Nonlinear pricing without market power; what could be learned from the contestable markets approach?

Eduardo Brou
NONLINEAR PRICING WITHOUT MARKET POWER; WHAT COULD BE LEARNED FROM THE CONTESTABLE MARKETS APPROACH?

Eduardo Brou*
ebrou@clix.pt

Keywords: nonlinear pricing, contestable markets, increasing returns to scale, asymmetric information.

Mots clés: nonlinear pricing, marchés contestables, rendements d’échelle croissants, information asymétrique.

1. INTRODUCTION

According to received economic theory, the implementation of price discrimination by a firm requires market power, (at least) in the short-term. However, casual observation, confirmed by empirical studies\(^1\), demonstrates that in industries that appear to be extremely competitive there is

\* Invited Adjunt Professor at Instituto Politécnico de Tomar. This paper is a version of chapter 6 of my Universidade Autónoma de Lisboa doctoral thesis. I would like to thank my advisor José Amado da Silva for helpful suggestions and comments.

widespread practice of price discrimination⁴, especially its second degree variant (Nonlinear Pricing); in fact, it seems clear that it is more difficult to identify competitive markets where there is no practice of a form of price discrimination, than the opposite: highly competitive industries such as air transport, removal companies, rental (car, equipment, space), hotels and restaurants are obvious examples (Dana, 1998, quoted in Baumol (2002))³. In these industries, the economic profit is approximately zero, the barriers to entry low, and firms are often mere price-takers without market power (no restriction of output of the industries in which firms operate⁴), mainly due to the conditions of free entry and exit⁵.

---

2 As Baumol (2002) refers to, this finding had already been made by Jules Dupuit in the mid-nineteenth century: “There is almost no industry where this phenomenon is not present” (1852-53, quoted in Ekelund e Hébert (1999)).

3 Many other examples of highly competitive markets may be mentioned: the computer industry (hardware), bookstores, transportation, ready-to-wear, theatres, laundries, advertising, automobiles. Wilson (1993) offers the following list of industries that employ Nonlinear Pricing (without specifying the degree of competition or contestability generally verified): Energy, Telecommunications, Express Mail, Air Transport, Transportation by Truck, Gas Transmission, Advertising, Rental (Equipment, Vehicles, Space, Rooms), Banking and Financial Services, Publications, Equipment (computers, copiers), Retail Sales, Hospitality, Cable TV.

4 See Levine (2000: p. 10) which underlines the relationship between market power and the capacity to restrict the output level of the industry: “To find market power, one will be required to do a conduct or structure analysis that is independent of multiple prices and which is designed to find restriction of output by intention or effect.” Obviously, the restriction of output is, in this framework, regarded as a vehicle to raise the company’s prices above the competitive level, and keep them there; this ability to keep prices above the competitive level (absent when there is free entry) constitutes itself as the true test of the existence of market power (which does not exclude other types of evidence that can be used to complement the argument that a certain firm possesses market power). We shall return to this subject, trying to correctly define the concept of market power (see Section 4, p. 58).

5 Certainly, in the industries mentioned, there will be a certain degree of differentiation between the firms, which would simplify the explanation of price discrimination maintenance in extremely competitive markets. Levine (2000) attempts to dismantle this argument, noting that it would be a great coincidence that differentiated firms possess only, and exactly, the amount of local monopoly power necessary to cover the sunk and common costs without generating profits or losses. Moreover, Baumol (2002) suggests that price discrimination will be mandatory in contestable markets, when certain conditions are met, without specifying any requirements regarding the differentiation of the firms’ product lines.
How to explain this phenomenon (price discrimination without market power), which is notable for its ubiquity, bearing in mind that one of the accepted fundamental prerequisites for the feasibility of price discrimination is the existence of market power?  

The core development of the present work is the explanation of the sustainability of Nonlinear Pricing, and price discrimination in general, in environments of perfect contestability (a concept that encompasses as a special case the notion of perfect competition), partial equilibrium models for various types of markets having been deduced. Thus, Nonlinear Pricing systems emerge as mechanisms that, in competitive environments, and with minimal central planning (centralisation of information), will allow greater social efficiency, constituting a “second-best invisible hand”; theory endows thus, a greater adherence to reality. The practical implications of this new competitive model are diverse, from which we highlight its influence in regulatory and competition policies.

The most relevant approaches towards, and recent contributions related to, this issue (and that inspired the present article, setting its path), are discussed in Section 2. Section 3 explains the sustainability of Nonlinear Pricing, and of price discrimination in general, in perfectly contestable environments, partial equilibrium models for various types of markets having been deduced. Section 4 revisits the question of market power, analysing alternative definitions and tests. Implications of the analysis for Economic Theory are discussed in the final section.

---

6 The literature (see, for instance, Stole (2007: p. 2226)) establishes that price discrimination is feasible only when certain conditions are satisfied: (i) firms have market power, at least in the short-term, (ii) consumers can be segmented directly or indirectly (there is a segmentation tool), and (iii) arbitrage is impossible, limited or controlled by the original producer (“Resale exclusion is essentially an elaboration of the requirement of monopoly power, because it prevents competition from customers acting as secondary suppliers in resale markets” (Wilson, 1993: p. 11)).
2. RECENT APPROACHES IN THE LITERATURE OF ECONOMICS AND ANTITRUST

In recent years, relevant literature has emerged dealing with the theme of the present article (and inspiring it), namely through its implications in Antitrust.

2.1. LEVINE (2000) and the IMPORTANCE OF COMMON COSTS

The article that launched the debate is Levine (2000), the main motivation of which was to prove, essentially to regulatory and antitrust agencies, that the presence of price discrimination is not by itself an indicator of market power (this is not a necessary or sufficient condition), since it is a phenomenon observed in competitive markets where firms are actors without any market power. Levine argues that the presence of common costs is what induces the need for firms to adopt the practice of price discrimination as a way to maintain their viability (and only the viability – zero economic profit) in the least production-restrictive way.

Levine does not explicitly separate the treatment of mono and multiproduct firms, and for the former, the appointment of common costs will certainly refer to fixed costs: “What producers of competitive products and services that share common costs do is to charge different customers who buy different units or different products different prices” (Levine, 2000: p.7); implicitly the first reference is directed to single product situations, while the second indicates multiproduct contexts.

---


8 Levine attributes the lack of market power of firms in a particular industry to free entry into said industry, according to the Theory of Contestability (Baumol et al. (1988)), see also footnote 4).
But Levine, in the multiproduct approach, which relies mainly on the case of joint production\(^9\), confuses the natural, predictable, and documented price difference between joint products with price discrimination\(^10\). Indeed, the products produced jointly are distinct, so that the price difference in competitive markets regarding these goods does not result from price discrimination as established robustly by Mill-Marshall joint supply theory\(^11\). There would be price discrimination in these joint production cases if the price differential in the supply of quantity and / or quality of each product set was not entirely justified by cost differences.

We consider that the great merit of Levine was to initiate a debate about the practice of price discrimination in competitive markets, and to underline the crucial role of common costs, which met an important development by William Baumol and Daniel Swanson (Baumol 2002 and Baumol and Swanson 2003).

\(^9\) Levine uses as a practical example the case of beef processors, among others, in line with one of the mentioned cases in John Stuart Mill’s pioneering approach (1848) (see footnote 10)

\(^10\) The pioneering approach on the Theory of Joint Supply belongs to John Stuart Mill in his 1848 work “Principles of Political Economy”, where Mill has established the principle concisely (Alfred Marshall just developed its graphical analysis (Ekelund and Hulett (1973)) . Indeed, the apparent difficulty in distinguishing price discrimination and joint supply situations dates back to at least the famous controversy between Taussig and Pigou on railway rates; Taussig (1891) held that the observed pattern of multiplicity of railway rates was conveniently explained by the Mill-Marshall theory of joint supply (Taussig argued that the supply unit, at a scale relevant to establishing the price, is not homogeneous), while Pigou (1912) explained this pattern through the presence of high common costs accompanied by the feasibility of price discrimination among customers, given the monopoly situation coupled with the presence of the necessary conditions for the practice of price discrimination (Pigou believed that the supply unit is homogeneous). See Ekelund and Hulett (1973) and Ekelund and Hébert (1997, Chapter 17) for a detailed description of the mentioned controversy.

\(^11\) While it is generally agreed that price differences between different products do not constitute price discrimination, Harold Demsetz raises such a possibility in cases where joint supply is monopolised, being able to check for diverse relations between joint product prices and their marginal costs; this would be a case of price discrimination if not dealing with different products (Demsetz 1973: p. 390-1). This observation could certainly lead us to the debate on used criteria to classify products as different or just horizontally differentiated, but it takes us away from the purpose of the present work.
2.2. BAUMOL (2002) and the COMPETITIVE DYNAMIC EQUILIBRIUM

Baumol (2002), inspired by Levine (2000), and already working in the polar (and hypothetical) environment of perfect contestability (which includes the perfectly competitive markets as a particular case), arrives at a central result that postulates the existence of competitive equilibria (in industries where economies of scale are important), with price discrimination:

“In a perfectly contestable market in which a seller can separate customers into distinct groups with different demand elasticities, and can prevent its product from being transferred from one customer to another, the firm’s equilibrium economic profits will be zero. But, except for cases constituting a set of measure zero, avoidance of losses will require discriminatory prices.” (Baumol, 2002, p. 3).

Baumol proves this proposition succouring in three observations, which derive from basic economic analysis:

- If consumers can be divided into subsets with different elasticity of demand, the discriminatory pricing generates higher profits than uniform pricing\(^{12}\)-\(^{13}\);
- The absence of barriers to entry and exit nullifies the possibility of positive economic profits, but makes it possible to avoid a negative economic profit (i.e., it does not prevent incumbent firms from covering all their costs, including common costs, fixed and recurring sunk costs);
- The zero profits a firm can earn in a contestable market equilibrium can only be obtained from its profit-maximising price vector, which is normally discriminatory.

Some remarks are necessary. In the case of a (natural) perfectly contestable monopolistic market, the novelty of the proposition is not

---

\(^{12}\) As Baumol and Swanson (2003: p. 673, footnote 24), recall, appealing to a notion of elementary optimisation theory), “…an unconstrained maximum is always at least as great as the corresponding maximum under constraint…” (profit maximisation constrained by the use of uniform pricing).

\(^{13}\) It is important to note that the use of Second Degree Price Discrimination (Nonlinear Pricing) may be a vehicle to overcome the inefficiencies posed by asymmetric information (especially in adverse selection and moral hazard environments) and thus avoid the “narrowing” or even destruction of the market.
significant, since Baumol *et al.* (1988: Chapter 8) reported that in such a market Ramsey discriminatory prices are sustainable (sufficient, but not necessary condition) against entry\(^{14}\). The great novelty of Baumol’s proposition (2002) relating to his previous work concerns the equilibrium in perfectly contestable markets that contain two or more firms (i.e., not natural monopolies); effectively, Baumol *et al.* (1988: Chapter 11) produced a theorem which stated that in such markets the only sustainable price would be equal to marginal cost. But according to Baumol (2002), usually at least some of the prices of incumbent firms will be greater than the marginal cost to permit coverage of sunk, fixed and common costs\(^{15}\); Baumol (2002) admits

---

\(^{14}\) This hypothesis is condensed in their Weak Invisible Hand Theorem (Baumol *et al.* (1988: chapter 8)). Note that, according to the authors, the concept of sustainability is an economically substantive generalisation of classical competitive equilibrium and, correspondingly, the notion of perfectly contestable market is an equally substantive generalisation of perfectly competitive markets (Baumol *et al.* (1988, p.271)). It should also be noted that, traditionally (see, for example, Braeutigam (1989, pp.1320-7), and this would be the case in Baumol *et al.* (1988), Ramsey Pricing in multiproduct environments does not imply nonlinear prices for each product, but rather linear prices according to the inverse elasticity rule for each product / market; it should be noted at the outset that efficiency can be further increased if the firm practices nonlinear pricing for each product (which puts into question the sustainability through Ramsey pricing with linear prices per product, as argued in the Weak Invisible Hand Theorem; indeed, Panzar and Postlewaite (1984) and Shaffer (1987) had already shown that nonlinear prices dominate linear prices in contestable natural monopolies (regarding non-decreasing marginal costs) via its Pareto superiority (being that a Pareto-inefficient rate is unsustainable), after Mirman and Sibley (1980) having had reported that a multiproduct monopolist with increasing returns to scale offers a nonlinear price tariff, and Willig (1978) having proved that it is always possible to construct a Pareto-superior tariff to one that employs linear prices different from marginal costs), also overcoming in this manner the issue of asymmetric information and its harmful effects on the performance of these markets.

\(^{15}\) Indeed, Baumol *et al.* (1988) considered that the practice of marginal-cost prices was economically viable under the assumption that economies of scale were “exhausted”, and the quantities demanded by the market at these optimal prices were superior to the Minimum Efficient Scale (MES) of the firm in question (see Baumol *et al.* (1988: p.359). At his recent approaches above (see p. 44), Baumol essentially targets “new economy” industries (Biotechnology, Software, Semiconductor, ...) that, due to significant sunk , fixed, and common costs, eventually “force” firms to offer certain products at production levels for which there are specific economies of scale (see also footnote 17).
that this phenomenon is an invitation to entry\textsuperscript{16}, which helps to produce a dynamic equilibrium (“churning” equilibria) with price discrimination\textsuperscript{17}. Zero economic profit for the industry means that the entry of firms into the industry requires the exit of the incumbent or of some newly-entrees.

\textsuperscript{16} Baumol (2002: p.19, footnote 14) notes that this invitation for entry has a limited practical applicability, since in this framework the entrant would have to replicate the incumbent’s product line almost completely, while still referring that this is one of the reasons why the threat of entry may reduce profit to zero without excluding price discrimination. Other factors also contribute to the fact that the referred invitation for entry is not so effective, such as uncertainty (the argument assumes that the entrant has very precise estimates of their marginal cost and demand elasticity – which allows for evaluating the demand response to a small reduction in price by the entrant), the established reputation of the incumbent, inertia of consumers to switch suppliers and the probability of a direct deterrent reaction by incumbents through their pricing (Baumol and Swanson, 2003: p.677).

\textsuperscript{17} The vision of Contestability Theory (see Baumol et al. (1988)) was that an oligopoly operating in a perfectly contestable market would be forced to operate with marginal-cost prices in equilibrium, as already mentioned. This first-best situation required that business costs were trans-ray convex (see Baumol et al. (1988: Ch 4D) to clarify the concept of trans-ray convexity, which is directly related to the existence of scope economies; we can briefly refer that this means that in the multiproduct context scope economies counterbalance the effect of specific economies of scale, and thus, multiproduct production is relatively less costly than specialised production), which, with their scale economies “exhausted”, their unit costs remained constant for significant expansions beyond MES, and that the quantities demanded by the market at optimal prices (made to marginal costs) were superior to the MES (Baumol et al, 1988, p.359). However, Baumol (2002) states that viability requires that at least some prices are higher than marginal costs (which does not necessarily imply price discrimination, see footnote 14, in his reference to Ramsey pricing), contradicting its previous position; as mentioned in footnote 15, the new conviction is due to the conjecture that certain products will have to be offered in the area of specific economies of scale due to the magnitude of sunk, fixed, and common costs. But Baumol fails to explain the sustainability of the practice of price above marginal costs (and therefore the sustainability of the supply of products with economies of scale in an oligopoly), because a competitor could increase the supply of product up to the area of scale diseconomies, where the marginal-cost price would generate solvency (assuming fixed capacity at its optimum level); recognising in a way this gap, Baumol notes that such a configuration will not be perfectly sustainable, constituting a call for entry despite the expected null economic profit postulated by free entry (Baumol, 2002: p. 18-20).
3. CONCEPT CLARIFICATION AND THE CRUCIAL NOTION OF SUSTAINABILITY

Baumol (2002), Baumol and Swanson (2003), and Levine (2000) based the outcome of competitive equilibrium with price discrimination – in the line of Ramsey pricing in multiproduct environments – on the need that companies have to cover their costs, especially common, fixed, and sunk ones (Levine particularly stresses the role of common costs).18

These authors do not express separation between markets of differentiated products and perfect substitutes because they are indeed dealing with price differentials between different, or even joint products manufactured by multiproduct firms traded in distinct markets.19,20

However, the observed phenomenon that we want to study in this work (and which actually constitutes price discrimination, in economic terms) is the generalisation of price differentiation between identical and similar products (quality differentials), not entirely explained by differences

18 Note that there is an apparent contradiction between the proposal of Baumol (2002) and the role of sunk costs arising out of Contestability Theory, of which Baumol was a major mentor (see Baumol (1982)); effectively, the absence of sunk costs constitutes a precondition for a market to be contestable. Baumol (2002) explains that conflict disappears, indicating that the sunk costs that are important for his analysis, and one of the main reasons for the business practice of price discrimination, are recurrent (e.g., continuous expenses for R & D in companies in “hi-tech” industries), which due to this characteristic do not constitute barriers to entry since they are “obligations” extended to incumbents of such industries, thus not conferring them with competitive advantage.

19 See Carbonneau et al. (2004: p.3) that in a critique of the extension of the basic model of price discrimination by Levine (2000), and apparently detecting such “confusion” (which in our opinion stems from the concept of Ramsey pricing in multiproduct environments and its traditional connection, which is direct in Nonlinear pricing, with price discrimination) report: “..., and differences in product offerings are not necessarily price discrimination.”

20 It should be noted that Baumol (2002) commences by addressing price discrimination for a product offered by a multiproduct firm (which will be an actual case of price discrimination), but in the discussion regarding sustainability of this practice succours to frameworks provided in Baumol et al. (1988) that for the multiproduct case do not contemplate nor directly concern the possibility of nonlinear prices for each product.
in costs (definition of price discrimination) seen in markets considered as competitive\textsuperscript{21}.

It is paramount, thus, to explain how price discrimination can be sustained without market power, which eventually will provide indications on the types of (partial) equilibria developed in such markets\textsuperscript{22}.

Firstly, it should be noted that the existence of absolute free entry and exit in a market (i.e. perfect contestability) is a sufficient condition for the firms operating in this market, regardless of any other factors, not having any power market\textsuperscript{23}. Following the investigation we set ourselves described in the previous paragraph, we also evaluate whether the indicators traditionally used to infer the existence of market power by a firm (e.g. Lerner Index) are effectively relevant.

\textsuperscript{21} Such phenomenon is described and exemplified earlier in this work (see Introduction).

\textsuperscript{22} It must be emphasised that we consider the assertions of Baumol (2002) and Baumol and Swanson (2003) to be valid with regard to price discrimination in multiproduct environments being mandatory in perfectly contestable markets (see pp. 44-6); in the course of our sustainability analysis we provide and explain the reasons why we believe these assertions are valid.

\textsuperscript{23} See footnote 14, where it is mentioned that the notion of perfectly contestable market is also a substantive generalisation of the perfectly competitive market (Baumol et al. (1988: p.271)).
3.1. Markets of homogeneous products / monoproduct firms

In the case of competition between monoproduct firms providing perfect substitutes we have actually to assume some type of extensive game (e.g., the Cournot game, as suggested in Wilson (1993)) in order to prevent in the short-term the outcome of uniform marginal cost pricing\(^{24}\), \(^{25}\), or to adopt the vision of Baumol (2002) (for which in the present case, a market populated by monoproduct firms applies mutatis mutandis) that it has generated a dynamic equilibrium (“churning equilibria”, see pp. 45-6), since in a perfectly contestable oligopolistic market, prices above marginal cost are not sustainable\(^{26}\).

\(^{24}\) “... Cournot competition among several firms drives marginal prices down, as compared to monopoly, but it does not eliminate nonlinear pricing. Profit margins vanish only if the number of firms is very large, since the percentage profit margin is \(1/n \eta (p, q, q)\).”...; consequently, the number of viable firms in an industry is limited whenever investment and fixed operating costs must be recovered from profits. Nonlinear pricing is then a predictable outcome, ...” (Wilson 1993, p.307 and p. 312), explaining aspects of the outcome of the Cournot game, being implicit in this argument that firms operate with excess economic capacity (excess capacity estimated at each instant the difference between economic capacity – production at minimum average total cost – and actual production (Amada da Silva (1991: p.50)), see also footnote 29), but not as high as it would be if they were limited to employing linear pricing. It should also be noted that this type of equilibrium is not sustainable under free entry and does not constitute a true equilibrium in contestable markets (see Baumol et al. (1988: p. 332)).

\(^{25}\) Mandy (1991), Hayes (1987), and Locay and Rodriguez (1992) presented models where binomial tariffs dominate uniform pricing in competitive homogeneous product markets; however, these studies were aimed at special cases where binomial tariffs facilitate transfers between individuals that are not possible with uniform prices.

\(^{26}\) “The conclusion, then, is that the structure of the equilibria described here can easily invite entry despite the zero expected profits for any pertinent industry as a whole. But those zero profits mean that the market has no room for both incumbents and entrants, so that after entry the demise of some firms can confidently be expected. Consequently, constant and unending strategic battle for survival must be common, with the time trajectory characterised by a ceaseless inflow of entrants, followed by a stream of exiting firms composed in part of entrants, in part of former incumbents.” Baumol (2002: p. 20), generally describing the dynamic equilibrium (“churning equilibria”) that is generated in contestable markets due to the non-sustainability of prices superior to marginal costs, which, in the outlined conditions are discriminatory due to a profit maximisation posture, which in view of the industry concerned, will never be positive.
It should be mentioned that we are, in the present work, omitting the case where consumers differ in their information on market prices (or the cost of acquiring it). Several competitive models, considering this situation, have been presented and analysed (Varian 1980, 1981, Salop and Stiglitz 1977, 1982, Rosenthal 1980, and Stahl 1989); a common result is that, in equilibrium, each firm offers a unique and uniform price (average prices increase with the proportion of non-informed consumers, or, from another perspective, the number of firms operating in the market), and price distribution in the market segments the consumer population between informed and non-informed buyers. It must be emphasised that said price discrimination occurs between firms and not at a firm level; the only exception is in Katz (1984) where by introducing additional assumptions to the model of Salop and Stiglitz (1977) (consumers who do larger purchases undertake a higher level of price research and tend to be better informed than consumers who carry smaller purchases), which is effectively a model of monopolistic competition, has equilibria with second degree price discrimination formally similar to that obtained in section 3.3.1. of this work (which focuses on the sustainability of Nonlinear Pricing in monopolistic competition).

3.2. Homogeneous product markets / multiproduct firms

Let us now consider the case of homogeneous product markets served by multiproduct firms.

In perfectly contestable markets, each multiproduct firm will be required to produce (efficiently) a cost minimising output vector\(^2\). Due to the presence of economies of scope, sufficient and necessary condition for the presence of multiproduct firms (see Baumol et al. (1988: p. 248-9)) (where one of its sources may be the joint production), the outputs forming the referred output vector are also interdependent\(^2\). Thus, for there to be a

---

27. “No firm will be able to operate inefficiently without risk of entrant takeover of its customers. A firm’s scale of operation can neither fall short of nor exceed the cost-minimising level, because that would make it vulnerable to entry.”; Baumol and Swanson (2003: p. 670), which underlines the “requirement” of the firm being operationally efficient and producing at a cost-minimising scale in a contestable market.

28. The real interdependence of outputs arises during joint production (either in fixed or variable proportions); when economies of scope are not based on joint production
situation where total average costs will be minimised, it is very likely that some products are produced in overcapacity and others in a situation of excess capacity\textsuperscript{29-30}.

of outputs, interdependence arises in the logic of cost minimisation. It should be emphasised that, where there is joint production, Nonlinear Pricing of each joint product allows for an easier coverage of joint production costs (since it would be much coincidence that the uniform (linear) market price for each joint product allowed exactly to cover the whole joint cost of each firm, considering varying or fixed proportions in non-symmetric firms (see footnote 32 for reference to the issue of symmetry); effectively, one can easily see that many joint products are offered through non-linear tariffs in competitive markets: ex., quantity and quality discounts in beef).

\textsuperscript{29} From among the various output vectors belonging to the M locus (“...the set of all output vectors that minimise ray average costs along their own rays”; see Baumol et al. (1988: pp. 58-9)), only some will be technologically feasible. For these vectors, and due to the interdependence of the outputs of the various products (already embodied in the feasibility of such vectors), the production of some products, will most likely take place in the area of increasing average incremental costs (specific diseconomies of scale; overcapacity), “counterbalanced” by the production of others who will be in the area of declining average incremental costs (specific economies of scale; excess capacity) (see Amado da Silva (1991: section 2.2.) for definitions of capacity and excess capacity; it should be emphasised that the concept of capacity that we use in this work is “economic capacity”; see also footnote 24)); it will be too much coincidence if the outputs of the various products meet a range of sufficiently small amplitude, such that all would be in the area of specific diseconomies of scale without their marginal costs outweighing average stand-alone costs, as claimed to be the only possibility by Baumol et al. (1988: p. 255-6) (see also footnote 30).

\textsuperscript{30} Baumol et al. (1988: p. 254-256) denied the possibility of some products being offered with specific economies of scale. However, these authors based such belief on a cost-benefit analysis the firm should make in considering to abandon the production of a product or product subset, where the cost will be provided by the income so exempted, accounted for sales of such products at prices equal to their marginal costs, and the benefit for the cost of production spared indicated by incremental costs. Thus, the ratio between the benefit and cost corresponds to the degree of scale economies relating to that product or product subset, which would be equal to or less than unity (lack of specific economies of scale) for the firm not to have advantages in abandoning the production of these products (it should be noted that this inability of the firm to offer products with some specific economies of scale implies that the proportion in which its various products are offered belongs to a fairly limited range, since no product can be offered at a level where its marginal cost of production, besides having to be greater than or equal to the average incremental cost, does not exceed its average stand-alone cost). However, as already referred, these authors assume that, as stated above, cost must be accounted for prices equal to the marginal cost, thus closing at the outset the possibility to practise linear or
Due to the pressure of free entry, the maximum profit generated by the firm will match, as already mentioned above, the zero economic profit, only allowing businesses to cover their total costs. Thus one realises that in the situation described in the preceding paragraph, the prices equated to marginal costs cannot cover the incremental costs of products offered with specific economies of scale, so that to have solvency, at least some of the prices of products offered with specific diseconomies of scale (formulated to marginal costs) would have to overcome their stand-alone costs (average); since this scenario implies the existence of cross-subsidisation, and considering that sustainability implies its absence (Baumol et al. (1988): p.202), the overall price of the products offered in excess capacity must be at least equivalent to their average costs (incremental), thus, superior to their marginal costs.

Having established the feasibility of maintaining prices higher than marginal costs, nonlinear tariffs Pareto-dominate, under these conditions, linear tariffs (see pp. 44-6, including footnote 14), since a nonlinear tariff constructed in accordance with the inverse-elasticity rule for each differential market being social efficiency maximising (incidentally, Baumol et al. (1988: p. 202) showed that a sustainable price vector must be non-dominated). The effect of competition is exerted in the form of increased elasticity of the demand profile (Wilson, 1993) facing each company in each market, assuming deterministic participation (in accordance with the canonical model of Musa and Rosen (1978), where participation is determined from the decision of the lowest utility consumer), as opposed to random participation models, such as in Rochet and Stole (2002) – where the decision to participate, although depending on the type of consumer, is random and not determined by the lowest utility consumer’s decision; the elasticity of the demand profile faced by a firm for each product (it should be noted that we assume the demands of the various products are independent) depends on the number of competitors and their tariffs, the nonlinear prices that stray from marginal costs (“..., to the system of non-discriminatory simple prices that are taken throughout the analysis to constitute the only form of prices available” (Baumol et al., 1988, p.333)), recognising that their analysis excludes the presence of nonlinear prices). Thus, the new ratio accounted for (i.e., with nonlinear pricing) will be less than the degree of specific economies of scale, so there may be situations in which the firm has no advantages to abandon the production of product subsets for which there are specific economies of scale.
firm establishing tariffs, for each product, minimising the overall cost and maximising income (which can only match the achievement of zero economic profit) \(^{31}\).

The contribution to the common costs will be made through the revenue generated by the products offered in overcapacity, the prices of which cannot exceed the respective stand-alone costs (average).

Each firm will “initially” have an optimal price-output vector that will lead to cost minimisation and profit maximisation, which corresponds, under free entry, to achieving breakeven. A competitor (or a candidate for entry) may have a different optimal price-output vector, which may consist in lower prices within a given market \(^{32}\); these lower prices would press

31 Assuming symmetry, the nonlinear tariff, in equilibrium, for a product offered with specific scale economies (product i, hereinafter) is obtained as follows: once sustainability in contestable markets requires that the multiproduct firm we are considering operates with outputs leading to the minimum cost, the amount of product i offered is determined and exogenous to the problem we are addressing; for this amount, determine (Ramsey Pricing, see Wilson (1993, Ch. 5)) the nonlinear tariff which can cover the incremental costs of that product; the elasticity of the demand profile faced by the firm (which will be the basis for determining the Ramsey number that allows the coverage of the incremental costs of the product offered with specific economies of scale) is given by multiplying the number of firms (in these conditions, sufficient to provide the market) by the elasticity of the market demand profile (up to the limit of elasticity that, with a unit Ramsey number, only allows for the recovery of incremental costs).

32 We are not, therefore, requiring symmetry between firms at this stage of the reasoning, which gives it added practical applicability (actually, symmetric multiproduct firms hardly exist in the “real world”). If the market were formed by symmetric firms, they would have in equilibrium to adopt a discriminatory pricing vector (at least for some products) that maximises profits (identical for all firms) which, however, only allows them to cover their total costs (see footnote 31); this discriminatory pricing vector is sustainable given that a potential entrant, besides having to replicate the incumbent’s entire product line (not to be at a disadvantage with respect to scale and scope economies; it is clear that the relative positions of the companies in respect of costs depend on the amount of fixed costs, so niche players, who generally avoid most of the sunk costs of its competitors, will not be disadvantaged, despite its narrow product line or reduced outputs – being that the usual strategy of these niche players encompasses hope for survival in these particular segments and expanding its investment to become effective rivals, although for a variety of reasons, including its own limitations, they often succumb (Baumol and Swanson (2003: p. 676)), if he tried to enter with marginally lower prices than
to an increased output of this firm in this market, but as there is interdependence of the outputs of its various products, this increase would push it away from the outputs leading to the operation at minimum cost of the other products, so it will be constrained to curb the increase in output in the first market. Meanwhile, the other firm, feeling to lose market share, would be obliged to react to the price level in the market not to stray too far from the optimum output, but the “story” is repeated due to its own interdependencies.

Thus, market forces impose not the adoption of the optimal vector output-prices for each firm, but rather an interval of discriminatory prices in equilibrium (identical for all firms), resulting in (better or worse) approximations to the optimal price-output vector for each company. Firms that operate closer to their optimal price-output vector have a competitive advantage, which is unlikely to be sustained, but rather temporary, once the equilibrium is unstable.

3.3. Differentiated product markets

Where there is differentiation between the lines of products offered by the firms, the conventional analysis conjectures that differentiation will be enough for each firm to face a non-perfectly elastic demand, which allows for them to have some control over its prices (but not on overall market

those of incumbents in a market (which could be profitable; see Baumol et al. (1988): p. 317-8 and earlier references in this work on this subject: pp. 45-6 and footnotes 15 and 16), he would alter his globally minimising cost position, given the aforementioned interdependence of the outputs.

33 If the firm loses significant market share in that market, it will alienate its minimising production cost vector. But if it changes significantly its nonlinear tariff in that market, repercussions will be expected in its position within other markets where it operates.

34 Thus, this interval of discriminatory prices results from a kind of “consensus” among the firms operating in that market; it follows that the equilibrium is not stationary nor modifiable just by exogenous shocks, but it is instead highly dynamic, which is in accordance with the frequent price changes (daily in nature in many instances; see for example the case of airlines) which occurs in highly contestable markets.

35 Moreover, it would be more appropriate to say that there is an ongoing process for achieving equilibrium through a “repeated non-cooperative game”.

prices, under contestability conditions). This control over their own prices does not permit positive economic profit under free entry and exit.

The obvious conjecture is that, given the position and elasticity of the demand curve faced by a firm, which depends on the degree of market entry freedom (of substitutes) and the substitutability of the competing products, that firm will be induced to practise price discrimination in order to increase its surplus. Surely this practice is complicated by the effect of competition, but the role of price discrimination remains essentially unchanged, with advantages for the firms, from a profit point of view, from the differentiation of quantity and quality (which will also lead to increased social efficiency). Under conditions of perfect contestability (which, as noted above – see footnote 14 – is an economically substantive generalisation of perfect competition), the surplus obtained by the firms does not, however, permit positive economic profit.

3.3.1. Differentiated product markets / monoproduct firms

In the case of differentiated product markets populated by single product companies, the heterogeneity between the offered products enables by itself the maintenance of imperfectly elastic demand curves under conditions of perfect contestability, in the tradition of the Theory of Monopolistic Competition (Chamberlin 1933).

We consider that, according to the model of Wilson (1993, Chapter 12.2) (it should be noted that it is not a random participation model), firms face induced demand profiles (by the effect of competition), maximising their profit through an optimal nonlinear tariff (inverse-elasticity rule for each differential market), and mandatory (see footnote 22), which under free entry only allows for zero economic profit (free entry influences the induced demand profiles)\textsuperscript{36}.

\textsuperscript{36} “The outcome from this jostling for competitive advantage is a set of tariffs, one for each firm, such that each firm’s tariff is the optimal response to the others’ tariffs. This outcome is called an equilibrium of the game of tariff competition between the competing firms. Direct calculation of equilibrium tariffs is complicated by the interaction between the firms, but a simple procedure usually suffices: as in real life, repeatedly construct each firm’s new optimal tariff in response to the others’ firms’ current tariffs. This procedure usually converges quickly to the set of
The excess capacity is not significant (in contrast to what is postulated by the theory of Monopolistic Competition; the firm, practising price discrimination, is not “obliged”, as the theory suggests, to operate in a linear price – output combination on its long-run average cost curve (CM)\(^3\), depending on the relative positions of the cost curves and induced demand (D), being that the latter, by the effect of free entry, is diverted more to the “left” and its elasticity increases, as Fig 1. tries to depict; its position will be such that will just permit, through the revenue earned by the practice of Nonlinear Pricing, to reach only zero economic profit\(^3\).

**Figure 1. Monopolistic Competition with Price Discrimination**

---

37 The well-known Chamberlinian tangency solution (of the demand curve with the long-run average cost curve) seems, however, to be engrained enough (although Baumol et al. (1988: p. 330) underline that this solution for the “large-group case” is not the only one to be sustainable if we relax the assumption of profit maximisation in any period of time, a fact also recognised by Amado da Silva (1991: p. 141-2) building several strategic alternatives that are not limited to the “large-group case”); by way of example, Economides and Wildman (1995) depart from this solution to establish that equilibria with Nonlinear Pricing (more precisely, with binomial tariffs, postulating a solution of tangency of an average producer surplus curve) dominate equilibria with linear prices in monopolistically competitive markets, providing an improvement from the point of view of social efficiency.

38 Basically, we will have a Ramsey Pricing situation, but where the regulatory role of the Ramsey number on the profit margins will be exercised by the actions of competitors on the (induced) demand faced by the firm.
The equilibrium with price discrimination enables a decrease in excess capacity relative to equilibrium with linear prices (approximately yielding $q_{dp}$, contrasting with the output $q_{cm}$ provided by the traditional model of monopolistic competition – tangency solution), affording greater social efficiency, even without accounting for the effect on consumer surplus of a possible offer of greater product variety.\(^{39}\)

It should be emphasised that, in this way, the configuration obtained by the introduction of price discrimination is sustainable, since under these structural conditions, social efficiency is maximised (see footnote 14 in relation to domination and consequent sustainability of Pareto efficiency maximising tariffs in environments with increasing returns to scale and non-decreasing marginal costs; in the area of decreasing marginal costs, firms become more vulnerable to cream-skimming strategies and by-pass: see Panzar and Postlewaite (1984: p. 14) and footnote 40 on this possibility) unlike what happened with the generalised Chamberlinian structures (those that were not comprehended by the “large-group case”; see Baumol et al. (1988): p. 331-4)\(^{40}\).

---

\(^{39}\) It is noteworthy that this possibility of having a greater product variety stems from the increase in the surplus that each producer may obtain by expanding its production through the practice of price discrimination relative to equilibrium with linear prices; these additional profits induce entry of new competitors (leading to the erosion of said profits, leading to zero economic profit in the Industry), providing similar products, thus increasing the variety of products within the market.

\(^{40}\) Clearly, in conditions of decreasing marginal costs, a rival or entry candidate could develop a product similar to that of a particular incumbent and try cream-skimming (see p. 53, footnote 32). In this situation, the targeted incumbent would have to change its optimal price discriminatory tariff through price reductions for high demand segments and price increases for low demand segments, in order to derail this type of competition or entry, and then resume the optimal discriminatory tariff; if the response of incumbents to such threats is fast enough, this type of competition is discouraged and may be considered that the optimal discriminatory tariffs are indeed sustainable; see Elhauge (2003a: Ch III.C.) for a similar reasoning in order to show that reactive price reductions above costs by incumbents are not predatory, providing many practical examples (such as the policy of some retailers that automatically equalize any lower price that the competition might eventually offer).
3.3.2. Differentiated product markets / multiproduct firms

In differentiated product markets populated by multiproduct firms, regarding the sustainability of price discrimination in perfect contestability conditions, the reasoning constructed for the case of homogeneous products essentially remains (see section 3.2), but enhanced by the existence of consumer lock-in. The maintenance of offered quantities by firms that generate the minimum cost and of the resulting discriminatory pricing vector is strengthened by the existence of product differentiation, which is a sufficient condition for the firm to face an imperfectly elastic demand.

4. THE QUESTION OF MARKET POWER

The existence of price discrimination implies that the demand curves for the various products of a firm (assuming multiproduct firm; yet we weave considerations that also apply to single product firms) are negatively sloped, which is traditionally considered a sign of market power, since the seminal contribution of Lerner (1934) (Baumol and Swanson 2003)\(^{41}\).

However, as we have seen, in perfectly contestable markets, firms are driven to practise price discrimination as to recover common, fixed, and sunk costs (recurring; see footnote 18), maximising their output and obtaining zero economic profit, without possessing the power to restrict total supply in this market or its overall prices; i.e., without having market power (see the following paragraphs, including footnote 42, for commonly accepted definitions of market power).

\(^{41}\) The Lerner Index, equated by the percentage excess of price over marginal cost, tested the degree to which a monopoly was exploited, which, for a profit maximizing monopolist, corresponds to the reciprocal of the firm’s demand elasticity. While Lerner has emphasized that its analysis was restricted to situations of monopoly and would seek, above all, to measure the deviations from the social optimum regarded to be obtained under conditions of perfect competition with products offered at marginal cost (see Amado da Silva (1991: section 3.4.) for a detailed analysis of the contribution of Lerner in the broader context of the relationship between monopoly power and excess capacity), many authors, including those in the field of Antitrust, have linked the presence of market power with demand conditions faced by firms in various market structures (see Baumol and Swanson (2003: footnote 8) for a listing of significant references).
The key to overcoming this apparent paradox lies in the pivotal competitive pricing definition: these would equal marginal costs in conditions of perfect competition (perfectly elastic demand curve, Lerner index zero), but in conditions of perfect contestability (which cover – see footnote 14 – perfectly competitive markets as a special case) it would be conferred by the discriminatory pricing schedules that firms practise as a mere survival strategy dictated by structural market conditions (Baumol 2002). The traditional approach (see footnote 41) did not foresee the latter cases, where prices higher than marginal costs do not necessarily indicate the presence of market power, following that the indiscriminate use of the Lerner index may indicate market power when it does not actually exist. Klein and Wiley (2003a, 2003b) specify that, in their view, a firm with positive Lerner Index has “economic” market power, taking some action over their prices (but that, under free entry, will yet fail to deliver positive economic profits) but not necessarily market power “from the perspective of Antitrust”, understood as the power to restrict output in the market and hence increase the market prices above the competitive level (and in this way earn a positive economic profit).

Thus, the definition of market power as the power to place price above the competitive level, remains valid, yet little functional, since the competitive level can manifest itself in the form of a set of discriminatory prices that are difficult to define and quickly and unequivocally calculated (Baumol and Swanson 2003).

The alternate definition, which states that a firm has market power if it can restrict the total market output (by restricting its output), influencing the market prices as a vehicle to increase the income or profits (above zero economic profit) (Baumol and Swanson 2003)\textsuperscript{42}, remains as valid and functional, continuing to indicate the power to impose an inefficient outcome that harms consumers (as opposed to the maximising social efficiency result achieved through the aforementioned price discrimination – see pp. 50-4). Moreover, Baumol and Swanson (2003: pp. 682-3) argue that the evidence of monopoly power\textsuperscript{43} on the part of a business should be

\textsuperscript{42} Elhauge (2003a: p.727) notes that both definitions are often used by courts and scholars, referring us to the work of Werden (1998) and Klein (1993) for confirmation.

\textsuperscript{43} Baumol and Swanson (2003), recognising that the distinction between monopoly power and market power is not clearly defined, note that monopoly power usually
the demonstration of persistent and substantially super-competitive economic profits\footnote{Baumol and Swanson (2003) suggest its measurement in terms of rate of return on total investment in the firm.} supported by a sustained restriction of its offer (in order to show that those profits are not attributable to a superior performance of the company, resulting in a distinctive competence or the assumption of an extraordinarily high risk) \footnote{However, Baumol and Swanson (2003) recognise that there are substantial practical difficulties in the economic profit estimation methodology, particularly regarding the availability of relevant data and differences in corresponding accounting conventions and economic concepts. Therefore, they note that in many cases the evaluation of market power will continue to use information on market shares, barriers to entry and the ability of competitors to expand output, which must be analysed when circumstantial evidence is the basis of the presumption of market power.} \footnote{Freedom of entry and exit (Rothschild and Stiglitz (1976)) obviously prevents the described phenomenon (positive economic profits supported by a sustained restriction of supply) from taking place since it is not sustainable.}.

\section*{5. FINAL CONSIDERATIONS}

Once the crucial issue of sustainability of price discrimination in competitive markets (or, more encompassingly, contestable markets; as has been referred (pp. 41, 51; footnote 14), the notion of perfectly contestable market is a substantive generalisation of perfectly competitive markets (Baumol et al. (1988: p. 271)) has been clarified, the presence of increasing returns to scale is thus reconciled with competition (contestability) (except for the polar case of natural monopoly, for which that fact was already established by Baumol et al. (1988) through the \textit{Weak Invisible Hand Theorem}; see p. 45, footnote 14). Keeping the already assumed, and restated above, perspective that a market is competitive if there is freedom of entry and exit,
regardless of the number of firms operating, we have seen in previous sections that the price sustainability above marginal costs in these markets enables firms to efficiently reconcile (specific) economies of scale with competition. The crossing between widespread increasing returns to scale in modern economies being verified and the intuition that competition is important in order to achieve efficient results constituted a paradox indeed; such crossing was deemed incompatible according to received economic theory (Heal 1999) (a competitive market is efficient, but increasing returns to scale precluded competitive equilibrium 47), and according to the Theory of Contestability, competition (in a broad sense) forced firms to adopt prices at marginal cost (except in the case of natural monopoly), which required, for the purpose of economic viability, for economies of scale to be “exhausted”, their unit costs remain constant in significant expansions beyond MES, and demanded quantities by the market at optimal prices (equated to marginal costs) be superior to the MES, which was clearly “heroic” (see also footnote 17). The explanation of the sustainability of Nonlinear Pricing in competitive (contestable) markets in the present article contributes, thus, to reconcile the role of competition in the presence of increasing returns to scale.

Once the issue of increasing returns to scale in competitive environments has been addressed, it is important to address, due to its ubiquity, the issue of asymmetric information in these environments, which, as already noted, is identified in various models as being one of the most important determining factors for the non-existence of competitive equilibria (see Rothschild and Stiglitz (1976), who established that in environments with asymmetric information there may not be a competitive equilibrium due to lack of sustainability in Nonlinear Pricing systems; this article constitutes the ultimate reference in this issue – hence this non-existence hypothesis being generically referred to (see Gale and Hellwig 2004 and Schmidt-Mohr and Villas-Boas 1999), as the Rothschild-Stiglitz Problem).

47 Heal (1999: p. xvi) refers, in this way, to the mentioned paradox: “It is puzzling that, although economies of scale are undoubtedly important in reality, our belief in the invisible hand, in the efficiency of competition, seems verified by observation and experience, although not supported by current theory. This suggests that our understanding of economies with increasing returns is far from complete. There may be a role for competition and markets in allocating resources in the presence of increasing returns that we have not yet understood”
Just as a (natural) monopolist simultaneously solves the issues related to the maximisation of social efficiency with increasing returns to scale, as well as asymmetric information using Nonlinear Pricing systems, we also consider that firms in competitive markets simultaneously solve the two mentioned problems through the same systems.

The issue of social efficiency maximisation was dealt with in the previous sections (see section 3.); note now that if firm prices are discriminated through their second degree variant, they will also overcome the problems posed by asymmetric information by inducing customer self-selection. In fact, as mentioned in the introductory section of this study, it is observed that Nonlinear Pricing is practised in industries close to the criteria of perfect contestability (see footnote 3, for the enumeration of some of these industries in such conditions).

Thus, firms operating in industries close to perfect contestability respond – through the use of Nonlinear Pricing systems – to the difficulties posed by asymmetric information, which in adverse selection and moral hazard situations could lead to the documented “narrowing” of such markets to the point of their total collapse. This gives rise, within conceptual frameworks, to the possibility of the Invisible Hand (the doctrine of which primarily relates to the concept of competitive economic equilibrium efficiency in classic environments (Heal 1999)) acting in non-classical environments with increasing returns to scale and asymmetric information, generating through competitive markets not a first-best situation, but rather in the realm of second-best (since there is the need to impart information rents to agents holding private information relevant to each market in question), which will nevertheless maximise social

---

48 Let us recall that the concept of contestability encompasses that of competition, and just as perfectly competitive markets do not exist in reality, there are no perfectly contestable markets; there are good enough approximations and that is what this is about.

49 See Riley (2001) for an investigation into the developments, for twenty-five years, within Economics of Information.

50 The decrease in social surplus in relation to the first-best situation corresponds to the value of restricted quantity / quality by nonlinear tariffs constructed according to Nonlinear Pricing rules; it should be stressed that this decrease will, in each case, be only the minimum sufficient for a nonlinear tariff contributing its share to the firm’s breakeven, which is dictated by competitive pressures.
efficiency in these situations. The market system will only transmit all necessary information for economic agents in the form of prices if they have the right incentives for revealing their entire utility functions and production possibilities (private information) in return for information rents. The result for each market is not first-best because of the need to grant information rents, but it is more efficient than if the Invisible Hand were “not to function” due to the presence of increasing returns to scale and asymmetric information, as is traditionally postulated.

REFERENCES


51 Compare the foreseen result in this new theoretical framework with that foreseen in the Theory of Monopolistic Competition for differentiated products (the famous excess capacity result), and more generally in Economics of Information, which postulated the “narrowing” of markets, or even its collapse, affected by asymmetric information; note that the result obtained in Rothschild and Stiglitz (1976), one of the most important and memorable articles on competitive markets affected by asymmetric information, of lack of equilibrium within these markets, subjected to conditions, was based on the alleged non-sustainability of price discrimination in competitive conditions: with such sustainability, the possibility of equilibrium existence in competitive markets with asymmetric information will be restored.


