NATIONAL RESEARCH UNIVERSITY
HIGHER SCHOOL OF ECONOMICS

Jacques-François Thisse, Philip Ushchev

MONOPOLISTIC COMPETITION WITHOUT APOLOGY

BASIC RESEARCH PROGRAM
WORKING PAPERS

SERIES: ECONOMICS
WP BRP 141/EC/2016

Monopolistic competition without apology*

Jacques-François Thisse[†] Philip Ushchev[‡]

June 14, 2016

Abstract

We provide a selective survey of what has been accomplished under the heading of monopolistic competition in industrial organization and other economic fields. Among other things, we argue that monopolistic competition is a market structure in its own right, which encompasses a much broader set-up than the celebrated constant elasticity of substitution (CES) model. Although oligopolistic and monopolistic competition compete for adherents within the economics profession, we show that this dichotomy is, to a large extend, unwarranted.

Keywords: monopolistic competition, oligopoly, product differentiation, the negligibility hypothesis

JEL classification: D43, L11, L13.

*We are especially grateful to a referee, C. Arkolakis and S. Redding for an impressive list of stimulating comments and suggestions. We thank M. Parenti for long and fruitful discussions on the subject matter. We also thank S. Anderson, D. Connolly, F. Di Comite, P. Fajgelbaum, A. Osharin, F. Robert-Nicoud, A. Tarasov and H. Vandenbussche for insightful discussions and remarks. The study has been funded by the Russian Academic Excellence Project '5-100'.

[†]Columbia University, Higher School of Economics (Russia), CORE-UCLouvain and CEPR. Email: jacques.thisse@uclouvain.be

[‡]NRU-Higher School of Economics (Russia). Email: ph.ushchev@gmail.com

1 Introduction

The absence of a general equilibrium model of oligopolistic competition unintentionally paved the way for the success of the constant elasticity of substitution (CES) model of monopolistic competition. This model, developed by Dixit and Stiglitz (1977), has been used in so many economic fields that a large number of scholars view it as virtually *the* model of monopolistic competition. The main thrust of this chapter is that monopolistic competition is a market structure in its own right, which encompasses a much broader set-up than what most economists believe it to be.

According to Chamberlin (1933), monopolistic competition is defined as a market environment in which a firm has no impact on its competitors (as in perfect competition) but is free to choose the output (or price) that maximizes its profits (as a monopolist). In other words, although one firm is negligible to the market, it is endowed with market power because it sells a differentiated product. For this to hold true, each firm must compete against the market as a whole or, to use Triffin's (1947) formulation, the cross-elasticity between any two varieties has to be negligible. According to the "Folk Theorem of Competitive Markets," perfect competition almost holds when firms are small relative to the size of the market. Hence, for a long time, economists debated heatedly whether Chamberlin's assumptions make sense. We will make no attempt to summarize this debate. Nevertheless, a few contributors raised fundamental questions that will be discussed later on.

We choose to focus on the two main approaches that have been developed to study monopolistic competition and explore the conditions under which these approaches lead to similar results. In the first, we consider an oligopolistic game in which firms compete in quantity (Cournot) or price (Bertrand). We then ask whether the sequence of Nash equilibria of these games converges to a competitive outcome when the number of firms grows indefinitely. If not, *monopolistic competition may be viewed as approximating a market in which strategic interactions among firms are weak.*

The second approach builds on Aumann (1964) who shows that the distribution of agents must be non-atomic for each agent to be negligible to the market. The same idea is applied to firms to account for Chamberlin's idea that a firm's action has no impact on its competitors. In other words, the supply side of the market is described by a continuum of firms whose mass is pinned down by the zero-profit condition. The next step is to check whether the Nash equilibria of these non-atomic games are identical to the competitive equilibria. When the answer is negative, monopolistic competition may be considered as a market structure per se. To put it differently, *monopolistic competition is the equilibrium outcome of a non-atomic game with an endogenous mass of players.*

Modeling monopolistic competition as a non-atomic game yields a framework easier to handle than standard oligopoly models while coping with general equilibrium effects, a task which hard to accomplish in oligopoly theory (Hart, 1985a). Furthermore, even though firms do not compete strategically, *general models of monopolistic competition are able to mimic oligopolistic markets with free entry within a general equilibrium framework*. This is in accordance with Mas-Colell (1984, p.19) for whom "the theory of monopolistic competition derives its theoretical importance not from being a realistic complication of the theory of perfect competition, but from being a simplified, tractable limit of oligopoly theory."

How product differentiation and consumer preferences are modeled has far-reaching implications for what is meant by monopolistic competition. In an influential review of Chamberlin's book, Kaldor (1935) objects to the idea that each firm is able to compete directly with all the others. According to Kaldor, firms are rooted in specific places. As a consequence, they have competitors that are close while the others are remote. Regardless of the number of firms in total, the number of firms competing for any particular consumer is small, so a decision made by one firm has a sizable impact only on the neighboring firms. Under these circumstances, monopolistic competition would not make sense. This argument is similar to the ideas developed by Hotelling (1929) and later on by Beckmann (1972) and Salop (1979). For these authors, *competition is localized*, meaning that a firm faces a limited number of direct competitors which operate in its vicinity. Lancaster (1979) puts forward the same idea in the context of a characteristics space where products are positioned while consumers have their own ideal varieties forming a constellation of points that belong to the very same space. These various strands of literature have given rise to a model of spatial competition with free entry. This model remains in the tradition of oligopoly theory: firm behavior is strategic because competition is localized while its global impact is diffused among firms through chain effects which link any two firms belonging to the same industry.

In contrast, if consumers have a love of variety, Kaldor's criticism ceases to be relevant. In this context, firms all compete together as they all strive to attract the entire population of consumers. This is why Chamberlin's model of monopolistic competition is henceforth associated with consumers who aim to consume many varieties, rather than consuming their ideal variety. After having attracted a great deal of attention in the 1930s, Chamberlin's ideas languished until Spence (1976) and, above all, Dixit and Stiglitz (1977) brought them back onto the scientific stage by proposing a model capable of being used in various economic fields. Spence developed a partial equilibrium setting, whereas the Dixit-Stiglitz model places itself in a general equilibrium context. Both modeling strategies are used in the literature. The former is more popular in industrial organization, whereas the latter is the workhorse of new trade and growth theories. This justifies our choice not to take a stance on choosing one particular strategy, but rather to deal with both.

On the production side, Chamberlin remained in the Marshallian tradition by assuming that firms face the U-shaped average cost curves. Since firms face downward sloping demand schedules and profits vanish under free entry, each firm produces at the tangency point of the demand and average cost curves. As a result, the equilibrium output level is smaller than the one that minimizes its unit costs, a claim dubbed the "excess capacity theorem." Under the severe conditions of the Great Depression, this was viewed as evidence that competition may generate a waste of resources. However, this argument overlooks the fact that, when consumers value product differentiation, a wider product range generates welfare gains that must be taken into account when assessing the (in)efficiency of monopolistic competition. Under these circumstances, there is a trade-off between scale economies associated with the production of varieties and the range of available varieties. This suggests the following question: *does the market over- or under-provide variety?*

This chapter reflects those various lines of research. However, its main emphasis will be on the models whose origin lies in the pioneering work of Dixit and Stiglitz (1977). Modeling monopolistic competition as a non-atomic game makes the corresponding market structure different from those studied in industrial organization. The upshot of the matter is that monopolistic competition encapsulates increasing returns and imperfect competition in a general equilibrium setting. Such a combination leads to a wide range of findings that may differ greatly from those obtained in a general competitive analysis, while permitting the study of issues that are hard to tackle within an oligopoly framework (Matsuyama, 1995).

The remainder of the chapter is organized as follows. First of all, there is a lot to learn from early contributions that are often disregarded in the modern literature. For this reason, Section 2 is devoted to those contributions, but we make no attempt to provide a detailed survey of what has been accomplished. In doing this, we follow the tradition of oligopoly theory and focus on partial equilibrium. Section 3 highlights the role of the negligibility hypothesis in the CES and linear-quadratic (LQ) models. Being negligible to the market, each firm treats parametrically market aggregates, which relaxes substantially the technical difficulties of working with imperfect competition in general equilibrium. In Section 4, we discuss a general set-up under the negligibility hypothesis and the "heroic assumption" that both demand and cost curves are symmetric (Chamberlin, 1933, 82). The focus is now on a *variable elasticity of substitution* (VES), which depends upon the individual consumption and mass of varieties. Under these circumstances, the VES model encompasses the whole family of models with symmetric preferences. Furthermore, the VES model of monopolistic competition is able to mimic key results of oligopoly theory. To a certain extent, we therefore find the dichotomy between oligopolistic and monopolistic competition unwarranted.

In Section 5, we make less heroic assumptions by recognizing that firms are heterogeneous. The literature on heterogeneous firms is huge and therefore, we are content to provide an overview of the main findings (Redding, 2011). In the spirit of the preceding sections, we depart from the CES which has taken center stage ever since Melitz's (2003) pioneering contribution. Section 6 is devoted to the classical question: does the market provide too many or too few varieties? As anticipated by Spence (1976), the numerous effects at work leave little hope of coming up with robust results, the reason being that the answer depends on the demand side properties. As a consequence, there is no need to discuss this question at length. Note, however, that the variety of welfare results casts some doubt on prescriptions derived from quantitative models that use CES preferences. Section 7 concludes and proposes a short research agenda.

A final comment is in order. This chapter is about the *theory* of monopolistic competition. This does not reflect any prejudice on our part, but dealing with econometric and applied issues would take us way beyond the scope of this chapter. We refer to De Loecker and Goldberg (2014) for a detailed survey of this literature.

2 Monopolistic competition as the limit of oligopolistic competition

There are (at least) three ways to model preferences for differentiated products. In the first, consumers are endowed with a utility $U(\mathbf{x})$ defined on the set X of *potential* varieties, which is continuous and strictly quasi-concave in $\mathbf{x}$ (see, e.g. Vives, 1999). It is well known that the convexity of preferences describes *variety-seeking* behavior. When preferences are symmetric, the convexity of preferences implies that a consumer has a love for variety, that is, she strictly prefers to consume the whole range of available varieties than any subset.

In the second approach, every consumer has one *ideal variety* and different consumers have different ideal varieties. In the spatial metaphor proposed by Hotelling (1929), a consumer's ideal variety is represented by her location in some geographical space (Main Street), while the variety provided by a firm is the location of this firm in the same space. Formally, the set X of varieties is defined by a metric space, such as a compact interval or a circle. Using a metric space allows one to measure the "distance" between any two locations, while the utility loss incurred by a consumer for not consuming her ideal variety is interpreted as the transport cost this consumer must bear to visit a firm, which increases with distance. Regardless of the number of available varieties, a consumer purchases a single variety. In this event, preferences are no longer convex, making it problematic to prove the existence of an equilibrium. However, ever since Hotelling (1929), it is well known that this difficulty may be obviated when there is a large number (formally, a continuum) of heterogeneous consumers.

A third approach was developed to account for taste heterogeneity, as in spatial models, but in a set-up that shares some basic features of symmetric models. This is achieved by using the random utility model developed in psychology and applied to econometrics by McFadden (1974). Interestingly, this approach looks at first sight like the second approach, but is isomorphic at the aggregate level to the first approach. Although discrete choice models have not been developed to study monopolistic competition per se, the results obtained under oligopoly can be used to study the market outcome when the number of firms is arbitrarily large.

The literature is diverse and, therefore, difficult to integrate within a single framework. In addition, some papers are technically difficult. In what follows, we use simple models to discuss under which conditions each of these three approaches leads to perfect or monopolistic competition when the number of firms grows indefinitely.

2.1 Variety-seeking consumers

2.1.1 Additive aggregate

There are two goods, a differentiated good and a homogeneous good. The homogeneous good x_0 is unproduced and used as the numéraire. The differentiated good is made available under the form of a finite number $n \geq 2$ of varieties, which are strong gross substitutes. Throughout this chapter, unless stated otherwise, each variety is produced by a single firm because firms seek to avoid the negative consequences of face-to-face competition, while each firm produces a single variety because there are no scope economies. The cost of producing x_i units of variety $i = 1, ..., n$ requires cx_i units of the numéraire.

There is a unit mass of identical consumers or, equivalently, a representative consumer who are each endowed with one unit of the numéraire. Like in mainstream oligopoly theory, consumers have quasi-linear preferences given by

$$U(\mathbf{x}) = \varphi\left(X(\mathbf{x})\right) + x_0, \tag{1}$$

where φ is twice continuously differentiable, strictly increasing, strictly concave over $\mathbb{R}_+$, and such that $\varphi(0) = 0$, while the sub-utility $X(\mathbf{x})$ maps the consumption profile $\mathbf{x} = (x_1, ..., x_n) \in \mathbb{R}^n_+$ into $\mathbb{R}_+$. The utility φ measures the desirability of the differentiated good relative to the numéraire. The concavity of $\varphi(\cdot)$ implies that the marginal utility of X decreases, and thus the marginal rate of substitution between X and x_0 decreases with X.

The sub-utility $X(x)$ is supposed to be symmetric and additive:

$$X(\mathbf{x}) \equiv \sum_{i=1}^{n} u(x_i), \tag{2}$$

where u is thrice continuously differentiable, strictly increasing, strictly concave over $\mathbb{R}_+$, and $u(0) = 0$. The concavity of $u(\cdot)$ amounts to assuming that consumers are variety-seekers: rather than concentrating their consumption over a small mass of varieties, they prefer to spread it over the whole range of available varieties. As a consequence, the elasticity of the sub-utility with respect to the per variety consumption level does not exceed one: $\mathcal{E}_{x_i}(u) \equiv x_i u'(x_i)/u(x_i) \leq 1$. The behavior of this elasticity plays a major role in shaping the welfare properties of monopolistic competition (see Dhingra and Morrow, 2016, and Section 8 of this chapter). Furthermore, it should be clear that the symmetry of lower-tier utility (2) means that the utility level is unaffected if varieties are renumbered.

Following Zhelobodko *et al.* (2012), we define the *relative love for variety* (RLV) as follows:

$$r_u(x) \equiv -\frac{xu''(x)}{u'(x)},$$

which is strictly positive for all $x > 0$. Very much like the Arrow-Pratt's relative risk-aversion, the RLV is a local measure of love for variety. Consumers do not care about variety if and only if $u(x_i) = x_i$, which means $r_u(x) \equiv 0$ for all $x > 0$. As the value of $r_u(x)$ grows, the consumer has a stronger love for variety. Therefore, how the RLV changes with the per variety consumption is crucial for the analysis of the equilibrium. Under the CES, we have $u(x) = x^{(\sigma-1)/\sigma}$ where σ, the elasticity of substitution between any two varieties, is a constant larger than 1; the RLV is given by $1/\sigma$. Other examples of additive preferences include the CARA (Behrens and Murata, 2007) and the addilog (Simonovska, 2015).

Let $\mathbf{p} = (p_1, ..., p_n)$ be a price vector. Utility maximization yields the inverse demand for variety i:

$$p_i\left(x_i, \mathbf{x}_{-i}\right) = \varphi'(X(\mathbf{x})) \cdot u'(x_i). \tag{3}$$

The Marshallian demands $x_i(\mathbf{p})$ are obtained by solving the following system of equations:

$$p_i = \varphi'\left[\sum_{j=1}^{n} u(x_j)\right] \cdot u'(x_i), \qquad i = 1, \ldots, n. \tag{4}$$

Combining (2) and (4) yields the Marshallian demand for variety i:

$$x_i(p_i, \mathbf{p}_{-i}) = \xi\left(\frac{p_i}{P(\mathbf{p})}\right), \tag{5}$$

where $\xi(\cdot) \equiv (u')^{-1}(\cdot)$, while $P(\mathbf{p})$ is the unique solution to the equation:

$$P = \varphi'\left[\sum_{j=1}^{n} u\left(\xi\left(\frac{p_j}{P}\right)\right)\right]. \tag{6}$$

Clearly, a price cut by firm i draws demand equally from all the other firms, which reflects the symmetry of preferences, while P plays the role of the Lagrange multiplier when the budget constraint is binding.

Bertrand competition. We consider a non-cooperative game in which the players are firms. The strategy of firm i is given by its price p_i and its payoff by its profits given by

$$\Pi_i^B(\mathbf{p}) = (p_i - c)x_i(\mathbf{p}) = (p_i - c)\xi\left(\frac{p_i}{P(\mathbf{p})}\right), \quad i = 1, ..., n. \tag{7}$$

A Nash equilibrium $\mathbf{p}^* = (p_1^*, ..., p_n^*)$ of this game is called a *Bertrand equilibrium*, which is symmetric if $p_i^* = p^B(n)$ for all $i = 1, ..., n$. It follows from (7) that $\Pi_i^B(\mathbf{p})$ is a function of p_i and $P(\mathbf{p})$ only. Therefore, the Bertrand game under additive preferences is an *aggregative game* in which $P(\mathbf{p})$ is the market statistic.

In the remainder of this chapter, we denote by $\mathcal{E}_z(f)$ the elasticity of a function $f(z)$ with respect to z. Differentiating (5) with respect to p_i and using (4) yields the price elasticity of the demand for variety i:

$$\mathcal{E}_{p_i}(x_i) = \frac{1 - \mathcal{E}_{p_i}(P)}{r_u[\xi(p_i/P)]}. \tag{8}$$

Since firm i's profit-maximizing markup is given by $m_i^B = 1/\mathcal{E}_{p_i}(x_i)$, (8) implies that *firm i's Bertrand-markup* may be written as follows:

$$m_i^B = \frac{1}{1 - \mathcal{E}_{p_i}(P)} \cdot r_u\left[\xi\left(\frac{p_i}{P}\right)\right], \tag{9}$$

where $0 < \mathcal{E}_{p_i}(P) < 1$ is shown to hold in Appendix.

Assume that firms treat P parametrically, so that $\mathcal{E}_{p_i}(P) = 0$. In this case, (9) boils down to $m_i^B = r_u(x_i) > 0$. Hence, even when firms are not aware that they can manipulate P, they price above marginal cost because their varieties are differentiated. When firms understand that they can manipulate P ($\mathcal{E}_{p_i}(P) > 0$), we have $1/[1 - \mathcal{E}_{p_i}(P)] > 1$. This new effect stems from the strategic interactions among firms through the market statistic P, which allows them to hold back their sales and to raise their profit. In summary, (9) highlights the existence of *two* sources of market power: *monopoly* power ($r_u(x) > 0$) and *strategic* power ($\mathcal{E}_{p_i}(P) > 0$).

We now show that the strategic power of firms vanishes as the number of firms unboundedly grows. Consider a symmetric Bertrand equilibrium $p_i^* = p^B(n)$ and find the equilibrium consumption $x^B(n)$, which is the unique solution to:

$$\varphi'[nu(x)] \cdot u'(x) = p^B(n).$$

The expression (A.1) in Appendix implies that

$$\lim_{n\to\infty} \mathcal{E}_{p_i}(P)\big|_{p_i = p^B(n)} = 0,$$

which means that *strategic interactions vanish in the limit.*

How does the monopoly term $r_u\left[x^B(n)\right]$ behave when n grows unboundedly? To check this, note that the budget constraint, together with $p^B(n) \geq c$, implies that $x^B(n) \leq 1/cn$. Therefore, when n tends to infinity, $x^B(n)$ converges to zero. Combining this with (9), we obtain:

$$\lim_{n\to\infty} m^B(n) = \lim_{n\to\infty} \frac{1}{1 - \mathcal{E}_{p_i}(P)|_{p_i=p^B(n)}} \cdot \lim_{n\to\infty} r_u\left[x^B(n)\right] = r_u(0).$$

Cournot competition. Firm i's profit function is now given by

$$\Pi_i^C(\mathbf{x}) = [p_i(x_i, \mathbf{x}_{-i}) - c]x_i = [\varphi'(X(\mathbf{x})) \cdot u'(x_i) - c]\, x_i,$$

A *Cournot equilibrium* is a vector $\mathbf{x}^*= (x_1^*, \ldots, x_n^*)$ such that each strategy x_i^* is firm i's best reply to the strategies $\mathbf{x}_{-i}^*$ chosen by the other firms. This equilibrium is *symmetric* if $x_i^* = x^C$ for all $i = 1, ..., n$. Using (3), we may restate *firm i's Cournot-markup* as follows:

$$m_i^C = r_u(x_i) + r_\varphi(X)\mathcal{E}_{x_i}(X), \tag{10}$$

where X is the market statistic (2).

As in the Bertrand game, there are two sources of market power, that is, strategic power and monopoly power. Under Cournot, the decomposition is additive, whereas the decomposition is multiplicative under Bertrand (see (9)). Despite this difference, both (9) and (10) show the importance of product differentiation for consumers through the value of the RLV.

Assume for simplicity that $r_\varphi(X)$ is bounded from above by a positive constant K; this property holds for the logarithmic and power functions. Since

$$\mathcal{E}_{x_i}(X) = \mathcal{E}_{x_i}(u) \cdot \frac{u(x_i)}{X},$$

and $0 < \mathcal{E}_{x_i}(u) < 1$ for all $x_i > 0$, it must be that

$$\lim_{n\to\infty} r_\varphi(X)\mathcal{E}_{x_i}(X)|_{x_i=x^C(n)} \leq \lim_{n\to\infty} \frac{K}{n} = 0.$$

Therefore, as in the Bertrand game, strategic power is diluted in an ocean of small firms selling differentiated varieties. As for the monopoly term in (10), the argument developed in the Bertrand case applies. It then follows from (10) that

$$\lim_{n\to\infty} m^C(n) = \lim_{n\to\infty} r_u\left[x^C(n)\right] + \lim_{n\to\infty} r_\varphi(X)\mathcal{E}_{x_i}(X)|_{x_i=x^C(n)} = r_u(0).$$

The limit of Bertrand and Cournot competition. The following proposition comprises a summary.

Proposition 1. *If there is $n_0 \geq 2$ such that a symmetric equilibrium exists under Cournot* *and Bertrand for all $n > n_0$, then*

$$\lim_{n\to\infty} m^B(n) = \lim_{n\to\infty} m^C(n) = r_u(0).$$

Since the strategic terms $\mathcal{E}_{p_i}(P)$ and $\mathcal{E}_{x_i}(X)$ converge to 0 when n goes to infinity, whether the limit of Bertrand and Cournot competition is perfectly competitive or monopolistically competitive is the same under both regimes and hinges on the value of $r_u(0)$. When $r_u(0) > 0$, a very large number of firms whose size is small relative to the market is consistent with the idea that firms retain enough market power to have a positive markup. To be precise, even when individuals face a very large number of varieties and consume very little of each variety, they still value diversity. It then follows from (8), that the price elasticity of a firm's demand is finite, which allows firms to retain monopoly power and to sustain a positive markup. On the contrary, when $r_u(0) = 0$, a growing number of firms always leads to a perfectly competitive outcome. Since both sources of market power vanish in the limit, the price elasticity of a firm's demand is infinite. Intuitively, consumers no longer care about diversity because their per-variety consumption is too low. In brief, *the love for variety must be sufficiently strong for monopolistic competition to emerge.*

2.1.2 Linear-quadratic preferences

In the case of two varieties, the LQ utility is given by:

$$U(x_1, x_2) = \alpha(x_1 + x_2) - \frac{\beta}{2}(x_1^2 + x_2^2) - \gamma x_1 x_2 + x_0, \tag{11}$$

where α, β and γ are three positive constants such that $\gamma < \beta$. In the case of $n > 2$ varieties, there are at least two different specifications of the LQ utility, which each reduces to (11) when $n = 2$:

$$U(\mathbf{x}) = \alpha \sum_{i=1}^{n} x_i - \frac{\beta}{2} \sum_{i=1}^{n} x_i^2 - \gamma \sum_{i=1}^{n} \sum_{j\neq i}^{n} x_i x_j + x_0, \tag{12}$$

or

$$U(\mathbf{x}) = \alpha \sum_{i=1}^{n} x_i - \frac{\beta}{2} \sum_{i=1}^{n} x_i^2 - \frac{\gamma}{n-1} \sum_{i=1}^{n} \sum_{j\neq i}^{n} x_i x_j + x_0, \tag{13}$$

where $\Sigma_i x_i^2$ is the Herfindahl-Hirschman index measuring the dispersion of the consumption profile $\mathbf{x}$, so that β measures the intensity of love for variety; α is the willingness-to-pay for the differentiated product, while γ is an inverse measure of the degree of substitution across varieties.

It is readily verified that under (12) the equilibrium markup tends to 0 when n goes to infinity, whereas the equilibrium markup is constant and positive under (13). In other words,

(12) leads to perfect competition and (13) to monopolistic competition.

To sum up, *whether the limit of oligopolistic competition is monopolistic or perfect competition hinges on preferences.* For example, under CARA or the LQ (12), we have $r_u(0) = 0$, and thus the limit of oligopolistic competition is perfect competition. In contrast, under the CES, we have $r_u(0) = 1/\sigma > 0$; the limit of the CES oligopoly model may thus be viewed as a "true" model of monopolistic competition.

2.2 Heterogeneous consumers: the spatial approach

In his review of Chamberlin's book, Kaldor (1935) argues forcefully that product locations in characteristics space, or firms' locations in the geographical space, mold market competition in a very specific way: whatever the total number of firms in the industry, each one competes more vigorously with its immediate neighbors than with more distant firms. Or, in the words of Kaldor (1935, p.390): "the different producers' products will never possess the same degree of substitutability in relation to any particular product. Any particular producer will always be faced with rivals who are nearer to him, and others who are farther off. In fact, he should be able to class his rivals, from his own point of view, in a certain order, according to the influence of their prices upon his own demand."

To develop the idea that some firms are close whereas others are distant, Kaldor used Hotelling's (1929) spatial metaphor. In spatial models of product differentiation, a consumer is identified by her "ideal" variety $s \in S \subset \mathbb{R}^n$, while the variety provided by firm i is denoted by $s_i \in S$. Hotelling (1929) uses the following spatial metaphor: firms and consumers are located in a metric space $S \subset \mathbb{R}^n$ where $d(s, s')$ is the physical distance between any two locations s and $s' \in S$. Because moving from one place to another involves a cost, space is sufficient to render heterogeneous consumers who are otherwise homogeneous. In a characteristics space à la Lancaster, $d(s, s')$ is the inverse measure of the degree of substitutability between the varieties s and s' (Lancaster, 1979). Besides the distance d, the other salient feature of the spatial model of product differentiation is the *transport rate* $t > 0$, which can be viewed as the intensity of preferences for the ideal variety, or the unit cost of travelling to a store. The taste mismatch of consumer s with variety s_i is now expressed by the weighted distance $td(s, s_i)$ between the consumer's ideal variety and firm i's variety. Even though the individual purchase decision is discontinuous - a consumer buys from a single firm - Hotelling (1929) finds it reasonable to suppose that firms' aggregated demands are continuous. Supposing that consumers are continuously distributed across locations solves the apparent contradiction between discontinuity at the individual level and continuity at the aggregated level.

Spatial models of product differentiation attracted a lot of attention in 1970s and 1980s. In this set-up, a consumer can purchase from any firm, provided she is willing to pay the transport cost, and thus the boundaries between firms are endogenous to firms' prices and locations. One of the earliest contributions is Beckmann (1972) who studied how firms equidistantly distributed over S compete to attract consumers who are uniformly distributed over the same space with a unit density. Each consumer buys one unit of the good up to a given reservation price, while transport costs are linear in the Euclidean distance. Accordingly, competition is *localized*, whereas it is *global* in models with symmetric preferences such as those discussed above. In the geographical space, the goods sold by any two stores are physically identical but differentiated by the places where they are made available. As a consequence, a consumer buys from the firm with the lowest full price, which is defined as the posted (mill) price plus the transport cost to the corresponding firm.

Assume that S is given by a one-dimensional market without boundary, e.g. the real line or a circle. In this case, firm i has only two neighbors located at a distance Δ on either side of s_i. When t takes on a high value, firm i is a local monopoly because it is too expensive for consumers located near the midpoint between firms $i-1$ and i to make any purchase. On the contrary, when t is sufficiently low, each firm competes with its two neighbors for the consumers located between them. As argued by Kaldor in the above quotation, the market power of a firm is restrained by the actions of neighboring firms. In other words, their (geographic) isolation avails them only local monopoly power, for firm i's demand depends upon the prices set by the neighboring firms $i-1$ and $i+1$:

$$x_i(p_{i-1}, p_i, p_{+1}) = \max\left\{0, \frac{p_{i-1} - 2p_i + p_{i+1} + 2t\Delta}{2t}\right\}. \tag{14}$$

If n firms are symmetrically located along a circle C of length L ($\Delta = L/n$), the equilibrium price is given by

$$p^*(n) = c + t\Delta = c + \frac{tL}{n}. \tag{15}$$

Hence, $p^*(n)$ decreases with the transport rate t because firms benefit less from their geographical separation. In the limit, when $t = 0$, distance does not matter anymore, implying that firms price at marginal cost. Thus, *the limit of the spatial model of monopolistic competition is perfect competition*. When fixed costs are taken into account, the free-entry equilibrium price can be shown to decrease when the market size is expanded by raising the consumer density over S.

Beckmann's paper went unnoticed outside the field of regional science. It is also worth mentioning the contributions made by Eaton and Lipsey (1977) who build a theory of market competition in a spatial economy. Again, despite the quality of the work, Eaton and Lipsey's contributions attracted a limited amount of attention. It was not until Salop (1979), who used the circular city model, that scholars in industrial organization started paying attention to spatial competition models, more or less at the same time as they came across Hotelling's

(1929) potential for new applications.[1]

Spatial models have proven to be very powerful tools because they account explicitly for the product specification chosen by firms, whereas the Chamberlinian and discrete choice models provide no basis for a theory of product choice and product design. Spatial models are appealing in two more respects. First, they capture consumer heterogeneity by means of a simple and suggestive metaphor, which has been used extensively to describe heterogeneous agents in several economic fields as well as in political science. Second, the spatial model of monopolistic competition is well suited for studying various facets of the market process, for example, by assuming that firms have a base product which is associated with the core competence of the firm. This product may be redesigned to match consumer requirements if the corresponding firm is willing to incur a cost that grows as the customized product becomes more differentiated from the base product. In this set-up, firms are multi-product and each variety is produced at a specific marginal cost. This problem may be studied by replacing Hotelling-like *shopping* models with *shipping* models, where firms deliver the product and take advantage of the fact that customer locations are observable to price discriminate across space (Macleod *et al.*, 1988; Eaton and Schmitt, 1994; Norman and Thisse, 1999; Vogel, 2011).

Unfortunately, spatial models become quickly intractable when they are cast in a general framework involving a non-uniform distribution of consumers and price-sensitive consumption of a variety. In those cases, showing the existence of a Nash equilibrium in pure strategies is problematic, especially when the location pattern is asymmetric.

2.3 Heterogeneous consumers: the discrete choice approach

There is a continuum of consumer types $\theta \in \mathbb{R}$. When n varieties are available, a consumer of type θ is described by a type-specific vector $\mathbf{e}^{\theta} = (e_1^{\theta}, ..., e_n^{\theta}) \in \mathbb{R}^n$ of match values with the varieties, which can also be viewed as the consumer's transport costs she bears to reach the varieties. Each consumer buys one unit of a single variety. More specifically, the indirect utility from consuming variety i by a θ-type consumer is given by

$$V_i^{\theta} = y - p_i + e_i^{\theta}, \tag{16}$$

where y is the consumer's income. Given prices, a consumer chooses her "best buy," that is, the variety that gives her the highest surplus net of its price.

We assume that each type θ is distributed according to the same continuous density $f(\cdot)$ and cumulative distribution function $F(\cdot)$. In this case, the market demand for variety i is given

[1] This is not yet the end of the story. Several results obtained by using the spatial competition model in industrial organization were anticipated by Vickrey in his *Microstatics* published in 1964.

by

$$x_i(\mathbf{p}) = \int_{-\infty}^{\infty} f(\theta) \prod_{k \neq i} F(p_k - p_i + \theta) \mathrm{d}\theta, \tag{17}$$

where, for any given type θ, $\Pi_{k \neq i} F(p_k - p_i + \theta)$ is the density of consumers who choose variety i at the price vector $\mathbf{p}$. The probability of indifference between two varieties being zero, each consumer buys a single variety. Because it embodies symmetry across varieties, (17) has a Chamberlinian flavor. However, even though consumer types obey the same distribution, they have heterogeneous tastes, for the probability that two types of consumers have the same match values is zero.

The Bertrand game associated with the demand system (17) can be studied along the lines of section 2.1. Anderson *et al.* (1995) show that a price equilibrium always exists if the density $f(\theta)$ is logconcave, while Perloff and Salop (1985) show the following result.

Proposition 2. *If either the support of the density $f(\theta)$ is bounded from above or $\lim_{\theta \to \infty} f'(\theta)/f(\theta) = -\infty$, then Bertrand competition converges to perfect competition as $n \to \infty$.*

Proposition 2 holds that the upper tail of the density of types is not "too" fat. Under this condition, as new varieties enter the market, it becomes more likely that two varieties are very close substitutes, implying the two producers get trapped into a price war. This in turn pulls down all prices close to the marginal cost. This can be illustrated in the case of a normal distribution where $f'(\theta)/f(\theta) = -\theta$, so that $p^B(n)$ tends to c. Otherwise, the tail is fat enough for a growing number of varieties to enter the market while remaining distant enough from each other, thus allowing firms to price above the marginal cost even when n is arbitrarily large. For example, when the match values are drawn from the Gumbel distribution, we obtain (Anderson *et al.*, 1992, ch.7):

$$\lim_{n \to \infty} p^B(n) = \lim_{n \to \infty} \left(c + \frac{n}{n-1} \varkappa \right) = c + \varkappa, \tag{18}$$

where $\varkappa$ is the standard-deviation of the Gumbel distribution up to $\sqrt{6}/\pi$. Since $\varkappa > 0$, the limit of Bertrand competition is thus monopolistic competition. In addition, as a higher μ signals a more dispersed distribution of tastes across consumers, (18) implies that *a more heterogeneous population of consumers allows firms to charge a higher price.* Alternatively, we may say that each variety has a growing number of consumers prepared to buy it even at a premium. Observe also that $p^B(n) = c$ for all n when consumers are homogeneous because $\varkappa = 0$, as in the standard Bertrand duopoly**.** Last, Proposition 2 is the mirror image of Proposition 1. The former states that individual preferences must be sufficiently dispersed for monopolistic competition to be the equilibrium outcome in a large economy, while the latter shows that a strong love for variety is needed for monopolistic competition to arise.

The model (16) can be easily extended to cope with the case where consumers have ideal varieties and a variable consumption level. This can be achieved by assuming that

$$V_i^\theta = \psi(y) - \phi(p_i) + e_i^\theta,$$

where both ψ and ϕ are increasing. A natural candidate investigated by Sattinger (1984) is obtained when $\psi(y) = \ln y$ and $\phi(p) = \ln p$. In this case, individual consumptions are variable and determined by price ratios, rather than price differences. Under the assumptions of Proposition 2, it is readily verified that $\lim_{n\to\infty} p^B(n) = c$. However, this ceases to hold under the Gumbel distribution where

$$\lim_{n\to\infty} p^B(n) = \lim_{n\to\infty} c\left(1 + \frac{n}{n-1}\varkappa\right) = c(1+\varkappa).$$

Again, taste dispersion allows firms to set prices higher than marginal cost.

Note, finally, that the symmetry of preferences may be relaxed by assuming the vector of match values is drawn from a multivariate distribution $F(x_1, ..., x_n)$, such as the probit where the covariance measures the substitutability between the corresponding two varieties. Though empirically relevant, it is hard to characterize the market equilibrium at this level of generality.

The above models are related to, but differ from, Hart (1985a). As in discrete choice models, Hart focuses on consumers who have heterogeneous tastes. However, unlike these models where consumers can switch between varieties, individual choices are restricted to a *given* and *finite* set of desirable varieties, which is consumer-specific. In equilibrium, every consumer chooses the quantity (which can be zero) to consume of each desirable variety. Hart (1985b) then shows that, in a large economy, a monopolistically competitive equilibrium exists if the taste distribution is sufficiently dispersed.

2.4 Where do we stand?

Summary. We have discussed three different families of models that describe preferences over differentiated products. In each case, the same conclusion emerges: the limit of Cournot or Bertrand competition may be monopolistic competition. Unlike what Robinson, Kaldor, Stigler and others have argued, a large number of firms need not imply perfect competition. As anticipated by Chamberlin, *when firms are many, their strategic power vanishes.* Nevertheless, *product differentiation may allow every firm to retain monopoly power over the demand for its variety* in an environment in which strategic considerations are banned.

Whereas the spatial models are very intuitive, the symmetric representative consumer models display a high degree of versatility. They both seem to belong to different worlds. This need not be so, however. The two families of models can generate the same market outcome. For this to happen, the market space of any variety must share a border with the market space of any other variety, while the distance between any two varieties must be the same. More specifically, if the number of characteristics is equal to $n-1$, where n is the number of varieties, each firm competes directly with every other firm (Anderson *et al.*, 1992, ch.4). To put it differently, a reconciliation between discrete choice theory, the representative consumer approach, and the spatial models of product differentiation is possible when *the number of product characteristics is sufficiently large relative to the number of product varieties.*

Syntheses. Two attempts at providing a synthesis of spatial and symmetric models are worth mentioning.[2] First, Chen and Riordan (2007) developed an ingenious synthesis of the spatial and variety-seeking models by using a *spokes network*. There are $\mathcal{N}$ potential varieties and a unit mass of consumers uniformly distributed over $n \leq \mathcal{N}$ spokes connected at the center $x = 0$ of the plane. Each spoke is the same length $\Delta/2$ and a single store is located at the endpoint $x = \Delta/2$. The distance between any two stores is thus equal to Δ. A consumer's ideal variety is described by her location along a particular spoke. Consumer variety-seeking behavior is captured by assuming that each consumer may purchase her second most-preferred variety chosen by nature with a probability equal to $1/(\mathcal{N}-1)$, so that this variety need not be available. When $n \leq \mathcal{N}$ varieties are available, the demand for variety $i = 1, ..., n$ is formed by consumers whose ideal variety is i and those who choose i as a second choice. Each firm has some monopoly power on its spokes, but competes symmetrically with the other firms. Hence, the model combines the above two transport geographies.

Assuming that all consumers buy their most- and second most-preferred varieties, the equilibrium price is given by

$$p^*(n) = c + t\Delta \frac{2\mathcal{N} - n - 1}{n-1}.$$

As the number n of varieties/spokes grows and reaches the value $\mathcal{N}$, the equilibrium price decreases toward $c + t\Delta > c$. Hence, regardless of the value of $\mathcal{N}$ *the limit of the spokes model is monopolistic competition*, whereas the limit of the circular model is perfect competition ($\Delta = 0$). This echoes what we have seen in the foregoing.

Second, Anderson and de Palma (2000) developed an integrative framework that links spatial and symmetric models. A consumer buys a fixed number $\bar{x}$ of units of the differentiated product (e.g., a given number of restaurant dinners per month) and has an entropy-like utility:

$$U(\mathbf{x}) = \sum_{i=1}^{n} x_i - \varkappa \sum_{i=1}^{n} x_i \log x_i + x_0 \qquad \text{s.t. } \sum_{i=1}^{n} x_i = \bar{x},$$

where the parameter $\varkappa > 0$ is a measure of the degree of differentiation across varieties. This specification corresponds to a special case of (1) in which $u(x) = x - \varkappa \log x$ and $\varphi(X) = X$. The *entropy* of a consumption profile $\mathbf{x}$ may be viewed as a measure of its dispersion. Therefore, the impact of the entropy term on the consumer's utility level tells us how differentiated varieties

[2] Other attempts include Hart (1985c), Deneckere and Rotschild (1992) and Ushchev and Zenou (2015).

are from the consumer's point of view.

Assume that identical consumers are uniformly distributed over the real line, while firms are equidistantly located over the set of integers $i = 0, \pm 1, ...$ Let $t > 0$ be the unit shopping cost. In this case, a consumer located at s has a *logit* demand given by

$$x_i(\mathbf{p}; s) = \bar{x} \frac{\exp\left[-(p_i + t\,|s - i|)/\varkappa\right]}{\sum_{k=-\infty}^{\infty} \exp\left[-(p_k + t\,|s - k|)/\varkappa\right]} > 0.$$

Competition is localized when $\varkappa = 0$. As $\varkappa$ rises from zero, market boundaries get blurred: a firm's spatial market is encroached on by its competitors; but this firm also captures customers from its rivals. In the limit, when $\varkappa \to \infty$ the market demand is equally spread across firms. For given prices, the individual demand for any variety is positive, as in Chamberlinian models, but decreases with the distance between the consumer and the variety-supplier. The market price is given by

$$p^*(\nu, t) = c + \varkappa \frac{(1+\phi)^2 \ln \phi}{2\phi(1+\phi) - \ln(1-\phi)^2},$$

where $\phi \equiv \exp(-t/\varkappa)$ is a measure of the degree of global competition in the market. Differentiating p^* with respect to t for any given $\varkappa$, or with respect to $\varkappa$ for any given t, shows that higher transport costs or a stronger love for variety lead to a higher price because the former weakens competition between neighboring firms while the latter means that varieties are more differentiated. When $\varkappa \to 0$, $p^*(\varkappa, t)$ boils down to the equilibrium price of the circular model, $p^*(0, t) = c + tL/n$, while $p^*(\varkappa, t)$ converges to $c + \varkappa n/(n-1)$ when $t \to 0$. Hence, when heterogeneity prevails along one dimension only, the equilibrium markups remain positive. However, $p^*(\varkappa, t)/\nu$ being homogeneous of degree zero, we have $p^*(0, 0) = c$ when heterogeneity completely vanishes.

3 The negligibility hypothesis in monopolistic competition

From now on, we assume that the supply side of the economy is described by a *continuum* of negligible firms whose *mass* is determined by free entry and exit. The *negligibility assumption* has several important implications. First, it captures the essence of the Chamberlinian idea of monopolistic competition, summarized in the following quote: "A price cut, for instance, which increases the sales of him who made it, draws inappreciable amounts from the markets of each of his many competitors, achieving a considerable result for the one who cut, but without making incursions upon the market of any single competitor sufficient to cause him to do anything he would not have done anyway." (Chamberlin, 1933, 83).

Second, because each firm treats the market as a given, it faces a given residual demand, very much like a monopolist. As a consequence, a firm can indifferently choose its profit-maximizing price and output. In other words, the negligibility assumption makes monopolistic competition immune to the difficult choice to be made between Cournot and Bertrand. Third, ever since Gabszewicz and Vial (1972), it is well known that the choice of a good produced by oligopolistic firms as the numéraire affects the equilibrium. Under the negligibility hypothesis, the choice of any particular variety as the numéraire has no impact on the market outcome. Last, one of the typical assumptions of monopolistic competition is that of free entry and exit. The role of this assumption is worth stressing. Indeed, positive (or negative) profits would affect individual incomes, hence firm demands. This feedback effect is precisely one of the major difficulties encountered when one aims to introduce oligopolistic competition in general equilibrium.

In what follows, we illustrate those ideas by discussing the CES and LQ models. Anderson *et al.* (2015) argue that these models can be viewed as aggregative oligopoly games in which "firms do not internalize the effects of their actions on the aggregate." To put it differently, the CES and LQ models may be viewed as sequential games in which a "fictitious Chamberlinian auctioneer" first chooses the value of the aggregate, while firms move second. As a result, the market outcome under monopolistic competition generates lower prices (or higher quantities) than those obtained under oligopolistic competition for the CES and LQ preferences.

3.1 The CES model of monopolistic competition

Even economists with minimal exposure to monopolistic competition have probably heard of the CES model. There is little doubt that this model has led to a wealth of new results (Matsuyama, 1995). For this reason, we find it useful to describe briefly how the CES model works.

3.1.1 The benchmark set-up

Firms and consumers. Labor is the only factor of production and is inelastically supplied in a competitive market; labor is chosen as the numéraire. There are L consumers endowed with y efficiency units of labor. They share the same CES utility function:

$$U = \left(\int_0^N x_i^{\frac{\sigma-1}{\sigma}} \mathrm{d}i \right)^{\frac{\sigma}{\sigma-1}}. \tag{19}$$

Maximizing U subject to the budget constraint yields the individual demand for variety i:

$$x_i = \frac{p_i^{-\sigma}}{\int_0^N p_i^{-(\sigma-1)} \mathrm{d}i} y, \qquad i \in [0, N]. \tag{20}$$

This expression implies that the supply of an infinitesimal interval of new varieties increases the denominator and, consequently, leads to a reduction in the demand for the existing varieties so long as their prices remain unchanged. In other words, the entry of new varieties triggers the "fragmentation" of demand over a wider range of varieties.

Let

$$P \equiv \left(\int_0^N p_i^{-(\sigma-1)} \mathrm{d}i \right)^{\frac{-1}{\sigma-1}}$$

be the CES *price index* of the differentiated good. The price index, which is the geometric mean of prices, decreases with the mass of varieties. Indeed, if a non-negligible range of new varieties Δ is added to the incumbent ones, we get

$$P \left(\int_0^N p_i^{-(\sigma-1)} \mathrm{d}i \right)^{\frac{-1}{\sigma-1}} > P_\Delta = \left(\int_0^{N+\Delta} p_i^{-(\sigma-1)} \mathrm{d}i \right)^{\frac{-1}{\sigma-1}}.$$

To put it differently, the price index falls as if competition among a larger mass of competitors were to lead to lower prices. In addition, as suggested by spatial models, the less differentiated the varieties, the lower the price index.

The market demand functions Lx_i may then be rewritten as follows:

$$X_i = p_i^{-\sigma} P^{\sigma-1} Ly. \tag{21}$$

Thus, a firm's demand accounts for the aggregate behavior of its competitors via the sole price index, and the game is aggregative. Since firm i is negligible to the market, it treats P as a parameter, whereas firms are price-makers, they are *price index-takers*. As a consequence, Triffin's condition $\partial X_i / \partial p_k = 0$ holds for all $k \neq i$. Furthermore, (21) implies that market demands are *isoelastic*, the price elasticity being equal to the elasticity of substitution σ. Finally, the market demand is still given (21) when individual incomes are redistributed because the demand X_i depends on the aggregate income only. As a consequence, the market demand is independent of the income distribution.

Firms share the same fixed cost F and the same constant marginal cost c. In other words, to produce q_i units of its variety, firm i needs $F + cq_i$ efficiency units of labor. Hence, firm i's profit is given by

$$\Pi_i(q_i) = (p_i - c)q_i - F. \tag{22}$$

Market equilibrium. A *symmetric free-entry equilibrium* (SFE) is a 4-tuple (x^*, q^*, p^*, N^*), which satisfies the following four conditions: (i) no firm can increase its profit by deviating from q^*; (ii) x^* maximizes a consumer's utility subject to her budget constraint; (iii) the product market clearing condition

$$q^* = Lx^*$$

holds; (iv) the mass of firms is pinned down by the zero-profit condition (ZPC). The Walras Law implies that the labor market balance

$$N^* \cdot (F + cq^*) = Ly$$

is satisfied.

The FOC shows that the equilibrium price is given by (the SOC is satisfied):

$$p^* = \frac{\sigma}{\sigma - 1}c,$$

which increases when varieties get more differentiated, as in the various models discussed in Section 2. The markup is constant and equal to

$$\frac{p^* - c}{p^*} = \frac{1}{\sigma}. \tag{23}$$

In other words, firm markups are the same in large/small/rich/poor countries, the reason being that firm demands are isoelastic. In game-theoretic terms, this means that firms have a dominant strategy - the reaction functions are flat, - a result that probably explains the lack of interest among researchers in industrial organization for the CES model of monopolistic competition.

A constant markup runs against the conventional wisdom that asserts that entry fosters lower market prices. The markup (23) is also independent of shocks on marginal cost and market size, which contradicts a growing number of empirical studies (De Loecker and Goldberg, 2014). Evidently, markups are variable under the CES when firms operate in an oligopolistic environment (d'Aspremont *et al.*, 1996). However, adopting this approach implies losing the flexibility of monopolistic competition.

The above criticisms need qualification, however. Even if the equilibrium price remains unchanged when the mass of firms increases, the consumption of the differentiated good is fragmented over a wider range of varieties. This in turn implies that each firm's profits go down. In other words, we come back, albeit very indirectly, to a kind of competitive effect as the entry of new firms has a negative effect on the profitability of the incumbents. Note also that the Lerner index increases exogenously with the degree of differentiation across varieties, which also agrees with one of the main messages of industrial organization, that is, product differentiation relaxes competition.

To determine the equilibrium firm size, one could substitute the equilibrium price into the demand function (21). By plugging prices and quantities into the ZPC, one could obtain the equilibrium mass of firms/varieties. It is in fact simpler, but strictly equivalent, to proceed in the reverse order by determining first the volume of production thanks to the free entry condition given by

$$\Pi_i = (p^* - c)q_i - F = \frac{c}{\sigma - 1}q_i - F = 0,$$

which yields

$$q^* = \frac{\sigma - 1}{c}F. \tag{24}$$

Thus, regardless of the mass of firms, they all have the same size. This result, which is a direct consequence of a constant markup, is one of the major weaknesses of the CES model: there is no scale effect as q^* is independent of the market size L.

It follows immediately from the labor market balance that

$$N^* = \frac{Ly}{\sigma F}. \tag{25}$$

Hence, when varieties are less (more) differentiated, the mass of firms is smaller (larger), while a firm's output is larger (smaller) because the market demand is less (more) fragmented. Furthermore, a higher degree of increasing returns is associated with larger output and fewer but larger firms.

There is no question that the CES model of monopolistic competition captures some fundamental features of imperfect competition. But, and this is a big but, it is at odds with the main corpus of oligopoly theory. Despite (or, perhaps, because of) its great flexibility in applications and econometric estimations, the CES model brushes under the carpet several effects that may deeply affect the results it gives rise to. Therefore, although this model is a natural point of departure in studying issues where imperfect competition and increasing returns are crucial, we find it hard to maintain that it can serve as a corner-stone of any sound theory. For this, we need alternative or more general models to test the robustness of the results. Note, finally, that how appealing a model is depends on what questions one is interested in and whether the features from which the CES model abstracts are important for the issue in question.

3.1.2 The weighted CES

Assume that the CES is modified as follows:

$$U = \left(\int_0^N (a_i x_i)^{\frac{\sigma-1}{\sigma}} \mathrm{d}i\right)^{\frac{\sigma}{\sigma-1}}, \tag{26}$$

where $a_i > 0$ are salience coefficients whose purpose is to account for asymmetries among varieties (Bernard *et al.*, 20010, 2011). If $x_i = x_j$, then $a_i > a_j$ implies that, everything else being equal, the utility of consuming variety i exceeds that of variety j. However, the consumer is indifferent between consuming a_i/a_j units of variety i and one unit of variety j. Therefore, the preferences (26) can be made symmetric by changing the units in which the quantities of varieties are measured. Nevertheless, changing the units in which varieties are measured implies that firms that are otherwise symmetric now face different marginal costs. To be precise, firm i's marginal cost is equal to c/a_i. This implies that *a CES with asymmetric preferences and symmetric firms is isomorphic to a CES in which preferences are symmetric and firms heterogeneous.* Accordingly, to discriminate between cost heterogeneity and the salience coefficients a_i, one needs data on prices and sales because prices reflect the heterogeneity in costs while the salience coefficients act like demand-shifters in the CES (Kugler and Verhoogen, 2012).

3.2 Monopolistic competition under linear-quadratic preferences

3.2.1 The benchmark set-up

We have seen that there are (at least) two versions of LQ preferences defined over a finite number of varieties, namely (12) and (13). Even though the former is not the limit of oligopolistic competition, it is associated with equilibrium values of the main variables that vary with the key parameters of the model (Ottaviano *et al.*, 2002):[3]

$$U(\mathbf{x}) = \alpha \int_0^N x_i \mathrm{d}i - \frac{\beta}{2} \int_0^N x_i^2 \mathrm{d}i - \frac{\gamma}{2} \int_0^N \left(\int_0^N x_k \mathrm{d}k \right) x_i \mathrm{d}i + x_0. \tag{27}$$

One unit of labor is needed to produce one unit of the homogeneous good x_0, which is sold under perfect competition. This good is chosen as the numéraire so that the equilibrium wage is equal to 1. A consumer's budget constraint is as follows:

$$\int_0^N p_i x_i \mathrm{d}i + x_0 = 1 + \bar{x}_0, \tag{28}$$

where $\bar{x}_0$, the initial endowment in the numéraire, is supposed to be large enough for the consumption of this good to be strictly positive at the market outcome.

Solving (28) for the numéraire consumption, substituting the corresponding expression into (27) and solving the FOCs with respect to x_i yields the individual inverse demand for variety i:

$$p_i = \alpha - \beta x_i - \gamma X \qquad i \in [0, N] \tag{29}$$

where

$$X \equiv \int_0^N x_k \mathrm{d}k$$

is the total individual consumption of the differentiated product. Varieties interact through the value of X, which determines the demand intercept $\alpha - \gamma X$, so that an hike in X renders the

[3] Papers that use the LQ model include Belleflamme *et al.* (2000), Nocke (2006), Foster *et al.* (2008) and Dhingra (2013).

inverse demand more elastic. As a consequence, *when choosing its output level each firm must guess what X will be*, meaning that the game is aggregative in nature.

Firm i's profit function is as follows:

$$\Pi_i = (p_i - c)q_i - F = L \cdot \left[(p_i - c)x_i - \frac{F}{L}\right],$$

so that maximizing Π_i with respect to q_i amounts to maximizing the bracketed term with respect to x_i. To ease the burden of notation, we assume that c is equal to zero, which amounts to rescaling the demand intercept.

The best reply function

$$x^*(X) = \frac{\alpha - \gamma X}{2\beta}$$

shows how *each firm plays against the market* as x^* decreases with X. Since the equilibrium values of x and X must satisfy the condition $Nx = X$, for any given mass N of firms, the consumption $x^*(N)$ is given by

$$x^*(N) = \frac{\alpha}{2\beta + \gamma N}, \tag{30}$$

which decreases with the mass of competitors. Using (30), (29) yields the price $p^*(N)$:

$$p^*(N) = \frac{\alpha\beta}{2\beta + \gamma N} = \beta x^*(N). \tag{31}$$

which also decreases with N. Thus, unlike the CES, entry generates pro-competitive effects. In addition, as suggested by product differentiation theory, the market price rises when varieties get more differentiated (lower γ).

The ZPC implies that the equilibrium mass of firms is given by

$$N^* = \frac{1}{\gamma}\left(\alpha\sqrt{\frac{\beta L}{F}} - 2\beta\right). \tag{32}$$

It is readily verified that N^* increases at a decreasing rate with the market size (L), the consumer's willingness-to-pay for the differentiated product (α), the degree of product differentiation ($1/\gamma$), whereas it decreases with the marginal and fixed costs (c and F).

Substituting (32) in (30) and multiplying by L gives the equilibrium output:

$$q^* = Lx^* = \sqrt{\frac{FL}{\beta}},$$

which increases with L at a decreasing rate, while a stronger love for variety allows more firms to enter the market, but they all have a smaller size. Plugging (32) in (31) gives the equilibrium price:

$$p^* = \sqrt{\frac{\beta F}{L}},$$

which decreases with L at an increasing rate.

Thus, market size and cost parameters matter for all the equilibrium variables under free entry. This makes the linear model of monopolistic competition a good proxy of an oligopolistic market. Notwithstanding the absence of an income effect, the linear model performs fairly well in trade theory and economic geography (Ottaviano and Thisse, 2004; Melitz and Ottaviano, 2008).

4 The VES model of monopolistic competition

Choosing an appropriate framework for studying imperfect competition involves a trade-off between allowing firms to have sophisticated behavior and capturing basic general equilibrium effects. In this section, we discuss a model that aims to find a prudent balance between those two objectives. The CES and LQ models, as well as the translog developed by Feenstra (2003), are all special cases. Firms are still symmetric, which allows one to insulate the impact of preferences on the market outcome and to assess the limitations of specific models.

4.1 Firms and consumers

Owing to their analytical simplicity, the CES and LQ models conceal a difficulty that is often ignored: working with a continuum of goods implies that we cannot use the standard tools of calculus. Rather, we must work in a functional space whose elements are functions, and not vectors.

Let $\mathbb{N}$, an arbitrarily large number, be the mass of *potential* varieties. As all potential varieties are not necessarily made available to consumers, we denote by $N \leq \mathbb{N}$ the endogenous mass of *available* varieties. A *consumption profile* $\mathbf{x} \geq 0$ is a Lebesgue-measurable mapping from the space of potential varieties $[0, \mathbb{N}]$ to $\mathbb{R}_+$, which is assumed to belong to $L_2([0, \mathbb{N}])$. Individual preferences are described by a *utility functional* $U(\mathbf{x})$ defined over the positive cone of $L_2([0, \mathbb{N}])$. In what follows, we assume that (i) U is symmetric over the range of potential varieties in the sense that any Lebesgue measure-preserving mapping from $[0, \mathbb{N}]$ into itself does not change the value of U, and (ii) U exhibits a love for variety. To determine the inverse demand for a variety, Parenti *et al.* (2014) assume that the utility functional is Fréchet-differentiable: there exists a unique function $D(x_i, \mathbf{x})$ from $\mathbb{R}_+ \times L_2$ to $\mathbb{R}$ such that, for any given N and for all $\mathbf{h} \in L_2$, the equality

$$U(\mathbf{x} + \mathbf{h}) = U(\mathbf{x}) + \int_0^N D(x_i, \mathbf{x})\, h_i \,\mathrm{d}i + \circ\left(||\mathbf{h}||_2\right) \tag{33}$$

holds, $||\cdot||_2$ being the L_2-norm.[4] The function $D(x_i, \mathbf{x})$, which is the marginal utility of variety i, is the same across varieties because preferences are symmetric. Parenti *et al.* (2014) focus on the utility functionals such that the marginal utility $D(x_i, \mathbf{x})$ is decreasing and twice differentiable with respect to x_i.

Maximizing the utility functional $U(\mathbf{x})$ subject to (i) the budget constraint

$$\int_0^N p_i x_i \mathrm{d}i = y,$$

and (ii) the availability constraint

$$x_i \geq 0 \text{ for all } i \in [0, N] \quad \text{and} \quad x_i = 0 \text{ for all } i \in]N, \mathbb{N}]$$

yields the inverse demand function for variety i:

$$p_i = \frac{D(x_i, \mathbf{x})}{\lambda} \quad \text{for all } i \in [0, N], \tag{34}$$

where λ is the Lagrange multiplier of the consumer's optimization problem. Expressing λ as a function of y and $\mathbf{x}$ yields

$$\lambda(y, \mathbf{x}) = \frac{\int_0^N x_i D(x_i, \mathbf{x})\, \mathrm{d}i}{y},$$

which is the marginal utility of income at the consumption profile $\mathbf{x}$ and income y.

Firm i maximizes (22) with respect to its output q_i subject to the inverse market demand function $p_i = LD/\lambda$, while the market outcome is given by a Nash equilibrium. Being negligible, each firm accurately treats the variables $\mathbf{x}$ and λ in (34) as parameters. Note the difference between the consumer and producer programs. The individual chooses a consumption level for all available varieties. By contrast, each firm selects an output level for a single variety. In other words, the consumer's choice variable $\mathbf{x}$ is defined on a non-zero measure set while firm i's choice variable q_i is defined on a zero-measure set. Thus, unlike in Aumann (1964), the key ingredient of monopolistic competition is the negligibility of firms rather than that of consumers.[5]

Plugging (34) into (22) and using the product market clearing condition, the program of firm i may be rewritten as follows:

$$\max_{x_i} \Pi_i(x_i, \mathbf{x}) \equiv \left[\frac{D\left(x_i, \mathbf{x}\right)}{\lambda} - c\right] L\, x_i - F.$$

Setting

$$D_i' \equiv \frac{\partial D(x_i, \mathbf{x})}{\partial x_i} \qquad D_i'' \equiv \frac{\partial D^2(x_i, \mathbf{x})}{\partial x_i^2},$$

[4] Formally, this means that we use the concept of Fréchet-differentiability, which extends in a fairly natural way the standard concept of differentiability to L_2.

[5] We thank Kristian Behrens for having pointed out this difference to us.

the FOC for profit maximization is given by

$$D(x_i, \mathbf{x}) + x_i D_i' = [1 - \eta(x_i, \mathbf{x})]\, D(x_i, \mathbf{x}) = \lambda c, \tag{35}$$

where

$$\eta(x_i, \mathbf{x}) \equiv -\frac{x_i}{D(x_i, \mathbf{x})} \frac{\partial D(x_i, \mathbf{x})}{\partial x_i}$$

is the elasticity of the inverse demand for variety i. The right hand side of (35) is variable, and thus each firm must guess what the equilibrium value of λ is to determine its profit-maximizing output. Parenti *et al.* (2014) show that the profit function Π_i is strictly quasi-concave in x_i for all admissible values of λc if and only if

(**A**) *firm i's marginal revenue decreases in* x_i.

4.2 The elasticity of substitution

We have seen that the elasticity of substitution plays a central role in the CES model of monopolistic competition. Many would argue that this concept is relevant for such preferences only. Parenti *et al.* (2014) show that such an opinion is unwarranted. More specifically, they use the elasticity of substitution function $\bar{\sigma}(x_i, x_j, \mathbf{x})$ between varieties i and j, which is conditional upon the consumption profile $\mathbf{x}$. At an arbitrary symmetric consumption pattern $\mathbf{x} = xI_{[0,N]}$, we have:

$$\sigma(x, N) \equiv \bar{\sigma}(x, x, xI_{[0,N]}).$$

In other words, along the diagonal *the elasticity of substitution hinges only upon the individual consumption per variety and the total mass of available varieties.*

To gain insights about the behavior of σ, we give below the elasticity of substitution for the two main families of preferences used in the literature.

(i) When the utility is additive, we have:

$$\sigma(x, N) \equiv \frac{1}{r_u(x)}, \tag{36}$$

where $r_u(x)$ is the relative love for variety (see Section 2.1). As implied by (36), σ depends only upon the individual per variety consumption.

(ii) When preferences are homothetic, $D(x, \mathbf{x})$ evaluated at a symmetric consumption profile depends solely on the mass N of available varieties:

$$\sigma(x, N) \equiv \frac{1}{\eta(1, I_{[0,N]})} \equiv \frac{1}{\mathcal{M}(\mathcal{N})}. \tag{37}$$

Using (36) and (37), we have the following: $\mathcal{E}_N(\sigma) = 0$ means that preferences are additive, while $\mathcal{E}_x(\sigma) = 0$ means that preferences are homothetic.

4.3 Market equilibrium

Assume that (**A**) holds. Then, for any given $N \leq Ly/F$, Parenti *et al.* (2014) show that there exists a unique Nash equilibrium such that (i) no firm can increase its profit by changing its output; (ii) each consumer maximizes utility subject to her budget constraint; (iii) the product markets clear; (iv) the labor market balance holds. Furthermore, this equilibrium is symmetric and given by

$$x^*(N) = \frac{y}{cN} - \frac{F}{cL}, \qquad q^*(N) = \frac{yL}{cN} - \frac{F}{c}, \qquad p^*(N) = c\frac{\sigma(x^*(N), N)}{\sigma(x^*(N), N) - 1}, \tag{38}$$

and thus the equilibrium markup is

$$m^*(N) \equiv \frac{p^*(N) - c}{p^*(N)} = \frac{1}{\sigma(x^*(N), N)}, \tag{39}$$

which generalizes the expression (23) obtained under the CES. First of all, (39) suffices to show that, *in monopolistic competition working with a variable markup amounts to assuming a variable elasticity of substitution and non-isoelastic demands.* Furthermore, as in oligopoly theory, all variables depend on the mass of active firms. In particular, the equilibrium per variety consumption $x^*(N)$ always decreases with N, whereas the impact of N on $m^*(N)$ is a priori undetermined. To be precise, since $\sigma(x^*(N), N)$ may increase or decrease with the mass of firms, entry may generate pro- or anti-competitive effects. This in turn shows why comparative statics may give rise to diverging results in models where preferences are characterized by different functions $\sigma(x, N)$. In a nutshell, monopolistic competition is able to mimic oligopolistic competition. Finally, since $q^*(N)$ decreases with N, there is a business stealing effect regardless of preferences.

Using (38) yields the operating profits earned by a firm:

$$\Pi^*(N) = \frac{c}{\sigma\left(x^*(N), N\right) - 1} Lx^*(N) - F, \tag{40}$$

Solving the ZPC $\Pi^*(N) = 0$ with respect to m yields a single equilibrium condition:

$$m^*(N) = \frac{NF}{Ly}. \tag{41}$$

Setting $m \equiv FN/(Ly)$, (39) may be rewritten as a function of m only:

$$m\sigma\left(\frac{F}{cL}\frac{1-m}{m}, \frac{Ly}{F}m\right) = 1. \tag{42}$$

This expression shows that *a variable elasticity of substitution* $\sigma(x, N)$ *is sufficient to characterize the market outcome under general symmetric preferences and symmetric firms.*

Note that (42) implies that σ *must* be a function, and not a constant, for the markup to be variable in our general framework. Since (42) involves the four structural parameters of the economy (L, y, c and F), how the market outcome varies with these parameters depends on how σ varies with x and N.

Although the above framework allows for very different patterns of substitution across varieties, it should be clear that they are not equally plausible. This is why most applications of monopolistic competition focus on different subclasses of utilities to cope with particular effects. Admittedly, making "realistic" assumptions on how the elasticity of substitution varies with x and N is not an easy task. That said, it is worth recalling with Stigler (1969) that "it is often impossible to determine whether assumption A is more or less realistic than assumption B, except by comparing the agreement between their implications and the observable course of events." This is what we will do below.

Spatial and discrete choice models of product differentiation suggest that varieties become closer substitutes when the number of competing varieties rises (Salop, 1979; Anderson *et al.*, 1995). This leads Feenstra and Weinstein (2016) to use the translog expenditure function, where $\sigma(N) = 1 + \beta N$ increases with N, to capture the pro-competitive effects of entry. Therefore, $\mathcal{E}_N(\sigma) \geq 0$ seems to be a reasonable assumption. In contrast, how σ varies with x is a priori less clear. Nevertheless, this question can be answered by appealing to the literature on pass-through.

If a firm's demand is not too convex, the pass-through of a cost change triggered by a trade liberalization or productivity shock is smaller than 100% for a very large family of demand functions (Greenhut *et al.*, 1987). More importantly, the empirical evidence strongly suggests that the pass-through is incomplete (De Loecker *et al.*, 2016). Which assumption about σ leads to this result? The intuition is easy to grasp when preferences are additive, that is, $m(x) = r_u(x) = \sigma(x)$. Incomplete pass-through amounts to saying that p/c increases when c decreases, which means that firms have more market power or, equivalently, varieties are more differentiated. As firms facing a lower marginal cost produce more, the per capita consumption increases. Therefore, it must be that $\sigma(x)$ decreases with x. In the case of general symmetric preferences, Parenti *et al.* (2014) show that the pass-through is smaller than 100% *if and only if* $\mathcal{E}_x(\sigma) < 0$ holds. In addition, the pass-through must be equal to 100% when preferences are homothetic because $\mathcal{E}_x(\sigma) = 0$.

This discussion suggests the following conditions:

$$\mathcal{E}_x(\sigma) \leq 0 \leq \mathcal{E}_N(\sigma). \tag{43}$$

Even though these inequalities do not hold for some preferences, it is convenient to assume here that (43) holds. Applying Propositions 1 to 4 of Parenti *et al.* to (43) then implies:

Proposition 3. *Assume that* (**A**) *and* (43) *hold. Then,* (i) *there exists a free-entry equilibrium for all* $c > 0$ *and* $F > 0$; (ii) *this equilibrium is unique and symmetric*; (iii) *a larger market* *or a higher income lead0s to lower markups, bigger firms and a larger number of varieties*; (iv) *the pass-through rate of a cost change is smaller than* 100%.

The pro-competitive effects associated with the extent of the market are intuitively plausible and supported by empirical evidence (Amiti *et al.*, 2016; De Loecker *et al.*, 2016). Nevertheless, one should bear in mind that the industrial organization literature highlights the possibility of anti-competitive effects (Chen and Riordan, 2008). Moreover, result (iv) of Proposition 3 may be used to study how firms react to a shock which affects aggregate productivity, as in Bilbiie *et al.* (2012). To capture the versatility of the market outcome in the present setting, Parenti *et al.* (2014) provide a complete description of the comparative static effects through necessary and sufficient conditions, which may be used to pin down the restrictions on preferences for the equilibrium outcome to be consistent with the stylized facts.

Last, since we focus on monopolistic competition, the markup (39) stems directly from preferences through only the elasticity of substitution. This stands in stark contrast to oligopoly models where the markup emerges as the outcome of the interplay between preferences *and* strategic interactions. Nevertheless, by choosing appropriately the elasticity of substitution as a function of x and N, monopolistic competition is able to replicate the direction of comparative static effects generated in symmetric oligopoly models with free entry, as well as their magnitude. Therefore, as conjectured by Mas-Colell (1984), *monopolistic competition may be considered as the marriage between the negligibility hypothesis and oligopolistic competition.*

4.3.1 Additive preferences

Let $u(\cdot)$ be a strictly increasing and strictly concave function. Assume that the utility functional is as follows:

$$U(\mathbf{x}) = \int_0^{\mathbb{N}} u(x_i)\mathrm{d}i. \tag{44}$$

Since $D(x_i, \mathbf{x}) = u'(x_i)$, the marginal utility of variety i is independent of the other varieties' consumption. This property suggests that additive models retain, at least partially, the tractability of the CES. And indeed, since $\mathcal{E}_N(\sigma) = 0$, the equilibrium condition (42) becomes simpler:

$$m = r_u\left(\frac{F}{cL}\frac{1-m}{m}\right). \tag{45}$$

The equilibrium markup m^* is a fixed point of the function $r_u(x)$, which maps the interval $[0, 1]$ into itself. If $r_u'(x) \geq 0$ for all $x \geq 0$, then the right-hand side of (45) is weakly decreasing, and thus m^* is always unique. When $r_u'(x) < 0$, showing uniqueness is less straightforward. However, if (**A**) holds, the right-hand side of (45) is a contraction mapping over $[0, 1]$, which implies that the equilibrium exists and is unique.

To illustrate, consider the CARA utility $u(x) = 1 - \exp(-\alpha x)$ studied in Behrens and

Murata (2007). Since the RLV is given by $r_u(x) = \alpha x$, (45) is the following quadratic equation:

$$m^2 + \frac{\alpha F}{cL} m - \frac{\alpha F}{cL} = 0,$$

the solution of which is as follows:

$$m^* = \frac{\alpha F}{2cL}\left(\sqrt{1 + 4\frac{cL}{\alpha F}} - 1\right). \tag{46}$$

Equation (46) gives a clue to understanding the asymptotic behavior of the market outcome: when the market is arbitrarily large, the equilibrium markup is arbitrarily close to zero. Thus, the economy features a competitive limit, which echoes what we saw in Section 2.1. Note that this is not so under the CES where $m^* = 1/\sigma > 0$ for all L.

Using (46) and recalling that $m = NF/(Ly)$ yields the equilibrium number of firms:

$$N^* = \frac{\alpha y}{2c}\left(\sqrt{1 + 4\frac{cL}{\alpha F}} - 1\right). \tag{47}$$

Plugging (47) into (38) pins down the equilibrium values of the remaining variables:

$$q^* = \frac{F}{2c}\left(\sqrt{1 + 4\frac{cL}{\alpha F}} - 1\right), \qquad p^* = c + \frac{\alpha F}{2L}\left(\sqrt{1 + 4\frac{cL}{\alpha F}} + 1\right). \tag{48}$$

Expressions (46) – (48) provide a complete solution of the CARA model. Furthermore, they imply unambiguous comparative statics with respect to L: *an increase in population leads to a drop in markup, price, and per variety consumption, an increase in firm size, and a less than proportional increase in the number of firms.*

Are these findings robust against the choice of alternative specifications for u? Zhelobodko *et al.* (2012) show that the following result: *if r_u is strictly increasing in x, then a larger market leads to a lower markup, bigger firms and a larger number of varieties, whereas the opposite holds when r_u is strictly decreasing in x.* Evidently, when r_u is constant, whence preferences are CES, L has no impact on the market outcome.

The above discussion also shows that the individual income y has no impact on the market solution. This led Bertoletti and Etro (2016) to work with indirectly additive preferences:

$$\mathcal{V}(\mathbf{p}; y) \equiv \int_0^{\mathbb{N}} v(y/p_i)\mathrm{d}i, \tag{49}$$

where v is strictly increasing, strictly concave, and homogeneous of degree zero. Such preferences mean that $\mathcal{E}_x(\sigma) = \mathcal{E}_N(\sigma)$, which is consistent with (43). Applying the RLV to v, Bertoletti and Etro show that the equilibrium price depends on y but not on L. Since $\mathcal{E}_N(\sigma) = 0$ for non-CES additive preferences and $\mathcal{E}_x(\sigma) = 0$ for non-CES homothetic preferences, indirectly additive preferences are disjoint from these two classes of preferences apart from the CES.

4.3.2 Homothetic preferences

There are several reasons for paying attention to homothetic preferences. First, such preferences retain much of the CES tractability. In particular, the marginal utility $D(x_i, \mathbf{x})$ of variety i is positive homogeneous of degree zero: $D(tx_i, t\mathbf{x}) = D(x_i, \mathbf{x})$ for all $t > 0$. By implication, an increase in income y leads to a proportional change in the consumption pattern $\mathbf{x}$ and leaves the relative consumptions x_i/x_j unchanged. Second, an appealing feature of homothetic preferences is that they can be easily nested into a multi-sectoral setting, for the aggregate price index is always well-defined.

Homothetic preferences were used by Bilbiie *et al.* (2012) to study real business cycles to capture the fact that both markups and the number of firms are highly procyclical variables, while Feenstra and Weinstein (2016) use *translog* preferences for studying international trade. It is well known that there is no closed-form expression of the translog utility functional, which is instead defined by the expenditure functional:

$$\ln E(\mathbf{p}, U) = \ln U + \frac{1}{\mathbb{N}} \int_0^{\mathbb{N}} \ln p_i \mathrm{d}i - \frac{\beta}{2\mathbb{N}} \left[\int_0^{\mathbb{N}} (\ln p_i)^2 \mathrm{d}i - \frac{1}{\mathbb{N}} \left(\int_0^{\mathbb{N}} \ln p_i \mathrm{d}i \right)^2 \right]. \tag{50}$$

Using (37), we find that under homothetic preferences the equilibrium condition (42) reduces to

$$m = \mathcal{M}\left(\frac{Ly}{F} m\right). \tag{51}$$

Under (43), $\mathcal{M}(N)$ is a decreasing function of N, and thus there exists a unique equilibrium markup.

Contrasting the properties of (51) with those of (45) provides an insightful comparison of the market outcomes generated by, respectively, homothetic and additive preferences. The most striking difference is that (45) does not involve y as a parameter. In other words, assuming additive preferences implies that per capita income shocks are irrelevant for understanding changes in markups, prices and firm sizes. This property of additive preferences justifies the choice of population size as a measure of the market size. In contrast, (51) involves both L and y through the product Ly, i.e. the total GDP. Another interesting feature of homothetic preferences is that, unlike (45), (51) does not involve the marginal cost c. This yields an important comparative statics result: *under monopolistic competition with non-CES homothetic preferences, the markup is variable but the pass-through is always* 1.

As in the case of additive preferences, we proceed by studying an analytically solvable non-CES example. We choose to work with translog preferences (50). In this case, $\mathcal{M}(N) = 1/(1+\beta N)$, while (51) is given by the following quadratic equation:

$$m^2 + \frac{F}{\beta Ly} m - \frac{F}{\beta Ly} = 0,$$

whose solution is

$$m^* = \frac{F}{2\beta Ly}\left(\sqrt{1+4\frac{\beta Ly}{F}} - 1\right). \tag{52}$$

Despite the differences between additive and homothetic preferences, the equilibrium markup (52) bears a remarkable resemblance to that obtained under the CARA utility, given by (46). In particular, (52) implies that $m^* \to 0$ in a large economy, i.e. when $Ly \to \infty$.

The equilibrium mass of firms can be determined by combining $m = FN/(Ly)$ with (52):

$$N^* = \frac{1}{2\beta}\left(\sqrt{1+4\frac{\beta Ly}{F}} - 1\right). \tag{53}$$

Plugging (53) into (38) and rearranging terms yields:

$$q^* = \frac{F}{2c}\left(\sqrt{1+4\frac{\beta Ly}{F}} - 1\right), \qquad p^* = c + \frac{cF}{2\beta Ly}\left(\sqrt{1+4\frac{\beta Ly}{F}} + 1\right). \tag{54}$$

Expressions (52)-(54) yield a complete analytical solution of the translog model and entail unambiguous comparative statics results: *an increase in GDP triggers a decrease in prices and markups, increases firm size, and invites more firms to enter the market.* The same holds for any symmetric homothetic preference satisfying (43). What is more, (52)-(54) are strikingly similar to (46)-(48). To be precise, the CARA and translog models yield the same market outcome up to replacing the population L by the total GDP Ly.

Figure 1 below shows the three subclasses of preferences used in the literature. The CES is the only one that belongs to all of them, which highlights how peculiar these preferences are.

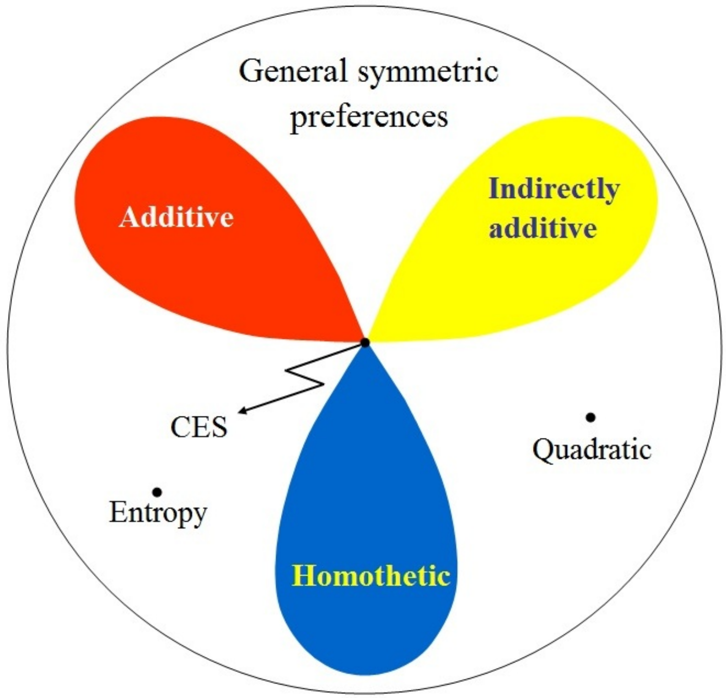

Fig. 1. The space of preferences.

5 Heterogeneous firms

In this section, we follow Melitz (2003) and assume that firms face different marginal costs. In this context, the key question is how the market selects the operating firms. We consider the one-period framework used by Jean (2002) and Melitz and Ottaviano (2008). Prior to entry, risk-neutral firms face uncertainty about their marginal cost while entry requires a sunk cost F_e. Once this cost is paid, firms observe their marginal cost drawn randomly from the continuous probability distribution $\Gamma(c)$ defined over $\mathbb{R}_+$. After observing its type c, each entrant decides whether to produce or not, given that an active firm must incur a fixed production cost F. Under such circumstances, the mass of entrants, N_e, is larger than the mass of operating firms, N.

Even though varieties are differentiated from the consumer's point of view, firms sharing the same marginal cost c behave in the same way and earn the same profit at the equilibrium. As a consequence, we may refer to any variety/firm by its c-type only. Furthermore, the envelope theorem implies that equilibrium profits always decrease with c. Hence, there is perfect sorting of firms by increasing order of marginal cost. In other words, there exists a value $\bar{c}$ such that all operating firms have a marginal cost smaller than or equal to $\bar{c}$, while firms having a marginal cost exceeding $\bar{c}$ choose not to produce. A consumer program may then be written as follows:

$$\max_{x_c(.)} \mathcal{U} \equiv N_e \int_0^{\bar{c}} u(x_c)\mathrm{d}\Gamma(c) \qquad \text{s.t.} \quad N_e \int_0^{\bar{c}} p_c x_c \mathrm{d}\Gamma(c) = y,$$

where $x_c \geq 0$ is the individual consumption of a c-variety. The mass of operating firms is then given by $N = N_e\Gamma(\bar{c})$. Since the distribution Γ is given, the equilibrium consumption profile is entirely determined by $\bar{c}$ and N_e. For homogeneous firms, the variable N is sufficient to describe the set of active firms.

A *free-entry equilibrium* $(c^*, N_e^*, q^*_{c\leq c^*}, x^*_{c\leq c^*}, \lambda^*)$ must satisfy the following equilibrium conditions:

(i) the profit-maximization condition for c-type firms:

$$\max_{x_c} \Pi_c(x_c, \mathbf{x}) \equiv \left[\frac{D(x_c, \mathbf{x})}{\lambda} - c\right] Lx_c - F;$$

(ii) the ZPC for the cutoff firm:

$$(p_{c^*} - c^*)q_{c^*} = F,$$

where c^* is the cutoff cost. At the equilibrium, firms are sorted out by decreasing order of productivity, which implies that the mass of active firms is equal to $N \equiv N_e\Gamma(c^*)$;

(iii) the product market clearing condition:

$$q_c = Lx_c$$

for all $c \in [0, c^*]$;

(iv) the labor market clearing condition:

$$N_eF_e + \int_0^{c^*} (F + cq_c)\mathrm{d}\Gamma(c) = yL;$$

(v) firms enter the market until their expected profits net of the entry cost F_e are zero:

$$\int_0^{c^*} \Pi_c(x_c, \mathbf{x})\mathrm{d}\Gamma(c) = F_e.$$

Although some entrants earn positive profits whereas others lose money, the last condition implies that total profits are zero. Hence, yL is the total income.

5.1 Additive preferences

Melitz (2003) and successors assume that consumers have CES preferences. This vastly simplifies the analysis because the equilibrium price of a c-type firm, $p^*(c) = c\sigma/(\sigma - 1)$, does not depend on the cost distribution Γ, although the price index does. Given that many properties derived under CES preferences are not robust, we focus below on additive preferences.

The inverse demand function (34) becomes $p_c(x_c) = u'(x_c)/\lambda$, which implies that the demand structure retains the simplicity of the homogeneous firm case. The profits made by a c-type firm are given by

$$\Pi(x_c; \lambda) = \left[\frac{u'(x_c)}{\lambda} - c\right] Lx_c - F.$$

Rewriting (39) for each type c implies that the equilibrium markup of a c-type firm is given by

$$m_c^* = r_u(x^*(c)) = 1/\sigma(x^*(c)), \tag{55}$$

which extends (23) to markets where firms are heterogeneous. It follows immediately from (55) that the elasticity of substitution is now c-specific in that it is the same *within* each type, whereas it varies *between* types. Furthermore, as firms of different types charge different prices, the individual consumption $x^*(c)$ varies with the firm's type, so that the equilibrium markup also varies with c. Evidently, a more efficient firm sells at a lower price than a less efficient firm. Therefore, consumers buy more from the former than from the latter, so that *a firm's markup increases* (decreases) *with its degree of efficiency when the RLV is increasing* (decreasing).

Given the second-order condition for profit-maximization ($r_{u'}(\cdot) < 2$), for each type c the expression

$$\bar{\pi}(c, \lambda; L) \equiv \max_{q \geq 0} \left[\frac{u'(q/L)}{\lambda} q - cq\right]$$

is a well-defined and continuous function. Since $\bar{\pi}(c, \lambda, L)$ is strictly decreasing in c, the solution $\bar{c}(\lambda; L)$ to the equation $\bar{\pi}(c; \lambda, L) - F = 0$ is unique. Clearly, the free-entry condition may be rewritten as follows:

$$\int_0^{\bar{c}(\lambda;L)} [\bar{\pi}(c, \lambda; L) - F] \, \mathrm{d}\Gamma(c) - F_e = 0. \tag{56}$$

Using the envelope theorem and the ZPC at $\bar{c}(\lambda; L)$, we find that $\bar{\pi}(c, \lambda; L)$ and $\bar{c}(\lambda; L)$ are both decreasing functions of λ, which implies that the left-hand side of (56) is also decreasing in λ. As a consequence, the above equation has a unique solution $\bar{\lambda}(L)$. Plugging this expression into $\bar{c}(\lambda; L)$ yields the equilibrium cutoff $c^*(L)$. In other words, the free-entry equilibrium, if it exists, is unique. The expression (56) also shows that $c^*(L)$ exists when the fixed production cost F and entry cost F_e are not too large.

We are now equipped to study the impact of market size on the cutoff cost. The ZPC at $\bar{c}$ implies that

$$\frac{\partial \bar{\pi}}{\partial L} + \frac{\partial \bar{\pi}}{\partial c} \frac{d\bar{c}}{dL} + \frac{\partial \bar{\pi}}{\partial \lambda} \frac{d\bar{\lambda}}{dL} = 0.$$

Rewriting this expression in terms of elasticity and applying the envelope theorem to each term, it can be shown that the elasticity of c^* with respect to L is, up to a positive factor, equal to

$$\int_0^{\bar{\theta}} [r_u(x^*(c^*)) - r_u(x^*(c))] \, R^*(c) \mathrm{d}\Gamma(c),$$

where $R^*(c)$ is the equilibrium revenue of a c-type firm. As a consequence, the elasticity of c^* is negative (positive) if r_u is increasing (decreasing). Therefore, we have:

Proposition 4. *Regardless of the cost distribution, the cutoff cost decreases with market size if and only if the RLV is increasing. Furthermore, the cutoff cost is independent of the market size if and only if preferences are CES.*

Thus, the number of firms selected when the market gets bigger depends only upon the behavior of the RLV. Intuitively, we expect a larger market to render competition tougher (the RLV increases), which in turn triggers the exit of the least productive firms. However, if a larger market happens to soften competition (the RLV decreases), then less productive firms are able to stay in business.

Zhelobodko *et al.* (2012) show that both the equilibrium mass of entrants and the mass of operating firms increase with L when the RLV increases. The same authors also establish that the equilibrium consumption $x^*(c)$ decreases with L for all $c < c^*$. Therefore, when the RLV is increasing, (55) implies that the equilibrium price $p^*(c)$ decreases for all the c-type firms which remain in business. Hence, when preferences are additive, prices move in the same direction in response to a market size shock whether firms are homogeneous or heterogeneous.

Note, finally, that even when preferences generate pro-competitive effects ($\mathcal{E}_x(\sigma) < 0$), the selection of firms associated with a bigger market may lead to a drop in aggregate productivity because the more productive firms need not gain more demand than the less productive firms. To be precise, Bertoletti and Epifani (2014) show that a hike in L may have a negative impact on the aggregate productivity if the elasticity of the marginal revenue is decreasing in q while the elasticity of substitution is decreasing in x.

5.2 Linear-quadratic preferences

Melitz and Ottaviano (2008) propose an alternative approach based on utility (27). Because such preferences generate linear demands which feature a finite choke-price α (see Section 3.2), there is no need to assume that firms face a positive fixed production cost ($F = 0$). In this case, total profits are generally different from zero. However, how profits/loses are shared does not matter because the upper-tier utility is linear.

Firm i operates if the demand for its variety is positive, that is,

$$p_i \leq p_{\max} \equiv \frac{\beta\alpha + \gamma N\bar{p}}{\beta + \gamma N}, \tag{57}$$

holds, where $\bar{p}$ is the average market price given by

$$\bar{p} \equiv \frac{1}{N}\int_0^N p_i \mathrm{d}i.$$

The market demand for a variety i is such that

$$q_i = \frac{L}{\beta}\left(p_{\max} - p_i\right). \tag{58}$$

Unlike the CES, the price elasticity of the demand for variety i is variable and equal to

$$\mathcal{E}_{p_i}(x_i) = \frac{p_i}{p_{\max} - p_i}. \tag{59}$$

The expressions (57) and (59) imply that the demand elasticity increases with its own price p_i, decreases with the average market price $\bar{p}$ (because varieties are substitutes), and increases when more firms are active.

Using (58), it is readily verified that the profit-maximizing price $p^*(c)$ set by a c-type firm must satisfy

$$p^*(c) = \frac{p_{\max} + c}{2}, \tag{60}$$

which boils down to (31) when firms are homogeneous.

Assume that the support of cost distribution Γ is a compact interval $[0, c_M]$, where c_M is large. The cut-off cost $\bar{c} \in [0, c_M]$ satisfies $p^*(c) = c$, that is, the least productive operating firm earns zero profits and prices at the marginal cost. Combining the cut-off condition $p^*(c) = c$ with (60) yields

$$\bar{c} = p_{\max},$$

so that the equilibrium price $p^*(c)$, output $q^*(c)$ and profits $\pi^*(c)$ of a c-type firm are given by:

$$p^*(c) = \frac{\bar{c} + c}{2}, \qquad q^*(c) = \frac{L}{\beta}\frac{\bar{c} - c}{2}, \tag{61}$$

$$\pi^*(c) \equiv \left[p^*(c) - c\right] q^*(c) = \frac{L}{4\beta}(\bar{c} - c)^2. \tag{62}$$

By implication, *firms with a higher productivity have more monopoly power and higher profits.*

It is well known that linear demands allow for a simple relationship between the variances of prices and marginal costs:

$$\mathbf{V}(p) = \frac{1}{4}\mathbf{V}(c),$$

which means *prices are less dispersed than marginal costs.* This result complements the discussion on incomplete pass-through in Section 4.2.

It remains to pin down $\bar{c}$, which is given by using the ZPC:

$$\int_0^{\bar{c}} \bar{\pi}(c) \mathrm{d}\Gamma(c) = F_e,$$

where $F_e > 0$ is the sunk entry cost. Using (62), we restate this condition as follows:

$$\frac{L}{4\beta}\int_0^{\bar{c}}(\bar{c}-c)^2\mathrm{d}\Gamma(c) = F_e. \tag{63}$$

The left-hand side of this expression is an increasing function of $\bar{c}$, which implies that (63) has a unique solution c^*. This solution is interior ($0 < c^* < c_M$) if and only if

$$\mathbf{E}\left[(c_M - c)^2\right] > \frac{4\beta}{L}F_e$$

holds. Therefore, when the population L is very small, all firms choose to produce. Otherwise, as implied by (63), a hike in L drives c^* downwards, which confirms the idea that firms do not pass onto consumers the entire fall in cost (see Proposition 3). In other words, *a larger market skews the distribution of sales toward the varieties that are more efficiently produced.*

Finally, (57) and (61) can be used to pin down the mass of active firms:

$$N^* = \frac{2\beta}{\gamma}\frac{\alpha - c^*}{c^* - \mathbf{E}(c|c \leq c^*)},$$

which indicates a decreasing relationship between c^* and N^*. In particular, an increase in L invites more firms to enter, even though a larger market pushes the least productive firms out of business.[6]

5.3 VES preferences

Working with general symmetric preferences and heterogeneous firms is tricky. Assume that the cutoff cost $\bar{c}$ and the number of entrants are given, so that the mass of active firms is determined. Unlike the CES, the equilibrium consumption of a given variety depends on the consumption levels of the other varieties. Hence, markets are independent across varieties. Unlike additive preferences, competition among firms is no longer described by an aggregative game. Unlike LQ preferences, a closed-form solution is not available. All of this implies that the way firms choose their output is through a non-atomic game with asymmetric players, which cannot be solved point-wise. Such an equilibrium $\bar{\mathbf{x}}(\bar{c}, N_e)$, which need not be unique, can be shown to exist if, when a non-zero measure set of firms raise their prices, it is profit-maximizing for the other firms to increase their prices, as in oligopoly games where prices are strategic complements (Parenti *et al.*, 2014). The corresponding free-entry equilibrium is thus defined by a pair $(\bar{c}^*, N_e^*)$ which satisfies the zero-expected-profit condition for each firm:

$$\int_0^{\bar{c}}[\bar{\pi}_c(\bar{c}, N_e) - F]\mathrm{d}\Gamma(c) = F_e, \tag{64}$$

and the cutoff condition:

[6] Behrens *et al.* (2014) undertake a similar exercise within a full-fledged general equilibrium model with CARA preferences and income effects.

$$\bar{\pi}_{\bar{c}}(\bar{c}, N_e) = F. \tag{65}$$

Thus, regardless of the nature of preferences and the distribution of marginal costs, the heterogeneity of firms amounts to replacing N by $\bar{c}$ and N_e because $N = \Gamma(\bar{c})N_e$. As a consequence, *the complexity of the problem increases from one to two dimensions.*

Dividing (64) by (65) yields the following new equilibrium condition:

$$\int_0^{\bar{c}} \left[\frac{\bar{\pi}_c(\bar{c}, N_e)}{\bar{\pi}_{\bar{c}}(\bar{c}, N_e)} - 1 \right] \mathrm{d}\Gamma(c) = \frac{F_e}{F}. \tag{66}$$

When firms are symmetric, we have seen that the sign of $\mathcal{E}_N(\sigma)$ plays a critical role in comparative statics. Since firms of a given type are symmetric, the same holds here. The difference is that the mass of operating firms is determined by the two endogenous variables $\bar{c}$ and N_e. As a consequence, understanding how the mass of active firms responds to a population hike requires studying the way the left-hand side of (66) varies with $\bar{c}$ and N_e. Let $\sigma_c(\bar{c}, N_e)$ be the equilibrium value of the elasticity of substitution between any two varieties supplied by c-type firms:

$$\sigma_c(\bar{c}, N_e) \equiv \bar{\sigma}[\bar{x}_c(\bar{c}, N_e), \bar{\mathbf{x}}(\bar{c}, N_e)].$$

In this case, we may rewrite $\bar{\pi}_c(\bar{c}, N_e)$ as follows:

$$\bar{\pi}_c(\bar{c}, N_e) = \frac{c}{\sigma_c(\bar{c}, N_e) - 1} L\bar{x}_c(\bar{c}, N_e),$$

which is the counter-part of (40), while the markup of a c-type firm is given by

$$m_c^*(\hat{c}, N_e) = \frac{1}{\sigma_c(\hat{c}, N_e)}.$$

Hence, the elasticity of substitution can be used for studying heterogeneous firms at the cost of one additional dimension, i.e. the firm's type c. Following this approach, Parenti *et al.* (2014) prove the following result.

Proposition 5. *Assume that $\bar{\pi}_c(\bar{c}, N_e)$ decreases with $\bar{c}$ and N_e for all c. Then, the equilibrium mass of entrants increases with L. Furthermore, the equilibrium cutoff decreases with L when $\sigma_c(\bar{c}, N_e)$ increases with $\bar{c}$ and N_e, whereas it increases with L when $\sigma_c(\bar{c}, N_e)$ increases with $\bar{c}$ but decreases with N_e.*

Given $\bar{c}$, the number of operating firms is proportional to the number of entrants. Therefore, assuming that $\sigma_c(\bar{c}, N_e)$ increases with $\bar{c}$ and N_e may be considered as the counterpart of the condition $\mathcal{E}_N(\sigma) > 0$ discussed in subsection 5.3. In response to an increase in L, the two effects combine to induce the exit of the least efficient active firms. However, Proposition 5 also shows that predicting the direction of firms' selection is generally problematic.

6 Equilibrium versus optimum product diversity: a variety of results

Conventional wisdom holds that entry is desirable because it often triggers more competition and enhances social efficiency. However, when the entry of new firms involves additional fixed costs, the case for entry is less clear-cut. What is more, when goods are differentiated the extent of diversity comes into play. In this context, the following question arises: does the market provide too many or too few varieties?

Spence (1976) casts doubt on the possibility of coming up with a clear-cut answer to this question because two opposite forces are at work. First, the entrant disregards the negative impact its decision has on the incumbents by taking away from them some of their customers (the "business stealing" effect). This effect pushes toward excessive diversity. Second, the entrant is unable to capture the entire social benefit it creates by increasing diversity because it does not price discriminate across consumers (the "incomplete appropriability" effect). This pushes toward insufficient diversity. As a consequence, the comparison between the market and optimal outcomes is likely to depend on the particular framework we believe to be a good representation of differentiated markets. Conventional wisdom holds that *spatial models fosters excessive diversity, whereas the market and optimal outcomes do not differ much in the case of symmetric preferences.* The reason for this difference is that a firm has few neighboring rivals in spatial models, which facilitates entry. On the contrary, when competition is global, the entrant must compete with many rivals, which dampens entry.

6.1 Additive preferences

The social planner aims to find the mass of firms and the consumption profile that maximize the common utility level and meet the labor balance constraint:

$$\max_{(\mathbf{x},\,N)} U(\mathbf{x}) \qquad \text{s.t.} \quad cL\int_0^N x_i \mathrm{d}i + NF = L. \tag{67}$$

Using additivity and symmetry, this program may rewritten as follows:

$$\max_{(q,\,N)} Nu(x) \qquad \text{s.t.} \quad N = \frac{L}{cLx + F},$$

which can be reduced to maximizing

$$\frac{Lu(x)}{cLx + F}$$

with respect to x. Applying the FOC yields

$$\mathcal{E}_x(u) = \frac{cLx}{cLx + F}. \tag{68}$$

Using the equilibrium condition (45), we obtain

$$1 - r_u(x) = \frac{cLx}{cLx + F}. \tag{69}$$

These two expressions show that firms care about consumers' marginal utility (see (69)), which determines the inverse demands, whereas the planner cares about consumers' utility (see (68)).

The equilibrium outcome is optimal for any L, c and F if and only if the utility $u(\cdot)$ satisfies both (68) and (69), that is, solves the following differential equation:

$$r_u(x) + \mathcal{E}_x(u) - 1 = 0. \tag{70}$$

Can this condition be given a simple economic interpretation? Let λ be the social value of labor, that is, the Lagrange multiplier of the social planner. Therefore, it must be that $u'(x) = \lambda cx$, so that

$$1 - \mathcal{E}_x(u) = \frac{u(x) - u'(x)x}{u(x)} = \frac{u(x) - \lambda cx}{u(x)},$$

and thus $1 - \mathcal{E}_x(u)$ may be interpreted as the "social markup" of a variety (Vives, 1999). Since $r_u(x)$ is a firm's markup, (70) means that *the market and social outcomes coincide if and only if the private and social markups are identical at the equilibrium consumption.*

It is readily verified that, up to an affine transformation, $u(x) = x^\rho$ is the only solution to (70). Furthermore, labor balance implies that each firm produces the optimal quantity. Accordingly, when firms are symmetric *the CES is the only model with additive preferences under which the market outcome is socially optimal.* Intuitively, under the CES everything works *as if* firms' marginal cost were $c\sigma/(\sigma - 1) > c$, while the market price equals $c\sigma/(\sigma - 1)$ in an otherwise perfectly competitive market. Under these circumstances, the amount $(p^* - c)q^* = cq^*/(\sigma - 1) > 0$ must be interpreted as a transfer from consumers to firms, which allows firms to exactly cover their fixed costs. Labor market clearing pins down the mass of firms, which is optimal because the surplus $(p^* - c)q^*$ generated by an additional variety is equal to its launching cost F. As a consequence, the market equilibrium coincides with the socially optimal outcome. Dhingra and Morrow (2016) extends this result to heterogeneous firms. But how robust are these interesting optimality properties?

Comparing (68) with (69) implies that the market delivers excessive variety if and only if the private markup exceeds the social markup at the equilibrium consumption level:

$$\mathcal{E}_x(u)\big|_{x=x^*} > 1 - r_u(x^*). \tag{71}$$

For example, under the CARA utility, (71) may be written as follows:

$$\frac{\alpha x^*}{1-\alpha x^*} > \exp(\alpha x^*) - 1.$$

Applying Taylor expansion to both sides of this expression yields

$$\sum_{k=1}^{\infty} (\alpha x)^k > \sum_{k=1}^{\infty} \frac{1}{k!} (\alpha x)^k ,$$

which holds for any positive value of x. As a consequence, under the CARA, the market provides too many varieties while firms' output is too small. In addition, Behrens *et al.* (2016) show that, when firms are heterogeneous, the more productive firms under-produce, whereas the less productive firms over-produce, and thus the average productivity at the equilibrium is lower than at the social optimum.

Since

$$\frac{\mathrm{d}}{\mathrm{d}x}[1-\mathcal{E}_x(u)] = \frac{\mathcal{E}_x(u)}{x}\left[r_u(x) - (1-\mathcal{E}_x(u))\right],$$

there is always excessive diversity, hence firms' output is too small, if and only if $\mathcal{E}_x(u)$ is decreasing. Equivalently, there is always insufficient diversity, hence firms' outputs are too large, if and only if $\mathcal{E}_x(u)$ is increasing. For example, under the preferences given by $u(x) = (x+\alpha)^\rho$, there are too few (too many) varieties in equilibrium if $\alpha > 0$ ($\alpha < 0$).

Furthermore, in a multi-sector economy where firms are heterogeneous, the upper-tier utility is Cobb-Douglas, while each sub-utility is CES, the equilibrium and optimum coincide if and only if the elasticity of substitution is the same across sectors. Otherwise, too much labor is allocated to the more competitive sectors (Behrens *et al.*, 2016). These results point to the lack of robustness of the CES welfare properties, which may lead to strong biases in policy assessment.

6.2 Homothetic preferences

Since homothetic preferences are also widely used in applications, it is legitimate to ask how the above results change when preferences are homothetic. Without loss of generality, we assume that U is homogeneous of degree one in $\mathbf{x}$. In the case of symmetric consumption profiles $\mathbf{x} = xI_{[0,N]}$, we have

$$U\left(xI_{[0,N]}\right) \equiv \phi(N,x) = X\psi(N),$$

where $\psi(N) \equiv \phi(N,1)/N$ and $X \equiv xN$. The ratio of the first-order conditions is given by

$$X\frac{\psi'(N)}{\psi(N)} = \frac{F}{cL},$$

which is equivalent to

$$\mathcal{E}_{\psi}(N) \equiv N\frac{\psi'(N)}{\psi(N)} = \frac{F}{cLx}.$$

As for the market equilibrium condition (41) can be reformulated as follows:

$$\frac{\bar{m}(N)}{1-\bar{m}(N)} = \frac{F}{cLx}.$$

The social optimum and the market equilibrium are identical if and only if

$$\mathcal{E}_{\psi}(N) = \frac{\bar{m}(N)}{1-\bar{m}(N)}, \tag{72}$$

while there is excess (insufficient) variety if and only if the right-hand side term of (72) is larger (smaller) than the left-hand side term.

Given $\phi(X, N)$, it is reasonable to map this function into another homothetic preference $\mathbb{A}(N)\phi(X, N)$, where $\mathbb{A}(N)$ is a *shifter* which depends only on N. Observe that the utility $\mathbb{A}(N)U(\mathbf{x})$ is homothetic and generates the same equilibrium outcome as $U(\mathbf{x})$, for the elasticity of substitution $\sigma(N)$ is unaffected by introducing the shifter $\mathbb{A}(N)$. To determine the shifter $\mathbb{A}(N)$, (72) is to be rewritten as follows in the case of $\mathbb{A}(N)\phi(X, N)$:

$$\mathcal{E}_{\mathbb{A}}(N) + \mathcal{E}_{\psi}(N) = \frac{m(N)}{1-m(N)}. \tag{73}$$

For this expression to hold, $\mathbb{A}(N)$ must be the solution to the linear differential equation in N

$$\frac{\mathrm{d}A}{\mathrm{d}N} = \left[\frac{m(N)}{1-m(N)} - \frac{N}{\psi(N)}\frac{\mathrm{d}\psi}{\mathrm{d}N}\right]\frac{A(N)}{N},$$

which has a unique solution up to a positive constant. Therefore, *there always exists a shifter $A(N)$ such that* (73) *holds for all N if and only if $U(\mathbf{x})$ is replaced with $A(N)U(\mathbf{x})$*. The shifter aligns the optimum to the equilibrium, which remains the same. Furthermore, it is readily verified that there is excess (insufficient) variety if and only if the right-hand side term of (73) is larger (smaller) than the left-hand side term. Thus, even when one restricts oneself to homothetic and symmetric preferences, there is, a priori, no reason to expect a robust result to emerge.

In sum, care is needed, for the choice of (additive or homothetic) preferences is likely to affect the nature of the prescriptions based on quantitative models of monopolistic competition. In particular, *CES preferences, which occupy center stage in the growing flow of quantitative models, must be used with care.*

7 Concluding remarks

Accounting for oligopolistic competition in general equilibrium theory remains a worthy goal rather than an actual achievement. This is why many scholars have embraced the CES model of monopolistic competition. Although this model has great merits, it leads to knife-edge results or to findings that clash with fundamental principles of microeconomics and industrial organization. In addition, recent empirical evidence pointing out the shortcomings of the CES is growing fast. This does not mean, however, that we need a totally new framework; the emphasis on the elasticity of substitution is warranted when recognizing that it is variable, rather than constant. By mimicking the behavior of oligopolistic markets, the VES model of monopolistic competition offers an alternative solution to some of the difficulties uncovered by Gabszewicz and Vial (1972) in their pioneering work on imperfect competition in general equilibrium.

Despite real progress, it should be clear that there is scope for more work. We provide here a short list of some major issues that should rank high on the research agenda. First, papers coping with several sectors typically assume a Cobb-Douglas upper-tier utility and CES lower-tier sub-utilities. Such a specification of preferences leaves much to desire as it does not allow for a genuine interaction across sectors because the income share spent on each product is given a priori. Behrens *et al.* (2016) is a worthy exception that should trigger new contributions.

Second, the demand side of existing models of monopolistic competition relies on the assumption of symmetric preferences, while heterogeneity is introduced on the supply side only. Yet, the recent empirical evidence gathered by Hottman *et al.* (2016) find that 50 to 75% of the variance in U.S. firm size can be attributed to differences in what these authors call "firms' appeal," that is, the demand side, and less than 20% to average marginal cost differences. As a consequence, one may safely conclude that it is time to pay more attention to the demand side in monopolistic competition theory. Besides the VES model, another step in this direction has been made by Di Comite *et al.* (2014) who embed taste heterogeneity into the LQ model. Absent a specific taste demand parameter, the model with heterogeneous costs and quality only explains 55% of the quantity variation in Belgian exports. Allowing for taste differences generates asymmetry in demand across countries offers a rational for the missing variability in sales.

Last, one may wonder what "heterogeneous firms" actually mean in a world where, despite a large number of producers, a handful of firms account for a very high share of total sales. There are at least two different modeling strategies that can be used to tackle this question. Ever since Melitz (2003), the first approach with firms operating under monopolistic competition but facing different marginal costs is dominant. However, as observed by Neary (2011), firms in this approach differ in types, not in kind, as all firms remain negligible to the market. An second line of research, developed by Shimomura and Thisse (2012), combines a continuum of negligible (non-atomic) players and a few large (atomic) players who are able to manipulate the market. Hence, firms now differ in kind. This leads to a hybrid market structure blending the features of oligopoly and monopolistic competition. Despite its empirical relevance, this approach has attracted little attention in the profession.

References

[1] Anderson, S.P. and A. de Palma (2000) From local to global competition. *European Economic Review* 44: 423 – 48.

[2] Anderson, S.P., A. de Palma and Y. Nesterov (1995) Oligopolistic competition and the optimal provision of products. *Econometrica* 63: 1281 – 302.

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

[4] Anderson, S.P., N. Erkal and D. Piccinin (2015) Aggregative oligopoly games with entry. University of Virginia, memo.

[5] Aumann, R.J. (1964) Markets with a continuum of traders. *Econometrica* 32: 39 – 50.

[6] Amity, M., O. Itskhoki and J. Konings (2016) International shocks and domestic prices: How large are strategic complementarities? NBER Working Paper No. 22119.

[7] Beckmann, M.J. (1972) Spatial Cournot oligopoly. *Papers and Proceedings of the Regional Science Association* 28: 37 – 47.

[8] Behrens, K. and Y. Murata (2007) General equilibrium models of monopolistic competition: A new approach. *Journal of Economic Theory* 136: 776 – 87.

[9] Behrens, K., G. Mion, Y. Murata and J. Südekum (2014) Trade, wages, and productivity. *International Economic Review* 55: 1305 – 48.

[10] Behrens, K., G. Mion, Y. Murata and J. Südekum (2016) Distorted monopolistic competition. Université du Québec à Montréal, mimeo.

[11] Belleflamme, P., P. Picard and J.-F. Thisse (2000) An economic theory of regional clusters. *Journal of Urban Economics* 48: 158 – 84.

[12] Bernard, A.B., S.J. Redding and P.K. Schott (2010) Multiple-product firms and product switching. *American Economic Review* 100: 70 – 97.

[13] Bernard, A.B., S.J. Redding and P.K. Schott (2011) Multi-product firms and trade liberalization. *Quarterly Journal of Economics* 126: 1271 – 318.

[14] Bertoletti, P. and P. Epifani (2014) Monopolistic competition: CES redux? *Journal of International Economics* 93: 227 – 38.

[15] Bertoletti, P and F. Etro (2016) Monopolistic competition when income matters. *Economic Journal*, forthcoming.

[16] Bilbiie, F., F. Ghironi and M. Melitz (2012) Endogenous entry, product variety, and business cycles. *Journal of Political Economy* 120: 304 – 45.

[17] Chamberlin, E. (1933) *The Theory of Monopolistic Competition*. Cambridge, MA: Harvard University Press.

[18] Chen, Y. and M.H. Riordan (2007) Price and variety in the spokes model. *Economic Journal* 117: 897 – 921.

[19] Chen, Y. and M.H. Riordan (2008) Price-increasing competition. *Rand Journal of Economics* 39, 1042-1058

[20] d'Aspremont, C., R. Dos Santos Ferreira and L.-A. Gérard-Varet (1996) On the Dixit-Stiglitz model of monopolistic competition. *American Economic Review* 86: 623 – 29.

[21] De Loecker, J. and P.K. Goldberg (2014) Firm performance in a global economy. *Annual Review of Economics* 6: 201 – 27.

[22] De Loecker, J., P.K. Goldberg, A.K. Khandelwal, and N. Pavcnik (2016) Prices, markups and trade reform. *Econometrica* 84: 445 – 510.

[23] Deneckere, R. and M. Rotschild (1992) Monopolistic competition and preference diversity. *Review of Economic Studies* 59: 361 – 73.

[24] Dhingra, S. (2013) Trading away wide brands for cheap brands. *American Economic Review* 103: 2554 – 84.

[25] Dhingra, S. and J. Morrow (2016) Monopolistic competition and optimum product diversity under firm heterogeneity. *Journal of Political Economy*, forthcoming.

[26] Di Comite, F., J.-F. Thisse and H. Vandenbussche (2014) Verti-zontal differentiation in export markets. *Journal of International Economics* 93: 50 – 66.

[27] Dixit, A. and J.E. Stiglitz (1977) Monopolistic competition and optimum product diversity. *American Economic Review* 67: 297 – 308.

[28] Eaton, B.C. and R.G. Lipsey (1977) The introduction of space into the neoclassical model of value theory. In: M. Artis and A. Nobay (eds.). *Studies in Modern Economics*. Oxford: Basil Blackwell, pp. 59 – 96.

[29] Eaton, B. C. and N. Schmitt (1994) Flexible manufacturing and market structure. *American Economic Review* 84: 875 – 88.

[30] Feenstra, R.C. (2003) A homothetic utility function for monopolistic competition models, without constant price elasticity. *Economics Letters* 78: 79 – 86.

[31] Feenstra, R.C. and D. Weinstein (2016) Globalization, markups, and U.S. welfare. *Journal of Political Economy*, forthcoming.

[32] Foster, L., J.C. Haltiwanger and C. Syverson (2008) Reallocation, firm turnover, and efficiency: Selection on productivity or profitability? *American Economic Review* 98: 394 – 425.

[33] Gabszewicz, J.J. and J.-P. Vial (1972) Oligopoly "a la Cournot" in a general equilibrium analysis. *Journal of Economic Theory* 4: 381 – 400.

[34] Greenhut, M.L., G. Norman and C.-S. Hung (1987) *The Economics of Imperfect Competition. A Spatial Approach.* Cambridge: Cambridge University Press.

[35] Hart, O.D. (1985a) Imperfect competition in general equilibrium: An overview of recent work. In K.J. Arrow and S. Honkapohja (eds.). *Frontiers in Economics.* Oxford: Basil Blackwell, pp. 100 – 49.

[36] Hart, O.D. (1985b) Monopolistic competition in the spirit of Chamberlin: A general model. *Review of Economic Studies* LII: 529 – 46.

[37] Hart, O.D. (1985c) Monopolistic competition in the spirit of Chamberlin: Special results. *Economic Journal* 95: 889 – 908.

[38] Hotelling, H. (1929) Stability in competition. *Economic Journal* 39: 41 – 57.

[39] Hottman, C., S.J. Redding and D.E. Weinstein (2016) Quantifying the sources of firm heterogeneity. *Quarterly Journal of Economics*, forthcoming.

[40] Jean, S. (2002) International trade and firms' heterogeneity under monopolistic competition. *Open Economies Review* 13: 291 – 311.

[41] Kaldor, N. (1935) Market imperfection and excess capacity. *Economica* 2: 35 – 50.

[42] Kimball, M. (1995) The quantitative analytics of the basic neomonetarist model. *Journal of Money, Credit and Banking* 27: 1241 – 77.

[43] Krugman, P.R. (1979) Increasing returns, monopolistic competition, and international trade. *Journal of International Economics* 9: 469 – 79.

[44] Kugler, M. and E. Verhoogen (2012) Prices, plant size, and product quality. *Review of Economic Studies* 79: 307 – 39.

[45] Lancaster, K. (1979) *Variety, Equity, and Efficiency*. New York, NY: Columbia University Press.

[46] MacLeod, W.B., G. Norman and J.-F. Thisse (1988) Price discrimination and equilibrium in monopolistic competition. *International Journal of Industrial Organization* 6: 429 – 46.

[47] Mas-Colell, A. (1984) *On the Theory of Perfect Competition*. The 1984 Nancy L. Schwartz Memorial Lecture, Northwestern University.

[48] Matsuyama, K. (1995) Complementarities and cumulative process in models of monopolistic competition. *Journal of Economic Literature* 33: 701 – 29.

[49] McFadden, D. (1974) Conditional logit analysis of qualitative choice behavior. In: P. Zarembka (ed.). *Frontiers in Econometrics*. New York: Academic Press, pp. 105 – 42.

[50] Melitz, M.J. (2003) The impact of trade on intra-industry reallocations and aggregate industry productivity. *Econometrica* 71: 1695 – 725.

[51] Melitz, M.J. and G.I.P. Ottaviano (2008) Market size, trade, and productivity. *Review of Economic Studies* 75: 295 – 316.

[52] Neary, P. (2010) Two and a half theories of trade. *The World Economy* 33: 1 – 19.

[53] Nocke, V. (2006) A gap for me: Entrepreneurs and entry. *Journal of the European Economic Association* 4: 929 – 55.

[54] Norman, G. and J.-F. Thisse (1999) Technology choice and market structure: Strategic aspects of flexible manufacturing. *Journal of Industrial Economics* 47: 345 – 72.

[55] Ottaviano, G.I.P. and J.-F. Thisse (2004) Agglomeration and economic geography. In: J.V. Henderson and J.-F. Thisse (eds.). *Handbook of Regional and Urban Economics, Volume 4*. Amsterdam: Elsevier. 2563 – 608.

[56] Ottaviano G.I.P., T. Tabuchi and J.-F. Thisse (2002) Agglomeration and trade revisited. *International Economic Review* 43: 409 – 36.

[57] Parenti, M., P. Ushchev and J.-F. Thisse (2014) Toward a theory of monopolistic competition. CEPR Discussion Paper N°10014. New version available at http://ssrn.com/abstract=2719560.

[58] Perloff, J.M. and S.C. Salop (1985) Equilibrium with product differentiation. *Review of Economic Studies* 52: 107 – 20.

[59] Redding, S.J. (2011) Theories of heterogeneous firms and trade. *Annual Review of Economics* 3: 77 – 105.

[60] Robinson, J. (1934) What is perfect competition? *Quarterly Journal of Economics* 49: 104 – 20.

[61] Sattinger, M. (1984) Value of an additional firm in monopolistic competition. *Review of Economic Studies* 51: 321 – 32.

[62] Salop, S.C. (1979) Monopolistic competition with outside goods. *Bell Journal of Economics* 10: 141 – 56.

[63] Shimomura, K.-I. and J.-F. Thisse (2012) Competition among the big and the small, *RAND Journal of Economics* 43: 329 – 47.

[64] Simonovska, I. (2015) Income differences and prices of tradables: Insights from an online retailer. *Review of Economic Studies* 82: 1612 – 56.

[65] Spence, M. (1976) Product selection, fixed costs, and monopolistic competition. *Review of Economic Studies* 43: 217 – 35.

[66] Stigler, G. (1969) *The Organization of Industry*. Chicago: The University of Chicago Press.

[67] Triffin, R. (1940) *Monopolistic Competition and General Equilibrium Theory*. Cambridge, MA: Harvard University Press.

[68] Ushchev, P. and Y. Zenou (2016) Price competitition in product variety networks. CEPR Discussion Paper N°10862.

[69] Vives, X. (1999) *Oligopoly Pricing. Old Ideas and New Tools*. Cambridge, MA: MIT Press.

[70] Vogel, J. (2011) Spatial price discrimination with heterogeneous firms. *Journal of Industrial Economics* LIX: 661 – 76.

[71] Zhelobodko, E., S. Kokovin, M. Parenti and J.-F. Thisse (2012) Monopolistic competition: Beyond the constant elasticity of substitution. *Econometrica* 80: 2765 – 84.

Appendix

That $\mathcal{E}_{p_i}(P) > 0$ is straightforward because $P(\mathbf{p})$ is increasing in any p_i. To show that $\mathcal{E}_{p_i}(P) < 1$ also holds, observe that $P(\mathbf{p})$ satisfies $P(t\mathbf{p}) < tP(\mathbf{p})$ for any $t > 1$. Indeed, (6) implies that for any given $\mathbf{p}$ and any given $t > 1$, $P_1 \equiv P(t\mathbf{p})$ and $P_2 \equiv tP(\mathbf{p})$ must solve, respectively, the following equations:

$$P_1 = \varphi' \left[\sum_{j=1}^{n} u \left(\xi \left(\frac{tp_j}{P_1} \right) \right) \right], \qquad P_2 = t\varphi' \left[\sum_{j=1}^{n} u \left(\xi \left(\frac{tp_j}{P_2} \right) \right) \right].$$

As $t > 1$, the right-hand side of the second equation is greater than the right-hand side of the first equation. Therefore, $P_1 < P_2$, that is, $P(t\mathbf{p}) < tP(\mathbf{p})$. This, in turn, implies

$$\mathcal{E}_t \left[P(t\mathbf{p}) \right] \big|_{t=1} < 1,$$

or, equivalently,

$$\sum_{i=1}^{n} \mathcal{E}_{p_i}(P) < 1.$$

Since $\mathcal{E}_{p_i}(P) > 0$, it must be that $\mathcal{E}_{p_i}(P) < 1$. Q.E.D.

Contact details:
Philip Ushchev
National Research University Higher School of Economics (Moscow, Russia).
Center for Market Studies and Spatial Economics, leading research fellow.
E-mail: fuschev@hse.ru