Chapter 8 sets out the economics of state aid law, which can be considered part of competition law in the wider sense. Having discussed the economics behind the major provisions of competition law, the next two chapters deal with what happens after an infringement has been found. Chapter 9 discusses the design of remedies for competition problems and infringements, an area that has attracted relatively little scholarly attention thus far compared with the identification of those problems—as one commentator noted, 'Everybody likes to catch them, but nobody wants to clean them.'[26] Chapter 10 deals with the quantification of damages arising from infringements, a topic of increasing interest as private damages claims against infringing parties are on the rise in many jurisdictions. Finally, in Chapter 11 we offer some thoughts on the use of economic evidence in competition cases in terms of best practice and future direction.

[26] Comment made by Tad Lipsky at the 2007 Federal Trade Commission hearings on s 2 of the Sherman Act (transcript of 28 March, p 47, at <http://www.ftc.gov/os/sectiontwohearings/docs/transcripts/070328.pdf>). The quote is attributed to William Baxter, the Department of Justice Assistant Attorney General behind the break-up of AT&T in 1982—see Chapter 9.

2

MARKET DEFINITION

2.1 WHY MARKET DEFINITION?

Two Italian producers of women's designer shoes want to merge. Their legal representatives notify the deal to the competition authority, saying it should be cleared because there is plenty of competition remaining post-merger. How should the authority go about assessing whether the merger is anti-competitive? Nowadays the basic approach to such mergers is well established in competition law. First, you define the relevant market. Is the market limited to Italian women's designer shoes or does it include other women's designer shoes, or indeed all women's shoes? Is the market national, or does it cover other countries where these shoes are sold? Second, you analyse the position of the merging parties in the relevant market. Do they have a substantially larger combined market share? Does the merger leave only a limited number of competitors in the market? If the answer to these last two questions is yes, the authority is more likely to find that the merger significantly reduces competition. (We address the topic of mergers more generally in Chapter 7.)

You can see how market definition matters here. If the market is limited to Italian women's designer shoes, and the merging parties have market shares of, say, 30% and 25% respectively, the merger is more likely to be blocked than if the market included all women's shoes, in which the parties have only 3% and 2.5%. A similar two-stage approach is taken in other types of competition case as well—for restrictive agreements and abuse of dominance, the relevant market is first defined, and then it is assessed whether the parties concerned have an appreciable or dominant position within that relevant market.[1]

2.1.1 AVOIDING THE SUBMARKET TRAP

The above basic approach may seem obvious to you. It is how almost all competition authorities around the world assess mergers and anti-competitive practices. But it wasn't always so obvious in the history of competition law. (Nor may it remain so obvious in future, as we explain at the end of this chapter and in Chapter 7.) Courts, authorities, and scholars struggled for a long time to develop the notion of market definition.[2] A high point of confusion about market definition was reached in 1962 with the US Supreme Court ruling in *Brown Shoe*, a merger between two shoe manufacturer–retailers (a case we also cited in Chapter 1 as causing confusion about the objectives

[1] We make two observations on terminology here. First, the expression 'relevant market' arose in US antitrust case law in the mid-twentieth century, where the term 'relevant' usually linked the market to a specific case or to a product of a defendant firm—for example, the market relevant for cellophane wrapping. Today, the concept has acquired a specific meaning that goes beyond the simple combination of two common words. Second, a number of commentators have advocated the use of the term 'market delineation', which arguably reflects more accurately what the exercise is about than 'market definition'. However, the latter term has gained the upper hand in common usage and we join the bandwagon in this book.

of competition law).[3] In this ruling the court endorsed the principle of submarkets within relevant markets. It first stated that 'The outer boundaries of a product market are determined by the reasonable interchangeability of use or the cross-elasticity of demand between the product itself and substitutes for it.' So far so good (we explain the concept of cross-elasticity in section 2.3 below). But the court went on: 'However, within this broad market, well-defined submarkets may exist which, in themselves, constitute product markets for antitrust purposes.'[4] The lower court had defined separate relevant markets ('relevant lines of commerce') for men's shoes, women's shoes, and children's shoes, but according to the Supreme Court this did not rule out the possibility that narrower relevant markets could exist at the same time, such as by price, quality or age/sex. In *Brown Shoe* itself such further submarkets were in fact rejected in the end, mainly on the basis that this would not change the analysis as the parties had similar market shares across those submarkets (eg, they had 9.6% of boys' shoes and 7.9% of girls' shoes; we discuss this aspect of market definition in section 2.7 on supply-side substitution and market aggregation). However, several subsequent cases in the lower courts followed the *Brown Shoe* principle and ended up defining submarkets within markets. For example, in *US v Mrs Smith Pie Co* (1976), the market was defined as that for all desserts (based on consumer research), but a separate submarket was then found for frozen dessert pies, and it was in this submarket that the competition concerns were investigated.[5]

It is plain to see that defining a relevant market and then allowing for submarkets within that market is not a very helpful step in the competition analysis. Commentators in the USA were quick to point this out. Some called the submarket principle 'an intellectual monstrosity' with 'little economic justification' (Hall and Phillips, 1964); others observed that 'either there is or there is not a market in which competition may be affected ... If the line of commerce is men's shoes, it should not also be men's golf shoes: if one boundary is right, the other must be wrong' (Hale and Hale, 1966). The submarket concept was eventually abandoned in the USA (although competition authorities in other parts of the world still fall into the submarket trap occasionally). The 1982 US Merger Guidelines paved the way for the hypothetical monopolist test, which is the test on which most competition authorities base their approach to market definition these days, and which we discuss extensively in this chapter.[6] Submarkets cannot exist under this test.

Under the hypothetical monopolist approach, market definition is an intermediate step in assessing the competitive constraints on companies. It analytically tries to separate two types of competitive constraint: competition from *other* products (or geographic areas), and competition from other suppliers of the *same* product (or in the same geographic area). The market definition stage of an inquiry focuses mainly on

[3] *Brown Shoe Co v United States* 370 US 294 (1962).
[4] Ibid., at 325.
[5] *US v Mrs Smith Pie Co* 440 F Supp 220 (ED Pa 1976).
[6] Department of Justice (1982), 'Merger Guidelines' 14 June

the first type, ie, competition from other products. Competition *within* the same product is assessed at a later stage in the inquiry, once the market has been defined. The way to achieve this analytical separation is to pretend that competition within products does not exist at the market definition stage, so as to focus only on competition *between* products. This pretending that competition within products does not exist is achieved by hypothetically monopolizing it (as explained in section 2.4).[7] If the relevant market is defined correctly in this way, the analysis can then focus on competition within that market, without worrying too much about outside products at that stage, or falling into the submarket trap. This is why, in our view, market definition can be a very helpful, and analytically sound, intermediate step in competition analysis.

2.1.2 AVOIDING THE PRODUCT CHARACTERISTICS TRAP

Market definition based on economic principles is about substitution and price pressure between products. It is not about the physical characteristics of products. Of course, physical features will often dictate whether products are substitutable—straw and twigs are not good substitutes for bricks—and they need to be taken into account in any market definition exercise, if only as a sense-check. But a market definition that relies solely on product characteristics can become economically unsound. Products with very different characteristics may still be substitutes in the eyes of the buyers—sugar and high-fructose corn syrup; plastic and glass bottles. One can have endless, and usually inconclusive, debates about market boundaries based on product features. Is chewing gum a relevant market because people like to chew something?[8] Are holiday parks with an indoor pool and sports facilities on site a separate market from holiday parks that have such facilities nearby but not in the park? (See section 2.5.) Moreover, market definition based on product characteristics overlooks the fact that price pressure between products does not require all customers to consider them substitutes. Babies and elderly people with dentures may not switch from bananas to other fruits, but that in itself does not make them a relevant market, as many other people may well see other fruits as substitutes. This may be sufficient for these other fruits to place a competitive constraint on bananas (see section 2.8). In this chapter we will see several examples of a reliance on product features possibly leading to erroneous conclusions about market definition.

[7] As we also explain later, the prevailing degree of competition within the product does have an impact on what happens after hypothetically monopolizing the market and hence on your conclusion on competition between products, so the analytical separation is not theoretically pure.
[8] This was the question in the Mexican predatory pricing case *Chicles Canel's SA de CV v Chicle Adams SA de*

2.1.3 BEWARE THE PRODUCT DIFFERENTIATION TRAP

Yet some caution is called for here. The economic theory of market definition works best when products are reasonably homogeneous—for example, apples versus bananas, or cellophane versus other wrapping materials—such that you can test whether one homogeneous group of products competes with another. Market definition works less well when product differentiation is an important feature of the market. Unfortunately from this point of view, many real-world products are differentiated—women's designer shoes, cars, breakfast cereals, mobile telephones, even apples come in different sizes and tastes. Differentiation typically means that there is a spectrum of products that are close but imperfect substitutes. For example, there is a whole range of cars from superminis and small family cars to executive and luxury cars. Likewise, corner shops, shopping centres, and airports are differentiated geographically. Delineating a relevant market on such a spectrum can be difficult, artificial and potentially misleading.

- Difficult: there may be no clear cut-off points along the spectrum. You can buy a new car at virtually every price multiple of €1,000 from around €7,000–€8,000—you get a Kia Picanto or Suzuki Alto for that—to well into the €150,000s, where you start buying variants of the Mercedes SL class, the BMW 7 series, or the Lexus LS series (and that's ignoring the Ferraris, Maseratis, and Rolls Royces that have prices an order of magnitude higher still).
- Artificial: a very wide market for all cars would include not-so-close substitutes such as a Skoda Fabia and an Alfa Romeo Spider in the same market; but very narrow markets—eg, the market for Porsche 911 Carreras—may not be informative for the competition analysis either.
- Potentially misleading: the reliance on market definition as an intermediate step in competition analysis creates the impression that the question is all about products being 'in, or out'—all products outside the relevant market are not close substitutes; all products within the relevant market are equally close substitutes. This is not the case in differentiated product markets. Product substitutability is a matter of degree, and market definition means you have to draw the line somewhere, in the knowledge that the product that is just inside the line may not be so different from the product that is just outside. (In fairness to the US Supreme Court, recognition of the complications that arise in differentiated product markets was probably one of the reasons behind the submarket principle in *Brown Shoe*.)

This need for caution does not imply that market definition is inappropriate in differentiated product markets. It is still often very useful in carrying out a market definition exercise. Several of the aspects of market definition that we discuss in this chapter—supply-side substitution and market aggregation (section 2.7), and chains of substitution (section 2.9)—can be specifically tailored to deal with the challenges of product differentiation. Nevertheless, the usefulness of market definition as an intermediate step is increasingly being questioned, and alternative approaches that focus directly on competitive constraints have been proposed, particularly in merger cases. We come

2.1.4 THE REMAINDER OF THIS CHAPTER

Section 2.2 gives an overview of the possible dimensions of a relevant market, in addition to the standard product and geographic dimensions. Section 2.3 discusses the basic economic principles of demand that are of relevance to market definition, in particular demand systems, substitution, and elasticities. Section 2.4 deals with the hypothetical monopolist test for market definition, as defined in its purest form in the 1992 US Horizontal Merger Guidelines. Section 2.5 explains critical loss analysis, the tool most frequently used to put the hypothetical monopolist test into practice. The subsequent sections address further specific aspects of market definition, namely the cellophane fallacy (2.6), supply-side substitution and market aggregation (2.7), price discrimination markets (2.8), and chains of substitution and further aspects of geographic market definition (2.9). Section 2.10 discusses market definition for complementary products (including aftermarkets and networks) and for bundles, while section 2.11 explains how market definition may depend on the relevant layer of the vertical supply chain. Section 2.12 addresses the question of product migration over time (for example, from narrowband Internet access to broadband) and how this relates to product substitution. Section 2.13 discusses how relevant markets may be defined by reference to features other than price, such as quality and innovation. Section 2.14 sets out some methods and techniques that are often used for measuring substitution and elasticities in practice. Finally, in section 2.15 we return to the question: 'Why market definition?'

2.2 DIMENSIONS OF THE RELEVANT MARKET

2.2.1 THE PRODUCT AND GEOGRAPHIC DIMENSIONS

When you read about relevant markets, you commonly see that they have both a product dimension and a geographic dimension. The relevant product market consists of the group of products that are considered close substitutes—for example, the market for broadband Internet services (whether over cable or fixed-telephony networks). The relevant geographic market is the area in which that group of products is sold or purchased—for example, the market for broadband Internet services in the UK, or in the Greater London area. However, a relevant market may have several dimensions in addition to product and geography. Often these dimensions are implicitly or explicitly considered as part of the product dimension, but sometimes they are overlooked. We list these dimensions here, before turning to the basics of market definition in section 2.3.

2.2.2 TIME OF PURCHASE OR CONSUMPTION

A train journey from Oxford to London at 7.30am is not much different from a train journey from Oxford to London at 10.30am (although the carriages tend to be somewhat more crowded at the earlier time), and yet the two may be in separate relevant markets. Travellers taking the 7.30am are likely to be more time-sensitive than price-sensitive.

you more likely to find a free seat on the 10.30, you are also significantly better off. Such peak versus off-peak market distinctions can be made in many transport, communications, and leisure markets, where peak may refer to the time of day or to the season. Time of purchase or consumption also matters for certain 'perishable' goods such as fresh fruits, newspapers, and sports broadcasting events. In the extreme this may also give rise to different relevant markets—for example, live broadcasting of sports events is generally seen as separate from deferred broadcasting of the same events (unless your team won the cup, in which case you may want to see the whole match over and over again).

2.2.3 PERIOD OF INVESTIGATION

This second additional market dimension is also related to time, but in a different way. It refers to the period of investigation, ie, whether the investigation is forward-looking, backward-looking, or both, and for how long. Merger assessments are forward-looking. They are generally concerned with how the merger will affect competition in the next one or two years. The analysis therefore focuses on competitive constraints in the next one or two years, and market definition should assist in identifying those constraints. Market analyses (and hence market definition exercises) carried out by national regulators under the EU telecoms regulatory framework focus on the next three years, as they must identify suitable regulatory measures for a period of three years.[9] In contrast, investigations into abuse of dominance often relate to the current situation or to the recent past, when the abuse took place. The question is then whether the company in question was dominant in that period. A forward-looking market definition in such cases may overlook the fact that a market has already been monopolized (and market power already exercised), and may erroneously conclude that there are close substitutes at the prevailing price (which is already a monopoly price). This is known as the cellophane fallacy, as explained further in section 2.6.

2.2.4 DIFFERENT CUSTOMER GROUPS

There may be different markets for the same product corresponding to different customers (or groups of customers). This may occur when some customers have a greater choice of alternatives than other customers, and suppliers can exploit this difference by targeting the latter, 'captive', customers with higher prices. This gives rise to price discrimination markets, a topic we discuss in section 2.8.

2.2.5 DIFFERENT DISTRIBUTION CHANNELS

The same bottle of beer can be in a different relevant market depending on whether it is sold in a supermarket or in a bar. The same holds for a pay-TV sports package that may be sold

[9] Council Directive (EC) 2009/140 of the European Parliament and of the Council of 25 November 2009 amending Council Directives (EC) 2002/21 on a common regulatory framework for electronic communications networks and services, 2002/19 on access to, and interconnection of, electronic communications net-

to residential customers and to bars. This is not to say that different distribution channels always form separate markets. The main criteria are again the difference in demand patterns and the ability of suppliers to exploit this difference—when you are in a bar you are unlikely to switch to buying a drink in the supermarket, even if it is five times cheaper to do so.

2.2.6 THE VERTICAL LAYER IN THE SUPPLY CHAIN

The same product—a car, a washing machine, a bunch of fresh flowers—may pass through a whole set of intermediaries on its way from producer to end-consumer. The relevant market can be different depending on which layer of the chain is considered (and this in turn will normally depend on where in the chain the competition problem in question arises). This poses difficult questions for market definition, and we discuss these further in section 2.11. One set of questions concerns the relationship between demand by end-consumers and demand by intermediaries—is demand by the intermediary derived from consumer demand? If consumers can switch to substitutes, can the product still be a must-stock item for the intermediary? Another set of questions relates to the importance of intermediaries vis-à-vis end-consumers—should competition law be mainly concerned with end-consumers as opposed to intermediaries? Does it matter if a whole layer of the chain is excluded from the market through the conduct in question if end-consumers are no worse off?

Not all market dimensions identified in this sub-section will be relevant to all competition cases, and sometimes the additional dimensions will be implicitly considered as part of the definition of the product market. Nonetheless, we suggest that it is useful to check explicitly which of these dimensions may lead to further separate markets in any specific case at hand.

2.3 THE DEMAND SIDE: SUBSTITUTION AND ELASTICITIES

A distinction was made above between two types of competitive constraint on companies—one from other products (which is assessed as part of market definition), and one from suppliers of the same product (which is mainly assessed at a later stage in the competition analysis once the relevant market has been defined). Another useful distinction is between demand-side substitution, supply-side substitution and new entry. Demand-side substitution refers to switching behaviour by customers. This is the most immediate competitive constraint that companies face. Supply-side substitution refers to suppliers in neighbouring markets using their existing production facilities to start producing the product in question (or service the geographic area in question). We explain supply-side substitution in section 2.7. It is usually a less immediate constraint on companies than demand-side substitution. Customers are generally quicker to vote with their feet if they are unhappy than wait for other suppliers to come to them. New entry by rivals takes even more time and investment than supply-side substitution, and is therefore

2.3.1 THE DEMAND CURVE, WITH CHART

In Chapter 1 we saw how customers' responsiveness to price can be captured by the demand curve, how this demand curve is affected by the prices of other, substitute, products, and how this forms the basis for market definition. We elaborate on that here. Figure 2.1 shows the demand curve that we didn't draw in Chapter 1. It represents a very simple demand schedule for, say, women's designer shoes (ignore for the moment any differences between the Italian, French, or Spanish varieties of this product). As with your imaginary price–quantity space in Chapter 1, price is on the vertical axis and quantity on the horizontal axis. According to the demand schedule, no customer buys the shoes if the price is 10, and for every price decrease of 1 monetary unit there is demand by one additional customer (this is a simplification: in demand schedules such as this one the quantity normally refers to number of units of the product and not to the number of customers, but let's just say every customer buys one pair of designer shoes). So if the price is 9 there is one customer willing to pay for the shoes, if the price is 8 there are two customers, etc. If we wrote the formula for this demand schedule (something we generally try to avoid in this book), it would be $q = 10 - p$, where q stands for quantity and p for price. We could also write it as $p = 10 - q$, which is equivalent. The latter is called the inverse-demand relation, since it expresses price as a function of quantity, while the demand relation expresses quantity as a function of price. You may have noticed that Figure 2.1 is drawn as the inverse-demand relation (as price is on the y-axis and quantity on the x-axis) rather than the demand relation. For this possibly counterintuitive way economists draw demand curves, blame Alfred Marshall, who laid the foundations for microeconomic analysis in the late nineteenth century (Marshall, 1890).

The simple demand curve in Figure 2.1 allows us to explore some of the main principles that underlie market definition. As noted in Chapter 1, the exact position of the demand curve in the price–quantity space is influenced by two external factors: income,

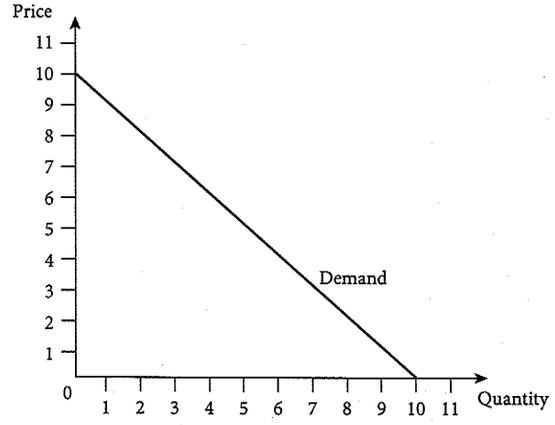

Fig. 2.1 The demand curve

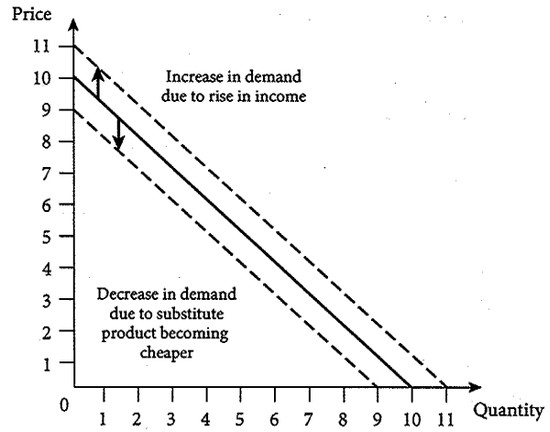

Fig. 2.2 The demand curve shifts

and demand for substitute products. Let's say there is an increase in income in the economy as a whole such that at every price there is now one additional customer who is willing (and able) to pay for women's designer shoes. The demand curve as a whole shifts 1 unit to the right (or 1 unit upwards, which is the same; we could write the new demand formula as $p = 11 - q$). This is shown in Figure 2.2. Now assume that a substitute product—say, mass-produced copies of designer shoes—is reduced in price, such that customers are lured away from designer shoes; at each price, there is now one customer less for designer shoes. This can also be seen in Figure 2.2.

What this really says is that demand for women's designer shoes is not just a function of the price of women's designer shoes, but also of income and of the price of substitute products. It is not unreasonable to assume that such a general relationship can hold for any product in the real world—how much you buy of a product is influenced by its price, by the price of alternative products, and by your income. Now picture a whole demand system in which demand for each good in the economy is expressed as a function of the price for that good, of total income, and of the prices of all other goods in the economy, ie, not just substitutes (economists have developed several types of demand system, but the most common is the Marshallian demand system, after the same Alfred Marshall). For market definition it is obviously the closer substitutes that matter most, but it is useful to bear in mind that within this demand system (but also in the real world) all products are to some extent substitutes of each other—any euro you don't spend on designer shoes you can spend on foreign holidays or French wine instead. This demand system forms the starting point for the hypothetical monopolist test, as we explain further in this chapter. Being aware of this demand system helps you understand the basics of market definition, and also some of the common mistakes that are sometimes made when defining relevant markets.

To help you picture this concept of the demand system, imagine a tablecloth. Each point on the cloth represents a product. Now lift the tablecloth at one specific point, and assume this is like raising the price of that product. The neighbouring products on the cloth are lifted up as well—these are the closest substitutes. The next closest substitutes go up to a lesser extent, and more remote products show no noticeable change. For example, when you lift the tablecloth at the point where apples are, then you may drag up pears and other fruits to some extent as well. Designer shoes are in the demand system too, but probably so far away from the apples that there is no noticeable effect on them from you lifting the cloth. The same would hold if you lifted the cloth at the point where the shoes are—apples are unlikely to be noticeably affected. Market definition is in essence about identifying those products that are close substitutes in the demand system—ie, those products that move together when you lift the tablecloth.[10]

2.3.2 DEMAND RESPONSIVENESS AND ELASTICITY

We noted above that demand-side substitution—ie, by customers—is the most immediate constraint on producers. What matters for market definition is how customers react to changes in price. Other non-price features, such as quality, are generally taken as given for the purpose of market definition (although they may sometimes be considered separately, as discussed in section 2.13). The demand curve in Figure 2.1 shows prices and quantities for a given product quality (changes in quality may have the effect of shifting the curve up or down, just like income and substitute products did in Figure 2.2).

The demand curve captures how customers react to price changes. There are two ways of looking at this: demand responsiveness and demand elasticity. Demand responsiveness can be inferred from the slope of the curve. In Figure 2.1 this slope has an angle of –1 (a 1-unit fall in price leads to a 1-unit rise in demand). If we draw a flatter curve, we would say that demand is more responsive to price (a small price change leads to a bigger change in quantity). The extreme would be a horizontal curve, where even a tiny increase or decrease in price leads to a loss or gain of all demand (this is the demand an individual supplier faces in perfect competition; recall from Chapter 1 that in this case the supplier is a price-taker and faces no choice but to supply at the market price). A steeper curve would represent less responsive demand. The extreme would be a vertical demand curve, where the same quantity is demanded regardless of what price is charged.

A more commonly used measure is the price elasticity of demand. This represents the percentage change in quantity following a 1% increase in price. An elasticity of –2 means that a 1% increase in price leads to a 2% fall in demand. Price elasticity can also be described as the percentage change in quantity divided by the percentage change

[10] Indeed, some authors have proposed a market definition test that focuses purely on which products rise above a certain threshold with the central product on the tablecloth. This would be an alternative to the hypothetical monopolist test, which, in terms of this analogy, focuses on which products prevent the monopolist from lifting the tablecloth. See ten Kate and Niels (2009).

in price. So if a 10% price increase results in a 20% fall in demand, the elasticity is, again, –2.[11] If a 10% price increase results in only a 3% fall in demand, the elasticity is –0.3. More accurately, this measure should be called the own-price elasticity of demand, as it relates the demand for a product to changes in its own price. Another measure that is relevant for market definition is the cross-price elasticity, which represents how demand for one product reacts to changes in the price of another product. If a 10% increase in the price of flights between London and Paris results in a 15% increase in demand for Eurostar train journeys, the cross-price elasticity of rail with respect to air travel between the two cities is 1.5 (15% divided by 10%). Note that the own-price elasticity is usually negative, since price and demand move in opposite directions (in Chapter 1 we said that 'Veblen goods' are the exception to this norm—they are more in demand as they become more expensive). The cross-price elasticity is positive when two products are substitutes—a price increase in one is associated with a demand increase for the other as a result of customers switching (like in the air–rail example above). The cross-price elasticity is negative for what economists call complementary goods—an increase in the price of gin not only reduces demand for gin itself but also demand for tonic water, as people will consume fewer gin-and-tonics. We explain in section 2.4 where in the market definition process own-price elasticities are of relevance and where cross-price elasticities play a role. We return to the role of complements in market definition in section 2.10.

Own-price elasticities play an important part in microeconomic theory and competition analysis. As we explain below in the context of the hypothetical monopolist test, elasticities are directly linked to the price that monopolists set when maximizing their profits. The more elastic the demand to start with, the lower the price that the monopolist can set (intuitively, this is because elastic demand means that customers are responsive to price). If the own-price elasticity lies between 0 and –1, demand is said to be inelastic. Customers do not respond much to price changes—a 10% price increase means customer demand falls by less than 10%. Where demand is inelastic, there is clearly scope for increasing price—indeed, a profit-maximizing monopolist would always do so. In contrast, demand is said to be elastic if the own-price elasticity is greater than 1 in absolute terms—say, –2 or –3 (because the own-price elasticity is a negative number, more elastic means a 'more negative', so smaller, number, which makes it difficult to explain this intuitively as we keep having to add the qualifier 'in absolute terms'). Elastic demand—eg, where a 10% price increase results in a 20% or 30% loss of sales—makes it less attractive for profit-maximizing companies to raise the price.

[11] Below and in section 2.4 we explain that if you raise the price by 10%, the elasticity itself may have changed, so this last interpretation of elasticity, while intuitive and therefore often used, is not strictly correct. From a theoretically pure perspective, an elasticity refers only to one specific point on the demand curve, and infinitesimal price changes from that point.

2.3.3 A HEALTH WARNING ON ELASTICITIES

Own-price elasticities must be used and interpreted with care. Competition practitioners (be they lawyers or economists) do not always bear this in mind. The reason is that the elasticity value depends crucially on the price level at which it is measured. Demand for a product can be inelastic when the price is low, and elastic when the price is high.

To see this, consider Figure 2.3, which shows the same demand curve we saw earlier. The (inverse) demand function for this curve was $p = 10 - q$. We defined the own-price elasticity as the percentage change in demand divided by the percentage change in price. We endeavour to keep algebra to a minimum in this book, but it is useful to explain the steps in calculating elasticities here—see also Table 2.1. The elasticity can be expressed as $\Delta q/q$ divided by $\Delta p/p$ (where Δ is the symbol for a very small change, so the first expression shows the change in q divided by q, which gives the percentage change[12]). This can be rewritten as $\Delta q/\Delta p$ times p/q. Of these last two expressions, the important one to explain our point here is p/q, ie, price divided by quantity.[13] The fact that the elasticity mathematically depends on the ratio between price and quantity implies that it matters where on the demand curve the elasticity is measured. It also implies that the elasticity increases as price increases, ie, as we move further up the

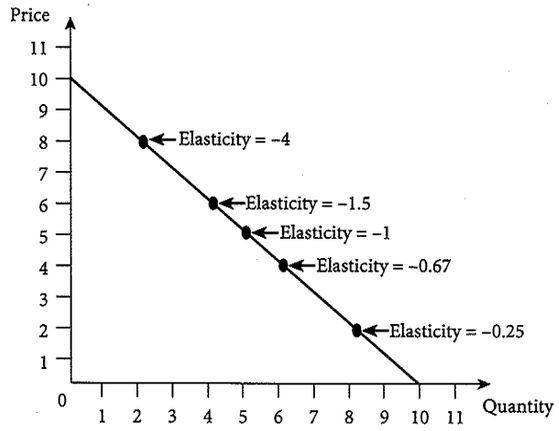

Fig. 2.3 Elasticities vary along the demand curve

[12] Well, it gives the per-unit change. Multiply this by 100% and you get the percentage change. Economists often do not make this distinction explicit as it is easier to speak of the percentage change, but leave out the 'multiply by 100%'.

[13] The other expression, $\Delta q/\Delta p$, represents the functional relationship between price and quantity and is the inverse of the slope of the curve in Figure 2.3 (the slope is $\Delta p/\Delta q$). Technically, $\Delta q/\Delta p$ is the first derivative of quantity with respect to price and equals −1 here (this is because the demand function is $q = 10 - p$; if the demand function had been $q = 10 - 2p$ then $\Delta q/\Delta p$ would have equalled −2).

Table 2.1 An elasticity calculation step by step

Description of the step	Result
Demand curve function	$q = 10 - p$
Inverse demand curve function (shown in Figure 2.3)	$p = 10 - q$
Percentage change in demand (we omit multiplying by 100%)	$\Delta q/q$
Percentage change in price	$\Delta p/p$
Own-price elasticity	$(\Delta q/q) / (\Delta p/p) = (\Delta q/\Delta p) \times (p/q)$
Slope (first derivative) of the demand curve	$\Delta q/\Delta p = -1$
Own-price elasticity at $p = 2$ and $q = 8$	$-1 \times (2/8) = -0.25$
Own-price elasticity at $p = 4$ and $q = 6$	$-1 \times (4/6) = -0.67$

demand curve. Take a point at the lower end of the curve, where price is 2 and quantity is 8. Our formula tells us that the elasticity at this point is −1 (see the last footnote) times 2/8, so −0.25. This implies highly inelastic demand. A bit higher up the curve, at a price of 4 and quantity of 6, demand is still inelastic, but less so—the elasticity is −1 times 4/6, so −0.67. At a price of 5 and quantity of 5 the elasticity equals −1. At any point above that, demand is elastic. So at a price of 6 and quantity of 4 the elasticity is −1.5, and at a price of 8 and quantity of 2 the elasticity is −4. All this is on the very same demand curve that has the same slope throughout.

You should therefore treat with some scepticism general statements such as 'the demand for petrol is inelastic' or 'the demand for foreign holidays is elastic'. The demand for petrol may be relatively unresponsive to price changes—and this may be reflected in a relatively steep slope of the demand curve—but whether demand for petrol is inelastic or elastic depends on the point of measurement, ie, whether petrol prices are already high or low.[14] The important implication of this is that the outcome of the hypothetical monopolist test also depends on the point of measurement—a theme to which we return below.

2.4 BRINGING IN THE SUPPLY SIDE: THE HYPOTHETICAL MONOPOLIST TEST

2.4.1 WHERE DOES THE HYPOTHETICAL MONOPOLIST COME FROM?

You saw before that in the demand system, like on the tablecloth, all products are to some extent substitutes and impose price pressure on each other. Market definition is about drawing the line somewhere such that the relevant market includes only the

[14] One exception to this is where the demand curve is 'iso-elastic', ie, it has the same elasticity along the whole of the curve. This is a demand curve specification that has some interesting theoretical properties but limited

closest substitutes which impose the most significant price discipline on each other. There are various methodologies to determine where and how the line is drawn—and some are more economically sound than others—but you should be aware that the line itself will always have an element of arbitrariness.

One way of drawing the line is to simply lift the tablecloth—raise the price of the product in question by, say, 5%—and include in the relevant market for that product those other products that move up in price as well up to a certain (arbitrary) point, say 4% (assuming that the output of those other products stays constant). This approach would look purely at demand interactions between products within the demand system—it does not say anything about any supply-side factors that might cause such a price change in the first place. The debates leading up to the hypothetical monopolist test in the USA, however, focused on the supply side. They centred around the question: which products constrain suppliers in the market from imposing a price increase? An early idea was that of the hypothetical price-fixing cartel—which firms (and products) would need to be included in the cartel such that a price increase would not be undermined by outside competitors? This idea of the hypothetical cartel then made way for the hypothetical monopolist—after all, monopolizing a market outright is generally more effective than cartelizing it. In this approach to market definition, it is the hypothetical monopolist who brings about the price rise in the demand system, and it is the impact on the profitability of the monopolist that determines the likelihood and sustainability of the price rise, and hence the boundary of the relevant market. If it is not profitable (or profit-maximizing—see below) for a hypothetical monopolist to raise price by a small amount—usually 5% or 10%—that must be because too much demand is lost to other products, and the nearest of those substitute products must be included in the market. In other words, so the logic of this methodology goes, a market is something worth monopolizing. In terms of the tablecloth analogy, the relevant market is formed by those products that prevent the hypothetical monopolist from lifting the cloth by 5%–10%.

First introduced in the 1982 US Merger Guidelines, the wording of the hypothetical monopolist test was refined in the 1992 Horizontal Merger Guidelines, published jointly by the Department of Justice and the Federal Trade Commission:

> A market is defined as a product or group of products and a geographic area in which it is produced or sold such that a hypothetical profit-maximizing firm, not subject to price regulation, that was the only present and future producer or seller of those products in that area likely would impose at least a 'small but significant and non-transitory increase in price', assuming the terms of sale of all other products are held constant. A relevant market is a group of products and a geographic area that is no bigger than necessary to satisfy this test.[15]

[15] Department of Justice and Federal Trade Commission (1992), 'Horizontal Merger Guidelines', s 1.0, reprinted in 4 Trade Reg Rep 104.

Each of these words in the definition has a specific meaning, and we will dissect it bit by bit. In the first sentence you will recognize the product and geographic dimension of the market (and you know from section 2.2 that you should be checking for additional dimensions as well). Also in the first sentence you see the hypothetical monopolist, ie, the 'only present and future' supplier of those products in that area. Importantly, the monopolist is a profit-maximizing firm (just like any firm in a microeconomics textbook), and there is no regulation that would prevent it from raising prices. The hypothetical monopolist test draws the line for market definition by reference to a 'small but significant and non-transitory increase in price'—now commonly known as 'SSNIP'—that the monopolist 'likely would' impose (assuming no price change for other products). We return to this below. Finally, the text implies that there is a distinction between 'a market' and 'a relevant market' as determined by the 'no bigger than necessary' criterion, a subtlety in the hypothetical monopolist test to which we also return later. There may be various 'candidate' markets, but only one relevant market (and no submarkets). Below we consider what a hypothetical monopolist 'likely would' do.

Before that, it is worth making some last comments on the notion of the hypothetical cartel. Although market definition now focuses on the hypothetical monopolist, thinking in terms of cartel price increases can still be useful in certain cases. In a US damages action in 1989 against a concrete producer cartel, the question arose as to whether concrete in West Los Angeles was a relevant market. The plaintiff argued that the success of the cartel itself proved the existence of a relevant market. The court agreed:

> As a purely logical matter, French [the plaintiff] is unquestionably correct. A price-fixing conspiracy confined to manufacturers of concrete would not have been able to succeed if concrete were not a distinct product market: when the cartel attempted to raise prices, customers would simply switch to sand, brick, gravel or some other construction material. Similarly, a price-fixing conspiracy confined to firms in West Los Angeles would not have succeeded if West Los Angeles were not a distinct geographic market: when the cartel attempted to raise prices, customers would simply take their business to East Los Angeles or San Diego or Phoenix. If a group of firms is able to fix prices, it is because their customers have nowhere else to turn. Every price-fixing conspiracy thus identifies directly, in a real world context, a group of firms which is insulated from outside competitive pressures. That is precisely what conventional market definition evidence attempts to identify artificially, by the collection and interpretation of economic data regarding the relationship between various demand curves, by common sense assumptions about the interchangeability of similar products, and the like.[16]

So the success of the cartel was seen as proof that the market was no wider than that for concrete in West Los Angeles.

The notion of the hypothetical cartel has recently made a reappearance in economic thinking on market definition. The 2010 Horizontal Merger Guidelines indicate that the US agencies may employ the concept of a hypothetical profit-maximizing cartel

instead of a hypothetical monopolist if the pricing incentives faced by the actual suppliers in the market differ substantially from those of the monopolist because they sell products outside the candidate market as well.[17] The example given is where the candidate market is one for durable equipment and the suppliers selling that equipment derive substantial revenues from selling spare parts and service for that equipment. We discuss such situations in section 2.10.

2.4.2 WHAT WOULD THE HYPOTHETICAL MONOPOLIST DO?

You start with the product where the competition concern arises in the first place—let's say women's designer shoes as in the merger at the start of this chapter (we return later to the importance of choosing the right starting point for the test—you might instead choose to start with Italian designer shoes if these are sufficiently differentiated from the other varieties). To define the boundaries of the market you then hypothetically monopolize the supply of women's designer shoes. All of a sudden, instead of multiple suppliers (including your two merging parties) there is only one. Luckily, economic theory tells us exactly what the monopolist is going to do in this candidate market.

As explained in Chapter 1, the monopolist will set its price and quantity at the profit-maximizing level. Here we explain in a bit more detail (and with a chart) how the profit-maximizing point is found. In Figure 2.4, which is the standard monopoly representation that you will find in any microeconomics textbook, the monopolist faces the entire (inverse) demand curve for women's designer shoes (since by assumption there are no other suppliers). The demand function is the same as before, represented by the equation $p = 10 - q$. Assume for now (like economists typically do when analysing this standard monopoly situation) that the monopolist does not need to worry about any shifts in the demand curve (as explained in section 2.3, the demand curve may shift up or down with changes in income or in the prices of other products, but the analysis here abstracts from such changes—in the hypothetical monopolist test this abstraction is covered by the assumption that 'the terms of sale of all other products are held constant').

Profits for the monopolist equal revenue minus costs. Some basic maths (which we don't explain here) means that maximum profit is achieved when marginal profit is zero (profits can no longer be increased further), and this occurs at the point where marginal revenue equals marginal cost. We now set out the steps in calculating this optimal point—see also Table 2.2. Revenue equals price times quantity—so $p \times q$, which equals $(10 - q) \times q$. This means that marginal revenue is $10 - 2q$ (the first derivative of revenue with respect to quantity). The marginal revenue curve is shown in Figure 2.4—it starts at the same point as the demand curve (price equals 10 at zero quantity), but then slopes downward twice as steeply. Now suppose that marginal cost equals 2 per unit of output (so total cost is 2q). This is also shown in the figure. Total profit is maximized at $q = 4$, since that is where marginal revenue equals marginal cost ($10 - 2 \times 4$ equals 2). At this point,

[17] Department of Justice and Federal Trade Commission (2010), 'Horizontal Merger Guidelines' 19 August, p.9

Table 2.2 Calculating the profit-maximizing point step by step

Description of the step	Result
Inverse demand function	$p = 10 - q$
Monopoly revenue = price × quantity	$(10 - q) \times q$
Marginal cost per unit	2
Monopoly profit = revenue – total cost	$(10 - q) \times q - (2 \times q)$
Monopoly marginal revenue (first derivative of revenue with respect to quantity)	$10 - 2q$
Monopoly profits are maximized where marginal revenue equals marginal cost	$10 - 2q = 2$, so $q = 4$ and $p = 6$
Monopoly profit at $q = 4$ and $p = 6$ (optimum point)	$(6 \times 4) - (2 \times 4) = 16$
Monopoly profit at $q = 5$ and $p = 5$	$(5 \times 5) - (2 \times 5) = 15$
Monopoly profit at $q = 3$ and $p = 7$	$(7 \times 3) - (2 \times 3) = 15$

price equals 6 (10 – 4), and the profit margin on each unit equals 4 (price minus marginal cost), so total profit on the 4 units sold equals 16. Total profit is shown in the figure as area A. Another way of describing the profit-maximizing point is that it is where the surface of area A is the greatest. If you set price a bit lower, at 5, total profit falls to 15 (as quantity increases to 5 and the profit margin per unit is 3). If you raise price to 7 and the profit margin to 5, quantity falls to 3 and total profit is also 15. Hence, as a monopolist you cannot do better than set the price at 6 and quantity at 4, making a profit of 16.

Recall from section 2.3 that at this profit-maximizing point of $q = 4$ and $p = 6$ the own-price elasticity is –1.5. So the monopolist ends up on the elastic part of the demand curve.

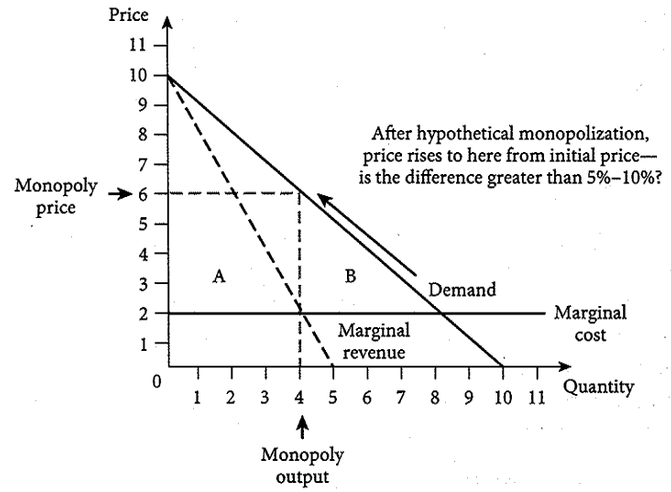

Fig. 2.4 Profit-maximization by the hypothetical monopolist

Contrast this with the perfect competition outcome, where price equals marginal cost, so p = 2, and quantity is 8, and the own-price elasticity equals –0.25. So you are on the inelastic part of the demand curve when there is perfect competition on the supply side. You can see how a monopolist would rapidly move away from there, raising prices and not losing too much custom in the process.[18] Area A in Figure 2.4 reflects a transfer of consumer surplus to producer profit when moving from perfect competition to monopoly. Area B reflects the consumer surplus that is lost because those units are no longer sold even though consumers were willing to pay more than the cost of producing them—area B is the deadweight welfare loss to society resulting from monopoly (the lost consumer surplus accrues to no one; it's wasted) compared with perfect competition.

Back to what this means for market definition. You have seen now at what price and quantity point you end up after hypothetically monopolizing the market. It is the point which the profit-maximizing monopolist 'would' choose. All you need to do next is compare that point with where you started before monopolizing the market, and measure how much the price has changed—is this change 'small but significant'? If the starting point was perfect competition, monopolization of the market in our example would lead to an increase from p = 2 to p = 6, so by 200%—clearly a large and significant increase in price. However, the starting point does not need to be perfect competition; the market pre-monopolization may have been characterized by some form of oligopolistic competition, such that price was already higher than marginal cost. If the market price pre-monopolization was, say, 5, the hypothetical monopolist would impose a price increase of 1 monetary unit to 6, so 20%, which is more than the 5%–10% range for SSNIP. If the market price was 5.75, the change to 6 would represent a 4% increase, so less than 5%–10%. If the market price was 6, monopolization would not result in any price increase. You can see that it therefore matters where exactly you start before applying the hypothetical monopolist test. If the product in question is already priced at the monopoly level, a hypothetical monopolist would not impose any further price increases as that would not maximize its profit. That does not imply that the market is not worth monopolizing, because it clearly is in this case. This result, and the erroneous conclusion you can draw from it, is known as the cellophane fallacy, which we discuss further in section 2.6.

2.4.3 WHY THE 'SS' IN SSNIP?

The term 'SSNIP' has become widely known in competition law. Indeed, the terms 'SSNIP test' and 'hypothetical monopolist test' are used interchangeably. Apart from raising an eyebrow at this somewhat strange abbreviation to describe a price increase, you may have wondered why it has to be a 'small but significant' increase in the first place. Is this for some scientific reason, or is it purely arbitrary? Would the test still work with a large price increase? Another question that springs to mind is why 'small but

[18] This is not always the case. Depending on the level of marginal cost, there are markets where even with perfect competition you would be on the elastic part of the demand curve, such that a hypothetical monopolist might not raise prices much further. Products like that are less likely to be worth monopolizing, as we discuss in section 2.12.

significant' is usually interpreted as 5% or 10%? Are there any reasons to prefer 5% over 10%, or vice versa? You will recall from the earlier sections that substitution between products is a matter of degree. It is not all or nothing, one–zero. If you lift the tablecloth high enough, many remote products will move up as well. If you raise the price of women's designer shoes far enough, there comes a point when even the cheaper mass-produced shoes become an attractive alternative. Any cut-off point for such a price increase is inherently arbitrary. You could perform a market definition test that asks if a hypothetical monopolist would raise price by 20%, 50%, or 100%. Conceptually, these are all valid tests. Why do most competition authorities use 5%–10% as the threshold? There are various reasons for this—policy-related, technical, and practical.

First, the policy-related reasons. Competition law is generally concerned with mergers and anti-competitive practices that reduce competition in the market and hence lead to higher prices. How high is too high? Some competition authorities consider that even a 5% increase is already of significant detriment to consumers. For example, the UK Competition Commission stated in its previous (2003) Merger Guidelines:

> The Commission will normally use 5 per cent for the SSNIP test, rather than the more common 5–10 per cent, because in many instances an increase in the price of a product of around 5 per cent (with all other prices unchanged) might reasonably be judged to have a significant effect on customers' expenditure on the product and so provides an appropriate level at which to consider the test. In addition, a 5 per cent increase in price might be expected to have an appreciable effect on a firm's profit margin, the main issue then being whether demand would be reduced to such an extent as to offset the effects of the higher margin.[19]

Of course the hypothetical monopolist is not real and does not impose price increases for real. But the SSNIP threshold is related to the more general concern about price rises. An authority, like the Competition Commission, that applies a 5% threshold can be said to be stricter than an authority that applies a 10% or 20% threshold. A higher threshold for the SSNIP means a greater chance of broader markets being defined and hence a lower likelihood of a finding of market power. Say you found that the hypothetical monopolist for flights between London and Paris would raise prices by 8% to maximize profits. A 5% threshold would mean the SSNIP test is passed and flights between London and Paris are a separate market (and hence you are more likely to consider the merger being investigated to be problematic). In contrast, under a 10% or 20% threshold the test is not passed, and you would include more substitute products in the relevant market. Again, there is no scientifically correct cut-off point, but most competition authorities would generally be concerned about price increases of 5%–10% (and some are concerned at only 5%), so there are policy reasons to define 'small but significant' as 5%–10%.

[19] Competition Commission (2003), 'Merger References: Competition Commission Guidelines', [2.8]. The 2010 Guidelines still prefer 5% but are more open to different thresholds: 'When applying the hypothetical monopolist test, the Authorities will normally use a SSNIP of 5 per cent, though they may sometimes use a higher or lower number.' Competition Commission and Office of Fair Trading (2010), 'Merger Assessment

Second, there are some technical reasons for using 5%–10% for the SSNIP rather than a higher threshold. How far the monopolist's profit-maximizing price lies above the initial price depends on the initial own-price elasticity. The more inelastic demand is (before hypothetically monopolizing the market), the more the monopolist will raise prices (and hence the more likely it is that the 5%–10% SSNIP threshold will be met). As explained in section 2.3, elasticities vary depending on the price point at which they are measured. Economists can usually estimate elasticities only at current price levels, not at the (hypothetical) monopoly price level. The estimated elasticities can therefore be used only to test very small price increases above the current price level—hence 5% or at most 10% for the SSNIP threshold. Beyond that it is difficult to know how quickly demand becomes more elastic (so how quickly the profit-maximizing price is reached)—a 20% SSNIP threshold would lead to uncertain results because the elasticity value might differ too much between the starting price point (where we can measure the elasticity) and the point to which price is increased to maximize profits (for which we don't have an estimate). Following this logic, the hypothetical monopolist test would be most accurately applied if the threshold used were 1% or even less. However, from a statistical perspective such a very small increase may not be noticeable or distinguishable from 'noise' in the price data. In statistics, a 5% cut-off point is often used to determine statistical significance, for no 'hard' reason other than 5% being sufficiently different from zero (it is 'significant') while at the same time still being a 'small' number. So, for pragmatic reasons, 5% (or 10%) is neither too small nor too large; it seems about right as a threshold for the SSNIP test.

Third, there are practical reasons for using a SSNIP of 5% or 10%. The hypothetical monopolist test is frequently applied using critical loss analysis (see section 2.5). This in turn often uses evidence from consumer surveys in which respondents are asked how they would react to a SSNIP. Surveys are difficult (but not impossible) to design in such a way that you get objective and meaningful answers (see section 2.14). Some people find it difficult to think in percentages. The clearest way of asking the SSNIP question may therefore be to refer to a 10% price increase, and relate this to a price the respondents actually, or typically, pay. For example, if a commuter train ticket costs €7.50, you can ask what consumers would do if it were raised by 10% to €8.25 (for a period of one or two years; see the discussion on the 'N' in SSNIP below). A hypothetical 5% increase to €7.88 (rounded) or 7.5% increase to €8.06 might be more difficult for survey respondents to relate to. At the same time, there are practical reasons why using 5% for a SSNIP is not necessarily better or worse. The degree of precision with which elasticities and customer switching rates after small price increases can be measured in practice is inherently limited. Any distinction between 5% and 10% may be easily outweighed by the uncertainties in the empirical estimates and would therefore be spurious.

2.4.4 WHY THE 'N' IN SSNIP?

The other aspect of the SSNIP is the 'non-transitory' nature of the price increase. The 'N' in SSNIP is as arbitrary as the 'SS' and the cut-off point is usually chosen for policy and

practical reasons. The main rationale for considering a 'non-transitory' as opposed to a transitory price increase is that competition authorities are not normally concerned with transitory market power. Positions of market power tend to be eroded over time (more on this in Chapter 3). If it suddenly starts to rain in a busy market place, vendors who sell umbrellas find themselves controlling a scarce good that is in strong demand, such that they can probably extract a hefty monopoly rent. The same is true for car hire and taxi companies when the European airspace is closed due to volcanic ash, and stranded passengers across Europe desperately try to get home.[20] Such a position of market power will not last, and does not usually merit a separate market definition by time of consumption (the first additional market dimension discussed in section 2.2). Only market power that persists over time is of concern. The small but significant increase in price must therefore be 'non-transitory'. One exception to this is the electricity generation market, where authorities have been concerned about very short, 30-minute, periods of market power and have defined relevant markets accordingly (these 30-minute periods recur more frequently and predictably than the periods of rain or volcanic ash clouds, so in reality in electricity generation the authorities are concerned about a series of regular short periods of market power over a longer time period).[21]

How long is non-transitory? Competition authorities usually refer to a period of one to two years. In some markets, customers may sign contracts with providers for a minimum period, say, six or twelve months. By interpreting non-transitory as one or two years, you would still take into account the fact that customers may switch once the contract expires, even if they can't switch before then. Longer-term contracts, say, for three or five years, are not taken into account—customers on these contracts will be in the hypothetical monopolist's pocket for the duration of the non-transitory period. The length of the period has implications for how strict a competition authority is. A three- or five-year period would imply a less strict approach than one or two years, as market dynamics are given more time to undermine the monopolist's price increase. The hypothetical monopolist for a commuter rail service may find it profitable to raise fares for one year, but not much longer, because at some point beyond one year too many customers would begin to use alternative modes of transport (eg, they might decide to buy a car). If the 'N' is taken as one year, this would still mean a separate relevant market for the commuter rail service, but if the 'N' is greater than one year the market is wider. In terms of the practical reasons for choosing a particular 'N', as with the 'SS', the analysis is likely to be more precise if you take one or two years than if you take longer periods. This is because market dynamics are difficult to predict further into the future—beyond one or two years any analysis is likely to become speculative. Likewise, the profit-maximizing price over a one-year or two-year period would depend on which costs are marginal (variable with output) over the period. Recall that the longer the time

[20] One of us paid €3,000 for a taxi ride from Rome to Paris during the April 2010 volcanic ash incident. Economists do not learn about economics only from textbooks.

[21] Competition Commission (2001), 'AES and British Energy'; and Office of Fair Trading and Ofgem (2005),

period, the more costs become variable, so it becomes more difficult to predict where the monopolist would set the profit-maximizing price.

The Horizontal Merger Guidelines in the USA do not specify the length of 'N'—they refer to 'lasting for the foreseeable future'. The UK Competition Commission's previous (2003) Merger Guidelines stated that the SSNIP 'is assumed to last for the foreseeable future', but then added that the authority 'will typically consider the extent of response which is likely to occur within a year of the price rise (although the exact time period will depend on the nature of the market considered)'.[22] The Canadian Competition Bureau's Merger Guidelines define non-transitory as one year (although with a footnote caveat that market characteristics may sometimes require a different period).[23] Curiously, the European Commission Notice on market definition refers to a 'permanent' increase in price.[24] In line with the description above, this would imply that the Commission takes greater account of long-term market dynamics and tends to define wider markets as a result. However, another possible explanation for the European Commission text would be that the term 'permanent' is simply an alternative (and somewhat inaccurate) way of capturing the concept of 'non-transitory'.

2.4.5 AN ITERATIVE PROCESS: FROM THE FOCAL PRODUCT TO THE 'SMALLEST' MARKET

We have described what the hypothetical monopolist test is: would the hypothetical monopolist impose a SSNIP? Here we discuss the importance of selecting the right start and end point for the test, ie, the right product and geographic area.

Let's begin with the end point. The US Merger Guidelines make a subtle distinction between the definition of a market—which is where a hypothetical monopolist would impose a SSNIP—and the relevant market—which is 'a group of products and a geographic area that is no bigger than necessary to satisfy this test'. In other words, the relevant market is the 'smallest' market in which the monopolist would impose a SSNIP.[25] A hypothetical mobile phone monopolist operating in the whole of Europe would be able to impose a SSNIP, so the whole of Europe would be a market. But a hypothetical mobile phone monopolist in Germany would probably also be able to impose a SSNIP. Which of these two candidate markets is the relevant market? Let's say it's the whole of Europe. We then find that there are so many mobile operators in Europe that none of them has significant market power, and conclude that there is no competition concern. This would be an erroneous conclusion. Within Europe there is at least one 'pocket' of

[22] Competition Commission (2003), 'Merger References: Competition Commission Guidelines', [2.7]. The 2010 Guidelines do not indicate how 'non-transitory' should be interpreted. Competition Commission and Office of Fair Trading (2010), 'Merger Assessment Guidelines', September.

[23] Competition Bureau (2004), 'Merger Enforcement Guidelines', September, [3.4].

[24] European Commission (1997), 'Notice on the Definition of Relevant Market for the Purposes of Community Competition Law', 97/C372/03, [17].

[25] The 2010 Horizontal Merger Guidelines no longer refer to this smallest market principle. We think that

market power—in this example, Germany—where a hypothetical monopolist (and possibly real companies as well) can raise prices. Therefore, Germany should be the relevant market. By taking the smallest market you avoid overlooking pockets of market power. It also means that you don't need to worry about submarkets within the relevant market.

So the end point for market definition is to find the smallest market in which the hypothetical monopolist would impose a SSNIP. But what is the starting point? Recall that market definition is not an end in itself, but only an intermediate step in the analysis of competitive constraints. For market definition to be informative, the starting point must always be the product in relation to which the competition case arises. If the case concerns mobile telephony services in Germany—say, two German operators wish to merge—the first candidate product and geographic area to hypothetically monopolize would be mobile telephony in Germany. If the concern arises from a proposed merger between two producers of frozen dessert pies in north-east USA, then that is the hypothetical monopolist's initial domain. The starting point for a relevant market is sometimes called the focal or labelling product (and geographic area).

The focal product can be narrowly defined. After all, you are trying to find the smallest market in which a SSNIP can be imposed in order to identify pockets of market power. Thus, in the merger case at the beginning of this chapter you can start with Italian women's designer shoes (both our merging parties own Italian brands), and hypothetically monopolize that market so as to test whether they face competitive constraints from French or Spanish designer shoes—if the monopolist would impose a SSNIP, the relevant market is indeed confined to Italian designer shoes. However, in section 2.1 we also warned that in markets with differentiated products the whole market definition exercise may become uninformative, especially if the focal products become too narrow (Porsche 911 Carreras, or Prada shoes). We recommend that in these situations you take a view on whether the products concerned are so differentiated that market definition loses value as an intermediate step (because you potentially end up finding that each differentiated product is worth monopolizing and hence in theory constitutes a separate relevant market). If differentiation is that significant, it may be more appropriate to focus the analysis directly on the competitive constraints and closeness of competition between the differentiated products, skipping market definition altogether (we return to this later in the chapter and in Chapter 7 on mergers). However, we think that there is often no need for skipping market definition, because in many differentiated product markets you can still find meaningful groups of products that are reasonably similar and that you can hypothetically monopolize to test the constraints from other groups—for example, top-of-the-range sports cars, Italian women's designer shoes, or even all women's designer shoes. There is an element of judgement that you may have to exercise here.

It is also important to bear in mind that in any competition case there can be more than one focal product (or area). Indeed, you should in principle define a relevant market for any product (or area) where you might suspect a possible competition concern. If a German and a UK mobile telephony operator wish to merge, you have to define a relevant market separately for mobile telephony in Germany as the focal product and mobile tele-

(Note that in this case you may well find that neither relevant market has the other in it, and hence that there are no competitive overlaps between the merging parties. This is one of the rare situations in which a narrow market definition, ie, separate national markets rather than a cross-border one, actually favours the merging parties; normally broader markets favour merging parties, and hence merging parties favour broader markets.)

Likewise, if a producer of frozen dessert pies merges with a producer of ice cream, you have to apply the hypothetical monopolist test to each of these products separately. The two exercises may well lead to the same market definition that has both products in it, for example, a market for all desserts. However, you might also find that substitution is strong in one direction but not the other. Say that frozen dessert pies are not worth monopolizing because consumers switch en masse to ice cream and other desserts if the price is raised, while at the same time, in the market with ice cream as the focal product, a hypothetical monopolist does find it profitable to impose a SSNIP because fewer consumers switch away to other desserts. So ice cream is in the market for frozen dessert pies, but frozen dessert pies are not in the market for ice cream. Alternatively, you might find that ice cream is not worth monopolizing but that fruit-flavoured yogurt is the nearest substitute, so the relevant market for ice cream contains ice cream and fruit-flavoured yogurt (more on the ranking of substitutes below). Again, frozen dessert pies are not in the market for ice cream. If you had started the market definition exercise from only one of these products, you would have observed only a partial view of substitution between the products. If you had defined the market starting only from ice cream, you might have erroneously concluded that the merger is of no concern because frozen dessert pies are in a separate market, as you would have overlooked the fact that, the other way round, ice cream does impose a competitive constraint on frozen dessert pies.

2.4.6 ASYMMETRIC MARKETS

Markets can be asymmetric in this regard—one product may be in the market for another product, even if that other product is not in its market (there is debate among economists about how often such asymmetry occurs, but it is a theoretical possibility and has been found in some competition cases, as discussed below). That is why it is important to make clear which is the focal or labelling product of a relevant market. It is also why a generally worded question such as 'Are frozen dessert pies and ice cream in the same relevant market?' is meaningless—the answer depends on what the focal product is. The right way of asking the question is: 'Is ice cream in the relevant market for frozen dessert pies?', and 'Are frozen dessert pies in the relevant market for ice cream?' In our hypothetical example, the answer to the first question was yes and the answer to the second question no. A case where asymmetric markets were found is the *Bayer–Aventis Crop Science* merger (2000), involving agricultural crop-protection products.[26] The European Commission found evidence of substitution from foliar and

[26] European Commission Decision of 12 July 2000 in Case COMP/M.2547

soil applications of fungicides and insecticides to seed treatment, but not the other way round (from seed treatment to the other applications). The Commission also found another instance of 'one-way substitution'—between two specific types of cereal crop-protection products. Asymmetric markets have also been found in various supermarket inquiries in the UK—larger stores were considered to constrain smaller stores, but not the other way round.[27]

A case where the authority focused too much on the question of whether two products are in the same market, as opposed to whether one is in the market for the other (focal) one, is the 2001 review of competition in the mobile market conducted by Oftel, the UK telecoms regulator (now Ofcom).[28] The focal product was mobile telephony, and one of the market definition questions addressed was whether fixed telephony was a substitute for mobile. The authority presented some consumer evidence suggesting that the constraint from fixed on mobile was limited (eg, the vast majority of mobile calls were short 'convenient' calls of the type that, by definition, could not be made from a fixed line—for example, a call to your partner from the supermarket to clarify the shopping list). But Oftel then also, incorrectly, discussed why mobiles posed limited constraints on fixed telephony—most users still saw fixed telephony as their main method of making calls as the quality of fixed telephony was higher, and fixed users said they would not switch to mobiles even after a large price increase in fixed telephony. It concluded on this basis that fixed and mobile telephony are in separate markets. However, these last pieces of evidence are not informative about whether mobile telephony is constrained by fixed; they would be more relevant in a competition review where fixed telephony was the focal product. A case where an expert made a similar mistake is the 2007 abuse of dominance case before the English High Court concerning local bus services. As pointed out in the judgment:

> [Counsel for the defendants] also made what I regard as a fair point of criticism of [the claimants' expert's] analysis. In any analysis of whether local buses form an exclusive product market, the usual approach is to hypothesise a small but significant non-transitory increase in price for bus services of 5 to 10% and determine what alternative modes of transport (if any) become a substitute. That is the right approach, whereas [the claimants' expert] appeared to regard it as equally relevant to consider whether buses were a substitute for cars. Buses may be competitively constrained by cars, but cars may not be competitively constrained by buses.[29]

The distinction between the focal product and other products is also of utmost importance in later rounds of the test, where the hypothetical monopolist controls multiple products and the question often arises as to whether the SSNIP then applies to all products or only to the focal product. The distinction is not explicitly made in the original formulation of the hypothetical monopolist test in the US Merger Guidelines,

[27] For example, Competition Commission (2005), 'Somerfield plc/Wm Morrison Supermarkets plc', September.

[28] Oftel (2001), 'Effective competition review: mobile', 26 September, Annex 1.

[29] *Chester City Council and Chester City Transport Limited v Arriva PLC* [2007] EWHC 1373 (Ch), [157]. We

as quoted above. Ten Kate and Niels (2009) therefore proposed a slightly amended wording of the test, as follows (the amendments to the original are in italics):

> A market *for a product at a specific location* is defined as *that* product or a group of products *containing that product* and a geographic area in which it is produced or sold *covering that location*, such that a hypothetical profit-maximizing firm, not subject to price regulation, that was the only present and future producer or seller of those products in that area likely would impose at least a 'small but significant and non-transitory increase in price' *for that product at that location*, assuming the terms of sale of all products not in that group or not in that area are held constant. The relevant market *for a product at a location* is the group of products and geographic area that is no bigger than necessary to satisfy this test.

2.4.7 GETTING TO THE SMALLEST MARKET

You have now seen what the starting and end points are for the hypothetical monopolist test. The way to get from one—the focal product and area—to the other—the smallest market in which a SSNIP would be imposed—is through a process of iteration. You would normally begin with the product market. The geographic market comes later (see section 2.9 on how it is important that the product and geographic market must remain linked in a meaningful way).

You hypothetically monopolize the focal product—say Italian women's designer shoes. The hypothetical monopolist will set the price at the profit-maximizing level (which is likely to be higher than the existing price level because the different producers of Italian women's designer shoes now no longer compete with each other). You then apply the SSNIP question to this candidate market: is this new profit-maximizing price more than 5%–10% higher than the existing price? If the answer is yes, you have found your relevant product market. If the answer is no—for example, the price increase would only be 3%—you must expand the market. Recall that if a SSNIP would not be imposed, that must be because consumers switch to other products in the demand system—it does not matter whether this is switching to close substitutes, such as French designer shoes, or spending income on more remote substitutes, such as theatre tickets or expensive bottles of wine. In keeping with the aim of finding the smallest market, you take the closest substitute for the focal product—say, French women's designer shoes—and bring it within the realm of the hypothetical monopolist. You then apply the SSNIP question again, this time to the new group of products controlled by the hypothetical monopolist. If the answer is now yes, you have found your relevant product market. If the answer is still no, you proceed to the third iteration of the test by including the second closest substitute in the realm of the hypothetical monopolist. And so on.

In reading this explanation of the iterative process, three questions may have occurred to you that we have not yet answered. First, how do I actually assess whether a SSNIP would be imposed? Second, if the test fails in the first iteration, how do I determine what the closest substitute is? Third, from the second iteration onwards, do I apply the SSNIP question to all products controlled by the monopolist or only to the focal product? We answer the second and third questions next in this section. The first question is addressed

in section 2.5 where we deal with critical loss analysis. Before that, however, it is worth pointing out that while in theory the hypothetical monopolist test can have several iterations, in practice you'll find that quite often one iteration is sufficient. If you conclude that a hypothetical monopolist of Italian designer shoes would not find it profit-maximizing to impose a small price increase, in all likelihood any real producer of such shoes—including the merged entity—would also not be able to raise price. Hence you have the answer to your competition question, and there is no need to delineate exactly where the boundaries of the relevant market are beyond Italian designer shoes.

2.4.8 RANKING SUBSTITUTES: CROSS-PRICE ELASTICITIES AND DIVERSION RATIOS

As you have by now read several times, many products within the demand system are to some extent substitutes for each other. But clearly some are closer substitutes than others. For the purpose of the hypothetical monopolist test, you need to rank substitute products by how close they are to the focal product. The closest substitute is normally the product that absorbs most of the sales that are lost by the hypothetical monopolist after imposing a price increase. That product poses the strongest competitive constraint on the focal product and is the one that the monopolist, if given the choice, would like to get its hands on most. Once that closest substitute product is grouped with the focal product, the hypothetical monopolist no longer cares about losing sales from the latter to the former because it controls both products. The most significant obstacle to raising the price of the focal product has thus been removed (internalized, as economists would say), and its price will increase more than in the first iteration. You are therefore more likely to find that the SSNIP threshold has been passed. Including the closest substitute first (in the second iteration of the test) is important to meeting the smallest-market criterion. If you included a more remote substitute, you might find that this group of products is still not worth monopolizing and hence continue expanding the market, potentially too widely.

How do you rank substitutes? The standard measure for the degree of substitutability between products is the cross-price elasticity, as we saw in section 2.3. It is defined as the percentage change in the quantity of one product divided by the percentage change in price of the other product. For substitute products, the cross-price elasticity is positive, while for complementary products it is negative. For any two products there are two different cross-price elasticities—the cross-price elasticity of demand for rail services with respect to the price of air travel, and the cross-price elasticity of demand for air travel with respect to the price of rail services. Just as markets can be asymmetric (as you saw above), cross-price elasticities can differ between two products (indeed, an asymmetry between markets is the result of such a difference in cross-price elasticities). So the cross-price elasticity of demand for rail services with respect to the price of air travel may be 1.5, while the other way round it may be 0.5. Which of the two elasticities are you most interested in? For the purpose of the hypothetical monopolist test, the

London and Paris because there is a proposed merger between two airlines operating on this route, in the hypothetical monopolist test we are analysing the effects of price changes in air travel between London and Paris (as imposed by the hypothetical monopolist). The relevant cross-price elasticity is therefore the one of demand for rail services between London and Paris with respect to the price of air travel—not the one of demand for air services with respect to the price of rail.

Now consider how cross-price elasticities work in the ranking of substitutes. Assume that the own-price elasticity for flights between London and Paris equals –3, and that the hypothetical monopoly airline would therefore not impose a SSNIP (see section 2.5 on critical loss analysis). Say there are currently 500,000 passengers per year on this route. A 10% price increase for one year would result in a 30% fall in passengers. Where do these 150,000 lost air passengers go? We know that the cross-price elasticity of demand for rail services between London and Paris with respect to the price of air travel is 1.5, so the 10% price increase by the monopoly airline would lead to a 15% increase in the demand for rail. The other possible substitute product is car journeys, including a ferry or Eurotunnel journey across the Channel. Say the cross-price elasticity of car journeys between London and Paris with respect to the price of flights is 0.5. Based on cross-price elasticities alone, rail would be the closest substitute. This is one way of ranking substitutes.

Another way would be to assess which substitute product absorbs the greatest proportion of the lost passengers. If 400,000 rail journeys are normally made each year, this increases to 460,000 after the 10% price increase in air travel (60,000 is 15% of 400,000). So 60,000 of the 150,000 lost air passengers switch to rail. If the total number of car journeys between London and Paris each year is 1.4 million, the 10% price increase for air travel means 70,000 (5% of 1.4 million) additional car journeys. By this measure, car journeys are a closer substitute for air travel than rail, because the greatest proportion of the lost air passengers switch to cars. This proportion is commonly known as the diversion ratio—the diversion ratio from air travel to car journeys is 47% (70,000 divided by 150,000); that from air travel to rail is 40% (60,000 divided by 150,000). Diversion ratios are frequently used in merger analysis, especially where differentiated products are concerned (see Chapter 7)—they are a useful measure of the closeness of substitute products, just as they are useful here for ranking substitutes for the purpose of the SSNIP test.

The ranking of substitutes should be straightforward in most cases. In the above example it was slightly complicated because the ranking by cross-price elasticity differed from that by diversion ratio. In practice this is not likely to occur often (diversion ratios are a function of the cross-price elasticity and the size of the market, so only where sizes of markets differ substantially—as here, where there are many more car journeys than rail or air journeys—do you get such differences in order). In any event, in cases such as these where there are two important close substitutes, the precise ranking may not make too much of a difference to the conclusion. Regardless of whether rail or car journeys are the closest substitute, the analysis has already found that air travel between London and Paris itself is not a market worth monopolizing. Hence, if the competition concern arose solely in air travel (say, a merger between two airlines

serving the route), you can quickly be reassured that the concern is limited because air travel faces competition from other modes of travel. If the competition concern arose across air and rail—for example, an airline forms a joint venture with a rail operator on the route—you would treat both air and rail in turn as the focal product and assess whether each is in the other's market. When looking at air as the focal product—which is one of the analyses you would undertake—which substitute do you include in the market for air once you have established that air travel by itself is not worth monopolizing? In this case you should probably try both ranking methods to avoid overlooking any competition concerns. Based on diversion ratios, car journeys (including the Channel crossings) are the closer substitute, and you might find that an air–car monopolist would impose a SSNIP. In this case you might conclude that air and rail are not in the same market and hence that the joint venture raises no competition concerns (this also depends on whether air is excluded from the market when you take rail as the focal product, which is the other analysis you would undertake). However, ranked by cross-price elasticities, rail is the closest substitute, and you might find that an air–rail monopolist would also impose a SSNIP (even if this price increase is somewhat smaller than the one imposed by the air–car monopolist). Air–rail is technically the smaller market of the two (a total of 800,000 passengers versus 1.5 million passengers), and hence would be the relevant market, and you would find that the joint venture raises potential competition concerns.

There are two further points to make in relation to the ranking of substitutes and the use of cross-price elasticities. First, in the context of market definition, any sales loss matters to the hypothetical monopolist since it makes the price increase less profitable. In other words, it does not matter to the monopolist whether sales are lost to close substitutes or remote substitutes. In the above example, 130,000 of the 150,000 lost air passengers switched to rail and car journeys; the remaining 20,000 may no longer make the journey at all (and spend their money on other things). Market definition is therefore not only about analysing switching to substitute products. When you design a consumer survey to assess market definition, it is important that you explore whether consumers simply buy less of the product, in addition to whether they switch to substitute products.

Second, it follows from the above that own-price elasticities are of greater interest in market definition than cross-price elasticities. The analysis of critical loss—explored in section 2.5—revolves around own-price elasticities, not cross-price elasticities. This has not always been understood in competition law. Before the hypothetical monopolist test was developed, markets were frequently defined according to whether the cross-price elasticity between two products was high or low—see, for example, the quote from the *Brown Shoe* ruling in section 2.1. The cross-price elasticity in itself provides insufficient information about market definition. Its role is in the ranking of substitutes for the purpose of the hypothetical monopolist test, and this is a rather secondary role. Nonetheless, a measurement of the cross-price elasticity may provide a useful sense-check on the own-price elasticity. This is because there is a theoretical relationship between the two (which we do not elaborate upon here)—within the Marshallian demand system, there is a general rule that the higher the sum of all of a product's cross-price elasticities, the higher (more negative) that product's own-price elasticity will be

If you see an estimate of own-price elasticity that suggests very elastic demand for a product (a 10% price increase causes a large sales loss), you should also expect to see that product having some high cross-price elasticities with other products, since that sales loss will be mainly absorbed by other products in the demand system.

2.4.9 THE SECOND ITERATION ONWARDS: WHICH OF THE MONOPOLIST'S PRODUCTS DOES THE SSNIP QUESTION APPLY TO?

This is a frequently asked question, and has not been answered very well in competition authorities' guidance or other texts. And yet it is straightforward: even if the monopolist controls multiple products, the SSNIP question should apply only to the focal (labelling) product as that is the product where the competition concern arises in the first place. Let's return to our Italian women's designer shoes. This is the focal product because two producers of these shoes want to merge. In the first iteration of the test you found that the hypothetical monopolist would impose a 3% price increase—insufficient to constitute a SSNIP. You find that French designer shoes are the closest substitute for the Italian variety. In the second iteration of the test the monopolist controls both Italian and French women's designer shoes. As a result, it now no longer cares about losing sales to French shoes when raising the price of Italian ones—the main constraint on raising Italian shoe prices has disappeared (in economic terms, it has been internalized by the monopolist). At the same time, the monopolist can now raise prices for both of these products if required to maximize profits (remember that in the first iteration the assumption was that the terms of sale of all other products, including French shoes, are held constant). Say the hypothetical monopolist would now impose a price increase on Italian shoes of 13%—a SSNIP. But French shoes increase in price by 4% (the profit-maximizing price, obtained in the same way as in Figure 2.4). So in this example there is no SSNIP on French shoes, despite the monopolization of their supply, the reason being that they face strong competition from other women's designer shoes that the monopolist does not control.

Must you therefore proceed to a third iteration of the test, where the group of products controlled by the monopolist includes all women's designer shoes? Suppose you did, and you found that the price of Italian shoes now increases by 15% and that of French shoes by 6% (the monopolist no longer cares about switching from Italian and French to other designer shoes, so can flex prices up a bit further), while the other shoes—monopolized for the first time in this third iteration—increase in price by only 2%. Following the same logic that made you proceed to the third iteration—ie, you require that *all* the hypothetical monopolist's prices must increase—you would eventually reach the conclusion that the market should include all women's shoes, in which the merging parties have a combined market share of only 5.5%. Alas, the logic and the conclusion are wrong. Already in the second iteration, you found a pocket of market power in Italian women's designer shoes. As soon as French shoes, their closest substitute, are added, the monopolist would impose a SSNIP (or even more) on the

Italian shoes. The relevant market for Italian women's designer shoes includes the Italian and French varieties, but no more. It is a market worth monopolizing, especially for the producers of Italian shoes (such as the two merging parties). The fact that you discover in the second iteration that French shoes in themselves are not a market worth monopolizing (even when taken together with Italian shoes) is irrelevant. You are focused on the scope for price increases in Italian designer shoes, because that is where the competition concerns from the merger arise. So in the second and any subsequent iterations of the hypothetical monopolist test, the SSNIP question is still applicable only to the focal (labelling) product.

If, instead, the merger is between a producer of French shoes and a producer of Italian shoes, you take both as the focal product and carry out two separate market definition exercises. The SSNIP question still applies to the focal product only. For the Italian shoes as focal product you reach the same conclusion as above—the relevant market is that for Italian and French shoes. When French women's designer shoes are the focal product, other designer shoes impose a significant constraint (the monopolist imposes only a 4% price increase, even if Italian shoes are included). The relevant market for French women's designer shoes becomes worth monopolizing only when all designer shoes are included—a 6% price increase is then imposed on the focal product, as we saw earlier.

2.4.10 PURITY, WITH SOME PRAGMATISM

Thus far we have considered the hypothetical monopolist test in its purest form, which is based on the definition of the test in the 1992 Horizontal Merger Guidelines. There are two major factors in reality that put this theoretical purity under pressure, and call for some pragmatic flexibility if the hypothetical monopolist test is to remain a useful concept for market definition. One factor is the need to put the test into practice. It is difficult enough to estimate elasticities and consumer responses to price changes (more on this in section 2.14). But once you have an estimate, identifying the profit-maximizing price that the monopolist would set gives rise to additional complications. It is easier—as we explain in the next section—to ask simply whether it would be profitable for the hypothetical monopolist to impose a SSNIP, rather than whether it would be profit-maximizing. This is the essence of critical loss analysis, a commonly used method to make the hypothetical monopolist test workable.

The other factor is that, as we mentioned before, in most markets there is a degree of product differentiation. The test works best when products are homogeneous. One pragmatic adjustment we have already discussed is that, for the purposes of the hypothetical monopolist test, you can to some extent ignore differentiation and group together products that may still be considered reasonably homogeneous (top-of-the-range sports cars, or women's designer shoes—no two products in these groups are exactly the same, but overall they constitute reasonably identifiable categories). There are markets, however, where the degree of differentiation is so high that the hypothetical monopolist test in its purest form becomes too uninformative. If that is the case, you should either skip market definition altogether and focus directly on competitive constraints (see section 2.15 and

Chapter 7 on mergers), or you could try some of the critical loss techniques for differentiated products that economists have developed relatively recently (see below).

2.5 CRITICAL LOSS ANALYSIS: 'COULD' OR 'WOULD'?

2.5.1 TWO DIFFERENT HYPOTHETICAL MONOPOLISTS

In section 2.4 we quoted the definition of the hypothetical monopolist test from the 1992 US Horizontal Merger Guidelines. Below are the definitions from the UK competition agencies' Merger Guidelines and the European Commission Notice on market definition. Can you spot the difference?

> In applying the hypothetical monopolist test, the Authorities will assess whether the hypothetical monopolist could profitably raise the price of at least one of the products in the candidate market by at least a small but significant amount over a non-transitory period of time (ie by a 'SSNIP'—a small but significant and non-transitory increase in price).[30]

> The question to be answered is whether the parties' customers would switch to readily available substitutes or to suppliers located elsewhere in response to a hypothetical small (in the range 5%–10%) but permanent relative price increase in the products and areas being considered. If substitution were enough to make the price increase unprofitable because of the resulting loss of sales, additional substitutes and areas are included in the relevant market. This would be done until the set of products and geographical areas is such that small, permanent increases in relative prices would be profitable.[31]

Rather than ask what a monopolist 'would' do to maximize profits—as in the US version—these other versions ask what a monopolist 'could' do profitably. The US version is the theoretically purer one—monopolists maximize profits. The other version is about breaking even—at what point does the monopolist make the same profits as before?

Having two different versions of the hypothetical monopolist test has caused much confusion among competition practitioners (lawyers and economists) in the last 10–20 years. Some may not even have been aware that there was a difference—the EU and UK guidance documents do not mention that their test is different from that in the USA, and even the US agencies switched versions between the 1982 and 1992 guidelines without stating this explicitly.[32] But does it really matter?

Theory tells us that it is not always the case that if a certain price increase is profitable, the hypothetical monopolist would actually impose it. A profit-maximizing monopolist would impose a smaller increase if that led to even higher profits. In fact it can be

[30] Competition Commission and Office of Fair Trading (2010), 'Merger Assessment Guidelines', September, [5.2.11].

[31] European Commission (1997), 'Notice on the Definition of Relevant Market for the Purposes of Community Competition Law', 97/C372/03, [17].

[32] The 1992 Guidelines in fact use both versions in different places. The difference between the two versions of the test is made explicit on p.12 of the 2010 Horizontal Merger Guidelines.

shown that the break-even price increase is twice as high as the profit-maximizing price increase (assuming linear demand and constant marginal cost, as we have in our examples). If you go back to Figure 2.4, the monopolist would go to the profit-maximizing price, but could raise price even higher and still make a profit. As a consequence, markets defined using the break-even approach tend to be narrower than those defined using the profit-maximization approach.

In practice, however, the break-even approach, when using critical loss analysis (discussed below), has some advantages, even if it is less theoretically pure. Identifying the price that maximizes the monopolist's profits is less straightforward than identifying the price at which the hypothetical monopolist makes the same profit as currently (ie, the profits made by the suppliers in the market before you hypothetically monopolized it). We note here that economists have also developed a practical means of applying the SSNIP question to the profit-maximizing hypothetical monopolist—this is through the critical elasticity.[33] Recall Figure 2.4. The lower (further to the right) you are on the demand curve, the more inelastic demand is, and the further the monopolist would increase price. There is a 'critical' elasticity level where the resulting price increase is exactly 5% (or 10%). You can apply the hypothetical monopolist test by estimating the actual elasticity and comparing this to the critical elasticity. However, the critical level depends crucially on the shape of the demand curve, ie, whether this is linear or of some other shape. A great advantage of standard critical loss analysis is that it does not depend at all on the shape of the demand curve, and you therefore do not need to make any assumptions about demand. Your results are more likely to be robust.

2.5.2 THE CONCEPT OF CRITICAL LOSS

If you raise the price of a product, two things happen. You sell less of the product, as customers walk away or reduce their purchases. But on the products that you do sell, you make a higher profit margin than before. These two effects therefore work in opposite directions. The first decreases your profits; the second increases them. Any business in the real world faces this trade-off when considering a price increase. And so does a hypothetical monopolist. How do you, and the hypothetical monopolist, know which of the two effects predominates? At what point do the two effects exactly cancel each other out, and leave you with the same amount of profits as before? This is what standard critical loss analysis is about—the critical sales loss is the percentage of sales at which the hypothetical monopolist makes the same profit before and after imposing a SSNIP. If the actual sales loss following the SSNIP exceeds the critical loss threshold, the SSNIP is unprofitable (conclusion: the market is wider). If the actual loss is below the critical loss, the SSNIP is profitable (conclusion: this is a relevant market).

You can already see from the description of the two effects which factors influence this critical loss threshold. The first effect is customers walking away or reducing purchases

following the 5% or 10% price increase. So it matters whether you select 5% or 10% as the SSNIP. As explained in section 2.4, either is valid in principle. With a 5% SSNIP you expect there to be a lower actual sales loss than with 10%, but the critical loss threshold will also be lower (see the discussion around Table 2.3 below). The reason why lost sales matter is that the monopolist made a profit margin on them before, and those profits are now lost. So the initial profit margin (as before, defined as price minus marginal cost, divided by price) also matters in determining the critical loss threshold. The second effect of the price increase is the higher margin on remaining sales. The higher margin equals the initial margin plus 5% or 10% (as that is by how much the price has increased; this assumes that marginal costs do not change over this range of output, a simplifying but often not unreasonable assumption). So, in sum, the critical loss threshold depends on only two factors: the percentage SSNIP on which you do the analysis, and the initial profit margin.

This is as far as we can take our verbal explanation of critical loss analysis. Some basic algebra is required to explain the formula for working out the critical loss. A chart also helps. Consider the following. There is a starting situation in which q units of the product are sold at price p. This gives rise to a price–cost margin, m, equal to price minus marginal costs, divided by price—ie, $m = (p-c)/p$. Then there is a SSNIP of x, which is equal to the change in p divided by p ($x = \Delta p/p$)—note therefore that x can be any percentage, even though it is usually 5% or 10%. Suppose that this SSNIP results in a decline in the quantity sold of Δq (Δq is a positive number if there is a sales loss). The two effects of the SSNIP on profits can then be expressed as follows. The fall in quantity, Δq, gives rise to a loss of profits since on each of the units previously sold the profit was m times p—ie, the loss of profit is Δq times m times p. The profit margin on the remaining sales is enhanced by x (the SSNIP)—this extra profit is $(q-\Delta q)$ times x times p. The critical loss question is: what is the maximum decline the quantity may suffer before the price increase becomes unprofitable? The answer: it is the Δq for which the profit gains from the price increase are equal to the loss of profits from the decline in quantity. In other words, it is where the hypothetical monopolist breaks even. A bit of algebra leads us from the break-even equation $(q-\Delta q)$ times p times $x = \Delta q$ times m times p, to the standard expression for the critical loss percentage: $\Delta q/q = x/(x+m)$.[34] Thus, in line with the verbal explanation earlier, the critical loss depends only on the SSNIP percentage used, x, and on the initial price–cost margin, m.

This situation is illustrated in Figure 2.5. F is the starting point, with a price p and a quantity q. Total profits for the monopolist are equal to the area ACFD (points A and C are on the marginal cost curve and a profit margin of p minus A is earned on every unit of q). Then price is raised from p to $(p+\Delta p)$ and output reduces to $(q-\Delta q)$. As above, it is assumed that marginal costs, c, do not change between these two quantities. The profits lost by the quantity reduction correspond to the shaded area BCFE. But there is also a profit gained by the price increase, as represented by the shaded area DEHG. Critical loss analysis is about comparing these two effects. If BCFE is larger in surface than DEHG, the price increase is unprofitable. If DEHG is larger, the increase is profitable. You can see that this depends crucially on the level of Δq. Given the values of q, p and Δp, the larger Δq is, the greater area BCFE

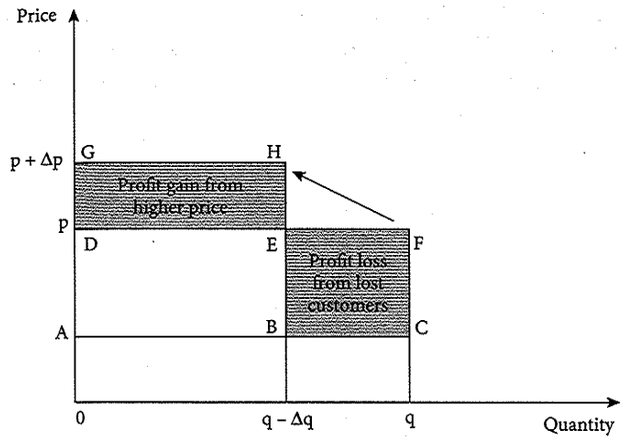

Fig. 2.5 The profit effects of a price increase

becomes, and the more area DEHG gets squeezed. The Δq for which areas DEHG and BCFE have exactly the same surface is the critical sales loss—this is where the monopolist breaks even, ie, would make the same profit before and after the price increase.

You may have noted that Figure 2.5 does not have a demand curve. It just has two price–quantity points, F and H. This is because determining the critical sales loss—and the formula $x/(x+m)$—does not depend on the shape or nature of the actual demand curve at all (other than that it is downward-sloping). It simply follows from an arithmetic comparison between profits before and after a price increase. This is one of the practical advantages of critical loss analysis. The profit-maximization version of the hypothetical monopolist test, while theoretically purer, relies crucially on assumptions that you would have to make regarding the shape of the demand curve in order the find the profit-maximizing price. All you need to know for critical loss analysis is the initial price–cost margin (see Chapter 10 for some practical tips on how to derive price–cost margins from financial information).

Table 2.3 shows the critical loss percentages for 5% and 10% price increases and for different values for the initial price–cost margin. Various points are worth noting in the table. When the initial margin is zero (price equals marginal cost, so no profits are made), the 5% or 10% price increase would be profitable for the monopolist regardless of the sales loss, as any remaining sales now generate a positive profit—hence the critical sales loss is 100%, at which point profit is zero again. The general point is that if you make a higher margin, every lost sale hurts you more. Furthermore, you can see that the critical sales percentage for a 10% SSNIP is higher than for a 5% SSNIP, although not quite proportionately so—the choice between 5% and 10% matters somewhat here (with 5% you tend to find slightly narrower markets than with 10%), but as noted before

Table 2.3 Critical loss values based on the formula x/(x+m)

Initial price–cost margin (m) (%)	Critical loss for 5% SSNIP—0.05/(0.05+m) (%)	Critical loss for 10% SSNIP—0.1/(0.1+m) (%)
0	100	100
10	33.3	50.0
20	20.0	33.3
30	14.3	25.0
40	11.1	20.0
50	9.1	16.7
60	7.7	14.3
70	6.7	12.5
80	5.9	11.1
90	5.3	10.0
100	4.8	9.1

we would suggest that your ultimate conclusions on market definition should not hinge upon the exact choice of the SS in SSNIP.

Another point is that many of the critical loss percentages in the table imply that a lot of customers do not have to switch—for a 5% SSNIP, the critical loss is less than 20% for any margins above 20%, which means that you may have a result where more than 80% of customers would continue to buy the product. Critical loss analysis is about what customers do at the margin, not what the average or typical customer does. It is still not a market worth monopolizing if switching by the other 20% makes the price increase unprofitable. In this case, those (marginal) customers with choice protect the other (typical) customers who do not have or do not exercise a choice. We also touch upon this general principle, and the difference between captive and non-captive customers, in section 2.8 on price discrimination markets. Moreover, while from one perspective the percentages in the table may seem low, from another they may seem high. They almost all exceed the SSNIP percentage itself (except where the margin is very high). Going back to the elasticities in Figure 2.3, what this implies is that critical loss analysis is not really about testing whether the elasticity is higher or lower than –1. In other words, –1 is usually not a critical level of elasticity—instead, as shown in the table, for an initial margin of, say, 40%, the break-even critical elasticity is –2.2 for a 5% SSNIP (11.1% sales loss divided by 5%) and –2 for 10% (20% divided by 10%).

2.5.3 HOLIDAY PARKS AND HOSPITALS: APPLICATIONS OF CRITICAL LOSS ANALYSIS

The 2001 merger between Gran Dorado (owned by Pierre & Vacances) and Center Parcs, reviewed by the Dutch competition authority (Nederlandse Mededingingsautoriteit, NMa), illustrates how you can get stuck focusing on product characteristics and how this can be overcome through critical loss analysis.[35] Gran Dorado and Center Parcs both operated self-catering accommodation in holiday parks with a range of facilities (indoor swimming pools, restaurants, playgrounds, etc) in the Netherlands, Belgium, and Germany. These parks are typically used by families for short breaks. The NMa approved the merger, but only after the parties had agreed to sell off a substantial number of parks.

Center Parcs and Gran Dorado were the two largest, and probably best-known, providers of holiday parks in the Netherlands and surrounding regions popular with Dutch short-breakers. Other, smaller players with similar offerings included Landal GreenParks, Zilverberk Parken, and Euroase Parcs. The NMa assessed whether there is a separate market for self-catering accommodation in holiday parks with a range of facilities, or whether other types of accommodation for short breaks are considered good substitutes by consumers. The NMa mainly emphasized the differences in product characteristics. It concluded that other types of holiday accommodation, such as family hotels, hotels for city breaks, hotels close to theme parks, and luxurious camping sites, should not be included in the relevant market because of the significant differences. For example, according to the NMa, parks with bungalows (as the holiday homes in the parks are known in the Netherlands) are mainly visited by larger groups, especially families with children, and also offer a range of facilities for entertainment; in contrast, a hotel primarily offers a place to sleep. It also noted that camping sites would be a substitute for holiday parks only in specific periods of the year and that there could be differences in comfort. Some of the holiday parks, particularly the larger ones, distinguish themselves from the smaller, more basic parks by offering a certain minimum set of facilities on site, such as a 'tropical' swimming pool and various restaurants. Furthermore, the NMa concluded that there are significant differences between the quality and price of the various types of holiday park. Based on these considerations, it provisionally narrowed the market down to those holiday parks that could be considered as four-season holiday villages, which included those operated by Center Parcs and Gran Dorado.

The differences in product characteristics that the NMa emphasized are no doubt of importance to holidaymakers. However, there are several product features that might make the holiday parks more interchangeable to others. For example, customer research carried out by the parties for their own commercial purposes (ie, not specifically for the merger inquiry) showed that for many visitors to the Center Parcs and Gran Dorado parks, walking and cycling through the surroundings outside the parks was one of the main activities during their short break (besides using the pool and other facilities inside the park). Indeed, these holiday parks are typically located in an attractive national park or coastal area, rather than, say, in the middle of an industrial estate. Seen from this perspective, other types of accommodation in such areas—including hotels, holiday homes, and holiday parks with fewer facilities—compete in the same market.

[35] NMa (2001), *Gran Dorado–Center Parcs*, Zaak 2209. See also Niels and van Dijk (2006). We advised the

Moreover, many alternative providers of short-break accommodation also offer access to facilities such as swimming pools and restaurants nearby, even if these facilities are not on the same site and are operated by third parties. Nor is the self-catering aspect unique to holiday parks. Many hotels now also offer self-catering accommodation (through so-called 'apart-hotels'), and two of Gran Dorado's bigger parks also offered hotel accommodation. An even broader perspective would be to include all accommodation for short breaks in the same market. Indeed, the parties regarded Eurodisney as a major competitor. Even if its product offering is quite different—hotel accommodation near a theme park—Eurodisney competes directly to attract families with children for their short breaks, which is inconsistent with the NMa's definition of the market. Indeed, the short-break brochures in which Grand Dorado and Center Parcs advertised also typically included Eurodisney and other options for short breaks. Finally, as to price differences, these were often more marked between holiday parks of a different quality than between a park and, say, hotel, of similar quality—ie, a short break in a top-of-the-range holiday park and a top-of-the-range hotel with similar facilities nearby could have a similar price.

Hence, focusing solely on product characteristics potentially leads to a discussion without end. (Does an indoor pool become a 'tropical' pool when it has a slide and some palm trees?) Yet an even greater shortcoming of this approach is that it fails to address the crucial question for market definition—namely whether a sufficient proportion of consumers consider these products to be substitutes. As noted earlier, products do not need to be perfect substitutes to be in competition with each other, since a small overlap of consumers who are willing to switch can be sufficient. With this in mind, the merging parties undertook a consumer survey to identify empirically the level of sales that a hypothetical monopolist would lose as a result of a SSNIP. This type of survey is often the easiest (and fastest) way to obtain some relevant data on switching behaviour, even though it has shortcomings (see section 2.14), and the results should always be interpreted with care.

The survey was held among close to 250 short-breakers in the Netherlands, ie, those who had been on any short break in the past three years. In order to focus the minds of the respondents (and hence increase the chance of relevant answers), they were first asked about their current behaviour and preferences in relation to short breaks. Then, before turning to the SSNIP question, respondents were asked which type of short break they were considering for the next two years. Those who (certainly or probably) considered visiting a holiday park with facilities and indoor pool (ie, of the Center Parcs or Gran Dorado type) were asked whether they would still do so after a 10% price increase lasting for two years or whether they would switch to an alternative type of short break, or not go on a short break at all (as noted earlier, the latter also represents a relevant sales loss to the monopolist). To help respondents interpret the question, they were not only informed of the 10% increase but were also given a numerical example of what this, on average, would do to the price they would pay.

The results of the survey indicated that the relevant market is broader than only holiday parks with a range of facilities and an indoor swimming pool. This is because 28%

of those who considered a short break in such a park said they would switch to another type of short break, or take no short break at all, after a 10% price increase. This sales loss was clearly higher than the estimated critical loss. Recall that the critical loss depends on the price–cost margin. A rough approximation indicated that the holiday parks had a margin in the range of 60%–80% (this is because a large part of their costs are fixed, ie, they do not vary with sales). Applying the formula, this generates a critical loss of 11%–14% (10% divided by 10% plus 60% equals 14%, and 10% divided by 10% plus 80% equals 11%). You can see that in cases like these there is no need to identify the price–cost margin with great precision—the broad margin range of 60%–80% generates a narrower critical loss range of 11%–14%, and the actual loss in this case is sufficiently in excess of the critical level for you to be confident that you are making the right comparison (the only uncertainty would be about the reliability of the survey response, not about the application of the critical loss analysis itself).

The survey also revealed that there was no clear closest substitute. Respondents who said that they would switch were asked what options they would switch to. They mentioned more than two alternatives on average (this in itself indicates that they see multiple substitutes). Stand-alone holiday homes and apartments had the highest diversion percentage, with 37% and 34%, respectively. Other types of holiday parks—without indoor pool or a pool nearby—were mentioned by 21%–26% of respondents. The different types of hotel—for city breaks, family hotels, theme park hotels, and others—all received scores between 18% and 25%. Given that the critical loss analysis indicated that the market should be defined more broadly than just holiday parks with facilities and indoor pools—the focal product—it was not really necessary to proceed to the second iteration of the hypothetical monopolist test including the closest substitute. In the end, the precise market definition was left open, as the NMa approved the merger on the condition that the parties divest a number of their parks.

Critical loss analysis was also used in a merger between two hospitals in the Netherlands in 2005.[36] The NMa commissioned a study on the demand elasticity of patients in order to determine the relevant geographic market in which these hospitals operated. The study estimated a 'time elasticity' of demand—ie, the sensitivity of patients with respect to travel time—but then translated this into a price elasticity of demand so as to compare the results with the critical loss formula (more on this in section 2.13). For one product market—clinical treatments—the study found an actual sales loss of just under 10% after a price increase of 10%, with a confidence interval of +/– 2.8% around this estimate (reflecting statistical uncertainty). Based on a margin estimate of around 70%—reflecting the fact that hospitals have relatively high fixed costs and low marginal costs—the study calculated the break-even critical loss to be around 12.5% (in line with Table 2.3). It concluded from this that the SSNIP was profitable and hence that the geographic market was not wider than the area immediately surrounding the

[36] NMa (2005), *Ziekenhuis Hilversum-Ziekenhuis Gooi-Noord*, Zaak 3897, 8 June. We advised the merging

two hospitals. However, had the profit-maximization approach been used, a margin of around 70% would imply a critical elasticity around −1.1 (critical loss of 11%). This would be well within the confidence interval around the actual loss, and hence cast some doubt on the conclusion that the SSNIP was profitable. In the end, based on various other factors including this one, the NMa concluded that the hospitals did compete in a wider geographic market and approved the merger. The main lesson from this example is that care should be taken when the actual loss estimates are very close to the critical loss percentage. In these situations it may be useful to cross-check the break-even result against the profit-maximization result.

2.5.4 CRITICAL LOSS IN MARKETS WITH DIFFERENTIATED PRODUCTS

If there is a merger between producers of the same product (as in our example of the Italian women's designer shoes), that product is the focal product for your hypothetical monopolist test. If the merger is between an Italian and a French producer, and there is some concern that these two products compete with each other, you take them both as the focal product and carry out separate tests for each. If both produce men's designer shoes as well, you would carry out four separate hypothetical monopolist tests. You can see that in markets with differentiated products—such as cars and breakfast cereals—you may end up having to perform a very large number of tests, one for each variety of the product. Earlier we explained two paths you could take to avoid this. First, you could to some extent ignore differentiation and group together products that may still be considered reasonably homogeneous (top-of-the-range sports cars, holiday parks with self-catering accommodation). In line with the smallest-market principle, you could do this conservatively and take relatively narrow groups to start with. The hypothetical monopolist test would then be applied to each of the groups as if they consisted of homogeneous goods. Second, you could skip the market definition stage altogether and focus directly on the competitive constraints and closeness of competition between the differentiated products (as discussed in Chapter 7).

Economists have recently developed some tools that provide a third alternative.[37] You can still define the relevant market through the hypothetical monopolist by applying critical loss analysis to a group of differentiated products. In essence this is the same as applying the SSNIP question in the second and subsequent iterations of the hypothetical monopolist test, but here you can potentially take any product in the group as the focal product. You effectively take a shortcut by starting with a monopolist who already controls all the products where competition concerns may arise. You can then apply the break-even SSNIP question to any one of the products in the group, or to all the products. The focus in this new critical loss test for differentiated products is on the diversion ratio between these products—the critical loss formula depends not just on

[37] See Farrell and Shapiro (2008), Deliard, Sagaard, and Thomassen (2008), and ten Kate and Niels (2010).

the SSNIP, x, and the initial margin, m, but also on the percentage of total sales loss after the SSNIP that is diverted to other products within the group as opposed to outside the group. This percentage can be called the intra-group diversion ratio (some economists call it the aggregate diversion ratio, but we believe that this term has more than one meaning). The higher the intra-group diversion ratio, the less the monopolist cares about losing sales of one of its products (as more of these lost sales are captured by the other products), and hence the more profitable the SSNIP.

We can show this in a formula. Recall that the critical loss formula for the standard case (homogeneous products) was x divided by x+m. With differentiated products the critical loss can be expressed as x divided by x+m(1−d), where d is the intra-group diversion ratio. The factor (1−d) has crept into the formula as the percentage of sales that is lost to outside products. To see how it makes a difference, take the critical loss for a 10% SSNIP and 40% margin from Table 2.3—equal to 20% in the standard formula. In the new formula, if d equals zero (ie, 100% of sales is lost to outside products), the formula becomes the same as before, and the critical loss is still 20%. But with a positive intra-group diversion ratio this changes. Say d equals 0.2 (20%). The new formula then gives a critical loss of 24%—this is higher than the 20% you had before, since it takes more sales loss before the monopolist starts to care. If d equals 0.5 the critical loss becomes 33%, and if d is 0.8 the critical loss is 56%. Thus, the more the products under the control of the hypothetical monopolist are each other's closest substitutes, the more the monopolist can profitably increase price. There is still some debate among economists and practitioners about whether in markets with differentiated products it is more informative to use these tools and still define relevant markets, or to skip the market definition stage altogether and use the tools for unilateral effects analysis, which are very similar, as discussed in Chapter 7.

2.6 THE CELLOPHANE FALLACY

2.6.1 BACK TO THE 1950S

The 'cellophane fallacy' is a widely known concept in competition law. Less well-known are the finer details of the original case, which is more than 50 years old. In 1956, the US Supreme Court ruled on a monopolization case brought by the government against Du Pont.[38] Du Pont had a 75% share of US sales of cellophane (the producer of the other 25% had a licence from Du Pont and paid it royalties), but cellophane constituted only 20% of all flexible packaging materials. The decision therefore turned on the definition of the relevant product market. The Supreme Court, unlike the government, considered that there was sufficient evidence of interchangeability between cellophane and other materials (such as paper, film, and foils). There were similarities in functionality, and large user industries such as meat production switched regularly between different

[38] *US v EI du Pont de Nemours & Co* 351 US 377 (1956).

types of packaging material. While cellophane was two to three times as expensive as its nearest competitors (glassine and grease-proof papers), the court found that further price increases were prevented by those other materials. Can you see the fallacy in this reasoning?

Cellophane production was already a (near-)monopoly at the time. Du Pont possessed the technology and competition from the other cellophane producer was restricted through an overt agreement and apparent tacit coordination between the two. If the prevailing price was already the monopoly price, a hypothetical cellophane monopolist (Du Pont nearly was one for real) would not raise prices any further, since at the monopoly price other competing products start to bite. But that does not mean that cellophane is not a market worth monopolizing—it could still be a relevant market of its own. Many commentators have pointed this out since. Indeed, three dissenting Supreme Court judges said it in the judgment itself:

> We cannot believe that buyers, practical businessmen, would have bought cellophane if close substitutes were available at from one-seventh to one-half cellophane's price. That they did so is testimony to cellophane's distinctiveness.[39]

The cellophane fallacy can be explained through Figures 2.3 and 2.4. We explained how own-price elasticities depend on where on the demand curve you are to start with. The hypothetical monopolist test assesses the price at which the monopolist maximizes profits, and asks whether this price is at least 5%–10% higher than the initial (pre-monopolization) price. You can see that if the price is already at (or very near to) the monopoly price, the monopolist would not impose a further SSNIP. The logic of the test would then dictate that the market is broader than the product in question, but this could lead to an erroneous conclusion that this product is not worth monopolizing.

2.6.2 IS IT A PROBLEM, AND IF SO, WHAT CAN BE DONE ABOUT IT?

The cellophane fallacy is not normally a concern in merger cases. Competition authorities review whether mergers reduce competition compared with the present situation. It is therefore valid to take the prevailing price level as the starting point for the hypothetical monopolist test where the market is reasonably competitive before a merger. The exception is where markets, pre-merger, are already characterized by a form of (tacit) price coordination between competitors, such that hypothetical monopolization would not lead to significant further price increases, hence giving rise to the cellophane fallacy again. The fallacy is most likely to arise in cases where there is already a dominant company, or more generally where it is suspected that the party or parties under investigation have already exercised some degree of market power. In any such case, competition authorities or claimants can, and often do, invoke the cellophane fallacy to refute attempts by the defendants to demonstrate a wider market based on a SSNIP question applied to prevailing prices. Are they right?

[39] Ibid. at [417]

In theory, the cellophane fallacy arises only when prevailing prices are at, or very close to, the monopoly level. If prices are more than 5%–10% below the monopoly level, you would find a hypothetical monopolist imposing a SSNIP—your analysis would not suffer from the cellophane fallacy. The question is: how do you know where prevailing prices lie relative to the monopoly price level? We suggest that the answer to this question cannot be inferred just from the existing market structure (except where you have an unregulated real profit-maximizing monopolist). You cannot say there is a cellophane fallacy just because the market structure does not look perfectly competitive. Prices in oligopolistic markets may lie anywhere between the competitive and monopoly levels. Even companies with very high market shares may still be constrained by competitors (even if small) from raising price all the way up to the monopoly level. In these cases, asking the SSNIP question based on prevailing prices can still provide useful insight into substitution and competitive constraints. You just have to bear in mind that some degree of market power (short of monopoly power) may already have been exercised at those prices. At the very least you can apply a one-way test. If, at the prevailing price level, a monopolist can profitably impose a SSNIP, you can be confident that the market is no wider. If, in contrast, you find that the monopolist cannot profitably increase price, you have to accept that there may be a cellophane fallacy and that the market may be defined too broadly. Ultimately, if there is a high probability that your analysis will suffer from a cellophane fallacy, you may have to conclude that market definition is not a useful intermediate step and instead focus directly on indicators of competitive constraint (as further discussed in Chapter 3).

A frequently cited solution to the cellophane fallacy is to start the hypothetical monopolist test from the competitive price level rather than the prevailing price level. Because the test has its foundations in microeconomic theory, it is tempting to take marginal cost as the competitive price level (since in perfect competition, prices equal marginal cost). However, as we discuss in greater detail in Chapter 3, there are many real-world markets in which prices are set above marginal costs, but which can still be considered competitive—prices above marginal costs may be required to cover fixed costs, or may reflect only a temporary position of pricing power that is eroded over time. Therefore, to get a better approximation of the competitive price level you would need to carry out an analysis of profitability over a longer time period. The problem is, however, that such an analysis makes the market definition stage rather redundant (Niels, 2004). After all, profitability sheds light directly on whether companies can persistently earn high returns without attracting entry or inducing consumer switching. This is precisely how competition law defines market power and dominance—profitability analysis is a direct indicator of dominance, which means that market definition as an intermediate step is no longer required. Indeed, the US Supreme Court itself cited profitability in the 1956 cellophane case as a reason to reject the monopolization claim and, implicitly, reject the presence of a cellophane fallacy (arguably the court referred to less sophisticated techniques of measuring profits than have been developed since—see Chapter 3):

> Nor can we say that du Pont's profits, while liberal (according to the Government 15.9% net after taxes on the 1937–1947 average), demonstrate the existence of a monopoly without

Cellophane was a leader over 17% in the flexible packaging materials market. There is no showing that du Pont's rate of return was greater or less than that of other producers of flexible packaging materials.[40]

2.6.3 THE 'REVERSE' CELLOPHANE FALLACY

Just as the cellophane fallacy leads to overly broad markets, economists have identified a phenomenon called the 'reverse' cellophane fallacy, which results in overly narrow markets. This occurs where the prevailing price level is too low to start with. Recall from Figure 2.4 and Table 2.3 that if price is very low (say, at or close to marginal cost), almost any price increase is profitable, and hence the hypothetical monopolist test is easily passed. When may prevailing prices be too low? We would suggest that situations like this are more exception than norm, but theoretical examples are where a predatory price war has been raging in the market in question, or where a product has been subject to a regulatory price cap that is below the point where a reasonable profit is made.

2.7 SUPPLY-SIDE SUBSTITUTION AND MARKET AGGREGATION

2.7.1 WHAT IF THE HYPOTHETICAL MONOPOLIST WERE NOT THE ONLY FUTURE SUPPLIER?

In section 2.3 we explained the distinction between demand-side substitution, supply-side substitution, and new entry. Thus far in this chapter we have dealt with demand-side substitution, ie, switching behaviour by customers. The only supply-side aspect of market definition that has come up at this stage is our hypothetical monopolist, ie, the supplier who imposes the price changes to which customers react. Supply-side substitution is about other suppliers' reactions to the monopolist's price changes. It is usually a less immediate constraint on companies than demand-side substitution (we said before that customers are generally quicker to vote with their feet if they are unhappy than wait for other suppliers to come to them). Nonetheless, it may also render price increases by the monopolist unprofitable.

In the formal definition of the hypothetical monopolist test in the 1992 US Horizontal Merger Guidelines—cited in section 2.4—supply-side substitution is expressly ruled out because the monopolist is the 'only present and future producer or seller of those products in that area'. This version of the test considers only demand-side substitution. However, other jurisdictions tend to also consider supply-side substitution as part of the market definition stage under certain conditions, discussed below. Ultimately, it should not matter too much during which stage of the analysis you consider supply-side substitution since all competitive constraints must be considered (recall that market definition is only

[40] Ibid, at [404]

an intermediate stage). In the US approach, those suppliers that are not in the market currently, but that could easily supply-side substitute into it, are taken into account at the point of calculating market shares in that market. This should give the same result as calculating shares in a wider market based on supply-side substitution.

Nevertheless, there are some traps you might inadvertently walk into when relying on supply-side substitution at the market definition stage. We discuss these below. First, we set out how the line is usually drawn between supply-side substitution and new entry (the latter does not form part of the market definition stage; more on this in Chapter 3).

2.7.2 PAPER AND BUSES: ILLUSTRATING THE CRITERIA FOR SUPPLY-SIDE SUBSTITUTION

Paper production is a frequently cited example of supply-side substitution. Paper is usually supplied in a range of qualities, from standard writing paper (uncoated) to high-quality paper (often coated) to be used, for instance, in art books. From a demand point of view, different qualities of paper are not substitutes—a pricey art book cannot be printed on low-quality paper. However, paper plants may be able to manufacture different qualities of paper, switching their existing production facilities at little cost within a short timeframe, if the commercial opportunity arises—for example, if the price of a paper variety they are currently not producing increases by 5%–10%. In such circumstances, the various qualities of paper may be included in the relevant market based on supply-side considerations. The European Commission uses the paper example in its 1997 Notice on the Definition of Relevant Market, and first dealt with the issue in a 1992 merger decision:

> Demand-side substitutability
>
> 17. There is limited substitutability between uncoated and coated papers because of their different characteristics (such as discolouring and printing quality). In addition uncoated paper is much cheaper than coated paper. The above table shows that the price of the cheapest coated paper is even higher than the price of the most expensive uncoated paper and that coated paper is on average 15% more expensive than uncoated paper.
>
> Supply-side substitutability
>
> 18. From the supply side, there is a rather high substitutability. Indeed, since the difference between coated and uncoated paper results from extra processing, the coating processing can be included whenever required. Since the different grades of paper result mainly from the blend used, the coating materials used and some other extra processing, it is relatively easy for a producer to switch from production of one paper type to another. This is particularly true with regard to the older non-integrated machines, which can be used for various kinds of paper production. The high speed, high capacity and highly integrated new mills are also able to switch from production of one paper type to another, although to a lesser extent than older non-integrated machines.[41]

[41] Case No IV/M.166, TORRAS/SARRIO (1992), OJ C58/30, [1992] 4 CMLR 241

Bus services are another common example of supply-side substitution, in the context of geographic market definition (paper is an example of product market definition). From a demand-side perspective, point-to-point bus routes would form separate markets. A hypothetical monopolist on the route from your home to your office would be able to increase fares (abstracting from competition from other modes of travel, such as cars, trains, taxis, or walking), because you would not travel somewhere else after a price rise. However, from a supply perspective, bus operators can shift or expand into new routes with relative ease, provided that they have an existing bus depot nearby (where buses can park overnight and receive maintenance). This means that geographic markets for bus services are often defined more widely, covering the whole area that can economically be served from an existing depot.

From these two examples you can see what kind of criteria matter for supply-side substitution. Supply-side substitution requires there to be no significant additional 'sunk' investments or costs of switching. It must be sufficiently swift and it must be of a sufficient scale to constrain the hypothetical monopolist. The first of these criteria refers to a commonly used concept in economics—that of sunk investments. Such investments are different from other investments in that they cannot be recovered upon exit from the market. A typical example is an investment in a new brand by a market entrant, which will lose its value once the company exits the market again (and the brand cannot be sold). In contrast, investments that are not sunk still have value at the point of exit. This matters for entry decisions—if you know you can get rid of your assets for a decent price in the event of your new business venture failing, you are more likely to go ahead than in the situation where you have to scrap the assets. The second criterion relates to timeliness of the supply switch. As with the 'N' in SSNIP there are no hard and fast rules for what is timely, but often the reference is to a one-year period, or simply to existing production facilities being used (existing paper plants or bus depots). Use of existing facilities implies swift entry, and also narrows the first criterion somewhat from no sunk investment to no major investment in new facilities (sunk or not). The 1992 US Horizontal Merger Guidelines call supply-side substitution 'uncommitted entry', which is contrasted with 'committed entry' (the 2010 Guidelines, at p 16, simply refer to 'rapid entrants'). In economic theory, industries without sunk costs and where entry is rapid are sometimes called 'contestable markets'—the possibility of 'hit and run' entry would prevent even a monopolist from raising prices in such markets (and hence, according to the hypothetical monopolist test, they may not be relevant markets in themselves in the first place).

To continue with the bus example, the provision of local bus services has some features of a contestable market. If the operator has a bus depot nearby, adding services to an existing route or opening new routes can be done swiftly (in the UK a new route can be opened 56 days after registration). Any investment in additional buses is not sunk, since there is often a reasonably liquid market for second-hand buses should the operator decide to exit the route (buses can also be leased or rented). With regard to the third criterion mentioned above—scale—in the provision of bus services even entry on a small scale can have a strong competitive impact on a particular bus route—for example,

if the entrant runs its service only on the busiest part of a route or during the busiest times of the day. This conclusion was reached in the 2004 OFT decision in the *Arriva/Wales and Borders Rail* case, where a regional rail franchise was acquired by an operator which also provided local bus services in the area:

> Buses may be bought outright, or leased. Bus depots (and out-stations [secure areas for overnight parking]) may be leased … An existing operator with a local depot is unlikely to face significant barriers to expansion and it is likely that larger operators would take up any profitable routes abandoned (wholly or partially) by Arriva. Where a route is unprofitable (wholly or at certain times) local councils may invite tenders from operators to run the route on a subsidised basis, in which case there is competition for, rather than on, these routes.[42]

2.7.3 THE RISK OF OVERSTATING COMPETITIVE PRESSURE FROM SUPPLY-SIDE SUBSTITUTION

There is something inherently odd about supply-side substitution. Demand-side substitution is about whether consumers see products as substitutes. If they do to a sufficient degree, you put the two products together in the relevant market. Supply-side substitution is not so much about the products as the suppliers of those products. The high-quality paper monopolist is constrained by the producers of low-quality paper because they can also produce high-quality paper, not by the low-quality paper itself. This oddity is one of the reasons why the US Guidelines prefer to focus on demand-side substitution, and then treat suppliers that can supply-side substitute (ie, enter in an 'uncommitted' way) as participants in the market and include them in the market share calculation. As noted earlier, this should ultimately result in the same conclusion as in the approach where you include supply-side substitution in the market definition itself. But there is a risk that you will overstate the competitive pressure from supply-side substitution as a result of this 'oddity'. We highlight the main pitfalls here.

Consider the following example. You have six producers of high-quality paper. The two largest ones (A and B) have 25% of production each, the other four 12.5%, as shown in Table 2.4. Considering demand-side substitution only, you find that a hypothetical monopolist for high-quality paper would impose a SSNIP—customers would not switch to low-quality paper. If the two largest producers were to merge, they would have a combined market share of 50%, sufficient to raise some competition concerns. But you haven't considered supply-side substitution yet. Suppose low-quality paper is produced by completely different companies, but you find that their existing plants and equipment can very easily be switched to produce high-quality paper. So on a supply-side basis the relevant market for the merger should include low-quality paper as well. As you can see from the table, the total market is now two-and-a-half times as large, and the merging parties have a combined market share of only 20%—not enough to warrant

[42] Office of Fair Trading (2004), 'Completed Acquisition by Arriva Plc of the Wales and Borders Rail Fran-

Table 2.4 Stylized example of supply-side substitution

	High-quality paper production (tonnes)	Market share in high-quality paper	Market share in all paper	Market share in high-quality paper plus available low-quality capacity
Producer A	100	25%	10%	21.7%
Producer B	100	25%	10%	21.7%
Producer C	50	12.5%	5%	10.9%
Producer D	50	12.5%	5%	10.9%
Producer E	50	12.5%	5%	10.9%
Producer F	50	12.5%	5%	10.9%
Market size (tonnes)	400	400	1,000	460
Low-quality paper production (tonnes)	600			
Low-quality capacity available for substitution (tonnes)	60 (10% of all low-quality capacity)			

competition concerns. This may well be the correct conclusion. But there is a risk that you have overstated the importance of supply-side substitution as a competitive constraint, and hence understated the scope for the merged entity's market power. You have effectively assumed that all production of low-quality paper can be readily switched to high-quality paper (you have added all 600 tonnes to the relevant market). This may not reflect reality. There may be various reasons why some of the capacity cannot be readily switched—producers of low-quality paper may not want to upset relationships with their existing customers, they may have long-term supply contracts with them, or they may still consider the low-quality market more lucrative despite the price increase in high-quality paper. In essence, this comes down to the scale of the supply-side substitution, the third criterion mentioned above. Suppose producers of low-quality paper would switch only 10% of their capacity to high-quality paper. Then you should count only that available capacity when you calculate market shares. You would find that according to this calculation the merged entity has a 43.5% market share.

2.7.4 THE RISK OF GETTING SUPPLY-SIDE SUBSTITUTION IN THE WRONG DIRECTION

Another potential pitfall with using supply-side substitution is to get the direction of the substitution wrong. This happens when you include all products or geographic areas which suppliers can substitute into, regardless of whether this substitution is towards

or away from, the focal product or area. This leads to overly broad markets, since only supply-side substitution *towards* the focal product or area in question is relevant in capturing the competitive constraints on that product or area.

An example of where such an error was made is the abuse of dominance case before the English High Court involving bus services, as referred to previously. The dispute concerned alleged predation on several local routes in Chester in the north-west of England—this was the focal area. The experts for the claimants and defendants had agreed that the geographic market should be defined based on supply-side substitution, including all existing bus depots that could economically serve routes in Chester (a drive time of 30 minutes was agreed upon as the maximum distance between the depot and the route—see section 2.9 for a discussion on isochrones analysis). However, the claimants' expert subsequently expanded the geographic market by including all areas that could be served from the depots, including areas in the other direction from Chester. This led to an overly broad market (in this case, the defendant had a higher market share in the wider market because it had substantial operations in areas adjacent to Chester). The court agreed with the defendants' expert that this was incorrect:

> [The claimants' expert's] starting point was to identify which depots can economically provide bus services to Chester, whether or not any such depot in fact does so; and that part of the exercise is one with which in principle [the defendants' expert] agrees. He then defined the market as comprising the areas that those depots serve or can serve, and he concluded that for practical purposes that meant it comprises the eight local authority districts in which the depots are situated; and with that approach [the defendants' expert] disagreed. Thus, for example, Arriva's Winsford depot (to the east of Chester . . .) is said to be capable of economically operating into Chester, and the Crew and Nantwich administrative district lying to its south-east (and increasingly remote from Chester) into which services from Winsford are also possible is therefore also said to fall within the geographic market . . .
>
> The effect of [the claimants' expert's] approach (unlike [the defendants' expert's]) is thus to include in his market depots from which he accepts it would *not* be possible for any operator economically to provide local bus services in Chester . . .
>
> I accept [the defendants' expert's] opinion . . . that when identifying a geographic market by reference to supply-side considerations it is a mistake to include all geographic areas into which suppliers can substitute regardless of whether this is towards or away from the focal market; and that only supply-side substitution towards the focal market is relevant to capture competitive constraints on that market. The [defendants' expert's] opinion is that [the claimants' expert] has committed the error of overlooking this principle and so has erroneously promoted a broader geographic market than is justified . . . I accept his opinion.[43] [Emphasis in original].

[43] *Chester City Council and Chester City Transport Limited v Arriva PLC* [2007] EWHC 1373 (Ch),

2.7.5 MARKET AGGREGATION

Market aggregation is a phenomenon closely related to supply-side substitution. It is one of those pragmatic adjustments to the theoretically pure hypothetical monopolist test that make its application more practical. We started this chapter with a merger between producers of Italian women's designer shoes. We later discussed the question of whether different types of shoes (such as French women's designer shoes) are seen as substitutes and impose price pressure on Italian shoes. But there is another distinction that must be considered. From a purely demand-side perspective, size 38 shoes and size 39 shoes are not substitutable. Thus, on this basis you should define a separate relevant market for each shoe size. Clearly that would be somewhat absurd. Your conclusions on competitive constraints would be exactly the same for each of these size-based markets, since each producer makes and sells all the sizes (except where producers specialize in very large or small shoe sizes)—the competitive conditions are the same. So you might as well aggregate all shoe sizes into one market. The same logic applies to various transport and communications markets, where demand-side considerations would lead to many different point-to-point markets, but where competitive conditions are often similar for all of these markets, and hence for practical reasons you can aggregate them.

2.8 PRICE DISCRIMINATION MARKETS

2.8.1 COMPARING APPLES WITH BANANAS

In the 1978 *United Brands* judgment, the General Court (then Court of First Instance) had to address the question of whether bananas are a separate market or part of a wider market for fresh fruits.[44] The case concerned banana distribution arrangements in the north-west of Europe. There was some statistical evidence that banana demand and prices are under pressure in the summer months when domestic fruits are in ample supply, and during the 'orange season' at the end of the year. The court agreed with the European Commission that these periods of substitution were too limited—bananas are available the whole year round, so candidate substitute fruits would have to be too, it held. Oranges were regarded as too distinct, and apples were seen as interchangeable only to a small degree. The court (and the Commission) reasoned as follows:

> This small degree of substitutability is accounted for by the specific features of the banana and all the factors which influence consumer choice.
>
> The banana has certain characteristics, appearance, taste, softness, seedlessness, easy handling, a constant level of production which enable it to satisfy the constant needs of an important section of the population consisting of the very young, the old and the sick.[45]

[44] Case 27/76, *United Brands Company and United Brands Continentaal BV v Commission*, 14 February 1978.
[45] Ibid., at [30], [31].

This reasoning may remind you of the holiday park merger discussed in section 2.5, which also sought to define markets primarily based on product characteristics. You can debate at length whether babies, elderly people, and the sick really can't replace bananas with other fruits (our own recent experience would suggest otherwise, but then supermarkets nowadays offer a much greater variety of fruits year round than they did in the 1970s). But more importantly, the court's reasoning misses the point about critical loss. Even if one particular 'section of the population' cannot switch away from bananas, there are other sections that can. If too many of those other customers would switch to apples, oranges and the like after a banana price increase, a hypothetical banana monopolist would not impose such an increase. The very young, the old, and the sick would thus be protected by the less captive customers. Crucially, this is because banana suppliers and retailers are unable to specifically target their captive customers (for a start, babies, the elderly, and the sick may not do their own shopping). If the monopolist raises price, it has to do so across the board to all customers. It cannot price-discriminate, ie, charge a different price to different groups of customers for the same product.

As discussed in section 2.2, price discrimination can give rise to separate markets by (groups of) customers. This may occur when some customers have a greater choice of alternatives than others, and suppliers can exploit this difference by targeting the latter, 'captive', customers with higher prices, without this being undermined by arbitrage (where the non-captive customers can somehow resell the product to the captive ones, or purchase it on their behalf). Banana suppliers are unable to achieve such price discrimination. But other suppliers can. A common example is the travel industry. A flight from London Heathrow to Milan Linate at 7.45am will have a mix of time-sensitive and non-time-sensitive passengers on board. The former have no choice but to be on that particular flight—they have a meeting in Milan at lunch time. The latter passengers could have chosen a different time to fly, or perhaps even a different destination (eg, a city break in Madrid instead of Milan). Airlines are able to exploit this difference since time-sensitive passengers tend to book their tickets less far in advance of flying (airlines tend to raise fares closer to the flight date), and are more likely to demand flexible tickets in case their meeting times change (airlines target this by offering more expensive flexible economy and business tickets). Competition cases involving airlines therefore frequently define separate markets for time-sensitive and non-time-sensitive passengers.[46] Returning to the bananas case, there was perhaps some economic merit in the court's argument that competition from seasonal fruits is not quite sufficient to prevent a banana price rise—it is just that this banana price rise would occur outside those seasons. Rather than applying the criterion that substitute fruits for bananas must be available all year round, the court could have followed the price discrimination logic and argued that bananas still form a separate relevant market during those times of the year in which other fresh fruits are in less ample supply. It would then probably still have found United Brands to be dominant in those periods.

[46] The European Commission has made this distinction in several airline merger cases, including *Lufthansa/*

Where suppliers cannot specifically target the more captive customers, there are no separate price discrimination markets. If you live two minutes' walking distance from a major supermarket, you are probably a captive customer—you wouldn't drive ten minutes to the other major supermarket in town instead. Fortunately for you, the supermarket has no way of distinguishing between you and the next customer in the queue who lives exactly between the two supermarkets and hence has more choice. The supermarket cannot discriminate against you—you and your captive immediate neighbours are 'protected' by other customers who do have a choice. This protection of customers who don't have a choice by those who do is a feature of demand that arises in many markets (this is not to say that supermarkets don't try to price-discriminate; many have developed sophisticated discount schemes through loyalty cards, voucher systems, and the like).

Clearly, price discrimination is a very common business practice (in Chapter 4 we discuss the circumstances in which it may be anti-competitive). It would not make sense to define separate price discrimination markets wherever different prices are charged. You might discover that almost every passenger on your London–Milan flight has paid a different fare, but you would not want to define a separate market for each passenger. Again, some pragmatism is required. The most consistent approach is to consider instances of price discrimination where the captive group of customers is significant (a group of customers worth having a monopoly over, such as time-sensitive passengers), and where the price difference is significant and sustainable for a non-transitory period.

2.8.2 A COMMENT ON THE RELEVANCE OF ABSOLUTE PRICE DIFFERENCES

You have seen that the hypothetical monopolist test is about reactions to relative price changes between products. It is not about absolute price differences. Branded soft drinks may provide a competitive constraint on own-label soft drinks, and hence be included in the market, despite being more expensive. Customers make a price–quality trade-off, so if the price difference between the two products becomes too narrow they may switch from own-label to branded. However, just as differences in product characteristics sometimes provide a useful sense-check on market definition, so do differences in price. For example, if you can buy a business ticket for a flight from London Heathrow to Milan Linate Airport at 7.45am on a Monday for €734, while an economy ticket from London Stansted to Milan Bergamo Airport at 11.55am on a Wednesday costs €82, chances are that these two flight tickets do not provide competitive pressure on each other. You would expect them to be in separate relevant markets. How can you establish this? It may be that there is a break somewhere in the chain of substitution from cheap to expensive flights between London and Milan (see the next section on chains of substitution). Or it may be that you have to define separate markets along various dimensions, such as by geography (the question being which airports you should include around London and which around Milan), by type of customer (time-sensitive versus non-time-sensitive passengers), or by time of flight (peak versus off-peak).

2.9 CHAINS OF SUBSTITUTION, AND SPECIFIC ASPECTS OF GEOGRAPHIC MARKET DEFINITION

2.9.1 CHAINS OF SUBSTITUTION

The relevant market for a focal product or area may be extended to include several indirect substitute products or areas through a chain of substitution. This phenomenon is consistent with the hypothetical monopolist test. Every link in the chain is in essence a next iteration of the test.

To see the logic of the test, consider the example of the holiday park merger in the Netherlands presented in section 2.5. The survey that was used to assist the product market definition also shed some light on the geographic scope of the market. Respondents were asked how far they are willing to travel to their destination for a weekend break (maximum three nights). The question was asked only for weekend breaks because holidaymakers tend to be unwilling to travel as far as they would for breaks of a week or longer—this would therefore identify the narrowest geographic market. On average, the maximum travel time for weekend breaks was four hours and the median answer was three and a half hours (the median is the middle one if you rank all observations in a sample from lowest to highest). This confirmed the view of the merging parties that people are willing to travel for around three to four hours to their weekend breaks. This travel time determines the catchment area of each park—we discuss the use of drive-time-based isochrones below. Respondents were also asked whether they would consider a destination abroad if all prices in the Netherlands for holiday parks with an indoor pool were increased by 10% for two years. It turned out that as many as 65% of customers who considered the holiday in question would indeed go abroad (the Netherlands is a small country—in a three-to-four-hour drive you will have crossed a border regardless of where you start). This high percentage sales loss, together with the maximum travel times mentioned above, indicated that the geographic market cannot be limited to the Netherlands only. Hence, from the perspective of Dutch holidaymakers, the geographic market for short breaks covers at least the Netherlands, Belgium, and some nearby parts of Germany (in particular the Eifel region), and the north of France. This is the first link in the chain of substitution.

Holiday parks in this broader area do not only compete for Dutch custom but also for visitors from neighbouring countries (the Dutch represented only around 45% of customers of Center Parcs at the time, while 30% were German, 15% French, and 10% Belgian). German customers within a three to four hour drive would come from places like Düsseldorf or Hamburg. For them, there are other popular short-break destinations within a three to four hour drive from where they live, including the Ostsee region, Thüringer Wald, or the Schwarzwald, which are in the other direction from the Dutch parks. This means that Center Parcs in the Netherlands and surrounding areas competes directly with holiday accommodation in those other destinations, despite the fact that Dutch holidaymakers are less likely to travel to those destinations. This is the

example, the Schwarzwald competes in turn with nearby destinations in France, Switzerland, and Austria. Following this logic, there could well be a geographic market for short-break destinations spanning large parts of western and central Europe.

Chains of substitution may also arise in product market definition. There may possibly be an all-cars market, from Skoda Fabias to Alfa Romeo Spiders and Porsche 911s, or an all-broadband-Internet-access market that covers all available download speeds from 2Mbit/s. This last example has been confirmed by several European telecoms regulators, including Ofcom, the UK regulator, in a 2010 market review:

> In 2008 we concluded that the [wholesale broadband access market] definition did not have an upper speed limit, i.e. there is a 'chain of substitution' through the available broadband internet access speeds. This means that for an asymmetric broadband internet access product of any given speed, there are lower or higher speed products (the next links in the chain) which are sufficiently close substitutes that products of all speeds are subject to a common pricing constraint.
>
> Current broadband packages available in the market tend to be at specific clusters of speed, such as 2Mbit/s, 8Mbit/s, and 20Mbit/s. One of the key characteristics of broadband packages is the download (and to some extent upload) speeds, with higher speed services commanding higher prices. Therefore for a given speed service, a 5–10% SSNIP would decrease the price differential between the speed of the service in question and the next service up. If there are sufficient consumers who switch up, it would render the SSNIP unprofitable and suggest a single product market between the two speeds.
>
> Our consumer survey shows that given a 10% increase in the price of the package consumers are currently paying, 14% of residential customers and 22% of business customers are willing to switch their broadband service to a different speed package. Given the critical loss factors it would suggest that the original price rise is not likely to be profitable . . .
>
> In addition, end users are almost as likely to switch up to a higher quality service as they would switch down to a lower quality service (around 6%–7%). This further suggests that consumers see the range of price/speed options as potential substitutes should the price of their package increase. As a result there is unlikely to be an identifiable break across the range of speeds available to warrant separate markets for low and high speed services within the current generation broadband access services available in the market.[47]

Ofcom thus found a continuous chain of substitution among prevailing Internet speeds. It then addressed the question of whether the chain might break following the recent introduction of 'super-fast broadband services' that use fibre rather than copper access and offer speeds of 40–50Mbit/s. The regulator considered that in the next few years there would still be one single market, but recognized that over time some breaks in the speed chain may occur because of the continued development and growth of Internet applications that require very high speeds (such as online games and Internet TV). This illustrates the importance of the second time dimension of relevant markets that we introduced in section 2.2: the scope of the market depends on the period of investigation. It also

[47] Ofcom (2010), 'Review of the wholesale broadband access market,' March, [2.90], [2.91].

highlights the complex market definition questions that arise in innovative industries where consumers migrate to new products over time. We return to this in section 2.12. Finally, the example shows a recognition that there may be breaks in the chain of substitution, and that such breaks must be analysed carefully. We discuss this now.

2.9.2 BREAKING THE CHAIN

You should perhaps be somewhat sceptical of chains of substitution that are too long. Remember that in every further iteration of the hypothetical monopolist test, the monopolist gets to control another substitute product (or area), so has a little bit more pricing power than previously. Before long, pricing power would be such that a SSNIP would be imposed for the focal product or (area)—remember that we said in section 2.4 that the SSNIP question still applies only to the focal product or area in these further iterations, so once the price increases there you have a relevant market. For a link in the chain to be strong, the next substitute must place substantial competitive pressure on the product or area that you have just added to the group controlled by the monopolist. If holiday accommodation in the Schwarzwald strongly constrains the holiday parks in the areas surrounding the Netherlands, the monopolist will not be able to increase prices of the latter by much, and this in turn constrains prices in the Netherlands.

One particular situation in which the chain of substitution may be strong is where you have one supplier that competes nationally with several regional suppliers and follows a national pricing strategy. In several European countries, a national telephony incumbent competes with a number of regional cable companies in the provision of telephony and Internet services. The cable operators do not compete directly with each other (often this is for legacy reasons; many of these operations used to be run as monopolies owned by local authorities). But each of them competes with the national operator. Now take the hypothetical situation where one of the bigger regional cable operators reduces the price of a bundled telephony/Internet/TV package (see section 2.10 on market definition for bundles). The national operator decides that it needs to match this decrease because it does not want to lose customers in that region. But that means reducing prices in the whole country, because it always charges the same everywhere (it advertises its packages at the national level, which is one of its competitive advantages over the regional operators). Hence, the other regional cable operators are also forced to lower their prices, so as to keep in line with the national operator. In effect, the first regional operator, through a chain of substitution via the national operator, has placed an indirect competitive constraint on all other regional operators. Clearly, if the national operator decides to abandon its policy of uniform pricing and to flex prices regionally instead (and if regulation allows this), the chain breaks down and different regional markets may be defined.

2.9.3 KEEPING AN EYE ON THE FOCAL PRODUCT

If you have defined the product market based on supply-side substitution, you should still

the geographic market. Take the following stylized example, based on an abuse of dominance investigation by the water regulator in England and Wales.[48] The investigation concerned the treatment of leachate. A somewhat less glamorous product than Italian women's designer shoes, leachate is the liquid that originates from rain percolating through the different layers of waste on a landfill site. It can be toxic and therefore needs to be treated appropriately before being discharged in the sewer. This occurs mainly at waste-water treatment works. The leachate is normally transported by tanker—'tankered'—from the landfill site to these treatment works. The complaint concerned access for companies tankering leachate to the waste-water treatment works which were owned by the incumbent regional water and sewerage company. There was discussion about the possibility of supply-side substitution by other treatment works that could treat any liquid waste but that did not currently treat leachate. Such supply-side substitution was deemed feasible overall, although there were some limitations on the available capacity of these other treatment works, and other types of liquid waste were more profitable to treat (recall the paper example in section 2.7). Let's ignore these limitations and assume that the product market is extended to the treatment of all liquid waste, on the basis that any treatment works can treat any type of liquid waste, including leachate. This brings us to the geographic market. Tankering leachate is expensive—it ceases to be economical if the distance between the landfill site and treatment works is greater than 30 miles (see also the sub-section below on transport costs). But most other types of liquid waste are more economical to ship over longer distances—say, up to 60 miles. Given that the product market has been defined as all liquid waste (in this stylized example), should you consider this longer distance when delineating the geographic market?

The answer is no. To see why, consider the (again slightly stylized) situation in which all waste-water treatment works that currently treat leachate and lie within 30 miles of the major landfill sites in the region are owned by the incumbent. Beyond the 30 miles are several other treatment works that could switch (supply-side substitute) to treating leachate, and therefore have been included in your relevant product market. But you can clearly see that leachate cannot be economically tankered to those supply-side substitutes because the distance is too great. If you had taken your product market—treatment of all liquid waste—as the starting point for your geographic market, you would have overlooked the pocket of market power in leachate treatment itself. Your geographic market definition should therefore always refer back to the focal product, ie, the product where the competition concern arose in the first place (in this case treatment of leachate only).

2.9.4 TRANSPORT COSTS AND TRADE PATTERNS AS A BASIS FOR GEOGRAPHIC MARKET DEFINITION

Geographic patterns of supply and demand are often driven by transport costs. It is therefore relevant to consider these costs when delineating geographic markets.

[48] Ofwat (2005), 'Investigation into charges for the treatment of tankered landfill leachate by United Utilities

Concrete is a classic example of an industry where geographic markets are highly localized because of transport costs—in-transit mixers generally have to reach the construction site within 90 minutes from loading at the plant to prevent the concrete from becoming hard and unusable. Other products may be more easily shipped—fresh flowers grown in the Netherlands make it in large quantities to homes in the USA and Japan. But even for those products, transport costs matter—air freight charges on Dutch flowers will add to the cost faced by the US purchaser, more so than on flowers grown in nearby Mexico, and this may ultimately affect market definition.

At its simplest, you can incorporate transport costs directly in the hypothetical monopolist test—if transport costs between two regions represent more than 5%–10% of the prevailing prices, a monopolist in one region could increase price by 5%–10% without attracting supply from the other region. The NMa applied this test when reviewing a merger between the only two major sugar producers in the Netherlands (which it ultimately approved).[49] Transport costs for industrial sugar were found to be around €0.08 per tonne/kilometre. With a prevailing price of around €650 per tonne, a 5% price increase might invite imports from a distance of 400 km (5% of €650 equals €32.5; transporting a tonne of sugar over a distance of 400 km costs €32, so becomes attractive with the new price). The NMa also took into account the fact that the new European internal market regime for sugar—promoting greater market integration—might lead to a fall in market prices to around €500 per tonne in the coming years. Assuming the same transport costs, the corresponding distance over which imports would be profitable after a 5% price increase is 300 km, and the NMa took this as the basis to expand the relevant market beyond the Netherlands.

When applying the hypothetical monopolist test with transport costs, the usual caution is required. For example, prevailing prices may already encapsulate the transport cost differential (this is akin to the cellophane fallacy discussed in section 2.6). Likewise, transport costs may change over time, as new technologies or modes of transport are developed (mix-at-site trucks can now deliver smaller amounts of concrete with less time pressure than in-transit mixers, so some concrete customers have greater choice than previously), or the conditions of competition in the transport market change (competition authorities have uncovered an international price-fixing cartel in the supply of air cargo services in recent years, which may have added to the transport cost of products like fresh flowers).[50] The period of investigation is a relevant dimension of the market here, as explained in section 2.2.

Sometimes it may be more relevant to consider transport costs from the perspective of the buyer rather than the supplier. In our holiday park example, the customers travel to the supplier rather than the other way round (this is a purer form of demand-side

[49] NMa (2007), *Cosun–CSM*, Zaak 5703/304, 20 April.

[50] US Department of Justice (2007), 'British Airways Plc And Korean Air Lines Co Ltd Agree To Plead Guilty And Pay Criminal Fines Totaling $600 Million For Fixing Prices On Passenger and Cargo Flights', press release, 1 August; and European Commission (2010), 'Antitrust: Commission fines 11 air cargo carriers €799 million in

substitution; where it is the supplier that moves, this is more closely related to supply-side substitution). When it is the customers that move, instead of focusing on transport costs as such, you would measure what in transport economics is commonly known as the generalized cost of travel. This is the sum of the monetary costs of a journey (the cost of the ticket if using public transport or the fuel if using the car) and its non-monetary costs. The latter consists mainly of the value of the total time spent on the journey. Calculating this value can take into account many different factors, such as opportunity costs (what else you could have done during that time), or the fact that some of the time spent has a greater 'cost' (waiting at a railway station is more 'costly' than sitting in a comfortable leather seat while on the move).[51]

In addition to using the analysis of transport costs or generalized costs of travel as part of the hypothetical monopolist test, you can use such analysis to delineate specific geographic areas within which you then assess the conditions of competition. In the example of the Dutch sugar merger, the authority drew circles of a 300km radius around each merging party's production facilities. The geographic market was thus extended to include producers in Belgium, France, and Germany—the so-called European 'sugar belt'. In this broader market the two Dutch producers had a share of around 20% (compared with nearly 100% in the Netherlands only). We discuss this further in the next sub-section on isochrones.

An alternative to assessing transport costs is to consider actual trade patterns. If there is a lot of movement of goods between two regions—by either suppliers or buyers—then it is likely that they form part of the same geographic market. A formal approach to this is the Elzinga–Hogarty test, named after two economists who first proposed that a geographic region is likely to be a separate market from other regions if there are few imports—'little in from outside' (LIFO)—and few exports—'little out from inside' (LOFI) (Elzinga and Hogarty, 1973). The test has been widely applied, including in several hospital mergers in the USA, where it was assessed whether hospitals attract patients from far afield.[52] The Dutch sugar merger case also included an analysis of current import and export patterns. The NMa found that imports into the Netherlands from the other countries in the 'sugar belt' were as yet relatively limited, but would be expected to increase under the new EU internal market regime. As with the SSNIP, exactly how you set the threshold for LIFO and LOFI is inherently arbitrary—Elzinga and Hogarty proposed two thresholds at 75% and 90% of supply (according to this logic, if imports and exports as a proportion of total supply are above the threshold, the geographic market is likely to be wider).

Using current trade patterns to define geographic markets can, however, lead to erroneous conclusions. In particular, it misses the basic logic of the hypothetical monopolist test, which is about demand and supply responses to small changes in price. Even if you observe few imports or exports at present, what matters is whether trade would

[51] More background on generalized cost can be found in Button (1993).

[52] One example is *FTC v Tenet Healthcare Corp* 17 F Supp 2d 937 (ED Mo 1998), rev'd 186 F 3d 1045 (8th Cir. 1999).

start to take place *after* the hypothetical monopolist imposes a SSNIP. Another criticism levelled at the Elzinga–Hogarty test—in the specific context of hospital mergers—is that it suffers from the 'silent majority fallacy' (Capps et al., 2001). There may be a significant difference between those patients who travel further afield and those who do not. So even if you observe that many people who live near one hospital travel to other distant hospitals, this does not mean that the first hospital has no market power—it may be that those who do not travel to other hospitals do not have the ability to do so (because of their condition, for example), and the first hospital may be able to exploit its captive customers. In essence this points to the possible existence of price discrimination markets, as discussed in section 2.8. Nonetheless, despite these valid criticisms of the Elzinga–Hogarty test, examining current trade patterns can still provide a useful sense-check for your geographic market definition, just like product characteristics and absolute price differences can provide useful sense-checks for product and geographic market definitions.

2.9.5 ATTACK OF THE ISOCHRONES

Until quite recently, geographic markets were often delineated by drawing circles of various sizes around production facilities or population centres to determine catchment areas—circles with a 30-mile radius around landfill sites for tankering waste or a 300km radius around sugar production facilities for the transport of sugar (as in the examples used earlier). You can see that circles are rather unsophisticated—they may accurately reflect distances as the crow flies, but not how far you can travel in a tanker filled with leachate, a truck filled with sugar, or an empty bus on its way from the depot to the starting point of a route in the early morning (the basis for defining local bus markets, as discussed in section 2.7). Modern technology has allowed for some further sophistication in this type of analysis. Mapping software is now regularly used to delineate more coherent geographic markets that take into account the characteristics of the underlying road network and travel speeds. Instead of circles you get isochrones. An isochrone is a line that joins together all of the points that can be reached within a constant journey time from a given starting location. To draw the isochrones, the drive time first needs to be determined—you can use evidence from surveys, analyse transport costs, or observe existing travel patterns (as discussed in the previous sub-section). There is an inevitable element of judgement in this, but you can always test the sensitivity of your results by drawing slightly broader and narrower isochrones to see whether the competitive landscape within the isochrones changes. Having determined the journey time, the second step is to construct the isochrones. The mapping software will contain a road matrix and a set of speeds for travelling on each road—speeds can be adjusted depending on the type of vehicle, the type of road, and the time of day.

What do you take as the central point of the isochrone? In line with the logic of the hypothetical monopolist test, this should reflect the focal product as closely as possible. In the bus example discussed in section 2.7, the competition concern arose with respect

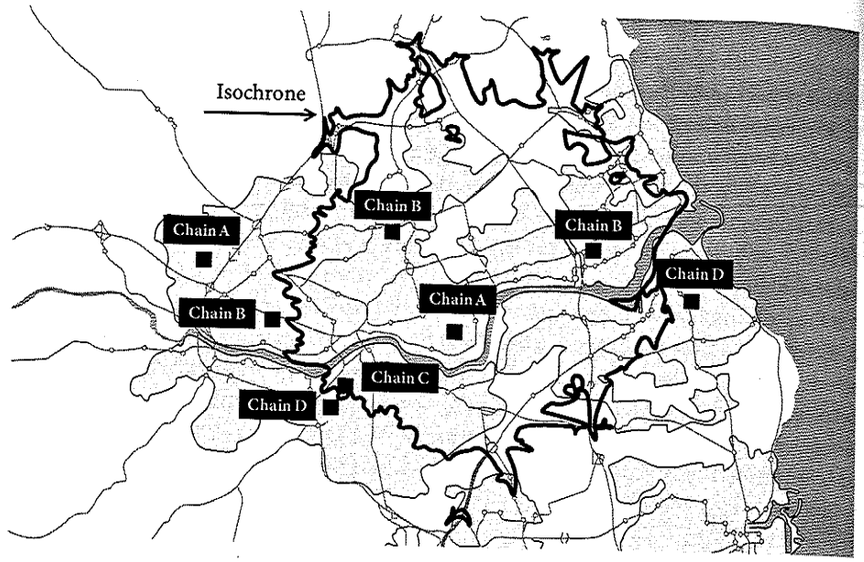

Fig. 2.6 20-minute isochrone for the cinema owned by chain A

to certain routes in the centre of Chester.[53] Based on supply-side considerations, the geographic market was found to include all existing bus depots within 30 minutes' drive time of the routes. In theory, you could draw an isochrone around every point on the route from which the bus can begin service in the early morning. Pragmatically, however, an isochrone around one particular main stop on the route (eg, the central bus station in Chester) will often provide sufficient insight. The next step in the competition analysis is then to analyse market shares and other indicators of competitive strength of the various operators in this geographic market.

In other cases you would draw the isochrones around the relevant outlets in question—each supermarket or holiday park. This gives the catchment area of each outlet, which provides useful information on the competitive constraints they face. For example, in a cinema merger in the UK—*Odeon/UCI* (2005)—the OFT and the parties drew isochrones for a 20-minute drive time around each cinema, a distance considered to reflect the cinema's core catchment area.[54] A 30-minute isochrone was then used as a sensitivity check. (The only part of the country in which isochrone analysis was not used was Central London, where people tend not to drive to the cinema.) Figure 2.6 shows the boundaries of the 20-minute isochrone for a cinema owned by one of the

[53] *Chester City Council and Chester City Transport Limited v Arriva PLC* [2007] EWHC 1373 (Ch).

[54] Office of Fair Trading (2005), 'Acquisition by Terra Firma Investments (GP) 2 Ltd of United Cinemas International (UK) Limited and Cinema International Corporation (UK) Limited', 7 January; undertakings accepted 9 May. We acted for the

merging parties (chain A) in one of the cities considered. This figure has been drawn using the MapInfo software. The irregular shapes and 'peaks' of the isochrone simply reflect the configuration and speed of the underlying road network. The other merging party (chain B) has two cinemas in the area, while only one competing cinema is present (chain C). The other competing operator in the city, chain D, has no presence in this isochrone. Strictly speaking, such isochrones do not accurately reflect the relevant geographic market, since they are supply-centred rather than demand-centred. Two cinemas do not compete because they are within 20 minutes' drive time of each other, but because they are both within reach (ie, within 20 minutes' drive time) of the same customers. One simple adjustment to the supply-centred isochrones is therefore to perform a 'population re-centring', which comes down to drawing the isochrones around major population centres and then assessing which outlets are within their reach. For pragmatic reasons, you do this only for the major population centres, rather than for every single customer as pure theory would dictate, since a focus on the major centres will usually give a sufficiently clear picture.

2.10 MARKET DEFINITION FOR COMPLEMENTS AND BUNDLES

2.10.1 SUBSTITUTES, COMPLEMENTS, AND BUNDLES

Recall from section 2.3 that substitute products have a positive cross-price elasticity while complements have a negative one. If a 10% increase in the price of flights between London and Paris results in a 15% increase in demand for Eurostar train journeys—a substitute—the cross-price elasticity of rail with respect to air travel between the two cities is 1.5 (15% divided by 10%). If a 10% increase in the price of gin reduces demand for tonic water—a complement in gin-and-tonics—by 10%, the cross-price elasticity of tonic water with respect to gin is –1. Substitutes and complements therefore have exact opposite properties. This has profound implications for the assessment of mergers (as further discussed in Chapter 7). If you merge two substitute products, you can expect them both to increase in price—this is commonly understood and lies at the heart of merger control in competition law. Less well understood is that merging two complementary products has the opposite effect: they decrease in price after the merger. This is why such mergers are often inherently benign to consumers of the product (they may still give rise to competition concerns of a different nature, such as portfolio and foreclosure effects—see Chapter 7).[55]

[55] That a monopolist with two complementary goods would reduce prices compared with the situation where each good is controlled separately was first identified by the French economist Antoine Augustin Cournot in 1838—the same Cournot from the famous oligopoly model discussed in Chapter 3 (Cournot, 1838). This 'Cournot effect' is also of relevance in the context of vertical restraints, where it leads to the problem of double

But first let us explore the implications of complements for market definition. If you expand the hypothetical monopolist's influence by adding a substitute product, prices will go up—the monopolist now no longer cares about losing sales to those substitutes. Following the same logic as above, if you add a complementary product instead, a profit-maximizing monopolist will set a lower price than before. A hypothetical monopolist in the supply of gin would selfishly raise price to extract maximum profits, without caring about the negative effect that this has on tonic water sales. If the monopolist controlled tonic water as well, it would care about lost tonic water sales, and hence would not raise the price of gin by as much when maximizing the profits of both products. This is exactly the opposite effect of adding a substitute product—a complementary product acts as a pricing constraint when it is inside the monopoly group, whereas a substitute product acts as a constraint when it is outside. How do you apply the hypothetical monopolist test here?

The test itself can be readily applied in cases where there are complementary products. The main difficulty lies in choosing the right starting point—in the first iteration of the test, does the monopolist control just one product, or does it control the two (or more) complements? The outcome of the test may differ substantially (a gin monopolist might impose a SSNIP while a gin-and-tonic monopolist might not). The choice will have to be determined by the specific market and the specific competition problem that you are considering. If the real companies in the market supply just one product, your hypothetical monopolist should do so as well—the production and sale of gin by spirits companies are usually separate from those of tonic water, so it is not unreasonable to take gin as the focal product if that is where the competition concern arises, and ignore the complementarities with tonic water at the market definition stage. In contrast, if the real companies normally control two complementary products already, or consumers in the market usually buy the complementary products together (in a bundle), it may be more appropriate to start the test from a hypothetical monopolist that controls both of those products. An obvious, if somewhat absurd, example is that of left shoes and right shoes. They are different products, and they are complements, not substitutes. Shoe producers supply and sell them together, and consumers buy them together. You can therefore safely start your market definition exercise with a hypothetical monopolist who produces both left and right shoes, and ask if it could impose a SSNIP on pairs of shoes.

The same logic holds for bundled products more generally. Bundling occurs when two or more products are sold more cheaply together than individually (this is known as mixed bundling), or when they are sold only together (pure bundling). Bundling is a very common business practice—with a left shoe comes a right shoe, with a car come four wheels and tyres and a stereo, with your mobile phone subscription comes a handset. Bundling has many economic justifications, but it may also raise competition concerns—we address this in Chapter 4. Here we explore how bundling may affect market definition. There is an economic case for defining relevant markets for bundles, as opposed to individual products, where: (i) bundling is pure rather than mixed—the two products are available only as a package, not separately (you can't buy a left shoe without a right shoe); (ii) the complements are consumed in fixed proportions (one left and one right shoe, one car and four wheels); and (iii) all suppliers in the market sell bundles. However, there are also markets where the picture is more mixed—only some suppliers sell bundles, or bundling is a relatively new phenomenon. An example is the various 'triple-play' and 'quadruple-play' packages that you can get from telecoms providers (fixed telephony, broadband Internet, TV, and mobile telephony). These packages have been on offer for several years and are steadily gaining ground among consumers, but still there are operators in the market with different combinations of services in the bundles (some are traditionally stronger in TV services, others in telephony services), and some customers purchase them while others still obtain individual services from different providers. Here there is no clear-cut case for either defining relevant markets at the level of the individual products or at the level of the bundles. It is probably good practice to do both. This allows you to explore the relative strengths of the different operators in each product separately, and at the same time to assess whether some operators are particularly well positioned in the offering of the bundles (not all providers of the individual services may be able to offer all services in the bundle), and identify any scope for market power in that way.

Below we illustrate how the hypothetical monopolist test might work when there are complementarities between products in two specific types of market—aftermarkets and two-sided platform markets.

2.10.2 MARKET DEFINITION FOR AFTERMARKETS

In most circumstances, you buy a product, you consume it, and that's the end of the story. In some markets, however, buying the product is just the beginning. In order to make full use of it, you subsequently have to buy associated products or services that are complements to the original product. Examples of this include razors and razor blades, printers and cartridges, game consoles and video games, and proprietary software and further upgrades to that software. The same holds for durable equipment (such as photocopiers, machinery, cars, and cash registers), and subsequent maintenance services and spare parts for that equipment. The market in which you buy the associated product or service is commonly referred to as the secondary market or aftermarket. Many competition cases have arisen in aftermarkets. Often they involve complaints that the supplier of the primary product is using its control over that product to exclude independent suppliers from selling the secondary product. A prime example is the *Kodak* litigation in the USA, which lasted from 1987 to 1998—Kodak was found guilty of monopolizing the market for the servicing of its high-volume copier and micrographics equipment.[56] It had a strong position in the supply of this equipment (the primary product), which required extensive, ongoing maintenance (the secondary product).

[56] The two main rulings in this case are *Eastman Kodak Co v Image Technical Services, Inc et al.* 504 US 451 (1992), and *Image Technical Services, Inc et al. v Eastman Kodak Co* 125 F 3d 1195 (9th Cir. 1997). A discussion

Many independent organizations serviced Kodak equipment at the time, in direct competition with Kodak's own national network of service technicians, but then Kodak stopped selling spare parts to these organizations, such that they could no longer provide maintenance services. We discuss abuse of dominance in Chapter 4. The issue addressed here is whether market definition can be a helpful intermediate step in assessing such practices.

Two main questions arise: Does the primary producer have the ability to foreclose the secondary market from competition? And if competition in the secondary market is foreclosed, does the primary producer have the ability to exploit its customers there (in other words, are customers worse off if Kodak keeps the aftermarket to itself)? In essence you want to know what the competitive constraints are on the primary producer. Market definition normally helps in identifying these competitive constraints. In this case, the main market definition question is whether the secondary market is a separate relevant market from the primary market, or whether there is a 'systems market' that includes both the primary and secondary product. However, regardless of how you delineate the market exactly, the source of any market power of the primary producer is its control over the primary product. So even if you conclude that the aftermarket is a separate market, you would still need to explicitly take into account the link between the primary market and the aftermarket.

A systems market would be more appropriate if customers are fully aware at the time of purchase that what they buy is a system, ie, they know that they need to purchase the secondary product or service as well, and take this into account when making the primary purchase decision. One example would be game consoles and video games. When choosing between a Nintendo Wii, a Sony PlayStation 3, and a Microsoft Xbox 360, you (or, more likely, your children) will normally consider which games are available for that console. If the market is a systems one, you could test whether a hypothetical monopolist controlling one system faces competition from other systems—does the Nintendo Wii face a pricing constraint from the PlayStation and Xbox? Does Kodak face strong competition from other producers of high-volume copiers and micrographics equipment? You would also take into account the general logic of complements we discussed earlier—if there were a genuine systems market, and companies priced on a systems basis, Kodak would be constrained from raising the price of the secondary product if that negatively affected sales (and hence profits) of the primary product, and vice versa. This market definition therefore helps you identify cases where any negative effects on customers are limited, ie, cases where system competition is so strong that it constrains any exploitation of market power in the secondary market.

There are other circumstances, however, in which system competition is not sufficiently strong to constrain market power in the secondary market, and where a systems market definition would therefore be less appropriate. This can occur where customers do not fully take the need for the secondary product into account when choosing the primary product, or have insufficient information at the time of the primary purchase. It can also arise when the primary product already has a large installed base of custom-

they were to move to another primary product). You would also be more doubtful about a systems market if there is evidence (as there was in *Kodak*) that the supplier of the primary product does not engage in system pricing. Finally, there may be a separate market for the secondary product if the competitive dynamics of the primary and secondary markets are very different, or where many secondary products are compatible with multiple primary products—for example, spare car parts such as tyres that fit on any make of car, or garages that service any make of car.

2.10.3 MARKET DEFINITION IN TWO-SIDED PLATFORM MARKETS

If you have followed developments in competition law in the last five years, chances are you have heard the term two-sided market. It is one of those economic concepts that suddenly became fashionable among scholars and practitioners. The economic literature on two-sided markets that has snowballed in recent years was originally inspired by a real competition case—the investigations since the late 1990s by competition authorities worldwide into interchange fees in the Visa and MasterCard credit card schemes.[57] We return to interchange fees in Chapter 5. Here we explore the economic properties of credit card schemes—in particular their two-sided nature and the implications for market definition.

Credit card schemes are one among many examples of networks or platforms that face demand from two groups of customer—hence the term two-sided demand—with positive demand externalities existing between these two groups (we explained the concept of externality in Chapter 1). In credit card schemes, these externalities arise between issuers—the participating banks that issue cards to their accountholders—and acquirers—the banks that sign up retailers (more generally referred to as 'merchants') to accept the card scheme. A card becomes more valuable to retailers the more consumers hold and use it, and a card becomes more valuable to have in your wallet the more retailers accept it in their shops. Other examples include video game consoles (externalities between users and game developers), PC operating systems (PC users and applications developers), newspapers (readers and advertisers), and nightclubs and speed-dating events (men and women).

In setting their pricing structure, platforms face the celebrated chicken-and-egg problem—they must 'get both sides on board', but in order to get participants on one side they need to have sufficient participants on the other side. This is not dissimilar to a situation where you have two complementary products—if demand for one goes up, then so does demand for the other. It makes commercial sense to set a pricing structure that optimizes the size of the network. This usually means setting relatively low prices

[57] Reserve Bank of Australia (2001), 'Reform of credit card schemes in Australia', December; Visa International—Multilateral Interchange Fee (Case COMP/29.373) Decision of 24 July 2002, [2002] OJ L318/17; and Office of Fair Trading (2005), 'Investigation of the multilateral interchange fees provided for in the UK domestic rules of Mastercard UK Members Forum Limited (formerly known as MasterCard/Europay UK Limited)',

for the more price-sensitive customer side, and relatively high prices for the less price-sensitive side. Likewise, it may imply setting lower prices to the side that is more important in terms of attracting the other side, ie, the side that generates larger network effects. For this reason, 'free' newspapers are free to users and recover costs from advertisers, and some speed-dating events are free to women and thus recover more of their costs from men. There is nothing inherently anti-competitive about these pricing practices. Indeed, to the extent that they have the effect of expanding the platform or network size, they may increase overall welfare compared with the situation where all sides are charged the same price.

A first step in the competition analysis is, as usual, to define the relevant market—does the platform face competitive constraints from other platforms? The hypothetical monopolist test can be used here. We illustrate this with a stylized application to a hypothetical credit card scheme. The first question that arises is what you take as the starting point—a hypothetical monopolist on one side, on the other side, or on both sides? It usually does not make much sense to consider a hypothetical monopolist on one side in isolation. In the context of a credit card scheme, if the question is whether the scheme competes with other payment systems, you want to test competition on both sides. The complementary and network effects between the two sides mean that no real card scheme would set prices on one side without considering the effects on demand from the other side. First, a reduction in the number of card transactions on one side would by definition mean a reduction in transactions on the other side—one retailer transaction necessarily involves one cardholder transaction (the two sides are like complements in this respect). Second, because of the network effects between the two sides, when setting prices, the monopolist would take into account the fact that a reduction in demand on one side also makes the product less attractive for the other side. Applying the hypothetical monopolist test to a credit card scheme is therefore similar to a second-round iteration of the test, where the monopolist controls two products. The main difference with the standard situation is that the two products are like complements, not substitutes.

How do you apply critical loss analysis? You follow the same logic as in standard critical loss analysis, but now take into account two sides of the market. So you test whether the hypothetical monopolist can profitably impose a 5%–10% increase in charges to retailers (known as merchant service charges), a 5%–10% increase in cardholder charges, or a combination of the two that results in an overall increase in the 'system price' of 5%–10%. To determine the critical loss, you need some estimate of the actual sales loss on either side that would result from the price increase, and an estimate of the profit margin.

Let's start with the margin. Suppose that the issuing banks that participate in the scheme (those banks that have relationships with cardholders) incur a cost per transaction of €1.20, and the acquiring banks (those that deal with retailers) a cost of €0.30 (such a cost imbalance between issuing and acquiring banks is not uncommon in payment systems, since the former incur costs such as fraud prevention, bad-debt write-offs, and cardholder loyalty awards). The total 'system cost' is therefore €1.50 per transaction. As to revenues, suppose that issuing banks [illegible]

from cardholder charges, while acquiring banks receive €1.50 per transaction from merchant service charges (this revenue imbalance is also not uncommon since retailers tend to have greater willingness to pay than cardholders—indeed, the combination of the cost and revenue imbalances provides the economic rationale for the interchange fees paid from acquirers to issuers, as discussed in Chapter 5). The total 'system price' per transaction is €2.50, so the initial price–cost margin equals €1.00, or 40%.[58]

Now suppose you have evidence on retailers' and cardholders' responses to price increases. A survey among retailers suggests that a 10% increase in merchant service charges would lead to a loss of 15% of total card transactions as the retailers where those transactions take place would cease to accept the card or discourage card payments by imposing a surcharge or minimum purchase value. Econometric evidence on cardholder elasticity of demand suggests that a 10% increase in cardholder charges would lead to a 30% drop in transactions—cardholders are apparently quite prone to switching to other payment methods. The last step in this hypothetical example is for you to compare the actual loss with the critical loss. You have to do this according to the system price and system cost. The merchant service charge (€1.50) represents 60% of the system price (€2.50). A 10% increase in that charge is therefore equivalent to a 6% increase in the system price. This 6% increase leads to a 15% loss in transactions. Applying the critical loss formula of section 2.6 gives you 6% divided by 6% + 40%, which equals 13%, so the actual loss exceeds the critical loss. The cardholder charges (€1.00) represent 40% of the system price. A 10% increase in those charges is therefore equivalent to a 4% increase in the system price. The critical loss for this is 4% divided by 4% + 40%, or 9%, so is again exceeded (transactions fall by 30%). Hence a 10% price increase is unprofitable both on the retailer side and on the cardholder side. This would imply that the market is broader than just credit cards.

2.11 MARKETS ALONG THE VERTICAL SUPPLY CHAIN

As noted in section 2.2, an additional market dimension is the vertical layer in the supply chain. The same product—a car, a washing machine, ice cream, a bunch of fresh flowers—may pass through a whole set of intermediaries on its way from producer to end-consumer. The relevant market may be different depending on which layer of the chain is considered. This in turn will normally depend on where in the chain the competition problem in question arises. If the concern relates to an exclusive distribution arrangement between a washing machine manufacturer and a large retail chain, the product market should be defined with reference to that upstream level of the supply chain. Washing machines are the focal product, and the first question to ask is whether the manufacturer has market power in the supply of washing machines (if it has, the

[58] You may work out that if both acquirers and issuers are to earn a 40% margin, the scheme requires an interchange fee paid by acquirers to issuers equal to €1.00 per transaction so as to make up for the cost–revenue imbalance in the scheme (the acquiring banks then receive €0.50 net per transaction while incurring a cost of [illegible]

arrangements are more likely to have anti-competitive effects—see Chapters 4 and 6). However, it will still be relevant to consider the final layer of the supply chain (the downstream market) as well, which is the layer where you purchase a washing machine from a retail outlet. This is for several reasons, as discussed below.

2.11.1 MARKETS IN DIFFERENT LAYERS AND THEIR INTERACTION

First, even if the upstream market between manufacturer and retail chain is the focal product where the exclusive arrangement arises, this arrangement is likely to have an effect on the final consumer market as well—you as a consumer may face less choice of where to buy your washing machine. The downstream market can therefore be an additional relevant market. In this market you want to analyse the competitive position of the retailer, and the effects of the arrangement on prices and other terms offered to consumers. Second, even if the focus is on the upstream market, the demand in this market—in this case the demand for washing machines—is often largely driven by the demand of final consumers. Whether different types of washing machine—high-capacity versus low-capacity; those with and without a tumble dryer—are close substitutes depends on the preferences of final consumers. The demand of retailers for washing machines is often a 'derived' demand from that of consumers. Retailers will buy and stock the products that consumers want.

Indeed, the main evidence you have on substitution will often relate to consumer behaviour rather than retailer behaviour. If that is the case, you have to exercise some care when applying the hypothetical monopolist test to the upstream market. Suppose you find that if the retail price of washing machines is raised from €500 to €525—a 5% increase—consumer demand falls by 10%. The own-price elasticity of demand for washing machines is –2. Is this price increase profitable? That depends on whether you look at it from the perspective of the retailer or the manufacturer. Recall Table 2.3, which shows the critical loss threshold depending on the initial price–cost margin and the percentage price increase. Suppose that the retailer obtains washing machines from the manufacturer at a wholesale price of €250. The retailer margin is therefore 50% (half of the retail price of €500). From Table 2.3 you can see that the critical loss threshold for a 5% price increase and 50% margin equals 9.1%. The 10% fall in demand (just) exceeds this critical level, which would lead you to conclude that the price increase is unprofitable. However, what you have just determined is that a hypothetical monopoly retailer of washing machines could not impose a SSNIP. The focal question here, in contrast, is whether a hypothetical monopoly manufacturer could profitably increase price. How can you use the consumer evidence from the downstream market to infer any conclusions on this last question in relation to the upstream market? You can use the principle that retailer demand is derived from consumer demand and apply critical loss analysis. The manufacturer might achieve the increase in the retail price from €500 to €525 by increasing the wholesale price from €250 to €275 (assuming that the retailer fully passes on this price increase of €25 to its customers). The wholesale price increase is 10%. The sales loss is still 10%—retailer demand is simply derived

purchase washing machines that they ultimately sell to consumers (with some delay usually). Now suppose that the cost of production of the washing machine is €125, which means that the manufacturer has an initial margin of 50% as well. As you can see from Table 2.3, for a 10% price increase the critical loss threshold is 16.7%, so for the hypothetical monopoly manufacturer this price increase is profitable and the market is no wider than washing machines.

Note that in this example we assumed that the retailer passes on 100% of the wholesale price increase. However, even if the pass-on rate is lower, your conclusion that the manufacturing layer is worth monopolizing remains unchanged, since a lower pass-on rate means that even fewer sales are lost, so the wholesale price increase is even more profitable. We discuss the economic principles of pass-on in more detail in Chapter 10 on the quantification of damages.

2.11.2 DO END-CONSUMERS ALWAYS MATTER MOST?

Retailer or wholesaler demand is not always fully derived from consumer demand. It may be the case that even where final consumers are willing and able to switch between different products, retailers cannot because some products are 'must-stock' items. For your flight from London to Paris, you can choose among several airlines. But when you book the ticket through a travel agent (something that probably occurs less and less with the rise of Internet bookings), you expect that agent to offer tickets for all the major airlines. So travel agents have less choice than final consumers—often it is the case that each major airline is a 'must-have' for each travel agent. What do you do in situations where consumers have choice (so you don't have to worry about them) but retailers or other intermediaries in the chain do not?

Competition policy is often, implicitly or explicitly, more concerned with effects on final consumers than effects on intermediaries. This is why the analysis of mergers and business practices upstream in a supply chain will frequently consider effects further downstream as well as the effects upstream (as in the washing machine example above). Sometimes you can go a step further and place the main weight of the analysis on the downstream market—you may not care what effect a practice or agreement has on certain intermediaries in the chain, as long as final consumers are not negatively affected. This may occur where the main effect of the practice or agreement is to redistribute profits from one layer in the chain to another—for example, from manufacturer to retailer, or vice versa. Such arrangements are very common and can lead to fierce commercial disputes in which competition law arguments are often invoked, but the final consumer may not always be affected. We discuss this further under vertical restraints in Chapter 6. Some policy judgement may be required in such cases. Is a particular intermediary layer in the value chain really worth preserving? Consider the example of the leachate abuse of dominance case in section 2.9.[59] The complaint was made by an intermediary acting as a broker

[59] Ofwat (2005), 'Investigation into charges for the treatment of tankered landfill leachate by United Utilities

between landfill sites and the providers of leachate tanker and treatment services. The incumbent water and sewerage company, which owned the treatment works and also had its own tanker operations, began to contract with some of the landfill sites directly, thereby 'squeezing' the intermediary broker out of the market. If you take a narrow view on market definition at the layer of intermediary brokerage services, you might find that the broker—given the nature of its activities—had no choice but to deal with the landfill sites and treatment works in that area. If you take a broader view, you might conclude that what matters is whether the landfill sites, as the customers in this market, have sufficient choice (the authority found that they did have some choice, including building their own treatment facilities). The position of the broker as intermediary matters less if you take that view—customers are unharmed, so why care if the broker goes out of business? (In market definition terms, you are implicitly saying that the leachate broker could dedicate itself to other activities instead.)

2.11.3 TURNING THE SUPPLY CHAIN UPSIDE DOWN FOR MARKET DEFINITION

Another issue related to defining markets in the context of a vertical supply chain arose in the European Commission's investigation into loyalty rebates offered by British Airways (BA) to travel agents.[60] When you deal with a case like this, the concern is usually about the effects that such rebates have on competition between airlines, and therefore ultimately on airline users, ie, the passengers. A sensible first stage in your analysis of these competitive effects would therefore be to define the relevant market with airline services as the focal product. You can then assess whether the airline has a dominant position in the provision of flights—if it has, such practices are more likely to produce anti-competitive effects than if it does not (we discuss this in Chapter 4). The European Commission took a different approach. It defined the relevant market the other way round, as that for the provision of air travel agency services, which are purchased from travel agents by airlines. It found that BA had a dominant position as a buyer in that market, with a (declining) market share as a buyer of over 40% of all relevant flights. Perhaps in this case the outcome might have been the same if the Commission had defined the market for the supply of airline services (where the 40% market share would also have been used as an indicator of dominance). But defining the market in the way the Commission did carries the risk of focusing the attention of the competition analysis on the wrong market, ie, it leads you to focus on the effects of the rebates in the travel agent market, while the real concern is with distortions to the supply of airline services to final customers. To capture this concern, defining a market for the supply of airline services is more informative.[61]

[60] *Virgin/British Airways*, (Case COMP/D-2/34.780) [2000] OJ L30/1.

[61] A separate competition concern identified by the Commission in this case was the potential distortive

2.12 PRODUCT SUBSTITUTION VERSUS PRODUCT MIGRATION

There are markets where consumers migrate from one product to another over time. Many modern examples come from the digital world: consumers migrate from dial-up (narrowband) Internet access to broadband Internet, from analogue TV to digital TV, from VHS to DVD (and now Blu-ray), from traditional cameras to digital cameras, and, to some extent, from letters and faxes to email. However, product migration occurs in other industries as well—for example, from wooden tennis rackets to those made of graphite and other high-tech materials, or from long-distance coach services to low-cost airlines for domestic and regional travel. Sometimes the shift from one product to the next may happen overnight, but often the old and new products live alongside each other for some time. In competition investigations, the question then regularly arises as to whether the two products form part of the same relevant market—are the old and new products regarded as close substitutes? As you have seen in this chapter, the hypothetical monopolist test is defined in terms of relative price changes (can the monopolist increase the price of one product by a small amount without losing too many customers to substitute products?). Product migration is not driven by small price changes. However, this is not to say that it should be ignored when defining relevant markets. Product migration can still have implications for whether a product is worth monopolizing.

Product migration is normally in one direction—from the old product to the new. The question of whether one is in the other's relevant market can arise in two directions, depending on which is the focal product. One direction is to ask whether the old product places a competitive constraint on the new product. At some point in time this becomes a less interesting question. Once you have switched to broadband Internet or digital TV, you are unlikely to ever switch back to the old product (narrowband or analogue TV, respectively). Nonetheless, in the early stages of the new product, the old one may still be very attractive to consumers, such that suppliers of the new product have to compete not only with each other but also with the suppliers of the old product. This may have been the case in the early years of broadband Internet access, when broadband providers had to keep prices low to attract narrowband users.

The more complicated question to ask is whether the new product imposes a pricing constraint on the old product. Two examples in which authorities had to address this question are the analysis of narrowband markets by Oftel, the predecessor to Ofcom, in 2003, and the analysis of the leased-lines market by OPTA, the Dutch telecoms regulator, in 2005. In the first of these cases the following statement was made:

> Oftel's view is that there is not a single market including both narrowband and broadband. While Oftel recognises that customers have moved from narrowband to broadband and that this is likely to continue to some extent in the future, it is not clear that this is substitution in response to a relative price change as such, as opposed to customers upgrading to a higher quality product that was not previously available.[62]

Hence, Oftel did not consider product migration to be a relevant form of substitution for the purpose of market definition, since it did not occur in response to relative changes in the prices of both products. It was therefore still concerned with narrowband services as a separate market from broadband. OPTA used the same reasoning with respect to the leased-lines market. A leased line is a permanently connected communications link between two premises, dedicated to a customer's exclusive use. At the time, business users had begun to gradually switch from leased lines to other data services, such as those based on Internet protocol technology. While these newer data services had more variable capacity and required more outsourcing of network management functions, customers considered them a lower-cost alternative to leased lines. Having noted a high degree of migration from leased lines to the newer data services between 2002 and 2004, OPTA nonetheless concluded that the movement between the products was not relevant for market definition:

> Switching and migration from service A to B does not automatically constitute demand substitution. Demand substitution requires that switching from A to B is caused by changes in the price difference between A and B. In this case, there is price pressure. Migration, however, can result from other factors, such as the emergence of a completely new service (B) or changes in user preferences. Consumers migrate as a result, where this migration no longer depends on further small (5% to 10%) changes in the price difference between A and B.[63]

Do these cases reflect an overly restrictive interpretation of the hypothetical monopolist test? What ultimately matters is whether the old product is a product worth monopolizing. Here, product migration does have relevance. Recall the charts earlier in this chapter. In Figure 2.2 you saw how the demand curve of a product can shift downwards if a close substitute product gains in popularity (or becomes cheaper). Product migration has precisely this effect. The rise of graphite tennis rackets meant that, over time, the demand for wooden rackets declined. The same happened to the demand for dial-up Internet access and VHS players. Going back to Figures 2.2 to 2.4 earlier in the chapter, you can see the consequences of the demand curve shifting downwards. The intersect between demand and marginal costs (which are assumed to remain the same) moves ever further towards the left-hand side of the demand curve. Remember from Figure 2.3 that this left-hand side is the elastic part of the curve. Now consider the more dramatic case in Figure 2.7, where demand has fallen substantially due to migration away from the product—this new demand curve can be expressed as $p = 3 - q$ (previously this was $p = 10 - q$, so for every price point there are 7 fewer buyers). In perfect competition, price would equal marginal cost, so $p = 2$, and quantity would be 1. Hence, even under perfect competition you would already have elastic demand (you may have worked out that at this point the elasticity equals –2). As it is, if perfect competition is the starting point in this example, a hypothetical monopolist would still impose a SSNIP (the monopoly price is 2.5 and monopoly output 0.5, so the price

[63] OPTA (2005), 'Ontwerpbesluit huurlijnen' (draft decision leased lines), 1 July, [280] (translated from Dutch).

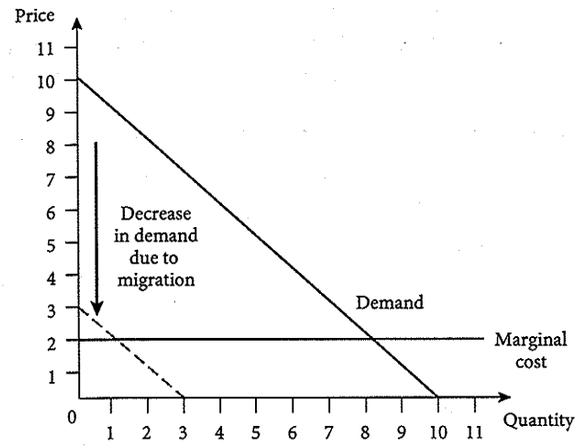

Fig. 2.7 Decrease in demand due to migration to other products

increase compared with a situation of perfect competition is still 25%). But you can see that if the starting price is higher than under perfect competition, or the demand curve falls even further, there comes a point where even a monopolist would not impose a SSNIP. That is the point when the old product is no longer worth monopolizing.

In practice, once it is recognized that product migration does matter for market definition as a matter of principle, the main question becomes whether migration places a sufficiently strong constraint. This depends in part on the speed of migration—does migration become significant within the period covered by the investigation? It may well be that in the examples of Oftel and OPTA such an analysis would have led to the same conclusions as those drawn by the two regulators—indeed, in both cases there were other indications that separate markets could be defined. One reason why the old product could still be worth monopolizing may be that those customers who have not yet migrated to the new one are in fact the less price-sensitive customers—in this case, while the monopolist has lost many customers due to migration, it can still profitably exploit the remaining ones. Such an effect is not captured in Figure 2.7. You might capture it by assuming that migration does not shift the whole demand curve down, as in Figure 2.7, but rather makes it rotate downwards from the point where price is zero and quantity is ten, such that the curves becomes steeper from that point. The profit-maximizing price then remains the same, and only quantity falls. These are all further refinements that you can incorporate into the analysis of your specific case. The important point in this sub-section is that product migration cannot be dismissed as irrelevant for market definition or for the analysis of competitive constraints—product migration is a relevant form of competitive pressure between products.

2.13 MARKET DEFINITION FOR FEATURES OTHER THAN PRICE

2.13.1 IT'S NOT JUST THE PRICE THAT MATTERS

The traditional hypothetical monopolist test focuses on the ability to raise price, and this is what we have discussed so far in this chapter. Raising prices is the most direct, and often most visible, way of exploiting a position of market power, and is usually the main concern of competition authorities. However, products do not compete only on price—they compete on non-price aspects such as quality, design and brand. A hypothetical monopolist can also exploit customers by downgrading those aspects of the product offering and thereby saving costs—for example, it could cut corners on quality or not improve the design of the product. This could equally be of concern to competition authorities, as consumers are negatively affected. In principle, therefore, the logic of the hypothetical monopolist test can be applied to non-price aspects as well. You could ask: would the hypothetical monopolist put into effect a 5%–10% reduction in quality? The difficulty tends to lie in establishing a clear metric for those non-price aspects (how do you measure quality?), and in identifying how a small change in a particular product feature affects the monopolist's profits (does lower quality mean lower costs? Does lower quality reduce demand?). But these difficulties have not prevented economists and competition authorities from using the hypothetical monopolist framework to define relevant markets according to product features other than price.

The UK Competition Commission's (CC) inquiry into the *Sportech/Vernons* merger in 2007 provides an example.[64] Between them, the merging parties operated the major weekly football pools in the UK, which consist of the stakes received from people playing in the pools. Football pools had been very popular, but had shrunk by a factor of 10 since 1994 when the National Lottery was introduced (you could say that this is a form of product migration as discussed in the previous section). The stakes in the pool are influenced by the entry prices set by the pool operators. Only a portion of the pool money goes into the prize fund—the payout ratio is in the range of 20%–25%. Most of the prize fund is paid out to the winning player as a jackpot. Thus, there are three important features of the product—the entry price, the payout ratio, and the size of the jackpot. The hypothetical monopolist test in its purest form would apply to the entry price only. However, the CC considered evidence on the effect of a small but significant deterioration in the merging parties' product offering more generally—ie, a small increase in the entry price, a small reduction in the jackpot, and a small reduction in the payout ratio. Note that for each of these features there is a clear metric (they are all expressed as monetary value) and a directly observable link with the profitability of the operators (the greater the payout, the smaller the profit left for the operator), which

[64] Competition Commission (2007), 'A report on the anticipated acquisition by Sportech plc of the Vernons

makes the test more straightforward to apply than for some other non-price features. The evidence indicated that not many customers actually switched between the two major pools in response to changes in the product offering. This suggested that the two did not really compete with each other, and this was the main reason why the CC concluded that the merger did not substantially lessen competition. As to competition with other gambling products, there was evidence that many customers would simply play the pools less if the product offerings worsened, but not switch to the National Lottery or other gambling products. The CC drew a somewhat confusing conclusion from this:

> Survey evidence shows that customers who stop playing the pools either save the money or spend it in a wide range of other ways. However, we considered that expanding the market definition to include all alternative uses of disposable income would not be appropriate. We therefore conclude, at this stage, that the market is no wider than the football pools . . .
>
> In our view, given the wide range of alternative spending or saving products to which football pools customers might switch in response to a price rise, a market definition which took account of all these alternatives would not be workable, even if we thought they were a constraint on pools prices.[65]

In line with the market definition principles set out in this chapter, you would instead conclude that the market *is* wider than football pools—customers who stop playing the pools are a relevant sales loss to the hypothetical monopolist. You cannot determine where exactly the market boundaries lie because there is no clear closest substitute for the football pools, but nor is there any need to do so, because the fact that a hypothetical monopolist in football pools cannot profitably worsen the product offering implies that the merging parties cannot either.

2.13.2 TIME ELASTICITIES IN GEOGRAPHIC MARKET DEFINITION

In section 2.9 we discussed transport costs as a factor in geographic market definition, and how in the case of individual consumers you can analyse the generalized cost of travel, which incorporates monetary and non-monetary costs (mainly the value of the time spent). Transport costs can determine the extent of the geographic market. Economists have taken this a step further and introduced the concept of time elasticity as a substitute for price elasticity in defining relevant markets, specifically in the context of hospital mergers (Capps et al., 2001). In many countries the government regulates the prices charged by hospitals and other healthcare providers, so the main concern when implementing the hypothetical monopolist test is not a price rise but a reduction in the quality of service. However, measuring the quality of healthcare delivery is notoriously difficult and patients generally have little information to judge the quality, which can complicate the market definition analysis. As an alternative to measuring changes in price and quality, the time elasticity approach measures how

many consumers would switch to competing healthcare providers in response to a hypothetical 5%–10% increase in travel time to a healthcare provider in a given geographic area. How many patients go to an alternative provider? The increase in travel time does not reflect what the hypothetical monopolist is really going to do—it is not actually going to relocate the hospital—but rather is an approximation for other reductions in quality of service. The willingness to change to alternative providers after hypothetical increases in the travel time gives an indication of the responsiveness in demand to changes in other aspects of the service. The analysis then makes certain assumptions about how the time elasticity relates proportionately to the price elasticity, which in turn allows the identification of a critical loss level above which the increase in travel time (and reduction in quality) becomes unprofitable because too many patients switch elsewhere.

You may get the impression that this is a rather roundabout way of applying the hypothetical monopolist test to non-price aspects of a product. Yet time elasticities have been used in several healthcare merger cases in the USA and Europe, as responsiveness to travel time to hospitals can be measured more easily than reactions to other non-price features of the service. In other markets, customer responses to changes in quality may be assessed more directly than in healthcare, and in some cases you might attempt to use a quality elasticity—the percentage change in demand divided by the percentage change in quality. Would our hypothetical monopolist for Italian women's designer shoes exploit its market power by cutting on quality and design? The principles would be the same as for the time elasticity, and based on the same logic as critical loss analysis. The main challenge is, again, identifying how a small change in quality affects the hypothetical monopolist's profits—if quality worsens, how much extra profit does the monopolist make through saved costs, and how much profit is lost due to customers switching? These questions are generally easier to answer for price changes than for quality changes.

2.13.3 MARKET DEFINITION IN INNOVATION MARKETS

In industries such as high-tech, IT, and pharmaceuticals, companies compete not only, or even mainly, on price, but also on who brings the most innovative products to market. Market power can be reflected in the control over innovative capabilities as much as in the control over prices. How can you capture this scope for market power during the market definition stage of your competition analysis? The US agencies—the DOJ and FTC—have come up with a rather 'innovative' approach: innovation capability is treated as a separate market in itself:

> An innovation market consists of the research and development directed to particular new or improved goods or processes, and the close substitutes for that research and development. The close substitutes are research and development efforts, technologies, and goods that significantly constrain the exercise of market power with respect to the relevant research and development, for example by limiting the ability and incentive of a hypothetical monopolist to retard the rate of [illegible] delineate an innovation market only when the capabilities to engage in the relevant research and development can be associated with specialized assets or characteristics of specific firms.[66]

This principle was applied in the DOJ's review of the proposed merger of General Motor's Allison Division and ZF Friedrichshafen, a German company, in 1993.[67] GM Allison and ZF were the world's two largest producers of medium and heavy automatic transmissions (used in buses, heavy trucks, and similar types of vehicle). ZF was dominant in Europe, whereas GM was dominant in the USA. For the purposes of assessing the merger, the DOJ defined both product markets and innovation markets. The production and sale of transmissions for buses and heavy trucks constituted the relevant product markets, and geographically they were confined to the USA (this was because it was a review by the DOJ; applying the same logic to Europe would probably have resulted in a separate European geographic market as well). In contrast, the innovation market was defined as the worldwide market for technical innovation in the design, development, and production of transmissions. This made a significant impact on the outcome of the case. While GM and ZF did not compete in many of the defined product markets in the USA—and hence the merger caused little concern in those markets—together they controlled most of the relevant worldwide innovation market. The DOJ was concerned that the merger would significantly lessen the competition in innovation. The proposed merger was ultimately abandoned.

Similar innovation markets are frequently considered in the context of technology transfer and licensing agreements. The European Commission tends to define technology markets in such cases:

> Technology markets consist of the licensed technology and its substitutes, i.e. other technologies which are regarded by the licensees as interchangeable with or substitutable for the licensed technology, by reason of the technologies' characteristics, their royalties and their intended use. The methodology for defining technology markets follows the same principles as the definition of product markets. Starting from the technology which is marketed by the licensor, one needs to identify those other technologies to which licensees could switch in response to a small but permanent increase in relative prices, i.e. the royalties. An alternative approach is to look at the market for products incorporating the licensed technology.[68]

Again, conceptually, the approach is very similar to the hypothetical monopolist test.

[66] Department of Justice and Federal Trade Commission (1995), 'Antitrust Guidelines for the Licensing of Intellectual Property', p11.

[67] *United States v General Motors Corp* Civ No 93-530, filed 16 November 1993.

[68] European Commission (2004), 'Guidelines on the application of Article 81 of the EC Treaty to technology transfer agreements', Commission Notice, 2004/C101/02, 27 April, [22]. At the time of writing, new guidelines

2.14 MARKET DEFINITION IN PRACTICE: MAIN EMPIRICAL METHODS

In this chapter you have been reading about the economic principles of market definition. You may still wonder how the conceptual framework of the hypothetical monopolist can be tested quantitatively in practice. Even where quantification is not feasible, the framework itself already helps you in asking the right questions about market definition. However, in many cases, some attempt at measuring whether a small price increase is profitable for the hypothetical monopolist can be made. Critical loss analysis is the basis for this, as explained in section 2.5. Determining the *critical* loss level is relatively straightforward—there is a simple formula for critical loss, and the main bit of empirical analysis you must carry out for this step is to identify the initial price–cost margin, which is one of the ingredients of the formula (see Chapter 10 on how you can obtain margin information from financial accounts). A greater challenge lies in empirically estimating the *actual* sales loss after a price increase, which you then compare with the critical loss to determine whether the price increase is profitable (the term 'actual' is of course not quite accurate because it is still hypothetical, but highlights the contrast with the 'critical' loss that is used as a threshold). This is where economists usually come in with their toolbox of empirical methods to measure customers' responsiveness to price. The aim of this book is not to explain in detail how these tools work; however, this section gives you a flavour of three of the main empirical tools used in market definition: regression analysis, surveys—both often used for measuring price elasticities—and price-correlation analysis. We also indicate how you, as a competition lawyer, can ask some critical questions when you are presented with the results of such empirical analyses. Chapter 10 provides a further discussion of regression analysis and other empirical methods in the context of quantifying damages.

2.14.1 USING REGRESSION ANALYSIS TO ESTIMATE ELASTICITIES

If you want to know how customers respond to price changes, one option is to observe actual prices and quantities for the product. How much of the product customers have actually purchased at different price levels tells you something about their 'revealed preferences' (as opposed to 'stated preferences', which is what you get from surveys, as discussed below). If you have a sufficient number of observations, you may be able to identify the demand curve and price elasticities. Imagine that you observe prices and quantities of the product—say, women's designer shoes—at regular intervals (monthly) over an extended period of time (two years). All observations are plotted in the price–quantity space in Figure 2.8. This is a highly stylized example—in reality you would not get such wide variation in the observed prices (economists like variation in the data, as that generally allows for more robust measurement). While eyeballing the figure already gives you some idea of the relationship between price and quantity (again, you would rarely see this in practice), you need to identify a more precise relationship that is also

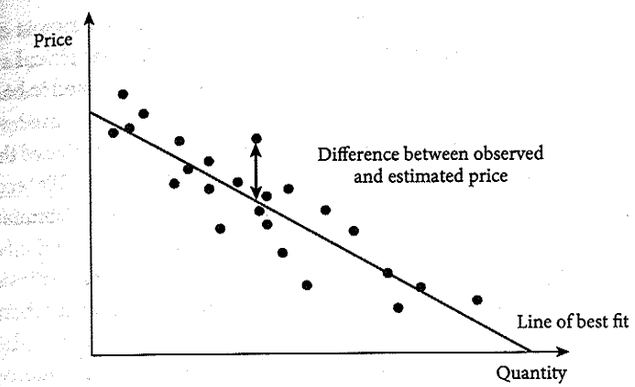

Fig. 2.8 The demand curve as the 'line of best fit'

statistically meaningful. This is where regression analysis comes in. Regression is a generic term for statistical methods that can be used to explain the variation in data using other factors. Economists (some of them specifically trained as econometricians—econometrics is the application of regression to economic data) use a regression to estimate the equation for the demand curve based on the observed prices and quantities.

Recall that the theoretical demand function in Figure 2.1 was expressed as $p = 10 - q$. Using regression analysis, you can estimate a demand function of this form that best matches the dots in Figure 2.8. You start by writing the equation for this line as $p = a - b \times q$ + error term, where a and b are the coefficients to be estimated—a is where the demand curve intersects with the y-axis, usually referred to as the 'constant term' ($a = 10$ in Figure 2.1), while b is the coefficient that represents the slope of the curve ($b = 1$ in Figure 2.1). The error term represents the distance from the estimated line of best fit to each of the dots (in Figure 2.8 this difference is illustrated for one particular observation). In other words, it reflects how far off the line you are at each given quantity (some observations are closer to the line than others). 'Line of best fit' means that you want to minimize the differences overall. Statistically, the trick to this involves several steps. First, you draw a line that you think fits reasonably well. Second, at each quantity you take the difference between the line and the actual observation (this can be positive or negative, depending on whether the observation is above or below the line you drew). Third, you take the square of each difference. Squaring has two rationales: you are treating points below and above the line the same (squaring a negative number gives a positive number), and you are giving more weight to the larger differences (as squares get bigger for higher numbers). Fourth, you take the sum of all the squares. Fifth, you follow the above four steps again for other possible lines, until you have found your line of best

approach is called 'ordinary least squares' (OLS) estimation. It is the most commonly used technique in econometrics and you will often see it when empirical studies are presented in competition cases (OLS and other techniques are explained in basic econometrics textbooks[69]). More sophisticated techniques are sometimes used where this suits the nature of the data, but the basic logic remains that the objective of the analysis is to find the line of best fit. When OLS was first developed in the 1930s, econometricians used to calculate the minimum sum of the squares by hand. Fortunately, these days you can use statistical packages such as Stata to do it for you.

In practice, when you are presented with the results of a regression analysis, you will often see that the demand curve is specified in logarithms ('logs') rather than 'levels'—instead of a function 'p = a – b × q + error term', economists often estimate the function 'log p = a – b × log q + error term'. This means that all the price and quantity data has been transformed into logarithms. The logarithm of a number to a given base is the power to which that base must be raised in order to produce that number (10 is often used as the base; $10^3 = 1{,}000$, so the logarithm of 1,000 to base 10 equals 3). The regression then simply tests the relationship between these transformed series instead of the original series. One reason for using this statistical trick is that the log specification is better at capturing the price–quantity interaction when the demand curve is not linear. Another reason is that the coefficient b directly represents the own-price elasticity, rather than just the slope of the curve—ie, it directly gives you the answer you are after (this is due to a mathematical relationship to do with the first derivative of a logarithmic function).

There is a potential problem with the interpretation of the observations plotted in Figure 2.8. Each observation reflects the quantity that was actually purchased at the particular price. These actual outcomes are not just reflective of the demand curve, they are the result of the interaction between demand and supply. Economists refer to this as the endogeneity of prices. When you look at the position of any two dots in the price–quantity field, you cannot tell whether the difference between the two is attributable to a shift *along* the demand curve, or to a shift *of* the demand curve *itself* to the left or to the right. Only shifts along the demand curve are relevant for the estimation of the slope of the demand curve. To isolate the shifts along the curve, economists look for instances where a difference between two observations has resulted from a shift of the *supply* curve rather than of the demand curve. These shifts in supply aid help to highlight all the changes in demand that take place along the same demand curve. A typical cause of a shift in the supply curve is production cost changes. Hence, cost data is often included in the estimation to identify supply changes. In commonly used economics jargon, the cost data is used as an 'instrument' in the regression analysis so as to properly estimate the slope of the demand curve. Using instruments addresses the endogeneity problem. Sometimes other variables can be suitable as instruments—for example, the weather

[69] Examples of basic textbooks include Wooldridge (2005) and Gujarati (2009). For technical expositions of how econometrics techniques can be applied to market definition we refer you to Davis and Garcés (2009), Bishop and Walker (2010), and Motta (2004).

conditions at particular times can be used as instruments to explain supply shocks in agricultural products (such as the 2010 ban on Russian wheat exports as a result of the drought). More complex statistical techniques and modelling can also be used to overcome the price endogeneity problem, such as panel data econometrics methods (we explain the basics of these techniques in Chapter 10).

To illustrate how regression analysis is used to estimate elasticities for market definition, consider the merger between Pan Fish and Marine Harvest, two farmers of Atlantic salmon, which was examined by the UK CC (and a number of other national competition authorities) in 2006.[70] An important question in the case was whether salmon farmed in Scotland constituted a product market that was distinct from salmon farmed in Norway. The CC used regression analysis to estimate own- and cross-price elasticities for Scottish salmon. The equation specified by the CC was of the form: 'log of quantity of Scottish salmon = a – b × log of price of Scottish salmon + c × log of price of Norwegian salmon + d × log of income + error term'. Note that this is the demand function where quantity is a function of price, rather than the inverse demand function that is shown in Figures 2.1 and 2.8, where price is a function of quantity. Plugging in the data it had on quantities, prices, and income, the CC could estimate the coefficients a (the constant term), b (the own-price elasticity of demand for Scottish salmon), c (the cross-price elasticity of Scottish salmon with respect to the price of Norwegian salmon), and d (the income elasticity). The CC recognized that prices might be endogenous, so it used exchange rate data as an instrument for Scottish and Norwegian salmon prices. The estimation results indicated that the own-price elasticity for the Scottish salmon was –3.5, and the cross-price elasticity with respect to the price of Norwegian salmon was 3. These estimates suggest that the demand for Scottish salmon is quite sensitive to price, and also that it responds positively and strongly to Norwegian salmon price changes. While the CC did not perform a formal critical loss analysis based on the own-price elasticity estimate, it considered that the own- and cross-price elasticities were high, and concluded that this evidence was consistent with Scottish and Norwegian salmon being in the same product market.

2.14.2 WHAT QUESTIONS CAN YOU ASK WHEN PRESENTED WITH REGRESSION ANALYSIS FOR MARKET DEFINITION?

Estimating elasticities through regression analysis is conceptually straightforward, and generally allows for more robust results than simply plotting a line. Nevertheless, there are often pitfalls and complexities with such analysis (endogeneity of prices, discussed above, is one). High standards need to be met before a regression analysis can be considered robust—economists have developed a reasonably clear idea of what constitutes 'good economic practice'. The econometrics toolbox may always be something of a black

[70] Competition Commission (2006), 'Pan Fish ASA and Marine Harvest NV merger inquiry—Final report',

box to you, and debates on which particular econometric method is most appropriate in the case at hand can be rather esoteric, but there are things you can do to shake and rattle the box by asking critical questions, and see if it still holds together. The critical questions fall into three main categories (see also Chapter 10 on how to assess empirical methods for quantifying damages, and Chapter 11 on best practice in using economic evidence in competition cases).

First, you can ask questions about the data: what is the data coverage in terms of time period and products or market participants? How frequent is the data (monthly, yearly)? How large is the dataset? Are there enough observations to estimate elasticities robustly using econometric methods? Is the data of good quality (are there many missing observations or measurement errors)? Data coverage must be sufficient to cover the relevant products and time period, and the more observations (and more variance between them) there are, the greater the likelihood of finding statistically significant results. You can see that in the extreme, if you have only two observations, chances are that the line drawn from one to the other will not accurately reflect the demand curve. Even 24 observations, as in Figure 2.8, is on the low side. Second, you can ask questions about the econometric approach: is the econometric method appropriate for the market concerned and in light of the available data (does OLS work, or is a more sophisticated technique needed)? What assumptions underlie the econometric approach? Is the equation specified correctly, and is it in line with economic theory and market reality? Does it solve the price endogeneity problem? If instruments are used for price, are they appropriate? How do the results vary if alternative approaches or specifications are used? The third category consists of questions about the elasticity estimates: are the estimated elasticity values plausible (is the own-price elasticity negative as theory would predict)? How similar or different are they if compared with other available elasticity estimates? Are the estimated coefficients statistically significant? Testing for statistical significance helps in understanding the uncertainty surrounding an estimate and informs about how much weight should be placed on the analysis. You should expect any econometric results to be accompanied by a range of statistical diagnostic tests, which indicate whether the results are statistically significant (one such test is the t-test) and whether they suffer from potential statistical problems such as endogeneity.

2.14.3 USING SURVEY EVIDENCE TO ESTIMATE ELASTICITIES

Customer surveys are an alternative approach to estimating demand elasticities, and are being used increasingly in competition cases. Their appeal is that they can be relatively cheap and 'quick and dirty' (although more sophisticated variants of customer surveys can be more elaborate and expensive), and that they can be used to ask the SSNIP question directly—how would customers react if prices were raised by a small amount? The responses give an indication of the demand loss resulting from a SSNIP, which in turn you can compare with the critical loss. Through surveys you can obtain information on customers' stated preferences, ie, what they say they would do after a price increase, as opposed to revealed preferences, reflecting what they actually did. To ensure that the

survey results are reliable, there are a number of rules of good practice you can follow when designing it. These rules have not been widely debated or formalized—a good deal of literature exists on how to undertake customer research, but surveys designed for competition cases tend to have their own specific characteristics. A useful first attempt to document good practice rules has been made by the UK competition agencies in a recent consultation on guidance for the design and presentation of consumer survey evidence in merger inquiries.[71]

A typical survey designed for the hypothetical monopolist test is directed at existing customers of the product in question—those are the customers who the monopolist cares about, and who may defeat an attempt to raise prices. If you have ever been asked to respond to a survey of this nature (by telephone, on the street, or over the Internet), whether for market definition or (more likely) general market research purposes, you may agree that it helps if the survey is not too long and follows a certain logical structure. First, the survey should ask some 'warm-up' questions—about the respondents themselves, about current behaviour and habits, and about why the customer chose the given product and whether alternative products were considered. These questions reveal certain characteristics of customers and help respondents in beginning to think about the product choice situation. Then you can ask the SSNIP questions—how would the respondent react to a SSNIP? This is followed by a number of questions about switching to alternative products. This basic survey structure is illustrated in Figure 2.9.

Several of the case examples discussed in this chapter have relied on survey evidence, including the Dutch holiday parks and hospital mergers (section 2.5). Another case where survey evidence was used to establish the sales loss for the hypothetical monopolist test is the acquisition by LOVEFiLM of Amazon's online DVD rental business, which effectively created a monopoly in this product in the UK—hence the question arose as to whether the relevant product market was limited to online DVD rentals.[72] Internet-based surveys were commissioned on behalf of the acquiring party, canvassing

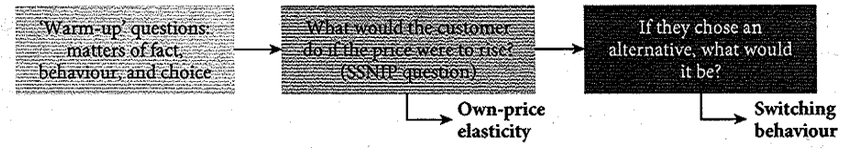

Fig. 2.9 Basic structure of a SSNIP survey
Source: Based on Walters and Reynolds (2008).

[71] Competition Commission and the Office of Fair Trading (2010), 'Good practice in the design and presentation of consumer survey evidence in merger inquiries—Consultation', May.

[72] Office of Fair Trading (2008), 'Anticipated acquisition of the online DVD rental subscription business of Amazon Inc. by LOVEFiLM International Limited', ME/3534/08, 8 May. We advised the acquiring party on this

the views of approximately 2,000 customers of online DVD rental services. First, questions about facts, behaviour, and choice were asked: which providers do you currently rent DVDs over the Internet from? Which of these would you consider to be your main provider? How much per month do you pay for your online DVD rentals from your main provider? The survey then asked the SSNIP question: 'Say that your online DVD provider decided to increase its prices, so that instead of paying £x per month, you had to pay £x + 10% (10% more) per month to receive exactly the same service, what would be your most likely response?' The 'x' was shown here as the price the respondent actually pays for online DVD rentals, as answered in an earlier question. The question on the screen also showed what £x + 10% actually is in money terms. Finally, questions about switching were asked: were you to cancel your online rental contract, how would you meet your demand to see films? The survey results showed that 30%–40% of respondents would switch after a 10% price increase. Since the critical loss threshold of 20–30% was below this actual loss, the relevant market was deemed to be wider than online DVD rental services. There was no clear nearest substitute—alternatives such as renting DVDs from 'bricks and mortar' rental shops, downloading films, and watching fewer films, were all given frequent mentions by switchers. But delineating the precise boundaries of the market was not necessary anyway, since the survey had already indicated that a hypothetical monopolist of online DVD rentals could not raise prices, and therefore neither could the real monopolist that was created by the acquisition. In the end, while there was some discussion about the interpretation of the survey results, the OFT cleared the merger because it found that internal business and strategy documents confirmed that the parties saw other products as their closest competitors, which was consistent with the market being broader than online DVD rentals.

If more time (and budget) is available, you may seek to carry out a conjoint or discrete-choice survey, which is a more sophisticated and generally more robust type of survey. Instead of asking respondents what they would do after a hypothetical price increase, a conjoint survey asks them to choose among products with different characteristics (price being one of these), which more closely resembles the actual choice-making situation that customers often find themselves in. The survey presents customers (or potential customers) with a menu of products that differ slightly from each other in their 'attributes', including price, functionality, and different aspects of quality. Two options are presented each time, and the respondent has to express a preference. By varying the product attributes in each option, a picture emerges of how customers trade off price and other attributes against each other. Econometric analysis can then be applied to the responses in order to estimate a price elasticity (or indeed to estimate an elasticity of demand with respect to any of the product attributes—conjoint analysis is often used for general marketing purposes as well).

Conjoint analysis was used for market definition in Ofcom's review of the pay-TV market.[73] In particular, it sought to understand whether channels containing premium

[73] Ofcom (2008), 'Pay TV second consultation—Access to premium content', 30 September, Annex 10.

Package A	Package B
- price: £15 per month	- price: £13.50 per month
- brand: Sky Sports	- brand: Sky Sports
- live FAPL games: YES	- live FAPL games: NO
- other football competitions: NO	- other football competitions: YES
- international cricket: YES	- international cricket: YES
- cricket featuring England: NO	- cricket featuring England: YES
- other sports: motor racing, darts	- other sports: rugby, golf, tennis, motor racing, darts, many others

Which of these options would you prefer?

Fig. 2.10 Stylized example of the choices presented in a conjoint survey

Source: Based on conjoint survey carried out in Ofcom (2008), 'Pay TV second consultation—Access to premium content', 30 September, Annex 10.

content, such as Football Association Premier League (FAPL) matches, constitute a separate market. The discrete-choice survey sought to test the importance of a variety of sports to households' decisions to subscribe to premium sports channels, and to inform on the level of substitution between sports. Respondents were presented with a series of pay-TV sports packages and in each case were asked which of two options they would prefer. Figure 2.10 shows a stylized example of one of these choice situations—two pay-TV sports packages which have as main differences the price and whether they included FAPL matches. After analysing the survey responses, Ofcom concluded that channels containing premium content do indeed constitute separate markets, since channels without such premium content are not seen as sufficiently close substitutes.

2.14.4 WHAT QUESTIONS CAN YOU ASK WHEN PRESENTED WITH SURVEY EVIDENCE FOR MARKET DEFINITION?

Not surprisingly, the use of surveys comes with the necessary health warnings, especially if they are of the 'quick and dirty' variety. There are a number of critical questions you can ask when presented with survey evidence. Where reasonably satisfactory answers can be given to these questions, surveys can make a useful contribution to the analysis, and some weight can be attached to the results.

A first set of questions relates to whether customers would actually do what they say in the survey response. A survey is a means of obtaining stated-preference data, and this may not necessarily reflect the actions that would be taken by customers in a real situation. Customers may overstate their switching when responding to a SSNIP question, leading to an overestimation of the sales loss and thus too wide a relevant market. A careful survey design (as discussed above) should mitigate this potential problem.

Careful phrasing of the questions is equally important—any confusion or ambiguity should be avoided, and questions should be neutral rather than leading. The representativeness of the survey is another critical aspect. Only when the sample is sufficiently large and representative of the relevant population as a whole can you obtain statistically meaningful results. For example, data from a survey may be biased if the survey is conducted in a location that tends to have different types of people passing through it at different times. Thus, carrying out a survey at a railway station on a weekday morning is likely to yield a different sample of travellers (mainly commuters) than if the same survey were carried out at the same railway station during the day at the weekend (mainly leisure travellers). This may bias the answers to the questions since the sample is not truly random and hence not representative. A lot of these potential issues can be resolved if the various parties involved in the case, in particular the merging parties and the competition authority, agree beforehand on the aim, design and wording of the survey (this may not always be feasible). This limits any disputes over the survey to the interpretation of the actual results, and not its design. A final set of questions you can ask relates to whether the survey results are consistent with the other evidence that is available in the inquiry. If yes, you can have confidence in the survey results—in both the holiday park and online DVD rental merger cases, the survey respondents' attitudes to price changes and switching corresponded with internal documents on what the companies themselves saw as their closest demand substitutes.

2.14.5 PRICE-CORRELATION ANALYSIS

Before the concept of the relevant market was developed for the purpose of competition law, economists used to regard a market as something where the 'law of one price' holds. The logic is that products that constitute a market should be priced very similarly. Any price differences within such a market would be removed through a combination of entry (by suppliers who can undercut higher prices), exit (by suppliers who can't compete at lower prices) and arbitrage (intermediaries buying low and selling high). Relevant markets in competition law are more practical and acknowledge that products can place competitive pressure on each other even if prices are not the same. Yet the idea of the law of one price still has its use for market definition, and lies behind price-correlation analysis. If products are close substitutes, and even if they have different prices, you would still expect those prices to move together over time—if some event causes the price of one of the products to change with respect to the other, this will trigger demand and supply substitution (customers switching, intermediaries engaging in arbitrage), and eventually prices will come back into line. The statistical term for such moving in parallel is correlation, as measured by the correlation index, which can take any value between +1 (when there is perfect positive correlation) and −1 (perfect negative correlation, ie, if one goes up the other goes down), with a correlation of zero meaning no correlation at all. The closer the correlation between the price series of two markets is to +1, the more likely it is that the two are in the same relevant market.

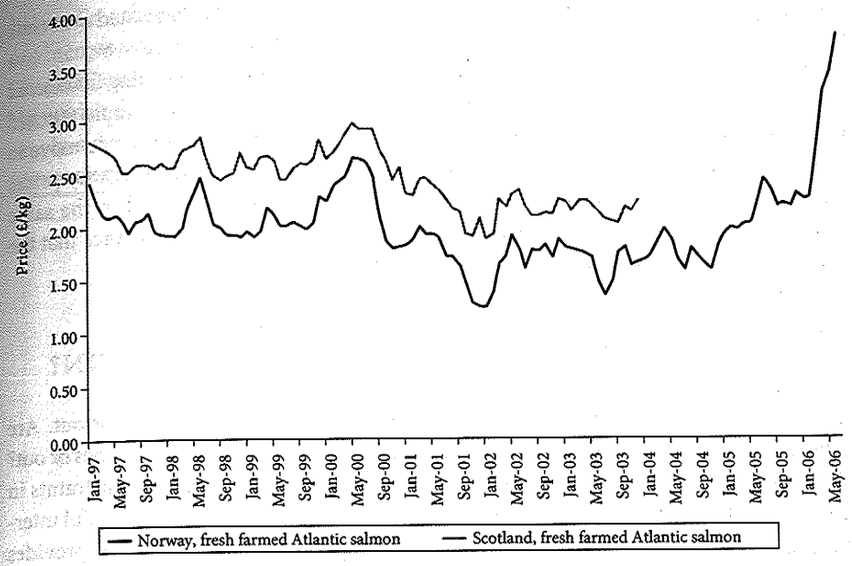

Fig. 2.11 Price correlation to support market definition: monthly farm gate prices of Scottish and Norwegian salmon

Source: Competition Commission (2006), 'Pan Fish ASA and Marine Harvest NV Merger Inquiry—Final Report', 18 December, Fig. 7.

The *Pan Fish/Marine Harvest* merger, which we saw earlier in this section, provides a good illustration of how price-correlation analysis can assist with market definition.[74] To answer the question of whether salmon farmed in Scotland constitutes a distinct product market from salmon farmed in Norway, the CC not only carried out regression analysis (see above) but also considered the correlation between the prices of Scottish salmon and Norwegian salmon over time. This is shown in Figure 2.11. You can see that prices of Scottish and Norwegian salmon tend to move in parallel, ie, they are highly correlated (the correlation coefficient is 0.91). Scottish salmon is consistently priced at a premium, but this does not alter the observation that the prices move in parallel, which suggests that Scottish and Norwegian salmon belong to the same relevant market.

Just like other empirical methods, price-correlation analysis comes with a number of health warnings. The CC was well aware of this: 'while informative, neither the correlation test, nor the extent of co-movement in prices (stationarity) test, can be viewed as definitive evidence of the existence of a relevant market.'[75] The main potential problem

[74] Competition Commission (2006), 'Pan Fish ASA and Marine Harvest NV merger inquiry—Final report', 18 December.

with the price-correlation test is that prices of two products can be correlated over time for reasons other than these products being in the same relevant market—you can have spurious correlation. A frequent cause of spurious correlation is prices being influenced by changes in costs that are common to both products. In the salmon example, it could be that salmon feed prices cause prices of Scottish and Norwegian salmon to move in parallel. However, in this particular case, there were no obvious reasons why prices might have moved in parallel, and the price-correlation results pointed in the same direction as the regression analysis and other evidence. This gives confidence that the conclusion on market definition is the correct one.

2.15 CONCLUSION: WHY MARKET DEFINITION?

Market definition is about whether a product (or geographic area) is in or out. Are French women's designer shoes in the market for Italian women's designer shoes or out? If they are in, you consider them as part of your analysis of competitive constraints in the market. If they are out, you don't. In this regard, market definition is a useful intermediate step in the competition analysis, and the hypothetical monopolist test provides a useful threshold for whether something is in or out—based on the question of which products constrain each other from raising price. Without such a threshold, there would be much confusion and lack of clarity on market definition, just like before the 1980s when the hypothetical monopolist test was introduced—submarkets within markets and market definitions based on product characteristics would probably creep up again. Furthermore, having market definition as a standard intermediate step in the analysis contributes to a degree of legal certainty. It assists companies in carrying out self-assessments of risks (or opportunities) under competition law—by defining the relevant market (or different scenarios for relevant markets), a company can calculate its own market share and that of rivals, and thus assess the likelihood of a finding of market power or lack of competition. Market shares measured within a relevant market are still a good (initial) guide to the existence of market power (see Chapter 3).

In this chapter we have explained the basic logic and some of the advanced features of the hypothetical monopolist test for market definition. We have also explained how exercising some judgement is inevitable when setting the threshold of whether something is in or out, and a measure of pragmatism is required when trying to apply the threshold in practice. Sometimes the required degree of judgement and pragmatism will be so high that market definition becomes less useful as an intermediate step. As explained in this chapter, a particular situation in which this may arise is when markets are characterized by a high degree of product differentiation. Asking whether something is in or out can be difficult, artificial, and potentially misleading when products are highly differentiated. The question becomes more about which products are each other's closest substitutes, than about which products are in or out. In these situations you can focus on the analysis of competitive constraints directly, and skip the market definition stage altogether. Rather than asking whether the relevant market contains

Italian, French, or all women's designer shoes, you ask whether Prada and Gucci, prior to a merger, provide a strong competitive constraint on each other. We return to this in Chapter 7 on mergers. Nonetheless, many of the economic concepts and tools underlying market definition that are discussed in this chapter—from demand elasticities and supply-side substitution to isochrones and complements—are also at the heart of the economic questions arising at subsequent stages of competition analysis.