number of firms $n(Q^m, \delta)$ such that $\delta^*(Q^m, n(Q^m, \delta)) = \delta$. For a larger number of firms the cartel would collapse. However, the firms can coordinate on a less

[18] Notice that, having assumed symmetric firms, the incentive compatibility constraint and the threshold discount factors are the same for each and every firm.

[19] For instance, it is easy to show that, in the linear Cournot model when firms implement the monopoly output the critical discount factor is $\delta^* = \frac{n^2+2n+1}{n^2+6n+1}$ and is therefore increasing in $n$.

collusive output (i.e. $\overline{Q}^C > Q^m$) such that the incentive compatibility constraint is satisfied. In general, when (11.25) holds as an equality, for given $\delta$ we have

$$\frac{d\overline{Q}^C}{dn} = -\frac{\frac{\partial \delta^*}{\partial n}}{\frac{\partial \delta^*}{\partial \overline{Q}^C}} \geq 0.$$

Hence, for $n \leq n(Q^m, \delta)$ the individual profits are decreasing in $n$ while the market price is $p^m$, whereas for $n > n(Q^m, \delta)$ both the individual profits and the market price are falling in $n$.

Finally, an informal argument that is often put forward refers to the impact of a larger and larger cartel on the monitoring activity that the firms have to perform to prevent cheating. It seems realistic that such activity may take more time the higher the number of firms to be scrutinized. We can include this further argument by considering that the length of the period in the repeated game framework may increase when more firms participate in the agreement and have to be monitored. A longer period, then, corresponds to a lower discount factor $\delta$, leading to a decreasing relationship $\delta(n)$. In this latter case, the incentive compatibility constraint would become $\delta^*(\overline{Q}^C, n) \geq \delta(n)$ and the effect of the number of firms on the maximal collusive allocation would be

$$\frac{d\overline{Q}^C}{dn} = -\frac{\frac{\partial \delta^*}{\partial n} - \frac{\partial \delta}{\partial n}}{\frac{\partial \delta^*}{\partial \overline{Q}^C}} \geq 0,$$

implying a stronger expansion in the cartel output when $n$ increases. Finally, when $n \to \infty$ both $\pi^P$ and $\pi^C$ tend to zero and the only sustainable output $\overline{Q}^C$ becomes the competitive one.

The effect of market size $S$ on collusive equilibria is twofold. Under constant marginal costs, market size and the scale of production multiplicatively affect the profits in each of the relevant states. Then, $S$ cancels out in the expression of the critical discount factor. In other words, under constant marginal costs the incentive compatibility constraints are unaffected by market size. On the other hand, the level of collusive equilibrium profits $\pi^C$ increase with market size.

To sum up, even the cartel equilibria display comparative statics properties similar to those already highlighted: the individual profits decrease, as does the market price, when the number of firms increases, and they tend to the perfectly competitive output when $n \to \infty$. Market size positively affects collusive profits while being neutral on the conditions for sustainability of the cartel. Moreover, the level of profits in a cartel are higher, for a given number of firms, than those of the oligopoly equilibria analyzed in the previous sections.

## 4 FREE ENTRY SYMMETRIC EQUILIBRIA

We can now endogenize the entry decision that determines how many of the $m$ potential entrants will decide to become active, sinking the entry cost $F$. In a symmetric setting, the post-entry profits depend on the number of active firms $n$ and are decreasing in it, as analyzed in detail across a wide set of models in the previous section. We can summarize the main findings in the relationship $\Pi(n, S, \gamma)$ between the individual profits, the number of firms $n$, the market size $S$ and the variable $\gamma$ that captures the intensity of price competition. This latter, therefore, can be referred to as the degree of substitutability among differentiated products, as in Section 3.2, as well as the mode of competition (price, quantity, collusion). Hence, the individual profits are decreasing in the number of firms, increasing in market size and decreasing in the intensity of competition.

The maximum number of firms $n^*$ in a symmetric free entry equilibrium (SFEE) is then captured by the two conditions:

$$\Pi(n^*, S, \gamma) \geq F \qquad (11.26)$$
$$\Pi(n^* + 1, S, \gamma) < F$$

The former ensures that all the active firms make non-negative net profits, whereas the latter implies that in a market equilibrium with $n^* + 1$ firms each one would not cover the sunk entry costs. Given the monotonicity of the individual profits in $n$ we can therefore write[20]

$$n^* = n(S, F, \gamma), \qquad (11.27)$$

where

$$\frac{\partial n^*}{\partial S} = -\frac{\partial \Pi / \partial S}{\partial \Pi / \partial n} > 0, \quad , \frac{\partial n^*}{\partial F} = \frac{1}{\partial \Pi / \partial n} < 0 \quad \text{and } n^*(\gamma') < n^*(\gamma) \text{ if } \gamma' > \gamma. \qquad (11.28)$$

Hence, our main predictions state that the number of firms in a symmetric free entry equilibrium is increasing in market size, decreasing in the sunk entry costs (economies of scale) and decreasing in the intensity of competition.[21] Interestingly, relaxed competition (a lower $\gamma$), as may arise if products are weak substitutes, or in the case that the industry is cartelized, is concomitant with an increased number of firms. We can further notice that if marginal costs are constant, market size multiplicatively increases the profits and therefore the number of firms depends on the ratio $F/S$ that captures the relevance of economies of scale with respect to market size. Then, an increase in market size, as it may derive from free trade agreements, leads to an increase in the number of firms and a fall in prices, making consumers better off.

The SFEE identifies the maximum number of firms sustainable given market fundamentals and the prevailing strategic behavior. More specifically, in differentiated products markets we have identified the maximum number of varieties sustainable in an SFEE, assuming that each variety requires to sink a cost $F$ to be produced, whereas the number of firms may be lower if some of them offer a portfolio of different varieties.[22]

[20] We consider here for convenience $n$ as defined on $\mathbb{R}^+$, ignoring the integer issue. Then, given the monotonicity of profits in $n$ the two conditions for an SFEE boil down to $\Pi(n^*, S, \beta) = F$.

[21] We express the relationship between $n^*$ and $\gamma$ to encompass both the case when $\gamma$ is defined over a compact interval (the substitutability parameter in the differentiated products models) and when it is a discrete index measuring the intensity of competition (as when comparing collusive and non-cooperative equilibria).

[22] This statement should be further qualified according to the different models of product differentiation. In general, if in a symmetric multi-product setting each firm offers $k$ varieties some cross-variety effects are internalized, and therefore the market price should be different (higher) than in the case of single-product firms. With higher individual profits in the symmetric $k$-varieties firms equilibrium some further entry should be profitable. Therefore, the number of multi-product firms should be larger than $n^*/k$, where $n^*$ is the SFEE number of single-product firms.

## 5 FREE ENTRY AND SOCIAL EFFICIENCY

Moving from the positive to the normative analysis, we are interested in evaluating whether the entry process leads to an optimal, excessive or insufficient number of firms. A frequent presumption is that guaranteeing conditions of free entry is desirable from a social point perspective. The analysis we have developed in the previous sections allows us to address this issue and to verify whether and under which conditions free entry leads to socially desirable outcomes. Spence (1976) and Dixit and Stiglitz (1977) have explored the issue in a monopolist competition set-up, finding that the number of varieties in a free entry equilibrium falls short of the social optimum. In a homogeneous product environment, instead, Von Weizsäcker (1980) and Perry (1984) established an opposite result, with too many firms entering with respect to the social optimum.

We discuss the social efficiency of SFEE following Mankiw and Whinston (1986) and Amir, De Castro and Koutsougeras (2014) and adopting the same two-stage game of the previous sections. We analyze a second-best welfare maximization problem where the social planner is assumed to control the number of firms but to be unable to affect or determine the behavior of the active firms once they enter. In the comparison of the equilibrium and the socially optimal number of firms we focus on the case when the fixed costs are non-negligible given market size, and the number of firms in either solution is finite.

We start with the case of homogeneous products and quantity competition and then move to a product differentiation and price competition environment. We can borrow from the analysis of symmetric market equilibria three conditions that we proved to hold under fairly general conditions in the Cournot model:[23]

1. In the symmetric equilibrium the individual output is decreasing in $n$: $q(n) > q(n')$ for $n' > n$.
2. Total output is increasing in the number of firms: $Q(n) = nq(n) < Q(n') = n'q(n')$ for $n < n'$.
3. The price cost margin is non-negative for any number of firms, and strictly positive for a finite number of firms: $P(Q(n)) - C'(q(n)) \geq 0$ for all $n$ and $P(Q(n)) - C'(q(n)) > 0$ for $n$ finite.

Given these features, the social planner maximizes total welfare by choosing the number of firms:

$$\max_n W(n) = \int_0^{Q(n)} P(s)ds - nC(q(n)) - nF \tag{11.29}$$

[23] In their paper, Mankiw and Whinston do not model explicitly the post-entry game and assume that certain features characterize the firm and aggregate pattern of the equilibrium strategies. We can, instead, explicitly refer to the properties of the equilibria developed in the previous sections. A similar approach can be found in Amir et al. (2014).

Let us define $n^W$ as the solution. Then, under 1–3, the SFEE number of firms is higher than the social optimum, that is $n^* > n^W$. The result can be easily proved by noting that the first-order conditions in problem (11.29) are:

$$\begin{aligned} W'(n) &= P\,(.)\left[n\frac{\partial q}{\partial n} + q(n)\right] - C(q) - nC'(q)\frac{\partial q}{\partial n} - F = \\ &= \Pi(n) - F + n\left[P\,(Q(n)) - C'(q(n))\right]\frac{\partial q}{\partial n}. \end{aligned} \tag{11.30}$$

Since in SFEE $\Pi(n^*) = F$, $\frac{\partial q}{\partial n} < 0$ by condition 1 and $P\,(Q(n^*)) - C'(q(n^*)) > 0$ for $n^*$ finite given condition 3, it follows that $W'(n^*) < 0$ and therefore $n^* > n^W$.

The economic intuition of the excessive entry result is straightforward: when an additional firm enters, it adds to the social welfare the profit $\Pi(n) - F$ but, at the same time, it steals output, and therefore profits, from the other firms, the last term in the derivative (11.30), second line. The *business-stealing effect*, captured by condition 1 above, creates a wedge between the private incentives of the entrant, and the social effect of entry, explaining why too many firms enter in an SFEE.[24] We can observe that when $F$ (or $F/S$) tends to zero then $P\,(Q(n^*)) - C'(q(n^*))$ and $\Pi(n^*)$ vanish, implying that an infinite number of firms enter in equilibrium and maximize welfare. In other words, the excessive entry result applies to the case of significant fixed costs and a finite number of firms, whereas it vanishes when fixed costs become negligible. A policy that expands markets, as it is a free trade approach, therefore can fix the excessive entry distortion and realign competitive market outcomes and social optimality.

The case of differentiated products adds an additional effect of entry on welfare, since more firms imply a larger set of varieties available to the consumers. Following Spence (1976) we capture this effect by assuming that the gross consumers' benefit is

$$CS(\mathbf{q}) = G\left[\sum_{i=1}^{n} f(q_i)\right] \tag{11.31}$$

where $\mathbf{q}$ is the vector of outputs, $f(0) = 0$, $f'(.) > 0$ and $f''(.) \leq 0$ for all $q_i \geq 0$ implies a preference for variety and $G'(z) > 0$, $G''(z) < 0$ for all $z \geq 0$ qualifies products as substitutes.[25] The social planner then solves the problem

$$\max_n W(n) = G\left[nq(n)\right] - nC(q(n)) - nF.$$

Contrary to the case of homogeneous products, when products are differentiated in general we cannot rank the number of firms in an SFEE and the socially optimal one. The reason is immediately evident from the first-order conditions of the problem:

$$\begin{aligned} W'(n) &= G'\left(nf'\frac{\partial q}{\partial n} + f\right) - C(q) - nC'(q)\frac{\partial q}{\partial n} - F = \\ &= \Pi(n) - F + n\left(G'f' - C'\right)\frac{\partial q}{\partial n} + G'\left(f - f'q\right) \end{aligned} \tag{11.32}$$

[24] Mankiw and Whinston show that, when the integer problem is taken into account, $n^* \geq n^W - 1$.

[25] Consumers' utility maximization implies that in a symmetric equilibrium the price is equal to $G'\,(nf(q))\,f'(q)$ and therefore the profits can be written as $\Pi = G'\,(nf(q))\,f'(q)q - C(q) - F$.

Condition (11.32) shows that an additional firm adds to total welfare the profits generated, $\Pi(n) - F$, and further affects total welfare with two additional terms. The first corresponds to the business-stealing effect already identified in the case of homogeneous products, and captures the fact that the new firm subtracts output and profits to the competitors, with a lower net social gain than the private firm and a bias towards excessive entry.

The last term is new and refers to the impact of an additional variety on consumers' surplus. $G'f$ is the marginal social effect of the new variety, whereas $G'f'q$ is the firm revenue. Since the firm does not internalize all the social benefit of the additional variety, the private incentives are lower than the social ones, leading to underprovision of varieties.

Without specific assumptions on preferences the two terms with opposite signs in (11.32) do not allow the identification of $W'(n^*)$ Therefore, we may have an excessive, insufficient or optimal number of firms entering in an SFEE. Under more specific assumptions on the utility function, we can generate examples where the ranking can be established. For instance, Dixit and Stiglitz (1977), using a CES utility function, obtain that the SFEE number of firms is lower than the welfare-maximizing one, reverting the case of excessive entry that characterizes a homogeneous product environment.

## 6 FREE ENTRY EQUILIBRIA WITHOUT SYMMETRY

Although a symmetric environment is a natural reference when analyzing long-run free entry equilibria, we may be interested in the effects of free entry in oligopoly markets when some kind of asymmetry has long-lasting effects. This may come from the existence of patents or other frictions in the adoption of process innovations that prevent the equalization of production techniques, from persisting advantages on the demand side coming from quality or brand image, to institutional features that affect the behavior of firms, such as, for instance, the coexistence of different ownership structures or the presence of state-owned firms. Since free entry equilibria suggest the pattern of adjustment when the entry process unfolds, asymmetric oligopolies are an interesting and relevant case to be addressed.

Once firms intrinsically differ, the number of firms is no longer a relevant statistic with which to describe, in a positive or normative sense, the long-run equilibria. However, many of the oligopoly models we have already considered in a symmetric setting share a particular property: that of being aggregative games, which allows us to deal easily with asymmetric environments.[26]

The profits of firm $i$ in an aggregative oligopoly game can be written as a function of a choice variable (action) $a_i$ and of the sum of the actions of *all* market participants $A = \sum_{j=1}^{n} a_j$; that is, $\Pi_i(a_i, A)$. A very simple illustration is the Cournot model already considered in Section 3.1. Setting $q_i = a_i$ we can write $\Pi_i(a_i, A) = P(A)a_i - C_i(a_i)$. We also recognize an aggregative structure in some of the models of product differentiation.[27] In the Singh and

[26] See Anderson, Erkal and Piccinin (2015) on free entry equilibria with aggregative oligopoly games.

[27] One can notice that address models with $n > 3$, such as the Salop circular road model described above, are not aggregative games, since the profits of each firm depend only on a subset of prices.

Vives (1984) linear model the prices enter additively in the demand function and therefore, setting $p_i = a_i$, the profits are written as:

$$\Pi_i(a_i, A) = (a_i - c)\frac{\alpha(1-\gamma) + \gamma A - \left[\gamma(n-1)+1\right] a_i}{(1-\gamma)\left[\gamma(n-1)+1\right]}.$$

Even the logit model shares the feature of an aggregative game, once we define $a_i = \exp(-\gamma p_i)$: the profits can be written as

$$\Pi_i(a_i, A) = (-\log(a_i)/\gamma - c_i)\frac{a_i}{A}.$$

To illustrate the main features of aggregative games, we use here the linear Cournot model $\Pi_i(q_i, Q) = (a - bQ - c_i)q_i$ as an example. The traditional setting describes the profit function as depending on own output and the aggregate of other firms' production $Q_{-i} = \sum_{j \neq i} q_j$; that is, $\Pi_i(q_i, Q_{-i}) = (a - b(q_i + Q_{-i}) - c_i)q_i$ and identifies the best reply

$$\widehat{q}_i(Q_{-i}) = \frac{a - c_i}{2b} - \frac{Q_{-i}}{2}.$$

Alternatively, following the aggregative setting we can identify the inclusive best reply first introduced by Selten (1970), where the optimal individual output is consistent with a given aggregate level of production:[28]

$$\widetilde{q}_i(Q) = \frac{a - c_i}{2} - Q\,.$$

Notice that an equilibrium exists only if $\sum_{i=1}^{n} \widetilde{q}_i(Q) = Q$; that is, if the sum of the inclusive best replies has a fixed point.[29] Further we can define firm $i$'s profits, when it and all firms choose their inclusive best reply, as a function of total output $Q$:

$$\Pi_i(Q, \widetilde{q}_i(Q) = \Pi_i^*(Q) = \frac{(a - c_i - bQ)^2}{b} \qquad (11.33)$$

that is strictly decreasing in $Q$. The function (11.33) plays a fundamental role in the analysis of free entry equilibria when asymmetries are admitted. Indeed, it allows the mapping of the total equilibrium output – in general the aggregate $A$ – into the profits of the individual firms,

---

[28] One can notice that both expressions come directly from the first-order conditions

$$\frac{\partial \Pi_i}{\partial q_i} = a - c_i - b(q_i + Q_{-i}) - bq_i = 0.$$

[29] Anderson et al. (2015) introduce a set of assumption that guarantee the existence and uniqueness of an equilibrium in inclusive best replies. Moreover, under these assumptions the nature of interaction (strategic substitutability or complementarity) of the original best replies translates into an analogous feature of the inclusive best replies.

where therefore $Q$ replaces the number of firms as the key driver of equilibrium profits in an asymmetric setting.

Continuing with our Cournot example, a free entry equilibrium (FEE) can be defined as a set of quantities $\left\{(q_i^*)_{i\in I}\right\}$ and a set of entrants $I \subseteq I_m$, where $I_m$ is the set of all $m$ potential entrants, such that

$$\Pi_i(Q_I^*) \geq F_i \text{ for all } i \in I \qquad (11.34)$$
$$\Pi_j(Q_I^* + q_j^*) < F_j \text{ for all } j \notin I$$

where $Q_I^* = \sum_{i\in I} \widetilde{q}_i(Q_I^*)$ is the aggregate output of the entrants $I$. Notice that we are not imposing symmetry in gross profits $\Pi_i$ or in the sunk costs $F_i$. As a final step, it is often argued that the marginal entrant in a free entry equilibrium gains zero profit, a condition that is shared by all firms in a symmetric equilibrium. Anderson et al. (2015) assume that, among the potential entrants, there is a subset $e \subset I_m$ of *symmetric* marginal firms[30] with identical profit function $\Pi_i = \Pi_e(q_i, Q)$ and entry cost $F_i = F_e$ for all $i \in e$. Some of these marginal firms may be active, belonging to the set $e_a \subset I$.

In a zero-profit free entry equilibrium (ZPFEE) a nonempty set of marginal firms $e_a$ is active and gains zero profit. More formally, a ZPFEE is an FEE with a set $I$ of active firms such that $e_a \subset I$ and $\Pi_i = \Pi_i^*(Q_I^*) = F_i$ for all $i \in e_a$, where $\Pi_i^*(.)$ is given by (11.33). The existence of a fringe of symmetric active marginal entrants allows the combination of the zero-profit condition of the marginal firms with a unique level of aggregate output $Q_I^*$ and with a variety of profit levels of the inframarginal (asymmetric) firms. Indeed, since $\Pi_i^*(Q)$ is decreasing in $Q$, from the zero-profit condition for the active marginal firms we obtain $Q_I^*$, and this latter determines the profits of the other inframarginal firms $\Pi_i^*(Q_I^*)$. The number of active marginal firms $e_a$ is then adjusted through the entry process to find the ZPFEE.

To illustrate these properties it is interesting to analyze how the ZPFEE varies when exogenous changes in the set of inframarginal firms occur, modifying their profit structure and, consequently, the optimal output they deliver to the market. Let us consider an exogenous shock that affects a subset $I_C$ of inframarginal firms (the changed firms), such as, for instance, a process innovation, or a merger, or a privatization, while leaving the other inframarginal firms in subset $I_U$ (the unchanged firms) unaffected. Hence, in the initial state, $I = I_C \cup I_U \cup e_a$.

Then, after the shock the set of active firms in a ZPFEE moves from $I$ to $I'$. All the changed and unchanged inframarginal firms remain active both before and after the shock, i.e. $I_C = I'_C$ and $I_U = I'_U$. The adjustment to the new ZPFEE works through a variation in the set of active marginal entrants: $e_a \neq e'_a$. Since $e'_a \neq \emptyset$ in the new equilibrium, $\Pi_i = \Pi_e^*(Q_{I'}^*) = F_e$ must hold for $i \in e'_a$ and therefore total output remains the same; that is, $Q_{I'}^* = Q_I^*$. Consequently, the profits of the unchanged inframarginal firms do not vary. Hence, for instance, a reduction in the marginal cost of the changed inframarginal firms $I_C$ leads them to produce more in the new ZPFEE, whereas the unchanged inframarginal firms $I_U$ maintain the same level of production. Since total output does not vary, the set of marginal firms shrinks as does their

[30] A possible justification of this key assumption rests on the following argument. The industry is populated by a set of larger firms that display rich strategies and, through them, are able to introduce some competitive advantage, i.e. asymmetry. Then, there is a fringe of small firms (the marginal entrants) that are not strategically sophisticated and adopt a standard and similar technology and are therefore less efficient than the larger ones.

overall production, adjusting the larger production of the changed inframarginal firms and maintaining total output $Q_I^*$ at the initial level.[31]

This property of the ZPFEE also encompasses the case of the "aggressive leaders" in Etro (2006), where one firm, the leader, is the inframarginal agent and the other symmetric firms, the followers, belong to the active marginal entrant group $e_a$. A change in the profits of the leader, for instance due to some investment, as Etro (2006, p. 150) writes, "does not affect the equilibrium strategies of the other firms, but it reduces their equilibrium number". Interestingly, in this setting with an endogenous number of followers, if the investment increases the marginal profit of the leader, this latter has an incentive to over-invest, no matter whether competition is in strategic complements or substitutes. Indeed, if the market equilibrium output does not change with its investment, whereas its market share and profits increase, the leader will over-invest. At the limit, if the investment is not costly, the leader has the incentive to produce more than the usual Stackelberg leader's output and to monopolize the market, preventing the entry of the followers.

This result of generalized over-investment is strikingly different from what happens when the number of followers (entrants) is given and exogenous. In the taxonomy proposed by Fudenberg and Tirole (1984), when the investment increases the marginal profit, the leader over-invests (top dog) if competition is in strategic substitutes but it under-invests (puppy dog) when it competes in strategic complements.

Aggregative games also greatly simplify the normative analysis of asymmetric environments. Starting with the case of homogeneous products, we observe that consumers' surplus depends on aggregate output only,[32] i.e. $CS = CS(Q)$, with $CS(0) = 0$, $CS'(.) > 0$ and $CS''(.) \leq 0$ for all $Q \geq 0$. Then, when a shock affects a subset of inframarginal firms while leaving total output $Q_I^*$ unchanged, consumers' surplus also does not vary. The only impact on social welfare comes from the variation in profits of the changed inframarginal firms $I_C$. Indeed, the profits of the unchanged inframarginal firms $I_U$ do not vary and the change in the number of active marginal firms from $e_a$ to $e'_a$ does not affect welfare, since they gain zero profits. We conclude that if a shock induces a profitable adjustment in a subset of firms and a change in their market shares, the only effect on welfare comes from the variation in the profits of the affected firms, quite in contrast with the impact in the short run when the number of firms does not vary.

To appreciate the result, let us consider the welfare impact of a merger between two firms absent any efficiency gain. The short-run effects are well known in the IO literature: the merged entity internalizes the negative externalities and contracts output; the outsiders react by expanding their production. The net effect is a fall in total output, consumers' surplus and total welfare, an increase in outsiders' profits and, in the case of constant returns to scale, a fall in insiders' profits.[33]

Once we consider entry and ZPFEE, however, the effects change significantly. Since additional active marginal firms enter in reaction to the short-run adjustments, total output,

[31] This neutrality outcome recalls a case of a competitive market where a fringe of identical firms with constant marginal costs makes the supply curve flat at some price $p$. Any efficiency improvement of the inframarginal firms affects the supply curve but the market equilibrium is always determined by $(p, D(p))$. The reduction in costs, then, is cashed in by the inframarginal firms as increased profits. These latter, in a sense, are Ricardian rents.

[32] This is true if firms' activities do not entail any externality, such as, for instance, different levels of pollution. If this were the case, the composition, and not only the total level of output would matter from a welfare point of view. In our discussion we are assuming that these composition effects do not arise in a homogeneous product market.

[33] See Salant, Switzer and Reynolds (1983).

consumers' surplus and outsiders' profits ($I_U$ and $e_a$) do not change. The insiders' ($I_C$) profits, due to their output contraction, are weakly lower. If, however, the merger allows the realization of efficiencies, insiders' profits, as well as their incentive to merge, increase, as does total welfare. This result brings with it a strong policy implication in favor of lifting ex ante merger control and authorization policies. Indeed, since the long-run private and social effects of a merger coincide, if private firms have an incentive to merge, then social welfare will rise, whereas socially damaging mergers would never be implemented given the lack of private incentives.[34]

In the welfare analysis of homogeneous product markets, we assumed that consumers' surplus depends only on total output but not on its allocation among the active firms. Moving to a differentiated products environment a similar assumption may be more problematic. Indeed, Anderson et al. (2015) show that in aggregative oligopoly games with differentiated products, a reallocation of a given aggregate among the different varieties, although neutral on the ZPFEE conditions, may affect total surplus and welfare. In other words, it may be that consumers' surplus not only depends on the aggregate, but also on its composition.

They show that the dependence of consumers' surplus on the aggregate only still persists with differentiated products if the demand functions satisfy the independence of irrelevant alternatives (IIA) property; that is, if the ratio of any two demands depends only on their own prices and not on the prices of other, unconsidered, alternatives. Notably, the logit model, as well as the demand functions derived from the CES utility function, satisfy the IIA and therefore the corresponding oligopoly game is not only aggregative, but also allows the expression of consumers' surplus as a function of the sum of the prices only.[35]

## 7 ENDOGENOUS SUNK COSTS AND PERSISTENT CONCENTRATION

The entry decision in the previous sections involved sinking a fixed set-up cost $F$ that was related to some initial indivisible investment. We have not further specified the nature of these outlays. Assuming that the level of the sunk cost $F$ is an exogenous parameter with respect to the entry and market strategies may be explained referring to technology (e.g. investment in a minimum efficient scale plant) or institutions (e.g. the payment of a license fee needed to operate). The sunk cost may vary, allowing us to extrapolate comparative statics properties, but for reasons orthogonal to the market strategies adopted by the active firms once entered. In this sense we can label the environments considered so far as characterized by *exogenous sunk costs.*

In this setting, the amplitude of the sunk costs $F$ compared to the size of the market $S$ is a fundamental driver in determining the maximum number of firms sustainable in a free entry equilibrium. The limiting case, when $F$ becomes negligible with respect to $S$, leads

---

[34] Notice that the hands-off policy implications of free entry on merger control are much stronger than the usual argument that low entry barriers may constitute a favorable element when analyzing a merger. In this latter case easy entry conditions may mean that pros are balanced with the cons of enhanced market power in the evaluation of a merger. In the ZPFEE case, free entry is instead sufficient to generate mergers only when they are welfare enhancing.

[35] It should be stressed that aggregative product differentiation models do not necessarily satisfy the IIA, as is evident, for instance, considering the linear model drawn from Singh and Vives (1984). In this case consumers' surplus depends not only on the aggregate price but also on its composition.

to convergence to a competitive equilibrium with an infinite number of firms, vanishing externalities and price converging to the marginal cost.

Although this paradigm can apply to several industries, there are many other sectors where a relevant part of the sunk costs arise due to specific market strategies of the firms, which in general we may connect to the effort of attaining a competitive advantage and market leadership. This is the case with investments in advertising that enhance the perceived quality of the product, or with R&D expenditures aimed at improving the efficiency of the technology or the quality of the products.[36] Similar effects take place in industries such as media and entertainment, where market leadership can be reached by securing specific, nonreproducible inputs such as, for instance, talent and premium content.[37] In all these examples, a competitive advantage is reached through enhanced efforts and, therefore, higher sunk costs. We label this second class of economic environments *endogenous sunk costs.*

When sunk costs react to market incentives, we may expect that the entry process, which is constrained by the need to repay all the sunk outlays, is affected. Indeed, market size, which drives the tendency to fragmentation in an exogenous sunk cost industry, has the additional effect of increasing the marginal return to market dominance, incentivizing leadership and endogenous sunk costs. A central result of the endogenous sunk cost case claims that if the incentives for effort are sufficiently high, an increase in market size does not lead to an increasingly fragmented market structure. There exists an upper bound to fragmentation such that, even at the limit, large firms and concentration persist.

We illustrate this result through a very simple model due to Schmalensee (1992)[38] that conveys the main ideas and intuition. In this setting we set the price $p > c$ fixed and concentrate on the investment in advertising $A_i$. The demand for product $i$ has a similar structure to that in discrete choice models: $D_i(A_i, A_{-i}) = S * P_i(A_i, A_{-i})$ where $S$ is market size and $P_i$ firm $i$'s market share. Moreover,

$$P_i(A_i, A_{-i}) = \frac{A_i^\gamma}{\sum_{j=1}^{n} A_j^\gamma} \tag{11.35}$$

where $\gamma \in [0, 2]$ is a parameter that measures the mobility of consumers in reaction to advertising outlays. Notice that $\frac{\partial D_i}{\partial A_i} = \frac{\gamma}{A_i} P_i * (1 - P_i)$.

The profit function of firm $i$, then, is

$$\Pi_i(A_i, A_{-i}) = (p - c)S \frac{A_i^\gamma}{\sum_{j=1}^{n} A_j^\gamma} - A_i - F \tag{11.36}$$

[36] A pathbreaking contribution in the theory and empirical analysis of these industries is due to Sutton (1999, 1998), the former referring to advertising-intensive industries and the latter to R&D-intensive sectors. See also Sutton (2007) for a comprehensive review.

[37] See on these examples Motta and Polo (1997, 2003).

[38] A full-fledged model based on quantity competition and investments in quality can be found in Sutton (1991, and 2007, Appendix B).

where the last two terms refer to endogenous sunk costs in advertising ($A_i$) and exogenous sunk entry costs ($F$). In this setting there exists a symmetric Nash equilibrium in advertising levels

$$A^* = (p-c)S\gamma\frac{n-1}{n^2} \tag{11.37}$$

that is increasing in market size $S$ and in consumers' reactivity to advertising $\gamma$.

Plugging into the profit function and taking into account that in a symmetric equilibrium $P_i = 1/n$, the zero-profit condition can be rewritten as:

$$\frac{1-\gamma}{n^*} + \frac{\gamma}{n^{*2}} - \frac{F}{S(p-c)} = 0, \tag{11.38}$$

where $n^*$ is a solution of the above equation; that is, the SFEE number of firms.

The last term refers to exogenous sunk costs $F$ and vanishes as the size of the market $S$ increases indefinitely. However, the first two terms, which are directly related to the endogenous sunk costs in advertising outlays, present a different pattern: they do not depend on market size.[39]

When $\gamma \leq 1$, corresponding to consumers poorly reacting to advertising, and therefore a weak competitive pressure for market leadership, the single positive solution $n^*$ of (11.38) increases indefinitely in market size $S$, reproducing a pattern we already observed in pure exogenous sunk cost models. However, for $\gamma \in (1, 2]$ the incentives to invest in market leadership bite and advertising increases in larger markets, pushing up the endogenous sunk costs. In this latter case

$$\lim_{S\to\infty} n^* = \frac{\gamma}{\gamma-1}.$$

The entry process in this case is predominantly governed by the endogenous sunk costs, and the number of sustainable firms is bounded above for any market size, implying persistent concentration.[40] Moreover, the endogenous sunk costs tend to rise more quickly when consumers are more responsive to advertising, increasing concentration. Interestingly, in exogenous sunk costs environments more intense competition is associated with a lower $n^*$ and a more concentrated market, although these features dilute and vanish when the market size increases indefinitely. This pattern of higher concentration when competition is harsher, instead persists in endogenous sunk cost industries even with growing market size.

[39] This feature, literally speaking, depends on the specific set-up of the very simple model we adopt. However, a general property of this class of models is that when market size increases indefinitely, gross profits and investment costs once we reach a certain number of firms tend to increase at the same rate. In this case, when $S$ increases, boosting the gross profits, the incentives to invest in market leadership increase accordingly and the endogenous sunk costs increase at the same rate, preventing entry of additional firms.

[40] Shaked and Sutton (1983) identify a second case when the number of firms does not increase when market size rises. When firms offer different qualities $x_i \in [\underline{x}, \overline{x}]$ and the burden of quality improvements falls on fixed rather than marginal costs, price competition squeezes the margins. With relatively similar prices the demand for lower-quality products vanishes and a limited number of firms survives (finiteness property).

## 8 FRICTIONLESS ENTRY AND CONTESTABILITY

The general result in the endogenous and exogenous sunk costs cases claims that there exists a maximum number of firms sustainable in a free entry equilibrium, and that it is decreasing in the amplitude of the sunk costs $F$ compared with market size $S$. A concentrated market, in turn, is associated with noncompetitive mark-ups and allocative inefficiency. At the limit, when the economies of scale are particularly relevant, then we might find that only one firm can operate in the market: a case of natural monopoly. The firm will set the monopoly price $p^m$, being able to cover the high fixed costs with the monopoly margins. A second, symmetric entrant, pushing the market price down to $p(2) = P(Q(2))$ with its additional output, would make losses, since by definition in a natural monopoly it would be unable to cover the fixed costs. Then, there is a range of fixed costs such that the monopoly price is charged and no entry occurs. Similar cases can be generated where a small number of firms can be sustained in a free entry equilibrium.

The contestable markets approach[41] challenges this view, arguing that when entry is frictionless, structural monopoly or oligopoly environments do not lead to monopoly or oligopoly pricing and the associated allocative distortions. Indeed, potential competition may exert a sufficient corrective effect on the incumbent, inducing it to set a (second-best) efficient price to prevent temporary (hit-and-run) entry. Allocative efficiency is therefore ensured by (potential) competition even when economies of scale are so relevant to preventing actual competition.

This striking result re-establishes in a free entry environment a central feature of the Bertrand result, which claims that no relationship exists between the number $n > 1$ of active firms and the (socially efficient) oligopoly equilibrium. Indeed, as the exogenous sunk cost paradigm extends to the free entry case the Cournotian result of smooth convergence to competitive equilibria, the contestable market approach brings to the stage of the free entry story a Bertrand-type flavour.

It is now time to specify in more detail what we mean by frictionless entry. As a general point, the incumbent firm and the (potential) entrant are, under any respect, perfectly identical.

Since we are considering a natural monopoly, the first issue to address is the nature and amplitude of the fixed costs. Let us consider the following example. On the supply side, suppose that, in order to operate in the industry, it is necessary to bear a total investment $F$ for an indivisible capital good that provides production services over a time horizon $T$. Let us divide this total time into $t$ periods, whose length we are going to specify below. The incumbent firm $I$, then, has to cover a fraction $f = F/t$ of the fixed costs in each of the $t$ periods it is active in the market, and has variable costs $C_I(q_I)$. Let us consider the case $f \in (\Pi_2, \Pi^m]$, where $\Pi_2$ are the gross profits from duopoly and $\Pi^m$ the monopoly gross profits. Under this assumption the number of firms sustainable in the market is $n^* = 1$; that, is the market is a natural monopoly.

The potential entrant $E$, if it is willing to enter, has to pay $F = t * f$ to purchase the capital good. If, after one period, $E$ decides to exit, the residual value of the capital good is $(t-1) * f$.

[41] See Baumol, Panzar and Willig (1982). To ease the exposition we present here the case of a contestable natural monopoly. The authors generalize the contestable market approach to natural monopolies, showing that second-best efficient allocations arise also in these cases when entry is frictionless. The case of multi-product firms and economies of scope is a third, relevant extension of the analysis.

Let $\alpha \in [0, 1]$ be the fraction of the residual value that can be cashed back by reselling the capital good or by using it in other markets. This parameter measures the sunkness of the initial investment, with $\alpha = 1$ corresponding to the case when the capital good can be efficiently recovered after exit and $\alpha < 1$ to some level of sunkness. If $E$ enters and produces, its costs are $C_E(q_E)$. It is evident that, since the incumbent can efficiently use the capital good in the market for the entire length of its economic life, the entrant is in a symmetric position on the supply side only if $\alpha = 1$ and $C_E(q) = C_I(q)$.

Turning to demand, for a given price $p$ the entrant's demand is $D_E(p) \leq D_I(p)$ where the equals sign corresponds to a symmetric position towards the customers, who are uncommitted and can switch to the entrant if the price $p_E$ is more attractive than the incumbent's price $p_I$.

The timing of the game is as follows: at $s = 0$ the incumbent sets a price $p_I$ *that cannot be changed* for a period of length $T/t$; just after $p_I$ is set the entrant posts its own price $p_E$; once the two prices are set, the customers choose which of the two firms to patronize and are supplied immediately; at $s = T/t$, before the incumbent changes its price, the entrant exits and resells (or reuses) the capital good, collecting $\alpha(t-1)f$.

Once the contestable market story is unbundled, two key ingredients become evident:

1. There is no administrative restriction on entry, as licenses or authorizations.
2. Demand and supply quantities adjust instantaneously while price changes take time.

In this environment, the incumbent sets a (limit) price that prevents the temporary entry of the competitor:

$$\widehat{p}_I = \frac{C_E\left(D_E(\widehat{p}_I)\right) + f\left[\alpha + t(1-\alpha)\right]}{D_E(\widehat{p}_I)}. \tag{11.39}$$

If we compare (11.39) with the second-best Ramsey price

$$p^{sb} = \frac{C\left(D(p^{sb})\right) + f}{D(p^{sb})}$$

we can immediately notice that the limit price set by the incumbent is second-best efficient if three further conditions hold:

3. The entrant has access to the same technology as the incumbent, with no restrictions coming from patents or privately owned know-how: $C_E(q) = C_I(q)$; moreover, it can instantaneously change the level of production at the desired level.
4. The customers see the entrant and the incumbent as offering perfect substitutes and have no restrictions or costs in switching from one to the other: $D_E(p) = D_I(p)$.
5. The fixed indivisible investment is not sunk and the entrant recovers the residual value of the capital good entirely: $\alpha = 1$.

Under assumptions 1–5 potential competition is able to discipline the incumbent and induces second-best efficient outcomes in markets plagued by substantial economies of scale and concentration. Intuitively, perfect symmetry of the incumbent and the entrant and frictionless entry allow the market to be supplied, indifferently, by either of the two firms. If the incumbent commits to a profitable price, it is temporarily replaced by the entrant through undercutting. In this case, the identity of the provider changes for a period, although the market remains a monopoly. To avoid undercutting, the incumbent is forced to adopt the efficient limit price equal to the average costs. This remarkable result is derived under a set of specific assumptions, and can be evaluated both with respect to their empirical relevance and theoretical robustness. On theoretical grounds, the limit price expression (11.39) clearly shows that substantial departures from the second-best efficient price occur when any of the assumptions are weakened.

Turning to empirical relevance, the contestable market approach inspired the liberalization of the airline industry in the USA in the late 1970s.[42] In this sector a market corresponds to a route, and therefore the large investments in aircrafts are not specific to a market: the aircrafts can be moved to other routes or resold in an efficient market. Alternatively, the carriers can lease the aircrafts. The other fixed costs, check-in and handling services, are specific to airports, and therefore to the routes served. In the market reform, the airports, rather than the carriers, supplied these services, leasing them to the carriers on a variable cost basis. Hence, Assumption 5 of no sunkness seems consistent with the empirical data, as well as the access to the same technology (Assumption 3). Price stickiness may derive from contractual constraints on fares posted in advance (Assumption 2), and lifting authorizations was a key measure of the reform (Assumption 1). However, Assumption 4 was the Achilles' heel of the reform, since slots were assigned under grandfather rights, and the peak-hour more profitable ones remained in the portfolio of the incumbents. Moreover, in the years after the reform the carriers reorganized the routes from a spoke-to-spoke to a hub-and-spoke pattern, enhancing their dominant role in large hubs and achieving high load factors. With $D_E(p) < D_I(p)$, after an initial phase of turbulence, the incumbents were able to profitably prevent entries and maintain dominance in their key hubs.

Hence, although intellectually brilliant, the contestable market approach can hardly be considered a general theory of free market equilibria due to its lack of robustness. Although potential competition is an important ingredient in entry games, its impact on the behavior of active firms has to be carefully evaluated from an empirical point of view.

## 9 CONCLUSIONS

In this chapter we have reviewed the different branches of the IO literature that analyzes free entry equilibria and the endogenous determination of market structure. A recurrent theme is the assumption of symmetric firms, which in a long-run perspective can be justified when the friction of access to technology and the features of demand allow all firms to refer to a common set of best practice techniques and to exploit the possibility of (horizontal) product differentiation. In this perspective, a very rich class of oligopoly models is characterized by significantly similar comparative statics properties of the market prices, quantities and profits when the number of active firms increases. Two limiting cases emerge, perfectly competitive and the monopolistic competitive outcomes, when the number of firms increases indefinitely. The monotone comparative statics tools allow the identification of the general conditions behind these results. Long-run market structures under free entry are determined by a small set of elements referring to technology (economies of scale) and preferences (market size), with an additional ingredient related to strategies and the intensity of price competition.

[42] See Bailey and Panzar (1981) and Fawcett and Farris (1989).

Hence, the general result of free entry equilibria provides a solid theoretical foundation to the traditional approach of industrial economics based on the structure–conduct–performance paradigm.

The normative properties of free entry equilibria show that in a homogeneous product setting the business-stealing effect is the key element that creates a wedge between the private incentives and the social planner, determining an excessive number of firms. When product differentiation is introduced, however, an opposite externality leading to underprovision of varieties is also introduced, since the private incentives to enter do not include the benefits of an increased number of substitute products on consumers.

While symmetric market games are a useful reference for the long-run evolution of markets, asymmetric settings may be relevant both in the long run, when frictions persist, and as a starting point from which to study the evolution of market structure under free entry. It is important to notice that some form of symmetry is also maintained in this framework, which exploits the aggregative nature of many oligopoly models, by assuming that the (relatively inefficient) marginal entrants are all alike. The zero-profit condition on the marginal entrants, together with the aggregative nature of the market games, then generates unconventional long-run effects when a shock hits the active firms. Indeed, in the new free entry equilibria the total output remains unchanged, while its composition varies, with the change in output of the firms affected by the shock absorbed by an opposite variation in the number of marginal entrants. With these results, a hands-off policy is implied.

Endogenous sunk costs related to market strategies provide a different pattern of adjustment characterized by persistent concentration even in very large markets, in contrast with the tendency to fragmentation when sunk costs are exogenous. Finally, we review the attempt to establish efficient entry equilibria even in markets characterized by huge economies of scale and structural concentration, including natural monopolies, by assuming frictionless entry and giving a role to potential competition. The contestable markets paradigm refreshes the features of Bertrand competition in a free entry set-up, in contrast with the Cournotian paradigm of the exogenous sunk costs approach. Once again, symmetry plays a role, since the effectiveness of potential competition in disciplining dominant firms rests on the assumption that the entrants can perfectly replace the incumbent during their temporary raid in the market.

## REFERENCES


Amir R. (2003), Supermodularity and Complementarity in Economics: An Elementary Survey, mimeo.

Amir R. and Lambson V. (2000), On the Effects of Entry in Cournot Markets, *Review of Economic Studies*, 67: 235–54.

Amir R., De Castro L. and Koutsougeras L. (2014), Free Entry versus Socially Optimal Entry, *Journal of Economic Theory*, 154: 112–25.

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

Anderson S., Erkal N. and Piccinin D. (2015), Aggregative Oligopoly Games with Entry, mimeo.

Bailey E. and Panzar J. (1981), The Contestability of Airline Markets During the Transition to Deregulation, *Law and Contemporary Problems*, 49: 125–45.

Bain J.S. (1956), *Barriers to New Competition*, Cambridge, MA: Harvard University Press.

Baumol W., Panzar J. and Willig R. (1982), *Contestable Markets and the Theory of Industry Structure*, San Diego, CA: Harcourt Brace Jovanovich.

Bulow J., Geanakoplos J. and Klemperer P. (1985), Multimarket Oligopoly: Strategic Substitutes and Complements, *Journal of Political Economy*, 93: 488–511.

Cournot A. (1838), *Recherches sur les Principes Mathématiques de la Théorie des Richesses*, Paris: Chez L. Hachette.

D'Aspremont C., Gabszewicz J. and Thisse J. (1979), On Hotelling "Stability in Competition", *Econometrica*, 47: 1145–50.
Dixit A. and Stiglitz J. (1977), Monopolistic Competition and Optimum Product Diversity, *American Economic Review*, 67: 297–308.
Etro F. (2006), Aggressive Leaders, *Rand Journal of Economics*, 37: 146–54.
Fawcett S. and Farris M. (1989), Contestable Markets and Airline Adaptability under Deregulation, *Transportation Journal*, 29: 12–24.
Friedman, J. (1971), A Non-cooperative Equilibrium for Supergames, *Review of Economic Studies* 38 (1): 1–12.
Fudenberg D. and Maskin E. (1986), The Folk Theorem in Repeated Games with Discounting or with Incomplete Information, *Econometrica* 54 (3): 533–54.
Fudenberg D. and Tirole J. (1984), The Fat Cat Effect, the Puppy Dog Ploy and the Lean and Hungry Look, *American Economic Review*, 74: 361–66.
Häckner J. (2000), A Note on Price and Quantity Competition in Differentiated Oligopolies, *Journal of Economic Theory*, 93: 233–39.
Hotelling H. (1929), Stability in Competition, *Economic Journal*, 39: 41–57.
MacManus M. (1962), Number and Size in Cournot Equilibrium, *Yorkshire Bulletin of Economic and Social Research*, 14: 14–22.
MacManus M. (1964), Equilibrium, Number and Size in a Cournot Oligopoly, *Yorkshire Bulletin of Economic and Social Research*, 16: 68–75.
Mankiw G. and Whinston M. (1986), Free Entry and Social Efficiency, *Rand Journal of Economics*, 17: 448–58.
Manski C. (1977), The Structure of Random Utility Models, *Theory and Decisions*, 8: 229–54.
Milgrom P. and Roberts J. (1990), The Economics of Modern Manufacturing: Technology, Strategy and Organization, *American Economic Review*, 80: 511–28.
Milgrom P. and Roberts J. (1994), Comparing Equilibria, *American Economic Review*, 84: 441–59.
Milgrom P. and Shannon C. (1994), Monotone Comparative Statics, *Econometrica*, 62: 157–80.
Motta M. and Polo M. (1997), Concentration and Public Policy in the Broadcasting Industry, *Economic Policy*, 25: 295–334.
Motta M. and Polo M. (2003), Beyond the Spectrum Constraint: Concentration and Entry in the Broadcasting Industry, in Baldassarri M. and Lambertini L. (eds.), *Antitrust, Regulation and Competition*, Basingstoke, UK: Palgrave Macmillan.
Novshek W. (1985), On the Existence of Cournot Equilibrium, *Review of Economic Studies*, 52: 85–98.
Perry M. (1984), Scale Economies, Imperfect Competition and Public Policy, *Journal of Industrial Economics*, 32: 313–30.
Quandt R. (1956), A Probabilistic Theory of Consumer Behavior, *Quarterly Journal of Economics*, 70: 507–36.
Roberts J. and Sonnenschein H. (1976), On the Existence of Cournot Equilibrium Without Concave Profit Functions, *Journal of Economic Theory*, 13: 112–17.
Salant S., Switzer S. and Reynolds R. (1983), Losses from Horizontal Mergers: The Effects of an Exogenous Change in Industry Structure on Cournot-Nash Equilibrium, *Quarterly Journal of Economics*, 98: 185–99.
Salop S. (1979), Monopolistic Competition with Outside Good, *Bell Journal of Economics*, 10: 141–56.
Scherer F.M. (1980), *Industrial Market Structure and Economic Performance*, Chicago, IL: Rand McNally.
Schmalensee R. (1989), Inter-industry Differences of Structure and Performance, in Schmalensee R. and Willig R. (eds.), *Handbook of Industrial Organization, Vol. 2*, 951–1009, Amsterdam: North Holland.
Schmalensee R. (1992), Sunk Costs and Market Structure: A Review Article, *Journal of Industrial Economics*, 40: 125–34.
Selten R. (1970), *Preispolitik der Mehrproduktunternehmung in der statischen Theorie*, Berlin: Springer-Verlag.
Shaked A. and Sutton J. (1983), Natural Oligopolies, *Econometrica*, 51: 1469–83.
Shubik M. and Levitan R. (1980), *Market Structure and Behavior*, Cambridge, MA: Harvard University Press.
Singh N. and Vives X. (1984), Price and Quantity Competition in a Differentiated Duopoly, *Rand Journal of Economics*, 15: 546–54.
Spence M. (1976), Product Selection, Fixed Costs and Monopolistic Competition, *Review of Economic Studies*, 43: 217–35.
Sutton J. (1991), *Sunk Costs and Market Structure*, Cambridge, MA: MIT Press.
Sutton J. (1998), *Technology and Market Structure*, Cambridge, MA: MIT Press.
Sutton J. (2007), Market Structure: Theory and Evidence, in Amstrong M. and Porter R. (eds.), *Handbook of Industrial Organization, Vol. 3*, Amsterdam: North Holland, pp. 2303–64.
Vives X. (1990), Nash Equilibrium with Strategic Complementarities, *Journal of Mathematical Economics*, 19: 305–21.
Vives X. (1999), *Oligopoly Pricing*, Cambridge, MA: MIT Press.
Von Weiszäcker C. (1980), A Welfare Analysis of Barriers to Entry, *Bell Journal of Economics*, 11: 399–420.