FAULL & NIKPAY

THE EU LAW OF COMPETITION

THIRD EDITION

Edited by

JONATHAN FAULL

Director General, European Commission
Visiting Professor, College of Europe, Bruges, and King's College, London
Emeritus Professor, Vrije Universiteit Brussel

ALI NIKPAY

Partner, Gibson Dunn & Crutcher LLP
Former Senior Director, UK Office of Fair Trading
Visiting Lecturer at University of Oxford

Assistant Editor

DEIRDRE TAYLOR

Of Counsel, Gibson Dunn & Crutcher LLP

OXFORD
UNIVERSITY PRESS

OXFORD
UNIVERSITY PRESS

1

THE ECONOMICS OF COMPETITION

Luc Peeperkorn and Vincent Verouden

A. Introduction

1.01 Nowadays, there is a clear awareness among competition policymakers, competition lawyers, and judges of the importance of economics for their daily work. In the EU, the US, and many other parts of the world, it is normal practice to discuss competition cases in terms of economic concepts such as market power, entry barriers, and sunk costs, and to evaluate cases according to their effects on the market. Competition policy is economic policy concerned with economic structures, economic conduct, and economic effects. It is for this reason that in a book on competition law an introduction to the economics of competition is of importance.

1.02 The growing acceptance and importance of economics in competition policy raises questions regarding the usefulness of economics, both for devising competition rules and for

deciding on competition cases. A word of caution is appropriate in this respect. Economic thinking and economic models have proved not to be perfect guides.

1.03 Economic theories and models are built on and around assumptions. This approach has the benefit of making explicit the various elements relied upon in arriving at a particular conclusion or insight. At the same time, these assumptions by definition do not cover (all) real-world situations. In addition, when the assumptions are changed the outcomes of the models may look very different. It is for these reasons that the application of economic theories may not always be able to give a clear and definite answer, for example as to what will happen in a market when companies merge, or when companies try to collude or engage in specific types of conduct.

1.04 The best that the application of economic principles can do in general is to provide a coherent framework of analysis, to provide relevant lines of reasoning, to identify the main issues to be checked in the context of certain theories of competitive harm, and possibly to exclude certain outcomes. The application of empirical methods may further help to test the relevance of theories of harm. In this way, economics helps to tell the most plausible story. In individual cases, it will be necessary first to find the concepts and the model that best fit the actual market conditions of the case and then to proceed with the analysis of the actual or possible competition consequences. Economic insights can also be useful in the formulation of policy rules, indicating under what conditions anti-competitive outcomes are very unlikely, very likely, or rather likely, and helping to devise safe harbours.

1.05 The competition policy practitioner is advised to follow the mainstream of economics in order to avoid too much contradiction and too many untested assumptions. This chapter gives a short introduction to the main insights of industrial economics.[1] It has the following structure:

- Section B briefly describes the main historical trends in the field of industrial economics;
- Section C describes the static welfare aspects of market power and introduces a number of (microeconomic) concepts that are commonly used in this context;
- Section D describes the dynamic welfare aspects of market power;
- Section E describes market definition as a method for identifying the extent to which products exert a competitive constraint on each other;
- Section F looks into the concepts of market power and market dominance in further detail and focuses on the ways in which market power may be maintained or enhanced through anti-competitive means; and
- Section G presents a number of empirical methods to verify the existence of competitive constraints and market power.

---

[1] Industrial economics or industrial organization can be described as applied microeconomics: it uses the models and concepts of microeconomics in an effort to understand the development of real-world markets and company behaviour. For an excellent introduction, see F. M. Scherer and D. Ross, *Industrial Market Structure and Economic Performance* (3rd edn, Boston: Houghton Mifflin, 1990). More recently, J. Church and R. Ware, *Industrial Organization—A Strategic Approach* (Boston: McGraw-Hill, 2000); D. W. Carlton and J. M. Perloff, *Modern Industrial Organization* (4th edn, Addison Wesley, 2004), M. Motta, *Competition Policy. Theory and Practice* (Cambridge: Cambridge University Press, 2004), and J. Lipczynski, J. O. S. Wilson, and J. Goddard, *Industrial Organization: Competition, Strategy, Policy* (3rd edn, Harlow: FT Prentice Hall, 2009). More technical and elaborate is M. Armstrong and R. Porter (eds), *Handbook of Industrial Organization* (Amsterdam: North Holland, 2007) and J. Tirole, *The Theory of Industrial Organization* (Cambridge, MA: MIT Press, 1988).

B. Structure, Conduct, Performance

(1) Early Developments

Interest in the issues of market power and cartels arose well before the twentieth century. **1.06** Descriptions of the dangers of monopoly can be found in ancient Greek written sources as well as in the Bible. Adam Smith in his *Wealth of Nations* (1776) made the famous remark that people of the same trade seldom meet, even for merriment and diversion, without it ending in a conspiracy to raise prices. In general, Smith warned against the negative effects of monopoly, both private monopoly and monopoly sponsored by government.

In the nineteenth century, neoclassical authors such as Augustin Cournot and Alfred Marshall **1.07** laid the basis for modern microeconomics with the development of simple models of perfect competition, monopoly, and duopoly. The hallmark of neoclassical economics is the paradigm of rational economic agents maximizing their utility (think of firms maximizing profit or consumers maximizing their welfare). The model of perfect competition was especially useful for developing a theory on general equilibrium for the whole of the economy. However, towards the end of the nineteenth century it became obvious that these models were unable adequately to describe market developments such as market concentration, the emergence of trusts, product differentiation, non-price competition, and advertising.

Research in the first half of the twentieth century also seems to indicate that companies were **1.08** not necessarily producing, as the model of perfect competition would predict, at minimum/ lowest average costs.[2] Instead, they were sometimes producing on a decreasing cost curve, that is, where there are increasing returns to scale, without, however, becoming much bigger. This phenomenon, known as the Great Cost Controversy, led several authors such as Piero Sraffa, Edwin Chamberlin, and Joan Robinson to write about imperfect and monopolistic competition, that is, those situations in between the two extremes of perfect competition and monopoly which more accurately describe how most markets function. In order to provide a rationale for imperfect and monopolistic competition, they were among the first to explore the role of product differentiation and advertising in their models.

(2) The Harvard School

Not satisfied with the limited, rather simple models mentioned in the previous section, **1.09** at around the time of World War II a number of economists such as John Clark, Edward Mason, and Joe Bain started to look for more empirically supported explanations of market phenomena.[3] They tried to develop a type of applied microeconomics. Instead of deduction based on assumptions, they wanted to take account of the richness of the real world. Data were gathered and by induction they tried to develop general insights concerning likely company behaviour, effects on the market, and possibilities for government intervention.

The main result of this so-called Harvard School, that dominated the industrial economics **1.10** scene for many years, is the Structure-Conduct-Performance (S-C-P) paradigm. In its simplest form it states that market structure determines companies' market behaviour which in turn determines market performance. Market structure, being the basis of the explanation,

---

[2] For the cost concepts used, see Section C, esp paras 1.31–1.35 and 1.58–1.77.
[3] See eg J. M. Clark, 'Toward a Concept of Workable Competition' (1940) 30 Am Econ Rev 241.

is seen as of paramount importance. In its most mechanistic form, the study of conduct becomes quite irrelevant. It is the structure that is responsible for the final market outcome. Studies were carried out for several industries collecting market structure data such as concentration ratios and the height of entry barriers. These data were linked to performance indicators such as profit levels. The general conclusion of these studies was that concentrated markets with entry barriers showed above average profitability. This approach fitted well with the general trend for structuralist theories and explanations developed in the social sciences in the 1940s, 1950s, and 1960s.

Figure 1.1 The simple S-C-P scheme

1.11   The main policy conclusion flowing from the simple S-C-P scheme (see Figure 1.1) has been that competition policy should concentrate on the structure of markets and on structural remedies, ensuring that markets do not become (overly) concentrated or that entry barriers be erected. This was reflected, for example, in the use of market concentration measures in assessing merger cases in the 1968 Horizontal Merger Guidelines issued by the US Department of Justice. Behavioural remedies to a competition problem were seen as ineffective without the necessary structural changes.

(3)   The Chicago School

1.12   A number of economists such as George Stigler, Harold Demsetz, and Yale Brozen questioned the S-C-P framework and its conclusion that market concentration in general leads to monopoly profits. This group of scholars, also known as the Chicago School, argued that competition policy should be less concerned with market structure and should focus more directly on the concept of economic efficiency (welfare) in evaluating business conduct or mergers.[4]

1.13   The Chicago School criticized the empirical studies underlying the S-C-P paradigm. By applying different techniques to the same data and by using improved or new data, they showed that the relationship between concentration, entry barriers, and monopoly profits was not so stable or strong and, at times, was even non-existent. More important, however, was their theoretical questioning of the S-C-P paradigm.

1.14   The Chicago School argued that the causal link is not between high concentration, on the one hand, and high profits, on the other. Instead, they argued that the causality runs as

---

[4]   In economics, the term 'efficiency' (or 'economic efficiency') generally refers to the extent to which welfare is optimized in a particular market or in the economy at large. Welfare is often conceived as the (weighted) sum of consumer surplus (the difference between consumers' willingness to pay for consumption and the price paid) and producer surplus (company profits): see Section C for further details. It should be noted that the weights accorded to consumer surplus and producer surplus imply a certain value judgement. The Chicago School proposed to use equal weights, arguing that not antitrust, but other laws should address the ways prosperity is used or distributed in society. See R. Bork, *The Antitrust Paradox: A Policy at War with Itself* (New York: Basic Books, 1978), ch 5. For a detailed account of the Chicago School, see H. Hovenkamp, *Federal Antitrust Policy: The Law of Competition and Its Practice* (2nd edn, St Paul: West, 1999), 60; M. W. Reder, 'Chicago Economics: Permanence and Change' (1982) 20 J Econ Lit 1.

follows: increased firm size leads to increased firm efficiency, which in turn leads to market concentration and ultimately to possibly higher profits. Central to this reasoning is the role of economies of scale and scope and a general belief that competition forces companies to become superior in terms of efficiency. The companies that succeed in this way will grow faster than others which may even go out of business. This may at times lead to higher concentration levels in the industry but, if this is the product of the market process which seeks and obtains efficiency, this is desirable from a competition policy point of view. It leads to more efficient firms, even when it would also result in profits in excess of the competitive norm. Monopoly profits would not be very likely to arise and certainly would not be durable, as it was argued that entry barriers are rarely very high and can be overcome in time. The more extreme statement of the Chicago School is that the only high and durable entry barriers are those created by the State, thereby telling governments to clean up their own act instead of pursuing vigorous competition policy.

These attacks of the Chicago School, that started in the 1960s but culminated in the 1970s and 1980s, brought back a greater reliance on the (self-correcting) forces of competition. High concentration is not necessarily bad and only in very particular circumstances is competition policy action called for. This fitted well with the general trend in the 1970s and especially the 1980s of seeing limits to the effectiveness of and scope for government interference.[5]   **1.15**

(4) More Recent Developments

The Chicago School returned in part to the deductive approach of the microeconomic models, focusing more on the theoretical underpinnings than on empirical testing. It highlighted the main theoretical weaknesses in the arguments of the Harvard School and it forced a reconsideration of the S-C-P framework that, as a consequence, has been extended and refined over the years. It has been recognized that a wide array of other basic conditions, such as consumer preferences and technological developments, influence the market structure and that these basic conditions may themselves change. Just as important, it has been accepted that conduct is not a negligible factor when it comes to explaining performance. In addition, it is recognized that conduct and also performance may help to shape the market structure. In other words, although the main causal link may still run from market structure to conduct to market performance, feedback mechanisms complicate the picture. In schematic terms, the resulting extended S-C-P framework can be illustrated as in Figure 1.2.[6]   **1.16**

This extended S-C-P framework is still important today in industrial economics and in competition policy, not as the perfect explanatory framework but as a good way to organize one's thoughts. Market structure is still the starting point for competition policy arguments and it is generally accepted that certain market structure conditions are a prerequisite for anti-competitive conduct and performance. However, these necessary conditions may not be sufficient. Conduct such as limit pricing or excess capacity creation to limit or prevent entry into the market, may play its own distinctive role. Structural conditions can be used to describe safe harbours: that is, situations in which anti-competitive behaviour or effects   **1.17**

---

[5] cf E. Fox, 'What is Harm to Competition? Exclusionary Practices and Anticompetitive Effect' (2002) 70 Antitrust LJ 371, 377.

[6] Adapted from Scherer and Ross, *Industrial Market Structure and Economic Performance* (n 1), Fig 1.1, p 5.
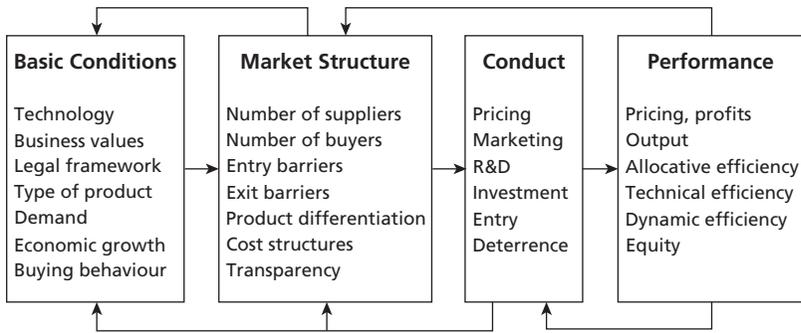
Figure 1.2 The extended S-C-P framework

are highly unlikely. However, to find anti-competitive situations, usually structural, behavioural, and performance aspects will have to be taken into account. Under Articles 101 and 102 Treaty on the Functioning of the European Union (TFEU) it is in general not enough to show that the market structure enables anti-competitive conduct: also the conduct itself and/or the likely negative effects that may result from this conduct must be shown. The same holds true under the EU Merger Regulation where, to assess the impact of a merger on competition, a purely structural analysis may not suffice.

1.18 The renewed attention to the behaviour of companies can in part be ascribed to a significant development in industrial economics since the mid-1980s, sometimes referred to as New Industrial Economics. The centre of attention has shifted to the possible *strategic behaviour* of companies in oligopolistic situations, trying to deduce, within the framework of more sophisticated microeconomic models and with the help of game theory, what the most likely company strategies are and whether or not anti-competitive strategies are likely.[7] It has cast more light on the (efficiency) rationales behind certain types of company behaviour, such as the use of vertical restraints, without however always leading to particularly robust outcomes. Rather, the models have led to the more general insight that whether or not business strategies are likely to have anti-competitive effects typically depends on the precise circumstances of the case. It has thus lent support to a more moderate, less ideological approach in antitrust where more emphasis has come to lie on the assessment of the facts of the case (case-by-case approach). This in turn has led to the further development, especially since the mid-1990s, of empirical techniques to verify the existence of market power and competitive constraints, a field generally referred to as Empirical Industrial Organization.[8]

1.19 Another even more recent reflection on conduct in economics is provided by what is commonly called 'behavioural economics'. Building on earlier ideas about the limits of economic agents' rationality and on insights from psychology, behavioural economics criticizes the neoclassical assumption that economic agents are able and willing in all circumstances to maximize their utility. It is found that in particular consumers have practical limits to process information, may be influenced in their decisions by how choices are presented, may have difficulties in anticipating their future needs, and may care more about losses than

---

[7] For a more detailed account, see Section C.5.

[8] For an overview, see L. Einav and J. Levin, 'Empirical Industrial Organization: A Progress Report' (2010) 24(2) J Econ Perspectives 145.

about gains. Consumers appear to exhibit a number of systematic biases in the way they access information about offers in the market, in the way they assess these offers, and in the way they subsequently act by purchasing a product or switching between products.[9] The main 'behavioural biases' concern processing power biases (including making use of rules of thumb), framing biases (including having a preference to choose the default option or the first or last option on a list), time inconsistency biases (including over- or underestimating how much a product will be used), and loss aversion biases (including the endowment effect of valuing a product more once it is owned than before it is owned).[10]

Firms may try to exacerbate and exploit these behavioural biases and to manipulate consumer choice in order to foreclose competitors and/or to increase consumers' willingness to pay, without necessarily increasing the utility derived from consuming the product.[11] This may in certain cases result in a reduction of (the intensity of) competition and an overall increase in prices. This does not necessitate a major shift in competition policy. In most markets, the ability of a firm to exploit such biases will be undermined and prevented by competition, for instance by competitors offering products which make a virtue out of not exploiting these biases. Competition remains vital to provide choice and quality and competition policy remains a crucial tool to make markets work well for consumers. But it may indicate that sometimes consumers are more easily harmed than would otherwise be expected and that a remedy, in order to be effective, should take the behavioural biases into account.[12] Behavioural economics also reminds us of the fact that competition policy is only one tool in the toolbox and that there will be situations where conduct of firms harms consumer welfare but where competition policy may not be well placed to solve the problem and where, for instance, consumer policy intervention requiring to reduce the complexity or increase the transparency of pricing of firms may be more effective.

1.20

C. Static Welfare Analysis of Market Power

(1) Introduction

In a nutshell, one could say that the economics of competition is about market power: what it is, how it is created or sustained, and what are its effects? This section provides a

1.21

---

[9] There are good reasons to expect that firms will in general be less inclined to have behavioural biases. Firms usually operate on a larger scale and can thus make use of economies of scale to process information and run their activities professionally. In addition, it may be expected that the market will discipline firms that make sub-optimal choices, by reducing their profits and market shares. Nonetheless, firms may also have behavioural biases, see M. Armstrong and S. Huck, 'Behavioral Economics as Applied to Firms: A Primer', Competition Policy International, Vol 6, No 1, Spring 2010. To the extent that these biases influence the likelihood of collusion, this is referred to in Section C.5(c) and (d).

[10] See M. Bennett, J. Fingleton, A. Fletcher, L. Hurley, and D. Ruck, 'What Does Behavioral Economics Mean for Competition Policy', Competition Policy International, Vol 6, No 1, Spring 2010.

[11] See E. Garcés Tolon, 'The Impact of Behavioral Economics on Consumer and Competition Policies', Competition Policy International, Vol 6, No 1, Spring 2010.

[12] eg if consumers have a bias towards remaining with the default option even if switching to alternative options is factually possible at little cost, a strategy to foreclose which would otherwise be unlikely to work could become effective. In case of a default bias, the remedy could be to prohibit the bundling, but could also be only to require that consumers are offered an explicit choice to avoid or reduce the bias. See the Art 9 remedy in the *Microsoft Internet Explorer* case, 'Commission welcomes Microsoft's roll-out of web browser choice', Press Release IP/10/216 (March 2010).

microeconomic perspective on these questions and introduces a number of concepts that are commonly used in this context.

1.22 The answer given by economists on the first question asked—what is market power?—concentrates on the power to raise price above the competitive level.[13] In the short run this means the power to raise price above marginal cost and in the long run above average total cost.[14] In other words, a company has market power if it has a perceptible influence on the price against which it can sell and, if by charging a price above the competitive level, it is able, at least for a significant period, to obtain 'supra-normal' profits (often referred to as 'monopoly' profits).

1.23 This makes it very clear that market power is not a black-and-white concept and that companies can have different degrees of market power. In principle, the appropriate measuring rod would be the net present value of the monopoly profits a company can make.[15] The net present value is today's value of the profit of this period and all future periods. It depends, therefore, on the monopoly profit per period, on the number of periods a monopoly profit can be sustained before entry or expansion by competitors takes the profit away, and on the discount rate against which future profits are evaluated.[16]

1.24 A firm with market power may raise its price by reducing its own output or by making competitors reduce theirs. As stated in para 1.22, this price increase should increase the firm's profits and do so for a significant period of time. What qualifies as a significant period of time will depend on the product and on the circumstances of the market in question, but under Article 102 normally a period of two years will be sufficient to find dominance.[17] Under the merger rules, the test is also, in practice, whether the merging companies involved will, in all likelihood, be able to obtain supra-normal profits for a period longer than two years. Under Article 101, shorter periods are also normally taken into account.

1.25 The second question about how market power is created or sustained brings us back to the question of the relevant elements of market structure and conduct. And so does the third question about its effects. This section is devoted to a static welfare analysis of these questions. By static it is meant that the state of technology is assumed to be constant and effects of market power on innovation and vice versa are ignored. The latter effects are dealt with in Section D, not surprisingly titled 'Dynamic Welfare Analysis of Market Power'.

---

[13] Obviously, any company can raise the price at which it sells if it disregards the effect that would have on its sales and profits. What is meant by the ability to raise price above the competitive level is the ability to do so *profitably*.

[14] The terminology used comes back and is explained in later parts in this section.

[15] In practice, however, the assessment of market power is rarely carried out by measuring profit margins. Instead, whether a particular firm has market power is generally addressed by investigating the factors that, in general, tend to determine these profit margins, in particular: (a) constraints imposed by the existing supplies from, and the position on the market of, actual competitors; (b) constraints imposed by the credible threat of future expansion by actual competitors or entry by potential competitors; and (c) constraints imposed by the bargaining strength of the firm's customers (see Section F.1).

[16] The net present value of a stream of profits is given by: $\text{NPV} = \sum_{i=1}^{n} \frac{1}{(1+\text{r})^i} \pi_i$, where n is the number of periods a monopoly profit is made, π_i is the profit in period i, r is the discount rate, and $\sum$ the summation sign for the different periods. As discount rate, usually the competitive rate of return on capital or the rate at which the company can lend money is taken, since this measures the opportunity cost of using the company's own funds.

[17] See Guidance on the Commission's enforcement priorities in applying Article [102 TFEU] to abusive exclusionary conduct by dominant undertakings, OJ 2009 C45/7, para 11.

Welfare economics is the branch of microeconomics concerned with the efficiency of the **1.26** company/the market/the economy.[18] A welfare economic analysis of the effects of market power concentrates on the effects on efficiency, both allocative and productive efficiency,[19] and therewith the effect on total welfare. The following subsections provide an explanation of these and other microeconomic concepts and analyse the market structures of perfect competition, monopoly, and oligopoly on their welfare effects.

(2) Basic Microeconomic Concepts

In this subsection, the following basic microeconomic concepts are discussed: consumer **1.27** surplus (short and long run) production costs, profit maximization, economies of scale, minimum efficient scale, entry barriers, and contestable markets.[20]

(a) Consumer Surplus

Consumer surplus is the net benefit consumers obtain by buying a certain good or service. **1.28** It is the difference between their willingness to pay, sometimes called their reservation price, and the price actually paid. As consumers have different preferences and incomes, some are normally willing to pay more than others for a certain good. Also, the higher the quantity of the good a particular consumer obtains, the lower in general his willingness to pay for an additional unit. These characteristics mean that a demand curve, which shows for an individual or a whole market the relationship between the willingness to pay and the quantity bought, is normally downward-sloping. This is shown in Figures 1.3a and 1.3b. Figure 1.3a shows an individual demand curve, where the individual consumer surplus (at a price level of 5) is presented by the shaded area. The individual demand curves add up to a market demand curve. The collective consumer surplus (at a market price of 5) is presented by the shaded area in Figure 1.3b.
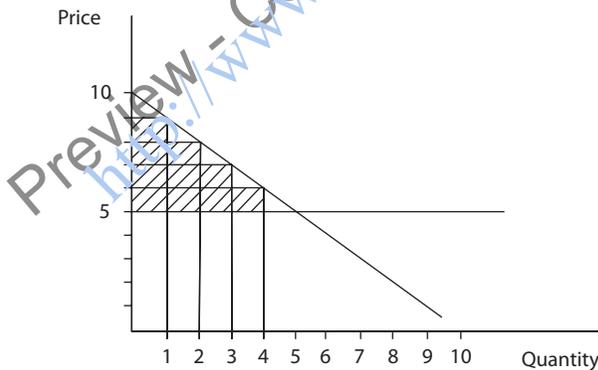

Figure 1.3a Individual demand curve

---

[18] T. Scitovsky, *Welfare and Competition* (London: Unwin University Books, 1952). For the term efficiency, see also n 4.

[19] For an explanation of these terms, see paras 1.35 and 1.63–1.64.

[20] For a more detailed exposition, see eg H. Varian, *Microeconomic Analysis* (3rd edn, London: W. W. Norton, 1992).
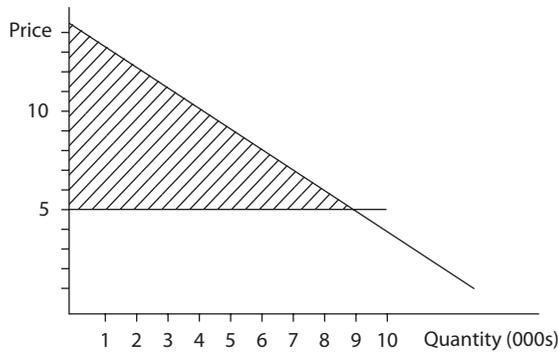
Figure 1.3b Market demand curve

(b) Production Costs

1.29   Production costs of a company can be represented as curves. These cost curves are, of course, not the same for different companies and different industries. Some firms are capital-intensive while others are labour-intensive, some have high fixed costs while others have high variable costs, some experience economies of scale while others have flat cost curves or even experience diseconomies of scale. However, there are some general characteristics to cost curves.

1.30   These general characteristics depend very much on whether one looks at the short or long run. In the short run, many production factors may be fixed, that is the producer is not able to vary the quantity of these factors used in response to demand changes. This is usually true for the buildings and other main capital goods and the production process adopted. But it may also be true for labour, at least in a downward sense when rules on firing make adaptation difficult and slow, and sometimes in an upward sense when, for example, training for specific capabilities takes a long time. Other inputs like raw materials, intermediate goods, and energy are often variable. In the long run, all factors become variable as plants, production processes, and personnel (including management) can be totally replaced.

(c) Short-Run Production Costs

1.31   The general characteristics of the short-run cost curves are best explained by what economists call the law of increasing and decreasing returns. Let us assume for the moment that we have only two factors of production, capital and labour. The former is fixed while the latter is variable.[21]

1.32   To produce, a company must employ labour to work with the available fixed capital. At first, employing more labour will lead to a more efficient use of capital (by increasing the utilization rate) and of labour, for example through specialization. If by adding an employee the average productivity per employee rises, the returns are increasing. In other words, the marginal productivity, that is, the change in total output resulting from the use of one more employee, is increasing. This means that the costs of producing a unit of output are decreasing. This is so for the average total cost (ATC), that is, all fixed and variable costs divided by

---

[21]   Variable costs as defined here are the same as avoidable costs; the costs that can be avoided by not producing the additional unit/that particular range of output.

total output, as well as for the average variable cost (AVC), that is, all variable cost divided by total output. It is also true for marginal cost (MC), that is, the cost of producing the last unit of output.

With the fixed capital as a constraint there comes a point where adding another employee will lead to less extra output when compared to adding the penultimate employee. The marginal productivity is declining and the returns start to decrease. The moment the marginal productivity starts to decline, the marginal cost starts to increase: producing one more unit of output becomes more expensive than the previous unit of output in terms of employee time used.[22] By adding more employees, the marginal cost will rise further and will cut the average variable and average total cost curves at their lowest point, as depicted in Figure 1.4. **1.33**
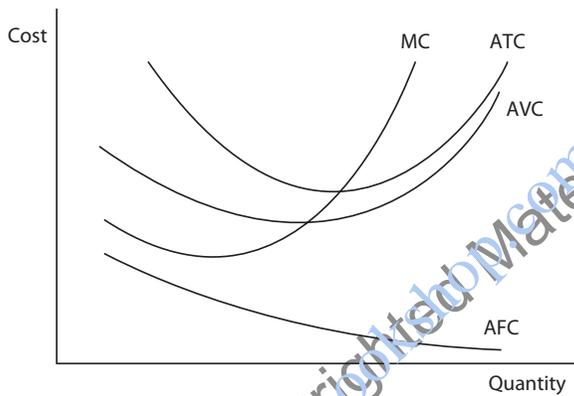
Figure 1.4 A company's short-run cost curves ATC, AVC, AFC, and MC

That the MC curve cuts the other two curves at their minimum is easily explained: when the extra costs incurred by producing one more unit of output are still lower, respectively, than the average variable cost or the average total cost, producing this extra output will further sink these averages. However, the moment that producing this extra unit has marginal costs that are higher than the respective average, the average will start to rise. **1.34**

In Figure 1.4, also, the average fixed cost (AFC) curve is depicted. This average will decline as long as output grows, as the fixed costs are spread over more units of output. In Figure 1.4, the cost curves are only drawn insofar as it is economically interesting. That means not too far left or right from the minimum of ATC. The further away from this minimum, the less efficient the company produces. At its minimum the company reaches productive efficiency.[23] **1.35**

(d) Profit Maximization

What range of the cost curves is economically interesting is linked to the goal of the company. Usually it is assumed that this goal is profit maximization. Certainly, in a competitive environment where profits are under pressure, a company is best advised to try to maximize its profits in order to survive in the long run. In a situation of fierce competition, profits will **1.36**

---

[22] It is assumed that the price of the production factor, in this example the wage rate, is constant and not influenced by the quantity demanded by the company.

[23] It is not relevant to see what happens if more and more employees are added to the fixed capital, making the average costs rise further and further and eventually leading to a decline in output. Nor is it interesting to see what happens when the company produces far below its optimal scale.

be rather low, just high enough to attract the required production effort, and a deviation from profit maximization will quickly lead to losses. In general, it is only when a company has a certain degree of market power that it can afford to pursue other goals such as sales maximization with a minimum profit constraint.[24]

1.37 To maximize its profits or, when times are bad, to minimize its losses, a company should produce up to the point where the additional costs of producing one extra unit of output are still covered by the additional revenue earned by this extra unit of output: producing less would mean that marginal cost is smaller than marginal revenue (MR) (MC is below MR), indicating that producing an extra unit of output will make it earn more. Producing more than the amount for which marginal cost equals marginal revenue would mean that marginal cost is higher than marginal revenue, indicating that by reducing output it will earn more. To maximize profits, therefore, the marginal cost should equal the marginal revenue (MC = MR). This rule holds good for companies with or without market power.

1.38 The MR curve will depend on the demand curve the company is facing. When the company operates in a perfectly competitive market it is a price taker: its output has no influence on the price in the market. If it raises its price above the market price, demand for its product will drop to zero. Its marginal revenue equals the market price. Graphically, this means the MR curve is a horizontal line at the level of the market price. In that situation, the MC curve represents the supply curve of the profit-maximizing firm: at each price level, the MC curve indicates the supply of a given firm (MC = MR = p).

1.39 If, on the other hand, the company faces a downward-sloping demand curve, meaning that by varying its output it can change the price at which it can sell, the MR curve will lie beneath the demand curve. Given that the demand curve is downward-sloping, the company has to lower its price if it wants to sell more units of output. This price decrease applies not only to the additional sales but to all its sales.[25] As a result, the additional revenue following the expansion of output is lower than the price at which the expansion takes place.

1.40 Let us assume for the moment that the company is a price taker. In Figure 1.5 this means that as long as the market price is below p1 the company is better advised not to produce at all: the price does not even cover the average variable costs. With a price above p1 the profit-maximizing company will produce the amount where its marginal cost (MC) equals the price (which is equal to MR). With a market price between p1 and p2, the company is in fact minimizing its losses, as the price does not yet cover all average total costs. When the price rises above p2, the company will make a profit, as the price exceeds the average total costs. With a price of p3, the profit will be the shaded area ABCD.

(e) Long-Run Production Costs

1.41 It was stated in para 1.30 that the cost curves depend very much on whether the short or long run is analysed. In the short run, the law of increasing and decreasing returns indicates that the ATC, AVC, and MC curves will first decline and then increase. An area where average costs are constant over a certain range of output is possible, but inevitably the cost curves will

---

[24] Whether a company with market power actually will deviate from the goal of profit maximization will depend on the incentives of management, the control of ownership over management, and in general the restraining influence of the capital markets. See also n 9.

[25] If there is no price discrimination.
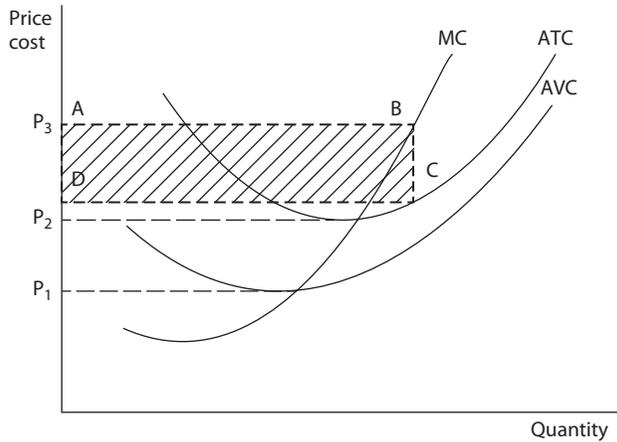
Figure 1.5 A company's cost curves and profit

rise as production is further increased. The optimal capacity utilization (the output level at which average costs are minimized) will not vary much in the short-run.

In the long run, when the fixed production factors are also variable, the picture looks different. If a company producing at its minimum short-run average total cost would like to double its output, it could do so by duplicating the existing plant. This means that the long-run ATC curve will have a flat section. The long-run average total cost is therefore in general depicted as in Figure 1.6. In the same picture, different short-run ATC curves are drawn belonging to different output levels. The long-run ATC curve represents the lowest short-run average total cost achievable for every level of output.[26]   **1.42**
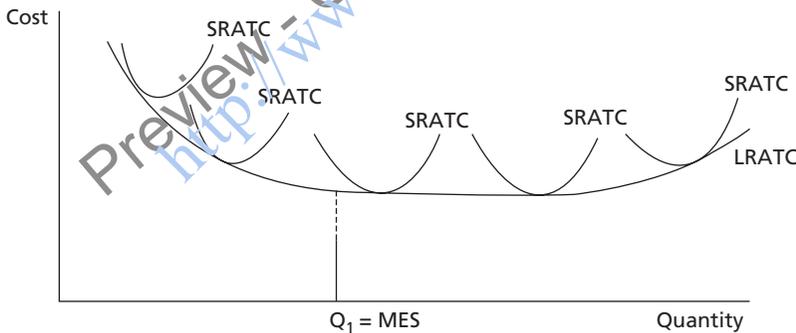
Figure 1.6 Long- and short-run ATC

(f) Economies of Scale and Minimum Efficient Scale

The long-run ATC curve drawn in Figure 1.6 illustrates two other important concepts: that of economies of scale and the minimum efficient scale (MES). In Figure 1.6 an output below   **1.43**

---

[26] The short-run costs are the real costs of a company, used eg when it has to calculate its profit or loss. The long-run costs in Figure 1.6 indicate the possibility frontier where the state of technology is assumed to be constant (static perspective); see Sections C.1 and D.1.

Q1 will be produced at higher average total cost than is attainable when more than Q1 is produced. Up to Q1, increasing capacity will lead to economies of scale: a higher capacity reduces the average costs. These economies of scale often result from the indivisibility ('lumpiness') of certain production factors: the bigger truck that transports more while still requiring only one driver, the bigger company that can afford to have a full-time specialist employed for every relevant area, the bigger plant that does not need to keep more spare parts in stock than the smaller plant. Economies of scale may also result from technical–physical relationships, such as the bigger oil tanker that requires relatively less steel to be built, or from economies of increased dimensions, such as the larger company that may obtain discounts when buying larger amounts of input or borrowing larger sums.[27] More generally, increased output brings a different, more efficient production process within reach.

1.44 Beyond Q1, no more economies of scale can be reaped. This point is called the minimum efficient scale. Although in practice not always easy to establish, it is an important concept helping to explain concentration in a market. The MES determines the maximum number of companies that can operate efficiently in a market, at least when producing below MES level results in significantly higher costs per unit of output. The extent to which producing below MES level results in higher costs is measured by the cost gradient, that is the steepness of the slope of the cost curve. For example, when the cost gradient is significant and the MES equals 10 per cent of total demand, there is room for at most ten efficient companies.

1.45 In the example of an MES of 10 per cent of market demand, it can also be expected that a company having capacity to produce 20 per cent of the market will be able to produce at the same low ATC. In theory, a company can have any size above MES and produce at the same low ATC. In order to produce more, its management could simply copy MES size units. In practice, however, it can be expected that above a certain size diseconomies of scale will also appear. Management may become too complex, the number of management layers will increase, and motivation may reduce. The long-run ATC may creep up when size continues to increase.

1.46 Economies of scale, especially when they are substantial, are an important explanation for concentration tendencies in a market. By indicating the maximum number of firms that can operate efficiently in the market, the MES determines the minimum concentration degree, at least in markets where products are relatively homogenous. As some companies will be above MES scale, the overall concentration in the market will usually be higher. This makes it quite clear that more companies in a market does not necessarily lead to a better market outcome as production may take place at sub-optimal levels and that protecting competitors is not the same as protecting competition or consumers' interests.

1.47 The economies of scale described in paras 1.43–1.46 are sometimes referred to as static internal economies of scale; internal because they are related to the plant or firm, static because they are not related to past production. Estimates of the importance of these static internal economies of scale depend to a degree on the method of measurement.[28] Econometric studies based on

---

[27] In the latter case, a distinction is made between real economies, based on actual cost savings on the side of the input producer/bank, and pecuniary economies that (merely) reflect a benefit at the expense of the input producer/bank resulting from a different balance of power.

[28] K. Junius, 'Economies of Scale: A Survey of the Empirical Literature', Kiel Working Paper No 813, Kiel Institute of World Economics (1997). See also European Commission, *The Single Market Review*, Subseries V, Vol 4: Economies of Scale (1997); and J. Stennek and F. Verboven, 'Merger Control and Enterprise

cross-section or time-series data on costs and profits tend to find only limited economies of scale. For example, Lyons found that in the UK for most of the 118 trades studied, MES was below 250 employees.[29] However, engineering estimates, that is, cost estimates by managers, engineers, etc, tend to give more weight to economies of scale. Surveys report important economies of scale in a number of capital-intensive sectors like motor vehicles, other means of transport, chemicals, machinery and instrument manufacturing, and paper and printing, that is, in particular in the production of industrial goods.[30]

In addition to static internal economies of scale, dynamic internal economies of scale are distinguished.[31] The latter refer to a lowering of the costs of production over time as a result of experience obtained on past cumulative output. They are also referred to as learning effects. In terms of Figure 1.6, these economies of scale lead to a downward shift of the long-run ATC curve. These economies of scale are not so much an explanation for concentration tendencies but may give rise to a first-mover advantage. The company or companies that entered the market first were possibly able to recoup the higher original costs while latecomers may have to sell immediately at lower prices dictated by the first entrants having gained some experience in the market. Such learning effects are more likely in new industries, especially when operating with a large amount of skilled labour, and less likely in mature industries with known technologies, especially when operating with a high level of fixed capital. **1.48**

A concept that is similar to, but distinct from, economies of scale is the concept of economies of scope. These economies refer to settings where the average total cost is reduced as a result of producing a larger product range. Economies of scope result if certain investments (to be) made for one product, for example electric razors, benefit the production and/or sales of an additional product, for example lady shaves. For instance, investments in R&D may provide results which are useful for a number of products and investments to establish a brand name may benefit the sales of various products sold under that brand name. The efficient use of such 'common costs' may create economies of scope. The existence of common costs will usually make the calculation of production costs per product more difficult, as the allocation of these common costs to the various products in the range introduces an element of arbitrariness. **1.49**

(g) Entry Barriers

As indicated in the previous paragraph, economies of scale are also an important element when describing another main concept of industrial economics, the concept of entry barriers. It was Bain who stressed the importance of entry barriers as a condition for companies with a significant market share to have market power and turn this into high (monopoly) **1.50**

Competitiveness: Empirical Analysis and Policy Recommendations' in *European Economy*, Reports and Studies No 5/2001, European Commission (2001).

   [29] B. Lyons, 'A New Measure of Minimum Efficient Plant Size in the UK Manufacturing Industry' (1980) 47 Economica 19.

   [30] C. Pratten, 'A Survey of the Economies of Scale', Economic Papers of the European Commission No 67 (1988); I. Gill and C. Goh, 'Scale Economies and Cities', World Bank Research Observer, Vol 25, No 2, 2010, 235–62. Gill and Goh mainly refer to older studies, reflecting that over the past years there has been little attention paid to the estimation of economies of scale.

   [31] The literature also distinguishes (static and dynamic) external economies of scale (see Junius, 'Economies of Scale' (n 28)). These refer to positive external effects resulting from firms being situated near each other. These economies play an important role in regional economics and trade theory. They are, however, less relevant from a competition policy perspective.

profits. Without entry barriers, easy entry would quickly eliminate such profits. Entry barriers, according to Bain, are 'the advantages of established sellers in an industry over potential entrant sellers, these advantages being reflected in the extent to which established sellers can persistently raise their prices above a competitive level without attracting new firms to enter the industry'.[32] In other words, the incumbent companies have certain advantages that allow them to increase their price above minimum ATC without attracting entry.

1.51　This definition of entry barriers is often used in competition policy as it indicates situations in which a competition concern may arise. In a market with entry barriers, further concentration through mergers may have to be stopped, especially when the incumbent firms already experience reduced competition. A competition authority will also have to be more alert to abuse of a dominant position in a case where a company with a high market share operates on a market shielded by entry barriers.

1.52　This definition of entry barriers, however, does not always give the right policy insights. When the question is raised whether a competition authority should stimulate or force entry in a particular market, another definition, first proposed by Stigler, is superior. He defined entry barriers as costs that new entrants have to bear, but which are not incurred by the incumbents.[33]

1.53　The difference with Bain's definition is most easily explained by the example of economies of scale. Economies of scale qualify as an entry barrier under Bain's definition. As new companies in general enter at a small scale, they will experience a cost disadvantage compared to the incumbents. This will allow the latter, when competition between them is already reduced, to keep their price above their own minimum average total cost and earn high (monopoly) profits.[34] However, the incumbents were also faced with scale economies when they entered. In addition, new entrants may be able to enter at minimum efficient scale. Scale economies therefore do not qualify as an entry barrier under Stigler's definition. Forcing entry by the competition authority will be inefficient when it increases the number of companies above the number of companies that can efficiently operate in the market, that is, when the incumbents are not much bigger than minimum efficient scale.

1.54　In addition to economies of scale, a number of other factors are sometimes mentioned in competition policy analysis as entry barriers, although these may not always qualify as such under Stigler's definition. Government regulations, especially when establishing exclusive rights, may work as an entry barrier, for example when only a limited number of licences are provided and State aid, when only available to incumbents, will work as an entry barrier. Import tariffs have the same effect on foreign suppliers. Intellectual property rights or ownership of absolutely scarce resources (eg platinum mines) may also inhibit access by those that cannot avail themselves of these patents or scarce resources. An essential facility, defined as a facility access to which is indispensable to be able to produce another good or service (eg the

---

[32] See J. Bain, *Barriers to New Competition* (Cambridge, MA: Harvard University Press, 1965), 3.

[33] G. J. Stigler, 'Barriers to Entry, Economies of Scale and Firm Size' in R. D. Irwin, *The Organization of Industry* (Chicago: Homewood, 1968). For a review of the various definitions of entry barrier, see R. P. McAfee, H. M. Mialon, and M. A. Williams, 'What is a Barrier to Entry?' (2004) 94(2) Am Econ Rev 461; and D. W. Carlton 'Barriers to Entry' in W. D. Collins (ed), *Issues in Competition Law and Policy* (Chicago: ABA Section of Antitrust Law, 2008).

[34] Where the entrant considers entry at large scale, it will seriously have to estimate the influence of its additional output on the market price. If it expects the price to drop to competitive levels entry may not be attractive.

railway track and the railway service), may work as an entry barrier if access to the facility is not open to competitors. Vertical links or vertical integration may make access more difficult and foreclose potential competitors. Economies of scope, that is, lower average total cost as a result of producing a larger product range, may also make entry more difficult. The same can be said of brand loyalty of customers, for example stimulated by high advertising outlays, as it makes customers less willing to switch to comparable or better offers. More generally, when a customer has to bear a high cost in order to switch to a new supplier, such switching costs may hinder entry of new suppliers. It should be added that many of these factors may work not only as an entry barrier but also as a barrier to expansion, preventing companies already in the market from expanding their output.

The question whether certain of these factors should be described as entry barriers partly depends on whether the necessary outlays are sunk costs. Sunk costs are those costs that have to be made to enter or be active on a market but that are lost when the market is exited. Advertising costs to build consumer loyalty will work as an entry barrier if an exiting firm cannot sell its brand name or use it somewhere else without incurring a loss. The more costs are sunk, the more potential entrants will have to weigh the risks of entering the market and the more credibly incumbents can threaten that they will match new competition as they will not leave the market.[35] High sunk costs invested in excess capacity may be an especially credible threat that the incumbent(s) cannot leave the market and will increase output and lower prices upon entry. **1.55**

(h) Contestability

It was Baumol, Panzar, and Willig who stressed the importance of sunk costs with their theory of contestable markets in the early 1980s.[36] A market is said to be contestable if there are no sunk costs or other entry barriers and consumers are willing to switch quickly, before incumbents can react, to the better offer of new entrants. Under these conditions, so-called hit-and-run entry is possible. When the incumbents charge a price above minimum average total cost, it becomes profitable to enter and to stay in the market for at least the time it takes before the incumbents lower their prices. The threat of such hit-and-run entry, in other words the existence of potential competition, will discipline the incumbents, even when they have very high market shares. **1.56**

At a conceptual level, the theory of contestable markets helped to underline and delineate the possible role of potential competition. In practice, not many markets are truly contestable. The important question is the degree to which markets are contestable. In general, entry requires sunk costs, sometimes minor and sometimes major, and incumbents are often in the position to react quickly, that is, before consumer loyalty wears down. Even in transport markets, where it is possible in theory to redirect assets, such as ships or planes, at short notice from one route to another, other entry barriers like the non-availability of necessary slots may delay or impede entry. Actual competition is therefore still to be preferred above potential competition. **1.57**

---

[35] This commitment element also applies to the entrant after market entry. The difference is that the incumbent is already in the market whereas the entrant still has to decide whether to enter.

[36] W. J. Baumol, J. C. Panzar, and R. D. Willig, *Contestable Markets and the Theory of Industry Structure* (New York: Harcourt Brace Jovanovich, 1982).

(3) Perfect Competition

(a) The Model

1.58   When market models are put on a market-power scale, perfect competition is the extreme at the low end. There is no company that holds market power and competition policy enforcers can quietly write books during office hours. Unfortunately, markets rarely fulfil the conditions of this model. However, the model is useful for two reasons. First, it highlights two very important welfare economic concepts, of allocative and productive efficiency. Secondly, in certain respects, it is useful as a *benchmark* against which to measure the competitiveness of actual markets.

1.59   In order to be called perfectly competitive, a market must have a number of characteristics, of which the following are the main ones: there must be many suppliers and many buyers, there are no entry barriers, the product is homogeneous, and there is full transparency. This means that the MES must be small compared to total market demand, so that many companies are able to operate in the market and produce at minimal costs. The condition of transparency means that suppliers and potential suppliers are aware of every change in demand and price and, as there are no entry barriers, swiftly react by expanding or reducing supply. The condition also implies that companies are aware of the most efficient production techniques and that no company is more efficient than the others.

1.60   A company operating under such conditions will be a price taker, as briefly indicated in the previous section. The price is determined by the market and a company's own output is so small compared to total output that a change in the company's output has no perceptible influence on the market price. As entry and exit are swift and without costs, the market will always quickly return to its equilibrium where the price exactly matches market demand and market supply, as shown on the right-hand side of Figure 1.7. If demand rises—graphically, this means the demand curve shifting to the right—the price will rise as the current output is not able to satisfy all demand. Entry of new firms or expansion of existing firms will immediately increase output until the equilibrium price is restored. A fall in demand—graphically, the demand curve shifting to the left—leads to firms reducing output or leaving the market until market output is sufficiently reduced and equilibrium restored.

1.61   At the equilibrium market price, every company in the market will produce at the same minimum average total cost and will make no profits. This is shown on the left-hand side of Figure 1.7. By 'no profits' it is meant that the company's income is just enough to cover the rewards that all factors of production, including capital, need to obtain in order to make them stay in this company. In economic terminology, they receive their opportunity cost (the money they would make elsewhere, ie on other markets), but no more. In other words, the situation of no profits allows for normal accounting profits that are necessary to make capital stay in the company. These normal profits are part of the ATC cost curve. However, no excess profits are made.

(b) The Outcome

1.62   Figure 1.7 deserves some further explanation as it shows a number of important issues. First, there is the difference between the market demand curve and the company's demand curve, that is, the demand the company faces for its own output. The market demand curve is downward-sloping, as explained in para 1.27 (see Figure 1.3b). The company's demand curve is (practically) horizontal at the level of the market price. At that price, the company,
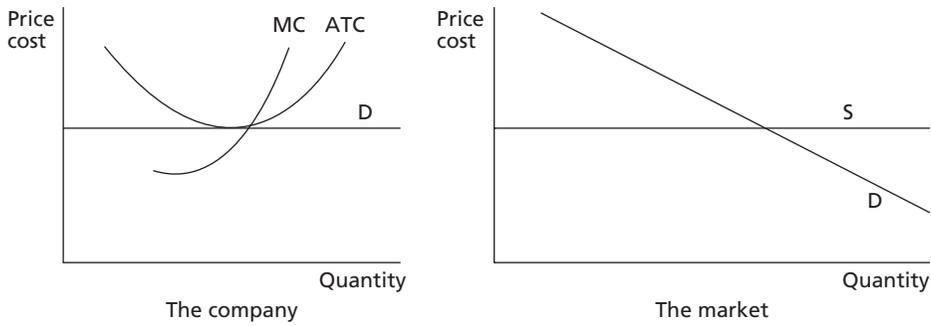
Figure 1.7 Perfect competition

given its small capacity, can sell as much as it wants. It does not need to lower its price to sell more. It is also not in a position to lower its price since that would lead to immediate losses as the price would go below marginal cost as well as average total cost. Increasing its price above the market price would lead to an immediate loss of all its sales and imply the company's exit from the market.

Secondly, Figure 1.7 highlights that in perfect competition there is *productive efficiency*: with given resources the maximum output is produced. This results from every company producing at the minimum average total cost. If a company is less efficient, it will make a loss and exit the market, in which a new efficient entrant will take its place. If a firm introduces a new cost-saving technique, this will be copied immediately by all the others, graphically represented by a downward shift of the supply curve, after which a new equilibrium will be realized at a lower price. **1.63**

Thirdly, Figure 1.7 indicates that in the equilibrium situation there is *allocative efficiency*: welfare is maximized.[37] If less output is produced than the market equilibrium quantity, welfare will be lower, because there will be buyers willing to pay more than the equilibrium price but who are not served. That means these buyers would be willing to pay more than it costs to produce more units and welfare could thus be increased by expanding output. Expanding output beyond the equilibrium would also lower welfare as the cost per unit would exceed the lower price level that would need to be set to sell the extra output; in other words, the extra costs would now exceed the willingness to pay of the marginal consumer. Productive resources are used at the wrong place: elsewhere, that is on other markets, they could be used to produce goods for which there is a higher willingness to pay. The allocative efficiency is reflected at company level by every company obtaining a price equal to its marginal costs (P = MC). **1.64**

(4) Monopoly

(a) The Model

Monopoly is at the other extreme of the market-power scale. In the fully-fledged monopoly model, the monopolist has the maximum achievable market power. One might expect that competition policy enforcers, when such a situation occurs, have to give up the possibility of writing books during office hours. However, this may not be the case as the analysis of pure **1.65**

---

[37] See also n 4.

monopoly situations is rather straightforward and markets rarely fulfil the conditions of this model. The model of monopoly is, however, very useful as it helps to highlight a number of important concepts and it provides the clearest example of what competition policy tries to prevent or remedy.[38]

1.66 In order to be called purely monopolistic a market must have a number of characteristics, the main ones being that there is only one supplier while there are many buyers and that there are entry barriers that practically prevent entry.

1.67 A company operating under such conditions will be a price setter. As it is the only supplier in the market, market demand is the demand for the company's product. By varying its output, the monopolist can determine the market price along the demand curve, which in a way is its only constraint. As entry is impossible, it can quietly try to maximize its profits or pursue other goals.

(b) The Outcome

1.68 Assuming that profit maximization is the monopolist's goal, it will produce that output where its marginal revenues equal its marginal costs (see para 1.37). In Figure 1.8 this is at quantity Qm. With a demand curve that is downward-sloping, its marginal revenue curve will also slope downwards and lie beneath the demand curve. The reason is simple. When the monopolist wants to sell an extra unit of output it has to lower the price somewhat. When price discrimination is assumed impossible, the monopolist has to lower the price not only for this last unit but for all units it wants to sell. This means that the marginal revenue at a particular output is the new price minus the cumulative price loss it has to take on all other units.[39] In Figure 1.8 it is further assumed that average total cost and marginal cost are constant (the ATC and MC curves of the monopolist are horizontal), that is, there are no fixed costs and no economies of scale. This assumption simplifies the drawings without changing the principal outcome of the model.
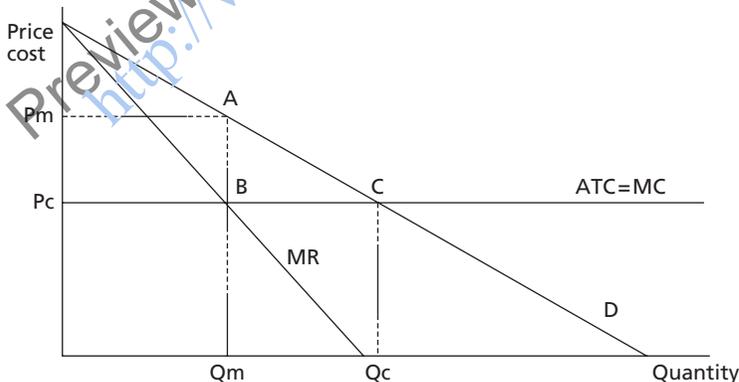

Figure 1.8 Pure monopoly

---

[38] This is not to say that competition policy is only concerned with static monopoly or market power. The dynamic point of view is also important. See esp Sections D and F.

[39] See also para 1.39.

Figure 1.8 clearly shows the main disadvantages of monopoly. The monopolist sells output Qm, which is less than the output Qc under competition. As a result, the price the consumers have to pay is higher: Pm compared to Pc. This has two main welfare effects. First, there is a loss of welfare as consumers acquire fewer products than before. The area ABC is what is generally called the dead-weight welfare loss of monopoly. It is 'dead-weight' in the sense that the consumer surplus is really lost: it is not acquired by anyone in the economy. Secondly, there is a transfer of income from consumers to monopolist. The monopolist makes a profit of PmABPc. This amount used to be consumer surplus, but with the higher price the consumers have to pay it is turned into profits for the monopolist. **1.69**

It can be debated whether the monopolist's profit should be counted as a welfare loss. One could argue that the transfer of income from consumers to monopolist does not change society's welfare as a whole, as some gain what many lose.[40] However, for a competition authority the case may be quite straightforward. First, there is a clear allocative inefficiency (the dead-weight loss referred to in the previous paragraph). As the monopoly price Pm is higher than the marginal costs, welfare could be increased by producing extra units. The consumers are willing to pay more for these units than it would actually cost to produce them. Secondly, insofar as the goal of competition policy is stated in terms of protection of competition to further the interests of the consumer (as is the case in most jurisdictions), there can be no doubt that monopoly profits—in particular, where they persist—must be seen as something negative which competition policy should try to avoid. **1.70**

Another question is whether the monopolist is technically efficient. In the example of Figure 1.8, the answer is 'yes'. The monopolist is producing at minimum ATC. But there are good reasons to believe a monopolist may not always be so efficient. Not feeling the heat of competition, the company may become slow and inefficient. Slack eats away part of the possible monopoly profits. Taking life easy instead of profit maximization may have become important, especially when the owners (shareholders) do not exercise effective control. It was Leibenstein who coined the phrase 'X-inefficiency', meaning internal inefficiency in the form of too high salaries, excessive corporate jets, a surplus of employees, etc. That this leads to an additional welfare loss is shown in Figure 1.9. **1.71**
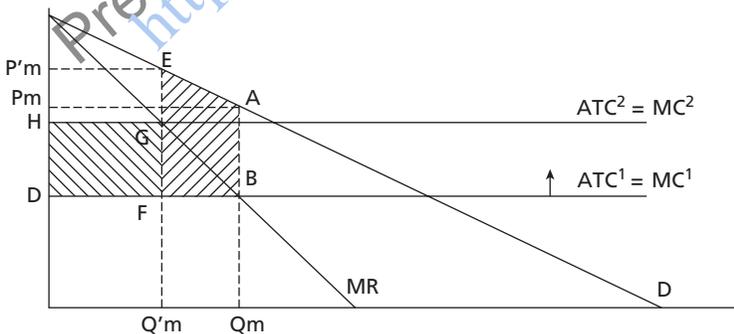
Figure 1.9 A monopolist with X-inefficiency

---

[40] See Bork, *The Antitrust Paradox* (n 4), ch 5. Bork and other scholars associated with the Chicago School held that the aim of antitrust policy should be to advance total welfare (the sum of consumer surplus and producer surplus), arguing that also shareholders are, in the end, consumers. Accordingly, the fact that monopoly prices entail a transfer from consumers to shareholders should not be a cause for illegality under the antitrust laws, only the fact that monopoly prices give rise to a dead-weight loss (reduction in output).

1.72   The X-inefficiency is reflected in higher ATC and higher MC curves. This results in a new equilibrium with the lower quantity Q'm and the higher price P'm. The consumer is paying for the higher costs with a higher price. There is an extra dead-weight welfare loss of EABF. In addition, productive factors of the area HGFD are lost to society, as what was previously monopoly profit has now been used to produce inefficiently.

1.73   A last welfare loss caused by monopoly could be named 'the price of success'. A monopoly position is very attractive and many resources may be wasted both by those who defend it and those who attack it. Those who defend it may try to erect and maintain entry barriers by keeping excess capacity, by excessive product differentiation, by political lobbying, by starting entry-delaying lawsuits, etc. Those who attack the monopoly position have to spend resources to overcome these barriers. In theory, all monopoly profits could be wasted in the struggle for a share of the pie.

1.74   Not everyone will recognize these costs as a welfare loss. When competition is seen as rivalry—a fight for temporary advantages, a struggle to gain market power before being overtaken by the next wave of competition—at least part of these costs may be seen as the necessary price to be paid for vigorous competition. However, most competition authorities will, for example, be rather suspicious when finding dominant companies running up costs to maintain excess capacity, in an attempt to keep competitors out of the market.

1.75   Monopoly may not only have negative effects but may also lead to certain advantages for consumers. First, it may be the case that economies of scale require such a size that only one company in the market can produce at minimal cost. This is what is called a natural monopoly.[41] Producing with more companies would necessarily lead to inefficient production. The dead-weight loss and the price asked by the monopolist may compare favourably with the welfare loss due to higher costs and the price level asked under competition. If protecting consumer welfare is the aim, then the only relevant question in such a static analysis is whether the price asked under monopoly is higher or lower than the price asked under competition. If the protection of total welfare is the aim, it would be preferable to have a monopoly not only when it would lead to a lower price, but also if it would lead to a higher price but the resulting production efficiency gains outweigh the dead-weight welfare loss created by the higher price.

1.76   A second possible positive effect of monopoly, which may be relevant in the context of innovation and patents, is that the prospect of monopoly profits may spark more effort on the part of companies to invest in innovation. The trade-offs in this area are considered in more detail in Section D.

1.77   A third possible positive effect has less relevance for competition policy. Some authors argue that the lower output and higher price of monopoly may counterbalance certain negative externalities in production or consumption. Less consumption of environmentally unfriendly products and less use of limited natural resources might actually increase welfare.

---

[41]   In the case where not even a monopolist has sufficient scale to produce at MES level, this has consequences for productive and allocative efficiency. Productive inefficiency will persist until demand grows and allows attainment of MES size. Similarly, as long as ATC is falling over the relevant output range, pricing at marginal cost, ie allocative efficiency, would result in a price below ATC and would thus result in overall losses, which would need to be recovered with the help of, eg, general taxation or two-part tariffs (composed of a fixed fee for the right to use or consume and a (low) marginal tariff based on actual usage or consumption).

(5) **Oligopoly**

(a) Introduction

The models of monopoly and perfect competition may be on opposite sides of the market-power scale, but they are remarkably similar in their emphasis on market structure and neglect of company behaviour. The only behavioural assumption that is introduced is that companies are profit maximizers. The C of the S-C-P paradigm can be ignored as the models are quite straightforward; the market structure leads *linea recta* to a specific market performance.

1.78

This is not the case for oligopoly, the most important intermediate market form on the market-power scale. Oligopoly is the market structure in which there are a few suppliers, at least two, while the maximum number of companies is not clearly determined. The main characteristic is that the companies in such a market realize or believe that their individual behaviour concerning output, price, etc has a perceptible influence on the market outcome and therefore may provoke reactions from the side of competitors. In the fisheries sector, a fisherman rightly ignores the influence of his catch on the market price of fish. In the oligopolistic car market, a large manufacturer cannot and will not ignore the impact its decisions have on the market and on its competitors and vice versa. This means that the C of the S-C-P scheme becomes more important in oligopolistic markets. It also means that competition issues in such markets are rather complicated. Competition policy enforcers faced with such markets are forced to write their books in the evening.

1.79

Oligopolistic markets are difficult to analyse. The outcome of oligopolistic behaviour can vary to such an extent that one of the more popular statements is that 'with oligopoly, anything goes'. The market price may be as low as under perfect competition or as high as under pure monopoly or anywhere between these two. The economic models of oligopoly reflect well this complexity. The outcome of oligopoly models is often highly specific to the exact assumptions used in the model. It is therefore important to identify and analyse the model specifications that best fit the actual market conditions. But even then the economic models of oligopoly may leave a wide range of possible outcomes.

1.80

This does not mean that competition policy has no function in oligopolistic markets. Experience shows that anti-competitive outcomes can certainly arise in such markets and that many markets are oligopolistic; they should probably therefore be the focus of competition policy. However, given the complexity of these markets and the limited guidance offered by economic models, the ambitions of the competition enforcer should be modest. Oligopoly cases are the clearest example of what was said in the introduction: that competition cases are concerned with identifying the most plausible story or explanation of the market outcome. A good part of this story will consist of analysing the factors that either enhance or decrease the scope for collusion and anti-competitive outcomes, and choosing the model and specifications that best fit the actual market conditions. Empirical verification is in this context of great importance.

1.81

In the limited space of this chapter, no more than a brief introduction to oligopoly theory can be provided.[42] The literature on oligopoly is vast and sometimes very technical. For a

1.82

---

[42] For more extensive introductions, see eg Scherer and Ross, *Industrial Market Structure and Economic Performance* (n 1) or Church and Ware, *Industrial Organization* (n 1); or, more technical, Tirole, *The Theory of Industrial Organization* (n 1).

layman it may be disappointing to see that the various models are often not very helpful for answering concrete policy questions. The models do not answer the questions of which market conditions will lead companies in an oligopolistic market to compete fiercely on all the important parameters (price, quality, and innovation) and of when competition will be replaced, on one or all parameters, by collusive behaviour. There is no super model that includes all possible relevant factors. Most models concentrate on the effects and interaction of a limited number of factors, abstracting away from more realistic settings. However, the literature provides useful insights in the main conditions relevant for anti-competitive effects to arise.

(b) Game Theory

1.83 Most advances in oligopoly theory have been made since World War II by using game theory, especially non-cooperative games.[43] Game theory studies situations of strategic interaction using mathematical models. A game theory model specifies the players in a game (eg firms in a market or individuals in an organization), the information they have (or do not have), the actions they can choose, the timing of these actions, and the pay-offs for each player that result from actions. In such a model, each player is assumed to choose a strategy (a plan of action) that maximizes his pay-offs based on the information available to him and his expectations about rivals' actions.

1.84 The main idea behind non-cooperative games, as opposed to cooperative games, is that the parties cannot make binding agreements. A non-cooperative game setting seems to be the appropriate framework to apply as competition rules make anti-competitive agreements unenforceable in court. Cartel members may make agreements, but these are not binding.

1.85 In non-cooperative game theory, fully rational oligopolistic behaviour requires an assessment of the potential actions of competitors. That is, the oligopolists take account of the interdependence of strategies. An equilibrium will therefore only exist when the decisions of companies lead to a 'self-reinforcing set of strategies in which each strategy is a best response to the other strategies'.[44] Such an equilibrium is called a Nash equilibrium, that is, 'a set of actions is in Nash equilibrium if, given the actions of its rivals, a firm cannot increase its own profit by choosing an action other than its equilibrium action'.[45] In other words, the game finds a stable outcome once every oligopolist sticks to its chosen strategy, for example concerning the price it sets for its own product, in light of the strategies chosen by the other oligopolists.

1.86 Game theory has been applied extensively to study the strategic behaviour of companies in oligopolistic markets. Game models with multiple stages are, for instance, appropriate to study situations in which a company, usually the incumbent, has a first-mover advantage over other market players. Such analysis has been applied to examine the scope for entry-deterrence strategies, such as creating excess capacity or product proliferation (introducing products in the market to deter entry, not because it is in itself profitable), as well as limit pricing (pricing low to signal that market conditions are not favourable for

---

[43] Game theory as a formal theoretical analysis started with the book by the mathematician John von Neumann and the economist Oskar Morgenstern, *Theory of Games and Economic Behaviour* (Princeton: Princeton University Press, 1944).

[44] D. Yao and S. DeSanti, 'Game Theory and the Legal Analysis of Tacit Collusion' (1993) The Antitrust Bulletin, 113–41.

[45] Tirole, *The Theory of Industrial Organization* (n 1), 206.

entry).[46] Also vertical restraints have been studied extensively for their impact on competition, by focusing on the strategic value that such restraints may provide.[47] For example, economic models have shown that single branding/exclusive dealing contracts may be tools for foreclosing markets, in particular when they render the anti-competitive strategy (foreclosure) more credible and time-consistent.[48] Similarly, delegating pricing decisions to exclusive distributors may allow producers credibly to commit to less competitive behaviour towards each other (to 'soften' competition), because exclusive distributors have different pricing incentives than distributors facing intra-brand competition.[49]

As noted in para 1.80, the outcomes of the models of strategic interaction tend to rely heavily upon the precise modelling assumptions. One of the important assumptions in this respect concerns the way in which companies are thought to compete. Two stylized modes of competition are often considered: Bertrand competition (price competition) and Cournot competition (quantity competition).[50] **1.87**

Under Cournot competition, it is assumed that each company in the market decides on its profit-maximizing output assuming the others' output will remain unchanged. In the (Nash) equilibrium, each company chooses a level of output that is optimal (profit-maximizing) in view of what the other market players produce. The equilibrium of this model features a market price below the monopoly level but (well) above marginal cost (the benchmark of perfect competition). **1.88**

Competition in output is often identified with situations where output or capacity decisions are the main drivers of the price level in the market. Conceptually, firms choose output or capacity and then, given the level of demand, adjust prices to sell this output. This might apply, for instance, to certain basic commodity industries, where price levels are primarily determined by the overall level of output in the market, but also to a variety of other markets, such as those for package holidays, hotel accommodation, and office space. In markets where output or capacity decisions are the most important strategic decisions of the firms, the important concern for firms is how their output decision influences market prices. **1.89**

Under Bertrand competition, it is assumed that each company in the market decides on its price assuming that other prices in the market will remain unchanged. In the corresponding Nash equilibrium, each company chooses a price level that maximizes profit in view of what the other market players charge. The equilibrium of this model features a market price equal **1.90**

---

[46] For an overview and discussion of such strategies, see Scherer and Ross, *Industrial Market Structure and Economic Performance* (n 1), ch 10; and Tirole, *The Theory of Industrial Organization* (n 1), ch 8.

[47] For an overview, see eg M. Waterson 'Vertical Integration and Vertical Restraints' in T. Jenkinson (ed), *Readings in Microeconomics* (Oxford: Oxford University Press, 1996); P. Rey and T. Vergé, 'Economics of Vertical Restraints' in P. Buccirossi (ed), *Handbook of Antitrust Economics* (Cambridge, MA: MIT Press, 2008), 353–91; or V. Verouden, 'Vertical Agreements: Motivation and Impact' in Collins, *Issues in Competition Law and Policy* (n 33).

[48] cf P. Aghion and P. Bolton, 'Contracts as a Barrier to Entry' (1987) 77 J Econ Theory 388.

[49] cf P. Rey and J. Stiglitz, 'The Role of Exclusive Territories in Producers' Competition' (1995) 26 RAND J Econ 431. Exclusive distribution provides retailers with a certain measure of market power and leads to higher retail margins, but at the same time sales volumes will react less strongly if the producers raise their wholesale prices as an increase in wholesale prices will be (partly) absorbed by the retailers' margin. This may make producers inclined to use exclusive distribution and carry through price increases, amounting to a 'softening of competition'.

[50] The two modes of competition are identified with the nineteenth-century economists Bertrand and Cournot, respectively.

to marginal cost in the case of homogeneous products[51] and a price that is higher in the case of differentiated products.

1.91 Competition in prices often refers to situations where firms set prices and adjust their production levels according to demand. Competition in markets for consumer products and for capital goods can often be characterized in this way. In many such markets, capacity and output levels are less determinative of the eventual market outcome. Rather, factors such as product differentiation (the products offered differ in the eyes of buyers with respect to one or more important parameters) provide each firm a certain margin of manoeuvre in its price-setting behaviour.[52]

1.92 This distinction in types of competition is, of course, stylized: there will be many cases where the type of competition cannot be characterized as being one or the other. Even in markets where 'output drives price' or where competition is 'mostly on price', it is not to be taken for granted that competition is Bertrand or Cournot. Most markets feature business decisions that go well beyond a choice of price or a choice of quantity at a given point in time. For instance, markets in practice may turn largely on innovation (both product and process innovation) or on building consumer loyalty. Nonetheless, it may be useful to consider the generic market types and to distil some key factors to be examined in each of these. It is the purpose of the economic models to clarify what factors are the more relevant in precise market settings. The role of the investigator is to see to what extent any given model is useful in light of the facts of the case.

(c) The Scope for Collusion Illustrated with the Prisoner's Dilemma

1.93 Another important area in which game theory has played a role in clarifying the main issues is that of collusion in oligopolistic markets. Collusion and collusive behaviour are used in this chapter as in the economic literature, that is, as any situation in which market players do not compete 'to the fullest' but instead charge higher prices than they otherwise would, provided other firms in the market do so as well. It therefore includes not only explicit collusion in the form of agreements or concerted action but also tacit collusion, whereby market players refrain from adopting a more competitive attitude (eg in terms of price setting) as this would trigger a rational reaction or retaliation from its rivals in later periods.[53] The latter is what lawyers define as (conscious) parallel behaviour. Collusion in the economic sense is possible without communication between the companies involved. Economists thus define collusion in terms of effects. This stands in contrast to legal definitions of collusion, which are usually limited to agreements and concerted practices, stressing the possibility for competition rules to provide a remedy for the situation.

1.94 Within the non-cooperative game setting, the game that provides most insight into the difficulties and possibilities of collusion is the prisoner's dilemma game.[54] The original example

---

[51] If there are cost differences between the companies, the price will be equal to the marginal cost of the second-most efficient firm.

[52] Products may be differentiated in terms of, eg, technical specifications, quality, brand image, level of service, or geographic location. The presence of switching costs can also induce buyers to consider products to be differentiated, since they would have to incur costs in order to switch to a competitor's product.

[53] cf Tirole, *The Theory of Industrial Organization* (n 1), 207.

[54] Games with this structure were devised and discussed by Merrill Flood and Melvin Dresher in 1950 as part of their work for the Rand Corporation. The title 'prisoner's dilemma' and the version with prison sentences as pay-offs are due to Albert Tucker (*Stanford Encyclopedia of Philosophy* at <http://plato.stanford.edu/entries/prisoner-dilemma/>). Much of the theory of tacit coordination derives from the work of G. Stigler, 'A Theory

used to explain the game went along the following lines. A murder is committed, and two suspects are arrested. Not having enough evidence, the police need them to confess in order to have a conviction for murder. If one of the prisoners testifies against the other, the first goes free if the other has not testified against him, and the second goes to jail for ten years. If neither testifies, both get a sentence of one year only for illegal possession of firearms. Lastly, if both testify, both go to prison for seven years.

The structure of the game is presented in the diagram in Figure 1.10, commonly referred to **1.95** as the 'pay-off matrix'. In this case, the pay-off matrix illustrates all the possible outcomes or pay-offs for the two suspects (the first number in each cell of the matrix provides A's jail sentence in years, the second B's sentence).

Player B

		N	T	
	N	1,1	10,0	N = not testify
Player A				
	T	0,10	7,7	T = testify

Figure 1.10 The prisoner's dilemma game

It is clear upon close examination that the rational strategy for both suspects in this case is **1.96** to testify. If B does not testify, it is better for A to testify. If B testifies, it is also better for A to testify. The same applies for the choice faced by B: testifying is optimal for B regardless of what A does. In the jargon of game theory, to testify is the dominant strategy. The result is that both suspects go to jail for seven years as they end up testifying against each other. The problem is one of commitment: the collectively optimal outcome for the two suspects (each serving only the light sentence) is not attained because the suspects cannot make a binding agreement not to testify.

This analysis can easily be extended to the study of oligopolistic behaviour of companies. **1.97** Although oligopolists are normally not confined to choosing between two prices, two quantities, etc, it can be assumed that the basic choice is between competing and colluding. Instead of the decisions being 'not testify' and 'testify' they could be labelled 'cooperate' or 'defect' in relation to collusive behaviour in the market. The pay-off matrix would then have the structure represented in Figure 1.11. The first figure provides company A's profit, the second company B's profit.

In the situation of Figure 1.11, the dominant strategy for both companies is to defect, in **1.98** other words to compete. When the other company in the market will restrict output and thereby ensure a high market price, it is advantageous not to restrict output. Similarly, when the other company will not restrict output it is also against one's own interest to restrict output. As a result, the two companies will not cooperate and will forgo the collective optimal outcome and end up with the equilibrium with the lower collective profit. The latter is the Nash equilibrium of the prisoner's dilemma game.

---

of Oligopoly' (1964) 72(4) J Pol Econ 44–61. The economic literature on tacit coordination is relevant to all forms of coordination not based on legally enforceable contracts.

Company B

		C	D	
	C	3,3	1,4	C = cooperate
Company A				
	D	4,1	2,2	D = defect

Figure 1.11 The prisoner's dilemma applied to a duopoly

1.99 If all oligopolistic markets followed the simple rules of the prisoner's dilemma game, there would not be many competition problems. Even duopolists would compete with each other down to the competitive price level. The prisoner's dilemma shows the basic instability present in many situations of collusion. The collusive outcome creates the possibility to free ride or cheat on the cooperative behaviour of the others, as witnessed in practice by the breaking down and erosion of many cartel agreements.

1.100 However, competition policy practice and simulation experiments show that a collusive outcome is attainable. In practice, to cooperate does seem to be the dominant or chosen strategy in a not insignificant number of cases. This is explained by a number of factors.

1.101 The first factor that makes a collusive outcome more likely is that oligopolists usually do meet each other many times in the marketplace; the game is not played once but more than once. Intuitively this means that although the prisoner's dilemma pay-off structure may indicate that it is rational to compete if one only looks at one round, such competitive behaviour may spoil future profits that could possibly be attained by collusion. Past behaviour and possible future profits become important when formulating a strategy.

1.102 In game theory one usually distinguishes in this context between games that are infinite versus games that are played a finite number of times. In a prisoner's dilemma setting that is played an infinite number of rounds, the players might come to a collusive outcome.[55] Whether a collusive outcome results depends on the balance for each player of the gains from competing in the first period against the loss of a part of the collusive (monopoly) profit for every period or at least a number of periods thereafter. The incentive to compete will be weighed by each player against the possible punishment the other players may inflict on him in the future if he does not cooperate. Such punishment will in turn depend on the possibilities and rationality of punishing possible competitors. The punishment may consist in returning to the competitive outcome on the market because all firms expand their output. The players may also try to reduce the attractiveness of competing for all players by limiting the scope for possible undetected competition and/or increasing the possibilities of punishment.[56] The exchange of sensitive market information may be used by the players to help to detect competition.

1.103 Also when the game is played a limited and not an infinite number of times, a collusive outcome may result in a non-cooperative setting with a prisoner's dilemma pay-off matrix.

---

[55] Such an infinitely repeated single-period game is called a supergame.
[56] The punishment strategy that is chosen by the players influences the pay-off that results after the competing behaviour has been detected. As the question of the best punishment strategy seems still unresolved, this is not discussed further here.

Theoretically collusion becomes, however, more difficult. This is explained by backward induction. In a one-period prisoner's dilemma-type game the best strategy for each player is, as explained before, not to cooperate. This means that in a multi-period game it is rational for both players not to cooperate in the last period. Given the certainty that both will not cooperate in the last period, it is also not rational to cooperate in the penultimate period, as there can be no reward in terms of cooperation in the last period, etc. Thus collusion will not be achieved in any period.

However, as soon as the players do not have full information but instead have imperfect information and have to make up their minds about their best strategy under uncertainty— the common situation in real markets—collusion again becomes a possible outcome. Players may not know the number of times the game will be played, may have to guess about the costs and possibilities of the others to punish, may assign probabilities to the possible strategies of the others, etc. This may make it rational to cooperate, at least until someone starts to compete. **1.104**

Different strategies can be imagined in repeated games. A most successful strategy in simulation that is also very simple is the so-called tit-for-tat strategy: cooperate on the first round and thereafter do whatever the other player did in the previous round. It has the advantage of starting with a cooperative strategy in the first round to try to reap the gains of collusion. In addition, it provides a quick reaction by hitting back when competitive behaviour is detected. After such punishment, it offers the other the possibility to restore the collusive equilibrium. **1.105**

A second factor that makes a collusive outcome more likely is that companies, also in a setting of a non-cooperative game like the prisoner's dilemma, may behave more as if they are in a cooperative game setting. Companies in general do not behave as nakedly rational as non-cooperative game theory usually assumes. Social constraints, moral codes of conduct, etc do influence behaviour. Business ethics may 'command' that oral non-binding agreements are kept; 'a man a man, a word a word'.[57] **1.106**

This also means that communication on future prices and output, sometimes described as 'cheap talk' as it does not involve binding commitments and does not change the pay-off matrix, may not be all 'cheap talk'. To discuss and hammer out agreements detailing how much each will produce and what price will be charged may be quite vital as companies may become rather nervous about their cooperative attitude when there is not enough communication. Communication may be essential to 'prevent' companies from starting to behave as rationally as the underlying non-cooperative game assumes.[58] **1.107**

---

[57] See also paras 1.19–1.20. This does not necessarily mean that players act irrationally, it may mean that their perspective becomes less 'self-regarding' and more 'other regarding'. One could be 'other regarding' and have as a moral principle that one does not want to be the first to cheat. When applied rationally, this leads to collusion. See eg E. Fehr and K. M. Schmidt, 'A Theory of Fairness, Competition, and Cooperation' (1999) 114(3) Quarterly J Econ 817. Also, sometimes certain deliberations do not enter the firm's pay-off matrix but may nonetheless influence the firm's strategy; from a manager's point of view it may be very rational to avoid being the one who 'ruins' the market if that would isolate him on the green on Saturday. See, for further details, H. Pellikaan, *Anarchie, Staat en het Prisoner's Dilemma* (Delft: Eburon, 1996), ch 8.

[58] Interestingly, it appears that face-to-face meetings help collusion more than communication via email, see Armstrong and Huck, 'Behavioral Economics as Applied to Firms' (n 9), 11.

1.108 From experiments with the prisoner's dilemma it is known that the narrow 'self-regarding' perspective is in general not realistic. The experiments of Flood and Dresher[59] in the early 1950s already show this. In a 100-round prisoner's dilemma experiment they find that even highly qualified players let their choice, while non-collusion is the dominant strategy in view of backward induction, be influenced by emotional considerations and feelings of revenge and that the players act in a surprisingly cooperative manner; in 60 rounds both cooperated while only in 14 rounds both defected.[60]

(d) Some Results

1.109 Real oligopoly situations are more complicated than the stylized games described in the previous sub-section. In an oligopoly there are usually more than two players, and each company has the choice not simply between competing or colluding but has to decide on a number of parameters that are important for competition; not just price or output, but also promotional activity, product differentiation, product and process innovation. On each of these parameters, there are not just two options and two pay-offs but usually a range of options and pay-offs. It is therefore not surprising that there is no super oligopoly model that by incorporating all the parameters and strategies provides clear-cut solutions to the oligopoly game. However, game theory helps to understand the inherent tension between competition and collusion within oligopolistic markets.

1.110 Game theory has helped to identify a number of factors which influence the scope for collusion between market players. One way or the other, they all have a bearing on the ease with which firms can establish the terms of coordination (eg to arrive at a 'focal price') and on the trade-off, outlined in para 1.102, that each player faces between the gains from competing in the short run against the loss of collusive profits for the subsequent period(s). Some of the more important factors are the following:[61]

- *The number of sellers*: the fewer the sellers, the easier it is to agree on the terms of collusion and to monitor adherence. Furthermore, the greater the number of sellers, the greater is the incentive to deviate given that each company has more market share to gain while its lower price will have less effect on the revenues from the output it already sells.
- *Market transparency*: the more transparent the market is in terms of, for instance, the availability of pricing data or market share data, the easier it becomes for the colluding firms to detect competitive behaviour.
- *Product differentiation*: the main reason why product differentiation makes tacit collusion more difficult is that it may exacerbate monitoring problems (eg as regards price setting by competitors). When product differentiation is also related to quality differences, companies producing high-quality products may have a greater incentive to deviate than low-quality firms: they may have more to gain from deviating and less to fear from retaliation by others.
- *Cost asymmetries*: the higher the disparities in terms of cost structure, the less likely it is that tacit collusion will result in a market. First, companies may find it difficult to agree on a 'common price': low-cost companies typically prefer a lower collusive level than high-cost

---

[59] See n 54.
[60] See W. Poundstone, *Prisoner's Dilemma* (New York: Doubleday, 1992), 106–16.
[61] The list of factors is based on M. Ivaldi, B. Jullien, P. Rey, P. Seabright, and J. Tirole, 'The Economics of Tacit Collusion', Report to DG Competition (available at <http://ec.europa.eu/competition/mergers/studies_reports/studies_reports.html>), with the exception of the item 'ringmaster'.

companies. Secondly, low-cost companies are more difficult to discipline. This reasoning also underlies the idea that 'maverick' firms make collusion more difficult.

- *Symmetry of market shares*: lack of symmetry in market shares is not by itself an indication that collusion is difficult to achieve in a market. However, when market shares are asymmetric in a given industry, this may be the result of different cost levels and/or differences in product characteristics. These more profound differences are factors that may affect the scope for collusion (see the previous two factors).

- *Frequency of interaction*: companies will find it easier to sustain coordination when they interact more frequently. This is because companies can react more quickly to a deviation by any of the other firms. Bidding markets featuring large and infrequent contracts are therefore less prone to collusion.

- *Entry barriers*: tacit collusion is more difficult to sustain when entry barriers are low. In deciding whether to adhere to the terms of coordination, companies make a trade-off between the short-term gains of deviating and the loss in future profits associated with collusion. The prospect of future entry tends to reduce the scope for future collusion, making the latter aspect less relevant in the trade-off.

- *Excess capacity*: the impact of capacity constraints (or the absence thereof) on the scope for tacit collusion is not so clear-cut. When companies are capacity constrained, they lack both the incentive to deviate (there is little scope for increasing market share), and the ability to react against another company that deviates from the collusion.

- *Demand growth*: in principle, demand growth increases the value of future gains from collusion and thereby the incentive to adhere to the terms of coordination. However, given that demand growth increases the prospect of future entry, it may also reduce the incentive to collude.

- *Innovation*: innovation makes collusion more difficult. The prospect of innovation reduces both the (expected) value of future coordination and the degree to which other firms can retaliate against a company deviating from the collusion.

- *The presence of a 'ringmaster'*:[62] the existence of a dominant firm acting as a price leader and as a swing producer, should changes in demand conditions require it, can be materially important in maintaining price discipline. Rival companies in such a market may choose not to contest the leadership position of the dominant firm, but instead prefer to live under the shelter of the price level maintained by that firm.

Game theory also puts into clearer perspective the role played by so-called facilitating **1.111** devices. A number of such practices that facilitate cooperation are described in the literature. Rees, for example, mentions the following facilitating devices: information exchange, trade associations, price leadership, collaborative research and cross-licensing of patents,

---

[62] The term 'ringmaster' was originally employed by T. Krattenmaker and S. Salop in a vertical restraints context ('Anticompetitive Exclusion: Raising Rivals' Costs to Achieve Power over Price' (1986) 96(2) Yale LJ 211). However, the intuition can also be extended to a scenario of an asymmetric Stigler-type detection/ punishment oligopoly (see Stigler, 'A Theory of Oligopoly' (n 54)), in which a dominant firm acts as the leader and enforcer of coordinated interaction (cf S. Salop, Assessment of Dominance: Unilateral and Co-ordinated Effects, speech delivered at the IBA Conference, Brussels, 8 November 2002). See also J. Harrington, 'How Do Cartels Operate?' (2006) 2(1) Foundations and Trends in Microeconomics 63:

> In the carbonless paper cartel, cartel member AWA had a market share in Europe of 30–35% and was the largest producer with capacity exceeding twice that of any other firm. It used its dominant position in the market to threaten aggressive pricing if firms did not comply with the collusive agreement.

most-favoured-customer (MFC) and meeting-competition (MC) clauses in sales contracts, resale price maintenance, basing point pricing, common costing books.[63] What all these devices have in common is the exchange of information as central element. This is obvious for the direct exchange of information between competitors or the exchange through an intermediary such as a trade organization (collection and dissemination of data, forecasting studies, common costing books, etc). But it is also the case when the exchange runs via the customers (price leadership, MFC and MC clauses, resale price maintenance, basing point pricing). These devices may all be used to limit the influence of factors that destabilize cooperative outcomes or strengthen the factors that support cooperative outcomes. This is done by limiting the gains of competing, by monitoring each other's behaviour thus making detection of competing easier, by better targeting the infliction of punishment, or by making it easier for firms to reach a view on the appropriate collusive strategy by reducing the effects of factors such as product heterogeneity, uncertainty about future cost, demand, or capacity, and technological change.

1.112 In terms of the prisoner's dilemma, such facilitating practices may reduce the gains of defecting/competing. They reduce the pay-off/profits that can be obtained from competing while the others act in a collusive manner. In the extreme case, the pay-off matrix may change so much that it is no longer a prisoner's dilemma type of game. Such a matrix is unlikely under most market conditions, but the oligopolists may take steps to worsen their possible gain from competing, to make the pay-off structure change from a prisoner's dilemma to a setting where restricting output is the dominant strategy. An example of this extreme case is given by the pay-off matrix represented in Figure 1.12.

Figure 1.12 A non-cooperative game with cooperation as the dominant strategy

1.113 In such a case, by reducing its price each player loses more in profits on its current sales than it could gain in profits from newly attracted sales. To restrict output has become the dominant strategy for both players and the high price outcome ensues without collusion. Competing is no longer an attractive option. For instance, the adoption by the oligopolists of an MFC plan, guaranteeing to pay customers retroactively any possible discount the company will give within, for example, the next year. Such a plan may significantly undermine an oligopolist's gain from competing, as the lower price offered to lure new customers away from its competitors will have to be awarded to all its customers during the previous year. If applied simultaneously by all, it will reduce each firm's gain from competing. However, to start colluding and to implement such facilitating practices that worsen the possible gain from competing, will itself require overcoming a prisoner's dilemma.

---

[63] R. Rees, 'Tacit Collusion' (1993) Oxford Rev Econ Pol 27, 35–7.

Although an accurate, predictive, and encompassing oligopoly model does not yet exist, what can one conclude as to the scope for collusion? The factor that is often taken as a starting point for competition policy analysis is the number of firms and their market shares and the resulting market concentration.

1.114

With a limited number of firms in the market, prices and profits may be significantly higher than they would be in a market with many firms. The economic literature does not provide a specific number of firms below which, or a particular level of market concentration above which, supra-normal prices and profits will be likely to arise.[64] As explained, the likelihood of such effects lowers as the number of firms increases. Both the likelihood and size of the possible effects can be expected to become rather small above a certain number of firms, possibly when there are more than ten or 12 main firms in the market.[65]

1.115

In the EU Merger Guidelines, the Commission creates safe harbours by formulating negative presumptions using the Herfindahl–Hirschman Index (HHI).[66] It is stated that the Commission is unlikely to identify horizontal competition concerns in a market with a post-merger HHI below 1,000—that is, in a market with at least ten competitors. In a market with a post-merger HHI between 1,000 and 2,000, where there will thus be at least five competitors, the Commission is unlikely to identify horizontal competition concerns if the HHI increase resulting from the merger remains below 250 and in a market with a post-merger HHI above 2,000 where the HHI increase resulting from the merger remains below 150. Above these thresholds, there is no negative or positive presumption of competition concerns. Other factors that may make collusion easier, more stable, or more effective, such as the possibility to monitor, deter, and raise entry barriers, will have to be evaluated. EU competition policy practice has been rather cautious and has challenged mergers on the basis of concerns of coordinated effects mainly in cases where there are two or three main companies in the market.[67]

1.116

---

[64] A specific number is mentioned by Selten in an article of 1973 and Phlips in later work which builds upon the ideas of Selten. See: R. Selten, 'A Simple Model of Imperfect Competition Where Four are Few and Six are Many' (1973) 2 Int'l J Game Theory 141; L. Phlips, *Competition Policy: A Game-Theoretic Perspective* (Cambridge: Cambridge University Press, 1995); and L. Phlips, 'On the Detection of Collusion and Predation' (1996) 40 Eur Econ Rev 495. Their conclusion is that '4 are few and 6 are many', ie when there are four firms or less in a market the likelihood of collusion will be one, whereas this likelihood drops to close to nil when the number of firms becomes six or more. However, their conclusion is only valid with the restrictive assumptions underlying their model. Selten's model, which is also the basis of Phlips's work, excludes the possibility of cheating. Each company decides beforehand whether it will cooperate. Once it has decided to cooperate, it sticks to its promise. This therefore resembles the situation of a cooperative game with enforceable agreements. The model shows that such binding agreements will be formed with a likelihood of one as long as the number of firms does not exceed four.

[65] Scherer and Ross, *Industrial Market Structure and Economic Performance* (n 1), 277 state: 'As a very crude general rule, if evenly matched firms supply homogeneous products in a well-defined market, they are likely to begin ignoring their influence on price when their number exceeds ten or twelve.' This number of ten or 12 companies does not necessarily include fringe companies and companies supplying niche markets, which can also be active on the market while not exerting important competitive pressure.

[66] The HHI is a measure of concentration defined as the sum of the squared market shares of all the firms in the market. Eg a market containing five firms with market shares of 40, 20, 15, 15, and 10 per cent, respectively, has an HHI of 2,550 ($40^2 + 20^2 + 15^2 + 15^2 + 10^2 = 2,550$). The HHI ranges from close to zero (in an atomistic market) to 10,000 (in the case of a pure monopoly).

[67] See eg the following merger cases: Case IV/M190 *Nestlé/Perrier*, OJ 1992 L356/1; Case IV/M308 *Kali & Salz*, OJ 1994 L186/38 and OJ 1998 C275/3; Case IV/M619 *Gencor/Lonhro*, OJ 1997 L11/30; Case COMP/M3099 *Areva/Urenco/ETCJV*, OJ 2006 L61/11; Case COMP/M.3333 *Sony/BMG (I)*, OJ 2005 L062/30; Case COMP/M.3333 *Sony/BMG (II)*, OJ 2008 C94/9; Case COMP/M.4980 *ABF/GBI Business*, OJ 2009 C145/09.

1.117 In the context of antitrust enforcement under Articles 101 and 102, competition authorities may want to concentrate on the detection of explicit collusion and in addition on the detection and analysis of facilitating practices. Investigation of facilitating devices offers the possibility to scrutinize conscious parallelism as closely as possible and take remedial action by, if necessary, disallowing the facilitating device. In terms of US antitrust practice, this means defining the 'plus' in 'conscious parallelism plus something else' that together restrict competition.

1.118 The Commission cartel decisions corroborate this analysis, as evidenced by an analysis of all cartel prohibitions adopted by the Commission under Article 101 in the period 2001–11. Although it is not always easy to establish the number of competitors in the relevant market(s) in Commission cartel decisions as it is not always considered necessary carefully to define the market(s) in cases of clear-cut price-fixing and market-sharing cartels, the number of cartelists seems in general to have been below 12 and in most cases the cartel consisted of between three and eight members while covering all or most of the market.[68] Exceptions to this rule are mainly found in decisions involving trade associations,[69] liner conferences,[70] or previously regulated industries such as the steel industry,[71] where for different reasons effective cartels were able to operate with a higher number of main players.

1.119 The (limited) number of firms in the market also plays a role where exchange of information is used as a facilitating device for parallel behaviour, as recently described in the Guidelines on horizontal cooperation agreements and as previously found in the *Fatty Acids* case and in the *UK Tractors* case, two cases where exchange of information was the sole competition

---

[68] See the following cases: *SAS/Maersk Air*, OJ 2001 L265/15; *Graphite Electrodes*, OJ 2002 L100/1; *Sodium Gluconate*, not yet reported; *Vitamins*, OJ 2003 L6/1; *Citric Acid*, OJ 2002 L239/18; *Belgium Breweries*, OJ 2003 L200/1; *Luxembourg Breweries*, OJ 2002 L253/21; *Zinc Phosphate*, OJ 2003 L153/1; *German Banks*, OJ 2003 L15/1; *Carbonless Paper*, OJ 2004 L115/88; *Specialty Graphite*, not yet reported; *Plasterboard*, OJ 2005 L166/8; *Methylglucamine*, OJ 2004 38/18; *Fine Art Auction Houses*, OJ 2005 L200/92; *Methionine*, OJ 2003 L255/1; *Austrian Banks—'Lombard Club'*, OJ 2004 L56/1; *Industrial and Medical Gases*, OJ 2003 L84/1; *Food Flavour Enhancers*, OJ 2004 L75/1; *Sorbates*, OJ 2005 L182/20; *Electrical and Mechanical Carbon and Graphite Products*, OJ 2004 L125/45; *Organic Peroxides*, OJ 2005 L110/44; *Industrial Copper Tubes*, OJ 2004 L125/50; *Copper Plumbing Tubes*, OJ 2006 L192/21; *French Beer*, OJ 2005 L184/57; *Raw Tobacco in Spain (The Tobacco Processors)*, OJ 2007 L102/14; *Hard Haberdashery/Needles*, OJ 2009 C147/23; *Choline Chloride*, OJ 2005 L190/22; *Monochloroacetic Acid*, OJ 2006 L353/12; *Thread*, OJ 2008 C21/10; *Italian raw tobacco*, OJ 2006 L353/45; *Rubber chemicals*, OJ 2006 L353/50; *Hydrogen peroxide*, OJ 2006 L353/54; *Methacrylates*, OJ 2006 L322/20; *Bitumen Nederland*, OJ 2007 L196/40; *Fittings*, OJ 2007 L283/63; *Synthetic rubber*, OJ 2008 C7/11; *Gas insulated switch gear*, OJ 2008 C5/7; *Elevators and escalators*, OJ 2008 C75/19; *Netherlands beer market*, OJ 2008 C122/1; *Hard haberdashery: fasteners*, OJ 2009 C47/8; *Bitumen Spain*, OJ 2009 C321/15; *Professional videotapes*, OJ 2008 C57/10; *Flat glass*, OJ 2008 C127/9; *Chloroprene rubber*, OJ 2008 C251/11; *Synthetic rubber*, OJ 2009 C86/7; *International removal services*, OJ 2009 C188/16; *Sodium Chlorate*, OJ 2009 C137/6; *Aluminium Fluoride*, OJ 2011 C40/22; *Candle waxes*, OJ 2009 C295/17; *Bananas*, OJ 2009 C189/12; *Car glass*, OJ 2009 C173/13; *Marine hoses*, OJ 2009 C168/6; *E.ON-GDF collusion*, OJ 2009 C248/5; *Calcium carbide*, OJ 2009 C301/18; *Power transformers*, OJ 2009 C296/21; *Heat stabilisers*, OJ 2010 C307/9; *DRAMs*, OJ 2011 C180/15; *Carbonless paper*, OJ 2011 C138/21; *Animal feed phosphates*, OJ 2011 C111/15 and 19; *LCD*, OJ 2011 C295/8; *Consumer detergents*, OJ 2011 C193/14; *Exotic fruit*, not yet reported; *CRT glass bulbs*, not yet reported; *Refrigeration compressors*, not yet reported.

[69] See the following cases: *SPO*, OJ 1992 L92; *CNSD*, OJ 1993 L203/27; *SCK/FNK*, OJ 1995 L312/79; *COAPI*, OJ 1995 L122/37; *FENEX*, OJ 1996 L181/28; *FEG and TU*, OJ 2003 L39/1; *Concrete Reinforcing Bars*, OJ 2006 L353/1; *French Beef*, OJ 2003 L209/12; *Raw Tobacco in Spain (The Associations of Tobacco Growers)*, OJ 2007 L102/14; *Industrial bags*, OJ 2007 L282/41; *Bathroom fittings and fixtures*, OJ 2011 C348/12; *Airfreight*, not yet reported.

[70] See the following cases: *TAA*, OJ 1994 L376; *Far Eastern Freight Conference*, OJ 1994 L378/17; *TACA*, OJ 1999 L95/1; *FETTCSA*, OJ 2000 L268/1.

[71] See the following cases: *Welded Steel Mesh*, OJ 1989 L260/1; *Steel Beams*, OJ 1994 L116/1; *Wirtschaftsvereinigung Stahl*, OJ 1998 L1/10; *Prestressing steel*, OJ 2011 C339/7.

infringement.[72] In these cases also other important elements of a game-theoretical analysis can be found. In *Fatty Acids*, the description of the parties' motivation contains many elements of a non-cooperative game with a prisoner's dilemma; companies are afraid of being misunderstood by their competitors, afraid of provoking price-cutting which again would make retaliation necessary, when business is stolen recouping it elsewhere would be detrimental to the current equilibrium, output needs to be controlled and monitoring of respective market positions is essential to allow 'orderly marketing'. In the *UK Tractors* decision, the Commission emphasized the context of a concentrated market, the creation of market transparency which is likely to destroy what hidden competition there remains in that market, the elimination of uncertainty about competitors' actions, the shortened reaction lag to price competition which greatly reduces the advantage of a company that tries to undercut, the situation that targeted punishment is made possible, and the possible effect that a reduction of intra-brand competition may have on inter-brand competition, which all fit very well in a game-theoretical explanation.

D. Dynamic Welfare Analysis of Market Power[73]

(1) Innovation and Welfare

The static welfare analysis described in Section C does not take dynamic aspects of competition, most notably innovation, into account. Technological developments are abstracted away, by assuming the level of technology as constant. This, of course, is at best reflective of reality in the short term and certainly not in the longer term. In the real world, product markets develop and change over time because of innovation; improved or new products and production processes are introduced. New or improved products will in general lead to greater consumer satisfaction and improved or new production processes will lead to lower production costs. In other words, these dynamic efficiencies lead to welfare gains. A proper welfare analysis of market power should thus not only take the static but also the dynamic negative effects and efficiencies into account and if the rate of innovation is affected by the market structure or the level of competition it may be necessary to assess any trade-off between static and dynamic negative and positive effects. **1.120**

There is agreement that competition is the driving force for static allocative efficiency. Competition forces companies in a market with a given technology to offer the best quality products at the lowest prices. However, it is also a generally accepted and well-substantiated point of view that innovation is the main source of increases in economic welfare. Starting with Solow, the literature has shown that technological innovation together with an increased ability (skill level) on the part of the labour force are the main driving forces behind productivity gains and welfare growth.[74] The most recent literature often speaks of total factor productivity (TFP) as the 'Solow' residual. It is the growth factor that remains after changes **1.121**

---

[72] Commission Guidelines on the applicability of Article 101 of the Treaty on the Functioning of the European Union to horizontal co-operation agreements, OJ 2011 C11/1, para 79; *Fatty Acids*, OJ 1987 L3/17; *UK Tractors*, OJ 1992 L68/19.

[73] This section is based on L. Peeperkorn, 'IP Licenses and Competition Rules: Striking the Right Balance' (2003) 26 World Competition 527.

[74] R. M. Solow, 'Technical Change and the Aggregate Production Function' (1957) Rev Econ and Statistics 312; Scherer and Ross, *Industrial Market Structure and Economic Performance* (n 1), ch 17; W. K. Tom, 'Background Note', Roundtable on Competition Policy and Intellectual Property Rights, Committee on Competition Law and Policy, OECD, October 1997, pp 21–2.

in the total amount of production factors (eg an increase in the size of the labour force) have been accounted for.[75] These studies have continued to show that increased ability of the workforce and innovation are the main drivers of growth.[76] This explains why societies in general try to spur the creation and dissemination of innovation. In the case of a choice between dynamic and static efficiencies, the former will quickly outweigh the latter.

(2) Different Views

1.122 This has led to the question whether innovation instead of price competition should be the focal point of competition policy and, if so, whether this should lead to a drastic revision of competition policy. This question goes to the heart of competition policy and questions its general validity when applied to markets for new and existing products. The assumption is that there may be a contradiction between innovation and (price) competition, or at least that by focusing on the preservation of (price) competition the rate of innovation may be harmed. Underlying this assumption is the view that (high) concentration may have a positive influence on the rate of technological progress.

1.123 There is no clear agreement in the economic literature concerning the benefit of competition for innovation and hence dynamic efficiency. There are economists who, in the footsteps of Schumpeter, claim that innovation is spurred by monopoly.[77] Monopoly profits may fund R&D and a high market share may help to appropriate the value of the resulting innovations. They argue that there is therefore a conceptual flaw in competition policy. Competition policy, by attacking monopoly and preventing market power from arising, may have a positive effect on static allocative efficiency but at the same time undermines dynamic efficiency. As the latter is much more important for welfare growth, it is argued that competition policy easily leads to unwanted policy results, that is, less growth and less welfare.

1.124 The Schumpeterian view has been contradicted by Arrow[78] and also by other economists, who have put forward a number of reasons why competition may provide more incentives for innovation than monopoly. A firm under competitive pressure will be less complacent and will have more market share to gain through innovation. In addition, in the case of a product invention, the new product will not cannibalize the firm's own market as it would under monopoly. It is also argued that innovation incentives depend not so much on the post-innovation profits per se, but on the difference between post-innovation and pre-innovation profits. The direct effect on welfare is also supposed to be better under competition, especially in the case of a process invention, as the innovation will be applied to a higher output than under monopoly.[79] Greater product market competition and a strict competition policy both work as an effective stick to foster innovative effort.[80]

---

[75] Depending on the definition and data availability, TFP includes or excludes improvements in the quality of the workforce. Where it is excluded, TFP mainly concerns technical and organizational innovation.

[76] B. Van Ark, M. O'Mahony, and M. P. Timmer, 'The Productivity Gap between Europe and the United States: Trends and Causes' (2008) 22(1) J Econ Perspectives 25; '2012 Productivity Brief—Key Findings', The Conference Board, 2012; C. Syverson, 'What Determines Productivity?' (2011) J Econ Literature, June, 326.

[77] J. A. Schumpeter, *Capitalism, Socialism and Democracy* (New York: Harper & Rowe, 1942).

[78] K. J. Arrow, 'Economic Welfare and the Allocation of Resources for Invention' [1962] The Rate and Direction of Inventive Activity: Economic and Social Factors 609.

[79] Static welfare analysis indicates that industry output is higher under competition than under monopoly. See Section C.

[80] P. Aghion, N. Bloom, R. Blundell, R. Griffith, and P. Howitt, 'Competition and Innovation: An Inverted-U Relationship' (2005) 120 Quarterly J Econ 701; S. Martin, 'Competition Policy for High Technology Industries' (2001) J Industry, Comp and Trade 441.

(3) Some Empirical Results

Empirical research on the relationship between market structure and innovation, usually the litmus test in the case of theoretical controversy, does not give unequivocal results but tends to support the view of Arrow. In general, competition and open markets provide the better incentives for innovation while monopoly and high concentration tend to limit and delay innovation.[81] There are some indications of an 'inverted U relationship' between concentration and the ratio of industry R&D to industry sales, with the highest R&D/sales ratios occurring where the four biggest companies in the industry sell 50–60 per cent of total industry sales.[82] It is thus in relatively de-concentrated markets that most is spent on R&D: if in a market the largest four firms collectively have no more than half the market, this means there are thus at least eight but generally more competitors in the market overall, implying that each firm will have only very limited market power. However, it is also clear that other factors such as the technological opportunity of the sector, that is, the ease of achievement of innovations and technological improvements in that sector, are more important than the level of concentration to explain R&D intensity. Nonetheless, using data for the UK and controlling for technological opportunity, Geroski also found higher seller concentration and increases in other monopoly-related variables to have a significant negative impact on the emergence of innovations.[83] In a study analysing reports in specialized technical literature covering the entire manufacturing sector, Acs and Andretsch found that the average small-firm innovation rate is higher than the large-firm innovation rate.[84] Other research points to the very important role of newcomers, especially where the invention of radically new products and concepts is concerned, and to the related interest in keeping entry barriers at modest levels.[85]

1.125

Further evidence on the positive relationship between competition and innovation comes from the comparison of the economic performance of countries with open and competitive vis-à-vis restricted market systems.[86] Typically, measures of competition intensity at the

1.126

---

[81] See Scherer and Ross, *Industrial Market Structure and Economic Performance* (n 1), ch 17; Tom, 'Background Notes' (n 74), 22; and 'Patents, Competition and Innovation', Background Note by the Secretariat, Competition Committee, OECD, September 2006, pp 27–38.

[82] Aghion et al, 'Competition and Innovation' (n 80).

[83] P. Geroski, 'Innovation, Technological Opportunity, and Market Structure' [1990] Oxford Economic Papers 42. See also Scherer and Ross, *Industrial Market Structure and Economic Performance* (n 1), ch 17.

[84] Z. J. Acs and D. B. Andretsch, 'Innovation, Market Structure and Firm Size' (1987) LXIX Rev Econ and Statistics 567.

[85] For all this literature it should be noted that research into the relationship between market structure and innovation is complicated by the fact that to a certain extent both are endogenous: both depend on more basic factors such as technological opportunities for innovation and demand conditions. This makes it difficult to identify the relationship between market structure and innovation in isolation.

[86] These findings are based on a large number of studies on the link between competition and productivity which led to the adoption of the Communication from the Commission, A Pro-Active Competition Policy for a Competitive Europe, COM/2004/0293 final/, available at <http://ec.europa.eu/competition/publications/archive.html>. See also L. Foster, J. Haltiwanger, and C. J. Krizan, 'Aggregate Productivity Growth: Lessons from Microeconomic Evidence' in C. R. Hulten, E. R. Dean, and M. J. Harper (eds), *New Developments in Productivity Analysis* (Chicago: University of Chicago Press, 2001); R. Disney, J. Haskel, and Y. Heden, 'Restructuring and Productivity Growth in UK Manufacturing' (2003) 113(489) Economic J 666; E. Bartelsman, J. Haltiwanger, and S. Scarpetta, 'Microeconomic Evidence on Creative Destruction in Industrial and Developing Countries', World Bank Policy Research Working Paper No 3464 (2004); L. Foster, J. Haltiwanger, and C. J. Krizan, 'Market Selection Reallocation and Restructuring in the US Retail Trade Sector In the 1990s' (2006) 88(4) Rev Econ and Statistics 748; A. Bravo-Biosca, 'Exploring Business Growth and Contraction in Europe and the US', NESTA Research Report (2010); C. Syverson, 'What Determines Productivity?' (2011) J Econ Literature 326;

economy-wide level are positively associated with economic growth. Specifically, product market competition has been found significantly to raise productivity growth rates. Greater product market competition causes not only the productivity level to increase at the firm level, but also reduces the differences in productivity levels between firms within a market by reducing the presence of less productive firms, thereby raising average productivity.[87] Competition not only increases productivity by providing more chances for entry and expansion of the more innovative firms at the cost of less productive firms (reallocation or selection effect, also sometimes referred to as the 'churn' process in industries), but it also incentivizes incumbent firms to improve their practices and to innovate.[88] There is also ample evidence that vigorous domestic competition promotes success in international markets. Comparative case studies in single industries in the US, Japan, and Europe show that import/export and competition (especially global competition with best-practice producers) enhances productivity. At the same time, firms with higher market power tend to be less productive in relative terms and significant increases in concentration are generally associated with reductions in efficiency and the level of productivity.[89]

(4) The 'New Economy'

1.127 In recent years, there has been a more refined debate as to whether the supposed different dynamics of competition in sectors undergoing rapid technological change requires a more or less fundamental revision of competition policy for those sectors.[90] For instance, Evans and Schmalensee argue that competition in important new industries centres on investment in intellectual property (IP). Firms engage in competition for the market through sequential winner-takes-all races to produce drastic innovations, rather than through price/output competition in the market and through incremental innovation.[91] They argue that firms will obtain considerable short-term market power, but ignoring their dynamic vulnerability may lead to misleading antitrust conclusions.

1.128 For competition policy, it would therefore be important to distinguish between industries where product markets are (continuously) destroyed and replaced through drastic innovations on the one hand and, on the other hand, industries where within product markets

---

and N. Bloom, C. Genakos, R. Sadun, and J. Van Reenen, 'Management Practices Across Firms and Countries' (2012) 26(1) Academy of Management Perspectives 12.

[87] For specific references, see n 86.

[88] What is described here as an effect of competition is not only an increase of the rate of innovation but also of the rate of dissemination and absorption of new technology and is linked to reducing what was termed X-inefficiency in Section C.4(b) on the static welfare effects of monopoly.

[89] T. J. Klette, 'Market Power, Scale Economies and Productivity: Estimates from a Panel of Establishment Data' (1999) 47(4) J Ind Econ 451; R. E. Caves and Associates, *Industrial Efficiency in Six Nations* (Cambridge, MA: MIT Press, 1992); A. Green and D. Mayes, 'Technical Efficiency in Manufacturing Industries' (1991) 101 Economic Journal 523.

[90] It is sometimes argued, often in a rather loose way, that the pace of technological change is increasing or has increased in recent times. There seems little evidence of this trend. Traditional indicators such as productivity growth rates have not shown a clear upward trend in the pace of innovation. Some claim that the rate of innovation is poorly measured by such indicators as many qualitative improvements are not captured: however, the same applies for the productivity figures of the past and to show a clear upward trend in the pace of innovation one should in that case show that qualitative improvements have become more important over time. It seems more likely that the impression that innovation is increasing in pace is only a matter of perception: changes in one's own time always seem more rapid and upsetting, just like the perception of speed will be stronger if one is near to a passing train than when one is looking at the train from a distance.

[91] D. S. Evans and R. Schmalensee, 'Some Economic Aspects of Antitrust Analysis in Dynamically Competitive Industries', NBER Working Paper 8268 (May 2001).

innovation develops incrementally. Evans and Schmalensee identified the following industries as having Schumpeterian dimensions: computer software, computer hardware, Internet-based businesses (portals, BtoB exchanges), communications networks, mobile telephony, biotechnology, and, to a lesser extent, pharmaceuticals.

This is again in the first place an empirical question. Evans and Schmalensee acknowledge that an initial phase with bursts of innovation may only characterize the infant stage of a new industry and may very well be followed by a long period of comparative stability and incremental innovation. They, for instance, refer to the car industry having had Schumpeterian aspects around 1910 and decades of stability afterwards. Other examples are the chemical and electronics industries that were described in the 1950s as 'new-economy'.[92] It seems most likely that also today's 'new economy' industries will turn into more 'normal and traditional' industries if they have not done so in good part already.

1.129

In addition, Evans and Schmalensee recognize that many of the sectors they assess as having Schumpeterian characteristics also have network effects and that these effects tend to reinforce the market leaders' position. A network effect is created when the consumption of a product by one customer enhances the value of consumption by other customers. The more customers who purchase the product, the higher its value to each of them. The classic example is the telephone. The more people who own a telephone, the more valuable having a phone is to each of them as the network they can call increases. Where such networks have a closed character, due to interoperability problems, network effects can make markets tip and become highly concentrated and can impose significant barriers to entry. Similarly, switching costs and lock-in may prevent displacement of market leaders.

1.130

In line with the general conclusion in the literature, Evans and Schmalensee do not contend that dynamically competitive industries should be immune from careful antitrust scrutiny, nor do they contend that the basic principles of antitrust should be modified.[93] Price fixing, foreclosure, market partitioning, etc can and will still harm consumers, also in the 'new economy'. However, as is the case for every industry, also for the new-economy industries, competition policy needs to take account of industry- or technology-specific characteristics. According to Evans and Schmalensee, in particular market definition and market power analysis have to be modified when applied to highly innovative sectors.

1.131

In their view, traditional market definition and market share analysis does not acknowledge that in Schumpeterian industries companies are constrained from doing harm to consumers by dynamic competition. An essential element of market-power analysis should be an examination of actual and potential innovative threats, including threats from alternative technologies. Where the market leader's position may not be based on durable assets such as production capacity but based on the quality of its current products and IP, it may therefore be fragile. They argue that in these industries a market share measures, at best, static market power. Static market power does not provide a useful measure of the real competitive constraints on the leading firms in these industries. They may not be constrained by the

1.132

---

[92] See D. E. Lilienthal, *Big Business: A New Era* (New York: Harper, 1952).
[93] See also eg 'E-Commerce and its Implications for Competition Policy', Discussion Paper 1, OFT, August 2000, p 1: 'e-commerce will not give rise to any entirely new forms of anti-competitive behaviour, nor will it raise any new issues that cannot be dealt with under the existing competition law framework. However, . . . there are . . . areas where detailed application of the rules may require some adjustment.'

behaviour of existing competitors as the latter are often few or absent and scale economies and network effects may form effective barriers to entry for similar products. The real and dynamic constraints may come from firms actually or potentially making significant R&D investments to replace the current products. The question whether these are around and how credible the threat might be, they argue, cannot be measured by market share. Dynamic competition may not be effective when the leading company owns all IP necessary for radical innovation or when it forecloses important distribution channels. It may be, though, that several companies make or can be expected to make significant R&D investments and that experts consider the outcome of the rivalry in doubt, in which case dynamic competition may be effective. In particular in such industries, during this initial phase where markets are continuously destroyed and replaced through drastic innovations, a company's market share may not reflect well its position on the market and may not serve as a good first indicator of its market power.

(5) Some Concluding Remarks

1.133 For an analysis of the competition dynamics in a particular industry, it is always necessary to take the characteristics of the industry into account. In principle, however, there seems to be no important conflict between innovation and competition policy aimed at product market competition and no conflict between protecting static and dynamic efficiency. Competition policy, by defending competition and open markets, will in general have a positive impact on both static and dynamic efficiency.[94] Companies under competitive pressure will be less complacent and will have more incentive to innovate and gain market share. Product market competition and a strict competition policy generally work as an effective stick to promote innovative effort.

E. Market Definition[95]

1.134 Antitrust analysis focuses on the question whether companies are, or will be, in a position to exercise market power. It is difficult to think of this question without reference to a proper context, without reference to a 'market'. For example, the analysis of a contract between two companies might indicate that certain clauses in the agreement restrict the competitive conduct of one or both of the parties. However, the effects of the clauses at issue can only be expected to have a significant impact, on any relevant market and hence on market variables such as price or output, if the companies concerned possess some market power. In order to identify the existence or creation of market power, one typically needs to proceed to an analysis of the market.[96]

1.135 What is the right context for antitrust analysis? What is the 'relevant market'? Though obviously related, the relevant market for antitrust purposes does not always coincide with the

---

[94] cf J. Baker, 'Beyond Schumpeter vs. Arrow: How Antitrust Fosters Innovation' (2007) 74 Antitrust LJ 575.

[95] This section builds on a text written by Kirti Mehta for the first edition of this book.

[96] As a general rule, market definition is more relevant where the analysis is prospective than in situations where the anti-competitive nature or effect can be analysed more directly, *ex post*. In the context of horizontal cartels, eg, the anti-competitive object of the behaviour is clear and does not require any market definition. In the latter case, one can still proceed to a definition of the market in order to evaluate the impact of the cartel, as is required in the context of claims for damages.

market as it is described in marketing reports or other business reports. Companies, when thinking of what constitutes the relevant market, naturally consider this question from a business perspective. For example, many European companies nowadays operate in several parts of Europe and the world, with a view to expanding their business. For them, the relevant geographic market is European, if not global. Similarly, many well-known companies would broadly describe their relevant area of activity as 'consumer electronics', 'health care', or 'automotive'. Market definition for antitrust purposes starts, however, from a different perspective: what options are open for the customers to acquire the product they wish to acquire? What alternatives do they have? Are they good alternatives? It is this perspective that determines, in large part, whether a company has the ability to exercise market power (eg profitably raise price) vis-à-vis its customers, or not.

Whether a company can exert market power depends on a number of factors. The availability of substitute products for the products offered by the company under consideration is only one of them. The strength of competitors, the presence of entry barriers, the presence of buyer power are other relevant factors. Nonetheless, it is useful to proceed in steps. The objective of defining a market, the first step, is to identify, both in the product and geographic dimension, those products that are capable of constraining the commercial behaviour of the company concerned in that they form sufficiently good substitutes for the product in question.[97] It thereby provides a context within which to assess the competition issue, be it the competitive impact of a given agreement, a certain type of company conduct, or a merger. **1.136**

Beyond providing context, it is clear that market definition also serves an important practical purpose. Once the market is defined it is possible to assign market shares to the various companies active in the relevant market, in order to obtain a first impression of their relative importance in the competitive process. Market definition thereby allows for a first screening of cases, to see whether there may be competitive issues.[98] **1.137**

The following subsections discuss the main principles of market definition, first in the product dimension and then in the geographic dimension. The section concludes with a number of further considerations on market definition. **1.138**

(1) Product Market Definition

The key concept in the definition of a relevant product market for antitrust purposes is *substitutability*, that is, the extent to which customers are able and willing to switch to other products in case of a price increase (or a corresponding non-price change such as a reduction in product quality or service). **1.139**

Two main avenues of substitution are often considered: demand-side substitution and supply-side substitution.[99] Demand-side substitution relates to the possibility of customers switching to alternative products that are already on offer. Supply-side substitution relates to the possibility of turning to products that are not yet offered by particular competitors, **1.140**

---

[97] cf Commission Notice on the definition of the relevant market for the purposes of [EU] competition law, OJ 1997 C372/5; see also J. Baker 'Market Definition: An Analytical Overview' (2007) 74 Antitrust LJ 129.

[98] While there can be some debate as to whether high market shares are good indicators of market power, there is less doubt on low market shares being good indicators of the absence of market power (provided the market has been properly defined). See also Section F.

[99] As will be discussed later (para 1.166), the US approach to supply-side substitution differs from that of the EU.

but that would readily be offered by them in the event of a higher price of the product in question.

(a) Demand-Side Substitution

1.141 The most immediate constraint upon the terms on which a firm supplies a product is the competitive pressure represented by adequate substitute products available (in the relevant geographic area). In the case of a price increase of the product in question, the customer would readily shift to such substitute products. In practice, the market definition problem thus reduces itself to determining the range of products that constitute good substitutes for the customer or, rather, for a sufficiently important number of customers.

1.142 The importance of demand-side substitution is underlined in the traditional description given in the EU to the concept of relevant product market: '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.'[100]

1.143 While the underlying idea is clear, in practice it is often rather difficult to determine whether products are good substitutes for each other by focusing on factors such as product characteristics, prices, or intended use alone. For some products it may be readily possible to identify a number of good substitutes on this basis, but more often than not these factors are unlikely to provide a clear basis for deciding which products should be considered part of the relevant market and which should not be considered part of the market. To take an example, different types of malt whisky may well be considered part of the same market. But can malt whisky be considered in the same relevant market as blended whisky? And what about vodka or gin? On the one hand, one would be tempted to say that malt whisky is different from blended whisky and very different from vodka or gin. On the other hand, there are a number of similarities as well: all the products are spirits, their (quality-adjusted) prices are comparable, and the products are consumed in rather similar circumstances.

1.144 Furthermore, it is important to note that not all customers are alike. Defining the market based on the 'average customer', where there are significant differences among customers, may lead to erroneous results. In determining whether other products constitute a competitive threat to the product in question, one needs to focus on the so-called *marginal customers*.[101] These are the consumers who are inclined to shift their demand, in whole or in part, to substitutes if the relative price of the product increases. If the proportion of marginal customers is sufficiently large relative to the other customers (called the infra-marginal consumers[102]), a relative price increase might well result in a substantial loss of sales.

1.145 Finally, even when one has reliable information about the actual degree of substitutability between products for the group of customers under consideration, on what basis does one conclude that the substitutability is high enough for products to form part of the same relevant market? Where should one draw a line between the products? What is the benchmark?

---

[100] See eg Market Definition Notice, para 2.

[101] Marginal consumers have, by way of definition, a willingness to pay for the product that is about equal to the price paid. If the price increased, they would probably stop buying the product (eg choose another product) or buy less of it.

[102] Infra-marginal consumers have a willingness to pay for the product that is higher than the price they have to pay for it and so they would substitute less, or not at all, if the relative price increases.

As it turns out, it is useful to apply a 'unifying principle' to market definition and to think of a relevant market in terms of what a company would do if it were the sole supplier of the set of products concerned. Specifically, one can ask the question: if a company had a monopoly for this set of products, would it want to raise the price of it? The logic is as follows: if not even a hypothetical monopolist can profitably raise the price on the products in question, then surely companies that control only part of the market cannot either. It makes no sense to look at such a 'market' in isolation; one has to look at something wider. By contrast, if a hypothetical monopolist can profitably raise price, then it becomes worthwhile to see whether any of the individual companies on the market (or all of them jointly) have monopoly power. This logic is embodied in the SSNIP methodology for the assessment of relevant markets, which we discuss next.

1.146

(b) The SSNIP Test

The need for a framework to assess economic substitutability has led to the development of the SSNIP ('small but significant non-transitory increase in price') test, also known as the 'hypothetical monopolist test'.[103] The SSNIP test links up with the purpose of market definition itself; that is, to identify the products that are capable of constraining the commercial behaviour of the company supplying the product under consideration. It proposes to make—in an iterative way—a distinction between products that would constrain a company, even if it were a monopolist, from raising the price for the product in question and those products that do not. Thus, the *benchmark* is whether it would be profitable for such a supplier (the 'hypothetical monopolist') to raise the price for the product concerned. The *methodology* is iterative.

1.147

Specifically, the SSNIP approach suggests the following line of inquiry: start with the product in question, postulate a hypothetical small but significant increase (eg in the range of 5–10 per cent) in the price at which that product is made available (the prices of the alternative products are held constant), and assess the likely reactions of customers to that increase.[104] If substitution away from the product by customers would be enough to make the price increase unprofitable because of the resulting loss of sales, then the product is not a relevant market by itself: not even a hypothetical monopolist would be able to profitably raise the price. There are other products that exercise a sufficient competitive constraint in that they prevent a company, even if it had a 'monopoly' on the product, from raising the price. [105]

1.148

---

[103] For a more elaborate discussion of the SSNIP test, see G. J. Werden, 'The 1982 Merger Guidelines and the Ascent of the Hypothetical Monopolist Paradigm' (2003) 71 Antitrust LJ 253. See also the 2010 US Horizontal Merger Guidelines, section 4.

[104] What price increase is significant or insignificant depends on the industry. In some markets, eg the market for crude oil, smaller price increases could be considered significant. However, taking a price increase that is too small would not capture the reactions of all the marginal customers, and might understate the extent of likely customer switching. Using a very large price increase would be likely to capture the reactions of significant portions of the infra-marginal customers. If used as a basis for defining the market, this would lead to very wide markets, hiding otherwise significant competition concerns.

Further, it must be noted that '5–10%' does not constitute a 'tolerance level' below which price increases would be acceptable (see also the 2010 US Horizontal Merger Guidelines, section 4.1.1).

[105] Note, however, that the logic 'If not even a hypothetical monopolist can profitably raise the price on the products in question, then surely companies that control only part of the market cannot either' need not hold when *in reality* the firm active in the postulated (narrow) market also sells substitute products *that are outside that postulated market* and where the firm has a high market share in those outside markets. In that case, the firm in question might want to raise price on the narrow market, even where a hypothetical monopolist (for which

1.149 If the price increase of the product is unprofitable, the next step of the SSNIP test is to consider the situation where a company would be the sole supplier of the product under consideration and also its next best substitute (the product to which the greatest proportion of customers switches when the price of the reference product goes up). Would such a company want to raise prices? If it would not, then these two products still do not constitute a relevant market, and it is appropriate to include additional substitutes. If raising prices were profitable, then the two products can be considered a relevant market, given that there are no other products that exert sufficient competitive pressure on the two products. More generally, the steps would be repeated until the set of products is such that small, lasting increases in relative prices would be profitable.[106] On such a set of products, a monopolist would find it profitable to raise prices, so it becomes relevant to check whether any of the individual companies on the market (or all of them jointly) possess monopoly power. Hence, we can rightly refer to the market as a 'relevant' product market for antitrust purposes.

1.150 In our spirits example, if, in the event of a price increase for malt whisky, customers would switch to blended whisky to such an extent that the price increase for malt whisky would not be profitable due to the resulting loss of sales, then the market would comprise at least all whiskies. The process would have to be extended to other available drinks (eg vodka, gin, jenever, etc) until a set of products is identified for which a price rise would not induce a significant enough substitution in demand. This would then be the relevant antitrust market from the perspective of malt whisky customers.

1.151 One might be left with the impression, from reading the previous text, that the SSNIP approach is a very 'quantitative' tool, and relies on the availability of detailed demand and cost data.[107] In our view, the complexity of the SSNIP test should not, however, be overemphasized. The most important aspect of SSNIP is its conceptual side, not its quantitative side.[108] Even when no detailed data are available, it is useful to think of the market definition question in terms of SSNIP. By asking a question which is directly linked to the purpose of antitrust analysis (is the exercise of market power an issue for this collection of products?), it brings a certain structure and consistency to the market definition exercise. The SSNIP concept provides for a framework within which to consider the question of economic substitution.

---

it is assumed that it does not sell anything outside the postulated market) would not. This possibility has to be kept in mind when applying the hypothetical monopolist test for the purpose of market definition.

This might raise the question of how any firm producing this product can ever be found dominant on this product or on a wider market given that not even a 'hypothetical monopolist' could profitably raise price by more than 5–10 per cent on the product concerned. The answer is simple. At each iteration, the SSNIP test assumes that prices of the products outside the postulated (narrow) market remain constant. This assumption may be incorrect in light of the nature of competition in the market. Especially in oligopolistic markets, producers of competing products may adjust their prices upwards in response to the price increase of the product concerned (see Sections C.5 and F.3 for further detail). The SSNIP test abstracts away from these competition aspects so as to focus purely on the question of the degree to which products are substitutes.

[106] If, for a given collection of products, a price increase is profitable, this is because the next best substitute does not exercise a sufficient constraining influence; hence, a wider market including the next best substitute could also be deemed to be the relevant market as on this wider market too a price rise will be profitable. It is for this reason that for competitive analysis the antitrust authorities normally seek to define the narrowest market among those that are deemed relevant markets.

[107] Cost levels matter in view of the profitability question ('would it be profitable to raise the price?').

[108] For similar views, see Baker, 'Market Definition: An Analytical Overview' (n 97).

In applying the SSNIP test, and in particular for the analysis of merger cases, the reference price to use will normally be the prevailing market price. However, special care needs to be taken in the context where the prevailing price has been determined in the absence of sufficient competition. In particular, for the investigation of abuses of dominant positions the fact that the prevailing price might already have been substantially increased by a given practice or conduct should be taken into account. If not, this would lead to overly wide markets being defined, and to an understatement of the firm's true market power. It is often the case that customers become more willing to switch to other products as the price of a given product increases. Assessing the degree of substitutability at this high price might wrongly suggest that more products are part of the relevant market and that therefore the relevant market is wide. This is the so-called *cellophane fallacy*.[109] In the context of horizontal mergers, the proper reference price arguably depends on the reason why there is insufficient competition.[110] When it is due to collective dominance (tacit or explicit coordination) pre-merger, it would be appropriate to start from the 'competitive' level (the price level in the absence of coordination), to identify the products relevant for maintaining collective dominance. When the high price is related to a single dominant position, the concern is that the merger may take away a next best substitute at current (high) prices.[111]

1.152

(c) Elasticity Concepts and the Diversion Ratio

An important concept in the assessment of demand substitution is the price elasticity of demand. The price elasticity of a product measures how demand for that product changes with the price of the product, keeping other prices constant (this elasticity is also called the *own-price elasticity*). In particular, it measures the percentage change in demand following a 1 per cent increase in its price. If the price elasticity is, for example, 2.0, this means that, following a 1 per cent price increase, demand for the product goes down by 2.0 per cent. The own-price elasticity is, normally speaking, negative: demand for a product falls when its price increases. However, it is common to leave the 'minus' sign out and speak of a high elasticity when the elasticity is high in absolute terms.

1.153

The (own-)price elasticity is in fact a summary indicator of the extent to which a product is subject to demand-side constraints. When the price of a product is raised, customers may, to various extents, switch away from it: they either switch to competing products, or they stop purchasing the product altogether. The (own-)price elasticity of a good captures both these movements. The higher this elasticity, the more the product is subject to demand-side constraints.

1.154

A related elasticity concept is the *aggregate price elasticity*, which measures how total market demand (combined demand for all products in a particular market) changes with a price increase of 1 per cent (keeping other prices constant).

1.155

---

[109] The cellophane fallacy is named after a case in 1956 where a US court overlooked this issue.

[110] See also the 2010 US Horizontal Merger Guidelines, section 4.1.2, and the UK Competition Commission Merger Guidelines, paras 2.09–2.10.

[111] After all, the objective of market definition is to identify the products that are capable of exerting some competitive pressure on the merging entities' products, in order to see whether a merger involving these products is problematic from a competition point of view. When the high price is related to coordination among the existing market players, the main concern is that the merger reinforces this coordination by making it less likely to break down in the future. See also Baker, 'Market Definition: An Analytical Overview' (n 97). When the high price is related to a single dominant position, there are no products exerting significant competitive pressure at the 'competitive' level (if there were, prices would not be that high). Instead, the focus should lie on identifying the products that exert competitive pressure at the higher price level, to see whether a merger involving these products allows the dominant company to further raise price.

1.156   The own-price elasticity of demand (or, more generally, the aggregate price elasticity for a group of products) provides direct input into the SSNIP test for market definition. For example, if the elasticity of the set of products one posits to be in the same relevant product market is equal to 1.5, the unit sales for the products will go down by approximately 7.5 per cent if prices for the products go up by 5 per cent (the usual SSNIP). Depending on the initial gross profit margins of the products involved, this may be profitable or not profitable. If initial margins are low, the price increase is more likely to be profitable.[112]

1.157   Another elasticity concept, the *cross-price elasticity*, is also relevant for analysing demand-side substitution, but from a different perspective. The cross-price elasticity measures how demand for a product changes when the price of some other product changes. For example, if the cross-price elasticity of product A vis-à-vis the price of B is 0.8, this means that, when the price of B goes up by 1 per cent, demand for product A goes up by 0.8 per cent. Similarly, there is a cross-price elasticity for product B with respect to the price of A.

1.158   Cross-price elasticities provide useful information on substitution patterns, but provide less direct input to the SSNIP test than the own-price elasticity. The SSNIP test is primarily concerned with the question of how much demand for product A changes with the price of A. This is measured by the own-price elasticity. The SSNIP test is only in the second instance also concerned with the question to which products demand switches. Accordingly, when, on the basis of the own-price elasticity, one concludes that a given product (or set of products) does not constitute a market on its own, an analysis of cross-price elasticities can point to the products that should be included in the relevant market. At the same time, own- and cross-price elasticities are linked. Generally, the higher the cross-price elasticity of product B with respect to the price of A, the more product B forms a competitive constraint for product A, and the less likely it is that product A is a relevant market on its own.[113]

1.159   A concept which is closely related to the cross-price elasticity is the *diversion ratio*.[114] The diversion ratio from product A to product B measures the proportion of the sales of product A that are captured by product B in the event of a price increase of product A. The diversion ratio and the cross-price elasticity are alternative ways to measure product substitution, with the former being viewed as somewhat more insightful.[115] It has become customary to define the 'next best substitute' of a product as that product for which the diversion ratio is highest.

---

[112]   An example is developed in Section G.2. One would expect the initial profit margin on any individual product and the own-price elasticity to be related: if the initial profit margin is low, this points to a high own-price elasticity. This insight does not necessarily extend to a set of products, however. Low margins observed on a set of products may be the result of competition within that set of products as opposed to demand-side substitution vis-à-vis products outside that set.

[113]   Some caution is necessary in interpreting cross-price elasticities, especially when the sales levels of products A and B are very different. Eg if the cross-price elasticity of product B vis-à-vis the price of A is 10.0, this means that when the price of A goes up by 1 per cent, demand for product B goes up by 10 per cent. If, however, the initial sales level of product A is 100 units and that of product B is only 10 units, then the 10 per cent increase in the demand for product B only represents one unit of B and, correspondingly, a decrease in demand of only one unit of A (on a total of 100, ie a 1 per cent decrease). In other words, a high cross-price elasticity does not automatically mean that the two products are in the same relevant product market. Furthermore, it is possible that the cross-price elasticity is high simply because the price of the product under consideration is itself already high (cf the cellophane fallacy problem discussed in the previous subsection). This consideration is, however, not specific to the cross-price elasticity, it is also relevant to the own-price elasticity.

[114]   For a presentation, see C. Shapiro, 'Mergers with Differentiated Products', Antitrust Magazine, Spring 1996, pp 23–30.

[115]   Also in view of the issues described in n 113.

(d) Supply-Side Substitution

1.160 Supply-side substitution relates to the possibility for customers to turn to products that are not yet offered, but that would readily be offered by companies (either new or existing) in the event of a higher price of the product in question.

1.161 Under the Commission's Market Definition Notice, supply-side substitution may be taken into account for market definition purposes in those situations in which its effects are 'equivalent to those of demand substitution in terms of effectiveness and immediacy'.[116] This requires that such alternative suppliers be able and willing to switch production to the relevant products and market them in the short term[117] without incurring significant additional costs or risks in response to small and permanent changes in relative prices (the SSNIP). When these conditions are met, the additional production that is put on the market may have a disciplinary effect on the competitive behaviour of the companies involved that is equivalent to that of demand substitution.[118] The products are then in general considered to be in the same relevant market, irrespective of whether there is substitutability from a demand perspective.

1.162 A classical example of the role of supply-side substitution is the case of paper.[119] Paper is usually supplied in a range of different qualities, from standard writing paper to high-quality papers used, for instance, to publish art books. From a demand point of view, different qualities of paper cannot be used for a specific use. For example, an art book or a high-quality publication often cannot be produced on lower quality paper. Similarly, office paper in A4 size is typically not substitutable with office paper in A3 size. However, it is possible that paper plants are prepared to manufacture the different qualities, and that production can be adjusted with negligible costs and in a short time frame. In the absence of particular difficulties in distribution, paper manufacturers are therefore able to compete for orders of various qualities. Under such circumstances, it makes sense not to define a separate market for each quality of paper and respective usage, but to view the various qualities of paper as part of one relevant market.

1.163 A practical question that arises in this context is how far one must take the argument that supply-side substitution warrants the grouping of various products into a broader market. Suppose, for example, that there is a product A that is produced by various companies, and a product B that is supplied by a number of other companies. Suppose further that only one of the B companies uses a production technology that allows it swiftly to switch production from product B to product A (the other B companies use a technology that only allows them to produce B). Would this be sufficient to conclude that product markets A and B constitute one relevant product market on the basis of supply-side substitution? Grouping the whole

---

[116] Market Definition Notice, para 20.

[117] A relevant time frame in this respect is often thought to be one year.

[118] When switching production is possible, but would require significant additional investments or time delays (eg due to the need to adjust existing tangible and intangible production assets), this possibility is not considered at the stage of market definition, but rather at the stage of considering potential competition. This is logical given that market definition is a step in the analysis identifying products that already constitute some form of competitive constraint on the product(s) in question. It makes sense, therefore, to limit attention to those companies that have the ability to provide a *swift* supply response, and to leave the more involved assessment of other entry to the stage of detailed competition assessment. Proceeding in this way also avoids the practical difficulty of having to assign hypothetical market shares to potential producers, of whom only an undefined proportion may become actual producers.

[119] An example which features in the Market Definition Notice, para 22.

A and B market into one would mean that all the single-purpose B companies are somehow viewed as constraining the A companies from raising prices, whereas in fact this is not the case. The same would apply when all B companies can switch production to product A, but in reality only a few will do so, given that the margins obtained on producing the B product are higher. In such circumstances, it is more appropriate to only take the B companies into account *to the extent* that they are able and willing swiftly to participate in the A market.

1.164 The response formulated to the previously described issue in the Commission's Notice is to note that it is appropriate to group products into one product market on the basis of supply substitutability, provided that *most* of the suppliers are able to offer and sell the various qualities under the conditions of immediacy and absence of significant increase in costs.[120]

1.165 Arguably, the principle that supply-side substitution may be taken into account for market definition in those situations in which its effects are 'equivalent to those of demand substitution in terms of effectiveness and immediacy' mandates that a cautious approach is also applied where margin differences (eg as in the case of branded versus private label products) limit supply substitution. Indeed, a useful line of inquiry for analysing supply substitution is suggested by the examination of the margins or gross returns in the production of supply substitutes as compared to the product in question. These margins should tend to equality if the supply substitutes are correctly identified, either because the prices and costs are the same or because quality-adjusted prices and costs tend to converge. Put differently, in the absence of switching barriers, the gross returns to the producers of supply substitutes cannot go too far out of line from those earned by producers of the product in question.

1.166 It is worth noting that the US approach to supply-side substitution is different from that applied in the EU. In principle, supply-side factors are not, as such, taken into account in the US in defining the scope of the relevant product market.[121] The alternative suppliers are instead considered to be *participating* in that market, even if they are not currently selling in the relevant market, in the sense that they would very likely provide a rapid supply response if prices were to rise (these firms are called 'rapid entrants').[122] In measuring such a firm's market share, the US agencies include its sales or capacity only to the extent that the firm's capacity is not 'committed or so profitably employed outside the relevant market, or so high-cost, that it would not likely be used to respond to a SSNIP in the relevant market'.[123]

1.167 In the previous example, the relevant market in the US would be the 'market for A'. The multi-purpose B company would be considered a *player* in this A market, but only to the extent it would be likely to switch production from B to A in the event of a price increase of A; the single-purpose B companies would not be viewed as players in the A market.

1.168 Having established the principle, the US Guidelines indicate, however, that if supply-side substitution 'is nearly universal among the firms selling one or more of a group of products, the Agencies may use an aggregate description of markets for those products as a matter of

---

[120] Market Definition Notice, para 21 (emphasis added).
[121] See also Baker, 'Market Definition: An Analytical Overview' (n 97).
[122] US 2010 Horizontal Merger Guidelines, section 4 ('Market definition focuses solely on demand substitution factors…').
[123] US 2010 Horizontal Merger Guidelines, section 5.2.

convenience'.[124] This *aggregation* of markets bears a resemblance to the approach taken in the Commission Notice to grouping markets on the basis of supply-side substitution when most of the suppliers are able swiftly to offer and sell the various products.[125]

(2) The Relevant Geographic Market

The relevant geographic market is traditionally defined as comprising '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 homogeneous and which can be distinguished from neighbouring areas because the conditions of competition are appreciably different in those areas'.[126] Despite this somewhat general wording, the main objective of defining a market, also in its geographic dimension, is to identify those competitors that are capable of constraining the commercial behaviour of the company under consideration, in that they supply products (or are able to do so in a short time frame) that are sufficiently good substitutes for the product in question. Also in the geographic dimension, it is possible to distinguish between demand-side substitution and supply-side substitution (although the latter term is less often used in this context).

1.169

(a) Demand-Side Substitution

The analysis of demand-side substitution in the context of geographic market definition focuses on the extent to which customers in a given geographic area are able and willing to switch to suppliers located in other areas. The conceptual approach to geographic market definition can again be based on the SSNIP test. One has to assess to what extent the customers of a given product or group of products would switch to suppliers located elsewhere in response to a hypothetical small but significant (in the range of 5–10 per cent), non-transitory increase in the price of the products in the area being considered (prices in other areas held constant). If substitution would be enough to make the postulated price increase unprofitable because of the resulting loss of sales, additional geographic areas are included in the relevant market. This would be done until the set of geographic areas is such that the postulated price increase would be profitable.[127]

1.170

In order to establish whether companies in different areas constitute an actual alternative source of supply for consumers, a number of relevant factors can be taken into account, such as transportation costs for the products involved, the need for (locally provided) sales support or maintenance services, the importance of national or local preferences, purchasing habits of customers, and product differentiation. All these factors have an impact on the attractiveness of products offered outside the geographic market under consideration for customers located within the relevant market.

1.171

(b) Supply-Side Substitution

Supply-side substitution relates to the possibility for customers to turn to products that are not yet offered by particular competitors, but that would readily be offered in the event of

1.172

---

[124] US 2010 Horizontal Merger Guidelines, section 5.1 fn 8. See also G. J. Werden, 'Market Delineation Algorithms Based on the Hypothetical Monopolist Paradigm', US DOJ Antitrust Division Economic Analysis Group Discussion Paper No 02-8 (27 July 2002), section 7.

[125] This requirement is also captured by another characterization of the relevant market sometimes used, and according to which a relevant market is a product space in which the 'conditions of competition are sufficiently homogeneous'.

[126] Market Definition Notice, para 8. See also Case 27/76 *United Brands* [1978] ECR 207, para 11.

[127] Market Definition Notice, para 29.

a higher price of the product in question. In the context of geographic market definition, this relates to the possibility of suppliers located outside a certain geographic area (swiftly) to start supplying into that area. Thus, whereas demand-side substitution relates to the prospect of customers (or their agents) turning to other areas to obtain the product demanded, supply-side substitution relates to the prospect of outside suppliers turning to the area under consideration to start offering their products.

1.173   In this context, it is important to investigate the various supply factors to see whether those companies located in distinct areas face significant impediments to developing their sales on competitive terms throughout the geographic market. Possible impediments may result from requirements for a local presence in order to sell in that area, the conditions of access to distribution channels, costs associated with setting up a distribution network, and the existence or absence of regulatory barriers such as administrative authorizations and packaging regulations.

1.174   Whereas demand-side substitutability is often seen as being the main form of substitution in the context of product market definition, the relative importance of demand-side and supply-side substitution is probably more in balance in the context of geographic market definition. In the product dimension, supply-side substitution relates to the ability of companies swiftly to change production from one product to another. The product areas which lend themselves to such substitution are probably limited in number, and may well fall short of the number of cases where companies are able swiftly to offer, in a different area, products they already produce.

(3) Specific Issues in the Context of Market Definition

1.175   It is worth addressing three specific situations where care has to be taken in the context of market definition.

(a) Chains of Substitution

1.176   In certain cases, the existence of *chains of substitution* may warrant a definition of a single relevant market, even where products or areas at either end of the market do not directly compete with one another. Consider, for example, a product with significant transport costs such as construction materials. In such a case, deliveries from a given plant are limited to a certain area around the plant because of transport costs. Such an area could, in principle, constitute the relevant geographic market. However, if the distribution of plants is such that there are considerable overlaps between the areas around the different plants, it is possible that the pricing of those products will be constrained by a chain substitution effect: prices in one area constrain prices in an adjacent area, which in turn constrain prices in another area (not adjacent to the first). If the 'chain' that links the three areas is strong enough, it would be appropriate to define the relevant market as including these three areas. Note that application of the SSNIP methodology would indeed identify the relevant market as such, whereas an overly strong emphasis on factors such as transport costs would not.

1.177   Chains of substitution may also be relevant in the context of product market definition.[128] Suppose that products A and C are single-purpose software programs each suitable for doing a different computing task and that product B is a dual-purpose software product that can be

---

[128]   It should be noted that chains of substitution are probably less prevalent in the context of product market definition than in the context of geographic market definition.

used for both tasks. Even if products A and C are not direct demand substitutes, it is appropriate to view them as belonging to the same relevant product market when their respective pricing is sufficiently constrained by substitution to product B (in the sense of the SSNIP concept) and vice versa. Product B can then be seen as forming the 'link' between products A and C.

(b) Price Discrimination

In certain markets, it is possible for suppliers to engage in price discrimination, that is, to charge different prices to different customers depending on their buyer characteristics.[129] Price discrimination is possible when suppliers can (explicitly or implicitly) identify to which group an individual customer belongs at the moment of selling the relevant product, and trade among customers (or arbitrage by third parties) is not feasible. In such cases, demand substitution (the ability of customers to obtain substitute products, or to obtain them elsewhere on better terms) may be impaired. If also supply substitution is difficult or impossible, it is appropriate to define a market by reference to the group of customers who may be the subject of such price discrimination. In terms of the SSNIP principle: the possibility of targeting a price increase raises the likelihood of such a price increase being profitable. **1.178**

Importantly, the chain of substitution effect described in the previous subsection no longer holds as a factor linking together distinct products or geographic areas when price discrimination is possible. For example, in the context of the software products example, if a hypothetical sole supplier of products A and B could identify customers by their specific software needs, it could increase price on the A product and, to those customers who are in need of the software function performed by A, also on the dual-purpose B product. It cannot even be excluded that, for a hypothetical sole supplier of products A and B, raising price on A alone might be profitable (some customers may switch to product B, but, given that B also belongs to the hypothetical monopolist, this need not be problematic). In market definition, the operational response to the possibility of price discrimination is to define markets by reference to the group of customers who may be the subject of such price differentiation (in the example, the customers in need of the software function performed by A). **1.179**

In the context of geographic markets, it is often the case that customers located close to the border are familiar with trading conditions across the border and ready to obtain the products needed there. Similarly, outside suppliers located near the border may be relatively quick at supplying across the border when the opportunity arises. When there is great demand- and supply-side substitutability at the borders, this would point towards a geographic market that is wider than the area delineated by the border if the SSNIP test is applied with a uniform price increase of 5 per cent in mind. An issue to be checked in such cases is whether a sole owner of the production or supply locations in the area could practise geographic price discrimination (in other words, whether a uniform price increase over the area is the appropriate benchmark). If the location of the production or supply locations is such that prices further inland could be different from (ie higher than) those near the border, then the area under consideration might be a relevant market after all.[130] In such a case, it might be **1.180**

---

[129] Bidding markets may be examples of markets where price discrimination is possible. In essence, these are markets where companies compete for specific contracts and where each customer receives, or may receive, a personalized offer.

[130] While it is true that customers located further inland could turn to the (lower priced) areas near the border, when these border areas are also under the control of the hypothetical monopolist, the incentive on the

appropriate to define the relevant geographic market as the original area under consideration, not wider.

(c) Captive Production

1.181 The definition of the relevant market involving intermediate products is often fairly complex. Intermediate product markets may feature both specialized producers and integrated producers captively producing all or a sizeable proportion of their output for internal use. The competitive constraints on a non-integrated supplier in such a market situation are not just the demand substitution possibilities of its customers (whether integrated or non-integrated), but also the supply possibilities of integrated producers who are currently only participating in the merchant supply a little, if at all.

1.182 In defining product markets for intermediate goods, it is customary first to focus on what is called the 'merchant market', that is, that part of the product market for which transactions take place between entities not belonging to one and the same group. This is because of the idea that, in response to a reduction in supply by any given company active on the merchant market, other non-integrated suppliers can normally be assumed to exert a competitive constraint by increasing their supply, whereas an integrated company may be more reluctant to increase supplies on the merchant market (if it is already active on it) or to become active on it (if it is not yet active).[131] Even when one decides that the integrated firm is likely to increase supplies or to become active, the question remains how much of its sales or capacity to take into account.[132] These factors make it appropriate to pay attention to the merchant market as such, especially at the early stages of the investigation.

1.183 At the same time, it is important not to lose sight of the relationship between captive sales and merchant sales in the overall market. It has to be realized (in the application of the SSNIP test) that the incentive to raise the price on the merchant market becomes less, the more the customers of such suppliers (the non-integrated downstream companies) would lose sales and market share to the integrated companies which would not be confronted with an increase in the price at which they can obtain input supplies. If it is the case that raising the price for merchant supplies would be unprofitable in view of the strong presence of integrated suppliers, this would plead in favour of looking at the captive sales and merchant sales as a whole. All in all, the best response to the complexity of market definition in the context of intermediate goods would seem to be to consider both possible market definitions (merchant market and combined market) and, when the companies involved have important market positions on either market, to proceed to a full analysis of competition.

---

part of the hypothetical monopolist to raise prices inland is higher than when the border areas are not under its control.

[131] The integrated firm's decision whether to (increase) supply on the merchant market is also a function of the impact this has on the profitability of its business activities further downstream (the stage that uses the intermediate product as an input). Such an impact may exist not only where increasing supplies into the merchant market implies cutting back on the internal use of the intermediate product (and, hence, reducing output of the downstream subsidiary), but also where supplying more of the intermediate product means more competition for the downstream subsidiary from non-integrated downstream rivals using the intermediate product.

[132] As with supply substitution in general, it would make sense to take the integrated firm's sales or capacity into account in measuring market shares only to the extent that the firm would be able and willing to respond to an increase in price in the merchant market (part of the firm's capacity may be committed or more profitably employed internally). See also Section E.1(d).

(4) Further Considerations

(a) Market Definition in Practice

At the start of this section, it was mentioned that market definition serves the purpose of putting the assessment of market power in a proper context. The more alternatives are available to customers, the less market power the companies supplying a given product are likely to have. Even when companies have a 'monopoly' on a given product, they may not have market power over that product when sufficient alternatives are present. By contrast, when there are few alternatives, it is opportune to see whether any particular company, or group of companies, has market power.

1.184

In many cases, the starting point of market definition is to describe clearly the product or service in question and to think of various conceivable markets. This then permits one to decide, from a summary examination of market shares on the various conceivable markets, whether in relation to the operation under analysis there are any competition issues, even on the narrowest conceivable market.[133] This allows for a first screening of cases, to see whether there may be competitive issues.

1.185

Having determined that an accurate market definition is needed, the SSNIP methodology suggests the following line of inquiry: start with the product under consideration and assess what proportion of the customers would switch, in whole or in part, from the product if its price were to be raised by a small but significant proportion, and to which substitutes would they switch. To obtain a first indication, an inquiry into the opinions, primarily of customers but also of competitors, can be undertaken concerning the extent to which the products under consideration are adequate substitutes. The accuracy of the inquiry can, in subsequent stages, be improved by addressing more customers and competitors (a wider base of respondents) and asking for more specific information. In this context, evidence of customer switching in the past would be particularly informative. Data on price-cost margins can shed further light on the question whether a 'hypothetical monopolist' would find it profitable to raise the price.

1.186

Various additional quantitative and empirical methods are available that can provide information on the degree to which products face demand-side constraints. These methods include the analysis of prices and price movements of the products under consideration to see to what extent they move together over time, the estimation of price elasticities, critical loss analysis, event analysis (to see whether particular events in the past shed light on the question of which products compete with one another), and the analysis of bidding data. These methods are presented in further detail in Section G.

1.187

(b) Defining the Market: Not an End in Itself

While one may debate the various alternative approaches to market definition, the essential point is that the market defined must seek to include the products (and the firms producing them) that represent a competitive constraint on the product(s) in question. Often the difficult issue in market definition is that, whatever the operational formulation or the test employed, the appropriate boundaries of the market cannot be decided precisely. Market

1.188

---

[133] This is not to say that the narrowest market is necessarily the one where the parties' market shares are the highest. After all, when product markets are defined very narrowly, there may be no competitive overlap in the first place. It also remains important to look at wider possible market definitions.

definition will indicate which products provide an immediate constraint on the product under investigation, but not that all these products are of equal constraining influence. In the context of differentiated product markets in particular, the issue of differing degrees of competitive pressure between products, even within one and the same relevant market, may be of great importance.[134]

1.189 The boundaries within which competition is at work cannot be fully captured by the classification of products into different 'markets'. This merely recalls the fact that market definition is not a goal in itself but an intermediate step for structuring the analysis. The aim of market definition is to analyse the economic substitutability of products in a structured way, not to represent a full analysis of competition among the companies supplying the products.

F. Market Power and Dominance[135]

1.190 Market power is often broadly referred to as the power to raise price above the competitive level. While the general idea behind this characterization is fairly clear, a number of comments can be made. The section starts by addressing some of the questions the concept of market power raises. It focuses in particular on the question of how to identify market power in a given market context. It then addresses the relationship between market power and the concept of 'dominance', as it is known in Article 102 and the EU Merger Regulation. The section ends with a discussion of ways in which market power may be maintained or enhanced through anti-competitive means, which is the main focus of competition policy.

1.191 From the outset it is important to clarify that market power is not a negative thing per se. Often companies obtain market power in entirely legitimate ways, for example by producing more efficiently than other players, by making better quality products, or by being more innovative—in short by providing benefits to consumers. Consequently, competition policy is not concerned with market power as such. Rather, it is concerned with the ways in which market power may be obtained, maintained, or enhanced (and subsequently exercised) through anti-competitive means, that is, to the detriment of consumers. While this in principle requires an individual analysis for each case, this does not preclude antitrust policy from relying upon certain presumptions regarding the effects once a certain degree of market power has been established. For instance, based on past experience it may be considered that certain specific types of conduct are so likely to increase or maintain market power when the firm already possesses market power that a (negative) presumption is warranted.[136]

---

[134] See also Section F.1. This insight has led to an increase in popularity of methods such as UPP (upward pricing pressure), primarily in the context of mergers with differentiated products. UPP focuses directly on the change in pricing incentives of the merging parties and may avoid the need to define markets in a context where this is.indeed inherently difficult. See Section G.4 for more details.

[135] This section builds on a text written by Kirti Mehta for the first edition of this book and, for the discussion of unilateral effects and tacit coordination, on V. Verouden, C. Bengtsson, and S. Albæk, 'The EU Notice on Horizontal Mergers: A Further Step Towards Convergence' [2004] Antitrust Bulletin 243.

[136] J. B. Baker and T. F. Bresnahan, 'Economic Evidence in Antitrust: Defining Markets and Measuring Market Power' in Buccirossi, *Handbook of Antitrust Economics* (n 47), 15.

(1) Market Power

(a) Concept

Market power can manifest itself in a number of dimensions, such as high prices, reduced output, reduced choice and quality, or diminished technological innovation. The former dimensions—price, output, and choice—are normally at the centre of the analysis as regards the static welfare impact of a given merger, agreement, or conduct. The latter dimensions—quality and innovation, but also choice—are of particular importance when it comes to assessing the dynamic welfare impact.

1.192

While the dynamic perspective of market power is arguably of great importance, antitrust analyses typically start by considering whether a company has (or will obtain) static market power. After all, without market power in the static sense, it is relatively unlikely that a company has market power in the dynamic sense.

1.193

The static notion of market power concentrates on the power to raise price above the competitive level. From a short-term perspective, the competitive price level is often taken to mean the marginal cost level. Market power then refers to the *ability* to sell a product at a higher price than it actually costs to produce at the margin. Where a company actually charges such a higher price it is said to *exercise* market power. If a company exercises market power, this implies that there is a certain welfare loss (also called *inefficiency*) stemming from the fact that some customers do not obtain the product although they have a willingness to pay for the product that is higher than it actually costs the company to make the product. From a longer term perspective, the competitive price level is often taken to mean the average cost level, where the cost benchmark includes a reasonable rate of return on investment.[137] Market power then refers to the ability to make supra-normal profits, that is, profits that are higher than customary in similar market settings, over a sustained period.

1.194

Obviously, any company can raise the price at which it sells. What is meant by the ability to raise price above the competitive level is the ability to do so *profitably*. This is only possible for a firm that does not face such pressure from its competitors that any reduction in its own output is easily made up for by the competitors. In such a case, the sales loss facing the company when it raises price above the competitive level is limited,[138] and increasing price above the competitive level may be profitable. The less competitors pose a competitive constraint on the firm in question, the more that firm is said to have market power. It follows that a situation of market power is unlikely to arise in a market where expanding output (or indeed entry) is easy, since in such conditions the pressure on prices charged by the incumbent firm(s) is rather persistent.

1.195

There are, in essence, three principal reasons why competitors may not easily make up for a reduction in output of the firm with market power. The first is product differentiation. Product differentiation means that the products that are being offered are imperfect substitutes for each other.[139] When rival producers offer alternative products, but these products

1.196

---

[137] The cost benchmark is sometimes taken to be forward-looking (ie what would it cost to start production now, with current technology).

[138] ie the demand faced by the firm is inelastic. Put differently, the firm-specific demand curve is downward-sloping (and not flat, as would be the case with perfect competition).

[139] Products may be differentiated in various ways. Differentiation may be based on brand image, technical specifications, product quality, or level of service. It may also find its origin in buyers having to incur switching costs to use a competitor's product. There may also be differentiation in terms of geographic location, based

are not as attractive as the ones offered by the firm raising the price (at least from the viewpoint of the customers of the firm raising the price), customers may prefer to stay with that company even when it raises the price. As a result, the firm in question has a certain leeway, or margin of manoeuvre, in its pricing behaviour. The more the products offered by competitors are close substitutes (or the more easily competitors can reposition their products), the less market power a company is likely to have.

1.197 The second reason why a company may have market power is that rival suppliers, even if they are offering equivalent or similar products to those of the company with market power, are not capable of supplying more in response to a price increase by the firm in question. A prime example is the situation where rivals have capacity constraints or face other barriers to expansion.[140] Rivals may, for example, have insufficient access to input supplies, relevant infrastructure, or distribution networks to provide a supply response. Also in these situations, the firm in question has a certain leeway to increase prices.

1.198 A third important source of market power is differences in productive efficiency. Where economies of scale or scope are important, a company with high production levels is able to produce more cheaply at the margin than companies operating at sub-optimal levels.[141] This source of market power translates into the inability of rivals to compete at low prices and allows the company with the cost advantage a possibly considerable margin to set prices.

1.199 An extreme case of market power is the situation where a firm has a monopoly on the relevant market, so that there are no rival companies to constrain the firm. A monopoly may be seen as entailing all or some of the sources of market power mentioned previously: strong product differentiation (the product in question basically forms a relevant product market by itself), inability of rivals to provide a supply response (entry barriers[142]), or substantial efficiency differences (no rival is able to supply at competitive prices).

1.200 The fact that setting price above competitive levels is only possible for a firm that does not face such pressure from its competitors so that any reduction in its own output is easily made up for by competitors, suggests an alternative (but equivalent) way of thinking of market power. In this perspective, market power relates to the ability of a firm significantly to influence, through its own output level, the aggregate output of the market.[143] The characterization captures quite well the three sources of market power previously identified. Where products are differentiated, the reduction of output by one firm is likely to lead to a reduction of aggregate output, since other rivals' products are not able to make up for the difference. Similarly, where competitors have capacity constraints or face other barriers to expand output, the reduction of output by one firm is likely to have an impact on total output. Finally,

---

on branch or store location. Eg location matters for retail distribution, banks, travel agencies, and petrol stations. Note that products can be imperfect substitutes even where they are part of the same relevant market. Substitutability is a matter of degree.

[140] More generally, rivals face increasing marginal cost levels when production levels go up.

[141] See Section C.2. As already indicated, market power is not in itself a bad thing. This holds, in particular, where the market power stems from superior efficiency.

[142] Monopoly positions are normally linked to barriers to entry, such as legal barriers to entry (patents on technology, brand names, statutory monopolies), technological barriers (extreme economies of scale or scope), or strategic entry barriers (related to the incumbent firm's behaviour or reputation). See Section C.2.

[143] cf B. Klein, 'Market Power in Antitrust: Economic Analysis after Kodak' (1993) 3 Supreme Court Econ Rev 43, 76. See also J. Azevedo and M. Walker, 'Dominance: Meaning and Measurement' (2002) 23 Eur Comp L Rev 363.

where a company has a significant cost advantage it can, at least within certain boundaries, determine the output level in the market.[144]

The previous discussion focused on the competitive constraints that may be exerted by actual or potential competitors. Competitive constraints may in certain circumstances also be exerted by customers, however. Even a company with a high market share may not be able to raise price when its customers have sufficient bargaining strength. Countervailing buyer power may stem, for instance, from the customers' size or their commercial significance for the company concerned, their ability to promote new entry or capacity expansion (eg by concluding long-term contracts giving rivals the prospect of significant sales volumes), to integrate vertically, and/or credibly to threaten to do so.[145] If countervailing power is of a sufficient magnitude, it may deter or defeat an attempt by the company profitably to increase prices.[146] **1.201**

(b) Identification of (Static) Market Power

The usual starting point for determining whether a company has market power is to consider the relative position of the company vis-à-vis its competitors on the market. Market shares, the main indicators used in this respect, often give at least some indication of the degree to which companies have, or do not have, market power. **1.202**

Market shares are used extensively for the purpose of identifying market power, not only because they are relatively simple measures, but also because the more direct methods to measure market power are difficult to use. The microeconomic definition of static market power—the ability to raise price over cost—suggests that one looks at the profit margin of a firm to find out whether this firm has market power. For example, the gross margin is a measure of the degree by which a firm's price exceeds marginal cost. This margin, while in principle ascertainable, is often difficult to assess in practice. Accounting costs, that is, the costs as they appear in the company's accounts, need not be accurate measures of the costs involved in producing additional units of output, which is the relevant economic benchmark.[147] Accounting costs are often based on aggregate costs calculated over the entire production, rather than cost levels at the margin. In addition, those costs that cannot be directly attributed to the production of a specific product or service (where common production factors are involved) are normally attributed according to standard accounting rules that have little connection with what it costs to increase production. **1.203**

If instead the elasticity of demand facing the firm is known with some precision, then that information could give some indication about the firm's margin. This idea underlies the **1.204**

---

[144] It is only after the price has risen above a certain level that the other companies become competitive and may start to produce or increase output.

[145] As becomes clear from this list of factors, the concept of customer bargaining power is closely linked to the ability and incentive of competing firms to enter or expand (in this case, with the assistance of the customers).

[146] Buyer power may not, however, be considered a sufficiently effective constraint if it only ensures that a particular or limited segment of customers is shielded from the market power of the dominant undertaking. Cf Article 102 Enforcement Priorities Guidance, para 18.

[147] The marginal cost of production and, on a longer perspective, the incremental cost of production are the relevant benchmarks when assessing economic efficiency. These cost concepts identify how much it actually costs to produce more at the margin and whether customers are left unserved, whose willingness to pay for the product exceeds the cost of production but who do not purchase the product as their willingness to pay is below the price charged.

so-called Lerner index, which is defined as the firm's gross margin in relation to the current price set by the firm. Economic theory predicts that, at its profit-maximizing point, a firm's margin will be the reciprocal of the elasticity of demand it faces: where the demand elasticity is low, the firm's margin is high, and vice versa. However, the elasticity of demand facing a firm is often not known either, at least not with sufficient precision.

1.205 Even when one has (directly or indirectly) established the relevant margin, the question remains as to what is a high margin. Industries may feature different gross and net margins depending on the level of fixed investment, the stage of the industry (growing or stagnant), or the degree of risk involved (the fact that margins turn out to be high *ex post* may be proper compensation for the risk incurred *ex ante*). Pharmaceutical companies, for example, often feature high gross and net margins on a limited number of products because R&D expenditure is both significant and risky, and marginal costs of production are low or negligible. One would therefore need to make comparisons with appropriate benchmarks, preferably from the same sector but in a different geographic area. It is clear, however, that finding appropriate benchmarks is one of the more difficult issues in the identification of market power through margin analysis.

1.206 Market shares are, by contrast, comparatively simple indicators of market power. The main question is, of course, how good market shares are as an indicator for market power. On the one hand, they are likely to contain some information on the competitive strength of each of the market players. In a competitive market with many players, each firm tends to be a price taker, that is, acts as if facing an infinitely elastic demand curve, irrespective of whether the total market demand is price elastic or inelastic. Similar technology, absence of scale or scope economies, and the commodity nature of products all tend to ensure that many companies are active in the market and that none of these firms has a high market share. Where, however, certain firms have relatively high market shares this may be an indication that such firms are either cost leaders or have product advantages in a differentiated product market. Alternatively, it may reflect a difference in production capacities. In such cases, the practical approach based on market shares can be considered a useful, if approximate, way of identifying firms with market power.

1.207 On the other hand, the observation that one or more companies in a market have significant market shares is compatible with a whole range of market settings, both competitive and less competitive ones. To take one example: a company may have a high market share for merely historical reasons and lack the ability to raise prices above any competitive level because other market participants face no problems in expanding output in response to a price increase by the former company.

1.208 Whether it is appropriate to use market shares as a proxy for market power also depends heavily on the quality of the definition of the 'relevant market'. In differentiated product markets, in particular, the degree of competition between the respective products may vary in ways not represented by market shares. It may well be that the company with the highest market share faces more competition than niche players with lower market shares. Market shares do not tell how close a substitute one product is vis-à-vis another product.

1.209 Another example of a market where market shares may be less informative is bidding markets. The fact that other firms did not make a sale in a particular bidding contest does not mean that these firms did not pose a significant competitive constraint on the winning

firm.[148] In addition, the link between market share and market power is probably less direct in bidding markets than in most other markets.[149] In bidding markets, each customer receives, or may receive, a personalized offer. Where this is the case, companies can decide to compete more aggressively on the margin, without this necessarily having a direct impact on the margins obtained on their existing customer base. When individual contracts are large and infrequent, the incentive to compete for each of them may be especially strong. Similar arguments can be raised in contexts where competition is 'for the market' instead of in the market.

It is clear that both the approach based on relative market positions and the more direct measurements of market power have certain drawbacks. In identifying whether a company has market power, it therefore remains indispensable to focus on the causes of market power, to focus on those factors that enable the company to raise price: the degree of product differentiation in the market, the presence of barriers to entry and expansion on the part of rivals, and differences in productive efficiency. It is only when such factors are present that one can persuasively say that a company has market power.   **1.210**

(2) Dominance

As the name suggests, the term 'dominance' refers to a strong form of market power. A distinction is commonly made between two forms of dominance: single dominance and collective dominance. The first refers to a situation where a single company has substantial market power, the second to a situation where a group of companies jointly hold such market power.   **1.211**

(a) Single Dominance

The traditional characterization of the term 'dominant position' in EU competition law is that it relates to a 'position of economic strength enjoyed by an undertaking which enables it to prevent effective competition being maintained on the relevant market by giving it the power to behave to an appreciable extent independently of its competitors, customers and ultimately of its consumers'.[150]   **1.212**

The latter part of this definition, referring to 'the power to behave to an appreciable extent independently of its competitors, customers and ultimately of its consumers' is closely related to the three factors giving rise to market power discussed in the previous subsection. Product differentiation in the market, the presence of barriers to entry and expansion on the part of rivals, and differences in productive efficiency all may provide a given company with substantial leeway in determining prices. In this context, it is important to recognize that no company can act entirely independently of competitors, customers, and consumers. It is only natural that a company, even when it is dominant, takes account of the fact that competitors may produce a bit more if it raises its price. Similarly, it will realize that customers are likely to consume less when the price goes up (the 'discipline of the demand curve'). Whether a company has market power and has the ability to set the price above the competitive level is a matter of degree. Whereas, legally speaking, a company either is or is not dominant, it is   **1.213**

---

[148] The same applies when considering market shares on an annual basis where the number of bidding contests in any given year is small. When the number of bids increases, one can expect market shares better to reflect competitive strength.

[149] For a critical analysis of this argument, see P. Klemperer, 'Bidding Markets' in Buccirossi, *Handbook of Antitrust Economics* (n 47).

[150] Case 27/76 *United Brands v Commission* [1978] ECR 207.

important to realize that, from an economic standpoint, the underlying variables determining the degree of market power form a continuum.

1.214 The Commission's Article 102 Enforcement Priorities Guidance specifies in this context that dominance relates:

> 'to the degree of competitive constraint exerted on the undertaking in question. Dominance entails that these competitive constraints are not sufficiently effective and hence that the undertaking in question enjoys substantial market power over a period of time. . . . The Commission considers that an undertaking which is capable of profitably increasing prices above the competitive level for a significant period of time does not face sufficiently effective competitive constraints and can thus generally be regarded as dominant.

The Guidance continues to explain that the assessment of dominance will take into account the competitive structure of the market, and in particular the following factors: (a) constraints imposed by the existing supplies from, and the position on the market of, actual competitors; (b) constraints imposed by the credible threat of future expansion by actual competitors or entry by potential competitors; and (c) constraints imposed by the bargaining strength of the undertaking's customers.[151]

1.215 The first element to consider when assessing dominance is probably the market share of the firm in question. In its Guidance, the Commission states that:

> low market shares are generally a good proxy for the absence of substantial market power. The Commission's experience suggests that dominance is not likely if the undertaking's market share is below 40 per cent in the relevant market. . . Experience suggests that the higher the market share and the longer the period of time over which it is held, the more likely it is that it constitutes an important preliminary indication of the existence of a dominant position . . . However, as a general rule, the Commission will not come to a final conclusion as to whether or not a case should be pursued without examining all the factors which may be sufficient to constrain the behaviour of the undertaking.

In other words, the importance of market shares is qualified by the extent to which they convey information on the ability of rivals to provide a competitive constraint on the dominant firm. The Court has held in this respect that, although the importance of market shares may vary from one market to another:

> the view may legitimately be taken that very large shares are in themselves, and save in exceptional circumstances, evidence of the existence of a dominant position. An undertaking which has a very large market share and holds it for some time, by means of the volume of production and the scale of the supply which it stands for—*without those having much smaller market shares being able to meet rapidly the demand from those who would like to break away from the undertaking which has the largest market share*—is by virtue of that share in a position of strength which makes it an unavoidable trading partner and which, already because of this secures for it, at the very least during relatively long periods, that freedom of action which is the special feature of a dominant position.[152]

1.216 Thus, dominance is said to exist only when the situation of substantial market shares is expected to be sustained over a period of time during which rival firms and entrants cannot be expected to bid away the dominant firm's market share through lower pricing and superior quality products and where there is insufficient countervailing power on the side of buyers.

---

[151] Article 102 Enforcement Priorities Guidance, paras 10–12.
[152] Case 85/76 *Hoffmann-La Roche v Commission* [1979] ECR 461, para 41 (emphasis added).

In general terms, the main factors that are taken into account in determining dominance all relate to the *ability* and *incentive* of the smaller competitors to increase their production or otherwise provide a constraining force.[153] Economies of scale and scope, control over input supplies, patents, or distribution networks, and other strategic advantages for the dominant firm (eg branding and reputation) are the more important ones in this respect. Such factors may make expansion of smaller firms or entry of new competitors difficult, either in the short term (eg when the new entity controls input supplies) or in the long term (eg when the possession of patent portfolios reduces the ability of competitors to innovate). Similarly, these factors may discourage smaller rivals from expanding and thereby affect their incentive to provide a competitive constraint. Here one can think, for instance, of cases where the company in question has control over the main distribution networks in an industry, leading to a significant reduction in competing rivals' incentive to invest in marketing effort or R&D. In such circumstances, an asymmetric market structure may prevail in which one firm dominates production and is the principal decision-maker with market power. **1.217**

It is worth noting that in economics there also exists a concept called the 'dominant firm', but that it has a meaning that is often more specific than the one commonly used in the EU competition context. It refers to a market situation where a single large actor faces a number of fringe competitors (often called the model of the 'dominant firm and the fringe').[154] In this model, the fringe competitors are price takers, so that they supply up to the point where their marginal costs equal the market price.[155] By contrast, the single large firm dominates production of the final good because of a cost advantage and acts strategically with respect to the fringe. The situation of a dominant firm in the total market can be depicted as in Figure 1.13. The dominant firm faces a firm-specific demand curve (ED, in the right-hand graph) which is obtained by deducting from market demand (DD, in the left-hand graph), at each price, the supply responses of all the other firms (SS) in the market. The dominant firm would maximize profits by producing where its marginal costs equal its marginal revenue; this is at the output Q_1, which implies the price P_1, leaving the balance of the output being produced by smaller firms. The small firms accept the price set since their supply response is limited (their marginal costs are increasing and exceed, at some point, the price charged by the dominant firm). **1.218**

(b) Collective Dominance

Collective dominance refers to a situation where a group of companies jointly hold market power. The companies derive this market power, on the one hand, from the fact that other firms in the market cannot challenge the joint market position and, on the other hand, from the fact that the firms have the ability (and probably the incentive) to suppress or limit competition among themselves by colluding. **1.219**

---

[153] As indicated in the previous section, even the concept of buyer power is closely linked to the ability and incentive of competing firms to enter or expand (in this case, entry or expansion with the assistance of customers).

[154] The model is due to K. Forchheimer, 'Theoretisches zum unvollständigen Monopole' (1908) 32 Schmollers Jahrbuch für Gesetzgebung, Verwaltung und Volkswirtschaft 1. See also M. Riordan, 'Anticompetitive Vertical Integration by a Dominant Firm' (1988) 88 Am Econ Rev 1232.

[155] The fringe players do not assume that their individual actions have an influence on the price level in the market, so that the marginal revenue of supplying more equals the market price.
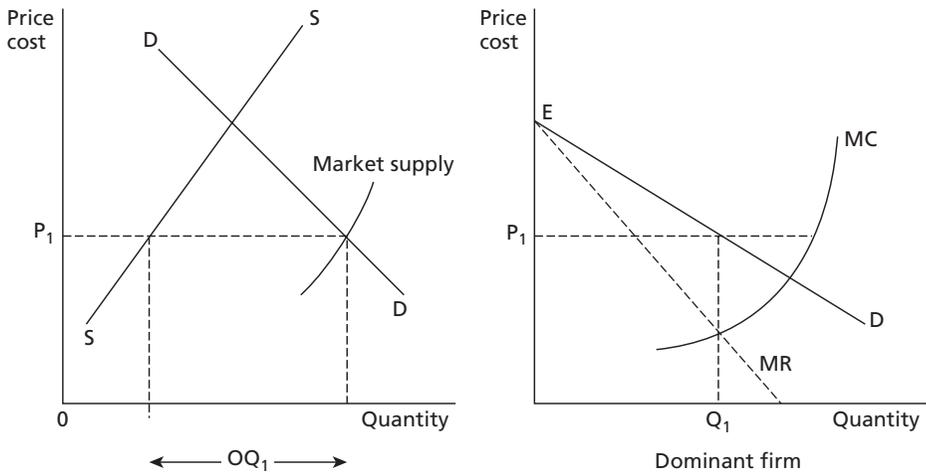
Figure 1.13 A dominant firm

1.220 Where the companies collude to raise prices, they can be said to *exercise* their joint market power.[156] Where companies collude to exclude other firms from the market (eg through a system of vertical agreements or rebates), such collusion may work to *increase or maintain* their joint market power.

1.221 Collusion can be understood either as explicit coordination to engage in a certain market conduct (eg coordination on price by way of explicit communication or a collusive agreement, even if it is non-enforceable in court) or as tacit coordination. The underlying mechanisms are to a large degree the same.[157]

1.222 As explained in Section C.5, the theory of coordination is anchored in economic models that explain how competitors can cancel the mutual competitive pressure by a coherent system of implicit threats. In a non-collusive setting, each competitor constantly has an incentive to compete. This incentive is ultimately what keeps prices low, and what prevents firms from *jointly* maximizing their profits. Coordination emerges when this short-run incentive is overruled by a stronger long-term incentive: each firm in the market exercises a self-imposed competitive restraint in the short run only because it knows that this restraint will be 'rewarded' in the long run by the other firms exhibiting similar restraint.

1.223 Coordination on prices is more likely to emerge in markets where it is fairly easy to establish the terms of coordination and where such coordination is sustainable.[158] Sustainability requires that there is: (a) sufficient market transparency, so that the coordinating firms are

---

[156] One could debate whether coordination on price can itself be viewed as a way to *achieve* market power (as is argued by P, Hofer, M. Williams, and L. Wu, 'Principles of Competition Policy Economics' [2004] Asia-Pacific Antitrust Rev 4) or that it rather should be viewed as the expression of the *exercise* of market power. Although ability and effect are difficult to disentangle in the context of collusion, it is true that, in principle, the ability to overcome the prisoner's dilemma, and thereby being able to collude, is not the same as actually colluding.

[157] The economic literature on tacit coordination, setting out the conditions under which coordination on price is feasible, is relevant to all forms of coordination that have to be maintained through means other than legally enforceable contracts, ie also to most forms of explicit coordination.

[158] Horizontal Merger Guidelines, para 41.

able to monitor to a sufficient degree whether the terms of coordination are being adhered to; (b) the existence of a disciplining mechanism to ensure adherence to the coordination; and (c) the absence of possible actions of outsiders, such as current and future competitors, as well as customers, that can jeopardize the results expected from the coordination.

The degree to which these conditions are fulfilled all vary with the characteristics of the firms, **1.224** markets, and products concerned. The reader is referred to Section C.5 for an overview of the most relevant factors.

(3) Enhancing Market Power

The existence of market power, in particular static market power, on the part of a single **1.225** firm is not a negative thing per se. As indicated in the introduction, companies often obtain market power in entirely legitimate ways, for example by producing more efficiently than other players, by making better quality products, or by being more innovative. Indeed, the prospect of obtaining (some) market power is a major determinant for companies to invest in product and process innovation. Consequently, competition policy is normally not concerned with market power as such. Rather, it is concerned with the ways in which market power may be maintained or enhanced through anti-competitive means.

One commonly distinguishes between two main ways in which companies may enhance **1.226** their (individual or collective) market power: through merger, and through exclusion.

(a) Merger with a Competitor: Unilateral vs Coordinated Effects

While the existence and extent of any negative impact of a merger on competition will **1.227** depend on many factors (eg the market position of the companies concerned, the strength of the competitors, the nature of the products, efficiencies), the immediate reason why a merger can have a negative impact is often the same: a merger may diminish the degree of competition in a market by removing important competitive constraints on one or more sellers, which consequently find it profitable to increase prices (or to reduce output, or to take other action to the detriment of consumers).

The first competitive constraint being removed is that which previously existed between **1.228** the merging firms. Whereas, before the merger, the merging parties exercised a competitive constraint on each other, in the sense that, if one party were to raise price, it would lose customers to the other party and vice versa, the merger lifts this particular constraint: part of the sales lost due to a price rise on one product will now flow to the product of the merger partner and, as a result, such a price increase may be profitable, while it would not have been profitable prior to the merger.

To illustrate, let us consider the example of high-quality cars and let us imagine that German **1.229** purchasers of cars essentially make a choice between brands A (say, an Audi), B (a BMW), M (a Mercedes), and V (a Volvo).[159] A reasonable starting point for any market analysis is to assume that pre-merger all producers are marketing their cars in a profit-maximizing way. Car manufacturers may pursue varying strategic objectives,[160] but let us assume that each producer tends to choose a selling price that is optimal in view of what the other producers

---

[159] This is a highly stylized example, which ignores, eg, the presence of other car manufacturers in the high-quality segment and the fact that each car manufacturer typically has several models within this segment.
[160] eg a market penetration or a product positioning objective.

are charging for their product. Accordingly, a reason why, for example, Audi is not charging more for its cars is that it realizes that it would probably lose too many sales to the other three producers. The reason that it does not decrease its price is that it would lose margin and not sufficiently increase volume. Each producer makes a trade-off between volume and margin.

1.230 How would a merger between, for instance, Audi and BMW change the picture? Suppose that Audi and BMW were competitively interdependent in the following way: if Audi were to increase prices by 10 per cent, half of the customers who would stop purchasing an Audi would instead purchase a BMW.[161] Similarly, if BMW were to increase prices by 10 per cent, one-third of the customers who would stop purchasing a BMW would instead purchase an Audi. The merger would change the marketing strategy of the new company fundamentally. After all, in deciding on the price of the Audi model, the fact that half of the Audi customers who would be lost following a price increase on Audi would turn up to buy a BMW would be a rather comforting thought for the new company's management. In the absence of other factors (such as new entry or the realization of efficiencies through the merger), the likely result of the merger would be an increase in the price of an Audi and, by analogy, also of a BMW.[162]

1.231 Such effects are not conditional on competitors changing their way of interacting in a given market (eg by starting to coordinate) but are instead the consequence of the merged firm's optimal response to the new market configuration where the merging firms no longer compete. The merged firm's behaviour is profitable *even if* rivals continue to compete in the same way as they would have done in the absence of the merger. Accordingly, such merger effects are often called 'unilateral' or 'non-coordinated'.[163]

1.232 This is not to say that competitors cannot also benefit from reductions in competitive pressure as a result of a unilateral price increase by the merging companies. In a way, a merger takes out a source of competition in a market. Other firms' likely responses to this may also be to increase prices, albeit perhaps to a lesser extent. Therefore, the incentive to raise prices on the side of the merging firms may lead to price increases for all firms present in the same market.

1.233 To come back to our car example: the moment that, as a result of the merger, both Audi and BMW have become more expensive, more customers will show up at the doorsteps of the Mercedes and Volvo dealers. The management of those two companies, confronted with more demand for their products, would make the usual trade-off between volume and margins. They would be likely to increase their prices and margins (even if a little), so as to benefit optimally from the increase in demand they face.

1.234 While competitors may react by raising their prices, it is important to note that it is not these reactions that make the unilateral price rise profitable in the first instance. In the case

---

[161] In other words, suppose that the *diversion ratio* from Audi to BMW (following a price increase of the Audi model of 10 per cent) equals 50 per cent.

[162] The point to remember is that an extra factor—the sales captured by the other model—enters into the pricing equation of the new entity, changing its pricing incentives for each model.

[163] The term 'unilateral' might leave the impression that the effects only relate to actions of a single firm, ie the merged entity. As will be developed further in the following paragraphs (1.232–1.233), competitors may also change their price or output levels in response to a merger. For this reason, some have suggested that 'unilateral effects' are better referred to as 'multilateral effects'. See eg J. Vickers, 'Competition Economics and Policy', speech delivered at Oxford University, 3 October 2002 (available at <http://www.oft.gov.uk/news-and-updates/speeches/2002/0702>). A decisive factor for effects to be 'unilateral' (or 'multilateral') is that they do not depend on companies in the market starting to coordinate. For this reason, the Horizontal Merger Guidelines and the Merger Guidelines of the UK Office of Fair Trading (OFT) use the term 'non-coordinated effects'.

of unilateral effects, the incentive of the new entity to raise its price stems entirely from the elimination of the competitive constraints that the two merger companies exercised on each other pre-merger, not from the new firm anticipating that its competitors will raise prices. While the magnitude of the price increase may depend on how the other remaining companies respond and vice versa, this is not the underlying reason for the price increase. This is different from the so-called 'coordinated effects' which may result from a merger. These refer to price effects (or other effects) which are profitable to the merging firms *only because* other companies in the market choose to refrain from competing in a strong manner, for example choose to coordinate.[164]

The precise nature of the competitive constraints between the parties that a merger eliminates can vary from merger to merger. In some mergers, it may be the fact that the merging entities produce relatively close substitutes that is the important aspect of the merger (our car example). In other mergers, the focus may be on the elimination of direct competition by the combination of important production capacities of the two firms.[165] In yet other mergers, it may be the combination of two market participants which previously provided important innovations and thereby influenced the nature of competition significantly. Unilateral effects analysis is therefore not confined to the context of price competition in differentiated product markets. **1.235**

It is also worth noting that the previously described effects have, by themselves little to do with the question whether the merging firms will become the largest player in a market. What matters is that the merger involves companies that, pre-merger, formed a significant competitive constraint on each other and that the market context is one where the remaining competitors do not form a fully effective competitive constraint. **1.236**

In the car example, what drove the result was the fact that a merger between Audi and BMW would eliminate competition between the two and that the two remaining companies would exert only a partial constraint on Audi and BMW. For example, in case of a price increase of 10 per cent on a Audi, half of the former Audi customers would go to BMW and not to Mercedes or Volvo. In this sense, Mercedes or Volvo exert only a partial constraint on Audi; the remaining part comes from BMW.[166] **1.237**

Consequently, a fundamental aspect in determining whether a merger should be considered anti-competitive is the degree to which the remaining companies exert a competitive constraint on the merging parties. **1.238**

---

[164] See Section C.5 for more detail on the scope for tacit coordination in a given market.

[165] In markets where output or capacity decisions are the most important strategic decisions of the firms, the important concern for firms is how their output decision influences market prices. In such circumstances, the merged firm may have an incentive to reduce output relative to the pre-merger levels, thereby raising the market price. This incentive is likely to increase, the larger the sales volume of the merged firm, since the corresponding price increase will benefit a larger base of sales. The combination of market shares from two previously independent firms will in some cases thus produce an incentive to reduce output or capacity. For more details, see M. Ivaldi, B. Jullien, P. Rey, P. Seabright and J. Tirole, 'The Economics of Horizontal Mergers: Unilateral and Coordinated Effects', Report to DG Competition (available at < http://ec.europa.eu/competition/mergers/ studies_reports/studies_reports.html>.

[166] This example can also serve to illustrate that it is not strictly necessary that the merging parties' products are 'closest' for the merger to produce a (noticeable) price effect. The incentive to raise price exists even if the two merging parties' products are not closest substitutes. However, it is true that the more the merging products are considered to be 'closest' by customers, the more likely it is that a noticeable effect will result (all else being equal). What matters is the degree to which the remaining products exert a competitive constraint on those of the merging parties.

1.239 In practice, therefore, there is a high degree of overlap between those cases where the merging parties end up being the dominant player in the market and the cases in which *significant* unilateral effects are likely to arise. This does not mean to say that market dominance (in the usual sense of the word of being the largest company in the market) is either a necessary or a sufficient condition for negative consequences to occur, but there is a strong correlation.[167]

(b) Exclusionary Strategies

1.240 A firm with market power may raise prices by reducing its own output or by making competitors reduce theirs. Strategies that seek to achieve the latter are commonly referred to as 'exclusionary'.[168] A company with market power may seek to exclude rivals in a variety of ways.

1.241 One important way to make rivals produce less is to raise their cost. This is a primary concern in the context of vertical restraints.[169] In the context of agreements between companies at different levels in the production or distribution chain (vertical agreements), antitrust concerns may arise when an agreement results in market foreclosure.[170] For example, it may be possible for a company to conclude exclusive agreements with the most important suppliers of raw materials or necessary infrastructure and thereby prevent competitors' access to these inputs or make such access more expensive for them (input foreclosure). When such foreclosure has the effect of significantly increasing the cost levels at which rivals can operate, it may increase the market power of the company having concluded the agreement and lead to higher prices downstream. This scenario is known as enhancing market power through *raising rivals' costs*.

1.242 Rivals' costs may also be increased through agreements that lead to the foreclosure of access to important sales channels (customer foreclosure). Such concerns typically arise in the context of exclusive dealing arrangements in the retailing or distribution sector, but may also apply in the context of loyalty rebates provided by dominant firms. When denied the necessary scale of operations, rival firms may be exposed to a higher cost level (be put on a higher point on the cost curve). Indirectly, and to the extent that customer foreclosure impacts upon the revenue streams of rivals and their ability to invest in R&D and cost reduction, it may also affect their ability to compete in the longer run.

1.243 Although vertical mergers differ from exclusive vertical agreements in that the divisions of the integrated firm can remain active as players in the intermediate goods markets, a vertical merger can modify the incentives of the integrated firm in its dealings with competitors upstream or downstream.[171] For instance, a vertically integrated firm, when deciding to

---

[167] See also the Horizontal Merger Guidelines, para 25, where it is stated that 'Generally, a merger giving rise to such non-coordinated effects would significantly impede effective competition by creating or strengthening the dominant position of a single firm, one which, typically, would have an appreciably larger market share than the next competitor post-merger'.

[168] The term 'exclusion' is broadly used for any (anti-competitive) practice which leads competitors to produce less; it is not limited to situations where competitors are forced to exit the market altogether. The same holds for the term 'foreclosure'.

[169] cf S. Salop and D. Scheffman, 'Raising Rivals' Costs' (1983) 73 Am Econ Rev 267; Krattenmaker and Salop, 'Anticompetitive Exclusion' (n 62).

[170] cf Guidelines on vertical restraints, OJ 2010 C130/1, para 100; Article 102 Enforcement Priorities Guidance, paras 19–20. For an elaborate analysis of foreclosure, see P. Rey and J. Tirole, 'A Primer on Foreclosure' in M. Armstrong and R. Porter (eds), *Handbook of Industrial Organization*, Vol 3 (Amsterdam: North-Holland, 2007), 2145–220, as well as the references in Section C.5(b).

[171] cf Guidelines on the assessment of non-horizontal mergers, OJ 2008 C265/6. For an elaborate presentation of the economic literature on non-horizontal mergers, see J. Church, 'The Impact of Vertical and

supply its competitors downstream with inputs, will take into account how these supplies affect the profits of its own downstream division. If the merged entity has substantial market power in the upstream market, it may have an incentive to raise the price level in that market as that will raise the costs of all non-integrated downstream firms, whereas the integrated firm has access to the input at the cost of production. The change in prices in the upstream market may thus reduce competitive pressure on the integrated firm in the downstream market, leading to overall increases in prices for downstream customers.

Assessing whether vertical integration or a vertical agreement has the effect of raising rivals' costs is in practice a fairly difficult matter. For input foreclosure to be a concern, it must generally be the case that the merging or contracting party involved in the input market has substantial market power: without such market power, it is difficult to see how price can be raised in the input market as a means to raise rivals' costs. One further needs to see to what extent rival companies lack sufficient alternative sources of supply and, where relevant, the ability to adopt counter-strategies (eg in the form of concluding their own contracts with players upstream, or to integrate vertically by way of merger). Furthermore, it is well recognized that vertical relationships may provide considerable scope for efficiency gains.[172] They may reduce transaction costs between the two companies[173] and better align the incentives of the companies in bringing a product to market.[174] As a result of such efficiency gains, competition in the market may intensify, rather than diminish. **1.244**

Apart from creating market power in a given market, vertical contracts or mergers can also serve to protect market power, by increasing entry barriers. Vertical linkages can raise the costs at which potential competitors can operate on a market (input foreclosure), or reduce the revenue streams that can be expected after entry (customer foreclosure). Because of foreclosure, potential competitors may have to enter two markets instead of one: entrants would also have to set up their own input production facilities or distribution system. When this is the result, a company with market power in either of the two relevant markets has become less exposed to potential competition. **1.245**

In settings where two or more products are often bought or used in combination, exclusionary conduct can also take the form of tying or bundling. 'Tying' occurs when customers who purchase one good (the tying good) are required also to purchase another good from the producer (the tied good). 'Bundling' refers to the way products are offered and priced by **1.246**

Conglomerate Mergers on Competition', Report for DG Competition, September 2004; M. Riordan and S. Salop, 'Evaluating Vertical Mergers: A Post-Chicago Approach' (1995) 63 Antitrust LJ 513; Rey and Tirole, 'A Primer on Foreclosure' (n 170).

[172] cf Riordan and Salop, 'Evaluating Vertical Mergers' (n 171), 523.

[173] Transaction costs can be understood as the usual costs of searching for a trading partner and of drawing up and enforcing contracts, but also as inefficiencies that result from not being able to write contracts as comprehensive as one might wish (incomplete contracts), which may reduce willingness to invest in assets which are specific to the vertical relationship. Mergers, but also exclusive contracts, can have the effect of restoring such incentives. See eg O. Williamson, 'Transaction Cost Economics' in R. Schmalensee and R. Willig (eds), *Handbook Of Industrial Organization* (Amsterdam: Elsevier, 1989).

[174] The incentives for the upstream and downstream companies are not necessarily well aligned. One classic example is the problem of double mark-ups. When the upstream and downstream markets are imperfectly competitive, both the downstream and the upstream company set a mark-up, as a result of which the joint mark-up may be too high from the point of view of the vertical structure as a whole. Depending on the market conditions, reducing the combined mark-up (ie the price) may allow the vertical structure significantly to expand output on the downstream market and increase profits.

the dominant undertaking. In the case of pure bundling, the products are only sold jointly in fixed proportions. In the case of mixed bundling, often referred to as a multi-product rebate, the products are also made available separately, but the sum of the prices when sold separately is higher than the bundled price. Tying and bundling can be used to 'leverage' a strong market position from one market to another.[175] The main antitrust concern in this context is again foreclosure, more particularly customer foreclosure. Such foreclosure may be inspired by the desire to gain market power in the tied goods market, to protect market power in the tying goods market, or a combination of the two.[176] As with vertical foreclosure concerns, it is fairly difficult to predict when bundling and tying are detrimental to competition, not least because bundling and tying also have a potential to lead to efficiency gains.

1.247 A final way in which rivals may be excluded in an anti-competitive way is through predatory pricing. Predation refers to the strategy of a (dominant) company to charge very low prices for its products in order to prompt the exit or marginalization of its rivals unable to sustain the losses incurred for a prolonged period. Following the exit or marginalization of rivals, the company would be in a situation of enhanced market power and be able to raise prices. While the idea of predation is rather straightforward, it is clear that there are quite substantial hurdles for such a strategy to work. Not only must rivals be marginalized or forced to exit, it must also be the case that, following their exit, there is no entry by new companies or re-entry by the old ones.

1.248 A complication with pursuing cases of exclusion, especially when they are of the customer foreclosure type, is that the type of behaviour pursued may closely resemble acts of normal competition. The concept of 'exclusion' is inherent in any process of competition. When companies seek to supply customers and are very successful in doing so, some rivals are 'excluded' in passing and may even have to exit the market. Such exclusion should, in principle, be of no concern to competition policy. Indeed, competition policy should ensure that the normal competitive process is able to perform its task in benefiting the companies that are the more efficient in producing goods and services and the more effective in catering for the customers' needs. This ground principle, however, also mandates that competition policy should keep an eye on companies that—though 'successful' in selling to customers—seek to exclude rivals in ways that are not compatible with the competitive process in the long run and are harmful to consumers. At the same time, the fact that real competition ('competition on the merits') and exclusionary practices are so difficult to disentangle only highlights the need to be cautious in intervening in free market processes out of concern that a given company is seeking to exclude rivals. Companies may restrict competition, but so may antitrust authorities—when their policies are too interventionist.

---

[175] There is no received definition of 'leveraging' but, in its most neutral sense, it is being able to increase sales in one market (the tied market), by virtue of the strong market position of the product to which it is tied or bundled (the tying market).

[176] See eg M. Whinston, 'Tying, Foreclosure and Exclusion' (1990) 80 Am Econ Rev 837. See also Non-Horizontal Merger Guidelines, paras 108, 111.

G. Empirical Methods for Market Definition and the Assessment of Market Power

Both market definition (an intermediary step in the analysis of market power) and the assessment of market power itself address the following central question: to what extent do companies compete with one another?

1.249

This question, in the majority of cases, is an empirical question. One needs to consider the specific facts of the case. In certain cases, where sufficient data are available, it is possible to apply quantitative, empirical methods to study this question.

1.250

In this section we will discuss the main methods that are available.[177] These methods are: the analysis of prices and price movements in the market, the estimation of price elasticities, critical loss analysis, the assessment of prices and market structure, event analysis, the analysis of bidding data, and, finally, techniques involving merger simulation.

1.251

As this review will show, empirical analysis does not need to be sophisticated, nor does it need to rely on having access to numerous data. Some methods are relatively simple. What matters most is that a method is chosen that is sound for the case under investigation. Help from econometricians is valuable in this respect, but using one's own common sense is also an important ingredient.

1.252

(1) Analysis of Prices and Price Movements

Prices are probably among the main competitive variables in any market. Analysis of prices, and of price movements, is therefore likely to provide useful first information on the degree to which products compete.

1.253

(a) *Price Correlation Analysis*

One intuitive tool for analysing prices is price correlation.[178] The main idea behind price correlation analysis is that, when two products are in the same relevant product market, over time their prices are likely to move together relatively closely. After all, when products are substitutes in the eyes of customers, the prices of these products are likely to constrain each other. Note that this does not mean that the prices themselves have to be at the same level; low-priced products of a lower quality may well constrain high-priced goods of a higher quality, and vice versa.

1.254

To illustrate, suppose we have the following monthly price levels for two products A and B, for the years 2010–12:

1.255

---

[177] For useful reference works, see S. Bishop and M. Walker, *Economics of EC Competition Law* (3rd edn, London Sweet & Maxwell, 2010); P. Davis and E. Garcés Tolon, *Quantitative Techniques for Competition and Antitrust Analysis* (Princeton: Princeton University Press, 2009); G. Niels, H. Jenkins, and J. Kavanagh, *Economics for Competition Lawyers* (Oxford: Oxford University Press, 2011). A recent overview of the use of empirical techniques in EU merger control is given in European Commission, 'Economic Evidence in Merger Control', Competition Committee, Working Party No 3, OECD, 15 February 2011.

[178] Price correlation analysis has been applied or discussed in various Commission cases, eg *Nestlé/Perrier*, OJ 1992 L356/1; *Procter&Gamble/Schickedanz*, OJ 1994 L354/33; *Gencor/Lonrho*, OJ 1997 L11/30; *CVC/Lenzing*, OJ 2004 L082/20; *Blackstone/Acetex*, OJ 2005 L312/60; *OMV/MOL* [2008], notification withdrawn; *Arjowiggins/M-real Zanders Reflex*, OJ 2008 C267/14; *Ryanair/Aer Lingus*, OJ 2008 C47/9; *Arsenal/DSP*, OJ 2009 C227/24; and *Outokumpu/Inoxum,* OJ 2013 C312/6.
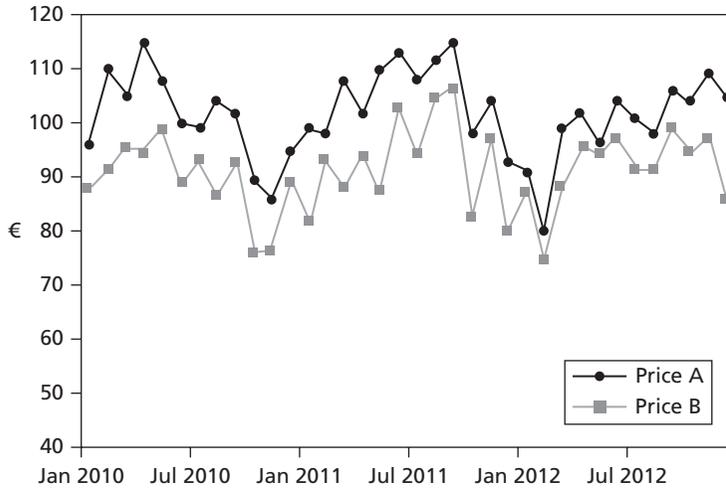
Figure 1.14 Price development of products A and B over time

From Figure 1.14 there appears to be some correlation between the two price series, but it is not perfect. In most months, the prices move in parallel (eg both prices make substantial drops in the second half of 2010, and again in early 2012), but in some months, they move in opposite directions.

1.256 The degree to which two prices move together is measured by the *correlation coefficient*, a measure that can take a value between −1 and 1. The coefficient is calculated on the basis of the deviations of the prices from their average values at each point in time.[179] When the correlation coefficient is equal to 1, the correlation is perfect. It is zero when the prices move independently of each other; it is −1 when the prices persistently move in opposite directions.

1.257 The fact that in Figure 1.14 there appears to be some positive correlation between the two price series is also conveyed by the correlation coefficient, which equals 0.77 in this case (positive value but less than 1).

1.258 When two products are in the same relevant product market, one would expect the correlation coefficient to be fairly high as this would be consistent with prices closely moving together over time. A first practical question that arises is: how high does the correlation coefficient need to be for the products to be considered in the same relevant market? One must have some idea of the relevant benchmark for comparison. One suggested way is to take two products that are known to be in the same relevant market (eg because of

---

[179] The correlation coefficient between two price series is equal to the covariance (joint variance) of the price series divided by the product of the standard deviations of the two individual price series. Specifically, if P_t^A denotes the price of product A at time t, and P_t^B the price of product B at time t, the covariance of the two price series is $(1/n) \sum (P_t^A - P^A)(P_t^B - P^B)$, where P^A and P^B are the average values of the price of A and B, respectively; n is the number of observations; and Σ the summation sign. The standard deviation of the price series is a measure of the variability of the price over time. For product A, it is the square root of $(1/n) \sum (P_t^A - P^A)^2$, for product B it is the square root of $(1/n) \sum (P_t^B - P^B)^2$. Thus, the correlation coefficient (r) is given by the following formula: $r = (1/n) \sum (P_t^A - P^A)(P_t^B - P^B)/\sqrt{(1/n) \sum (P_t^A - P^A)^2} \sqrt{(1/n) \sum (P_t^B - P^B)^2}$. The correlation coefficient is always pair-wise (eg between two series of prices).

their identical product characteristics), and to see how much their prices are correlated. This approach is known as benchmarking. In our previous example, the idea would be to compare the correlation of 0.77 with the correlation between product B and another product C known to be in the same market as B.

However, one issue with benchmarking is that, if one takes two close substitutes with a price correlation of, say, 0.90, a third product may be perhaps not so correlated (not so close), but still be close enough to be in the same relevant market. Therefore, the 'benchmark' obtained from two products in the same relevant market should not be used too strictly, but rather as a rough indication. **1.259**

Another important point to be aware of is that prices may be correlated for reasons that have nothing to do with competition between these products. For example, if prices follow the same trend (eg upwards, due to general inflation), this would show up in the correlation coefficient. The problem becomes particularly relevant when two products are made with the same major input. The classic example is prices at the petrol station. It may very well be that prices at petrol stations in Sweden and Portugal are highly correlated but this probably says very little about the relevant geographic market for petrol distribution. Rather, the correlation is likely to be the result of developments in the price of crude oil. Correlation driven by this type of factor is called 'spurious' (spurious in the sense that there is correlation due to reasons unrelated to substitutability). **1.260**

One must also think of the proper time dimension. The degree of correspondence between two prices depends on the speed with which prices can react to each other. Prices may constrain each other, but only with a certain delay. This would be the case, for example, for a commodity that is traded both at the spot market and on a supply contract basis, with quarterly revisions of the supply price. Prices in the contracted market may react to spot market prices, but can do so only at the revision dates or after the contracts have expired. In this case, the daily or weekly correlation in prices may turn out to be relatively low, while the correlation between prices measured at a quarterly basis is probably higher. In such an instance, the appropriate correlation coefficient to look at would be the one based on quarterly prices. **1.261**

While it is certainly useful to have a look at price correlation (in particular by looking at the graphs), one must realize that it does not provide the full answer to the question whether products belong to the same relevant market. If there is a high correlation between the prices of two products, this simply means that, on average, when the price of one product went up, the price of the other product went up as well, and vice versa. It does not directly address the question of *how many* customers would switch in the event of a price increase on a product (or group of products), which is the central question for market definition purposes (the SSNIP test). It is true that a high correlation coefficient suggests a strong competitive relationship in this sense, but it can only be taken as indicative evidence. **1.262**

(b) Extension: Stationarity/Co-Integration

As noted previously, measuring price correlation may give rise to misleading ('spurious') results if, for example, the prices of two products follow the same general (upward or downward) trend. A price series following a certain trend is in fact a special instance of a price series that is *non-stationary*: the series cannot be said to move around a stable mean over time.[180] **1.263**

---

[180] Another example of a price series that is non-stationary is one where a random price movement at one point in time appears to have effects that persist (eg 'random walk').

When working with price series that are non-stationary, there is a high risk that the correlation coefficient is unreliable.

1.264 Co-integration analysis is a rather technical way of analysing price series that are non-stationary.[181] It starts from the idea that two price series that are non-stationary may still be connected to each other, that is, 'co-integrated' (eg price A is usually 40 per cent higher than price B). The intuition underlying co-integration analysis is the same as that of price correlation analysis: when two products are in the same relevant product market, their prices are likely to move together over time. This can be translated into analysing the difference between two price series (in absolute or relative terms) to see whether that difference follows a stable pattern, that is, is stationary. The statistical test used to analyse whether a series is stationary is rather involved, and typically requires expert input.[182]

(2) Analysis of Price Elasticities of Demand

1.265 The price elasticity of a product measures how demand for that product changes with the price of the product (this elasticity is also called the *own-price elasticity*). In particular, it measures the percentage change in demand following a 1 per cent increase in the price. If the price elasticity is for example 2,[183] this means that, following a 1 per cent price increase, demand goes down by 2 per cent.

1.266 As indicated in Section E.1, the own-price elasticity is a summary indicator of the extent to which a product is subject to competitive constraints (due to customer reactions and the presence of competitors). When the price of a product is raised, customers switch away from it: they either switch to competing suppliers, or they stop purchasing the product altogether. The (own-) price elasticity of a good captures both these movements. The higher the own-price elasticity, the more the product is subject to competitive constraints. Alternatively, the lower the own-price elasticity, the higher the degree of market power for the supplier concerned.[184]

1.267 Price elasticity analysis provides direct input into the SSNIP test for market definition. For example, if the aggregate elasticity[185] of the set of products one posits to be in the same relevant product market is equal to 1.5, the unit sales for the products will go down by approximately 7.5 per cent if prices for the products go up by 5 per cent (the usual SSNIP). Depending on the initial gross profit margins of the products involved, this may be profitable or not profitable. If these margins are around 40 per cent, the 5 per cent price increase represents a 12.5 per cent (= 5/40) increase in the profits made on the 92.5 per cent (100% – 7.5%) of sales retained. Comparing the profit gain on retained sales (0.925 × 12.5% = 11.6%) with the profit loss on sales lost (0.075 × 100% = 7.5%[186]), the price increase would be profitable.

---

[181] Co-integration tests have been applied or discussed in a relatively small number of Commission cases, eg *Gencor/Lonrho*, OJ 1995 C314/14; *CVC/Lenzing*, OJ 2004 L82/20; *Blackstone/Acetex*, OJ 2005 L312/60; *Ryanair/Aer Lingus*, OJ 2008 C47/9; and *Arjowiggins/M-real Zanders Reflex*, OJ 2008 C267/14.

[182] See the next subsection for more information on the subject of statistical testing.

[183] Because demand normally decreases if price increases, the own-price elasticity is in principle a negative number. However, it is customary to use the absolute term, ie to present it as a positive number, which is the approach used in this text.

[184] Note that the own-price elasticity of a product is normally greater than 1 (in absolute terms). If it were less than 1, eg 0.5, the supplier of the good could make more money by raising its price (a price increase of 1 per cent would result in only 0.5 per cent less demand, and hence lead to a net increase in profit).

[185] See Section E.1(c). The aggregate elasticity measures how total market demand (combined demand for all products in a particular market) changes with a price increase of 1 per cent.

[186] The full margin (100 per cent) is lost on the units no longer sold.

The *cross-price elasticity* measures how demand for a product changes with the price of some other product. For a set of products, there is an array of cross-price elasticities, each corresponding to an individual pair of products. The cross-price elasticity between competing products is normally positive (if the cross-price elasticity is zero, then the products concerned are not competing). Generally, the higher the cross-price elasticity of B with respect to the price of A, the more product B forms a competitive constraint for product A. Cross-price elasticities are thus particularly helpful in evaluating the 'closeness' of substitute products (relevant both for market definition and for evaluating possible unilateral effects arising from mergers). **1.268**

Information on elasticities can be obtained in various ways.[187] Some (rudimentary) information can result from customer surveys that ask the question: 'in the face of a 5 per cent price increase for product X, and assuming that the price of alternative products did not change, would you switch? If so, by how much?' If, out of a sample of 100 respondents, 25 indicate that they would switch away half of their demand to other suppliers, this could indicate that the own-price elasticity of the product in question is about 2.5 (assuming the respondents are more or less of equal size). The same question can also be asked for a group of products to see what the elasticity is for the group as a whole.[188] **1.269**

An issue with surveys is that the results obtained from a sample of customers should be representative for the larger group of customers. This is not always easy to achieve, if only for practical reasons (one may need a substantial group of respondents to have representative results). Further, the questions asked should be sufficiently accurate that they leave relatively little room for misinterpretation. Finally, the question is—by definition—hypothetical: 'what would you do if'. The answers from respondents to a survey are unlikely to be as well thought through as business decisions in the case of real price increases. With these caveats, however, surveys remain a useful tool, and certainly a good starting point.[189] **1.270**

Further (and more affirmative) information on switching behaviour can be obtained from looking at actual decisions to switch in the past. If there are quite a few respondents who indicate that they have switched in the past to take advantage of price differences between products, this signals that the elasticity for a particular product (or set of products) is likely to be substantial. **1.271**

To avoid the problem of surveys, information on switching behaviour can also be obtained from looking at historic market sales and price data. These data may reveal a certain pattern, namely that, on average, falls (or increases) in the sale price are followed by a certain increase (or fall) in sales. From this it may be possible to distil the price elasticity of demand. **1.272**

---

[187] The Commission considers price elasticities in most cases in a more qualitative manner, eg on the basis of surveys. More sophisticated regression techniques have been applied or discussed in, eg, *Procter&Gamble/Schickedanz*, OJ 1994 L354/33; *Guinness/Grand Metropolitan*, OJ 1998 L288/24; *TetraLaval/Sidel*, OJ 2004 L43/13; *Omya/Huber*, OJ 2007 L72/24; *Pernod Ricard/V&S*, OJ 2008 C219/5; *TomTom/Teleatlas*, OJ 2008 C237/8; *Friesland/Campina*, OJ 2009 C 75/06; *Unilever/Sara Lee Body Care*, OJ 2012 23/10.

[188] A small but significant minority of switching customers may already be enough for the own-price elasticity of a product to be substantial. See Section D.

[189] For useful guidance on the use of surveys, see eg 'Good practice in the design and presentation of consumer survey evidence in merger inquiries' published by the UK OFT and the UK Competition Commission in 2011 (available at <http://www.oft.gov.uk/shared_oft/consultations/merger-inquiries/Good-practice-guide.pdf>).

1.273 To illustrate, suppose that in addition to information on monthly prices in the period 2010–12 (depicted in Figure 1.14), we also have data on the monthly quantities of product A bought. The observed quantities and prices of product A may look like those shown in Figure 1.15.[190]
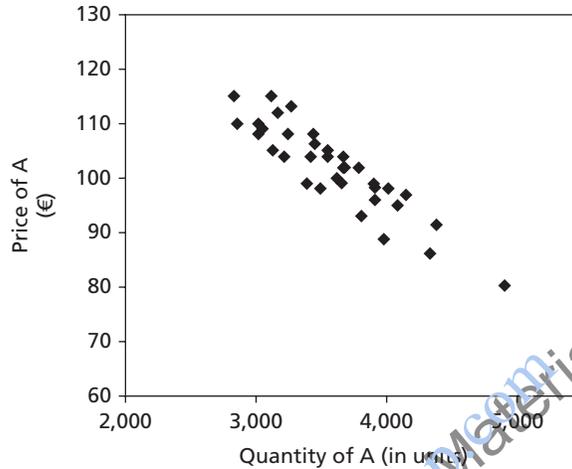

Figure 1.15 Observed prices vs quantities

1.274 Each 'dot' in Figure 1.15 represents a combination of the observed price and quantity of product A in a particular month (there are 36 dots). The dots indicate a negative relationship between prices and quantities. We can even recognize something like a 'demand curve' in the graph, by drawing a line that 'best fits' the points on the graph. An example of such a curve is depicted in Figure 1.16. It is downward-sloping, and seems to have a slight curvature. Note however that, strictly speaking, one cannot call the curve a 'demand curve' (a curve expressing demand as a function of price), unless one is confident that there are no other (important) factors that influence or explain the observed demand for product A at a certain price of A.[191] This aspect usually requires a lot of care in empirical analyses in order to avoid false inferences from the data.

---

[190] In economics, it is customary to display prices on the vertical axis and quantities on the horizontal axis, even when quantities are thought to be a function of prices, rather than the opposite.

[191] As will be discussed in paras 1.282–1.287, the price of substitute products and other factors may substantially influence demand for a given product. This will have to be taken into account. A more fundamental issue is that in each given period the observed prices and quantities are a reflection of the *equilibrium* in the market (ie the situation where supply equals demand), where quantity and price are jointly determined in a way that does not necessarily mimic the demand curve as such. It is only possible to interpret the dots in Figure 1.15 as a demand curve if we are confident that the observed relation between quantity and price reflects the consumer response to price changes, rather than a mixture of changes in demand and supply conditions. Otherwise, one is faced with the problem of *identification*: how to identify the true causal relationship between demanded quantity and price, when they are jointly determined and both affected by multiple factors. For this purpose, more advanced econometric analysis is generally needed.
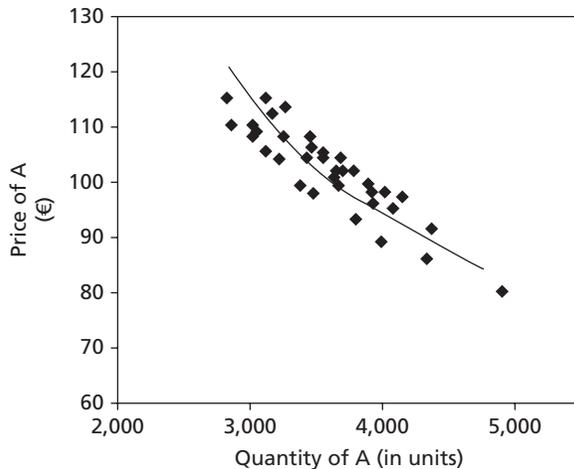
Figure 1.16 A line of best fit

The slope of the demand curve (assuming this curve is determined correctly) gives informa- **1.275** tion on the demand elasticity. The greater the slope of the demand curve (the steeper the demand curve) at a given price level, the lower the price elasticity at that price: a change in price has little impact on the quantity bought.[192]

If there are no other factors to take into account, an estimate of the price elasticity for prod- **1.276** uct A is obtained from the demand curve that best fits the data points observed. It is quite common to start with the assumption that the elasticity is constant across (the relevant range of) the demand curve, so that there is only one value to be estimated. This assumption determines the shape of the curve.[193] The assumption is not entirely innocuous, as the price elasticity of a product tends to increase when the price increases (demand often becomes more elastic at higher prices). However, its appropriateness can be checked later.[194]

---

[192] Note that the price elasticity and the slope cannot be 'equated', however. The slope relates quantity changes (in units) with price changes (in euros). An elasticity is about relating percentage changes, which is different. Eg for a given price increase of 1 per cent, a drop in sales of 100 units starting from a level of, say, 5,000 units is not the same as a drop of 100 units at a level of 3,000 units. The former drop is lower in percentage terms than the latter (2 per cent vs 3.33 per cent). In general terms, the relation between elasticity and the slope of the demand curve is as follows. For a certain unit change in price (Δp) and corresponding unit change in demand (Δq), the elasticity is approximately $(\Delta q/q)/(\Delta p/p) = (\Delta q/\Delta p) \times (p/q)$, ie the (inverse) slope of the demand curve multiplied by price level p divided by quantity q.

[193] Assuming that the elasticity is constant across the demand curve amounts to assuming that the relationship between the *logarithms* of quantities and prices is linear. As indicated in n 192, the price elasticity of demand and the slope of the demand curve are related, but not identical. In order to identify (and estimate) elasticities more easily, price and quantity data are usually transformed into logarithmic values (essentially expressing prices and quantities in terms of growth rates compared to a certain base). The slope of a curve in the resulting plot does indicate an elasticity: the slope relates a percentage change in the price with a percentage change in the quantity.

[194] Other simplifications are used as well, especially when simultaneously evaluating the price elasticities (both cross- and own-price elasticity) of various products. Eg the demand functions for differentiated products are sometimes assumed to follow a *discrete choice* model or an *AIDS* (Almost Ideal Demand System) model. The purpose of these initial—and generally testable—assumptions is to model consumers' behaviour in order to unveil the substitution patterns between different products. Also, and this is especially true in the case of discrete choice, these models reduce the number of parameters to be estimated.

1.277 The standard statistical tool used by economists to find and evaluate a relationship between observed data points is *regression analysis*.[195] Broadly speaking, in its most common uses, regression analysis aims at identifying a line through data points that provides the best fit, that is, which minimizes the differences between the actual observations and the plotted line.[196] It then evaluates whether the differences between the actual observations and the plotted line are substantial, in view of the number of data points available. The better the 'fit', the more 'precise' the estimated relationship can be deemed to be.

1.278 How confident can we be that the resulting elasticity estimate is precise and reliable? In general, the more data points one has, and the better the fit, the more one can be confident of having found a reliable estimate. The extent to which elasticity estimates obtained from regression are 'precise' in a statistical sense is answered in the following way. There is a 'true' price elasticity of demand, and there is the elasticity estimate found by drawing the line through the available data. In econometrics, when establishing a relationship between variables, it is recognized that there may still be other (small) factors, and measurement errors on the variables, that may produce an observed relationship that is not exactly identical to the 'true' relationship. Taken together, these other factors and the measurement errors form a certain 'chance' component in the observations (this produces the 'scatter' in the graph).

1.279 Technically, econometricians say that they estimate the following 'model':

$$Q_t^A = \alpha - \beta \cdot P_t^A + \varepsilon_t,$$

where Q_t^A stands for the quantity (in logarithms[197]) of product A bought in month t (t = January 2010,..., December 2012), P_t^A the price (in logarithms) of product A in that month, β represents the 'true' relationship between the quantity of A and the price of A, and where εt is the error term, that is, the 'chance' component at time t causing the quantity of A observed to deviate from the 'normal' level at price A (the quantity predicted by the model). Parameter α is a constant term to improve the fit. Under conditions of normality,[198] and as long as the error is not 'systematic' (eg as would be the case if there were still some other relevant, but omitted variable in the model), it can be shown statistically that, 95 per cent of the time, the 'true' elasticity lies within about *two standard deviations* of the elasticity

---

[195] For a good yet non-technical overview, see P. Kennedy, *A Guide to Econometrics* (6th edn, Cambridge, MA: Wiley-Blackwell, 2008), ch 1.

[196] To measure difference, one can use the absolute differences between the observations and the plotted line, or other measures of difference. The most practical method has proved to be to take the squared differences, and to draw a line such that the sum of the squared differences is minimized. This method is called Ordinary Least Squares (OLS). Statistical tests have been developed for the OLS method, and its variants.

[197] The use of logarithms allows us to interpret the coefficient β as an elasticity, see n 193.

[198] Normality conditions in this context means that the error term follows a 'normal distribution' with a mean equal to zero (ie the error is on average zero). The 'normal distribution' is a distribution of values with a certain shape. The term 'normal' is not taken by chance, in fact. It appears that many things, especially in nature (eg the height of oak trees), follow some normal distribution. A normal distribution is thought to result when the variable itself (height) is the result of many small and independent events influencing the variable (the amount of rainfall during each month, the amount of sunlight, branches breaking off during storms, young couples carving their names into the tree, etc). The—more prosaic—events in economics relate to measurement error, small random events determining demand, etc.

estimate obtained from the data sample.[199] The standard deviation of the estimated elasticity is an estimate of the variability of the elasticity estimate.[200]

If we conducted a regression analysis on the data shown in Figure 1.15, we would find that the own-price elasticity of product A would be estimated to be 1.40.[201] Allowing for the 'chance' component that may have produced this result, it can be said with 95 per cent confidence that on the basis of this regression the 'true' elasticity lies between 1.17 and 1.65.[202] This interval is called the *95 per cent confidence interval*. In principle, the narrower the confidence interval (the more closely it surrounds the estimate of 1.40), the more precise the estimate can be considered to be. **1.280**

'Preciseness' and 'reliability' are, however, relative concepts. As indicated (para 1.275), a problematic issue in the interpretation of the curve in Figure 1.15 arises from the fact that it relates the observed quantities of product A only to the observed prices of A. There may be other factors that influence the quantities of A bought, not just the price of A. When this is the case, then the relationship found by mechanically comparing observed quantities and prices of A (as carried out previously) is unlikely to be the correct one: the found elasticity estimate is then called 'biased'. And when the elasticity estimate is itself 'biased', the confidence intervals surrounding the estimate do not mean much either. **1.281**

For instance, when a product B is a good substitute for product A, an obvious factor influencing the demand for A is the price of product B. The way to obtain correct (unbiased) elasticity estimates is to add the price of product B into the analysis as a possible explanatory variable for the demand for A. By explicitly adding the 'price of B' to the analysis, the real effect of the 'price of A' on 'quantities of A bought' is identified. In graphical terms, the picture becomes three-dimensional, with on the vertical axis 'quantities of A bought' and on the two ground axes 'price of A' and 'price of B'. Econometric estimation (regression) finds the line that best fits all the data points in the three-dimensional plot.[203] The slope of the (new) line with respect to the price of A provides an estimate of the own-price elasticity of product A. The slope of the line with respect to the price of product B gives an estimate of the cross-price elasticity of product A with respect to the price of B. **1.282**

---

[199] The factor 'two' (in '*two* standard deviations') is in fact closer to 1.96, and is linked to the assumption of normality (see n 198). With a distribution different from the normal distribution, one would need a different factor. The same holds if one were to take a different confidence level (eg with 90 instead of 95 per cent, the factor becomes 1.64).

[200] Remember that there is a 'chance' component in the whole exercise, so that the estimate obtained is itself also influenced by chance. Hence, even though we end up having only one estimate of the coefficient (based on the sample), one can speak of a certain (intrinsic or underlying) variability of the estimate.

[201] Estimate obtained using a statistical software package.

[202] Note that this does not mean that the true elasticity lies in the interval with 95 per cent probability. Either the true elasticity lies within the interval (in which case the probability of the true elasticity lying in the interval is 100 per cent) or it does not (in which case the probability of the true elasticity lying in the interval is zero).

[203] Technically, econometricians now estimate the following 'model': $Q_t^A = \alpha - \beta \cdot P_t^A + \gamma \cdot P_t^B + \varepsilon_t$ where Q_t^A stands for the quantity (in logarithms) of product A bought in month t (t = January 2010,…, December 2012), P_t^A P_t^B the price (in logarithms) of product A in that month, β the price (in logarithms) of product B in that month, β and γ represent the 'true' relationships between, on the one hand, the quantity of A and, on the other hand, the price A and B respectively, and where ε_t is the error term, ie the 'chance' component in the observations at time t. Parameter α is a constant term to improve the fit.

1.283 Suppose the prices of product B over the period 2010–12 are those depicted in Figure 1.14. Using this information with the data on the prices and quantities of product A shown in Figure 1.15, the own-price elasticity of product A would be estimated to be 2.23. Allowing for the 'chance' component that may have produced this result, the 'true' elasticity is between 2.09 and 2.37, with about 95 per cent confidence. Note that demand for product A thus turns out to be more elastic than that suggested by the previous regression (2.23 is greater than 1.40). This is consistent with the fact that the prices of products A and B were quite correlated (see Section G.1), suggesting they might be in the same relevant market. On average, increases in the price of A were accompanied by increases in the price of B, limiting the actual sales loss of product A from an increase in its price. However, when prices of B are held constant (the 'all else being equal' aspect inherent in the notion of 'elasticity'), the sales loss of A is higher.

1.284 Regression analysis can also be used to help us to test hypotheses. For example, a regression analysis could be used to test whether two products are, in fact, substitute products. One hypothesis that can be tested is this: products A and B are not substitutes, which means that the cross-price elasticity is (close to) zero. A regression analysis can help us to test this hypothesis by providing an estimate of the cross-price elasticity between products A and B along with the standard deviations of the estimate. In our example, the regression produces an estimate of the cross-price elasticity with respect to the price of B equal to 0.98, with a 95 per cent confidence interval between 0.86 and 1.11. Given that the confidence interval is such that it does not include zero, it can be concluded with 95 per cent confidence that the true coefficient is not zero (in other words, one can be rather confident that the two products are indeed substitutes). In this case, econometricians say that the found coefficient is in statistical terms *significantly different from 0*, or in short 'statistically significant'.[204]

1.285 In a case where the confidence interval is such that it includes zero, it cannot be excluded with 95 per cent confidence that the true coefficient is in fact different from zero. Suppose, for example, that we had found a cross-price elasticity of 0.21 and a confidence interval between –0.05 and 0.47. In that case, econometricians would say that the found elasticity estimate of 0.21 is statistically not significantly different from zero. In other words, although the found estimate of the cross-price elasticity is positive (0.21), one would not be able confidently to say that the two products are in a competitive relationship.

1.286 There are three broad reasons why estimates may not be significantly different from zero. The first obvious possibility is that the coefficient being estimated is indeed zero or close to zero. Secondly, the data set may be too small to be confident that the result is different from zero: small data sets usually lead to wide confidence intervals, and this shows up in the estimate being 'statistically insignificant'. Thirdly, the differences between the actual

---

[204] Closely related to confidence intervals are the concepts of *t-statistic* and *p-value*. Whereas confidence intervals depict the range of values around the obtained estimate for which we can be 95 per cent certain that it will contain the 'true' coefficient, the *t-statistic* is the transformation of the obtained estimate into a test variable (think of *t-statistic* as meaning 'test statistic'), which is known to follow a certain standard probability distribution. Hence, we can test its significance and, accordingly, that of the corresponding elasticity estimate. When the *t-statistic* is larger than the critical value 'two', it is said to be significantly different from zero at the 95 per cent confidence level (on the number 'two', see n 199). The *p-value* is the probability that an estimate as large as or larger than the one obtained from the sample is obtained, when the true elasticity is in fact zero. When the *p-value* is low (eg below 5 per cent), it is unlikely that the true elasticity is indeed zero. At this point, one can conclude that the elasticity is significantly different from zero.

observations and the plotted line are substantial (ie the 'fit' is not good enough), so that the confidence interval around the estimate includes 'zero'. The statistical significance test alerts us that one of these situations applies.

As noted, estimates obtained from regression analysis are likely to be biased whenever variables that have a significant impact on the dependent variable are omitted from the analysis. In order to have reliable results, it would be necessary to check whether there are any omitted variables left. The added value of regression analysis is that it allows account to be taken of many factors that may potentially have an influence on the variable to be explained. Sometimes it is possible to think of potentially omitted variables. For instance, one could see if there have been promotion campaigns for either product A or B, and include such information in the analysis. In this way, one could check whether the influence of promotion campaigns is statistically significant. Alternatively, one could carry out some (econometric) checks to see whether the differences between the observed data points and the plotted line follow some systematic (yet unexplained) pattern, which would suggest that there may still be other factors at play. In a similar vein, one needs to bear in mind that the observed quantities and prices may not reflect the demand curve as such but rather the relation between quantities and prices in equilibrium. In this case, one is faced with the problem of *identification*, that is, how to identify the true causal relationship between demanded quantity and price, when they are jointly determined and both affected by multiple factors (including supply-side factors). To address this concern, more advanced econometric analysis is typically needed.[205]

A final remark relates to the relation between statistical significance and economic significance. The two concepts are obviously related, but not identical. For example, the estimate of a cross-price elasticity may, through the wealth of data available, be statistically distinguishable from zero, but it may still be very low in economic terms (eg 0.15). Similarly, while an own-price elasticity estimate may, due to a lack of data, not be statistically different from zero, it may still be quite high and important (eg 3.0). It is important to ask oneself why an estimate may be statistically significant or insignificant, and to keep an eye on the value of the estimates to see whether they are important and whether some economic implications could be derived from them.

(3) Critical Loss Analysis

Critical loss analysis is another method addressing the market definition question: would a hypothetical monopolist want to raise price on a set of products?[206] It addresses the SSNIP test from the other angle: rather than evaluating actual or likely demand-side responses to a price increase (eg through estimation of price elasticities), it looks at the supply side and asks: given a price increase of X per cent, what would the percentage loss in unit sales have to be to make the price increase unprofitable? If the actual loss of sales is larger than this amount, then a price increase is unlikely to be profitable. If it is less, it is profitable.

1.287

1.288

1.289

---

[205] For a useful description of the problem of identification and possible solutions, see J. B. Baker and T. F. Bresnahan, 'Economic Evidence in Antitrust: Defining Markets and Measuring Market Power' in Buccirossi, *Handbook of Antitrust Economics* (n 47).

[206] Critical loss analysis has been applied or discussed in a relatively small number of Commission cases, eg *Ineos/Kerling*, OJ 2008 C 219/15.

1.290 For example, if the gross profit margin (the difference between price and marginal cost) is 40 per cent, a 5 per cent price increase represents a 12.5 per cent (= 5/40) increase in the profits made on the sales that continue to be made.[207] At the same time, the full margin (100 per cent) is lost on the units no longer sold. Let the percentage of sales lost be denoted by L. Then the gain of the price increase is equal to 12.5% × (100-L); the loss is 100% × L. The critical loss is given by that L for which there is no net gain: 12.5% × (100-L) = 100% × L. The critical loss is therefore equal to 11.1 per cent.[208]

1.291 Critical loss analysis provides a benchmark with which the estimate of the actual sales loss in the case of a price increase can be compared. In the context of market definition, when an estimate of the price elasticity for a group of products in the candidate relevant market is available, one can compare the estimated sales loss (based on the price elasticity) following a 5–10 per cent price increase with the critical loss benchmark or threshold. If the former is higher than the latter, this indicates that the price increase would be unprofitable and that the relevant market should be wider. If not, the candidate market is an antitrust market (and the market assessed may even have been taken too large).

1.292 If no estimate of the price elasticity is available, one can still see whether the critical loss analysis suggests that the elasticity would have to be unrealistically low (or high) for the products to be in the same (or a different) relevant antitrust market. Note that a critical loss easily translates into a 'critical elasticity': if the critical loss in the context of a 5 per cent price increase is 11.1 per cent (as in the previous example), this means that the critical elasticity is 11.1%/5% = 2.2.

1.293 The critical loss benchmark for a given product (or group of products) depends on the price-cost margin on the product(s) and on the hypothesised price increase. The larger the margin, the smaller the critical loss will be. This is not surprising given that it is much more costly to lose sales when margins are high than when they are low.

1.294 One common misunderstanding is that, because high margins mean that the critical loss benchmark for a given group of products is low, it follows that the relevant market is probably wider than that group of products. This may indeed be the case, but one must keep an eye on what causes the high margins in the first place. Notably, high margins may be the result of a degree of product differentiation. In such a case, the critical loss may be low, but so is—in all likelihood—the actual loss in the case of a price increase.[209] A comparison of critical loss and (likely levels of) actual loss therefore remains preferable in many cases.

(4) UPP

1.295 A method that has become increasingly popular, especially in the context of assessing the likely impact of mergers in differentiated product markets, is the UPP ('upward pricing

---

[207] This assumes that the price-cost margin is constant over the sales base. The price-cost margin, also called gross profit margin, is the difference between price (p) and the incremental cost (c) of supplying one more unit of output, expressed as a percentage of price: (p-c)/p.

[208] A general formula for critical loss is given by: *Critical Loss* = $\Delta p/(\Delta p + m)$, where Δp denotes the percentage price change, and m the price-cost margin (in per cent). The formula only holds good when the price-cost margin m is constant over the sales base (in other cases, it is an approximation). In the example, it gives 5%/ (5% + 40%) = 11.1%.

[209] cf M. Katz and C. Shapiro, 'Critical Loss: Let's Tell the Whole Story' (2003) Antitrust Magazine, ABA Section of Antitrust Law 49–56.

pressure') method.[210] The method tries to gauge how pricing incentives *change* when a group of products is sold by one firm (the merged entity) instead of being sold by individual firms that make independent pricing decisions. UPP can also be used for the purpose of applying the hypothetical monopolist test (SSNIP test) in the context of market definition,[211] even if, in practice, it is rarely used in this way. Indeed, rather than being a tool for market definition, UPP has so far been advocated (and used) primarily in merger cases as a tool for *avoiding* market definition in a context where this is inherently difficult (differentiated product markets) and for focusing directly on whether the merger will generate upward pricing pressure.

1.296 The UPP method relies heavily on the concept of diversion ratios and as such is very close in spirit to other methods focusing on the elasticity of demand and closeness of substitution. However, UPP reinterprets the difference in pricing incentives between the single firm and the independent firms from the *cost* side, in particular, from the angle of opportunity costs. Before the merger, if one of the two merging parties were to increase its sales (by reducing its price), the value of sales lost by the other firm would not be taken into account.[212] After the merger, however, this impact on the other firm is taken into account and, indeed, viewed as a cost (an opportunity cost) to increasing sales. This amounts to an increase in the marginal cost of production, which tends to lead to higher prices or 'upward pricing pressure'. Unless efficiency gains due to the merger are sufficiently large to offset the increased opportunity cost, one can be confident that the merger will likely lead to a net UPP.

1.297 To illustrate, consider a merger between firms A and B which both sell differentiated mobile phones. Assume that they sell mobile phones at a pre-merger price of €100. The marginal costs of production are €75, leaving a gross margin (contribution to profit) of €25. When one firm, say firm A, decreases the price by €2.50 it will sell an additional 10,000 mobile phones. It will thereby negatively affect the other firms in the market, including firm B. Assume firm B sells 4,000 mobile phones less as a result (this is another way of saying that the diversion ratio is 0.40) and loses the profit margin of €25 it used to make on those sales, that is, it loses €100,000 profit in total.[213] When firm A merges with firm B to form a single firm, pricing incentives change. The profit margin lost by firm B becomes an opportunity cost for the merged entity when deciding on the optimal price of product A. The value lost by the firm B of €100,000, averaged over the 10,000 unit increase in sales, translates into an additional opportunity cost of €10 per mobile phone of brand A.[214]

---

[210] J. Farrell and C. Shapiro, 'Antitrust Evaluation of Horizontal Mergers: An Economic Alternative to Market Definition' (2010) 10(1) BE J Theoretical Econ; S. Salop and S. Moresi, 'Updating the Merger Guidelines: Comments' (2009), available at <http://www.ftc.gov/os/comments/horizontalmerger-guides/545095-00032.pdf>; G. Werden, 'A Robust Test for Consumer Welfare Enhancing Mergers Among Sellers of Differentiated Products' (1996) 44 J Industrial Econ 409. One of the (few) cases where UPP has so far been used in EU merger control is Case *Hutchison 3G Austria/Orange Austria*, OJ 2013 C224/6.

[211] J. Farrell and C. Shapiro, 'Recapture, Pass-Through, and Market Definition' (2010) Antitrust LJ 585.

[212] cf Section F.3(a) on the unilateral effects of mergers.

[213] As the products are differentiated (eg through brands), firm A's increase in sales of 10,000 stem from sales to customers drawn away from competing firms (including firm B) as well as sales to entirely new customers.

[214] Expressed as a percentage of the pre-merger price of brand A, the increase in opportunity costs amounts to 10 per cent. This ratio is also known as GUPPI (Gross Upward Pricing Pressure Index). Formally, the GUPPI for product A equals $D_{AB} \times m_B \times P_B/P_A$, where D_{AB} = diversion ratio from product A to product B, m_B = the percentage margin on product B, P_B = the price of product B, P_A = the price of product A. In terms analogous to the 2010 US Merger Guidelines, the GUPPI for product A equals the value of sales diverted to product B (the increase in profit on product B) divided by the lost revenues on product A.

1.298 Higher opportunity costs can be viewed as higher marginal costs for product A, which tends to lead to higher prices. Unless efficiency gains due to the merger exceed this level, the merger will probably lead to price increases. Suppose that we know (eg from past experience in the market) that cost increases are typically passed through by firm A at a rate of about 60 per cent, and let us assume that there are no efficiencies. On this basis, we can anticipate that the merged firm would increase the price of mobile phone A by about €6 (= 60% × 10), that is, carry through a 6 per cent increase in price.

1.299 In its original form, the UPP method compares the increase in the opportunity cost of production (the value of sales lost by the merging partner divided by the volume gain) with the efficiency gain from the merger to see whether there is a net UPP. Typically in merger reviews, however, the assessment of efficiencies is undertaken only in a second step, that is, once it has been established that the merger is likely to give rise to 'significant' anti-competitive effects absent efficiencies. In practice, therefore, UPP-type methods are primarily used as screens, to assess whether the merger is prima facie likely to produce significant price effects absent efficiencies. Where applicable, a more direct comparison with the expected efficiencies is then undertaken at a later stage of the investigation.

1.300 When does one say that a predicted UPP is sufficiently large to cause a 'significant' price increase? To answer this question one would need to know the relevant pass-through rate, that is, the extent to which an increase in the marginal cost of a product translates into a higher price for it. Precise estimates of the pass-through rate are not typically available, especially during the earlier stages of the investigation.[215] One sensible way to proceed which has been proposed in the literature is to proxy the pass-through rate using a default value, for example 50 per cent.[216] For instance, if one deems a 5 per cent predicted increase in the price level of any single product of the merging firm to be prima facie problematic (not yet taking into account efficiencies), this would mean that one should be worried about a predicted increase in the opportunity cost (expressed as a percentage of the pre-merger price of the product in question[217]) of 10 per cent. Of course, different levels result if one starts from different 'problematic' price increases.

1.301 To summarize, three ingredients go into a UPP analysis: diversion ratios, profit margins, and pre-merger prices. Combining these with 'default' pass-through rates turns the UPP method into a useful screen to separate mergers that require additional scrutiny from mergers that probably do not. At a second stage, UPP can be used to draw more precise conclusions, based on more detailed analysis of, for example, the nature and type of competition in the market, the likely efficiencies, and the likely pass-through rate.

(5) Event Analysis

1.302 Relevant information for the purpose of market definition and impact assessment can also be derived from the analysis of past 'events' or 'shocks' occurring in the industry.[218] The idea

---

[215] Note that for the purpose of assessing the magnitude of price effects on the merging parties' products, one needs product-specific pass-through rates, not industry-wide pass-through rates.

[216] cf Farrell and Shapiro, 'Antitrust Evaluation of Horizontal Mergers' (n 210). A 50 per cent pass-through rate is the rate that applies to a context of Bertrand price competition with a market demand that depends linearly on the price of the products concerned.

[217] This ratio is equal to the GUPPI index (cf n 214).

[218] This type of analysis has been applied in some Commission cases, eg *Procter&Gamble/Schickedanz*, OJ 1994 L354/33; *Kimberley-Clark/Scott*, OJ 1996 L183/1; *Blackstone/Acetex*, OJ 2005 L312/60; *Ineos/Kerling*,

is to consider the event, and to see how customers and/or companies reacted to it. Typically, but not necessarily, this analysis would involve some type of econometric analysis.

The 'events' can be of various types. An important type of event is past market entry. For instance, if, following market entry by company A, company B lost many sales, but company C's sales remained constant, then it may be concluded that A and B's products are in the same relevant market, and C's products are probably not. This analysis may also be applied on a more general basis, to see which products, rather than others, are closer substitutes for one another. If B's sales reacted strongly, but C's sales much less, then one could conclude that products A and B are closer substitutes than products A and C. **1.303**

Other examples of 'events' include supply shortages, shocks in input prices, regulatory intervention, technological change, and promotional and advertising activity. For example, if a promotional activity on one branded good (eg a strong advertisement campaign, or heavy discounting) resulted in a capture of market share of one other brand in particular, this may be taken as evidence that those two goods are in close competition with each other. **1.304**

Exchange-rate developments, given that they relate to trade between countries, may provide some insight into the question of geographic market definition. For example, if in the past, following a strong depreciation of the US dollar persisting for a lengthy period, US exports of the product under consideration did not increase, this could be taken as an indication that the US and the EU formed separate geographic markets for the product. Obviously, with the arrival of the euro, the 'exchange-rate event' is likely to become applicable less often in EU investigations, but, in cases involving both euro and non-euro countries, it remains a potential source of information. **1.305**

(6) Assessment Methods Relating Price to Market Structure

A promising avenue for investigating whether products or companies are the subject of significant competitive constraints opens up where it is possible to compare markets with one another, either a comparison between different markets (eg different geographic markets) or a comparison of markets over time (eg following entry or exit in the market or other changes in market structure[219]). **1.306**

(a) Price Concentration Analysis

An example of comparing markets with one another is price concentration analysis.[220] The object of study of price concentration analysis is to see whether prices are systematically higher in markets where there are a few players (high market concentration), than in markets where there are many players (low market concentration). **1.307**

OJ 2008 C219/15; *Ryanair/Aer Lingus*, OJ 2008 C47/9; *Lufthansa/SN Airholding*, OJ 2009 C295/10; *Arsenal/DSP*, OJ 2009 C227/24 in the context of merger control; and COMP/37.507 *Generics/Astra Zeneca*, OJ 2006 L332/24 in the context of Art 102. Obviously, the industry under investigation must have witnessed an 'event' in order to apply this technique. A useful presentation of the technique is provided by M. Coleman and J. Langenfeld, 'Natural Experiments' in Collins, *Issues in Competition Law and Policy* (n 33).

[219] When inferences are drawn from discrete events such as entry and exit, the assessment bears similarities to the event analysis method described in the previous section.

[220] This technique has so far been considered by the Commission in relatively few cases. Examples are *Nordic Capital/Mölnlycke Clinical/Kolmi*, OJ 1998 C39/19; *StatoilHydro/ConocoPhillips*, OJ 2008 C201/5 and a number of cases involving airlines (to investigate whether certain city-to-city routes constitute separate relevant markets), eg *Ryanair/Aer Lingus*, OJ 2008 C47/9. The closely related technique of comparing the level of discounts and the number of bidders participating in tenders for contracts is discussed in Section G.6.

1.308   Figure 1.17 provides an example of what appears to be a positive relationship between market concentration and price for a sample of distinct geographic areas. Where such a positive relationship can be established, this is an indication that the product market under consideration is indeed a relevant antitrust market[221] and that an increase in market concentration (eg through a merger) may lead to price increases. If market concentration is high due to the presence of a firm with a very large market share, it is also an indication that this firm is exerting market power and can be deemed dominant in the market.
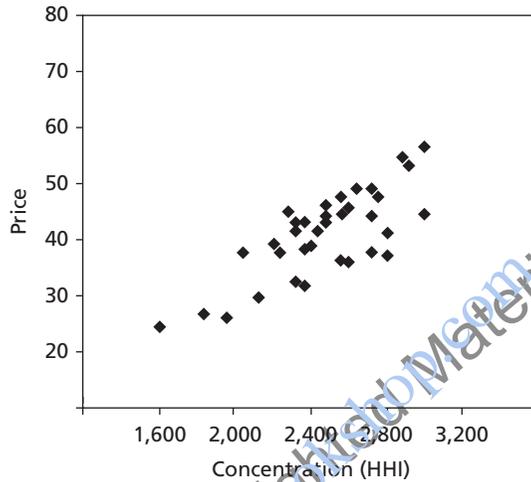

Figure 1.17   Price vs concentration

1.309   When there is no clear relationship between concentration and price, this signals that in the more concentrated markets there is no more market power than in less concentrated markets, for instance due to very low entry barriers. It can also signal that the 'markets' (on the basis of which concentration is measured) are themselves not really relevant product or geographic markets, but rather part of a broader relevant product or geographic market. For instance, one would expect to find little relation between the number of malt whisky producers and the price of malt whisky, if the relevant market in reality includes both malt whisky and blended whisky.

1.310   Figure 1.17 appears to suggest a positive relationship between concentration and price. The robustness of this conclusion can be (and typically should be) further investigated using econometric methods. Through regression analysis, one can seek to identify a line that best fits all the data points in the plot. The greater the slope of this line, the stronger the relationship (all else being equal). Econometric tests can then be performed, on the basis of the 95 per cent confidence interval around the estimated price concentration relationship, to check whether the relationship is indeed significant from a statistical point of view.

1.311   Importantly, in this context, the regression analysis allows for taking account of factors other than concentration that also affect price. For example, if it is the case that in certain countries the costs of running a business are high, so that prices are relatively high and fewer firms are

---

[221]   cf the 2010 US Horizontal Merger Guidelines, section 4.

active, then this would show a certain positive relationship between price and concentration regardless of the intensity of competition in the market. It is hence necessary to take such other factors into account, because analysis of the relation between price and concentration by mechanically comparing those two variables alone is likely to provide misleading results for the purpose of antitrust analysis.[222]

To make meaningful comparisons, it is necessary to compare 'like with like'. When the products whose prices are being compared are not identical across the regions, price disparities might be the result of differences in the product characteristics and costs, rather than of differences in the degree of competition present in the market. Incorporating product characteristics and costs directly into the analysis may be difficult when data on these variables are difficult to obtain. In such cases, it is preferable to work with margin data, given that margins typically better control for differences in product characteristics and costs than prices. **1.312**

(b) Direct Evaluation of Competitive Constraints

An important variant of the previous method is to perform the analysis not only on the number of market players, but also on the identity of the market players. This essentially amounts to analysing whether the presence of firm A typically goes hand in hand with lower prices charged by firm B, and vice versa.[223] If this is the case, this gives an indication of the likely price impact of a merger between companies A and B. Again, econometric methods (regression analysis) can be used to estimate the order of magnitude of the price effect and to check whether the relationship found is in fact significant from a statistical point of view, properly controlling for other relevant factors that may have an influence on prices charged in the market. **1.313**

(7) Analysis of Bidding Data

Certain markets can be characterized as bidding markets. In essence, these are markets where companies compete for specific contracts. The term 'bidding market' covers both situations where customers use formal bidding rules (as is the case in public procurement) and situations where customers simply elicit bids from sellers during negotiations. **1.314**

Analyses of bidding data are often helpful in evaluating the nature of competitive interaction among firms in the marketplace. They can be used to assess market definition by helping to identify the firms that participate or compete in a bid. They can also be used to assess market power by identifying the firms whose presence is most important in determining the outcomes of bidding situations. **1.315**

A particular issue in the context of bidding markets is the question of what role market shares play in the competition assessment. In each particular bidding contest, there is normally only one winner. The fact that another firm did not make a sale in a particular bidding contest does not mean that this firm did not pose a significant competitive constraint on the winning firm. In such a case, market shares (which give an indication of the firms' success in bids) may not be a good reflection of the competitive significance of firms, especially when **1.316**

---

[222] In this context, one must also be aware of potential feedback effects, eg higher prices in the market attracting entry, which may bias the estimates.

[223] Such direct evaluation of competitive constraints has been performed in eg *Ryanair/Aer Lingus*, OJ 2008 C47/9; *StatoilHydro/ConocoPhillips*, OJ 2009 C201/6; and in a number of cases involving bidding markets (cf Section G.7).

the number of bids in a given year is small (when the number of bids increases, one can expect market shares better to reflect competitive strength).

1.317 In addition, the link between market share and market power is probably less direct in bidding markets than in most other markets.[224] In bidding markets, each customer receives, or may receive, a personalized offer. Where this is the case, companies can decide to compete more aggressively on the margin, without this necessarily having a direct impact on the margins obtained on their existing customer base. Especially when individual contracts are large and infrequent, the incentive to compete for each of them may be strong.

1.318 Accordingly, in bidding markets it is useful to seek direct information on the importance of the respective market players in the bidding process, and to see whether market shares overstate or understate market power. Three forms of bidding analysis are often applied, mostly with a view to establishing which firms have been competing strongly against each other for certain types of contract.[225]

1.319 *Frequency of encounter analysis* consists in counting how often specific firms meet. For example, if firm A meets firm B more than 80 per cent of the time in those bids in which it participates, but meets firms C and D only 30 per cent and 20 per cent of the time, respectively, this can be an indication that firms A and B are 'close' competitors for the customers they supply.[226]

1.320 *Runner-up analysis* seeks to provide more accurate information on the 'closeness' of competitors by looking at the number of times company A has come second when company B has won a bid, and vice versa. The more often two companies have put in the two most competitive bids, the more they represent the main competitive threat to each other.

1.321 *Price impact analysis (discount analysis)* investigates whether the number (and possibly the identity) of bidders present in a bid has a significant impact on the prices (or discounts) being offered. When prices are, on average, higher when the number of bidders is low, this indicates that the number of bidders in the market matters, and that a merger may lead to price increases. One can also investigate whether the prices offered by company A tend to be lower when company B is also bidding (and vice versa). This would give an indication of the likely price impact of a merger between companies A and B.

1.322 Also in this context, one should compare 'like with like'. When the contracts are particularly diverse in nature or size, it is probably better to compare discounts than actual prices (discounts normally vary less with differences in the actual contract to be performed). Even then, however, one still needs to be aware of factors influencing the level of discounts, such as the value of the deal (higher values usually attract greater discounts).

---

[224] But see Klemperer, 'Bidding Markets' (n 149), for a critical discussion.

[225] Bidding data have been analysed by the Commission in a considerable number of cases, eg *Boeing/ McDonnell Douglas*, OJ 1997 L336/16; *PriceWaterhouse/Coopers & Lybrand*, OJ 1999 L50/27; *Philips/Agilent*, OJ 2001 C92/10; *Buhrmann/Samas Office Supplies*, OJ 2003 C117/5; *GE/Instrumentarium*, OJ 2004 L109/1; *Oracle/Peoplesoft*, OJ 2005 L218/6; *IBM/Telelogic*, OJ 2008 C195/05; *Syniverse/BSG*, OJ 2008 C101/25; *AEE/ Lentjes*, OJ 2009 C101/6; *WPP/TNS*, OJ 2009 C83/6; *Panasonic/Sanyo*, OJ 2009 C322/3; *Cisco/Tandberg*, OJ 2010 C36/7; *Oracle/Sun Microsystems*, OJ 2010 C91/7; *Western Digital Ireland/Viviti Technologies*, OJ 2013 C241/6; and *UPS/TNT* (2013), not yet reported.

[226] Note that such a pattern may be perfectly compatible with a market context where all four firms have equal market share (25 per cent). Eg companies C and D may meet each other more often (and secure more wins) in bidding contests for other customers.

A systematic way to investigate the relationship between discounts and the number (or identity) of bidders, and properly to control for other factors influencing the level of discounts, is to carry out a regression analysis. Econometric tests can then be performed to see how precise the relationship found to exist between the number (identity) of bidders and discounts can be deemed to be, on the basis of the 95 per cent confidence interval, and to test whether the relationship is indeed significant from a statistical point of view.

1.323

In certain industries, the number of bidders taking part in any particular bid is determined by the customer itself. If so, and when the number of potential bidders exceeds the number of bidders usually invited, the impact of the observed relationship between the number of bidders and the discount is likely to be small. This is likely to show up in an estimate for the relationship that is insignificant.

1.324

(8) Merger Simulation

Merger simulation is a more recent technique to simulate the impact of mergers in specific markets.[227] Two ingredients go into this technique: information on demand ('demand elasticities') and an assumption about the nature of competition in the market ('a model').

1.325

The idea behind merger simulation is that if one knows the demand elasticities, and knows the model according to which companies compete, it is possible to predict how prices will change once two firms in the model have merged.

1.326

Also when data on certain model parameters are not available (eg the precise cost levels of the firms, or possibly even the price elasticities of some of the products), it may be possible to 'retrieve' these parameters by fitting the market outcome as is predicted by the model for the situation pre-merger (eg in terms of market shares or prices) to the market outcome actually observed pre-merger. This step is called 'calibrating the model'. With all the model parameters available, it is then possible to 'recalculate' the model, but with two firms in the model having merged.

1.327

Merger simulation has so far been developed for three main industry settings: differentiated product markets (where companies are assumed to compete on prices à la Bertrand), commodity markets (where companies are assumed to compete on output à la Cournot), and bidding markets (where competition between firms can be modelled as an auction).

1.328

Provided it is carried out properly, the main advantage of merger simulation is that it casts some light on the magnitude of effects that can be expected following the merger, and on the question of whether they will be substantial or minimal. In that sense, the technique is a useful companion to merger analysis that mainly relies on the (qualitative) analysis of the change in market structure. Especially in industry settings where market shares are not

1.329

---

[227] Merger simulation was first used by the Commission in the case of *Volvo/Scania*, OJ 2001 L143/74. In this case, the Commission decided that, in view of the novel character of the approach and some not fully resolved issues on the reliability of the results and the data, it would not rely on the simulation results for deciding the case. In *Lagardere/Natexis/VUP*, OJ 2004 L125/54, the Commission did rely on the results, but only as part of the wider body of evidence. Merger simulation studies have further been considered in *Philip Morris/Papastratos*, OJ 2003 C212/4; *Sydkraft/Graninge*, OJ 2003 C240/4; *Oracle/Peoplesoft*, OJ 2005 L218/6; *BHP Billiton/Rio Tinto* [2008], notification withdrawn; *EDF/British Energy*, OJ 2009 C38/4; *Kraft Foods/Cadbury*, OJ 2010 C29/3; *Unilever/Sara Lee Body Care*, OJ 2012 23/10; and *Outokumpu/Inoxum*, OJ 2013 C312/6.

necessarily informative (in particular, in differentiated product markets, where market definition itself is a difficult exercise, and in bidding markets), merger simulation can provide added value.[228]

1.330 In addition, merger simulation can allow for the explicit consideration of merger efficiencies. When one expects the merger to produce significant cost savings (notably, in the form of marginal cost savings), the model can be recalculated on the basis of the new, lower cost level for the merged entity. Merger simulation is thereby a means directly to assess the *net* impact of a merger on the market. Potentially, this is a major advantage of merger simulation in comparison with more traditional, market structure-based analyses of competition.

1.331 The main weaknesses of merger simulation are also well known. The 'model' content in the exercise is very high, possibly to the detriment of the empirics. In its purest form, empirical analysis is about observing things, and drawing inferences that are consistent with what is observed. Merger simulation also considers data, but draws inferences partly on the basis of a model, which is not the same. For example, when merger simulation involves calibration to obtain information on the value of parameters pre-merger, it obtains such estimates on the basis of a model (the model imposes a 'structure' on the data). Also, for its predictions, merger simulation clearly relies on the correctness of the specific model being used.

1.332 It is therefore important that one follows a strict approach in the application of merger simulation techniques when assessing mergers. Leading experts in this field commonly emphasize that it is essential that the model used and the estimates obtained provide a good 'fit' for the industry at hand, in that they 'explain' the past history of the industry at a fairly high level of generality,[229] and that sensitivity analysis should be conducted. When the model fits the industry, merger simulation has a number of potential advantages. As a general rule, however, it appears best not to rely on merger simulation alone, but to use it as part of a wider body of evidence.

---

[228] Merger simulation may better incorporate the fact that demand substitutability is a matter of degree. The products do not have to be regarded as either 'in' or 'outside' the market.

[229] G. Werden, L. Froeb, and D. Scheffman, 'A Daubert Discipline for Merger Simulation' (2004) 18 Antitrust Magazine 89–95 (the name 'Daubert' refers to the doctrine of the same name of the US courts with respect to the admissibility of expert economic evidence).