

The Nature of Services and the Implications for Competition Policy

Annemieke Karel*

May 21, 2004

Abstract

Though ‘services’ is a very broad concept and any generalisations should be made with caution, it is widely acknowledged that services differ from goods in many respects. Analyses of key elements of market structure, conduct and performance and the interface between these elements show us that competition in services markets is equally distinctive. We find that traditional competition policy instruments generally do not reflect competition and performance in most services markets. Alternative policy measures should recognise the importance and disciplining power of non-price competition and the fact that quality and reputation are the main variables in services competition.

JEL classification: D82, K21, L11, L80.

Keywords: Antitrust, Industrial Organisation, Services markets.

*The author acknowledges useful comments from Prof. Lewis Evans and Prof. Neil Quigley. Corresponding author: ISCR, PO Box 600, Wellington, New Zealand. Email: annemieke.karel@vuw.ac.nz New Zealand Institute for the Study of Competition and Regulation

Contents

1	Introduction	3
2	Market Structure	7
2.1	Monopolistic competition and oligopoly	7
2.2	Asymmetric information	9
2.2.1	Moral hazard	9
2.2.2	Adverse selection	11
2.3	Entry factors	14
3	Market Conduct	16
3.1	Non-price competition	16
3.2	Quality discrimination	18
3.3	Strategic firm behaviour	21
4	Market Performance	22
4.1	Product diversity	23
4.2	Advertising	25
4.3	Collusion	27
5	Evaluation of Competition Policy Instruments	30
5.1	Traditional performance measures	31
5.2	Concentration ratios	32
5.3	Market definition	35
5.4	Other forms of regulation	39
6	Conclusion	41

The Nature of Services and the Implications for Industrial Organisation and Competition Policy

1 Introduction

Produced by “unproductive labour” , services were “[. . .] of course, not part of the stock of wealth” in the minds of the great economic thinkers of the eighteenth and nineteenth century.¹ Two centuries later, services make up approximately 64 percent of world GDP² and rank high on the international political agenda.

Despite the growing importance of services markets, there seems to be little consensus on the definition of services. Still utilised today is the definition of Fisher (1935) and Clark (1940), defining services as “the tertiary sector”, which simply includes all forms of economic activity not included in the primary or secondary section (i.e. agriculture or manufacturing).³

⁴ In a seminal paper, Hill (1977) defines a service as: “a change in the condition of a person, or of a good belonging to some economic unit, which is brought about as the result of the activity of some other economic unit, with the prior agreement of the former person or economic unit”. He further distinguishes services affecting goods, services affecting persons, and labour services.⁵ Other services classifications can be found in Katouzian (1970) and Sampson and Snape (1985). The former distinguishes ‘new services’

¹Smith (1776); Marshall (1890).

²CIA Factbook (2003).

³See, for example, Productivity Commission (2002).

⁴Fisher (1935) specified that products of the tertiary sector consisted of “[. . .] facilities for travel, amusements of various kinds, governmental and other personal and intangible services, flowers, music, art, literature, education, science, philosophy and the like [. . .]”.

⁵Hill (1977) notes that capital services do not fit the definition developed in the paper as they are not an autonomous economic unit and thus do not provide services as such.

(almost coinciding with Fisher’s definition of tertiary products), ‘complementary services’ such as banking, finance, transportation, wholesale and retail trade, and ‘old services’ (activities which flourished before industrialisation and have declined since, like domestic service). In an influential paper, Sampson and Snape (1985) classify services based on how and where the services are produced and traded, in particular they consider whether the transactions occurred with or without movement of either or both the producer and/or consumer.⁶

Essentially extending Hill’s (1977) definition of services, Miles and Boden (2000) argue that the so-called tertiary sector can be seen as effecting changes in the state of the environment, artefacts (produced by the manufacturing sector), people, and symbols (i.e. information). They list a number of characteristics that are often seen as typical of services, the so-called “peculiarities of services” (see Table I).

Though many of the activities we think of as services usually have several of the characteristics listed in Table I in common, one can easily think of exceptions; professional services, for example. The multinational accounting and/or consultancy firm certainly has a very different ‘labour organisation process’ than does the local craftsman, yet we label both as ‘services’. Controversy exists also over the most frequently cited distinguishing feature of services: their intangible or immaterial nature. If a dentist gives someone false teeth or fillings - very material products - can we still consider this a service? What about the consultant that provides his services by means of a written report or cd-rom?

Hill (1997) states that “the habit of describing services as intangible products is an invention of economists” and adds that “services have noth-

⁶The work of Sampson and Snape (1985) is likely to have served as a basis for Article I of the General Agreement on Trade in Services (GATS), which sets out a comprehensive definition of trade in services in terms of four different modes of supply: cross-border, consumption abroad, commercial presence in the consuming country, and presence of natural persons.

Table I: Services characteristics

Service production	
<i>Technology and plant</i>	Low levels of capital equipment; heavy investment in buildings.
<i>Labour</i>	Some services highly professional (especially requiring interpersonal skills); others relatively unskilled, often involving casual or part-time labour. Specialist knowledge may be important but rarely technological skills.
<i>Organisation of labour</i>	Workforce often engaged in craft-like production with limited management control of details of work
<i>Process</i>	Production is often noncontinuous and economies of scale are limited.
<i>Features of production</i>	Some services state-run public services; others often small-scale with high preponderance of family firms and self-employed.
<i>Organisation of industry</i>	
Service product	
<i>Nature of product</i>	Immaterial, often information-intensive. Hard to store or transport. Process and product hard to distinguish.
<i>Features of product</i>	Often customised to consumer requirements.
<i>Intellectual property</i>	Hard to protect (can rarely be patented, though copyright or design rights may be possible), easy to copy many service innovations. Reputation is often crucial.
Service consumption	
<i>Delivery of product</i>	Production and consumption coterminous in time and space; often client or supplier has to move to meet the other party.
<i>Role of consumer</i>	Services are consumer intensive, requiring inputs from consumer into design/production process.
<i>Organisation of consumption</i>	Often hard to separate production from consumption. Self-service in formal and informal economies commonplace.
Service markets	
<i>Organisation of markets</i>	Some services delivered via public sector bureaucratic provision. Some costs are invisibly bundled with goods (e.g. retail sector).
<i>Regulation</i>	Professional regulation common in some services.
<i>Marketing</i>	Difficult to demonstrate products in advance.

Source: Miles and Boden (2000).

ing in common with intangibles, despite the practice of describing them as such". He emphasises two essential characteristics of services, namely that services cannot be produced without the agreement, co-operation and possible active participation of the consuming unit(s), and that the outputs produced are not separate entities that exist independently of the producers or consumers. He clearly distinguishes services from both tangible and intangible goods. *Goods* are entities that exist independently of their owner and preserve their identity through time. If ownership rights can be estab-

lished it follows that they can also be transferred from one economic unit to another, which implies that goods must be exchangeable. The owner of a good derives some economic benefit from owning it, in contrast to a ‘bad’, which has a negative exchange value. *Intangible goods* are described as intangible entities which have all the economic characteristics of goods, that are originally produced as outputs by persons, or enterprises, engaged in creative or innovative activities of a literary, scientific, engineering, artistic, or entertainment nature. In contrast, *services* are not separate entities that exist independently of the direct interaction between producers and consumers. They can therefore not be stocked and ownership cannot be transferred. In addition, quantification of services is much less informative than it is in the case of other commodities. Not only are services difficult to measure, it appears that quality is the more relevant variable, not quantity. Hill’s notion of services appears to be the most accurate one presented in the services literature to date and will therefore act as a reference point for the remaining part of this paper.⁷

Though there has clearly been much thought on services and their increasing role in today’s economies, an important issue that is not only missing in Table I, but in the services literature in general, is services *competition*. If we accept that services are different from goods and even intangible goods, that their production and consumption is different, and that their markets are different, it seems only natural to think that competition within these markets is different as well. If this is the case, there are important implications for competition policy in services markets. It is the purpose of this paper to examine the nature of services competition and to determine whether this justifies an adaptation of current competition policy mea-

⁷As becomes clear from Table I, “services” are a wide concept, encompassing a broad range of activities that have a number of common features, but may at the same time differ in market-specific characteristics. It should be noted that the structure, conduct, and performance analyses that follow include parts that do not necessarily apply to *all* services markets. Where this is the case, it will be clearly indicated.

tures. Section 2 looks at the characteristics of the prevailing market structures in most services industries, particularly asymmetric information and the existence of barriers to entry. Section 3 analyses the kind of non-price competition present in services markets and the associated firm behaviour. Section 4 combines economic theory and empirical evidence to picture the performance in services markets, focussing on product diversity, advertising, market power, and collusion. Section 5 evaluates current performance measures and other competition policy instruments and their appropriateness in providing a framework for efficient competition in services markets. Finally, Section 6 sums up the main findings and concludes on the nature of services competition and the implications for competition policy.

2 Market Structure

2.1 Monopolistic competition and oligopoly

The market structures observed in most services industries are best described as either monopolistic competition or oligopoly. The basic features of these market structures are listed in Table II. The most obvious structural

Table II: Market structure features

<i>Monopolistic competition</i>	<i>Oligopoly</i>
many small buyers	many small buyers
many small sellers	few large sellers
heterogeneous products	hetero-/homogeneous products
buyers may (not) be informed	buyers may (not) be informed
free entry	entry factors

feature of services markets is the fact that the ‘products’ are heterogeneous. The very nature of services production implies that no two services are exactly the same, as service provision at any time is determined by the provider

and consumer at that particular point in time.

Second, most services markets are characterised by many small buyers, in the sense that no one of them can individually influence the market price. Where there are many sellers, the service market in question is described as monopolistic competition. Monopolistically competitive services markets are for example markets for hairdressing, carwashes, travel bookings, etc. These markets are characterised by a large number of providers, no one of which is large relative to the overall market, and all of which try to differentiate their service from those offered by other providers. Where there are increasing returns to scale, network effects, high sunk cost, or (other) determinants of entry, a few large providers may serve the market instead, in which case we speak of an oligopoly. Section 2.3 will elaborate on possible barriers to entry that may cause an oligopolistic market structure in various services markets.

A final characterising feature of both monopolistic competition and oligopoly is the possible presence of asymmetric information. This is a particular issue in services industries. Service suppliers, for example, typically have more information about the quality of the service provided than most customers. In some cases, customers cannot even evaluate the quality of the service after purchase/consumption.⁸ They can therefore not accurately assess the price/quality relation offered by any service supplier, which allows the latter to price above marginal cost. So-called “moral hazard” problems like this are often present in services markets. In addition, high-quality services may be driven out of the market due to “adverse selection”. Section 2.2 analyses these issues of asymmetric information in services markets.

⁸The term used in the economic literature for these type of products is ‘credence goods’. Consumers are usually unable to observe the quality of the credence good (service) they received, or, if this quality is observable *ex post*, they cannot - or only with a time lag - judge whether it is the *ex ante* needed one (think of medical services, for example).

2.2 Asymmetric information

2.2.1 Moral hazard

Moral hazard can be defined as opportunism characterised by an informed person's taking advantage of a less-informed person through an unobserved action.⁹ Moral hazard problems are particularly present in services markets, as these markets are per definition characterised by asymmetric information. The customer-provider relation can be characterised by moral hazard in both ways. First, moral hazard occurs due to the customer's inability to observe actions taken by the service provider (i.e. the *provider's* actions are hidden).¹⁰ In other words, because customers cannot observe the quality of the service provided, service providers are tempted to under-perform/overcharge unless they face an incentive to provide proper services.

Standard principal-agent theory tells us moral hazard problems like this can at least partly be overcome by designing contracts that limit the informed party's opportunistic behaviour. As the customers cannot directly observe the variable they would want to observe (i.e., the exact quality of the service), any payment for the provision of the service is necessarily based on correlated, observable variables like the outcome or the time spent on the service provision. For example, instead of the quality of the service t , the principal (customer) observes a signal $x = \mu(t) + \varepsilon$.¹¹ Because he cannot tie the contract to t , the principal makes it conditional on x . In a simple linear example, the agent's (service provider's) wage would be of the form $w = \alpha x + \beta$. Assuming the agent will maximise his own expected utility given the contract and will actually prefer the contract to an outside option,¹² the principal can maximise his expected utility with respect to α , thereby optimising x , the signal of the true (but unobservable) quality of

⁹Perloff (2001).

¹⁰See Holmstrom (1985).

¹¹See Holmström and Milgrom (1991).

¹²In the principal-agent literature, these conditions are respectively called the 'incentive compatibility constraint' and the 'individual rationality constraint'.

the service t .

Sometimes, however, it can be preferable for principals *not* to specify certain quality signals in contracts with agents as such specifications may divert focus away from the real quality of the service to the observable feature only. Holmström and Milgrom (1991) argue that where multiple performance aspects are important to the principal, it is better to pay a fixed wage instead of conditioning the wage on the measurable aspect only. Assuming there is one unmeasurable and one measurable attribute to the provision of the service, $\mu(t_1, t_2)$ is reduced to $\mu(t_2)$ in $x = \mu(t) + \varepsilon$. A wage structure like $w = \alpha x + \beta$ would then result in an optimal effort x , *based only on* (t_2). Holmström and Milgrom (1991) use the example of a home renovation, where some desirable attributes of the contractor's performance such as courtesy, attention to detail, or helpful advice, are unmeasurable. The contractor would focus on the measurable attribute, timely completion, at the cost of the unmeasurable attributes if his wage was tied to measurable performance aspects only. A fixed wage would not provide him with this incentive and would more likely result in an even distribution of effort and thus a more desirable outcome for the principal. The same analysis applies to other services.¹³

Moral hazard problems can also present themselves in the other direction - especially in financial markets, it is the actions of the *customer* that are hidden. The less-informed service provider cannot control or even observe the behaviour of customers, which allows those customers to behave more opportunistically than they would under full information. The classic example is the insurance market. Without insurance, homeowners, for example, have ample incentives to protect their homes from burglary, fire, etc.

¹³Exceptions are for example legal services, that are often based on contingent contracts - contracts that are contingent upon variables that are both relevant to the delivery of the service and are observed so that contracts can be enforced. Lawyers often sell their legal services on an outcome contingent basis: if they win their case they collect a share of the award, if they lose they may charge nothing. See Holmström (1985), p188.

and thus to provide locks, fire alarms, etc. Insurance, however, changes the marginal benefit of such precautions as the insurance company will cover the costs of replacing/repairing the house in case anything happens, *whether the precautionary measures are in place or not*.¹⁴ Also, as the insurance company cannot observe the amount of precautions taken, it cannot base its premium on that and consequently cannot give financial incentives to customers that induce them to increase the amount of precautions they take (i.e., reduce the risk for the insurance company). In other words, the private benefit of 'correct' behaviour becomes smaller than the social benefit under asymmetric information.

Similarly, banks are subject to moral hazard when holding deposits and providing financial credit, as investment is usually made subsequent to financing. As banks have only limited control over the actual use and repayment of the funds, debtors have low incentives to act in society's best interest to the extent that this behaviour does not coincide with their private interest after the financing has taken place.^{15 16}

2.2.2 Adverse selection

Another asymmetric informational problem that might arise in services markets is 'adverse selection'. 'Adverse selection' is the term used in economic literature to describe the process in which asymmetric information lowers the quality of goods or services supplied in the market.¹⁷ Service suppliers

¹⁴This is a simplified example to illustrate the principle of moral hazard. In reality we often observe clauses in insurance contracts that prescribe a certain level of care to reduce the risk to the insurance company.

¹⁵Of course this argument assumes a static setting. In a more dynamic setting, firms borrowing from the financial market will take future borrowing possibilities into account and will value a good reputation as a debtor.

¹⁶This analysis applies not only to firms and individuals, but also to countries (think of countries borrowing from the World Bank and the International Monetary Fund).

¹⁷Resulting from Akerlof's (1970) classic analysis of the market for lemons (low-quality cars).

typically have more information about the quality of the service provided than most customers have.¹⁸ The uninformed buyer has an equal chance of purchasing a low or a high-quality service and will determine his willingness to pay based on the *expected* value of the service (the sum of his reservation prices times the probabilities of actually purchasing the associated levels of quality). If the high-quality service providers' reservation price lies below the buyers' expected value, equilibrium in the market is determined by the intersection of supply and demand at a price equal to the buyers' expected value.¹⁹ However, if the service providers' reservation price exceeds the buyers' expected value, the former will be unwilling to sell the service and as a result, the low-quality services will drive the high-quality services out of the market. For example, car maintenance services are difficult to evaluate for customers so they base their reservation price on the expected value of the service they will receive. If the high-quality providers' reservation price exceeds this expected value, they will find it no longer profitable to stay in business and leave the market. The only suppliers left are low-quality service providers and when consumers realise that at any price below the high-quality providers' reservation price, they will only be able to receive low-quality maintenance on their cars, equilibrium in the low-quality market segment (the only market segment left) is simply determined by the supply of and demand for low-quality repair services.²⁰

¹⁸At the very least, customers have less information *prior* to the consumption of the service, but in many cases (think of experts' services like medical or legal services, the so-called "credence goods") customers often cannot determine the extent of the service that was needed and how much was actually performed. See Wolinsky (1993).

¹⁹Although this outcome is efficient, the allocation of funds differs because low-quality providers receive more than they would under full information, at the expense of high-quality providers who would benefit from buyers' higher reservation price for high-quality goods if they could only distinguish between services of different quality (i.e., in a full information situation).

²⁰In practice, this means that certain people will choose to (attempt to) repair their cars themselves or to keep driving without any maintenance at all (neither option seems socially desirable!).

Carlton and Perloff (2000) rightfully remark that not all markets with asymmetric information degenerate so that only the lowest-quality item is sold. However, markets characterised by asymmetric information are likely to be inefficient compared to markets with full information, in that quality levels may be lower.²¹ The cause of this inefficiency is in essence an externality problem. By raising the quality of his service, an individual service provider raises the average quality in the market and thereby consumers' expected value of *all* products, as they cannot distinguish between the higher-quality service and the other services offered in the market. Though the individual service provider bears all the cost of the quality increase, he has to share the benefits with all the other providers in the market because the quality increase allows everyone to charge a higher price (which equals the increased expected value of the consumers). Clearly, there is no incentive for individual service providers to produce and sell higher-quality services in markets with asymmetric information.

There are a number of ways to overcome this externality, however.²² Guarantees or warranties allow sellers to credibly convey the information to consumers that their products are in fact of high quality. Due to the perceptive element in services consumption, however, these measures may be less suited for services quality signaling. Service providers are more likely to rely on reputation. In fact, reputation and goodwill are crucial assets in most services markets, as they provide a strong signal with respect to the quality of the services provided. Another important quality signaling tool are standards and certifications - in services markets, particularly in markets for expert services, this usually comes in the form of proof of a particular education. Alternatively, membership of an industry organisation can be used as a certificate signaling a certain quality threshold. Though

²¹It should be noted that in the case of services markets, comparisons with situations of full information may not be appropriate because asymmetric information is inherent in service provision.

²²See, for example, Carlton and Perloff (2000) and Holmström (1985).

these measures may not always fully eliminate all the information asymmetries, it should be noted that these asymmetries are inherent to the nature of services, irrespective of who provides them or to what extent they are regulated.

2.3 Entry factors

Entry factors are a structural market feature of oligopoly, but not of monopolistic competition. The relevant factors in services industries stem from sunk costs, network effects, and collusion. However, the extent to which these factors translate into barriers to entry in services markets is not always entirely clear, not in the least because opposing views exist on what exactly constitutes a barrier to entry. In the tradition of Bain (1968), barriers to entry are defined as the extent to which, in the long run, established firms can elevate their selling prices above the minimal average costs of production and distribution without inducing potential entrants to enter the industry. He distinguishes three main sources of barriers to entry/advantages of established over potential entrant firms: product differentiation, absolute cost advantages, and economies of scale. Stigler (1968b), on the other hand, argues that a barrier to entry is merely “[...] a cost of producing (at some or every rate of output) which must be borne by a firm which seeks to enter an industry but is not borne by firms already in the industry”. Hence, according to Stigler’s definition, setup costs, advertising outlays and economies of scale are *not* barriers to entry, so long as the incumbent firms face(d) the same costs as the entrants do.²³

Baumol and Willig (1981) use essentially the same definition as Stigler (1968b) and conclude that fixed costs do not constitute barriers to entry, but sunk costs can. Specifically, they state that “[...] their role as barriers to entry depends on the risk to which they subject the entrant”, implying

²³Stigler (1968b) recognises that product differentiation only acts as a barrier to entry “[...] if the costs of differentiation (design, advertising, etc.) are higher for a new firm than an existing firm”.

that the prospective of incurring large sunk costs (and thus large potential losses in case of market exit) may make firms reluctant to enter a market in which they are uncertain of success.²⁴ Sunk costs may also act as a barrier to entry if the extent to which costs are sunk determines the cost of capital - potential entrants may face higher rental cost as capital markets work imperfectly and providers demand a risk premium on top of the 'normal' cost of capital.²⁵ Sutton (1991) does not appear to dispute the argument that sunk setup costs may act as an exogenous barrier to entry, but does oppose the view that sunk advertising outlays merely shift this barrier, as long as such outlays can induce some fixed fraction of consumers to choose some particular firm's offerings at a price level that exceeds that firm's unit variable cost. This theory is known as the natural oligopoly theory.²⁶

Sunk setup and advertising costs are present in most, if not all, services markets. Following Stigler (1968) and Baumol and Willig (1981), we conclude that these costs constitute a factor but not generally a barrier to entry.

Another entry factor in some service industries is the presence of positive network externalities. This basically means that the value of the network increases with the number of users. The larger (and thus more valuable) the network, the larger the so-called 'switching costs' for users, i.e., the more difficult/expensive it is for them to switch to a competing supplier. The entry factors created by these switching cost present themselves in services markets like telecommunications, software and transport services.

Finally, entry can be determined by collusion among service providers. Though some doubt the possibility of collusion in markets characterised by intense non-price competition (see section 4.3), a number of services markets, particularly markets for professional services, are to a greater or lesser

²⁴Perloff (2001), p285.

²⁵Unlike the incumbents, entrants do not have records of the success rate of past sales efforts or a captive clientele that may reduce the risk premium.

²⁶See section 3.1.

extent regulated by professional associations. The main rationale for the existence of these associations is to ensure the quality of the services provided to overcome the informational asymmetries that exist between professionals and clients. They often do this by maintaining qualitative entry restrictions.²⁷ It should be noted, however, that the entry-restricting powers of these organisations seem to be weakening due to increasing government intervention in services markets to stimulate competition.²⁸ Moreover, Fershtman and Pakes (1999) argue that, in comparing a heterogeneous market equilibrium with and without collusive possibility, the collusive industry is not only less concentrated and offers both more and higher quality products to consumers, but also that that positive effect of collusion on the variety and quality of products marketed more than compensates consumers for the negative effect of collusive prices, so that consumer surplus is larger in the collusive environment. In other words, to the extent that collusion is feasible in services markets, it does not necessarily imply a decrease in welfare.

3 Market Conduct

3.1 Non-price competition

Non-price competition is an essential feature of services markets, whether they can be characterised as monopolistic competition or as oligopolistic, as reputation is often the crucial asset of a service provider. Non-price efforts are believed to generate demand, especially where the 'product' cannot be evaluated before consumption (as is the case in services industries). Though advertising is a main form of non-price competition, the term should be interpreted as encompassing all forms of non-price competition.

In modern economic literature, two distinctions are made regarding product differentiation. Where consumers show differences in subjective tastes,

²⁷See also section 5.4.

²⁸Noether (1986): "Both popular and professional writings suggest that the American Medical Association (AMA) no longer wields the authority that it held even 10 years ago".

we find producers engaging in horizontal product differentiation. Hotelling (1929) developed the first horizontal differentiation model, which is still the most-referenced model today. According to the Hotelling model, in the simple case of uniform pricing, all consumers are distributed uniformly along a straight line and sellers are located at different points along this line. In this model, it is assumed that customers have no preference for either seller except on the ground of price plus the cost of transporting the goods from the seller's location to their own location. Salanié (2000) rightfully remarks that the product could generally present itself in numerous varieties other than location, the obvious example being quality. This makes Hotelling's (1929) model very suited to analyse markets in which firms are selling products or services of different qualities to a population of consumers differing in their taste for quality.²⁹ Where consumers recognise that different services are actually different qualities of the same service, however, quality becomes an attribute of vertical product differentiation. Shaked and Sutton (1987) define vertical differentiation as follows: "given any two distinct products, if they were sold at the same price, then all consumers would choose the same one (the 'higher quality' product)". In other words, consumers are willing to pay different prices for (what they perceive to be) different quality levels.

In practice, services can be differentiated both horizontally and vertically, as consumers differ in income, tastes, or both.³⁰ The endogenous sunk cost (ESC) model developed by Sutton (1991) analyses market situations that are characterised by horizontal and vertical differentiation.³¹ It basically concludes that the classical negative relationship between market size and concentration is not necessarily valid for industries in which advertising or other sunk outlays play a significant role. This conclusion is based on the notion that when sunk outlays enhance consumers' willingness-to-pay,

²⁹In the original analysis of Gabszewicz and Thisse (1979) this difference rests on differences in income.

³⁰Shaked and Sutton (1987).

³¹See Appendix A for a more detailed presentation of the model.

thereby increasing the size of the market, competition will lead to escalating investment in sunk costs by the dominant firms in the market, thereby preventing the industry from converging to a fragmented structure. These key findings hold in an extended version of the model that allows for differences in consumer tastes. Dick (2002) and Ellickson (2003) respectively show that the ESC model applies to banking services and retail industries (supermarkets) and that measures of quality other than advertising lead to similar outcomes - as long as they are increased primarily through sunk investments. Further empirical research should be able to tell whether the ESC model applies to services markets in general, but given that it allows for both horizontal and vertical differentiation - the main features of competition in services - this is likely to be the case. The implications of these findings will be analysed in sections 4 and 5.

3.2 Quality discrimination

As long as service providers know more about the quality of their services than their customers do, they have at least *some* market power. In most services industries this is reinforced by consumer loyalty and search costs. We have seen that asymmetric information may create difficulties for consumers in distinguishing between services of different qualities. Similarly, it may cause different consumers to have different beliefs about services of equal quality. Where this is the case, service providers, having market power, can successfully engage in second degree price discrimination. When the same price schedule is offered to all consumers, they sort themselves through a process of self-selection - consumers who (perhaps falsely) believe some services are of higher quality than other, similar services, pay more. Perloff (2001) notes that price discrimination is only possible if some, not all, consumers face asymmetric information, i.e. where some consumers are informed and some are not. If all consumers would be informed or all would be uninformed about the quality of various services, service providers would

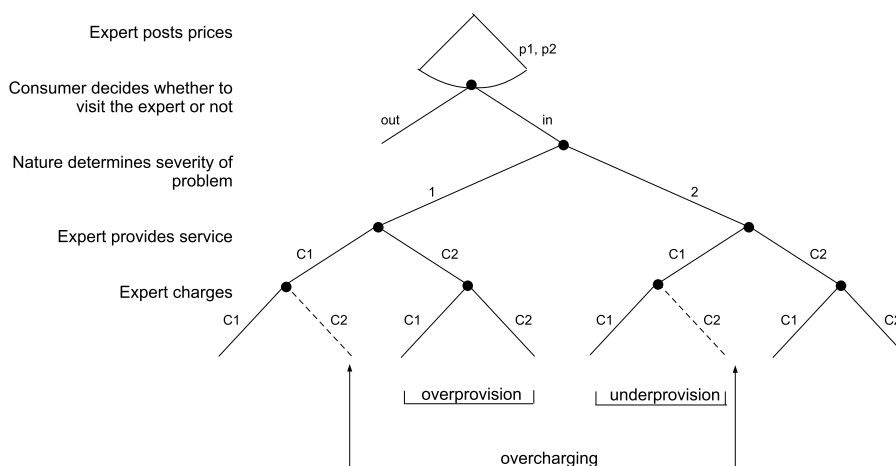
be best off charging a single price.

Second-degree price discrimination appears to be particularly suitable in credence goods/services markets. In the case of expert services provided by, for example, medical doctors, lawyers or mechanics, consumers are usually unable to observe the quality of the service they received, or, if this quality is observable *ex post*, they cannot judge whether it is the *ex ante* needed one. Credence goods literature has mostly focused on two major inefficiency sources: inefficient treatment and overcharging.³² Figure 1 is a simple game-tree taken from Dulleck and Kerschbamer (2003) which illustrates these concepts. First, the expert posts his prices (p_1, p_2) . The consumer observes these prices and decides whether to visit the expert or not (*in, out*). Next, nature determines the severity of the consumer's problem/service demand ($k = 1, 2$). The expert provides either high- or low-quality service ($c_k = c_1, c_2$), and finally charges for either high- or low-quality service (e_1, e_2).

Dulleck and Kerschbamer (2003) argue that if the assumption of homogeneous consumers is relaxed (to allow for the possibility of second-degree price discrimination), three classes of tariffs or price schedules can be distinguished in the basic model represented in Figure 1. One type of tariff contains a higher mark-up for the expensive treatment ($p_2 - c_2 > p_1 - c_1$), other tariffs have a higher mark-up for the cheap treatment ($p_2 - c_2 < p_1 - c_1$), and a third group of tariffs has equal mark-ups ($p_2 - c_2 = p_1 - c_1$). The authors find that only under tariffs where the differences in the intervention prices reflect the differences in treatment costs (equal mark-up tariffs) will the expert perform a serious diagnosis and recommend the appropriate treatment. Under tariffs where the intervention prices depart from the equal mark-up rule the expert will recommend and provide the most profitable treatment without

³²There are two types of inefficient treatment. Undertreatment means a simple procedure is provided while a high quality intervention is required and overtreatment is the provision of high-quality service where a simple one would have sufficed. Overcharging means a high-quality price is charged while a low-quality service is provided. See Dulleck and Kerschbamer (2003).

Figure 1: Credence goods



a serious diagnosis. In other words, high quality diagnosis and appropriate treatment is sold to the most profitable market segment only.

Masson and Wu (1974) argue there is ample evidence of price discrimination in the market for physicians' services, even after accounting for charity pricing for low-income groups and the possible collusive power of the type of industry organisations that are common in most medical professions.³³ They take a somewhat different approach than Dulleck and Kerschbamer (2003) and argue that price discrimination for physicians' services is enabled by different information/search costs and concerns about quality for rich and poor consumers. They assume the opportunity cost of search is higher for the rich than for the poor, and that the rich tend to search less because their medical expenditures constitute a smaller proportion of their income (i.e., the elasticity of demand for physicians' services is different among income groups: the higher the income, the lower the elasticity at every price level).³⁴ In addition, it is thought that if quality of treatment (service) is a

³³For example, the American Medical Association (AMA).

³⁴See Masson and Wu (1974), p67.

normal good, the rich may concentrate on finding a higher quality doctor.

Though in principle quality discrimination is possible in any service market (as long as the service provider has some degree of market power), empirical evidence shows it is mostly found in markets for credence goods/services. Many professional services, and medical services in particular, seem to be subject to second-degree price discrimination.

3.3 Strategic firm behaviour

In oligopolistic services markets in particular, non-price tools, such as advertising and product quality, provide the major source of interfirm rivalry.³⁵ In many studies of oligopolistic competition, the standard Cournot assumption is adopted: firms suppose that their decisions will have no impact on their rivals' future actions.³⁶ In other words, it is often assumed that the *conjectural variation* is 0. The conjectural variation measures the elasticity of rivals' advertising (or output, or price) with respect to firm i 's advertising, i.e. it is the percentage change in all other firms' advertising outlays (z_{-i}) that firm i expects in response to a 1 percent change in its own advertising outlay (z_i):³⁷

$$\alpha_i = \frac{\Delta z_{-i}/z_{-i}}{\Delta z_i/z_i} = \frac{z_i}{z_{-i}} \frac{\Delta z_{-i}}{\Delta z_i}. \quad (1)$$

Scherer (1980) feels that an advertising conjectural variation of zero (the standard Cournot assumption) may be plausible for two reasons. First, it takes some time for rivals to respond to the firm's changes in advertising messages, so the firm may behave as if the rivals will not respond. Second, rivals actually may be unresponsive to changes in the firm's advertising, especially in the face of increased advertising on the part of the firm, because of the uncertainties of the effect of the firm's campaign. Unfortunately

³⁵Roberts and Samuelson (1988).

³⁶In the classic Cournot model, it is assumed that each firm believes that its rivals hold output constant. To many authors, this assumption appears implausible (see, for example, Martin (1994), p123).

³⁷Martin (1994), p127.

little is known about possible conjectural variations in services markets. We can, however, infer some cautious results from the extensively studied US cigarette market, which, like many services markets, is an advertising-intensive market.

Roberts and Samuelson (1988) find the conjectural variations to be negative in the US cigarette market. In other words, in response to an *increase* in advertising by one supplier, all other suppliers would *decrease* their advertising. This result is surprising, and is unlikely to apply to services markets, where reputation is crucial for firms' survival and where any increase in non-price efforts is likely to be met by rivals. In other words, if anything, we would expect a *positive* conjectural variation in advertising-intensive markets. This expectation is met by Seldon, Banerjee and Boyd (1993), who call Roberts and Samuelson's (1988) finding "[. . .] startling on two counts". Not only is a negative conjectural variation in an industry where product differentiation is important intuitively surprising, it also contradicts empirical evidence. They critique the estimation method used by Roberts and Samuelson (1988) and find positive advertising conjectural variations instead.

In the terminology of Tirole (1988) this means that advertising messages (and other non-price efforts) are *strategic complements*. That is, it is in a firm's best interest to, following an increase in advertising outlays by its rival(s), increase its advertising outlays accordingly. Intuitively, this is what we would expect to be the case in most services markets. Our intuition is confirmed by Atle Berg and Kim (1994), who find positive conjectural variations in the Norwegian banking sector. However, at this moment, sufficient empirical research to be fully conclusive on this matter is lacking.

4 Market Performance

Ultimately, the decisions in antitrust cases are taken based on market performance. The question is not whether the market is concentrated, but whether this concentration predisposes multi or unilateral market power.

For merger evaluation it is the potentiality that matters. We assume the ultimate focus of competition authorities is the efficiency of markets and the extent to which the outcome of the competitive process is socially optimal. However, market structure and performance are the outcome of the workings of competition and market conduct and can therefore not be analysed separately. The following section thus serves two purposes: to conclude our discussion on the nature of services competition and to lay the foundations for our analysis on the implications for competition policy in the subsequent section.

4.1 Product diversity

Many studies on product differentiation have concerned themselves with the question of optimal product diversity. As Dixit and Stiglitz (1977) remark, “modeling the desirability of variety has been thought to be difficult and several [...] approaches have been adopted”. Gabszwick and Thisse (1986), for example, base their ideas on the Hotelling (1929) spatial approach and compare performance under horizontal and vertical product differentiation. They find that in the case of horizontal differentiation, there exists no perfect equilibrium, as no price equilibrium can exist if any two sellers are located sufficiently close to each other.³⁸ The result is that horizontal product differentiation entails unstable price and product competition. The model used by Gabszwick and Thisse (1986) suggests more stability is to be expected under vertical product differentiation, though the authors note that it is unknown whether this conclusion stands up in a more general setting. Modifications of the model generate the same result, however. Salop (1979) analyses a variant of the traditional Hotelling (1929) model in which the economy consists of two industries, rather than one.³⁹ In this general equilibrium setting, it is

³⁸Here, too, ‘location’ should be interpreted to encompass a wide range of quality features.

³⁹One industry is monopolistically competitive with differentiated brands and decreasing average costs; the other is a competitive industry producing a homogeneous commodity.

found that the optimum variety depends on the difference between the average surplus and the surplus of the marginal consumer, relative to fixed costs; “the value of adding an extra brand (and respacing the others) effectively converts marginal consumers to average ones, at fixed cost”.⁴⁰

Other studies of optimal product differentiation are based on Lancaster’s (1975) “characteristics” approach. In this “indirect analysis of consumer behaviour”, the consumer is assumed to derive his actual utility or satisfaction from characteristics which cannot in general be purchased directly, but are incorporated in goods.⁴¹ Lancaster concludes that a socially optimal degree of product differentiation does exist, but that its conditions are difficult to recognise with respect to the actual number of varieties. More importantly, however, is the finding that in the case of constant returns to scale, market imperfections do not tend to give a non-optimal degree of product differentiation, whereas they do under increasing returns to scale. Specifically, under constant returns to scale, the optimal number of product varieties is not finite unless consumer preferences are - where preferences form a continuum, optimal product differentiation is essentially unbounded. Under increasing returns to scale, however, the optimal degree of product differentiation *is* bounded and can, in principle, be determined. This is the reason why most studies of optimal product diversity are confined to the case of increasing returns to scale.⁴²

In services industries, however, economies of scale are limited (see Table I) as input and output are often the same. This implies that, given

See Salop (1979), p142.

⁴⁰Salop (1979), p152.

⁴¹Lancaster (1975), p567.

⁴²See, for example, Dixit and Stiglitz (1977), who use a direct approach, noting that the convexity of indifference surfaces of conventional utility functions already embodies the desirability of variety. Assuming economies of scale, they compare the market equilibrium with the constrained (by the assumption that each firm must have non-negative profits) and the unconstrained optimum under constant and variable intra- and intersector elasticities of substitution.

that preferences are continuous, product differentiation in monopolistically competitive or even oligopolistic services markets is, contrary to economic intuition,⁴³ *not* necessarily non-optimal.

4.2 Advertising

The advertising literature is characterised by debates that include advertising effects on price-elasticity, competition, and transparency. The optimal amount of advertising in static as well as dynamic settings is also unresolved. It is beyond the scope of this paper to review all these issues here. Instead, we will touch upon a few issues that are particularly relevant for the analysis of performance in services markets.

Advertising is generally thought to be excessive. Spence (1977), for example, compares the market equilibrium with the social optimum under monopolistic competition, and finds that, in case of constant elasticity of demand, markets tend to oversupply non-price services such as advertising (see Appendix B). Netter (1982) finds empirical evidence that in some markets, competitive advertising may be mutually destructive and thus non-optimal from a welfare point of view.

In a seminal paper, Dorfman and Steiner (1954) determine the optimal amount of advertising for differentiated markets. Equilibrium is found when the slope of the demand curve (i.e. the price-sensitivity of demand) is equal to the rate at which sales increase in response to increases in average cost incurred in order to increase quality. In this model, quality tends to be higher “the greater the sensitivity of consumers to quality variation [...], the lower the sensitivity of consumers to price variation [...], and the lower the effect on average costs of quality changes”. Criticism of these results is perhaps best presented in Becker and Murphy (1993), which disputes the assumption adopted in the Dorfman-Steiner theorem that the effect of

⁴³As Dixit and Stiglitz (1977) put it: “The fact that in a Chamberlinian equilibrium each firm operates to the left of the point of minimum average cost has been conventionally described by saying that there is excess capacity.”

advertising on the quantity demanded is independent of (instead of positively related to) the price-elasticity of demand. In other words, what counts is the effect of advertising on the utility from *marginal* units of product x . Becker and Murphy (1993) determine whether advertising is socially optimal by means of the ‘surplus criterion’:

$$\frac{dS}{dA} \gtrless 0 \quad \text{as } V_a - x \frac{dp_x}{dA} - \frac{dT}{dA} \gtrless 0, \quad (2)$$

where S is total surplus, x the product advertised, p_x the price of x and T any revenue from the sale of advertising A . The term V_a is the marginal utility to consumers from advertisements for x (net of any reduction induced by these advertisements in the utility from other goods).⁴⁴

An extensive body of literature also exists on the dynamic effects of advertising and how this effects the social optimum. Most of this work is based on a seminal paper by Arrow and Nerlove (1962), which argues that advertising expenditure should be treated in the same way as investment in a durable good. The authors assume current demand is determined by a stock of goodwill, which summarises past advertising and, like capital stock, depreciates over time. They extend the classic Dorfman-Steiner model on optimal advertising expenditures by considering the effects of current advertising on future demand.⁴⁵

Another notion of advertising is that of strategic firm behaviour. This notion is mainly based on the idea that advertising can function as a sig-

⁴⁴Many economic discussions assume that advertising is given away to consumers ($dT/dA = 0$), that advertisements do not directly provide utility ($V_a = 0$), and that advertising raises the price of advertised goods (dp_x/dA has a positive sign) which is why most of them conclude that advertising is generally excessive. However, Becker and Murphy (1993) argue that “[...]advertising does affect the utility of consumers and is often sold - sometimes at a negative price” and determine whether advertising is socially optimal by means of the ‘surplus criterion’ given by (2).

⁴⁵In an interesting extension of the Arrow-Nerlove model to oligopolistic markets, Fershtman (1984) argues that the optimal stock of goodwill can be described as a function of the current market share. See Appendix C for a short explanation of Fershtman’s argument.

nal of strategic commitment, i.e., a promise or warning to rivals that the advertising firm will stay in the market because it will want to recoup the sunk investments it made by advertising. In addition, such investments may deter entry by raising potential entrants' costs.⁴⁶

Though solid empirical evidence of the amount and effect of advertising in services markets is lacking, it is possible to cautiously draw a few conclusions from the above theories and available empirical evidence on advertising-intensive industries. Arguments of excessive advertising usually refer to persuasive advertising. Where advertising has the effect of reducing informational asymmetries it may be more efficient and thus closer to the social optimum. Claims of excessive advertising also argue that competitive advertising may do little more than cancel out. However, empirical results indicate that this is mostly the case among firms producing consumer goods, particularly non-durables.⁴⁷ Studies of the US cigarette industry indicate that in advertising-intensive markets, advertising does not only rearrange market shares but may also increase total market demand.⁴⁸ Whether advertising in a particular service market is excessive or not thus depends on the degree of competition in the market, the type of advertising, the extent of asymmetric information, and the effects on demand for variety and quality.

4.3 Collusion

Quite a number of studies have analysed collusion in markets where non-price competition plays an important role. Most of that literature starts from the basis of some degree of market power exists, stemming from the product differentiation and informational asymmetries characteristic of non-price competition. Two seminal papers by Dixit (1979) and Lancaster

⁴⁶In the terminology of Fudenberg and Tirole (1984), this type of entry-detering behaviour would be classified as a 'top dog' strategy.

⁴⁷Netter (1982).

⁴⁸See Roberts and Samuelson (1988).

(1998), in particular, compare various equilibria obtained by collusion on price, quality and/or entry, and price alone respectively.

Dixit (1979) focuses on the interaction between collusion over some or all of the strategic variables entry, quantity and quality and competition in the others and compares the various equilibria with each other as well as with the social optimum. Dixit finds that when the objective is to maximise average profit, the collusive control of entry essentially amounts to full collusion (i.e. the two cases yield the same welfare results). When the objective is to maximise total profit, Dixit finds that in the social optimum, as well as in the equilibrium in case of full collusion, *and* the equilibrium in case of collusion over entry and quantity, the choice of quantity (given quality) occurs at the minimum of the cost per unit of contribution to the industry sub-utility. Making the additional assumptions of linear costs and increasing quality increases the price elasticity of demand, Dixit further finds that the equilibria in the asymmetric cases of collusion over entry and collusion over both entry and quality are second-best in terms of welfare, and that enforcing competition in only one dimension may actually reduce welfare.

Lancaster (1998) examines the effect of collusion on non-price efforts in the market by analysing the result of an increased price (due to regulation or cartel agreement) on non-price competition. Reasonable assumptions on the behaviour of other economic parameters⁴⁹ lead Lancaster to conclude that the effect of a price constraint above the equilibrium level on the degree of non-price competition depends on the presence of economies of scale in the non-price variable.⁵⁰ Economies of scale are limited in most services industries, which implies that the term q_{zz} is close to zero in these markets.

⁴⁹Lancaster (1998) assumes demand is downward-sloping, sales efforts or advertising activities increase sales. In addition, he assumes non-decreasing marginal cost and some monopoly markup and that price increases hardly affect the influence of non-price competition on sales (and certainly do not increase it).

⁵⁰In the presence of economies of scale, an imposed price increase away from the equilibrium would lead a firm to cut back on its non-price efforts, and a price ceiling set under the equilibrium price would increase the firm's sales efforts.

Consequently, a price increase resulting from collusion would, if anything, induce greater non-price competitive effort.

This finding seems consistent with some authors' doubts about the possibility of successful cartelisation in markets characterised by non-price competition.⁵¹ After all, the incentive to cheat on any price agreement by supplying non-price services is considerable since non-price activities are particularly difficult to detect. Where non-price competition shapes a considerable part of the market (as is the case in services markets), this incentive is even larger. Ginsburg (1993) states that: “[...] cartelization cannot be profitable - and is therefore a good deal less likely to occur - in a market where non-price competition can play a significant role”. Ginsburg refers to the intense non-price competition and insignificant excess profits in regulated transport industries and reasons agreement on non-price variables and detection of cheating would be too difficult and too costly for firms, even in concentrated markets. Other empirical evidence also implies that collusion has not been an important phenomenon in most services markets.⁵²

Ivaldi et al (2003) add that a situation of vertical differentiation is similar to a situation of asymmetric costs of production. It is argued that the firm that can offer a better quality at the same costs (or, equivalently, the same quality at lower costs) would have more to gain from cheating and would have less to fear from a possible retaliation from the other firms. They conclude that “when firms are differentiated by levels of quality, collusion is more difficult, the larger the competitive advantage of the high-quality firm”.

Successful collusion on *non*-price activities is also argued to be unlikely, again because monitoring output and detecting potential cheating is very difficult.⁵³ Finally, it is sometimes argued that, as with technological com-

⁵¹Exceptions are markets that are affected by regulation that inhibits entry, similar to the effects some professional bodies may have.

⁵²See, for example, Gaynor and Haas-Wilson (1999) for an analysis on competition in health care markets.

⁵³In addition, Ginsburg (1993) points at the difficulties of agreement on non-price vari-

petition, firms in services markets *must* engage in intense non-price rivalry to secure the survival of the enterprise as reputation and goodwill are a service provider's dearest assets in that they signal the quality of the provided service (see section 2.2.2).

It can thus be concluded that for service providers, in the absence of regulatory barriers to entry, actual collusion on either price or quantity is unlikely to be successful in most services markets due to the high incentive to cheat (in turn due to the low probability of detection) and the need for continuous non-price competition. Where collusion *is* possible, for example due to the presence of industry associations or other regulatory entry barriers, Dixit's analysis may well apply (i.e., we might observe collusion on several variables at the same time and enforcing competition in only one area may actually reduce welfare).

5 Evaluation of Competition Policy Instruments

Performance evaluation in services markets may be inaccurate because of the difficulties that arise when defining markets, measuring output, and determining the intensity of competition. The nature of services and the prominent role of non-price competition in services markets require more than the traditional performance measures and market definition instruments and this should be taken into account by competition authorities when evaluating firm behaviour and/or performance in these markets.

ables from an intra-firm perspective. These difficulties appear to increase with the size of the firm. Joyce (1989) finds evidence that the decision-maker within the firm is more likely to collude if he or she has a substantial ownership stake in the firm, which is of course more likely in a small than in a large enterprise. In addition, as more individuals within the firm are affected by the decision to collude, agreement on non-price variables becomes increasingly difficult, profitability per person decreases and the probability of detection increases.

5.1 Traditional performance measures

Sosnick (1958) lists a number of traditional market performance norms suggested by various writers: “(1) Operations should be efficient. (2) Promotion expenses should not be excessive. (3) Profits should be at levels which reward investment and efficiency and induce innovation. (4) Output should be consistent with a good allocation of resources. (5) Prices should not intensify the cyclical problem. (6) Quality should conform to consumers’ interests. [...] (10) Entry should be as free as the nature of the industry permits.” Most of these criteria can be reduced to the behaviour of the parameters price, quantity, and profit. Though these parameters may provide some indication of performance in many markets today, it can be argued that they do not fully reflect competition in services markets and that additional parameters should be analysed to obtain a complete picture.

The reasons for this position are multifold. Because of the nature of services, output in services markets is generally difficult to measure. Apart from the fact that a rather low number of countries actually reports on services trade, available data are aggregated at best, full of discontinuities at worst, and systematic recompilation by international organisations has simply not yet developed. In addition, most of the available data show values only and do not distinguish between movements attributable to price change and movements attributable to changes in quantities or qualities.⁵⁴ The resulting lack of useable data poses a problem for analysts, policymakers, and businesses alike. Although some progress has been achieved by international organisations and national statistical offices in recent years, the area still is not highly developed. It is unlikely that this fundamental problem will be resolved in the future due to the very nature of services and the difficulties associated with measuring output.

Spence (1977) argues that profits may be an equally poor indicator of market performance in markets characterised by non-price competition. Ba-

⁵⁴see Whichard (2001).

sically, he states that performance may be inadequate even in the absence of excessive profits, not because prices do not accurately reflect costs, but because a failure in the area of non-price competition may cause costs to be too high or too low.⁵⁵ It should be noted that even in a competitive services market, price may exceed marginal cost, as in the presence of asymmetric information, price is often used to signal a certain level of quality (i.e. a high price ‘ensures’ high quality).

Though ‘quality’ is among the performance norms listed by Sosnick (1958), it has not received nearly as much attention in antitrust decisions as the traditional performance measures price, quantity and profit. However, the prominent role of non-price competition in services implies that these indicators become less reliable, as quality is an important determinant of performance. As modern economic literature reveals, the inclusion of quality in economic analysis is likely to result in outcomes quite different from outcomes based on traditional performance evaluation alone. Quality and diversity may increase utility and affect both the position and shape of the demand curve by increasing total demand and possibly changing demand elasticities. This would have serious implications for welfare and a great impact on the outcomes of cost-benefit analyses.

5.2 Concentration ratios

As performance varies with structure, an important set of performance indicators are actually measures of market structure. Concentration ratios and other measures of market power are widely used in economic research as well as in antitrust policy. One of the most common concentration ratios is the Herfindahl-Hirschman Index (HHI), which is the sum of squares of the market shares of all firms in the industry:

$$(s_1)^2 + (s_2)^2 + (s_3)^2 + \dots + (s_n)^2, \quad (3)$$

⁵⁵See also section 3.2.

where s_1 is the market share of firm 1, s_2 is the market share of firm 2, etc. Another often quoted concentration ratio is the C4, which is the share of industry sales accounted for by the largest four firms in the market.⁵⁶ Related to these measures is the price-cost margin given by the Lerner Index:

$$\frac{P - MC_i}{P} = \frac{s_i}{\varepsilon_{QP}}, \quad (4)$$

which is often used as a measure of market power.⁵⁷ The difference between price P and firm i 's marginal cost MC_i as a fraction of the price measures that firm's mark-up, which is equal to that firm's market share s_i divided by the elasticity of demand. In other words, the greater the mark-up a firm is able to obtain without losing too many customers, and thus the greater that firm's market share, the greater its market power. The aggregated Lerner Index shows the industry average of firms' price-cost margins using share weights

$$\frac{P - \bar{c}}{P} = \frac{\sum_i (s_i)^2}{\varepsilon_{QP}}, \quad (5)$$

where \bar{c} is industry-wide marginal cost and $\sum_i (s_i)^2$ is equal to the HHI.

Carlton and Perloff (2000) note that, in addition to measurement difficulties, concentration measures have two serious problems. One is that most commonly used ratios are not exogenous measures of market structure. Factors like profitability have an influence on market structure, and most concentration measures depend on the profitability of the industry. The result is that they do not describe market structure in a way that suggests properties of the market that definitely indicate performance. A second problem is that many concentration measures are biased because of improper market definitions. This issue will be discussed extensively in section 4.4.

⁵⁶The share of sales accounted for by the largest eight firms, the C8, is also used sometimes.

⁵⁷It should be noted that this measure is more applicable in some industries than others. For example, in industries with high fixed costs, one may expect price to exceed marginal cost - this implies recovery of investment rather than market power.

In addition to the above problems, concentration measures are unlikely to accurately reflect performance in services markets. Some argue that actual performance may be much closer to the competitive mark than examination of concentration ratios alone would suggest (Martin, 1994).⁵⁸ As pointed out above, cartels are inherently unstable because firms have an incentive to cheat on the agreement by engaging in non-price competition. Especially where non-price competition plays an important role in shaping a market, as it does in services industries, agreements on price or entry are likely to be undermined by the incentive of firms to engage in non-price activities. Thus, although high concentration ratios may alarm antitrust authorities, they do not accurately reflect the amount of competition in the market, as intense non-price competition may act as a restraint on market power, i.e., on the ability to raise price or preclude entry. In fact, the endogenous sunk cost model of Sutton (1991) tells us that where competition is based on investment in sunk outlays like advertising, we may expect to see a concentrated market structure as the outcome of a vigorous competitive process. Empirical evidence from retailing and banking sectors (see Ellickson (2003) and Dick (2002), respectively) indicates that this is particularly true for services markets.

In evaluating market performance in services industries, particularly with respect to potential collusion, competition authorities should thus be cautious to base their decisions on concentration ratios. Insofar as competition authorities can identify markets in which there are opportunities for firms to engage in non-price competition without the chance of easy detection by rival firms, “[. . .] enforcement agencies should be sceptical about the prospects for cartelisation” (Ginsburg, 1993). In addition, in the same way

⁵⁸Martin (1994) argues that in industries characterised by rapid technological development, firms must engage in intense rivalry, resulting in performance being much closer to the competitive mark than suggested by concentration ratios. His argument can be equally well applied to markets where reputation, rather than technological development, is crucial to firms’ survival. See Martin (1994) p132.

merger control takes into account the facilitating effects of potential competition in the post-merger market (either through new entry or expansion of existing firms), when evaluating proposed mergers, antitrust authorities should take into account the beneficial effects of non-price competition on market performance.⁵⁹ It is desirable that antitrust authorities recognise the prominent role of non-price competition in services markets requires a careful approach to the application of concentration ratios.

5.3 Market definition

One of the most important conditions for proper enforcement of competition policy is appropriate relevant market definition, as it is “[...] the critical underpinning for the evaluations of substantial lessening of competition, the calculation of concentration ratios and the evaluation of import competition and barriers to entry”.⁶⁰ Competition authorities throughout the world normally define a market in both its product and geographic dimension. The European Commission Notice on the Definition of the Relevant Market (1997) provides definitions for these two dimensions, that are very similar to the notions of relevant markets existing in other countries:

“A **relevant product market** comprises all those products and/or services which are regarded as interchangeable or substitutable by the consumer, by reason of the products’ characteristics, their prices and their intended use.”

“The **relevant geographic market** comprises the area in which the undertakings concerned are involved in the supply and demand of products or services, in which the conditions of competition are sufficiently

⁵⁹However, others argue the reverse might be the case if non-price competition is not included in market analysis. In a paper on hospital mergers, Connors (2003) argues that consolidation may affect non-price factors in the same way that price competition is often affected, i.e., larger players may lose their incentive to attract and maintain customers (patients) and quality may decrease. It should be added that this depends on the ability of consumers to detect quality differentials.

⁶⁰Australian Merger Guidelines (1999).

homogeneous and which can be distinguished from neighbouring areas because the conditions of competition are appreciably different in those areas”.

In addition, the Australian Merger Guidelines (1999) explicitly distinguish a functional and a time dimension. The recently amended New Zealand Merger and Acquisitions Guidelines (2004) also provide a customer dimension: where a significant group of buyers within a relevant market is likely to be subject to price discrimination, the Commerce Commission considers whether it would be appropriate to define additional markets based on particular uses for a good or service, particular groups of buyers, or buyers in particular geographic areas that are captive to those products, and in the face of a price increase unable to switch.

Defining the relevant market may be quite complex in services industries. For example, there are many different providers of health services, or financial services. Are (the providers of) acute care services in the same market as primary care services or physiotherapy?⁶¹ The boundaries of a relevant market are defined by three competitive restraints on firm behaviour: demand substitution, supply substitution, and (to a lesser extent) potential competition. We argue that these criteria need to be reassessed in the face of increasing non-price competition.

Supply substitution reflects the ability of suppliers in related markets to switch production to the relevant products (services) and market them in the short term without incurring significant additional costs or risks in response to small and permanent changes in relative prices.⁶² Potential competition refers to the possibility of new entry that may increase the relevant market.

⁶¹When evaluating the proposed merger between the Westpac Bank and the Bank of Melbourne, the Australian Competition and Consumer Commission identified distinct product clusters for deposits, home loans, small business banking, credit cards, personal loans and transaction accounts.

⁶²The European Commission Notice on the Definition of the Relevant Market for the Purposes of Community Competition Law, 1997.

Where supply side substitutability and potential competition are thought to have a disciplinary effect on the competitive behaviour of the firms involved, they are equivalent to the demand substitution effect.⁶³ It can be argued that the disciplinary effect of supply substitution in services markets is limited. Though supply substitution may be technically feasible in some cases, the importance of intangible assets as reputation and goodwill (and the required investments in these assets), as well as specialist skills, in services markets may limit supply substitution (and potential entry, for that matter) having a truly disciplinary effect on the firms considered. The same is true for the effect of potential entrants. However, the disciplinary effects stemming from fringe firms *already in the market* should not be underestimated. To the extent that services markets can be described by the endogenous sunk cost model of Sutton (1991), we may expect a varying number of fringe firms in the market. Though these firms may operate in market niches or provide lower quality services than the dominant firms in the market,⁶⁴ they may well limit the extent to which those dominant firms can set price above marginal cost.

Demand substitution measures the response in demand to a “Small but Significant Non-transitory Increase in Price”. This so-called SSNIP test seeks to identify the smallest market within which a hypothetical monopolist or cartel could impose such a price increase and defines this as the relevant market. Traditionally, demand substitution is regarded as constituting the most immediate and effective disciplinary force on the suppliers of a given

⁶³It should be noted that it is generally thought that the competitive restraints arising from supply side substitutability and potential competition are less immediate than those of demand substitution and though such constraints are taken into account, primary attention is focused on demand substitution.

⁶⁴Dick (2002) finds that in the banking industry, which can be well explained by the endogenous sunk cost model of Sutton (1991), the dominant firms provide a higher level of quality than fringe banks.

product, “[...] *in particular in relation to their pricing decisions*”.⁶⁵ However, given the prominent role of non-price competition in services markets and the fact that demand in these markets is based on price *and* non-price variables, we may expect considerably smaller demand substitution in services markets than in markets mainly characterised by price competition. However, limited demand substitution is regarded (at least partial) proof of a limited relevant market. By definition, the smaller the market, the larger is each firm’s market share. But large market shares may be associated with dominance, with the possible result that competition authorities will interpret firms’ actions to be anti-competitive.⁶⁶ In addition, a proposed merger may be thought to obtain a dominant position after merging and consequently may be denied or accepted on an inaccurate notion of market size.

The SSNIP test is difficult to apply in markets characterised by a high degree of product differentiation as there might not be a obvious chain of substitutes, which could potentially result in multiple competing market definitions.⁶⁷ Hence, the methods used to measure demand substitution clearly need to be reassessed in the face of growing services markets and the associated increasing importance of non-price competition. Instead of the often-used SSNIP test, competition authorities should look at alternative measures that recognise that demand is determined by both price *and* quality, and acknowledge that sole focus on price-elasticity may generate inaccurate market definitions and potentially inefficient policy decisions. Though some do not believe that an alternative sensible methodology to the SSNIP test exists,⁶⁸ the New Zealand Merger and Acquisition Guidelines

⁶⁵The European Commission Notice on the Definition of the Relevant Market for the Purposes of Community Competition Law, 1997.

⁶⁶Dominance is not regarded as anti-competitive as such, however when firms are considered to be dominant, their actions are thought to have the ability to restraint trade or competition and may therefore be considered anti-competitive.

⁶⁷See, for example, Rubinfeld (2000).

⁶⁸See NERA (2001), p19.

(2004) recognise that in differentiated product markets, market boundaries are often difficult to delineate precisely, as the different products can vary in the degree to which they are substitutable. As services are differentiated by nature, the following applies especially to services markets: “In these markets, the Commission considers the varying degree of competitive constraint on the products of the combined entity. Less close substitutes impose a lesser competitive constraint than closer substitutes”.⁶⁹ Where there is no obvious break in the chain of substitutes, antitrust authorities should focus upon competition analyses and the impact of a potential acquisition on localised prices, rather than precise market definition.

5.4 Other forms of regulation

There are a number of ways in which regulators can influence competition in services markets in addition to the general competition framework in place. In a recent Communication Report, the European Commission (2004) evaluates various methods that are used to regulate markets for professional services specifically. Where necessary (for example to guarantee quality, improve information provision, and/or guarantee sufficient supply), regulators can regulate prices, advertising, entry, and business structure, though the European Commission remarks that in all cases of regulation a proportionality test should be applied.⁷⁰ This section will shortly discuss each of these policy instruments.⁷¹ It should be noted that while the report of the European Commission concerns professional services only, some of the discussed regulatory measures could equally well be applied in other services markets.

Though some professional associations argue fixed prices will ensure low

⁶⁹The New Zealand Mergers and Acquisitions Guidelines (2004), p17. The ‘Commission’ is the New Zealand Commerce Commission.

⁷⁰In other words, regulators need to ensure that the intervention is proportional to the problem at hand and check the availability of other, less restricting options that may lead to the same result.

⁷¹For an extensive analysis on regulation of professional services see European Commission (2004).

prices, economic theory suggests that price regulation is unlikely to ensure prices that are lower than at competitive levels. It is also argued that maximum prices might protect consumers from excessive charges in markets with high entry barriers and a lack of effective competition. However, empirical evidence does not indicate that this is actually the case.⁷² An alternative to fixed prices is said to be ‘recommended’ prices, though it can effectively be argued that their main benefit - information provision - can also be obtained with alternative methods of providing price information that do *not* potentially facilitate the co-ordination of prices between service providers.

Particularly, professional services are often subject to some form of advertising regulation in addition to the general legislation that prevents untruthful or deceptive advertising. Sometimes advertising is prohibited completely, in other cases specific types of advertising content are proscribed. The ongoing debate on advertising regulation concentrates around two main arguments. On the one hand, advertising restrictions may contribute to lessen information asymmetries present in services markets, to the extent that it can guarantee truthful and non-manipulative information provision. Opponents of advertising restriction argue that if anything, it may hamper information provision and reduce competition by increasing the costs of gaining information. Empirical evidence on advertising restrictions in legal services seems to favour the latter argument, though no general comments can be made on the subject until more research has been conducted in this area.⁷³

The main rationale for qualitative restrictions on entry is to ensure the quality of the services provided. Such restrictions can take the form of minimum periods of education, professional examinations, and experience, and have traditionally been maintained by professional associations such as the American Medical Association. Experience has shown, however, that excessive licensing restrictions have in some cases led to higher prices without

⁷²Idem.

⁷³Stephen and Love (2000).

ensuring higher quality, and, conversely, that the loosening of restrictions in some professions has led to lower prices without any apparent detriment to quality. Fortunately, regulators recognise that most entry restrictions would not pass the proportionality test nowadays and that alternative mechanisms to guarantee quality should be used instead.⁷⁴

A final regulatory instrument to influence competition in services markets is business structure regulations, which may, for example, restrict business ownership and collaboration with other service providers. Any such restrictions are in place to ensure service providers' personal responsibility and liability towards customers and to avoid conflicts of interest. They may have negative consequences for dynamic efficiency, however, to the extent that they may inhibit providers from developing new services or cost-efficient business models. Also here, regulators should consider whether the restrictions are justified in terms of the fulfillment of their objective.

6 Conclusion

Though 'services' is a very broad concept and any generalisations should be made with caution, it is widely acknowledged that services differ from goods in many respects. In addition, we have come to understand that the production and consumption of services have distinctive elements. In the previous discussion, we have attempted to answer the question whether this distinctive working of services markets also implies a distinctive competitive process and, if so, if this requires a separate set of competition policy measures. In answering this question, we have analysed the main elements of market structure, conduct and performance that prevail in most services markets and the interface between these elements.

We have seen that most services markets can be described as either monopolistic competitive or oligopolistic. Key structural aspects that are strongly related to the nature of services are a high degree of product differ-

⁷⁴European Commission (2004), p16.

entiation, sunk costs, and asymmetric information. Though such a market structure tends to give service providers a certain degree of market power, most services markets appear to be characterised by intense non-price competition.

One feature of this form of competition is that service providers engage in both horizontal and vertical product differentiation, according to consumers' preferences. Another is the possibility of quality discrimination, which, according to empirical evidence, is particularly apparent in markets for medical and other experts' services. Finally, in oligopolistic services markets, non-price competition is likely to contain elements of interdependence, resulting in a positive conjectural variation that shows advertising outlays are strategic complements.

Market performance is invariably linked with market structure and conduct. The most important performance aspects in services markets are product diversity, advertising, and collusion. We have seen that in the presence of constant economies of scale, the optimal number of service varieties is essentially unbound. And although advertising is thought by some to be excessive, it can contribute to greater transparency in most services markets, which are inherently characterised by asymmetric information. As advertising outlays appear to be strategic complements in services markets, intense non-price competition through (sunk) investments in reputation and goodwill may result in concentrated market structures. This theory is known as the endogeneous sunk cost model or natural oligopoly theory as provided by Sutton (1991).

Possibilities for collusion seem limited in most services markets, as the focus on non-price competition complicates monitoring output and detecting cheating. Industry organisations, which we often find in medical professions, may affect entry conditions and quality standards. Where this is the case, the economic literature points out that enforcing competition in only one of these areas may result in second-best outcomes, particularly where regula-

tory intervention would pose a barrier to entry.

We found that the traditional competition policy tools inaccurately reflect competition and performance in services markets. Measurement problems and lack of data cause price, output, and profit to be less reliable measures. More importantly, quality, the main variable in non-price competition, receives insufficient attention in most competition analyses, leaving the authorities with an inaccurate image of competition in the market. In addition, whereas concentration ratios usually alarm competition authorities and point at inefficient market performance, high concentration rates may actually be an outcome of vigorous non-price competition in services markets. Finally, the main instrument of market definition, the SSNIP test, appears to be unsuited to evaluate competition in differentiated markets. The test relies on measurement of output and price-responsiveness, both of which are difficult to measure and only partially indicative of performance in services markets. Instead competition authorities are best to focus on competition analyses that include barriers to entry and search for alternatives that account for the disciplining power of non-price competition. Any alternatives and possibly service-specific regulations should of course be subject to a proportionality test to ensure efficiency and effectiveness.

Appendix A

The basic Endogeneous Sunk Cost (ESC) model

As long as sunk costs are exogenous (for example setup costs), concentration in both homogenous and heterogenous product markets declines indefinitely as market size increases and rises with what Sutton calls the “toughness of price competition”⁷⁵. In these types of industry, bigger markets allow sunk costs to be spread over larger sales (Ellickson, 2003). However, when sunk outlays such as advertising or R&D activities enhance consumers’ willingness-to-pay, thereby increasing the total size of the market, competition will lead to escalating investment in sunk costs by the dominant firms in the market. It is these endogenous sunk costs that prevent an industry to converge to a fragmented structure, however large the market may be. This market structure is referred to as a “natural oligopoly”.

Building on the theory of vertical product differentiation and assuming that increases in fixed outlays raise consumers’ willingness-to-pay, Sutton defines a three-stage game in which firms respectively decide on entry (at cost σ), choose a value of vertical attribute u (quality) at additional cost $A(u)$, and compete in a Cournot game in which all firms but one offer the same level of perceived quality \bar{u} , while one deviant firm offers some level u . The firm’s total fixed outlays are a combination of the setup cost σ and the advertising outlay $A(u)$, introducing a as the cost per message:

$$F(u) = \sigma + A(u) \tag{6}$$

where $A(U)$ is of the form

$$A(u) = \frac{a}{\gamma}(u^\gamma - 1), \quad \gamma > 1. \tag{7}$$

⁷⁵Sutton uses the term “toughness of price competition” to refer to the function linking concentration to prices or unit margins. This function will be affected by such features of the market as the physical nature of the product (homogeneous versus differentiated products) and the climate of competition policy.

Higher values of γ correspond to more rapidly diminishing returns to increases in outlays. The elasticity of the function $F(u)$ (i.e., the extent to which we have increasing, constant, or decreasing costs in u) can then be written as

$$\frac{u}{F} \frac{dF}{du} = \gamma \left[1 - \frac{\sigma - \frac{a}{\gamma}}{F} \right]. \quad (8)$$

As $u \rightarrow \infty$, $F(u) \rightarrow \infty$ and this elasticity tends to γ , but for any finite u (and thus $F(u)$), the elasticity lies above or below γ as $\sigma < \frac{a}{\gamma}$ or $\sigma > \frac{a}{\gamma}$. When $\sigma = \frac{a}{\gamma}$, the elasticity is constant for all u .

The relationship between market size S and the equilibrium number of firms N is specified by the equation

$$N + \frac{1}{N} + 2 = \frac{\gamma}{2} \left[1 - \frac{\sigma - \frac{a}{\gamma}}{S} N^2 \right]. \quad (9)$$

The right-hand side of this equation takes the value $\frac{\gamma}{2}$ as $S \rightarrow \infty$. Sutton denotes the unique N that solves for this as $\tilde{N}(\frac{\gamma}{2})$. Assuming zero profits, i.e., $F = S/N^2$, (9) reduces to

$$N + \frac{1}{N} + 2 = \frac{\gamma}{2} \left[1 - \frac{\sigma - \frac{a}{\gamma}}{F} \right], \quad (10)$$

which describes a locus in (N, F) space that is upward sloping, vertical, or downward sloping according as σ is greater than, equal to, or less than $a\gamma$. The market size / market structure relationship is different in each of these cases, which is illustrated in Figure (2):

- $\sigma = a\gamma$

For a sufficiently small market size S , where the marginal cost of advertising outweigh the marginal benefits, i.e., where

$$\left. \frac{d\Pi}{du} \right|_{u=\bar{u}=1} \leq \left. \frac{dF}{du} \right|_{u=1}, \quad (11)$$

equilibrium corresponds to that of the symmetric Cournot model where all firms offer a common level of perceived quality $\bar{u} = 1$ and no additional advertising outlays are incurred. As S increases, however, a

number of firms N is reached at which advertising begins - this so-called "switch value" is determined by replacing the inequality in (11) by an equality, writing this condition in explicit form, multiplying across by \bar{u} and solving the resulting expression

$$2S \frac{(N-1)^2}{N^3} = \gamma \left[F - \left(\sigma - \frac{a}{\gamma} \right) \right] \quad (12)$$

for N . In this case, the switch value coincides with $N_\infty = \tilde{N}(\frac{\gamma}{2})$. Further increases in market size involve only increasing levels of advertising, with no further change in market structure.

- $\sigma < a\gamma$

As S increases, N first increases until the switch value (where the market switches from non-advertising to advertising) is reached. Beyond this level, N falls as S increases and as $S \rightarrow \infty$, N converges asymptotically to the value $N_\infty = \tilde{N}(\frac{\gamma}{2})$.

- $\sigma > a\gamma$

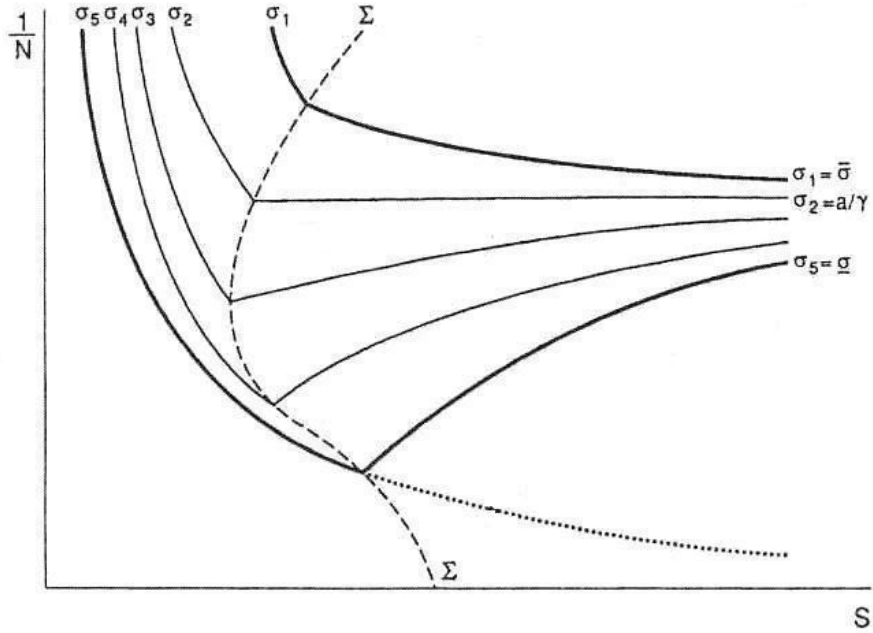
A similar analysis indicates that the number of firms N increases (i.e., concentration decreases) in the case of $\sigma > a\gamma$, as market size S increases, regardless of advertising levels.

Figure (2) shows the relationship between market size and market structure for these different values of σ . The $\sum \sum$ locus connects the switching values that indicate where the market switches from non-advertising to advertising.

Extensions and generalisations of the basic model

Two extensions of the basic model involve a dual market structure due to differences in consumer tastes, and first-mover advantages in a sequential game. In the first case, it is assumed that two groups of consumers exist, one being more or less sensitive to advertising (the retail segment) and one choosing suppliers largely on the basis of price alone (the non-retail

Figure 2: Concentration and market size



Source: Sutton (1991).

Note: $\bar{\sigma} = \sigma_1 > \sigma_2 = \frac{a}{\gamma} > \sigma_3 > \sigma_4 > \sigma_5 = \sigma$.

segment). Here, there are two critical values of market size S . First, as S increases, some critical value S_1 is reached where one firm begins to advertise. Sutton states that in this regime, consumers are indifferent between advertised and non-advertised goods as $p = u\bar{p}$. However, as S increases further, a second critical value S_2 is reached beyond which the (p, u) combination offered by the advertising firm(s) is strictly preferred by retail consumers ($p < u\bar{p}$). From this point on, the two sub-markets are essentially independent. In the retail segment, further increases in (sub-)market size will only increase levels of advertising, while the number of firms will stay constant. The retail market will develop as described in the basic model and hence,

will remain concentrated no matter how large the market becomes. The non-retail segment, on the other hand, will evolve according to the predictions of the traditional exogenous sunk cost model, in which the number of firms will increase with market size and consequently, concentration will converge to zero.

When allowing one or more firms to enjoy first-mover advantages, a strategic asymmetry is introduced which is best captured by a game of sequential entry. In an illustrative example, Sutton (1991) shows that the first mover can successfully preempt entry for γ close to 1 by only slightly increasing its advertising outlays above those that would be undertaken in a symmetric (duopoly) equilibrium⁷⁶. For larger values of γ , as market size S increases, a critical market size is reached beyond which a second firm enters. The leader's advertising levels (and profitability) will fall but will still be higher than the entrant's. Further increases in S lead to growth in both firms' advertising levels, but, Sutton remarks, the size disparity between the two remains.

Tested for different values of the parameters σ , a , and γ , the relationship between market size and market structure as described by (9) in the basic model and used in the extensions above has two important features, that can easily be inferred from Figure (2). First, the so-called 'convergence property' does not hold, i.e. increases in market size do not lead to an indefinite fall in the level of concentration. Secondly, the market size/market structure relationship is not necessarily monotonic. Empirical evidence suggests that industries (or sub-markets) that are relatively insensitive to advertising are well explained by the exogenous cost model (like the salt and sugar industry), while more advertising-intensive industries like the frozen food and the clothing industry appear to conform well to the predictions of the endogenous sunk cost theory.⁷⁷

⁷⁶Where γ is close to 1, duopoly profits are approximately zero.

⁷⁷Sutton (1991).

Appendix B

The foundation of modern economic literature on non-price competition is laid down in Chamberlin's (1948) model of monopolistic competition, in which quality variation is treated as a monopolistic element that distinguishes producers from one another within what is essentially one market. Building on this theory is the model of Spence (1977). In this model, Spence compares the supply of non-price services in various optima and equilibria, allowing for free entry. Basically, he states that performance may be inadequate even in the absence of excessive profits, not because prices do not accurately reflect costs, but because a failure in the area of non-price competition may cause costs to be too high or too low. Spence assumes demand depends upon prices or quantities (X) and upon non-price expenditures⁷⁸ (A) according to the specification:

$$B(X, A) = G \left[\sum_{i=1}^n \phi_i(x_i, a_i) \right] \quad (13)$$

where G is a concave function of all products and non-price activities produced in the industry. Assuming symmetry, constant elasticity, denoting the costs per firm as $c(x, a)$ and the number of firms as n , total surplus is

$$T = G(s) - nc(x, a).$$

As

$$s = n\phi(x, a),$$

total surplus can be written as

$$T = G(s) - s \frac{c(x, a)}{\phi(x, a)}.$$

From this it follows that the surplus is maximised when c/ϕ is minimised with respect to x and a , i.e. when

$$G'(s) = \min_{x, a} \left(\frac{c}{\phi} \right). \quad (14)$$

⁷⁸I.e., expenditures that are made for non-price competition purposes.

However, firm profits equal revenues (price p_i times quantity x_i) minus costs. As the inverse demand for firm i is given by

$$p_i(X, A) = G'(s) \frac{\delta \phi_i}{\delta x_i}(x_i, a_i),$$

the profits of the representative firm are thus

$$\pi = G'(s)x\phi_x(x, a) - c(x, a)$$

and equilibrium occurs when

$$G'(s) = \min_{x,a} \left(\frac{c}{x\phi_x} \right). \quad (15)$$

A crucial assumption in Spence's model is free entry. The threat of entry forces firms to adopt strategies that minimise $\frac{c}{x\phi_x}$. Because ϕ is concave in x (i.e., demands are downward sloping), $x\phi_x < \phi$. This implies that the minimum of $\frac{c}{\phi}$ is less than the minimum of $\frac{c}{x\phi_x}$ and thus, that profits are negative at the optimum. In essence, in the constant elasticity case, markets tend to oversupply non-price services. Relaxing the assumption of constant elasticity, Spence shows that market performance is determined by whether the ratio of revenues to incremental surplus declines or increases in x and a , i.e.:

$$\frac{\delta}{\delta a} \left(\frac{x\phi_x}{\phi} \right) < / > 0 \text{ and } \frac{\delta}{\delta x} \left(\frac{x\phi_x}{\phi} \right) < / > 0. \quad (16)$$

According to Spence, a market tends to oversupply non-price services that increase this ratio, and to undersupply those that do not.

Appendix C

In an interesting extension of the Arrow-Nerlove model to oligopolistic markets, Fershtman (1984) finds that the optimal stock of goodwill \hat{G}_i can be described as a function of the current market share $S_i(t)$:

$$\hat{G}_i(S_i) = (p - m_i)\phi(p) \left[\frac{\alpha S_i(1 - S_i)}{r + \delta} \right] \quad (17)$$

where

$$(p - m_i)\phi(p)$$

is the markup (price minus marginal cost) on the quantity supplied and

$$\frac{\alpha S_i(1 - S_i)}{r + \delta}$$

is a discounted measure of the optimal stock of goodwill given market share S_i and the goodwill-elasticity of demand, α . Now we can derive the optimal advertising policy for the i th firm at time t , since the market share $S_i(t)$ is given:

$$\begin{cases} a_i(t) = 0 & \text{if } G_i(t) > \hat{G}_i S_i(t) \\ a_i(t) = \delta \hat{G}_i S_i(t) & \text{if } G_i(t) = \hat{G}_i S_i(t) \\ a_i(t) = \infty & \text{if } G_i(t) < \hat{G}_i S_i(t). \end{cases} \quad (18)$$

In other words, where actual goodwill exceeds the optimal amount of goodwill (as a function of market share $S_i(t)$), it is best not to advertise. In the opposite case, firms's optimal advertising policy is to advertise 'infinitely' until actual goodwill equals optimal goodwill. Where this is already the case, firms are best of maintaining their current levels of advertising.

References

- Akerlof, G.A. (1970), "The Market for 'Lemons': Quality Uncertainty and the Market Mechanism", *Quarterly Journal of Economics*, Vol. 84, No. 3, pp. 488-500
- Arrow, K.J. and M. Nerlove (1962), "Optimal Advertising Policy under Dynamic Conditions", *Economica*, New Series, Vol. 29, No. 114 (May 1962), pp. 129-142
- Atle Berg, S. and M. Kim (1994), "Oligopolistic Interdependence and the Structure of Production in Banking: An Empirical Evaluation", *Journal of Money, Credit and Banking*, Vol. 26, No. 2, pp. 309-322
- Australian Merger Guidelines, June 1999
- Australian Productivity Commission (2002), "Australia's Service Sector: A Study in Diversity", Staff Research Paper
- Bain, J.S. (1956), "*Barriers to New Competition: Their Character and Consequences in Manufacturing Industries*", Cambridge, Harvard University Press
- Bain, J.S. (1968), "*Industrial Organization*", 2nd ed., John Wiley and Sons, Inc., New York, United States
- Baumol, W.J. and R.D. Willig (1981), "Fixed Costs, Sunk Costs, Entry Barriers, and Sustainability of Monopoly", *The Quarterly Journal of Economics*, Vol. 96, No. 3, pp. 405-431
- Becker, G.S. and K.M. Murphy (1993), "A Simple Theory of Advertising as a Good or Bad", *The Quarterly Journal of Economics*, Vol. 108, No. 4, pp. 941-964
- Boden, M. and I. Miles (2000), "*Services and the Knowledge-Based Economy*", Continuum, London and New York

- Carlton, D.W. and J.M. Perloff (2000), “*Modern Industrial Organisation*”, 3rd ed., Addison-Wesley Longman, Inc., United States
- Clark, C. (1940), “*The Conditions of Economic Progress*”, 3rd ed. (1957), Macmillan and Co Ltd, New York, USA
- Connors, J.R. (2003), “A critical misdiagnosis: How courts underestimate the anticompetitive implications of hospital mergers”, *California Law Review*, Vol. 91, No. 2, p.543
- Dick, A.A. (2002), “*Market Structure and Quality: An Application to the Banking Industry*”, Federal Reserve Board, Washington, D.C., United States
- Dixit, A. (1979), “Quality and Quantity Competition”, *The Review of Economic Studies*, Vol 46, No. 4 (Oct. 1979), pp. 587-599
- Dixit, A. and J.E. Stiglitz (1977), “Monopolistic Competition and Optimum Product Diversity”, *The American Economic Review*, Vol. 67, No. 3, pp. 297-308
- Dorfman, R. and P.O. Steiner (1954), “Optimal Advertising and Optimal Quality”, *The American Economic Review*, Vol. 44, No. 5., pp. 826-836
- Dulleck, U. and R. Kerschbamer (2003), “*Price Discrimination in Markets for Experts’ Services*”, Working Paper No: 0312, Department of Economics, University of Vienna, Austria
- Ellickson, P.B. (2003), “*Supermarkets as a Natural Oligopoly*”, Duke University, version February 27, 2003
- European Commission (1997), Commission Notice on the definition of the relevant market for the purposes of Community competition law, OJ C 379 (9/12/1997)

- European Commission (2004), Communication from the Commission, Report on Competition in Professional Services, 9 February 2004, Brussels
- Fershtman, C. (1984), “Goodwill and Market Shares in Oligopoly”, *Economica*, New Series, Vol. 51, No. 203, pp. 271-281
- Fisher, A.G.B. (1935), “*The Clash of Progress and Security*”, Macmillan and Co Ltd, New York
- Fudenberg, D. and J. Tirole (1984), “The Fat Cat Effect, the Puppy Dog Ploy and the Lean and Hungry Look”, *American Economic Review*, Vol. 74, No. 2, pp. 361-366
- Gabszewicz, J.J. and J.-F. Thisse, “On the Nature of Competition with Differentiated Products”, *The Economic Journal*, Vol. 96, No. 381, pp. 160-172
- Gaynor, M. and D. Haas-Wilson (1999), “Change, Consolidation, and Competition in Health Care Markets”, *The Journal of Economic Perspectives*, Vol. 13, No. 1, pp. 141-164
- Ginsburg, D.H. (1993), “Nonprice competition”, *Antitrust Bulletin*, Vol. 38, Iss. 1 (Spring 1993), pp. 83-111
- Hill, T.P. (1977), “On Goods and Services”, *Review of Income and Wealth*, Vol. 23, Issue 4 (Dec. 1977)
- Hill, T.P. (1997), “Tangibles, intangibles and services: a new taxonomy for the classification of output”, *Canadian Journal of Economics*, Vol. 32, No. 2 (April 1999)
- Holmström, B. (1985), “The provision of services in the market economy” in: Inman, R.P. (ed.), “*Managing the service economy - Prospects and problems*”, Cambridge University Press

- Holmström, B. and P. Milgrom (1991), “Multi-task principal-agent analyses: incentive contracts, asset ownership, and job design”, *Journal of Law, Economics and Organization*, Vol. 7, pp. 24-52
- Hotelling, H. (1929), “Stability in Competition”, *Economic Journal*, Vol. 39, in: *The Collected Economic Articles of Harold Hotelling*, (1990) edited by Darnell, A.C., Springer-Verlag, New York
- Ivaldi, M., B. Jullien, P. Rey, P. Seabright and J. Tirole (2003), “The Economics of Tacit Collusion”, Final Report for DG COmpetition, European Commission, March 2003, IDEI, Toulouse
- Joyce, J.M. (1989), “Effect of Firm Organizational Structure on Incentive to Engage in Price Fixing”, *Contemporary Economic Policy*, Vol. 7, No. 4, pp. 19-35
- Katouzian, M.A. (1970), “The Development of the Service Sector: A New Approach”, *Oxford Economic Papers*, Vol. 22, No. 3, pp. 362-382
- Lancaster, K.J. (1975), “Socially Optimal Product Differentiation”, *The American Economic Review*, Vol. 65, No. 4, pp. 567-585
- Lancaster, K.J. (1998), “Non-Price Competition”, topic in John Eatwell, Murray Milgate, and Peter Neuman eds. (1998), “*The New Palgrave: A Dictionary of Economics*”, Stockton Press, New York, United States
- Marshall (1890), “*Principles of Economics*”, Macmillan and Co., Ltd., 8th ed. (1920), London, United Kingdom
- Martin, S. (1994), “*Industrial economics: economic analysis and public policy*”, 2nd ed., Prentice-Hall Inc., New Jersey, United States
- Masson, T. and S. Wu (1974), “Price Discrimination for Physicians’ Services”, *The Journal of Human Resources*, Vol. 9, No. 1, pp. 63-79

- Netter, J.M. (1982), "Excessive Advertising: An Empirical Analysis", *The Journal of Industrial Economics*, Vol. 30, No. 4, pp. 361-373
- Noether, M. (1986), "The Growing Supply of Physicians: Has the Market Become More Competitive?", *Journal of Labor Economics*, Vol. 4, No. 4, pp. 503-537
- Perloff, J.M. (2001), "*Microeconomics*", 2nd ed., Addison Wesley Longman, Inc., United States
- Roberts, M.J. and L. Samuelson (1988), "An empirical analysis of dynamic, nonprice competition in an oligopolistic industry", *RAND Journal of Economics*, Vol. 19, No. 2 (Summer 1988)
- Salani, B. (2000), "*The Microeconomics of Market Failure*", Massachusetts Institute of Technology, United States
- Sampson, G.P. and R.H. Snape (1985), "Identifying the Issues in Trade in Services", *The World Economy*, Vol. 8, Issue 2, pp.171-81
- Schmalensee, R. (1976), "A Model of Promotional Competition in Oligopoly", *The Review of Economic Studies*, Vol. 43, No. 3 (Oct. 1976), pp. 493-507
- Seldon, B.J., S. Banerjee, and R.G. Boyd (1993), "Advertising Conjectures and the Nature of Advertising Competition in an Oligopoly", *Managerial and Decision Economics*, Vol. 14, No. 6, pp. 489-498
- Shaked, A. and J. Sutton (1987), "Product Differentiation and Industrial Structure", *The Journal of Industrial Economics*, Vol. 36, No. 2, pp. 131-146
- Smith, A. (1776), "*An Inquiry into the Nature and Causes of the Wealth of Nations*", Cannan, E. ed. (1904), Methuen and Co., Ltd. 5th ed., London, United Kingdom

- Sosnick, S.H. (1958), "A Critique of Concepts of Workable Competition", *The Quarterly Journal of Economics*, Vol. 72, No. 3, pp. 380-423
- Spence, M. (1977), "Nonprice Competition", *The American Economic Review*, Vol. 67, No. 1, Papers and Proceedings of the Eighty-ninth Annual Meeting of the American Economic Association (Feb. 1977), pp. 255-259
- Stephen, F.H. and J.H. Love, "*Regulation of the Legal Profession*", in B. Bouckaert and G. De Geest (eds) (2000), *Encyclopedia of Law and Economics*, Volume III: The Regulation of Contracts, Cheltenham
- Stigler, G.J. (1968), "The Organization of Industry", Richard D. Irwin, Inc., Illinois, United States
- Sutton, J. (1991), "*Sunk Costs and Market Structure - Price competition, Advertising, and the Evolution of Concentration*", Massachusetts Institute of Technology, United States
- Tirole, J. (1988), "*The Theory of Industrial Organisation*", Massachusetts Institute of Technology, United States
- Whichard, O.G. (2001), "Measurement and classification of Service Sector Activity: Data Needs for GATS 2000" in: Stern, R.M. (2001), "*Services in the International Economy*", University of Michigan Press, United States
- Wolinsky, A. (1993), "Competition in a market for informed experts' services", *RAND Journal of Economics*, Vol. 24, No. 3, pp. 380-398