5

The Relationship between Market Structure and Price

Merger investigations usually seek to determine whether the change in market structure caused by a merger will have a significant impact on the market outcomes for consumers. The outcome of most direct concern will be price although quality or choice effects may also be important though typically longer term and usually more difficult to assess. At the core of merger assessment then is the expected relationship between the number and size of firms operating in the market, market structure, and the prices or qualities that result from the competitive process.

Economic theory predicts that market structure affects prices. Under reasonably general conditions, a reduction of the number of players will result in an increase in market prices all else equal. This prediction forms the basis for the “unilateral effect” of a merger, where, post-merger, the new merged firm will usually have a unilateral incentive to raise prices above their pre-merger levels. This unilateral effect in turn may lead others to have an incentive to raise prices and this in turn usually reinforces the original unilateral incentive to increase prices. We term the former effect a “unilateral” effect since it is the incentive a single firm has to unilaterally increase prices. We term the latter effect a “multilateral” incentive since it involves independent actions by multiple parties each of which enjoy an incentive to increase prices following a merger. We will see that such effects are fairly generic in mergers involving firms that produce substitutes for one another.

This chapter explores the frameworks which can help competition agencies when they try to identify this effect in practice. Practically all models of competition predict that a change in market structure will have consequences for market prices. Still, empirically assessing the actual relation between structure and price is by no means always an easy activity. Nonetheless, we will see that several empirical strategies can be used to approximate the extent of an increase in prices that will result when concentration occurs. Understanding the underlying theoretical rationale for the relationship between structure and price will be important in order to design an appropriate way of empirically measuring the effect, so we will spend some time describing the basic underlying theory. We then present examples of methodologies that estimate the effects of market structure on price. Our examples are designed
5.1 Framework for Analyzing the Effect of Market Structure on Prices

We begin our discussion of the relationship between market structure and market outcomes by discussing the effect that the number of active firms has on the ruling equilibrium price or prices in the market under a variety of assumptions about the nature of competition. We then progress to examine methods which can be used to move our model from being purely a theoretical analysis into a framework that is appropriate for undertaking empirical work. Specifically, in the next section, we examine the potential drivers of the decision to enter a market and consider the effect that such entry has on the competitive process and also how we can learn about market power by observing such entry decisions.

We structure our study of the effect of market structure on prices by considering the following two-stage game. At stage one, firms decide whether or not to enter the market. If they enter, they incur a cost which is sunk (irrecoverable) at stage two. We call \( N \) the number of firms that decide to enter the market at stage 1. At stage two, the \( N \) active firms compete among themselves in prices or quantities. The game is represented in figure 5.1. In what follows, we follow the economics to provide practical guidance on appropriate data sets and techniques that could be considered for such analysis and also to point to a number of the potential problems previous investigators have faced when trying to identify the way in which price is likely to change with market structure.

Identifying the relationship between price and market structure is hard for a number of reasons. For example, whichever particular methodology we choose, we will need to address the difficult issue of identifying a causal link between market structure and the market outcomes—in particular price. In addition, if we adopt a longer-term perspective, both the number of active firms and the number and type of potential entrants may, on occasion, also constrain pricing power. If so, then in assessing the likely effect of a change of market structure one may also want to evaluate the constraint exerted by potential competitors. These are just two of a large number of potential difficulties analysts in competition agencies often come across. We outline a number of other difficulties below and then go on to describe potential solutions to these problems that the literature has developed.

Figure 5.1. A two-stage game.
5. The Relationship between Market Structure and Price

literature in analyzing such a game by starting with the examination of stage two and then proceeding “backward” to discuss stage one, the entry stage in the next section.¹

5.1.1 Theoretical Predictions about the Effect of Structure on Prices

Many economic models of competition can be embedded into this general two-stage structure and each will predict a relationship between market structure and market prices. We will establish the result for three important cases, namely the models in which firms are (1) price-takers, (2) oligopolists competing in prices, and (3) oligopolists competing in quantities. By examining these three canonical cases we are able to examine the mechanisms by which market structure can affect equilibrium prices. For example, we will see that, generally, a merger between two firms producing substitutes will tend to result in higher prices. Such results form the theoretical backbone of the investigations of the unilateral and multilateral effects of mergers.

5.1.1.1 Market Structure among Price-Taking Firms

The structure of market supply can be important for economic efficiency in a price-taking environment since where production takes place will usually matter for the aggregate costs incurred to produce any given level of output. That is, the number of firms that are producing will usually have an effect on the total costs of production. That in turn matters because the pricing pressures that firms face are determined by the intersection of market demand and market supply, which in a price-taking environment is determined by the industry’s marginal costs of production. A reduction in the number of firms will, except in special circumstances, reduce the aggregate supply to the market and hence induce the price to rise. Higher prices in turn induce increases in supply from at least one remaining active firm that, if it suffers from diseconomies of scale, will nonetheless find it profitable to produce extra output despite higher unit costs. Because of the potential diseconomies of scale, a lower number of firms may result in higher prices required to sustain a given level of aggregate output. Generally therefore, assuming a price-sensitive demand and firm-level diseconomies of scale, an equilibrium involving a reduced set of firms will involve lower quantities and higher prices.

A price-taking firm operates in a homogeneous product environment where quantity is usually the firm’s decision variable. It solves the following profit-maximization problem,

\[
\max_{q_i} p_i q_i - C(q_i),
\]

¹Technically, we examine equilibria of such games using “backward induction” to find the pure-strategy subgame perfect Nash equilibrium of the game; see your favorite game theory textbook.
where \( C \) is the total cost function describing the total costs of producing a given level of output \( q_i \) such that, for example,

\[
C = \begin{cases} 
   cq_i + \frac{1}{2}d q_i^2 + F & \text{if } q_i > 0, \\
   0 & \text{if } q_i = 0.
\end{cases}
\]

In this model, beyond the first unit of production, marginal costs increase with production and there is a limit to the efficient production scale. Solving the maximization problem describes the optimal quantity that this firm will want to supply at each announced price:

\[
q_i^* = \begin{cases} 
   \frac{p - c}{d} & \text{if } p_i q_i^* - C(q_i^*) \geq 0 \text{ at } q_i^* = \frac{p - c}{d}, \\
   0 & \text{otherwise}.
\end{cases}
\]

Next, suppose there are \( N \) symmetric active firms, each of which have produced positive amounts so that their (the firm’s) supply function can be summarized as \( q_i^* = (p - c)/d \), we may sum to give the market supply function:

\[
Q_{\text{Supply Market}} = N \left( \frac{p^* - c}{d} \right).
\]

If we further assume linear individual demands and \( S \) identical consumers so that the market demand is \( Q_{\text{Demand Market}} = S(a - bp^*) \) and that equilibrium price \( p^* \) is determined by the intersection of supply and demand, we may write

\[
Q_{\text{Supply Market}} = N \left( \frac{p^* - c}{d} \right) = S(a - bp^*) = Q_{\text{Demand Market}},
\]

which is an equilibrium relationship that we may solve explicitly to give the equilibrium price:

\[
p_i^* = \frac{Nc + Sda}{N + Sbd}.
\]

Note, in particular, that the equilibrium price depends on \( N \), that is the market structure, and also on the cost and demand parameters including the size of the market. Note also that with symmetric single-product firms, market structure can be completely described by the number of firms. Richer models will require a more nuanced description.

While the main aim of this section is to note that our various models imply that price is a function of market structure, it would be nice to see an analytical result which fits well with our intuition that prices should fall when the number of competitors goes up. In fact, looking at the equation for the equilibrium price in
price-taking environments makes it quite difficult to see immediately that a decrease in \( N \) obviously always leads to an increase in price. Fortunately, the result is easier to see if we consider the familiar picture with linear market supply and linear market demand equations (we leave the reader to draw the diagram as an exercise). Reducing \( N \) and having firms exit the market shifts the market supply curve leftward, which will clearly generally result in an increase in equilibrium market price. In contrast, entry will shift the aggregate market supply curve rightwards and, in so doing, reduce equilibrium prices. For those who favor algebra, one can easily calculate the derivative of the equilibrium price with respect to the number of firms \( N \) to see the negative relation between the two in this example.\(^2\)

### 5.1.1.2 Market Structure in a Cournot Setting with Quadratic Costs

Consider next an oligopoly in which firms that entered the market compete in quantities of a homogeneous good, the Cournot model. In this market exit does two things. First, it reduces the number of firms so that total market output tends to be reduced. Second, it increases the amount that any incumbent firm will produce due to the shape of each individual firm’s equilibrium supply function. The net effect on total output, and hence prices, is therefore potentially ambiguous. It depends on the relative effect of an increase in firm output and a decrease in the number of firms. Usually, we expect the impact of losing a firm not to be compensated for by the expansion in output produced as a result by surviving rivals. In that case, price will rise following the exit of an incumbent firm and fall following entry of a new player.

Let aggregate market demand be

\[
Q = S(a - bp),
\]

where \( S \) is the size of the market, so that the corresponding inverse aggregate demand equation is

\[
p(Q) = \frac{a}{b} - \frac{Q}{bS}.
\]

Assuming again a quadratic cost function,

\[
C(q_i) = cq_i + \frac{1}{2}dq_i^2 + F,
\]

and \( N \) profit-maximizing firms that exhibit the following first-order condition for profit maximization:

\[
p(Q) + p'(Q)q_i - C'(q_i) = 0,
\]

where

\[
Q = \sum_{i=1}^{N} q_i.
\]

\(^2\)Doing so allows us to check the conditions required on the parameters \((a, b, c, d)\) to ensure that the linear supply and demand curves cross.
5.1. Framework for Analyzing the Effect of Market Structure on Prices

Solving this equation for $q_i$, the firm’s reaction function is

$$q_i = \frac{S(a - bc) - \sum_{j \neq i} q_j}{2 + bsd},$$

which in fact is identical for each $i = 1, \ldots, N$.

We use the Cournot–Nash equilibrium assumption under symmetry, which allows us to assume that each firm will produce the same amount of output in equilibrium, $q_1 = q_2 = \cdots = q_N = q^*$. The symmetry assumption implies that all $N$ first-order conditions are entirely identical,

$$q^* = \frac{S(a - bc) - (N - 1)q^*}{2 + bsd},$$

and that allows us to solve them all by solving this single equation for $q^*$. A little more algebra allows us to express the equilibrium quantity supplied by each firm as

$$q^* = \frac{S(a - bc)}{1 + N + bsd}.$$

Plugging the resulting aggregate quantity $Nq^*$ in the demand function, we can retrieve the equilibrium market price:

$$p^* = p(Nq^*)$$
$$= p\left(\frac{NS(a - cb)}{1 + N + dbs}\right)$$
$$= \frac{a}{b} - \frac{1}{b} \left(\frac{NS(a - cb)}{1 + N + dbs}\right)$$
$$= \frac{a}{b} - \frac{1}{b} \left(\frac{N(a - cb)}{1 + N + dbs}\right).$$

As with price-taking firms, we see that prices are generally dependent on market structure.

The algebraic relationship between price and the number of firms is not obviously negative. The magnitude of the actual predictions from the model will once again depend on the assumptions about the cost symmetry of firms and the shape of the demand. In the simple case of symmetric firms with decreasing returns to scale and a linear demand, a reduction in the number of firms leads to a reduction in total output and an increase in price.

---

3The first-order condition can be expressed as

$$\frac{a}{b} - \sum_{j \neq i}^{N} \frac{q_j}{bs} - \frac{1}{b} s q_i - c - d q_i = 0 \iff aS - \sum_{j \neq i}^{N} q_j - 2q_i - bSc - bsdq_i = 0$$

from which the expression in the text immediately follows.
5. The Relationship between Market Structure and Price

5.1.1.3 Market Structure in a Differentiated Product Price Competition Setting

As the third of our examples we now consider the case of differentiated products Bertrand competition, in which existing firms in a market produce differentiated products and compete in price for potential customers.

In pricing games where firms produce goods that are substitutes, optimal prices increase in the prices of rivals under fairly weak conditions. That means that if a firm’s rival raises its price, the best response of the firm is to also raise its own price. The reaction functions of two firms producing substitute goods and competing in prices are plotted in figure 5.2.

Assuming that firm 1 produces product 1 at marginal cost $c_1$, the firm’s profit-maximization problem can be expressed as

$$
\max_{p_1} (p_1 - c_1) D_1(p_1, p_2; \theta),
$$

where $D_1(p_1, p_2; \theta)$ is the demand for product 1 and $\theta$ is a consumer taste parameter. The first-order condition for this problem can be written

$$
\frac{\partial \Pi_1^{\text{Single}}}{\partial p_1} = (p_1 - c_1) \frac{\partial D_1(p_1, p_2)}{\partial p_1} + D_1(p_1, p_2) = 0.
$$

Solving this equation allows us to describe firm 1’s reaction function,

$$
p_1^* = R_1(p_2; c_1, \theta),
$$

that is, its optimal choice of price for any given price of firm 2. In a similar way, we could derive the reaction function for firm 2,

$$
p_2^* = R_2(p_1; c_2, \theta).
$$

Figure 5.2. Reaction curves and static Nash equilibrium in a two-firm industry and in a single-firm industry.
5.1. Framework for Analyzing the Effect of Market Structure on Prices

This positive relation between the optimal prices of competing firms selling substitutes is the basis for the unilateral effect described above whereby, after a merged firm increases the prices of the substitutes goods it produces, competitors that produce other substitute goods will follow the price increase, turning this price increase into an all-market phenomenon.

We now show analytically why a merging firm combining the production of two substitutes has the incentive to increase both prices post-merger. This result is derived from the fact that the merged firm can appropriate the profits generated by the increase in the demand of the second substitute good if the price of the first good is increased. This ability to get the profits generated by both goods will result in higher equilibrium prices for both goods, all else equal.

Suppose we have one multiproduct firm which produces both the two goods 1 and 2. Such a multiproduct firm will solve the following profit-maximization problem:

$$\max_{p_1, p_2} (p_1 - c)D_1(p_1, p_2) + (p_2 - c)D_2(p_1, p_2).$$

The first-order conditions for this problem are

$$\frac{\partial \Pi^\text{Multiproduct}}{\partial p_1} = (p_1 - c)\frac{\partial D_1(p_1, p_2)}{\partial p_1} + D_1(p_1, p_2) + (p_2 - c)\frac{\partial D_2(p_1, p_2)}{\partial p_1} = 0$$

and

$$\frac{\partial \Pi^\text{Multiproduct}}{\partial p_2} = (p_1 - c)\frac{\partial D_1(p_1, p_2)}{\partial p_2} + D_2(p_1, p_2) + (p_2 - c)\frac{\partial D_2(p_1, p_2)}{\partial p_2} = 0.$$

One approach to these equations is to calculate the solution \( (p_1^{\text{Multiproduct}}, p_2^{\text{Multiproduct}}) \) by solving the two simultaneous equations and then consider how those prices relate to \( (p_1^{\text{Single}}, p_2^{\text{Single}}) \). We will do that for a very general case in chapter 8. Here, however, we follow a different route. Namely, instead of calculating the equilibrium prices directly, we can instead evaluate the marginal profitability of increasing prices to the multiproduct firm at the prices \( (p_1^{\text{Single}}, p_2^{\text{Single}}) \) that would have been chosen by two single-product firms. Doing so allows us to evaluate whether the multiproduct firm will have an incentive to raise prices. Note that we can write

$$\frac{\partial \Pi^\text{Multiproduct}}{\partial p_1}(p_1^{\text{Single}}, p_2^{\text{Single}}) = 0 + (p_2^{\text{Single}} - c)\frac{\partial D_2(p_1^{\text{Single}}, p_2^{\text{Single}})}{\partial p_1}$$

and

$$\frac{\partial \Pi^\text{Multiproduct}}{\partial p_2}(p_1^{\text{Single}}, p_2^{\text{Single}}) = (p_1^{\text{Single}} - c)\frac{\partial D_1(p_1^{\text{Single}}, p_2^{\text{Single}})}{\partial p_2} + 0.$$
5. The Relationship between Market Structure and Price

since at \( p_i = p_i^\text{Single} \) profits on the single product are maximized and the first-order condition for single-product maximization holds. So,

\[
\text{sign} \left( \frac{\partial \Pi^\text{Multiproduct}(p_1^\text{Single}, p_2^\text{Single})}{\partial p_1} \right) = \text{sign} \left( \frac{\partial D_2(p_1^\text{Single}, p_2^\text{Single})}{\partial p_1} \right)
\]

and

\[
\text{sign} \left( \frac{\partial \Pi^\text{Multiproduct}(p_1^\text{Single}, p_2^\text{Single})}{\partial p_2} \right) = \text{sign} \left( \frac{\partial D_1(p_1^\text{Single}, p_2^\text{Single})}{\partial p_2} \right).
\]

These equations give us an important result, namely that if goods are demand substitutes, so that

\[
\frac{\partial D_1(p_1^\text{Single}, p_2^\text{Single})}{\partial p_2} > 0 \quad \text{and} \quad \frac{\partial D_2(p_1^\text{Single}, p_2^\text{Single})}{\partial p_1} > 0,
\]

then this “two-to-one” merger will very generally result in higher prices for both goods. For example,

\[
\frac{\partial \Pi^\text{Multiproduct}(p_1^\text{Single}, p_2^\text{Single})}{\partial p_1} > 0
\]

means that the multiproduct firm will have higher profits if she raises the price of good 1 above the single-product price.

This incentive to raise prices is what is commonly referred to as the “unilateral” effect, or more accurately, the unilateral incentive by merging firms to raise prices after the merger. This incentive is created by the fact that the merged firm would retain revenues on the consumers switching to the alternative product after a price hike. In contrast we can also conclude that if both goods are demand complements, then prices will usually fall following a merger.

Graphically, we can represent the unilateral effect of a two-to-one merger of firms producing substitute goods (see figure 5.2).

The prices that result from a joint maximization of profits made on goods 1 and 2 are higher than the prices that are obtained when profits are maximized for each one of the products separately whenever goods are substitutes.

Notice, as explained above, that this result will hold if there were other firms in the market producing other products. If the prices \( p_1 \) and \( p_2 \) increase, other firms will also increase the prices of their goods as long as they also have upward-sloping reaction functions with respect to \( p_1 \) and \( p_2 \). This in turn will further cause a further incentive to increase in the prices of \( p_1 \) and \( p_2 \) and so on until the process settles at higher prices for all substitutable products. How much higher the prices are compared with a situation in which there are single-product firms will depend on the concentration and ownership structure in the market, i.e., on which firm(s) produce(s) which products. Generally, a more concentrated ownership structure will lead to higher prices, everything else constant.
5.1. Framework for Analyzing the Effect of Market Structure on Prices

This important prediction will be more closely analyzed in the context of merger simulations and we will formalize this result for a fairly general case in chapter 8. Merger simulation has some disadvantages but it does have the advantage that it allows us to explicitly model the way in which merger effects depend on the shape of demand. By doing so carefully we can reflect both the range of choices that the consumer faces and also the substitution opportunities that exist given the consumer’s taste. Chapter 9 discusses the estimation of different models of demand functions that are useful for merger simulation exercises.

In this section, we have illustrated how the most common theoretical frameworks used to characterize competition predict that market structure and in particular the number of players should be expected to affect the level of prices in the market. In particular, in the case of price competition among substitute products, the prediction of the effect of an increased concentration of ownership on the price level of all competing products is unambiguously that price will rise. The European Commission Merger Regulation explicitly mentions the case when a merger will have a negative effect on competition, and therefore on prices, quantity, or quality, because of the reduction in the competitive pressure that firms may face after the merger.\(^4\)

In particular, the regulation states that:

However, under certain circumstances, concentrations involving the elimination of important competitive constraints that the merging parties had exerted on each other, as well as a reduction of competitive pressure on the remaining competitors, may, even in the absence of a likelihood of coordination between the members of the oligopoly, result in a significant impediment to competition.

In practice, the nature and extent of the resulting price change is an empirical question that needs to be addressed using the facts relevant to each case. Not all mergers will be between firms producing particularly close substitutes and some may even involve mergers between firms producing complements. As a result, the magnitude of the likely impact of market structure on prices must be evaluated. In what follows, we describe several methods to empirically determine the relevance of the relationship between market structure and price in specific cases. Although it will not always be possible to perform such detailed quantitative assessments, these techniques highlight the type of evidence that will be relevant for a unilateral effect case and provide guidance on how to assess market evidence even when less quantitative in nature.

5.1.2 Cross-Sectional Evidence on the Effect of Market Structure

One way to look at the possible relation between market structure and prices is to look at the market outcomes (e.g., prices) in situations where the market structure differs. That is, an intuitive approach to evaluating whether a “three-to-two” merger

\(^4\)EC Merger Regulation, Council Regulation on the control of concentrations between undertakings 2004/1.
will affect prices is to examine a market or set of markets where all three firms compete and then look at another market or set of markets where just two firms compete. By comparing prices across the markets we might hope to see the effect of a move from having three active competitors to having just two active competitors. As we will see, such a method while intuitive does need to be applied with great care in practice since it will involve comparing markets that may be intrinsically different. That said, if we do have data on markets with differing numbers of active suppliers, looking at whether there is a negative correlation between the number of firms and the resulting market prices is likely to be a good starting point for analysis.

5.1.2.1 Using Cross-Sectional Information

Using cross-sectional information can be a good starting point for an empirical assessment of the effect of market structure on prices, provided that one can argue that the different markets that are being compared are at least broadly similar in terms of cost structure and demand. Consider a somewhat extreme but illustrative example. Suppose we want to analyze the effect of the number of bicycle shops on the price of bicycles in Beijing. It is pretty unlikely to be very helpful to use data about the price of bicycles in Stockholm, which has fewer bicycle shops, to address the impact of bicycle shop concentration on bicycle prices. Stockholm would have fewer shops and higher prices than Beijing. Even ignoring the likely massive cross-country differences in regulatory environment, the probably huge differences in tastes, market size, and the likely differences in the cost and quality of the bikes involved, the comparison would be effectively meaningless. No matter how concentrated Beijing’s market became, there is no obvious reason to believe that equilibrium prices would provide a meaningful comparison with Stockholm’s prices for the purposes of evaluating mergers in either Stockholm or Beijing. Even comparing Paris and Amsterdam, where more people favor bicycles as a mean of transportation, may well not be appropriate.

The lesson is that when comparing prices across markets we need to make sure that we are comparing meaningfully similar markets. With that important caveat in mind, there are nonetheless many cases in which cross-market comparisons will be indicative of the actual link between the number of firms competing and the price.

One famous U.S. case in which this method, along with more sophisticated methods, was used involved the proposed merger between Staples and Office Depot. This merger was challenged by the FTC in 1997. The resulting court case was reputedly

---

5 The discussion of FTC v. Staples in this chapter draws heavily on previous discussion in the literature. See, in particular, those involved in the case (Baker 1999; Dalkir and Warren-Boulton 1999) and also Ashenfelter et al. (2006). There is some debate as to the extent of the reliance of the court on the econometric evidence. See Baker (1999) for the view that econometrics played a central role. Others emphasize that the econometrics was supplementary to more traditional documentary evidence and testimony.

the first in the United States in which a substantial amount of econometric analysis was used by the court as evidence. The merging parties sold office supplies through very large shops (hence they are among the set of retailers known as “big box” retailers) and operated as specialist retailers, at least in comparison with a general department store. Their consumers were mostly small and medium size enterprises which are too small to establish direct relations with the original manufacturers as well as individuals. The FTC proposed that the market should be defined as “consumable office supplies sold through office superstores.” Examples of consumable office supplies include paper, staplers, envelopes, and folders. This market definition was somewhat controversial since it (i) excluded durable goods such as computers and printers sold in the same stores since they are “nonconsumable,” (ii) excluded consumable office supplies sold in smaller “mom and pop” stores, in supermarkets, and in general mass merchants such as Walmart (not specialized office superstores). To those skeptical about this market definition, the FTC’s lawyers suggested gently to the judge that “one visit [to an office superstore] would be worth a thousand affidavits.” Since we have considered extensively the process of getting to market definition in an earlier chapter, we will leave the discussion of market definition and instead focus on the empirical evidence that was presented. While some of the empirical evidence is relevant to market definition, its focus was primarily on measuring the competitive pricing effects of a merger. The geographical market was deemed to be at the Metropolitan Statistical Area (MSA) level, which is a relatively local market consisting of a collection of counties.8

By 1996, there were only three main players on the market: Staples, with a $4 billion revenue of which $2 billion was in office supplies and 550 stores in 28 states; Office Depot, with a $6.1 billion revenue of which $3 billion was in office supplies and 500 stores in 38 states; Office Max, with a $3.2 billion revenue of which $1.3 billion was in office supplies and 575 stores in 48 states. The merger far exceeded the threshold for scrutiny in the United States in terms of HHI and market shares, at least given the market definition.

The FTC undertook to compare the prices across local markets across the United States at a given point in time to see whether there was a relationship between the number of suppliers present in the market and the prices being charged. They used three different data sources for this exercise. The first data set came from internal documents, particularly Staples’s “1996 Strategy Update.” The second data set contained prices at the SKU (product) level for all suppliers. The last data set

---

7 The evidence suggests Judge Hogan did indeed drive around visiting different types of stores such as Walmart, electronics superstores, and other general supplies stores. He concluded that “you certainly know an office superstore when you see one” and accepted the market of office supplies sold in office superstores as a relevant “submarket.” See Staples, 970 F. Supp. at 1079 also cited in Baker and Piotofsky (2007).

8 Some MSAs are nonetheless quite large. For example, the Houston Texas MSA is about 150 miles (around 240 km) across.
was a survey with a comparison of average prices for a basket of goods as well as specific comparisons for given products.

The first set of cross-market comparisons came from the parties’ internal strategy documents. The advantage of internal strategy documents that predate the merger is that they consist of data produced during the normal course of business and, in particular, not as evidence “developed” to help smooth the process of approval of the merger being considered. If the firm needs the information in a particular document to be reliable because it intends to make decisions involving large amounts of money by using them, then it will usually be appropriate to give such documents considerable evidential weight. In particular, such documents should probably receive far more weight as evidence than protestations given during the course of a merger inquiry, where there can be a clear incentive to present the case in a particular light. In this case, the internal strategy documents provided an informal cross-market comparison of prices by market structure. The results are presented in table 5.1 and suggest that when markets with only Staples in are compared with markets with Staples and Office Depot stores in, then prices are 11.6% lower in the less concentrated market.

In addition to the internal documents, the FTC also examined advertised prices from local newspapers in order to develop price comparisons across markets. In particular, the FTC performed a comparison of Office Depot’s advertised prices using the cover page of a January 1997 local Sunday paper supplement. In doing so the FTC tried to choose two markets which provided an appropriate comparison. Ideally, such markets will be identical except for the fact that one market is concentrated while the other is less concentrated. In some regards it is easy to find “similar” markets; for instance, we can fairly easily find markets of similar population to compare. However, at the front of our minds in such an exercise is the concern that if two markets are identical, then why do we see such different market structures? With that caveat firmly in mind, the results are provided in table 5.2 and show considerably higher prices in the market where there is no competition from other office supply superstores.

Table 5.1.  Informal internal across-market price comparison.

<table>
<thead>
<tr>
<th>Benchmark market structure</th>
<th>Comparison: OSS market structure</th>
<th>Price reduction</th>
</tr>
</thead>
<tbody>
<tr>
<td>Staples only</td>
<td>Staples + Office Depot</td>
<td>11.6%</td>
</tr>
<tr>
<td>Staples + Office Max</td>
<td>Staples + Office Max + Office Depot</td>
<td>4.9%</td>
</tr>
<tr>
<td>Office Depot only</td>
<td>Office Depot + Staples</td>
<td>8.6%</td>
</tr>
<tr>
<td>Office Depot + Office Max</td>
<td>Office Depot + Office Max + Staples</td>
<td>2.5%</td>
</tr>
</tbody>
</table>

5.1. Framework for Analyzing the Effect of Market Structure on Prices

Table 5.2. Price comparison across markets.

<table>
<thead>
<tr>
<th>Product</th>
<th>Orlando, FL (three firms)</th>
<th>Leesburg, FL (Depot only)</th>
<th>Percentage difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Copy paper</td>
<td>$17.99</td>
<td>$24.99</td>
<td>39%</td>
</tr>
<tr>
<td>Envelopes</td>
<td>$2.79</td>
<td>$4.79</td>
<td>72%</td>
</tr>
<tr>
<td>Binders</td>
<td>$1.72</td>
<td>$2.99</td>
<td>74%</td>
</tr>
<tr>
<td>File folders</td>
<td>$1.95</td>
<td>$4.17</td>
<td>114%</td>
</tr>
<tr>
<td>Uniball pens</td>
<td>$5.75</td>
<td>$7.49</td>
<td>30%</td>
</tr>
</tbody>
</table>

Source: Figure 2 in plaintiff’s “Memorandum of points and authorities in support of motions for temporary restraining order and preliminary injunction.” Public brief available at www.ftc.gov/os/1997/04/index.shtml.

5.1.2.2 Comparing Price Levels of Multiple Products across Markets

Whenever an authority compares prices across multiproduct retailers the investigator immediately runs into the problem of determining which prices should be compared. If there are thousands of products being compared, it is important that parties to the merger evaluation do not have the flexibility to pick the most favorable comparisons and ignore the rest. In this section we consider the element of the studies which explicitly recognized the multiproduct nature of the cross-market pricing comparisons.

The third cross-market study in the Staples case used a Prudential Securities pricing survey which compared prices in Totawa, New Jersey (a market with three players), with prices in Paramus, New Jersey (a market with two players). Since it was difficult to compare prices of 5,000 with 7,000 items, it built a basket of general office supplies that included the most visible items on which superstores usually offer attractive prices. It found that on the “most visible” items, prices were 5.8% lower in the three-player market than in the two-player market.

When comparing price levels across retailers or across multiproduct firms, one is always faced with the problem of trying to measure a price level relating to many products, often thousands of products. Sometimes, the different firms or suppliers will not offer the same products exactly or the same combination of products so that the comparison is not straightforward. A possible solution is indeed to construct a basket of products for which a price index can be calculated. A famous example of a price index is the Stone price index, named after Sir Richard Stone, which can be calculated for a single store $s$ using the formula

$$\ln P_{st} = \sum_{j=1}^{J} w_{jst} \ln p_{jst},$$

where $w_{jst}$ is the expenditure share and $p_{jst}$ is the price of product $j$ in store $s$ at time $t$. This formula gives a price index for each store and its value will depend on
5. The Relationship between Market Structure and Price

the product mix sold in that particular store. For the purpose of comparing prices across stores, we may therefore prefer to use an index where the weights do not depend on the store-specific product mix, but rather depend on the general share of expenditure within a market, such as

$$\ln P_{st} = \sum_{j=1}^{J} w_{jt} \ln p_{jst},$$

where $w_{jt}$ is the expenditure share of product $j$ in the market rather than at the particular store. Naturally, there is a great deal of scope for arguments with parties about the “right” price index.\(^9\) One could, for instance, reasonably argue for keeping the composition of the basket constant over time as price increases might make people switch to cheaper products. In such a case, the price index would not capture all price increases and would also not necessarily reveal the loss in quality. In the FTC v. Staples case, the FTC reportedly solved the choice of index by choosing one which the opposing side’s expert witness had himself proposed, thereby making it rather difficult to critique the choice of index too much. Such a strategically motivated choice may not always be available and, even if it were, may not be desirable since there is quite an extensive literature on price indices, not all of which are equally valid in all circumstances.

Discussions about the “right” price index to use can appear esoteric to nonspecialists and therefore a general rule is probably to check that conclusions are robust by exploring the data using a few different indices. Doing so will also have the advantage of helping the investigator understand the patterns in the data if she reflects carefully on any substantive differences that arise.

To construct price indices that are representative, extensive data are needed covering a large range of products and suppliers. Price data can be obtained through a direct survey by the investigators as long as the suppliers are unaware of the action, or the investigatory authority is clear there are no incentives to strategically manipulate observed prices. Alternatively, one could solicit internal company documents that may provide own-price listings of products at different points of time in different stores or markets. Firms do tend to have documents (and databases) with comprehensive list prices. Unfortunately, in some industries, list prices are only weakly related to actual prices once rebates and discounts are taken into account. If such discounts are important in the industry, it is usually advisable to take them into account when calculating the final net price. Allocating rebates to the sales can be a challenging exercise and one should not hesitate to ask companies for the data and clarification as to what rebates apply to which sales. Sometimes, the quality of the data will determine the level of minimum aggregation possible with respect to the products and the time unit used. Finally, one should also inquire about internal

---

\(^9\) For a review of the price index literature, see, for example, Triplett (1992) and also Konüs (1939), Frisch (1936), and Diewert (1976). For a recent contribution, see Pakes (2003).
5.1. Framework for Analyzing the Effect of Market Structure on Prices

documents on market monitoring as very often those will reveal relevant information about competitors’ observed behavior.

Unless our price data come from internal computer records generated ultimately from the point of sale, the investigative team is unlikely to have either quantity or expenditure data. Unfortunately, such data are often important for price comparisons—either for computing price indices explicitly or more generally helping to provide the investigators with appropriate weighting to evidence about particular price differences. If a price comparison suggests a problem but the prices involve goods which account for 0.00001% of store sales, probably not too much weight should be given to that single piece of evidence taken alone. On the other hand, it may be possible to examine the prices associated with a relatively small numbers of goods whose sales are known to account for a large fraction of sales.

In 2000, the U.K. Competition Commission10 (CC) undertook a study of the supermarket sector.11 Several data sources were used to compare the prices of specific products and of a basket of products across chains and stores. To construct the basket, the CC asked the twenty-four multiple grocery retailers such as Tesco, Asda, Sainsbury’s, Morrisons, Aldi, M&S, and Budgens for details of prices charged for 200 products in 50–60 stores for each company on one particular day before the start of the inquiry: Thursday, January 28, 1999. The basket was constructed using 100 products from the top 1,000 sales lines, picking “well-known” products across each category and 100 products chosen at random from the next 7,000 products “although the choices were then adjusted as necessary to reflect the range of reference product categories.”12 The main difficulty was comparability: finding “similar” products sold across all supermarket chains. The CC also asked for sales revenue data for each product in order to construct sales-weighted price indices.

The inquiry also used internal company documents in which firms monitored the price of competitors. Aldi, for instance, had daily price checks on major competitors as well as weekly, monthly, and quarterly reports on prices of certain goods for selected competitors and across the whole range in discounters. Asda had three different weekly or monthly price surveys of competitors.13 The aim of collecting all these data was to compare prices across local markets with different market structures. To accomplish this, the CC’s economics staff plotted all the stores on a map and visually selected 50–60 stores that faced either “intense,” “medium,” or “small” amounts of local competition. This appears to be a pragmatic if slightly ad hoc approach with the advantage that the method did generate cross-sectional variation. Recent developments in software for geographic positioning (known as geographic information systems) greatly facilitate characterizing local competition.

10 In its previous guise as the U.K. Monopolies and Mergers Commission.
12 See paragraph 2 in appendix 7.6 of the CC’s supermarket final report.
13 See appendix 7.4 of the CC’s supermarket inquiry report.
5. The Relationship between Market Structure and Price

As always in empirical analysis, getting the right data is a first important step. With very high-quality data on a relevant sample, simple exercises such as the cross-sectional comparisons can be truly revealing. In the FTC v. Staples office supplies case, all the results from the cross-sectional comparison pointed to a detrimental effect of concentration on prices. Markets with three suppliers are cheaper than markets with two suppliers, which are in turn cheaper than markets with a single supplier. This was supported by the comparison across market using different data sources. The evidence was enough to indicate that a merger might be problematic in terms of prices to the final consumer.

Still, although local markets in the United States (and particularly neighboring markets such as those used for many of the comparisons) are probably close enough for the comparisons to make sense, the merging parties still claimed that price differences were due to cost differences in the different areas and in particular that price differences were not caused by the lack of additional competitors. The strength of any evidence needs to be evaluated and the “cost difference” critique suggests that the cross-market correlation between market structure and prices may be real but the explanation for the correlation may not be market power. To address this potentially valid critique, the FTC undertook further econometric analysis to take account of possible market differences, and it is to that we now turn.

5.1.2.3 Endogeneity Problems in Cross-Sectional Analysis

Results obtained from a simple cross-sectional comparison across markets with different market structures are informative provided the comparisons involved are sensible. However, such studies will rarely be entirely conclusive by themselves since they are vulnerable to the criticism that, although there might be a link between market structure and price, this link is not causal. For example, if two markets have in truth different costs, then we will tend to see both fewer stores and higher prices in the high cost market. In such a situation an investigator could easily and erroneously conclude that a merger to increase concentration would increase prices. Such a situation is of particular difficulty since costs are often difficult to observe and provides yet another example of an “endogeneity bias.”

To summarize the problem consider a regression equation attempting to explain prices as a function of market structure:

\[ p_m = \alpha + N_m \gamma + \varepsilon_m, \]

where \( p_m \) is the price in market \( m \) and \( N_m \) is the number of firms in market \( m \). Suppose that the true data-generating process (DGP) is very closely related:

\[ p_m = \alpha + N_m \gamma^{\text{true}} + u_m, \]

with the determinants of prices other than “market structure,” \( N_m \), captured in the unobserved component, \( u_m \). For instance, costs will affect prices but are not explicitly controlled for, so their effect is a component in the error term. If high costs
5.1. Framework for Analyzing the Effect of Market Structure on Prices

cause high $u_m$ and therefore high prices as well as low entry (low $N_m$), then we have $E[u_mN_m] < 0$, i.e., the “random” term in the equation will not be independent of the explanatory variable. This violates a basic condition for getting unbiased estimates of the regression parameters using our standard technique of OLS (see chapter 2). We will find that markets with fewer firms will be associated with higher prices, but the true cause of the high prices is not the market structure but rather the higher costs. One must therefore beware “false positives” when using across-market data variation to identify the relationship between market structure and prices. False positives are possible when there is a factor such as high cost that will positively affect prices and that will also independently negatively affect entry and the number of firms. If this happens, we will find a negative correlation between price and market structure that is due to variation in costs (or other variable) and not to differences in pricing power.

False negatives can also occur when using across-market data variation. This happens when there is an omitted factor that increases both prices and the number of firms in the market. For instance, a high demand for reasons we do not see (e.g., demographics, tastes) will result in high prices and also in a large number of firms. In this case, we will tend to find a positive correlation between price and number of suppliers that is due to variation in demands across markets. Again such a positive correlation is not down to differences in pricing power, but may act to make pricing power more difficult to identify. Specifically, we may find no correlation at all when there is in fact a negative correlation due to pricing power. This is because the “endogeneity” bias now acts to bias our estimate of $\gamma^{\text{true}}$ upward—toward zero or even above zero.

The endogeneity bias in the cross-sectional comparisons of markets with different structures ultimately occurs when there is a component that we do not account for that affects both prices and the number of firms or in other words it affects both prices and entry.

To illustrate where the endogeneity concern comes from using a theoretical model, consider the equilibrium price in a Cournot model with quadratic costs such as described above:

$$p_m = \frac{a_m}{b} - \frac{1}{b} \left( \frac{N_m(a_m - c_m b)}{1 + N_m + dbS_m} \right),$$

where $S$ is the size of the market, $a$ and $b$ are demand parameters, and $c$ and $d$ are the cost parameters. The demand and costs parameters are unobserved and their effect is therefore included in the error term of the pricing regression. In this model, if we use the free entry assumption to solve for the equilibrium number of firms $N$, we get

$$N^*_m = \frac{a_m - c_m b}{2} \sqrt{\frac{2S_m(2 + dbS_m)}{bF} - 1 - dbS_m}.$$
And the point to note is that both $p$ and $N$ are correlated with both demand and costs. Thus the unobserved components of both demand and costs will both emerge in the pricing equation’s residual and also be a determinant of the number of firms, $N$.

Sometimes, analysts will be able to convincingly argue that endogeneity is not an issue. Often, it will be advisable to try to control for it. In the following section we illustrate one way of attempting to do so.

### 5.1.3 Using Changes over Time: Fixed-Effects Techniques

Fixed-effects techniques were introduced in chapter 2 and are closely related to the natural experiment techniques discussed in chapter 4.\textsuperscript{14} In both cases, one observes how the outcome of interest (for example price) for similar observations changes over time following changes in the explanatory variable for only some but not all the observations, thereby identifying the effect of that explanatory variable on the outcome of interest. The great advantage of these techniques is that we do not need to control for all the remaining explanatory variables that are assumed to remain constant. Fixed effects are also technically very simple to implement. When used properly, fixed effects are a powerful empirical method that provides solid evidence. But as in many empirical exercises, the ability to produce regression results with easy-to-use software can mean that the technique appears deceptively simple. In reality, the investigator must make sure that the conditions necessary for the validity of the method are satisfied. In this section we discuss fixed effects and highlight when this very appealing technique may be properly used and when, on the contrary, one must be wary of applying it.

#### 5.1.3.1 Fixed Effects as a Solution for Endogeneity Bias

To identify the effect of market structure on the level of prices one must control for each of the determinants of price and obtain the pure effect of the number of competitors on price. The difficulties are both that the number of variables that one needs to control for may be large and that at least some of the variables (particularly cost data) are likely to be difficult to observe. Comprehensive data are therefore unlikely to be available. One way to proceed in the face of this issue is to choose a reasonably homogeneous subset of observations and look at the effect of the change in market structure on that subset. For example, we may look over time at the effect of a change in market structure affecting the price at a particular store. Such an approach uses “within-store” and “across-time” data variation. This kind of data variation is very different from the across-store or across-market data variation used in the previous section to identify the relationship between prices and the number

\textsuperscript{14}The econometric analysis of fixed-effects estimators and other techniques for panel data are widely discussed in the literature. For example, readers may wish to consult Greene (2007), Baltagi (2001), or Hsiao (2003).
5.1. Framework for Analyzing the Effect of Market Structure on Prices

of stores. If we have just one store, we could use the data variation from that one store and the only data variation would be “within store across time.” However, if we have many stores observed over time, then we can combine the cross-sectional information with the time series information that we have for each store. Data that track a particular sample (of firms, individuals, or stores) over time are referred to as panel data. Panel data sometimes offer good opportunities for identification because we can use either cross-sectional or a cross-time data variation to identify the effect of market structure on prices. A panel data regression model for prices can be written

\[ p_{st} = \alpha_s + x_{st} \beta + \varepsilon_{st}, \]

where \( s \) indicates the cross-sectional index (here, the store) and \( t \) indicates the time period so that the price \( p_{st} \) is store-time specific as are the explanatory variables, \( x_{st} \). Allowing for a store fixed effect \( \alpha_s \) in the regression controls for a particular price level to be associated with each store. By introducing this store-specific constant and looking at the effect of a change of structure (i.e., a variable in \( x_{st} \)) on that store, we control for all store-specific time-invariant store characteristics. For example, if our data are fairly high frequency and costs change slowly, then the store’s cost structure may be sufficiently constant across time for this to be a reasonable approximation. Similarly, the fixed effect may successfully control for the impact of store characteristics such as a particularly good location persistently affecting demand and hence prices. Controlling for these unobserved characteristics by using the store fixed effect will help address the concern we highlighted with the cross-sectional evidence, that, for example, the costs in a particular location are high and this is therefore associated with both high prices and low entry. Thus store fixed effects may help alleviate “endogeneity bias.” Such an approach to alleviate endogeneity is often used when the researcher has panel data.\(^{15}\) Of course, one still needs to account for time-varying effects but permanent structural differences across stores are at least accounted for.

To be clear, the fixed-effects technique will only work to the extent that there is not any substantial time-varying change in demand or costs within stores that affect both the number of local stores and prices. If there are, then the fixed-effects approach may not help solve the problems associated with endogeneity bias.

To illustrate this method let us return to our discussion of the *FTC v. Staples/Office Depot* case. In that case, the FTC had product level data from 428 Staples stores in 42 cities for 23 months available. To make the data set manageable, a monthly price index was constructed for each store, based on a basket of goods. The FTC proposed the following fixed-effects regression:

\[ p_{smt} = \alpha_s + x_{smt} \beta + \varepsilon_{smt}, \]

where as before \( s \) indicates store, \( t \) indicates the time period, \( m \) indicates market or city, \( p \) is the price variable, and \( x \), in this instance, is a set of dummy indicators for

\(^{15}\) For a review of the history of panel data econometrics, see Nerlove (2002). (See, in particular, chapter 1 of that book, entitled “The history of panel data econometrics, 1861–1997.”)
the presence of nearby stores such as an Office Depot within five miles (OD_{5\text{miles}}^{smt}) or the presence of a local Walmart or other potentially relevant competitor stores. The latter coefficients turned out to be insignificant so we will focus on the effect of the Office Depot store. Note that the regression has a store-specific fixed effect $\alpha_s$, which means that the changes in the $x$ variables are considered “holding the store effect constant.” Specifically, if a single store experiences nearby entry, we will see that either its price drops or it does not. For those stores which experience no change in prices over time, the store fixed effect will absorb all of the variation in prices and so that variation will not be used to help identify the value of the parameters in $\beta$. That is, in contrast to the cross-sectional data variation, the store fixed-effects regression uses primarily the “within-store” data variation, albeit using the within-store data variation across the whole sample (see also the discussion on this point in chapter 2).

The fixed-effects regression was meaningful in this case because there was enough informative variation in the data. Prices varied across time and across stores but it is notable that they varied more across stores than across time. Since the store fixed effects will account for all the time-invariant variation across stores, only the relatively small amount of within-store data variation may be left once the fixed effects are allowed for. Fortunately, there was some variation across time within a store in prices and also in the presence of competitors in some of the stores’ market. Enough stores experienced entry by nearby rival stores to ensure that it was possible to identify the effect of that change in market structure on prices.

The effect of the presence of a competitor (i.e., an Office Depot store) on Staples’s prices can be calculated using the expression:

$$
100 \frac{\hat{p}_{smt}(\text{OD}_{5\text{miles}}^{smt} = 1) - \hat{p}_{smt}(\text{OD}_{5\text{miles}}^{smt} = 0)}{\hat{p}_{smt}(\text{OD}_{5\text{miles}}^{smt} = 0)} = 100 \frac{\hat{\beta}_{OD}}{\hat{p}_{smt}(\text{OD}_{5\text{miles}}^{smt} = 0)},
$$

where $\hat{p}_{smt}(\text{OD}_{5\text{miles}}^{smt} = 1)$ denotes the predicted price level at store $s$ in market $m$ at time $t$ when the $x$ variable associated with the indicator for whether there is an Office Depot within five miles takes on the value 1 and $\hat{p}_{smt}(\text{OD}_{5\text{miles}}^{smt} = 0)$ is defined analogously. This expression provides the predicted percentage decrease in prices at a Staples store which results from having an Office Depot within five miles, all else equal.

The defendants’ expert found only a 1% effect of the presence of an office supply superstore on the price and claimed that the difference with the cross-sectional results was due to the endogeneity bias caused by comparing stores in different markets. He argued that the difference between the cross-sectional and fixed-effects estimates arose because the panel data estimates controlled for store-specific costs that were not observed directly and hence not controlled for in either the cross-sectional regression or the panel data regression unless fixed effects were included. However, in the event Baker (1999) argues there were several problems

---

16 Specifically, $100 \times \hat{\beta}_{OD} / \hat{p}_{smt}(\text{OD}_{5\text{miles}}^{smt} = 0) = 1\%$. 

5.1. Framework for Analyzing the Effect of Market Structure on Prices

with the defendant’s expert regression. First, the FTC view was that the expert had somewhat arbitrarily drawn circles around stores at 5 miles, 10 miles, and 20 miles and constructed dummies for the presence of stores within that range. The FTC argued that internal documents suggested that companies priced according to pricing zones that were not circles and could sometimes be quite large and as large as the MSA area. While generally an approach of drawing circles around stores would seem a highly plausible way to proceed, the regression aims to capture the data-generating process for prices. Here the documents reveal the nature of competitive interaction and so the specification should be guided by the documentary evidence. Including the count of stores within the MSA tripled the price effect to a range of about 2.5–3.7%. Thus the FTC argued that the merging parties' preferred results were (1) not robust to slight changes in specification and (2) did not reflect the documentary evidence. In addition, Baker (1999) reports that the defendant's expert had dropped from their sample observations from California, Pennsylvania, and a few others for reasons that were not entirely clear. When included back in the data set, the effect was estimated to be three times larger again, between 6.5% and 8.6% depending on the detail of the specification. Thus in sum, the FTC expert concluded that a reasonable estimate was that prices of Staples stores were on average 7.6% lower when an Office Depot store was in the MSA, which was also consistent with their findings using only cross-sectional data variation.

5.1.3.2 Limitations of Fixed Effects

Fixed-effect regressions attempt to control for the bias generated by the presence of endogeneity or omitted explanatory variables. These problems can be potentially severe in cross-sectional comparisons and the use of panel data provides an opportunity to at least partially address the endogeneity problem. Fixed-effect regressions control for firm- (or store-) specific characteristics and compute the effect of a change in the variable of interest for a particular firm (or store) only. However, because we force the effect to be measured only within firm (or store), we are, albeit deliberately, no longer fully exploiting the cross-sectional variation.

Suppose, for instance, that there is very little variation in market structure over time, i.e., no entry or exit, and we estimate a specification which includes in $x$ a count of the number of nearby stores. When we estimate

$$p_{st} = \alpha_s + x_{st}\beta + \varepsilon_{st},$$

we will estimate $\beta = 0$ because the store fixed effect will explain all the observed variation in prices and there will be no additional variation in the data allowing us to tell apart the store-specific fixed effect and the effect of local market structure, which did not change for any given store. In an extreme case, when there is literally no time series variation in market structure, our regression package will either fail or else tend to print out estimates of standard errors which involve very large numbers.
The reason is that we have tried to estimate a model which is simply not identified unless there is time series variation in the local market structure variables. It is very important to realize that such a finding does not necessarily mean the variation in prices across markets is not at least partly caused by the variation in market structure. In our office supplies example, stores with a competitor nearby may have lower prices and this is just showing up in the difference in the level of the store fixed effect \( \alpha_i \). Cross-sectional variation may be explained by omitted variables but it might also be due to lack of competition near some stores. Thus, it may be appropriate to consider fixed-effects estimates as low-end estimates when most of the data variation is cross sectional.

In sum, the fixed-effects regression identifies the coefficients in \( \beta \) by using the variation in the data within a group of observations, for example, across time for a given store as well as the across-store variation to the extent that the specification restricts the slope coefficients to be the same across stores (see the extensive discussion in chapter 2). If there is not enough within-store data variation, the regression will be uninformative about slope parameters. In fact, if most of the variation in the data is across groups because little changes within the groups, then by using the fixed effects we effectively lose all the information in the data into the fixed effects. One lesson is that analysts must be aware of the source of information, i.e., the source of the variation in the data set, when choosing the appropriate econometric technique. Fixed-effect estimators will help correct an endogeneity problem but to do so there must be sufficient within-group variation in the data. A second lesson is that fixed-effects estimators can be used to test cross-sectional evidence but the results must be interpreted carefully—a concern raised by a cross-sectional relationship between market structure and price may not be allayed by a finding that the relationship does not survive to the fixed-effects model. A mistaken belief that is the case can mean that a case handler erroneously finds there is no problem with a merger when in fact it is just that there is very little identifying variation in the explanatory variables in her data set.

### 5.1.4 Using Time and Cross-Sectional Variation

When the variation in the data is mostly cross sectional, fixed-effects techniques that follow a store or a firm over time may not be very informative. Moreover, we have argued that it may be a mistake to take out all of the cross-sectional variation in the data when evaluating the effect of a merger by introducing the fixed effects. We may be controlling for endogeneity but in doing so we might be taking out much of the effect of interest. As a result it will sometimes be useful to revisit cross-sectional data variation. To do so, we can use our panel data set but carefully choose the technique in order to ensure that we use the cross-sectional variation in the data to identify the effect of market structure on prices appropriately.
5.1. Framework for Analyzing the Effect of Market Structure on Prices

5.1.4.1 Explaining the Variation in the Data

One approach is to break up the price variance in the sample into a part which varies over time, a part which is firm or store specific, and an idiosyncratic part particular to a time and firm or store. To do this we can run the following regression:

\[ p_{st} = \alpha_s + \tau_t + \varepsilon_{st}, \]

where \( \alpha_s \) is the store-\( s \)-specific effect, \( \tau_t \) is the time-\( t \)-specific component, and \( \varepsilon_{st} \) is the store- and time-specific component for each observation. We can then run the store-specific effect on a set of regressors, including measures of rivalry from competitors:

\[ \hat{\alpha}_s = x_s \beta + u_s. \]

This method will have the merit of exploiting all the variation in the data but if we omit variables that are linked with the structure of competition as well as with the price, then we will still have an endogeneity problem just as in the cross-sectional analysis. Specifically, if the covariance between the regressors \( x_s \) and the error term \( u_s \) is not zero, then OLS estimators will be biased. Assuming we have an endogeneity problem only in one variable, then the sign of the endogeneity bias will be the sign of covariance between the regressor and the error terms. Instrumental variable techniques can help alleviate such biases.

5.1.4.2 Moulton Bias

Regression analysis typically assumes that every observation in the sample is independent and identically distributed (i.i.d.). This means that observations in the sample \( (Y_i, X_i) \) are independent draws from the population of possible outcomes. If we use panel data of a cross section over time and there is little change in the variables over time, then the observations are not really independent but are in fact closely related. If so, then we are doing something close to drawing the same observation in each time period. For example, suppose we have monthly data for twenty stores over two years but that during those two years very little changes in terms of the structure of competition and prices. The regression assumes we have 24 \( \times \) 20 = 480 different independent observations but in fact it is closer to the truth to say that we only have twenty independent observations since there is barely any variation over time and the information in the data mostly comes from the cross-sectional variation across the twenty stores. Our 480 pairs \( (p, N) \), where \( p \) is price and \( N \) is the number of competitors, are not i.i.d. The consequence is that the standard errors computed by the standard formula in a regression package will underestimate the true value of uncertainty associated with our estimates, i.e., the precision of the estimated effect will be overstated. As a result, we are more likely to find an effect when there is in reality not enough information to establish one. Correcting this problem involves
modeling the error structure to account for the correlation across observations. Alternatively, the technique described above in which we computed the predicted cross-sectional variation in the outcome variable (prices in our example) and related it to possible determinant of prices including the variable of interest (number of competitors in our example) provides a way of ensuring that standard errors are computed based on the relevant number of independent observations.

5.1.5 Summary of Good Practice

The above discussion has we hope provided a focused discussion of the challenges of identifying price-concentration relationships. Along the way the discussion has illustrated some important elements of good practice when attempting to use empirical techniques to identify the effect of one variable on another. Because those good practices are very important to ensure the quality of the results, we proceed to summarize them.

Collect meaningful data. From the beginning of the investigation, it is important to gather data on the relevant variables for a representative sample. One should not hesitate to contrast data from different sources and check whether other evidence such as that coming from company documents does fit the picture that emerges from the empirical analysis.

Check that there is enough variation in the data to identify an effect. Empirical work will only be as good as the data used. If there is not a lot of variation in the variable of interest in the sample that we examine, it will be very difficult to determine the effect of this variable on any outcome. Variation can be cross sectional or across time and it can be explainable or idiosyncratic. Giving considerable thought to the process that is generating the data, i.e., thinking about the determinants of the observed outcome, will be vital both in terms of understanding the data and also in determining the best econometric methodology to use.

Beware of endogeneity. Once it is established that there is enough variation in the data to estimate an effect, one must be able to argue a causal link between the variable of interest and the outcome. In order to do this, it is important to make sure that all other important determinants of the outcome that could bias the coefficient of the variable of interest are controlled for. If they cannot be controlled for, other methods of identification should be tried or else one must explain why endogeneity is not likely to be a problem. Often it will be possible to sign the expected bias emerging from a particular estimation technique. When we change the estimation technique to control for endogeneity, our estimation results should change in the expected direction.

\(^{17}\) See Kloek (1981) and Moulton (1986, 1990). In practice, statistical packages have options to help correct for Moulton bias. For example, STATA has the option “cluster” to its “regress” command. For a more technical discussion of Moulton bias, see, for example, Cameron and Trevedi (2005).
Perform robustness analysis. Once a regression is run, it is important to make sure that the resulting coefficients are relatively robust to reasonable changes in the specification. For example, results should not be crucially dependent on the exact composition of the sample, except perhaps in deliberate or well-understood ways. They should also not depend on a particular way of measuring the explanatory variables unless we know for sure that it is exactly the correct way to measure them. In general, good results are robust and show up to a higher or lesser extent across many sensible regression specifications.

Use more than one method. One good way to generate confidence in the results of empirical analysis is to use more than one method and show that they all tend toward the same conclusions. If different methods produce divergent results, one should have a convincing explanation of why this happens.

Do not treat econometric evidence as “separate” from the investigation. First, no single source of evidence is likely to be entirely compelling and generally econometric evidence in particular runs the risk of being treated skeptically by judges who are extremely unlikely to be expert econometricians. That risk increases when the results are presented as some form of a mysterious “black box” analysis. Always look for graphs that can be drawn to illustrate the data variation generating the econometric results. Second, when econometric analysis proceeds in a vacuum, disconnected from the rest of the case team and hence the facts of the case, the results are unlikely to capture the core elements of the data-generating process and, as a result, the analysis is fairly unlikely to be either particularly helpful or robust.

In our case study, the FTC’s evaluation of the merger of office supplies superstores, the FTC did manage to produce convincing evidence that the number of players and the prices were negatively related. The summary of their findings is presented in table 5.3. This table presents as convincing a case that the merger between the two superstores will increase prices by more than 5% as is likely to arise in practical case settings. The results are consistent and robust and therefore easy for a nonspecialist judge to accept as credible. In fact, on June 30, 1997, the FTC got a federal district court judge to grant a preliminary injunction blocking the proposed merger between Staples and Office Depot. Subsequently, the parties gave up on their merger plans. That sounds like good news for empirical work in antitrust. However, before coming to that view it is very important for all to realize that such activity probably cannot become the benchmark for the level of evidence required by antitrust authorities in all but the most important cases. The fact that the analysis in Staples took two expert witnesses and about six Ph.D. economists to undertake means it is resource intensive. While the first time will always be harder than the second and third times, the decision...
5. The Relationship between Market Structure and Price

Table 5.3. Estimating merger effects using different sources.

<table>
<thead>
<tr>
<th>Forecasting method</th>
<th>Estimated price increase from merger</th>
</tr>
</thead>
<tbody>
<tr>
<td>Noneconometric forecast: internal strategy documents</td>
<td>5–10%</td>
</tr>
<tr>
<td>Estimate from simple comparisons of average price levels in cities where Staples does/does not compete with Office Depot</td>
<td>9%</td>
</tr>
<tr>
<td>Cross-section, controlling for the presence of nonsuperstore retailers</td>
<td>7.1%</td>
</tr>
<tr>
<td>Fixed effects, with nonsuperstore retailers in (California and rest of United States)</td>
<td>7.6%</td>
</tr>
<tr>
<td>Weighted average of two regional estimates</td>
<td>9.8%</td>
</tr>
</tbody>
</table>

*Source:* FTC results from a variety of specifications as reported in Baker (1999).

In the more recent Ryanair and Aer Lingus case (which provides a European example of such analysis) runs to more than five hundred pages of careful analysis.

5.2 Entry, Exit, and Pricing Power

In the previous section, we discussed some techniques for determining the impact that market structure has on the level of prices. A great deal of our discussion revolved around the problem of endogeneity, or the fact that the number of firms is potentially not exogenously determined but rather is determined in part by the expected profits that firms think they would make if they enter, and this in turn may be related to prices. Cost and demand factors may simultaneously affect both prices and structure. In simple economic models of the world where entry is assumed relatively unfettered by barriers, high profits will attract entry, which in turn induces higher market output and lower prices. If entry is relatively free, we will expect the process of competition to work, driving prices down to the great benefit of consumers. That said, there is a variety of sources of barriers to entry. Some entry barriers are natural—you cannot enter the gold-mining business unless you have access to gold deposits. Some entry barriers are regulatory—you cannot enter the market for prescribing drugs without the requisite qualifications. On the other hand, oligopolistic firms may strategically

---

18 Case no. Comp/M.4439. This decision is available at http://ec.europa.eu/comm/competition/mergers/cases/decisions/m4439_20070627_20610_en.pdf. See, in particular, Annex IV: Regression analysis technical report.
19 Of course, such regulatory barriers may aim to solve another problem. Free entry into prescription writing may reduce the costs of getting a prescription for a patient, but one might worry both about the suitability of the resulting prescriptions and the total costs of prescribing if the patient’s cost of drugs is subsidized by a national health care system.
5.2. Entry, Exit, and Pricing Power

seek to raise entry barriers and thereby deter entry. For example, firms may try to influence the perceived profits by potential entrants in a way that may deter such entry even if the existing firm is making substantial profits. This section turns to the analysis of entry and potential entry and examines in particular the way in which strategic entry deterrence may take place. In doing so, we hope to illustrate how to inform the sometimes difficult question of whether entry is likely to play the role of an effective disciplinary force.

5.2.1 Entry and Exit Decisions

Entry is the first decision a firm faces. Entering a market involves investment in assets and at least a portion of those investment costs will typically become sunk costs. On the other hand, a firm may choose not to enter and by doing so will save the sunk costs and leave its resources available for other purposes. The simplest model of firm entry behavior therefore posits that a firm will enter a market if the net profits obtained from doing so are at least as large as the best alternative use of its capital. Of course, since sunk costs incurred on entry today will typically generate a stream of profits over some future time horizon, when estimating net profits, we will often want to consider the net present value of the stream of profits it hopes to generate in the future. Such an approach will be familiar to accountants in the form of discounted cash flow (DCF) approaches to evaluating profit opportunities and to financial analysts evaluating whether stocks are appropriately valued. The difference with standard accounting and financial market practice here is that we must typically study such entry opportunities in strategic environments. To that end, in this section we study methods which may help us to analyze such strategic situations.

We begin by studying a practical example for a two-firm game in which each firm must, like an entry game, make a 1 or 0 decision—in this case to exit or stay in the market. Exactly the same methods can apply to the analysis of entry games, although the data required are necessarily more prospective if a firm has not yet decided to enter the market. After this illustrative example, we return to consider the entry game.

For illustration, consider the competition between two air carriers Prime Air and Lean Air considering entering an intercontinental route. Assume initially, Prime Air was awarded the only available slots to link both airports. Having been awarded a monopoly, Prime Air expects to make handsome profit. However, Lean Air announces shortly after that it will also start flying to a sufficiently close airport that has refurbished its facilities for international flights. Prime Air management had thought that the airport in question was too small and remote so that the entry

---

20The observant reader will detect that this is a fictional example, but it is one that has parallels in numerous real-world cases, particularly in the airline industry and local bus markets. For a wonderful and richly historical illustration of these kinds of calculations in a practical setting, see the Harvard Business School case, British Sky Broadcasting versus Sky Television (HBS case number 5-799-078).
The Relationship between Market Structure and Price

Table 5.4. Matrix representation of the strategic situation between the two competitors.

<table>
<thead>
<tr>
<th>Prime Air</th>
<th>Fight</th>
<th>Exit</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lean Air</td>
<td>(V_{Lean}^\text{Fight}, V_{Prime}^\text{Fight})</td>
<td>(V_{Lean}^\text{Exit}, V_{Prime}^\text{Exit})</td>
</tr>
<tr>
<td>Exit</td>
<td>(V_{Lean}^\text{Exit}, V_{Prime}^\text{Fight})</td>
<td>(V_{Lean}^\text{Exit}, V_{Prime}^\text{Exit})</td>
</tr>
</tbody>
</table>

of a competitor was effectively impossible. However, the smaller airport managed to solve its problems with new infrastructure and thereby was able to facilitate entry by Lean Air.

Suppose, following a frequent reality in such cases that what followed was a massive war of attrition between the incumbent Prime Air, which had previously thought it would be a monopolist and Lean Air as the new, perhaps low-cost, entrant. For example, after the announcement of Lean Air’s entry, we might see both Prime and Lean rush to increase their investment in fleets and marketing, spending substantial sums. Alternatively, or in addition at the end of the process, we might see a merger proposal from the two companies. We consider each of these possibilities below, once we have laid out a suitable analytical framework for analysis.

5.2.1.1 Net Present Values

We can study this strategic situation by examining the incentives of Prime Air and Lean Air to continue and to fight one another instead of exiting the market. Specifically, we will use the normal form of the game represented in the $(2 \times 2)$ matrix reported in table 5.4, where $V_j(a_j, a_{-j})$ is the payoff of firm $j$ under a choice of action $a_j$ when its rival chooses action $a_{-j}$.

Suppose now each firm carefully constructed a financial model detailing its expected net present value of economic profits in each circumstance. As evidence in an investigation such financial forecasts by a firm can be credible evidence, but will probably only be so if they are not prepared for the purposes of the investigation, but rather provided the basis for actual investments, or possibly if they are built using updated data but assumptions that predate an investigation. In such circumstances there is no obvious immediate incentive to “manage” the information provided and also strong incentives to make the forecast as reliable as possible since actual money is at stake.

To illustrate, notice that to put numbers in our $2 \times 2$ matrix of payoffs we require a financial model that calculates each firm’s payoffs in each strategic situation. For example, the net present value of Lean Air profits at the moment of entry could be calculated for the case where Prime Air remained active but their market share fell to, say, 70% after the entry of Lean Air. If so, that net present value is calculated
5.2. Entry, Exit, and Pricing Power

following the formula:

$$\text{NPV} = \sum_{t=1}^{T} \frac{\Pi_t}{(1 + r)^t} - S_0,$$

where $\Pi_t$ is the economic profit realized at the end of time period $t$ when both firms decide to remain active and fight, $S_0$ is the initial cost which will be sunk at the start of period 0, and $r$ is the discount rate reflecting the companies' time value of money. For the purposes of this kind of calculation it will often be appropriate to use cash flows since we do not want to have artificial accounting adjustments showing up in economic profits.

Such cash flow numbers can be used to actually put numbers into normal form games of the form that all economists are used to considering as theoretical constructs. Obviously, any net present value forecast provided by a company in, say, a merger case where there was a failing firm, argument would require considerable careful scrutiny.

5.2.1.2 Nash Equilibrium Strategies in a Static Framework

Once we calculate net present values of expected profits or payoffs, we can use the numbers to resolve a game such as the one represented in table 5.4. Depending on the Nash equilibrium generated by the values of the payoffs, both firms may choose to stay, or one of them will exit.

Assume Prime Air reacted to the entry by committing to spending large sums of money very quickly, perhaps in marketing or a predation strategy pursued by expanding output. Such a strategy may have been aimed at changing the payoffs of the game by increasing its own expected payoff while at the same time decreasing the returns to Lean Air. One interpretation of such a strategy is that it could be aimed at credibly committing to the (Fight, Fight) outcome. By credibly signaling an intention to fight, Prime Air could in turn convince Lean Air that it faced a strategic reality of almost certainly being forced out of the market. If so, then a merger (takeover) proposal from Prime Air may prove attractive to both parties relative to the cost of progressing down the path of war of attrition.\(^{21}\)

5.2.1.3 Multiple Equilibria and Equilibrium Selection

Following Bresnahan and Reiss (1990, 1991a,b), consider the general form of an entry game, as described in table 5.5.

First we note that throughout the analysis we assume $\Pi^D < \Pi^M$. If $\Pi^D > 0$, the Nash equilibrium is unique and both firms enter the market. If $\Pi^M < 0$, there are no profits to be made in the market and neither firm enters the market. But, for

\(^{21}\)For example, the U.K. Competition Commission regularly considers such issues in bus merger inquiries. See www.competition-commission.org.uk/inquiries/subjects/bus.htm.
instance, if $\Pi^D < 0 < \Pi^M$, there are two possible equilibria in the market, one in which firm 1 enters alone and another in which firm 2 enters alone. In general therefore, for at least some fixed values of payoffs (those where $\Pi^D < 0 < \Pi^M$ in the duopoly game) there will be several possible Nash equilibria in an entry game ($N$ equilibria in an $N$-firm game, where $\Pi^D < 0 < \Pi^M$). That means this model of firm behavior is, at least for some parameter values, unfortunately generating not one but several possible predictions about what will happen in the world.

One approach to such a situation is to make even stronger assumptions about firm behavior than those required for Nash equilibrium. A Nash equilibrium requires that each firm is playing a best response to its competitors’ actions so that no player has an incentive to change its action. In the cases of multiple Nash equilibria, to tie down predicted behavior further we would need to make additional assumptions that remove one outcome as a possibility. That said, sometimes it may not be necessary to determine the exact equilibrium outcome. For example, an antitrust investigator may be content with the prediction that only one firm will survive in the market and there might be no need to know precisely which one will survive.

In reality, many situations will produce multiple Nash equilibria and the payoffs to firms will differ across the different outcomes. For that reason we expect that firms will attempt to influence which outcome does in fact occur. For example, firms will sometimes play quite sophisticated games in which they try to affect the perception of their competitors about what their payoffs are in order to influence their choices and increase the odds of a particularly favorable equilibrium. By playing such games, firms may be able to increase the perceived barriers to entry into their markets and successfully limit rival entry.

An important example of such behavior involves product announcements that are well ahead of their actual launch. For instance, some software firms announce new releases of software, sometimes months or even years in advance. Some commentators have alleged that in doing so they are playing the “FUD” card, that is spreading “fear, uncertainty, and doubt” about whether rivals’ products will be successful. Such a strategy would involve sending messages to potential customers that a particular firm will be the eventual “winner” in a market, so, for example, consumers should not risk buying what will ultimately not be a successful product. It is further alleged that an example of such behavior occurred in the mid 1980s when Microsoft’s operating system MS-DOS was successful in the operating system market. The price of
MS-DOS increased from $2–5 a copy in 1981 to $25–28 in 1988, even though the allegations suggest relatively few improvements were made to the product. A rival company, Digital, developed an alternative DR-DOS that was released in its 3.31 version in March 1988. In May 1990 as DR-DOS version 5 was being released, it is reported that Microsoft announced that MS-DOS version 5 would be available in the next few months. In fact, MS-DOS 5.0 was not released until June 1991. If, as a result, customers delayed their purchases of a new version of their operating system by waiting for the announced new product, then perhaps the alleged early announcement of the release of MS-DOS 5.0 was effective. It has also been alleged that Microsoft used a similar strategy when it stated that its new Windows 3.1 operating system released in 1991 would not be able to run on DR-DOS. Digital claimed that the beta version of Windows 3.1 released in 1991 contained code that generated error messages when it was running on DR-DOS.22

The purpose of strategies such as premature announcements or apparently sinking investments is to change the perception of consumers and/or competitors about the likely final outcome of the competitive process. When there are several possible equilibria, convincing consumers and competitors that a particular outcome is the most likely is often the way to make it actually happen. When performing empirical analysis, one should pay attention to the motives of a firm’s choices of action and how those choices affect the competitive process. Actions whose sole purpose and benefit to the firm is the potentially exclusionary effect of other players should be closely scrutinized.

5.2.2 Market Power and Market Structure

In the previous section we saw that there can be several possible equilibrium outcomes in a market. In this section we concentrate on methods that exploit observed entry decisions in markets of different sizes to extract information on the (1) effect of entry on profitability, (2) the magnitude of fixed costs, and (3) the extent of market power.

5.2.2.1 Determinants of Entry

Following the framework laid out in Bresnahan and Reiss (1990, 1991a,b), we can use the two-stage game structure of entry and then competition among the active firms that we outlined at the beginning of this chapter (see figure 5.1). Doing so allows us to consider the outcomes of the second stage of the game to be equilibrium

---

The Relationship between Market Structure and Price

prices $P_N$ and a level of sales $D(P_N)$ associated with equilibrium of that second stage if $N$ firms decide to enter the market at stage 1. Those in turn allow us to describe the profits for each firm that will result if $N$ firms enter the market. Specifically, define

$$
\Pi_N = [P_N - AVC]D(P_N) - F,
$$

where AVC represents average variable cost, $D(P_N)$ firm demand, and $F$ fixed costs. If marginal costs are constant, then AVC is equal to marginal cost and independent of output and hence prices. If we further assume that $S$ is the total market size and that (1) market demand is a scaled version of a representative individual’s demand and (2) that the equilibrium of the subgame is symmetric so that total demand is shared equally between active firms, then we can write firm demand as

$$
D(P_N) = d(P_N) \frac{S_N}{N},
$$

where $d(P_N)$ represents the units demanded per customer.

Profits as expected depend on the price, which in turn depends on the number of competitors and the costs. But they also crucially depend on market size. We see that as the market size increases, the potential total profits of a firm increase. If firms enter until profits are nonpositive and we ignore the integer constraint so that $\Pi_N = 0$, we can rearrange the expression to describe the minimum market size per firm necessary for entry:

$$
S_N = \frac{S}{N} = \frac{F}{[P_N - AVC]d(P_N)}.
$$

The minimum market size is the one that will provide enough customers for $N$ active firms to each make a profit, covering their operating and fixed costs. The higher the fixed costs, the bigger the potential market needs to be, all else equal. Similarly, the higher the margin that the entrant can extract per customer, the smaller the market size for the entrant needs to be everything else constant.

We further describe the role of market size in the entry decision in figure 5.3, where $S_M$ is the minimum market size for a single firm to break even and $S_D$ is the minimum market size for the operation of two firms in the market. The variable costs per units sold are assumed to be constant, implying that prices and marginal costs do not change with market size. $V_N$ is the variable profit per representative customer with $N$ firms operating in the market so that $V_N = (P_N - AVC)d_N(P_N)/N$.

We have seen earlier in the chapter that a wide class of oligopoly models describing competition in the subgame will predict that additional players will reduce prices and increase output for each given potential market size, $S$. Intuitively since margins fall on entry of additional players a monopolist may be willing to enter, and charge monopoly prices, at a market size of $S_M$. A second firm, however, would end up only being able to charge duopoly prices so that her margin per customer would be lower. As a result, to recover sufficient monies to cover her fixed costs, the duopolist
will not be willing to enter at a market size of $2S^M$ but rather will only enter at some higher market size $S^D > 2S^M$. In terms of figure 5.3, the variable profit per customer will decrease as prices decline and this is reflected in the figure by the slope of the duopolist’s profit line being shallower than the slope of the monopolist’s profit line. Given the fixed costs, the second entrant will thus require a bigger size of market than the incumbent needed to enter the market. Higher fixed costs for the entrant would further exacerbate this situation since it would shift the duopoly profit line downward. Similarly, if the marginal costs of the entrant were higher it would also reinforce this effect.

Note that this figure also suggests that firms may be able to behave strategically in a number of ways in order to prevent or delay entry. For example, a monopolist may attempt to change the average profit per customer for competitors, increase the required fixed costs $F$ (those costs which will be sunk on entry), or increase the costs of expansion so that the necessary scale of operation for an entrant becomes too large compared with the actual market size. Heavy investments in advertising or in customer-specific infrastructure are potentially ways to achieve this result. What we obtain is a strategic shift downward of the average return per customer that renders monopoly a viable alternative for a given market size. By strategically increasing costs, the incumbent lowers its own profits but prevents or delays entry by a competitor. Even if its own profits are lower, compared with the alternative of a duopoly the monopolist is still better off. Because entry deteriorates the profitability of incumbents, firms in a market may engage in strategic behavior to diminish their own payoffs and that of potential entrants in order to increase the minimum scale they will need to be profitable. Figure 5.4 illustrates in principle how this can be done.

When entry significantly decreases the degree of incumbents’ market power, the new firm will need a larger scale of operation than was needed for the first
5. The Relationship between Market Structure and Price

The equation $S_i^M + F_i - F_j - B$ represents the relationship between market structure and price. Spending on barriers by monopolist lowers profits but can deter entry until bigger market size. Figure 5.4 illustrates strategic behavior to prevent entry. Source: Based on figure 2 in Bresnahan and Reiss (1990).

Entrant which could benefit from high margins. We explore more formally the relation between the likelihood of entry and market size in different competitive environments.

5.2.2.2 Entry and Market Size

In this section we describe how we can use information on market size and margin behavior to predict the number of firms that will enter a market. We also see that we can alternatively use information on market size and the number of firms to learn about margin behavior, i.e., the extent of market power being exploited.

Theory tells us that a larger number of firms will normally be associated with lower prices, lower margins, and, if market demand is price sensitive, higher quantity demanded. A larger number of firms will then normally be associated with a higher total demand and a lower profitability per customer. This is equivalent to saying that we expect sufficiently large markets to be associated with more firms than smaller markets and, as the number of firms increases, each additional firm will need a bigger increment in market size to cover its fixed costs. Under simple assumptions, a sufficient condition for a larger increment in market size to be associated with entry is that the profitability per customer decreases with the number of firms in the market. If this is so, new firms will only be able to enter if the market is large enough to accommodate them at lower per unit margin levels. In contrast, if per unit margins are constant and independent of the number of players, then we will find that the market size necessary to support $N$ firms is linear in the number of firms and the equilibrium number of firms will increase proportionally with market size. We
might expect such a situation where firms are involved in highly competitive markets since then margins will not drop as the number of firms increases. We illustrate this later by considering a game with price-taking firms. If per unit margins decline sufficiently fast with entry of new firms, we will need larger and larger market sizes to sustain one additional firm. This case is illustrated in figure 5.5.

We present the relation between the entry decision and the market size in general terms. Let us define a general function that describes the way in which margins change with the number of firms, \((p_N - AVC_N) = h(N)\). Since margins typically fall with the number of competitors that enter, let us assume that \(\partial h / \partial N < 0\). In that case, if we consider a symmetric equilibrium to the second stage of our game, then we can describe sales by firm \(i\) as

\[
q_i = d(p_N) \frac{S}{N} = g(N) \frac{S}{N},
\]

where \(d(p_N)\) is the firm’s demand per consumer at price \(p_N\), which in turn depends on \(N\) and so we can define a function \(g(N) \equiv d(p_N)\). Further, define \(f(N) = h(N)g(N)\), which represents the total margin per unit sold times the size of the per firm demand. The zero profit condition, i.e., the breakeven point, is defined as

\[
\Pi_N = (p - AVC)q_i - F = f(N) \frac{S}{N} - F = 0,
\]

which may be rearranged to solve for the variable we will take as exogenous and may form an important component of a data set, market size:

\[
S = \frac{NF}{f(N)} \equiv \phi(N; F).
\]

Note that this relationship says that the market size \(S = \phi(N; F)\) required to support \(N\) active firms increases in the number of firms as we would expect whenever

\[
\frac{\partial \phi(N; F)}{\partial N} = F \left( \frac{f(N) - Nf'(N)}{(f(N))^2} \right) > 0.
\]

This condition in turn holds provided \(f(N) - Nf'(N) > 0\), i.e., assuming positive margins per customer, this will hold whenever \(f'(N) < 0\). That is, if markups per customer decline with the number of firms, then, as \(N\) increases, the markup per customer \(f(N)\) becomes smaller and the necessary market size for sustaining \(N\) active firms increases. If \(f(N)\) is constant and does not move with \(N\), then the number of firms will increase proportionally with market size. If, on the other hand, the \(S = \phi(N; F)\) is convex in \(N\), then inverting the relationship would imply that \(N\) will be concave in \(S\), i.e., the number of firms we observe will be a concave function of the potential size of the market. We will get such an outcome when margins per customer \(f(N)\) drops sufficiently fast when \(N\) increases. We next illustrate these effects within the three formal economic models we examined at the start of this chapter.
5. The Relationship between Market Structure and Price

5.2.2.3 Entry in Price-Taking Competition

In price-taking environments firms must decide whether to enter given the existing price, knowing that the price will not be affected by its entry in the market. The fact that prices do not decrease following the entry of a single firm actually promotes entry with regards to other competitive environments for a given price level. On the other hand, the lower prices and margins generally present in price-taking markets will discourage the entry of the less efficient firms and encourage only efficient entry.

If we assume an efficient scale for firms, which means that we assume the presence of increasing marginal costs and some fixed costs, then in a price-taking and free-entry environment the theoretical prediction is that the number of firms will increase proportionally with market size. If the market size doubles, the equilibrium number of firms in the market doubles.

To see why, recall that in a price-taking model, firms that enter the market set their quantities to solve the following profit-maximization problem:

$$\max_{q_i} p_i q_i - (cq_i + \frac{1}{2}d q_i^2) - F_i$$

so that

$$q_i^* = \frac{p_i - c}{d}.$$  

where we have assumed $C(q_i) = cq_i + \frac{1}{2}d q_i^2 + F$. The equilibrium prices and quantities are, respectively,

$$p_i^* = \frac{Nc + Sda}{N + Sbd}$$

and

$$q_i^* = \frac{1}{d} \left( \frac{Nc + Sda}{N + Sbd} - c \right),$$

where $S$ is the number of customers in the market. Given this, we can substitute the equilibrium prices and quantities into the profit equation to solve for the equilibrium number of firms in the market:

$$\Pi_i = p_i^* q_i^* - (cq_i^* + \frac{1}{2}d q_i^2) - F = (p_i^* - c) \left( \frac{p_i^* - c}{d} \right) - \frac{d}{2} \left( \frac{p_i^* - c}{d} \right)^2 - F$$

$$= (p_i^* - c)^2 \left( \frac{1}{d} - \frac{1}{2d} \right) - F = \frac{1}{2d} \left( \frac{Nc + Sda}{N + Sbd} - c \right)^2 - F = 0.$$
So that with a little algebra, the equilibrium number of firms can be described as

\[ N^* = S \frac{d}{d} \left( \frac{b(c + \sqrt{2dF}) - a}{c - (c + \sqrt{2dF})} \right), \]

which is a linear function of the market size \( S \) (all else equal) with a slope that is given by demand and technology parameters.

Before turning to study oligopolistic environments, we end this section by noting that this linearity result does not immediately hold in the case of price-taking under constant returns to scale, where, once the fixed costs are covered, there is no further efficiency requirement and hence the cost function does not place a limit on plant or firm size. In that case, the number of firms may increase at a lower pace than market size and we cannot determine \textit{ex ante} how many firms will operate in the market. The basic difficulty is that if firms have constant returns to scale at all levels of output, the size of the firm is fairly fundamentally not determined by the simple theoretical model we have presented. In such a situation it is possible to get predictions which effectively contradict our explicit behavioral assumption, for example, the model may predict a monopoly and yet we have assumed such a monopoly would be a price-taking firm.

### 5.2.2.4 Entry in Cournot Competition

In Cournot competition, the entry of a firm will cause an increase in output that will lower the equilibrium market price and the expected margins for all firms in the market. As a result, successive potential entrants will need larger and larger additional market sizes to sustain their profitability. In Cournot competition, the margin is partly determined by the number of competing firms in the market. As more firms enter, the price decreases toward the marginal cost and the margin is reduced. Since the profitability per customer decreases with the number of firms, firms need increasingly higher numbers of customers to achieve the breakeven point upon entry. For this reason, the number of firms will increase proportionally less than the market size so that if, for example, we double the size of the market, the number of firms will less than double. This situation is illustrated in figure 5.6.

For a formal derivation, assume the inverse market demand equation

\[ p(Q) = \frac{a}{b} - \frac{1}{b} \frac{Q}{S} \]

and a cost function with constant marginal costs, \( C(q_i) = cq_i + F_i \). The firm solves the profit-maximization problem:

\[ \max_{q_i} p_i(Q)q_i - cq_i - F_i. \]
And in a symmetric equilibrium we can describe equilibrium prices and quantities, respectively, as

\[ p_i^* = \frac{a}{b} - \frac{1}{b} \frac{N q_i^*}{S} \quad \text{and} \quad q_i^* = \frac{S(a-bc)}{N+1}. \]

As in the previous section, we substitute the optimal quantity back into the profit equation:

\[
\Pi_i = p_i^* q_i^* - cq_i^* - F \\
= \left( \frac{a}{b} - \frac{N}{bS} \left( \frac{S(a-bc)}{N+1} \right) - c \right) \left( \frac{S(a-bc)}{N+1} \right) - F \\
= \frac{a-bc}{b} \left( \frac{S}{N+1} \right) - F.
\]

For the firm to break even, we need at least \( \Pi_i = 0 \). If we solve for the corresponding equilibrium number of firms, we obtain

\[ N^* = (a-bc) \sqrt{\frac{S}{bF}} - 1. \]

The number of firms is therefore concave in market size \( S \).

The Cournot equilibrium derived above is somewhat special in that, to make the algebra simple, we assumed constant marginal costs. Constant marginal costs are the result of constant returns to scale and, as we noted previously, such a technology effectively imposes no constraint on the scale of the firm. An alternative assumption would be to introduce convex costs, i.e., we could assume that at least eventually decreasing returns to scale set in. In that case, while we will still obtain the same result of concavity for smaller market sizes, we will find that as market size increases
the relationship becomes approximately linear. Such a feature emphasizes that in the limit, as market size gets big, the Cournot model becomes approximately competitive and close to the case described for the price-taking firms with decreasing returns. With a large number of firms, the effect of the diseconomies of scale sets in and the size of an individual firm is then mainly determined by technological factors while the number of active firms is determined by the size of the market.

5.2.3 Entry and Market Power

The previous sections explained the basic elements of the entry game and described particularly how market size, demand, technology, and the nature of competitive interaction will determine expected profitability and this in turn will determine the observed number of firms. An interesting consequence of these results is that they suggest we can potentially learn about the intensity of competition by observing how entry decisions occur. Bresnanan and Reiss (1990, 1991a,b) show that for this class of models, if we establish the minimum market size required for the incumbents to operate and the minimum market size for a competitor to enter, we can potentially infer the market power of the incumbents. In other words, we can potentially use the observed relationship between the number of firms and the size of the market to learn about the profitability of firms. Specifically, we can potentially retrieve information on markups or the importance of fixed costs. Consequently, we can learn about the extent to which margins and market power erodes as entry occurs and markets increase in size.

5.2.3.1 Market Power and Entry Thresholds

In this section, we examine the change in the minimum market size needed for the \(N\)th firm \(s_N\) as \(N\) grows. Particularly, we are interested in the ratio of the minimum market size an entrant needs to the minimum market size the previous firm needed to enter, \(s_{N+1}/s_N\). If entrants face the same fixed and variable costs than incumbents and entry does not change the nature of competition, then the ratio of minimum market sizes a firm needs for profitability is equal to 1. This means the \((N + 1)\)th firm needs the same scale of operation as the \(N\)th firm to be profitable. If on the other hand entry increases competitiveness and decreases margins, then the ratio \(s_{N+1}/s_N\) will be bigger than 1 and will tend to 1 as \(N\) increases and margins converge downward to their competitive levels. If fixed or marginal costs are higher for the entrant, then the market size necessary for entry will be even higher for the new entrant. If \(s_{N+1}/s_N\) is above 1 and decreasing in \(N\), we can deduce that entry progressively decreases market power.

Given the minimum size \(s_N\) required for entry introduced above

\[
s_N = \frac{S}{N} = \frac{F}{[P_N - AVC]d(P_N)}.
\]
We have
\[
\frac{s_{N+1}}{s_N} = \frac{F_{N+1}}{F_N} \frac{[P_N - AVC_N]d(P_N)}{[P_{N+1} - AVC_{N+1}]d(P_{N+1})}.
\]
If marginal and fixed costs are constant across entrants, then the relation simplifies to
\[
\frac{s_{N+1}}{s_N} = \frac{[P_N - c]d(P_N)}{[P_{N+1} - c]d(P_{N+1})}
\]
so that the ratio describes precisely the evolution in relative margins per customer.

5.2.3.2 Empirical Estimation of Entry Thresholds

Bresnahan and Reiss (1990, 1991a,b) provide a methodology for estimating successive entry thresholds in an industry using data from a cross-section of local markets. In principle, we could retrieve successive market size thresholds for entry by observing the profitability of firms as the number of competing firms increases. However, profitability is often difficult to observe. Nonetheless, by using data on the observed number of entrants at different market sizes from a cross section of markets we may learn about the relationship.

First, Bresnahan and Reiss specify a reduced-form profit function which represents the net present value of the benefits of entering the market when there are \(N\) active firms. The reduced form can be motivated by plugging in the profit function the equilibrium quantities and prices obtained from an equilibrium to a second-stage competitive interaction between a set of \(N\) active firms, following the game outlined in figure 5.1, and, say, the price-taking or Cournot examples presented above. The profit available to a firm if \(N\) firms decide to enter the market can then be expressed as a function of structural parameters and be modeled as
\[
\bar{\pi}_N(X, Y, W; \theta_1) = V^N(X; \alpha, \beta)S(Y; \lambda) - F^N(W; \gamma) + \epsilon = \tilde{\pi}_N + \epsilon,
\]
where \(X\) are the variables that shift individual demand and variable costs, \(W\) are variables that shift fixed costs, and \(Y\) are variables that affect the size of the market. The error term \(\epsilon\) captures the component of realized profits that is determined by other unobserved market-specific factors. If we follow Bresnahan and Reiss directly, then we would assume that the \(\epsilon\)s are normal and i.i.d. across markets, so that profitability of successive entrants is only expected to vary because of changes in the observed variables. Note that this formulation assumes that firms are identical and is primarily appropriate for analyzing market-level data sets. A generalization which is appropriate for firm-level data and also allows firms to be heterogeneous in profitability at the entry stage of this game is provided by Berry (1992).

Bresnahan and Reiss apply their method to several data sets each of which documents both estimates of market size and the number of firms in a cross section of small local markets. Examples include plumbers and dentists. To ensure independence across markets, they restrict their analysis to markets which are distinct.
5.2. Entry, Exit, and Pricing Power

geographically and for which data on the potential determinants of market size can be collected. The variables explaining potential market size, \( Y_m \), include the population of a market area, the nearby population, population growth, and number of commuters. The variable used to predict fixed costs for the activities that they consider is the price of land, \( W_m \). Variables included in \( X_m \) are those affecting the per customer profitability. For example, the per capita income and factors affecting marginal costs. The specification allows variable and fixed costs to vary with the number of firms in the market so that later entrants may be more efficient or require higher fixed costs.

Denoting market \( m = 1, \ldots, M \) we may parameterize the model by assuming

\[
S(Y_m; \lambda) = \lambda' Y_m,
\]

\[
V_N = X_m' \beta + \alpha_1 - \left( \sum_{n=2}^{N} \alpha_n \right),
\]

\[
F_N = W_m' \gamma_L + \gamma_1 + \sum_{n=2}^{N} \gamma_n.
\]

In order to identify a constant in the variable profit function, at least one element of \( \lambda \) must be normalized, so we set \( \lambda_1 = 1 \). Note that changes in the intercept, which arise from the gammas in the fixed cost equation, capture the changes in the level of profitability that may occur for successive entrants while changes in the alphas affect the profitability per potential customer in the market. The alphas capture the idea, in particular, that margins may fall as the number of firms increases. Note that all the variables in this model are market-level variables so there is no firm-level heterogeneity in the model. This has the advantage of making the model very simple to estimate and requiring little in the way of data. (And we have already mentioned the generalization to allow for firm heterogeneity provided by Berry (1994).) The parametric model to be estimated is

\[
\Pi_N(X_m, Y_m, W_m, \varepsilon_m; \theta_1)
\]

\[
= \Pi_N + \varepsilon_m
\]

\[
= V^N(X_m; \alpha, \beta)S(Y_m; \lambda) - F^N(W_m; \gamma) + \varepsilon_m
\]

\[
= (X_m' \beta + \alpha_1 - \sum_{n=2}^{N_m} \alpha_n)(\lambda' Y_m) - W_m' \gamma_L - \gamma_1 - \sum_{n=2}^{N_m} \gamma_n + \varepsilon_m.
\]

where \( \varepsilon_m \) is a market-level unobservable incorporated into the model. A market will have \( N \) firms operating in equilibrium if the \( N \)th firm to enter is making profits but the \((N + 1)\)th firm would not find entry profitable. Formally, we will observe \( N \) firms in a market if

\[
\Pi_N(X_m, Y_m, W_m; \theta_1) \geq 0 \quad \text{and} \quad \Pi_{N+1}(X_m, Y_m, W_m; \theta_1) < 0.
\]
Given an assumed distribution for $\varepsilon_m$, the probability of fulfilling this condition for any value of $N$ can be calculated:

$$P(\Pi_N(Y, W, Z; \theta_1) \geq 0 \text{ and } \Pi_{N+1}(Y, W, Z; \theta_1) < 0 \mid Y, W, Z; \theta_1)$$

$$\implies P(\tilde{\Pi}_N(Y, W, Z; \theta_1) + \varepsilon \geq 0 \text{ and } \tilde{\Pi}_{N+1}(Y, W, Z; \theta_1) + \varepsilon < 0 \mid Y, W, Z; \theta_1)$$

$$= P(-\tilde{\Pi}_N(Y, W, Z; \theta_1) \leq \varepsilon < -\tilde{\Pi}_{N+1}(Y, W, Z; \theta_1) \mid Y, W, Z; \theta_1)$$

$$= F_\varepsilon(-\tilde{\Pi}_{N+1}; \theta_1) - F_\varepsilon(-\tilde{\Pi}_N; \theta_1),$$

where the final equality follows provided the market-specific profitability shock $\varepsilon_m$ is conditionally independent of our market-level data $(Y_m, W_m, Z_m)$. Such a model can be estimated using standard ordered discrete choice models such as the ordered logit or ordered probit models. For example, in the ordered probit model $\varepsilon$ will be assumed to follow a mean zero normal distribution. Specifically, the parameters of the model $\theta_1 = (\lambda, \alpha, \beta, \gamma L)$ will be chosen to maximize the likelihood of observing the data (see any textbook description of discrete choice models and maximum likelihood estimation).

If the stochastic element $\varepsilon$ has a cumulative density function $F_\varepsilon(\varepsilon_m)$, then the event “observing $N$ firms in the market” corresponds to the probability that $\varepsilon_m$ takes certain values. Figure 5.7 describes the model in terms of the cumulative distribution function assumed for $\varepsilon_m$. Note that in this case, if figure 5.7 represents the actual estimated cut-offs from a data set, then it represents a zone where $N$ firms are predicted by the model to be observed, and note in particular that the zone shown is rather large: the value of the cumulative distribution function $F(-\tilde{\Pi}_{N+1}; \theta_1)$ is reasonably close to zero while $F(-\tilde{\Pi}_N; \theta_1)$ is very close to one. Such a situation might arise, for example, when there are at most three firms in a data set and $N = 2$ in the vast majority of markets.

To summarize, to estimate this model we need data from a cross section of markets indexed as $m = 1, \ldots, M$. From each market we will need to observe the data $(N_m, Y_m, W_m, Z_m)$, where $N$ is the number of firms in the market and will play the role of the variable to be explained while $(Y, W, X)$ each play of the role of explanatory variables. Precise estimates will require the number of independent markets we observe, $M$ being sufficiently large; probably at least fifty will be required in most applications. If we assume that $\varepsilon_m$ has a standard normal distribution $N(0, 1)$ and
independent across observations, we can estimate this model as an ordered probit model using maximum likelihood estimation.\(^{23}\)

The regression produces the estimated parameters that allow us to estimate the variation of profitability with market size, variable profitability, and fixed profits. Partial results, those capturing the determinants of variable profitability in the market for doctors, are presented for illustration in Table 5.6. Note that the results suggest that there is a significant change in profitability between a monopoly and a duopoly market. However, after three firms, further entry does not seem to change the average profitability of firms.

From those results, we can retrieve the market size \(S_N\) necessary for entry of successive firms. We present the results in Figure 5.8.

Looking first at the results for plumbers and tire dealers, the results suggest first that plumbers never seem to have much market power no matter how many there are. The estimated relationship between \(N\) and \(S\) is basically linear. In fact, the

\(^{23}\)For an econometric description of the model, see Maddala (1983). The model is reasonably easy to program in Gauss or Matlab and the original Bresnahan and Reiss data set is available on the web at the Center for the Study of Industrial Organization, www.csio.econ.northwestern.edu/data.html (last verified May 2, 2007).
results suggest that even a monopolist plumber does not have much market power, though it may also be that there were not many markets with just one plumber in. Somewhat in contrast, tire dealers appear to lose their monopoly rent with the second entrant and thereafter the relationship between the number of players and market size appears approximately linear as would be expected in a competitive industry. The results for doctors, chemists, and plumbers and tire dealers appear to fit Bresnahan and Reiss’s theory very nicely. Somewhat in contrast, in the dentist's results, while there is concavity until we observe two firms, the line for dentists actually shows convexity after the third entrant, indicating that profitability increases after the third entrant. Such a pattern could just be an artifact resulting from having too little data at the larger market sizes, in which case it can be ignored as statistically insignificant. However, it could also be due to idiosyncracies in the way dentist practices are organized in bigger places and if so would merit further scrutiny to make sure in particular that an important determinant of the entry decision for dentists is not missing from the model. A problem that can arise in larger markets is that the extent of geographic differentiation becomes a relevant factor and if so unexpected patterns can appear in the \((N, S)\) relationship. If in such circumstances the Bresnahan and Reiss model is not sufficient to model the data, then subsequent authors have extended the basic model in a variety of ways: Berry (1992) to allow for firm heterogeneity and Mazzeo (2002) and Seim (2006) extended the analysis and estimation of entry games to allow for product differentiation. Davis (2006c) allowed for some forms of product differentiation and also in particular chain entry so that, for example, each firm can operate more than one store and instead of choosing \(0/1\) firms choose \(0, 1, \ldots, N\). Schaumans and Verboven (2008) significantly extend Mazzeo's model into an example of what Davis (2006c) called a “two-index” version of these models. While most of the entry literature uses a pure strategy equilibrium context suitable for a game of perfect information, Seim’s paper introduces the idea that imperfect information (e.g., firms have private information about their costs) may introduce realism to the model and also, fortuitously, help reduce the difficulties associated with multiplicity of equilibria. There is little doubt that the class of models developed in this spirit will continue to be extended and provide a useful toolbox for applied work.

A striking general feature of Bresnahan and Reiss’s (1990) results is that they find fairly consistently that market power appears to fall away at relatively small market sizes, perhaps due to very relatively low fixed costs and modest barriers to entry in the markets they considered. Although the results are limited to the data they considered their study does provide us with a powerful tool for analyzing when market power is likely to be being exploited and, at least as important, when it is not.

The framework developed by Bresnahan and Reiss (1990) assumes a market where firms are homogeneous and symmetric. This assumption serves to guarantee
that there is a unique optimal number of firms for a given market size. The methodology is not, however, able to predict the entry of individual firms or to incorporate the effect of firm-specific sources of profitability such as a higher efficiency in a given firm due to an idiosyncratic cost advantage. But, if we want to model entry for heterogeneous firms, the resulting computational requirements become rather greater and the whole process becomes more complex and therefore challenging on an investigatory timetable. Sometimes such an investment may well be worthwhile, but at present, generally, most applications of more sophisticated methods are at the research and development stage rather than being directly applied in actual cases.

Although agencies have gone further than Bresnahan and Reiss in a relatively small (tiny) number of cases, the subsequent industrial organization literature is important enough to merit at least a brief introduction in this book. For example, if an agency did want to allow for firm heterogeneity, then a useful framework is provided in Berry (1992). In particular, he argues persuasively that there are important elements of both unobserved and observed firm heterogeneity in profitability, for example, in terms of different costs, and therefore any model should account for it appropriately. Many if not all firms, agencies, and practitioners would agree with the principle that firms differ in important ways. Moreover, firm heterogeneity can have important implications for the observed relationship between market size and the number of firms. If the market size increases and efficient firms tend to enter first, then we may observe greater concavity in the relationship between $N$ and $S$. Berry emphasizes the role of unobserved (to the econometrician) firm heterogeneity. In his model the number of potential entrants plays an important role in telling us about the likely role being played by unobserved firm heterogeneity. Specifically, if firm heterogeneity is important we will actually tend to observe more actual entrants in markets where there are more potential entrants for the same reason that the more times we roll a die the more times we will observe sixes. For a review of some of the subsequent literature see Berry and Reiss (2007).

### 5.2.4 What Do We Know about Entry?

Industrial organization economists know a great deal about entry and this book is not an appropriate place to attempt to fully summarize what we know. However, some broad themes do arise from the literature and therefore it seems valuable to finish this chapter with a selection of those broad themes. First, entry and exit are extremely important—and in general there is a lot of it. Second, it is sometimes possible to spot characteristics of firms which are likely to make them particularly likely entrants into markets, as any remedies section chief (in a competition agency) will be able to tell you. Third, entry and exit are in reality often, but not exclusively, best thought about as part of a process of growth and expansion, perhaps followed by shrinking and exit rather than one-off events. This section reviews a small number of the important papers on entry in the industrial organization literature. In doing so
we aim to emphasize at least one important source of such general observations and also to draw out both the modeling challenge being faced by those authors seeking to generalize the Bresnahan and Reiss article and also to paint a picture of the dynamic market environment in which antitrust investigations often take place.

5.2.4.1 Entry and Exit in U.S. Manufacturing

Dunne, Roberts, and Samuelson (1988) (DRS) present a comprehensive description of entry and exit in U.S. manufacturing by using the U.S. Census of Manufactures between 1963 and 1982. The census is produced every five years and has data from every plant operated by every firm in 387 four-digit SIC manufacturing industries.\footnote{The Standard Industrial Classification (SIC) codes in the United States have been replaced by the North American Industrial Classification System (NAICS) as part of the NAFTA process. The system is now common across Mexico, the United States, and Canada and provides standard definitions at the six-digit level compared with the four digits of the SIC (www.census.gov/epcd/www/naics.html). The equivalent EU classification system is the NACE (Nomenclature statistique des Activités économiques dans la Communauté Européenne).}

An example of a four-digit SIC classification is “metal cans,” “cutlery,” and “hand and edge tools, except machine tools and handsaws,” which are all in the “fabricated metal products” three-digit classification. In the early 1980s, a huge effort was undertaken to turn these data into a longitudinal database, the Longitudinal Research Database, that allowed following plants and firms across time. Many other countries have similar databases, for example, the United Kingdom has an equivalent database called the Annual Respondent Database.

The first finding from studying such databases is that there are sometimes very high rates of entry and exit. To examine entry and exit rates empirically, DRS defined the entry rate as the total number of new arrivals in the census in any given survey year divided by the number of active firms in the previous survey year:

\[
\text{ENTRY RATE} = \frac{\text{New arrivals this census}}{\text{Active firms}_{t-1}}.
\]

Similarly, DRS defined the exit rate as the total number of firms that exited since the last survey year divided by the total number of firms in the last survey year:

\[
\text{EXIT RATE} = \frac{\text{Exits since last census}}{\text{Active firms}_{t-1}}.
\]

Table 5.7 presents DRS’s results from doing so.

First note that the entry rate is very high, at least in the United States, on average in manufacturing. Between 41 and 52\% of all firms active in any given census year are entrants since the last census, i.e., all those firms have entered in just five years! Similarly, the exit rate is very high, indeed a similar proportion of the total number of firms. Even ignoring entry and exit of smallest firms, the turnover appears to be very substantial. On the other hand, if we examine the market share of entrants and exitors, we see that on average entrants enter at a quarter to a fifth of the average share.
Table 5.7. Entry and exit variables for the U.S. manufacturing sector.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Entry rate (ER):</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>All firms</td>
<td>0.414</td>
<td>0.516</td>
<td>0.518</td>
<td>0.517</td>
</tr>
<tr>
<td>Smallest firms deleted</td>
<td>0.307</td>
<td>0.427</td>
<td>0.401</td>
<td>0.408</td>
</tr>
<tr>
<td>Entrant market share (ESH):</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>All firms</td>
<td>0.139</td>
<td>0.188</td>
<td>0.146</td>
<td>0.173</td>
</tr>
<tr>
<td>Smallest firms deleted</td>
<td>0.136</td>
<td>0.185</td>
<td>0.142</td>
<td>0.169</td>
</tr>
<tr>
<td>Entrant relative size (ERS):</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>All firms</td>
<td>0.271</td>
<td>0.286</td>
<td>0.205</td>
<td>0.228</td>
</tr>
<tr>
<td>Smallest firms deleted</td>
<td>0.369</td>
<td>0.359</td>
<td>0.280</td>
<td>0.324</td>
</tr>
<tr>
<td>Exit rate (XR):</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>All firms</td>
<td>0.417</td>
<td>0.490</td>
<td>0.450</td>
<td>0.500</td>
</tr>
<tr>
<td>Smallest firms deleted</td>
<td>0.308</td>
<td>0.390</td>
<td>0.338</td>
<td>0.372</td>
</tr>
<tr>
<td>Exiter market share (XSH):</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>All firms</td>
<td>0.148</td>
<td>0.195</td>
<td>0.150</td>
<td>0.178</td>
</tr>
<tr>
<td>Smallest firms deleted</td>
<td>0.144</td>
<td>0.191</td>
<td>0.146</td>
<td>0.173</td>
</tr>
<tr>
<td>Exiter relative size (XRS):</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>All firms</td>
<td>0.247</td>
<td>0.271</td>
<td>0.221</td>
<td>0.226</td>
</tr>
<tr>
<td>Smallest firms deleted</td>
<td>0.367</td>
<td>0.367</td>
<td>0.310</td>
<td>0.344</td>
</tr>
</tbody>
</table>

Source: Dunne et al. (1988, table 2). The table reports entry and exit variables for the U.S. manufacturing sector (averages over 387 four-digit SIC industries).

scale of existing firms in their product market and therefore account for only 14–17% share of the total market between the years surveyed. Exiting firms have very similar characteristics. The fact that entering and exiting firms are small gives us our first indication that successful firms grow after entry but unless they maintain that success, then they will shrink before eventually exiting. At the same time other firms will never be particularly successful and they will enter small and exit small having not substantively changed the competitive dynamics in an industry. Small-scale entry will always feature in competition investigations, but claims by incumbents that such small-scale entry proves they cannot have market power are usually not appropriately taken at face value.

The figures in table 5.7 report the average (mean) rates for an individual manufacturing industry and Dunne et al. also report that a large majority of industries have entry rates of between 40 and 50%. Exceptions include the tobacco industry with only 20% of entry and the food-processing industries with only 24%. They found the highest entry rate in the “instruments” industry, which has a 60% entry rate. Finally, we note that DRS find a significant correlation between entry and exit measures, an observation we discuss further below.
5.2.4.2 Identifying Potential Entrants

There are a number of ways to evaluate the set of potential entrants in a market. Business school strategy teachers often propose undertaking a SWOT (strengths, weaknesses, opportunities, and threats) analysis and such analyses do sometimes make their way into company documents. After a company has undertaken such an analysis, identified potential entrants will often be named under “threats,” while markets presenting potential entry opportunities may be named in the opportunities category. Thus information on potential entrants may come from company documents or, during an investigation, from surveys and questionnaires of customers or rivals (who may consider backward integration), and/or senior managers (the former may have the experience and skills necessary to consider setting up rival companies). Alternatively, sometimes we can examine the issue empirically and in this section we provide a couple of well-known examples of doing so.

First, let us return to Dunne et al., who found that the average firm produces in more than one four-digit product classification and that single-plant firms account for 93–95% of all firms but only 15–20% of the value of production. The latter figure implies that multiplant firms account for an 80–85% share of total production. Such observations suggest examining entry and exit rates by dividing potential entrants into three types: new firms, diversifying firms entering the market with a new plant, and diversifying firms entering the market using an existing plant.

Table 5.8 shows the entrants by type. Note that in any survey year, most entrants are new firms opening new plants while diversifying firms opening a new plant are a relatively rare event as it is much more common for diversifying firms to enter by diversifying production at their existing plant. On the other hand, when a diversifying firm enters with a new plant, it enters at a much larger scale than the other entrant types, at a whopping 90% or more of the average size of the existing firms in three of the survey years considered. Thus while entry by a multiproduct firm opening a new plant is a relatively rare event, when it happens it will often represent the appearance of a very significant new competitor.

For an example of how this can work, consider the U.K. Competition Commission’s analysis of the completed acquisition by Greif Inc. of the “new steel drum and closures” business of Blagden Packaging Group, where new large-scale entry played a very important role.25 The CC noted that the merger, on its face, was likely to result in a post-merger market share (of new large steel drums and closures in the United Kingdom) of 85%, with the merger increment 32%. On the face of it, since imports were negligible pre-merger, this merger clearly appeared to raise substantial concerns unless there were some mitigating factors such as a very high demand

---

25 Closure systems are the mechanism by which the contents of a drum can be poured or pumped out and the drum resealed. The CC found the market in closures was global so that the area of concern was only steel drums. The CC (2007a) “found that, over the past five years, both Greif and Blagden lost more custom to each other than to any other competitor in the world.”
### Table 5.8. Entry variables by types of firms and method of entry.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Entry rate</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>0.307</td>
<td>0.427</td>
<td>0.401</td>
<td>0.408</td>
</tr>
<tr>
<td>NF/NP</td>
<td>0.154</td>
<td>0.250</td>
<td>0.228</td>
<td>0.228</td>
</tr>
<tr>
<td>DF/NP</td>
<td>0.028</td>
<td>0.053</td>
<td>0.026</td>
<td>0.025</td>
</tr>
<tr>
<td>DF/PM</td>
<td>0.125</td>
<td>0.123</td>
<td>0.146</td>
<td>0.154</td>
</tr>
<tr>
<td>Entrant market share</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>0.136</td>
<td>0.185</td>
<td>0.142</td>
<td>0.169</td>
</tr>
<tr>
<td>NF/NP</td>
<td>0.060</td>
<td>0.097</td>
<td>0.069</td>
<td>0.093</td>
</tr>
<tr>
<td>DF/NP</td>
<td>0.019</td>
<td>0.039</td>
<td>0.015</td>
<td>0.020</td>
</tr>
<tr>
<td>DF/PM</td>
<td>0.057</td>
<td>0.050</td>
<td>0.058</td>
<td>0.057</td>
</tr>
<tr>
<td>Entrant relative size</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>0.369</td>
<td>0.359</td>
<td>0.280</td>
<td>0.324</td>
</tr>
<tr>
<td>NF/NP</td>
<td>0.288</td>
<td>0.308</td>
<td>0.227</td>
<td>0.311</td>
</tr>
<tr>
<td>DF/NP</td>
<td>0.980</td>
<td>0.919</td>
<td>0.689</td>
<td>0.896</td>
</tr>
<tr>
<td>DF/PM</td>
<td>0.406</td>
<td>0.346</td>
<td>0.344</td>
<td>0.298</td>
</tr>
</tbody>
</table>

*NF/NP, new firm, new plant; DF/NP, diversifying firm, new plant; DF/PM, diversifying firm, product mix. Source: Dunne et al. (1988, table 3). Entry variables by type of firm and method of entry. (Averages over 387 four-digit SIC industries.)*

elasticity. However, toward the end of the merger review process, a new entrant building a whole new plant was identified: the Schuetz Group was constructing a new plant at Moerdijk in the Netherlands, including a new steel drum production line “with significant capacity.” The company described the facility as consisting of a floorspace of 60,000 m² located strategically and ideally located between Rotterdam and Antwerp with a capacity of 1.3 million drums annually per shift. The total U.K. sales of new large steel drums were estimated to be approximately 3.7 million in 2006. This new entrant, whose plant was not operational at the time of the CC’s final report, was deemed likely to become an important competitive constraint on the incumbents once it did open at the end of 2007 or early 2008. This appears to be one example of a diversifying firm entering a market by building a new plant of significant scale, although the diversification is relative to the U.K. geographic market rather than the activities of the firm per se.

---

27 See paragraph 8.4 of CC (2007).
29 In this case, Schuetz was already involved in some closely related products in the United Kingdom; specifically, it was a U.K. manufacturer of intermediate bulk containers but not new large steel drums. Schuetz was also already active in steel drums and a number of other bulk packaging products elsewhere in the world.
### Table 5.9. Number and percentage of markets entered and exited in large cities by airlines.

<table>
<thead>
<tr>
<th>Airline</th>
<th># of markets served</th>
<th># of markets entered</th>
<th># of markets exited</th>
<th>% of markets entered</th>
<th>% of markets exited</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Delta</td>
<td>281</td>
<td>43</td>
<td>28</td>
<td>15.3</td>
<td>10.0</td>
</tr>
<tr>
<td>2 Eastern</td>
<td>257</td>
<td>22</td>
<td>22</td>
<td>8.9</td>
<td>8.9</td>
</tr>
<tr>
<td>3 United</td>
<td>231</td>
<td>36</td>
<td>25</td>
<td>12.8</td>
<td>10.8</td>
</tr>
<tr>
<td>4 American</td>
<td>207</td>
<td>12</td>
<td>12</td>
<td>5.8</td>
<td>5.8</td>
</tr>
<tr>
<td>5 USAir</td>
<td>201</td>
<td>17</td>
<td>17</td>
<td>8.5</td>
<td>8.5</td>
</tr>
<tr>
<td>6 TWA</td>
<td>174</td>
<td>23</td>
<td>23</td>
<td>13.2</td>
<td>13.2</td>
</tr>
<tr>
<td>7 Braniff</td>
<td>112</td>
<td>20</td>
<td>20</td>
<td>17.9</td>
<td>17.9</td>
</tr>
<tr>
<td>8 Northwest</td>
<td>75</td>
<td>7</td>
<td>7</td>
<td>8.0</td>
<td>8.0</td>
</tr>
<tr>
<td>9 Republic</td>
<td>69</td>
<td>6</td>
<td>6</td>
<td>13.0</td>
<td>8.7</td>
</tr>
<tr>
<td>10 Continental</td>
<td>62</td>
<td>5</td>
<td>5</td>
<td>14.5</td>
<td>8.1</td>
</tr>
<tr>
<td>11 Piedmont</td>
<td>61</td>
<td>14</td>
<td>14</td>
<td>23.0</td>
<td>3.3</td>
</tr>
<tr>
<td>12 Western</td>
<td>51</td>
<td>7</td>
<td>7</td>
<td>11.8</td>
<td>13.7</td>
</tr>
<tr>
<td>13 Pan Am</td>
<td>45</td>
<td>1</td>
<td>1</td>
<td>2.2</td>
<td>2.2</td>
</tr>
<tr>
<td>14 Ozark</td>
<td>28</td>
<td>4</td>
<td>4</td>
<td>14.3</td>
<td>14.3</td>
</tr>
<tr>
<td>15 Texas Int’l</td>
<td>27</td>
<td>6</td>
<td>6</td>
<td>11.1</td>
<td>22.2</td>
</tr>
</tbody>
</table>

Source: Berry (1992, table II). The number and percentage of markets entered and exited in the large city sample by airline.

Interestingly, the fact that entry does not usually happen at the average scale of operation for the industry is at least somewhat at odds with the assumption of U-shaped average cost curves that predict that most firms should have approximately the same efficient scale in the long run, as proposed in the influential Viner (1931) cost structure theory of the size of the firm. Indeed, one could in extremis argue that these data seem to suggest that theory applies to only 2% of the data!

Berry (1992) provides an industry study where it proves possible to provide evidence on the set of people who are likely to be potential entrants. He extensively describes entry activity in the airline sector by using data from the “origin and destination survey,” which comprises a random sample of 10% of all passenger tickets issues by U.S. airlines. While Berry’s data involve only data from the first and third quarters of 1980, it enables him to construct entry and exit data for that relatively short period of nine months. Specifically, to look at entry and exit over the period he constructs 1,219 “city-pair” markets linking the fifty major cities in the United States. City-pair markets are defined as including tickets issued between the two cities and do not necessarily involve direct flights, but (realistically) assuming that the 10% ticket sample gives us a complete picture of the routes being flown, it enables entry and exit data to be constructed (albeit under an implicitly broad market definition where customers are willing to change planes). The results are provided in table 5.9, which again reveals that there is a lot of entry and exit activity taking place.

---

30 See chapter 2 and, in particular, chapter 4 of Viner (1931), reprinted in Stigler and Boulding (1950).


Specifically, if we look at the results by markets, we see that entry and/or exit occurred in more than a third of all markets, which implies significant dynamism in the industry since this entry relates to only a nine-month period. Furthermore, table 5.10 reports that 3.37% of markets, i.e., forty-one city-pair markets, experienced both entry and exit over these nine months. The existence of apparently simultaneous entry and exit indicates that firm heterogeneity probably plays a role in the market outcomes: some firms are better suited to compete in some of the markets.

Berry (1992) examines whether airport presence in one of the cities makes an airline carrier more likely to enter a market linking this city. He finds that this is indeed the case. As illustrated in table 5.11, only rarely is there entry by someone not already operating out of or into at least one of the cities concerned. In this case, if one wants to estimate the likelihood of entry in the short term, potential entrants should be defined as carriers that already operate in at least one of the cities.

To conclude this section, let us say that although the DRS study describes only the manufacturing sector of the U.S. economy of the 1960s to the 1980s, the study
remains both important and insightful more generally. In particular, it provides us with a clear picture of the extensive amount of entry and exit that can occur within relatively short time periods. If entry and exit drive competition, and most importantly productivity growth, then protecting that dynamic process will be extremely important for a market economy to function, vital if the new entrants are drivers of innovation. The facts thus outlined suggest in particular that while antitrust authorities can play a very important short-term or even medium-term role in considering whether market concentrations should be allowed to occur, the effect of an increase in concentration which enhances market power may last only a relatively few years provided there are no substantial barriers to entry which act to keep out rivals attracted by the resulting high profits. Making sure that profitable entry opportunities can potentially be exploited by new or diversifying firms, i.e., ensuring efficient entrants face at least a fairly level playing field, thus provides one of the most important functions of competition policy.

5.3 Conclusions

- Most standard models of competition predict an effect of market structure on the level of prices. Generally, all else equal, an increase in concentration or a decrease in the number of firms operating in the market will be expected to raise market prices and decrease output. In the case of firms competing in prices of differentiated products which are demand substitutes, this effect is unambiguously predicted by simple models. Whether such price rises/output falls are in fact material, and whether all else is indeed equal, are therefore central questions in most competition investigations involving changes in market structure.

- One way to examine the quantitative effect of changes in market structure on outcomes such as prices and output is to compare the outcomes of interest across similar markets. The (impossible) ideal is to find markets that differ only in the degree of concentration they exhibit. In reality we look for markets that do not differ “too much” or in the “wrong way.” In particular, an analyst must be wary of differing cost or demand characteristics of the different markets and when interpreting such cross-market evidence an analyst must always ask why otherwise similar markets exhibit different supply structures. In the jargon of econometrics, cost and demand differences across markets that are not controlled for in our analysis can result in our estimates suffering from endogeneity problems. If so, then our observed correlation between market structure and price is not indicative of a causal relationship but rather our correlation is caused by an independent third factor.
5.3. Conclusions

- When the data allow, econometric techniques for dealing with the endogeneity problem can be very useful in attempting to distinguish correlations from causality. Such techniques include the use of instrumental variables and fixed effects. However, any technique for distinguishing two potential explanations for the same phenomenon relies on assumptions for identifying which of the contenders is in fact the true explanation. For example, when using the fixed-effects technique, there must at a minimum be both (1) within-group variation over time and (2) no other significant time-varying unobserved variables that are not accounted for in our analysis. The latter can be a problem, in particular, when using identifying events over time such as entry by nearby rivals. For example, sometimes prices rise following entry when firms seek to differentiate their product offerings in light of that entry.

- Entry increases the number of firms in the market and, in an oligopoly setting, is generally expected to lower prices and profitability in the market. Factors which will affect whether we observe new entry may include expected profitability for the entrant post entry, which in turn is determined by such factors as the costs of entrants relative to incumbents, the potential size of the market, and the erosion of market power due to the presence of additional firms. Moreover, incumbents can sometimes play strategic games to alter the perceived or actual payoffs of potential entrants in order to deter entry.

- The economics literature emerging from static entry games has suggested that the relationship between market size and the number of firms can be informative about the extent of market power enjoyed by incumbents. To learn about market power in this way, one must, however, make strong assumptions about the static nature of competition. In particular, such analyses largely consider entry as a “one-off” event, whereas entry is often best considered as a “process” as firms enter on a small scale, grow when they are successful, shrink when they are not, and perhaps ultimately exit.

- Relatedly, many markets are dynamic, experiencing a large amount of entry and exit. A considerable amount of the observed entry and exit only involves very small firms on the fringe of a market. However, a large number of markets do exhibit entry and exit over relatively short time horizons on a substantial scale. The existence of substantive entry and exit can alleviate the concerns raised by actual, or, in the case of anticipated mergers, potential market concentration. However, the importance of entry as a disciplining device on incumbent firms also underlines the need for competition authorities to preserve the ability of innovative and efficient new entrants to displace inefficient incumbent firms.
Identification of Conduct

In the previous chapter, we discussed two major methods available for assessing the effect of market structure on pricing and market power, the question at the heart of merger investigations. The broader arena of competition policy is also concerned with collusion by existing firms or the abuse of market power by a dominant firm. For example, the U.S. Sherman Act (1890) is concerned with monopolization. In Europe, since the Treaty of Rome (1957) contains a reference to “dominant” firms, collusion is known as the exercise of joint or collective dominance while the latter is known as “single” dominance. Any such case obviously requires a finding of dominance and in order to determine whether a firm (or group of firms) is dominant we need to know the extent of its individual (collective) market power.

In this chapter, we discuss methods for identifying the presence of market power and in particular whether we can use data to discriminate between collusive outcomes, dominant firm outcomes, competing firms acting as oligopolies, or outcomes which sufficiently approximate perfect competition. That is, we ask whether we can tell from market outcomes whether firms are imposing genuine competitive constraints on one another, or instead whether firms possess significant market power and so can individually or collectively reduce output and raise prices to the detriment of consumers.

Abuses of monopoly power (single dominance) are forbidden in European and U.S. competition law. However, the range of abuses that are forbidden differs across jurisdictions. In particular, in the EU both exclusionary (e.g., killing off an entrant) and exploitative abuses (e.g., charging high prices) are in principle covered by competition law while in the United States only exclusionary abuses are forbidden since

---

1 For a tour de force of the evolution of U.S. thinking on antitrust, see Shapiro and Kovacic (2000).
2 The term “dominant” appears in the Treaty of Rome, the founding treaty of the European Common Market signed in 1957 and has played an important role in European competition policy ever since. The term is unwieldy for most economists, as many are more familiar with cartels, monopolies, and oligopolies. Today there are two relevant treaties which have been updated and consolidated into a single document known as the consolidated version of the Treaty on European Union and of the Treaty Establishing the European Community. This document was published in the Official Journal as OJ C. 321 E/1 29/12/2006. The latter treaty is a renamed and updated version of the Treaty of Rome. The contents of Articles 81 and 82 of the treaty are broadly similar to the contents of the first U.S. antitrust act, the Sherman Act (1890) as updated by the Clayton Act (1914). The laws in the European Union and the United States differ, however, in some important areas. In particular, under the Sherman Act charging monopoly prices is not illegal while under EU law, it can be. In addition, jurisprudence has introduced differing legal tests for specific types of violations.
6.1. The Role of Structural Indicators

the Sherman Act states that to “monopolize, or attempt to monopolize,” constitutes a felony but it does not say that to be a monopolist is a problem. The implication is that a monopolist may, for example, charge whatever prices she likes so long as dominant companies do not subsequently protect their monopolies by excluding others who try to win business. In Europe, a monopolist or an industry collectively charging prices that result in “excessive margins” could in principle be the subject of an investigation.

When discussing collusion (joint dominance) it is important to distinguish between explicit collusion (cartels) and tacit collusion, since the former is a criminal offense in a growing number of jurisdictions. In Australia, Canada, Israel, Japan, Korea, the United Kingdom, and the United States the worst forms of cartel abuses are now criminal offenses so that cartelists may serve time in jail for their actions. Events that increase the likelihood of explicit or tacit coordination are also closely watched by competition authorities due to their negative effect on the competitive process and consumer welfare. For example, a merger can be blocked if it is judged likely to result in a “coordinated effect,” i.e., an increased likelihood of the industry engaging in tacit collusion.

We begin our discussion of this important topic by first revisiting the history and tradition of the “structure–conduct–performance” paradigm that dominated industrial organization until the emergence of game theory. While such an approach is currently widely regarded as old fashioned, we do so for two reasons. First it provides a baseline for comparison with more recent work motivated primarily by static game theory. Second, the movement toward analysis of dynamic games where evolution of market shares may sometimes occur slowly over time, and empirical evidence about early mover advantages in mature industries may, in the longer term, restore the flavor of some elements of the structure–conduct–performance paradigm. For example, some influential commentators are currently calling for a return to a “structural presumption,” where, for example, more weight is given to market shares in evaluating a merger (see, in particular, Baker and Shapiro 2007).

6.1 The Role of Structural Indicators

The structure–conduct–performance (SCP) framework—which presumes a causal link between the structure of the market, the nature of competition, and market

---

Footnotes:

3 U.S. cartelists have served jail time for many years since cartelization became a criminal offense (in fact a felony) after the Sherman Act in 1890. Outside the United States, experience of criminal prosecutions in this area is growing. Even where active enforcement by domestic authorities is limited, in a number cases the very fact that legislation has passed criminalizing cartel behavior has enabled U.S. authorities to pursue non-U.S. nationals in U.S. courts. The reason is that bilateral extradition treaties sometimes require that the alleged offense is a criminal one in both jurisdictions.

4 See, for example, the work by Sutton (1991), Klepper (1996), and Klepper and Simons (2000), and in the strategy literature see Markides and Geroski (2005) and McGahan (2004).
outcomes in terms of prices, output, and profits—has a long history in industrial organization. Indeed, competition policy relies on structural indicators for an initial assessment of the extent of market power exercised by firms in a market. For example, conduct or mergers involving small firms with market shares below a certain threshold will normally not raise competition concerns. Similarly, mergers that do not increase the concentration of a market above a certain threshold are assumed likely to create minimal harm to consumers and for this reason we enshrine “safe harbors” in law. This provides legal certainty, the benefits of which may outweigh any potential competitive damage. Those structural thresholds are useful to provide some discriminating mechanism for competition authorities and allow them to concentrate on cases that are more likely to be harmful. However, “structural indicators” such as market shares are now treated only, as the name emphasizes, as indicators and are not considered conclusive evidence of market power. It is possible that the pendulum will swing back slightly to place more presumptive weight on structure in future years, though it is not clear it will do so at present. However, even if it does, the lessons of static game theory that drive current practice and that we outline below will remain extremely important. Most specifically, in some particular situations, a high market share may provide an incumbent with very little market power.

6.1.1 Structural Proxies for Market Power

Most of the structural indicators that competition authorities consider when establishing grounds for an investigation or to voice concerns are derived from relationships predicted by the Cournot model. For example, the reliance on market shares, concentration ratios, and the importance attributed to the well-known Herfindahl–Hirschman index (HHI) can each be theoretically justified using the static Cournot model.

6.1.1.1 Economic Theory and the Structure–Conduct–Performance Framework

In antitrust, good information on marginal cost is rare, so it is often difficult to directly estimate margins at the industry level to determine the presence of market power. However, if we are prepared to make some assumptions we may have alternative approaches. In particular, we may be able to use structural indicators to infer profitability. For example, under the assumption that a Cournot game captures the nature of competition in an industry, a firm’s margin is equal to the individual market share divided by the market demand elasticity:

\[
\frac{P(Q) - C_i'(q_i)}{P(Q)} = \frac{s_i}{\eta D},
\]

5 For a very nice description of the numerous market share thresholds enshrined in EU and U.K. law, see Whish (2003).
where \( s_i \) is the market share of the firm and \( \eta^D \) the market demand elasticity. Furthermore, under Cournot, the weighted average industry margin is equal to the sum of the squared individual market share divided by the market demand elasticity:

\[
\sum_{i=1}^{N} s_i \left( \frac{P(Q) - C'(q_i)}{P(Q)} \right) = \frac{1}{\eta^D} \sum_{i=1}^{N} s_i^2.
\]

To derive these relationships, recall that in the general Cournot first-order condition for a market with several firms is

\[
\frac{\partial \pi_i(q_i, q_{-i})}{\partial q_i} = P \left( \sum_{j=1}^{N} q_j \right) + q_i P' \left( \sum_{j=1}^{N} q_j \right) - C'(q_i) = 0.
\]

If we denote \( Q = \sum_{j=1}^{N} q_j \) and we rearrange the first-order condition, we obtain the firm’s markup index, also called the Lerner index, as a function of the firm’s market share and the elasticity of the market demand:

\[
\frac{P(Q) - C'(q_i)}{P(Q)} = q_i \frac{P'(Q)}{P(Q)} \iff \frac{P(Q) - C'(q_i)}{Q \frac{P'(Q)}{P(Q)}} = \frac{s_i}{\eta^D}.
\]

This relationship can be used in a variety of ways. First, note that if we are prepared to rely on the theory, the Cournot–Nash equilibrium allows us to retrieve the markup of the firm using market share data and market demand elasticity. The markup will be higher, the higher the market share of the firm. However, the markup will decrease with the market demand elasticity. That means that a high market share will be associated with a high markup, but that a high market share is not in itself sufficient to ensure high markups. Even a high market share firm can have no market power, no ability to raise price above costs if the market demand is sufficiently elastic. An important fundamental implication is that while high market shares are a legitimate signal of potential market power, high market shares should not in themselves immediately translate into a finding of market power by antitrust authorities. Naturally, measuring the nature of price sensitivity will be helpful in determining if this is, in the particular case under consideration, a factually relevant defense or just a theoretical argument.

There are estimates of average markups in many industries, often constructed using publicly available data. Domowitz et al. (1988), for example, estimate average margins for different industries in the United States using the Census of Manufacturing data and find that the average Lerner index for manufacturing industries in the years 1958–81 is 0.37.

### 6.1.1.2 The Herfindahl–Hirschman Index and Concentration Ratios

There is a long tradition of inferring the extent of market power from structural indicators of the industry. Firm size and industry concentration are the most commonly
Table 6.1.  HHI measures of market concentration: comparison of CR(4) and HHI measures of market concentration.

<table>
<thead>
<tr>
<th>Firm</th>
<th>Share</th>
<th>Share²</th>
<th>Firm</th>
<th>Share</th>
<th>Share²</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>20</td>
<td>400</td>
<td>1</td>
<td>50</td>
<td>2,500</td>
</tr>
<tr>
<td>2</td>
<td>20</td>
<td>400</td>
<td>2</td>
<td>20</td>
<td>400</td>
</tr>
<tr>
<td>3</td>
<td>20</td>
<td>400</td>
<td>3</td>
<td>5</td>
<td>25</td>
</tr>
<tr>
<td>4</td>
<td>20</td>
<td>400</td>
<td>4</td>
<td>5</td>
<td>25</td>
</tr>
<tr>
<td>5</td>
<td>20</td>
<td>400</td>
<td>5</td>
<td>5</td>
<td>25</td>
</tr>
<tr>
<td>—</td>
<td>—</td>
<td>—</td>
<td>6</td>
<td>5</td>
<td>25</td>
</tr>
</tbody>
</table>

CR(4) 80  CR(4) 80
HHI 2,000 HHI 2,950

used structural indicators of profitability and both are thought to be positively correlated with market power and margins. The two most common indicators of industry concentration are the $K$-firm concentration ratio and Herfindahl–Hirschman index (HHI).

The $K$-firm concentration ratio (CR) involves calculation of the market shares of the largest $K$ firms so that

$$C_K = \sum_{i=1}^{K} s_{(i)},$$

where $s_{(i)}$ is the $i$th largest firm’s market share.

The HHI is calculated using the sums of squares of market shares:

$$HHI = \sum_{i=1}^{N} s_i^2,$$

where $s_i$ is the $i$th firm’s market share expressed as a percentage so that the HHI will take values between 0 and 10,000 ($= 100^2$). As illustrated above, in the Cournot model, the HHI is proportional to industry profitability and can therefore be related to firms’ market power.

The HHI will be higher if the structure of the market is more asymmetric. The examples in table 6.1 show that the HHI is higher for a market in which there are more firms but where one firm is very large compared with its competitors. Also, given symmetry, a larger number of firms will decrease the value of the HHI.

The result that a market with few firms, or a market with one or two very big firms, may be one where firms can exercise market power through high markups is intuitive. As a result the HHI is used as a preliminary benchmark in merger control where the data on a post-merger situation cannot be observed. Both U.S. and EU merger guidelines use the HHI screen for mergers which are unlikely to be of much
concern and to flag those that should be scrutinized. This is done by using the pre- and post-merger market shares to compute the pre- and post-merger HHI. Respectively,

\[ \text{HHI}^{\text{Pre}} = \sum_{i=1}^{N} (s_i^{\text{Pre}})^2 \quad \text{and} \quad \text{HHI}^{\text{Post}} = \sum_{i=1}^{N} (s_i^{\text{Post}})^2, \]

where, since post-merger market shares are not observed and we need a practical and easy-to-apply rule, post-merger market shares are assumed to simply be the sum of the merging firms’ pre-merger market shares. In initial screening of mergers, these values are assumed to be an indicator of the extent of margins before and after the proposed merger and the effect of the merger on such margins. Specifically, in the EU merger guidelines, mergers leading to the creation of a firm with less than 25% market share are presumed to be largely exempt from anticompetitive effects. The regulations use an indicative threshold of 40% as being the point at which a merger is likely to attract closer scrutiny. Mergers that create a HHI index for the market of less than 1,000 are also assumed to be clear of anticompetitive effects. For post-merger HHI levels between 1,000 and 2,000, mergers that increase the HHI level by less than 250 are also presumed to have no negative effect on competition. Changes in the HHI index of less than 150 at HHI levels higher than 2,000 are also declared to cause less concern except in some special circumstances. Similarly, the U.S. Department of Justice Horizontal Merger Guidelines also use a threshold at 1,000, a region of 1,000–1,800 to indicate a moderately concentrated market, and “where the post-merger HHI exceeds 1,800, it will be presumed that mergers producing an increase in the HHI of more than 100 points are likely to create or enhance market power or facilitate its exercise.”

To see these calculations in operation, next we present an example of the package tour market using flights from U.K. origins. The first and second firms in the market, Airtours and First Choice with 19.4% and 15% market shares respectively, merged to create the largest firm in the industry with a combined 34.4% market share. The HHI index jumped from approximately 1,982 before the merger to around 2,564 after the merger, an increase of 582. Such a merger would therefore be subject to scrutiny under either the EU or U.S. guidelines. Of course, in using such screens, we can only calculate market shares on the basis of a particular proposed market definition. In practical settings, that often means there is plenty of room for substantial discussion.

---


7 Ibid. In the United Kingdom, the Enterprise Act 2002 empowers the OFT to refer mergers to the CC if they create or enhance a 25% share of supply or where the U.K. turnover of the acquired firm is over £70 million. As an aside, some argue that it is not immediately clear that the term “share of supply” actually does mean the same as “market share.”

8 See the U.S. Horizontal Merger Guidelines available at www.usdoj.gov/atr/public/guidelines/hmg.htm, section 1.5.

over whether a merger meets these threshold tests even though lack of data means it is not always possible to calculate even a precise HHI number so that results near but on opposite sides of the thresholds are not appropriately treated as materially different outcomes.

A practical disadvantage of the HHI is that it requires information on the volume (or value) of sales of all companies, as distinct from a market share which requires estimates of total sales and the sales of the main parties to a merger (the merging companies). Competition agencies with powers to gather information from both main and third parties may usually be able to compute HHI, at least to an acceptable degree of approximation provided they can collect information from all the large and moderately sized players. Very small companies will not usually materially affect the outcome. On the other hand, some significant agencies (e.g., the Office of Fair Trading in the United Kingdom) do not currently have powers to compel information from third parties (while those which do may hesitate to use them) so that even computing a HHI can sometimes face practical difficulties.

It is important to note that it is not the practice to prohibit a merger based on HHI results alone. It is useful to use the HHI as a screening mechanism, but the source of the potential market power should be understood before the measures available to competition authorities are applied. That said, market shares and HHIs will certainly play a role in the weighing up of evidence when deciding whether on balance a merger is likely to substantially lessen competition.

---

**Table 6.2.** HHI calculations for a merger in the package tour industry.

<table>
<thead>
<tr>
<th>Company</th>
<th>$s_i$</th>
<th>$s_i^2$</th>
<th>Adjustments for merger</th>
</tr>
</thead>
<tbody>
<tr>
<td>Airtours</td>
<td>19.4</td>
<td>376.36</td>
<td>−376.36</td>
</tr>
<tr>
<td>First Choice</td>
<td>15.0</td>
<td>225</td>
<td>−225</td>
</tr>
<tr>
<td>Combined</td>
<td>34.4</td>
<td>1,183.36</td>
<td>+1,183.36</td>
</tr>
<tr>
<td>Thomson</td>
<td>30.7</td>
<td>942.49</td>
<td></td>
</tr>
<tr>
<td>Thomas Cook</td>
<td>20.4</td>
<td>416.16</td>
<td></td>
</tr>
<tr>
<td>Cosmos Avro</td>
<td>2.9</td>
<td>8.41</td>
<td></td>
</tr>
<tr>
<td>Manos</td>
<td>1.7</td>
<td>2.89</td>
<td></td>
</tr>
<tr>
<td>Kosmar</td>
<td>1.7</td>
<td>2.89</td>
<td></td>
</tr>
<tr>
<td>Others (&lt; 1% each)</td>
<td>8.2</td>
<td>$9 \times (8.2/9)^2$</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>100</td>
<td>1,982</td>
<td>2,564</td>
</tr>
</tbody>
</table>

*Source:* Underlying market share data are from Nielsen and quoted in table 1 from the European Commission’s 1999 decision on *Airtours v. First Choice*. These calculations treat the market shares of “Others” as being made up of nine equally sized firms, each with a market share of $8.2/9 = 0.83$. The exact assumption made about the number of small firms does not affect the analysis substantively.
6.1. The Role of Structural Indicators

6.1.1.3 Beyond the SCP Framework

Theoretical developments in industrial organization, particularly static game theory, clearly illustrated important limits of the SCP analysis. In particular, static game theory suggests that the relationship between structure, conduct, and performance is not generally best considered to be a causal relationship in a single direction. In particular, the causality between market share and market power is in no way automatically established. Even though the Cournot competition model predicts that markups are linked to the market shares of the firms, it is very important to note that high industry margins are not caused by high HHIs, even though they coincide with high HHIs. Rather, in the Cournot model, concentration and price-cost markups are both determined simultaneously in equilibrium. This means that they are ultimately both determined by the strategic choices of the firms regarding prices, quantity, or other variables such as advertising and by the structural parameters of the market, particularly the nature of demand and the nature of technology which affects costs. If the market demand and cost structure are such that optimization by individual firms leads to a concentrated market, high margins may be difficult for even the most powerful and interventionist competition agency to avoid. Under Cournot, for example, firms that are low-cost producers will have high market shares because they are efficient. Their higher markup is a direct result of their higher efficiency.

The pure SCP view of the world that structure actively determines conduct which in turn determines performance has been subjected to a number of serious critiques. In particular, as figure 6.1 illustrates, game theorists have argued that in the standard static (one shot) economic models, market structure, conduct, and market performance typically emerge simultaneously as jointly determined outcomes of a model rather than being causally determined from each other. Such analyses suggest that a useful framework for analysis is one that moves away from the simple SCP analysis, where the link between structure and market power was assumed to be one-way and deterministic, to one in which firms can endogenously choose their conduct and in return affect the market structure.
Although we have stressed the lack of established causality between structure and performance in static models, it is important to note that many dynamic economic models push considerably back in the other direction. For example, in the previous chapter we examined the simplest two-stage models, where firms entered at the first stage and then engaged in competition, perhaps in prices. In that model, structure—in the sense of the set of firms that decided to enter—is decided at the first stage and then does indeed determine prices at the second stage. A complete dismissal of SCP analysis might therefore lead agencies in the wrong direction, but the extreme version of SCP, the view that “structure” is enough to decide whether a merger should be approved, is difficult to square with (at least) a considerable amount of economic theory.

The importance of structural indicators in determining the extent of market power and the anticompetitive effects of a merger has gradually decreased as the authorities increasingly rely on detailed industry analysis for their conclusions. Still, structural indicators remain important among many practitioners and decision makers because of their apparent simplicity and their (sometimes misunderstood) link with economic theory.

6.1.2 Empirical Evidence from Structure–Conduct–Performance

The popularity of the simple SCP framework lies in the fact that it provides a tool for decision making based on data that are usually easily obtained. This has real advantages in competition policy, not least because legal rules based on structural criteria can provide a degree of legal certainty to parties considering how particular transactions would be treated by the competition system. Critics, however, point to disadvantages, particularly that certainty about the application of a simple but inappropriate rule may lead to worse outcomes than accepting the \textit{ex ante} uncertainty that results from relying on a detailed investigation of the facts during a careful investigation.

In considering the debate between the advocates and critics of SCP style analysis, and its implications for the practice of competition policy, it is helpful to understand an outline of the debate that has raged over the last sixty years within industrial organization. We next outline that debate.\textsuperscript{10}

6.1.2.1 Structure–Conduct–Performance Regressions

SCP analysis received a substantial boost in the 1950s when the new census data in the United States that provided information at the industry level were made available to researchers. These new data allowed empirical studies based on interindustry

\textsuperscript{10}For a classic survey of the profit–market power relationship and other empirical regularities documented by authors writing about the SCP tradition, see Bresnahan (1989).
6.1. The Role of Structural Indicators

comparisons to flourish. The influential study by Bain (1951)\textsuperscript{11} compared the profitability of firms in different types of industries and attempted to relate industry concentration to industry profitability.

Bain primarily compared group averages of high and low concentration industries, where concentration was measured as the eight-firm concentration ratio, $C_8$ and each firm’s profit rate was measured using accounting data as the ratio of “annual net profit after income taxes to net worth at the beginning of the year.” A simple comparison of average industry profit rates between industries with an eight-firm concentration ratio above and below 70% gave a striking difference in average profit rates of between 12.1 and 6.9%, while a test of the difference in means suggested a significant difference with a \( p \)-value of less than 0.001.

Subsequent authors, controlling for other potential determinants of profitability, ran the following simple cross-industry regression:\textsuperscript{12}

\[
y_i = \beta_0 + \beta_2 H_i + \beta_i X_i + \epsilon_i,
\]

where \( y \) was a measure of profitability (performance) such as the Lerner index \( (P - MC)/P \) or the accounting return on assets. The variable \( H_i \) denotes a measure of industry concentration, perhaps the HHI index, and \( X_i \) denotes a set of variables that measure other factors thought to affect profits such as barriers to entry, the intensity of R&D, the minimum efficient scale, buyer concentration, or product differentiation proxied by the advertising-to-sales ratio. The literature consistently if not entirely universally found\textsuperscript{13} \( \beta_2 > 0 \) and interpreted the positive coefficient as evidence of market power being exploited by firms in more concentrated industries.

One potential implication of such a relationship, were it capturing a causal relationship indicating that structure causes high margins and profits, would be that if we broke firms up, reducing concentration, profits but also margins would fall and that would help consumer welfare. Such a policy conclusion relies extremely heavily on the causal nature of the estimated relationship between structure and margins or profitability. We now follow the literature, spurred in particular by Demsetz (1973), in examining whether this relationship is causal.

\textsuperscript{11}This paper (with minor corrections reported in the following issue) tests one of the testable hypotheses proposed in Bain (1950). Bain’s work was also reported in his classic book on industrial organization (Bain 1956). Specifically, Bain had access to industry concentration data from the 1935 Census of Manufactures on 340 industries of which 149 had profit data available from the Securities and Exchange Commission (SEC) from its publication “Survey of American Listed Corporations 1936–40.” To resolve geographic market definition issues, he further selected only those industries classified as both “national” and in which each manufacturer “as a rule” was involved in production of all the products covered by an industry classification, as defined in the U.S. publication “Structure of the American Economy.” Doing so left a total of 83 industry-level observations on both profit and concentration. These were further reduced down to a total of 42 industries (355 firms), for example because SEC profit data did not cover a large proportion of industry output.

\textsuperscript{12}For a review of the literature from the 1950s to 1970s, see Scherer (1980).

\textsuperscript{13}As Pelzman (1977) describes (and cited by Clarke et al. (1984)), “With few exceptions, market concentration and industry profitability are positively correlated.”
To establish the kind of concern we might have with such regressions, let us follow Cowling and Waterson (1976) and examine the Cournot competition model in which firms compete by setting their quantities. We showed earlier that in an industry characterized by a Cournot equilibrium, the relationship between an indicator of profitability, the Lerner index, and market share should be positive. Additionally, the estimated coefficient is, according to theory, one over the market price elasticity of demand.

To capture the relationship between margins and market share, we might imagine running a regression reflecting the determinants of profits for the firm along the lines of

$$ y_{if} = \beta_0 + \beta_1 s_{if} + \beta X_{if} + \varepsilon_{if}, $$

where $i$ is the indicator for the industry and $f$ is the indicator for the firm. Variable $s_{if}$ captures the firm’s market share and $X_{if}$ other determinants of firm profitability. Now Bain and his followers were working primarily with industry-level data rather than firm-level data since only industry-level data were available from the census at that time. We can nonetheless aggregate up to the industry level and consider what we should expect to see in their regression equations. The traditional way to aggregate across firms would be to generate a weighted sum across firms within an industry using market shares as weights. Doing so gives

$$ \sum_{f=1}^{F} s_{if} y_{if} = \beta_0 \left( \sum_{f=1}^{F} s_{if} \right) + \beta_2 \sum_{f=1}^{F} s_{if}^2 + \beta \sum_{f=1}^{F} s_{if} X_{if} + \varepsilon_{if}, $$

where $\sum_{f=1}^{F} s_{if} y_{if}$ is a measure of industry profitability for industry $i$ and $\sum_{f=1}^{F} s_{if}^2$ is the HHI. Note also that $\sum_{f=1}^{F} s_{if} = 1$.

Strikingly, this is in fact the regression of average industry profits run by Bain and his colleagues using industry-level studies. One interpretation of the literature’s regression is therefore that it is exactly what you would expect to see if the world were characterized by a Cournot model. That observation provides the basis for an important critique of the SCP literature since, while the Cournot model suggests that we will observe a positive relationship between market performance (profitability) and market concentration, the relationship is not causal in the sense of running from concentration to profitability. In particular, since the only way in which firms can differ in a Cournot model is by being more efficient, and lower-cost firms will achieve higher market shares, any policy attempting to lower concentration will in fact at best end up moving production from efficient low-cost firms to inefficient high-cost ones. The opponents of SCP conclude that such a policy is highly unlikely to improve welfare and in fact very likely to actively harm consumers and generate higher prices!

For completeness, before moving on be sure to note that the market demand elasticity is also a determinant of industry profitability according to this static economic
6.1. The Role of Structural Indicators

model. This has the fundamental implication that measures of concentration are not, alone, decisive in determining whether firms in an industry are likely to be able to exploit market power. A high market share in a market where demand is very price sensitive may not endow a firm with market power.

6.1.2.2 Empirical Caveats of SCP Analysis

We have noted that the validity of the SCP framework depends heavily on being able to interpret the relationship between structure and profits (i.e., the positive coefficient $\beta_2$ in our industry regression estimation) as a causal relationship. If it is causal, then concentration causes the high margins. We have seen that, at least in static models, the number of firms, the degree of concentration, and profitability are simultaneously determined in the market where the underlying primitive factors are the variables determining demand, technology, and the strategic choices made by firms. The advent and application of static game theory naturally lead to such a conclusion.

That said, there are a number of other critiques of SCP that have implications for empirical work. Since this has been such an important force in antitrust and industrial economics history, in the next section we expand on those criticisms, which gave rise to new methodologies to identify and quantify market power.

6.1.3 Criticism of Empirical Estimations of the Effect of Structural Indicators

There are two main sources of criticisms of the studies that relate profitability to structural indicators such as industry concentration. One is an econometric criticism and states that the causality between industry concentration and market power cannot usually be established by simple correlations. The other criticism relates to the difficulty of obtaining economically meaningful measurements of firms’ profits. The latter is a topic which all competition agencies that attempt to measure firms’ profitability regularly encounter.

6.1.3.1 Firm Heterogeneity

The most extreme critique of the SCP framework denies that market structure plays any independent role at all in determining firms’ profitability and, as we have described, follows Demsetz (1973) in ascribing the positive relationship solely to efficiency. Demsetz went on to suggest that the competing market power and efficiency explanations of the observed relationship could be examined using within-industry data variation. Specifically, he argued that the efficiency hypothesis should introduce a difference between the rate of return earned respectively by large and small firms. On the other hand, at least in a homogeneous product market, a pure market power story where all firms were equally efficient would find that large and
small firms alike would earn high returns in concentrated markets. Demsetz (1973) provided results which favored the efficiency hypothesis (and which were critiqued by Bond and Greenberg (1976)).

Ravenscraft (1983) ran the following regression using the FTC’s 1975 data on lines of business for a cross-section of individual firms:

$$y_{if} = \beta_0 + \beta_1 H_i + \beta_2 s_{if} + \beta X_{if} + \epsilon_{if},$$

where $H$ again represents the HHI and where the regression varies from the cross-industry regression in the fact that it uses firm-level data and so is able to allow roles for both market share and also industry concentration.

Putting aside the extremely heroic assumption that all industries can be well approximated by a single Cournot model, with in particular a single price elasticity of demand, this regression may help distinguish the effect of concentration from the indirect effect caused by low-cost firms having high market shares. The reason is that there is no role for the HHI in the firm-level Cournot relationship between the Lerner index and the market share, while the latter controls for firm heterogeneity in the industry-level relationship. This regression involves comparing profits achieved by firms with similar market shares in industries with different degrees of concentration.

The SCP model predicts a positive coefficient for $\beta_1$, implying a positive relation between industry concentration and firm profits. There can be many lines of business per firm and the data reveal firms with as many as 47 different lines of business. The average is 8 lines of business per firm. A total of 261 lines of business were considered. A summary of Ravenscraft’s results is provided in table 6.3.

First consider the industry-level regression shown on the left in table 6.3. The results indicate a weakly positive effect of industry concentration on the firm’s profit. However, once we move to firm-level data and the firm’s market share in that line of business is included in the regression, the effect of the level of concentration in the industry seems to disappear. In fact, it has an insignificant but negative coefficient. Critics of the SCP framework suggest that this provides further evidence that profitability may be linked to the firm’s market share but not to industry concentration, once we have controlled for important differences in market share which in

Table 6.3. Relationship between market share and profitability using firm-level data.

<table>
<thead>
<tr>
<th>Variable name</th>
<th>Industry data</th>
<th>Firm-level data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Industry concentration</td>
<td>0.0375</td>
<td>-0.0222</td>
</tr>
<tr>
<td></td>
<td>(1.67)</td>
<td>(-1.77)</td>
</tr>
<tr>
<td>Firm market share</td>
<td>—</td>
<td>0.1476</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(5.51)</td>
</tr>
</tbody>
</table>

Source: Ravenscraft (1983). t-statistics are reported in parentheses.
6.1. The Role of Structural Indicators

turn reflect cost differences across firms. Their suggested explanation is that more efficient firms are more profitable and also tend to be larger thereby generating a positive relation between a firm’s market share and profitability. The common causal driver of both profitability and market shares is therefore not the industry structure but rather efficiency.

Such a position in its rawest form appears fairly extreme as a critique of the SCP. First, the literature noted that there are a number of potential explanations unrelated to efficiency for the observed within-industry relationship between profitability and market shares. For example, on the demand side product differentiation can endow firms with market power and also drive differentials in market shares, while on the cost side economies of scale may generate market power which is subsequently exploited so that we can benefit from productive efficiency (low costs of production) but suffer the consequences in terms of allocative efficiency (i.e., high-market-share firms may set high markups). In addition, firms that need to incur large fixed costs to enter and operate in an industry will do so only if they can expect large operating profits and a large scale of operations. This will limit the number of entrants and create a link between concentration and operating profitability that is not necessarily linked to a strategic exercise of market power.

Second, as we have seen in chapter 5, many simple dynamic economic theory models do predict that under oligopolistic competition, profits will decrease with the number of firms. In any two-stage game with entry followed by Cournot competition, the theoretical prediction is that concentration is a factor that will tend to increase market power, all else equal (see, in particular, section 1.4 of Sutton (1991)). The conclusion of this conceptual debate appears to be that the positive relationship between profitability and market structure is robust across industries, but probably has complex causes that may differ significantly across industries. Competition agencies must therefore pay careful attention to market power while also making sure that where evidence of efficiency benefits of concentration via product and process innovation is available it is taken into account. A relevant question is also how much of these efficiencies translate into actual consumer benefit.

For a discussion of these topics see the analysis of the model and U.K. data provided by Clarke et al. (1984), who note, for example, that introducing a U-shaped total variable cost function to the Cournot model suggests that the measured Lerner index (using observed average margins—see below) would be related to both the level and square of market shares so that fitting a linear model in terms of market share alone would result in omitted variable bias. They concluded that, “If anything, the evidence...is more sympathetic to the traditional market power explanation of profitability–concentration correlations at the industry level than to Demsetz. We find no evidence for the U.K. that differences between small and large firm profitability tend to be larger in high concentration industries” (p. 448). They concluded that the relationship between (gross profits to revenues) and the HHI was positive but with a declining slope, \( \frac{\Pi}{R} = (1/\eta_j) \{0.170 + 2.512H_j^2 - 1.682H_j^2\} \), where \( \eta_j \) denotes the market elasticity of demand in industry \( j \).
6. Identification of Conduct

6.1.3.2 Measuring Profitability

Measuring profitability can be difficult. Margins in an economic sense are rarely very cleanly observed. Margin and profitability figures taken from published accounting documents often include imputations of fixed costs and estimation of depreciation that may well not bear much relation to the economic concepts used to calculate economic costs. Also, accounting profits may be subject to intertemporal or interproduct allocations of revenues and expenses that do not correspond to meaningful economic concepts. (See chapter 3 for a discussion of the differences between accounting and economic profits.)

To take a specific example, the SCP studies generally approximated price-cost margins with \( \frac{R - TVC}{R} \), where \( R \) is revenues and TVC is the total variable cost. If we divide both elements of the ratio by quantity \( Q \), we obtain \( \frac{AR - AVC}{AR} \), where \( AR \) is the average revenue per unit and \( AVC \) is the average variable cost. This ratio will be similar to the Lerner index if and only if the average variable cost is similar to the marginal cost.

Fixed costs may also play a role in determining the structure of the market, the number of competitors, and the profitability of firms because firms need to recover and make a return on their investments. Ignoring fixed costs will reduce the analysis to short-term comparative statics and may be an inadequate framework for comparing structural equilibria across industries. For example, if we consider pharmaceutical markets we will find both high concentration and high margins but these might (or might not) be driven by very high R&D, drug approval, and marketing costs. Relatedly, Sutton (1991, 1998) challenges the profession to confront his observation that levels of fixed costs will often be choice variables so that market structure and the size of variables often treated as entry barriers (R&D and advertising levels) may be jointly determined.

Many of these points and others were made in the widely cited contribution by Fisher and McGowan (1983), which is sometimes considered to have effectively led industrial organization economists to conclude that efforts at measuring profitability were hopeless and should be abandoned. In coming to a rounded view on such matters it is worth noting that the financial markets do place a great deal of emphasis on financial accounting data while competition authorities can generally also obtain management accounting data, the kind of data often used to at least partially inform investment decisions within organizations. The skepticism of the industrial organization academic community, which has resulted in profit data rarely being used in academic industrial economics, is simply not shared by other professional groups. The response by most other groups has rather been to attempt to adapt financial data to provide evidence on the economic quantities they attempt to measure. For example, in finance, cash flows are often used in firm valuation models rather than data directly from profit and loss statements, while whole organizations have

\[15\] For a somewhat contrary view, see Geroski (2005) and also OFT (2003).
developed to train analysts in other areas to derive useful information from such accounting evidence. (For example, globally the chartered financial analyst (CFA) qualification is fairly widely recognized as certifying an individual’s ability to do so in an investment context.)

6.1.3.3 A Recent Case Example: U.K. Groceries

In its investigation of the supermarket sector, the U.K. Competition Commission ran a regression of store profit margin on local area concentration. The CC ran regressions attempting to relate store (variable) profit margin to local concentration where the latter was measured using a variety of local concentration measures. The CC motivated its regression by reference to two-stage games where entry occurred at the first stage and then competition occurred within the local area, i.e., games of the form studied in chapter 5. Specifically, it constructed a measure of variable profit margin of the form

$$\pi_j = \frac{\text{Store variable profit}}{\text{Store revenues}}$$

for a cross section of stores. The margin data relate to the period May 2005 to April 2006 for Tesco, Asda, Morrisons, and Sainsbury’s stores with a net sales area of larger than 280 m² (and below 6,000 m²). The CC estimated the regression equations using instrumental variables (particularly local population) in an attempt to control for endogeneity of market structure by using its exogenous correlate population. A selection of the CC’s parameter estimates are reported in table 6.4, indicating that the CC found that store profit margins are correlated with local market structure. In particular, the CC considered this evidence that markets were local. The issue of efficiency explanations for the relationship was explicitly considered, although variables relating to local cost conditions were not available and so could not be included.

Slade (2006) provided an expert report evaluating the CC’s regressions. She notes that (i) the profitability analysis is difficult to perform because profits and market structure are jointly determined and therefore it is difficult to determine causality; (ii) the regression specifications are free of endogeneity problems under some assumptions, but not others; but that (iii) all the statistical tests they perform indicate that the endogeneity problems has been “overcome or is minor.” She concludes, “in spite of potential difficulties, I find the CC regressions to be surprisingly

---


18 In particular, local market population measures may potentially be correlated with store profitability. If so, local population would not be a valid instrument.
6. Identification of Conduct

Table 6.4. Profit margins and local market structure in U.K. supermarkets.

<table>
<thead>
<tr>
<th>Number of competing fascias over 280 m²</th>
<th>Number of competitor stores within 10 min</th>
<th>Combined net sales area (thousands) of competitors within 10 min</th>
<th>Share of competitors' net sales area in total net sales area within 10 min</th>
</tr>
</thead>
<tbody>
<tr>
<td>Effect on store profit margin</td>
<td>-0.0096&lt;sup&gt;a&lt;/sup&gt;</td>
<td>-0.0034&lt;sup&gt;a&lt;/sup&gt;</td>
<td>-0.0026&lt;sup&gt;a&lt;/sup&gt;</td>
</tr>
<tr>
<td>t-statistics</td>
<td>(-3.05)</td>
<td>(-2.93)</td>
<td>(-3.06)</td>
</tr>
</tbody>
</table>

<sup>a</sup> $p < 0.01$.

Source: Table 1, appendix 4.6, CC’s Groceries Market Investigation (2007).

robust to changes in market structure, the functional form of the equation, and the choice of instruments. The evidence is thus supportive of the hypothesis that very local market conditions are important determinants of grocery store profits.”

Margaret Slade’s observations fit a common theme that emerges regularly when considering the appropriate treatment of econometric evidence: we either need robustness of results or else we need to have clearly established that one set of regression evidence is materially better than alternative specifications. Panel data were not available in the supermarket case. However, it is important to note that where panel data are available we can use the identification strategies we previously discussed extensively in chapter 5 when examining regressions of price on market structure.

In the next section we discuss new methodologies that have been developed that try to identify the extent of market power within an industry by looking at the firms’ behavior within an industry. Such analyses use predictions of economic theory regarding firms’ behavior in different competitive environments to identify the nature of competition, firms costs, and also the resulting profitability using firms’ observed behavior.

6.2 Directly Identifying the Nature of Competition

Cross-industry regression analysis relied on the assumption that all industries are characterized by similar empirical relationships. However, game theory quickly brought home to researchers that apparently small details in industry characteristics such as institutional or technological characteristics can, at least in the theory, be of great significance in the determination of industry equilibrium outcomes. For instance, how much market information is accessible to firms can play an important role.

<sup>19</sup> For a nice review of identification methodologies by one of the key innovators, see Bresnahan (1989).
6.2. Directly Identifying the Nature of Competition

role in the likelihood or at least nature of a collusive outcome (see, for example, Stigler 1964; Green and Porter 1984). If so, then cross-industry comparisons used in the Bain-style regression analysis will have a hard time identifying the links between observed market characteristics and outcomes since they compare environments that are in fact not easily comparable. On reflection it seems highly plausible that computers, pharmaceuticals, aircraft manufacturers, and shampoo manufacturers are incredibly different industries and therefore empirical work which captures all of their differences in a single equation is probably a rather optimistic exercise.\footnote{That said, and in recognition of the substantial contributions made by the authors writing in the SCP tradition, it is important to note that such illustrative characterizations are in truth caricatures of at least the best empirical work that was undertaken in this area.}

As a result, since the late 1970s the majority of empirical research in industrial organization has evolved to focus mainly on industry-specific studies. There are a few exceptions, notably Sutton (1991, 1998), but they remain a fairly rare activity, if one that is probably growing. Our interest in the relationships between structure, conduct, and performance remains, but the dominant methodology in industrial organization now involves examining particular industries in substantial detail.\footnote{This progression to detail is sometimes noted with amusement or even on occasion with frustration by academic colleagues from other fields within economics who in private will remark that prominent industrial organization researchers, whom they refer to as “the [radio/TV/movie/breakfast cereal/cement guy/woman]” are doing very detailed work which looks on the face of it to be rather like toying at the margins of an economy. Such observations are striking to many economists when we compare industrial organization papers to those studying “the knowledge economy” or trade economists studying “world trade flows.” On the other hand, local studies can sometimes illustrate very big ideas. For example, the study of the general topic of the diffusion of technology is often traced back to work on hybrid corn by Griliches (1957).}

To do so, industrial organization economists have developed specific methodologies that use observed data on cost drivers, prices, and quantities in order to infer the nature of competition in the market. An empirical model which nests a number of possible theoretical models can be used to discriminate among these potential competitive environments, given a suitable identification strategy. The starting point of any such analysis is careful examination of institutional structure, market history, and also the basic patterns revealed in the data. Once a researcher is ready to begin to model the process which has generated the data, the approach is to specify a structural model based on the potential behavioral assumptions that we would like to discriminate between. For example, we might want to consider a model of competition where firms individually profit maximize and then choose the firms’ strategic variables (price or quantities) that will explicitly characterize the competitive outcome in a Nash equilibrium. On the other hand, we might suspect collusion between the firms and thus our alternative model for the process which has generated the data would be a collusive model where firms might choose quantities in an attempt to maximize industry profitability. Whenever we want to distinguish
6. Identification of Conduct

two models that might have generated the same data, some feature of the data must allow us to tell them apart. We will call this our “identification strategy.” (See also chapter 2.)

In the next section, we first review the key concept of identification by examining the classic case of identification of demand and supply equations in multi-equation structural systems. We then progress to explain the extension of the methodology that has been developed for some classic models from industrial economics. For example, when we want to tell apart standard competitive models from collusive models. We will examine some classic cases, but many of the lessons of identification are important ones far more generally since the analysis of identification is exactly what allows us to test between competing economic models.

6.2.1 Structural Models of Supply and Demand

The study of identification is fundamental to distinguishing between economic models. We saw in chapter 2 that the identification of supply and demand is the canonical example which demonstrates the difficulties typically faced when developing an identification strategy. We begin this section by revisiting the identification of supply and demand in structural models. Doing so provides an important stepping stone toward the analysis of the problem of using data to identify firm conduct.

6.2.1.1 Formalizing a Structural Models of Supply and Demand

The basic components for any structural model of an industry are the demand function and the supply function, where the reader will recall that in oligopoly settings the “supply” function is best thought about as a “pricing equation” since it represents the price at which a firm is willing to supply a given quantity of output.\(^\text{22}\) In a homogeneous product industry we will face a single market demand curve and similarly we can derive a single market supply curve. We observe the equilibrium market outcome, the price–quantity combination that equates aggregate firm supply with aggregate consumer demand. This market outcome is the result of factors affecting both the demand and the supply functions.

The theoretical analysis (whose prestigious origins we describe in our discussion of identification in chapter 2)\(^\text{23}\) will show that when a variable affects both supply and demand, we will only be able to separately identify the magnitude of the effect on price–quantity outcomes that occurred through movement of demand and the effect which occurred through movement of supply curve in particular circumstances. In

\(^{22}\) See chapter 1 for a review of the determinants of market outcome. In oligopolistic competition the supply function is referred to as the pricing function to distinguish it from the perfect competition supply function which is given by the marginal cost curve. In oligopolistic markets, the supply function does not trace the marginal cost curve since profit maximization implies pricing above marginal cost at each quantity.

\(^{23}\) For a more formal treatment of identification in linear systems, see, for example, Johnston and Dinardo (1997).
6.2. Directly Identifying the Nature of Competition

contrast, we will usually be able to observe the so-called “reduced-form” effect, that is, the aggregate effect of the movement of the exogenous variables on the equilibrium market outcomes (price, quantity). The reduced-form effects will tell us how exogenous changes in demand and cost determinants affect market equilibrium outcomes, but we will only be able to trace back the actual parameters of the demand and supply functions in particular circumstances.

Let us assume the following demand and supply equations, where $a_t^D$ and $a_t^S$ are the set of shifters of the demand and supply curve respectively at time $t$:

Demand: $Q_t = a_t^D - a_{12} P_t,$
Supply: $Q_t = a_t^S + a_{22} P_t.$

Further, let us assume that there is one demand shifter $X_t$ and one supply shifter $W_t$ so that

$a_t^D = c_{11} X_t + u_t^D$ and $a_t^S = c_{22} W_t + u_t^S.$

The supply-and-demand system can then be written in the following matrix form:

$$\begin{bmatrix} 1 & a_{12} \\ -a_{22} & 1 \end{bmatrix} \begin{bmatrix} Q_t \\ P_t \end{bmatrix} = \begin{bmatrix} c_{11} & 0 \\ 0 & c_{22} \end{bmatrix} \begin{bmatrix} X_t \\ W_t \end{bmatrix} + \begin{bmatrix} u_t^D \\ u_t^S \end{bmatrix}.$$  

Let $y_t = [Q_t, P_t]'$ be the vector of endogenous variables and $Z_t = [X_t, W_t]'$, the vector of exogenous variables in the form of demand and cost shifters which are not determined by the system. We can write the structural system in the form $Ay_t = CZ_t + u_t,$ where

$$A = \begin{bmatrix} 1 & a_{12} \\ 1 & -a_{22} \end{bmatrix} \quad \text{and} \quad C = \begin{bmatrix} c_{11} & 0 \\ 0 & c_{22} \end{bmatrix},$$

and $u_t$ is a vector of shocks

$$u_t = \begin{bmatrix} u_t^D \\ u_t^S \end{bmatrix}.$$  

The “reduced-form” equations relate the vector of endogenous variables to the exogenous variables and these can be obtained by inverting the $(2 \times 2)$ matrix $A$ and performing some basic matrix algebra:

$$y_t = A^{-1} C Z_t + A^{-1} u_t.$$  

Let us define $\Pi = A^{-1} C$ and $v_t = A^{-1} u_t$ so that we can write the reduced form as

$$y_t = \Pi Z_t + v_t.$$  

Doing so gives an equation for each of the endogenous variables on the left-hand side on exogenous variables on the right-hand side. Given enough data we can learn about the parameters in $\Pi$. In particular, we can learn about the parameters using changes in $Z_t$, the exogenous variables affecting either supply or demand.
6. Identification of Conduct

6.2.1.2 Conditions for Identification of Pricing Equations

The important question for identification is whether we can learn about the underlying structural parameters in the structural equations of this model, namely the supply and demand equations. This is the same as saying that we want to know if, given enough data, we can in principle recover demand and supply functions from the data. We examine the conditions necessary for this to be possible and then, in the next section, we go on to examine when and how we can retrieve information about firm conduct based on the pricing equations (supply) and the demand functions thus uncovered.

Structural parameters of demand and supply functions are useful because we will often want to understand the effect of one or more variables on either demand or supply, or both. For instance, to understand whether a “fat tax” will be effective in reducing chocolate consumption, we would want to know the effect of a change in price on the quantity demanded. But we would also want to understand the extent to which any tax would be absorbed by suppliers. To do so, and hence understand the incidence and effects of the tax we must be able to separately identify demand and supply.

As we saw in chapter 2, the traditional conditions to identify both demand and supply equations are that in our structural equations there must be a shifter of demand that does not affect supply and a shifter in supply that does not affect demand. Formally, the number of excluded exogenous variables in the equation must be at least as high as the number of included endogenous variables in the right-hand side of the equation. Usually, exclusion restrictions are derived from economic theory. For example, in a traditional analysis cost shifters will generally affect supply but not demand. Identification also requires a normalization restriction that just rescales the parameters to be normalized to the scale of the explained variable on the left-hand side of the equation.

Returning to our example with the supply-and-demand system:

\[ Ay_t = CZ_t + u_t. \]

The reduced-form estimation would produce a matrix \( \Pi \) such that

\[
\Pi = A^{-1}C = \begin{bmatrix} 1 & a_{12} \\ 1 & -a_{22} \end{bmatrix}^{-1} \begin{bmatrix} c_{11} & 0 \\ 0 & c_{22} \end{bmatrix} = \frac{1}{-a_{22} - a_{12}} \begin{bmatrix} -a_{22}c_{11} & -a_{12}c_{22} \\ c_{11} & c_{22} \end{bmatrix}
\]

so that our reduced-form estimation produces

\[
Q_t = \frac{-a_{22}c_{11}}{-a_{22} - a_{12}} X_t - \frac{a_{12}c_{22}}{-a_{22} - a_{12}} W_t + v_{1t},
\]

\[
P_t = \frac{-c_{11}}{-a_{22} - a_{12}} X_t - \frac{c_{22}}{-a_{22} - a_{12}} W_t + v_{2t}.
\]
6.2. Directly Identifying the Nature of Competition

The identification question is whether we can retrieve the parametric elements of the matrices $A$ and $C$ from estimates of the reduced-form parameters. In this example there are four parameters in $\Pi$ which we can estimate and a maximum of eight parameters potentially in $A$ and $C$. For identification our sufficient conditions will be

- the normalization restrictions which in our example require that $a_{11} = a_{21} = 1$;
- the exclusion restrictions which in our example implies $c_{12} = c_{21} = 0$.

For example, we know that only cost shifters should be in the supply function and hence are excluded from the demand equation while demand shifters should only be in the demand equation and are therefore excluded from the supply equation.

In our example the normalization and exclusion restrictions apply so that we can recover the structural parameters. For instance, given estimates of the reduced-form parameters, $(\pi_{11}, \pi_{21}, \pi_{12}, \pi_{22})$, we can calculate

$$\frac{\pi_{11}}{\pi_{21}} = \left(\frac{-a_{22}c_{11}}{-a_{22} - a_{12}}\right) / \left(\frac{-c_{11}}{-a_{22} - a_{12}}\right) = a_{22}$$

and similarly $\pi_{21}/\pi_{22}$ will give us $a_{12}$. We can then easily retrieve $c_{11}$ and $c_{22}$.

Intuitively, the exclusion restriction is the equivalent of the requirement that we have exogenous demand or supply shifts in order to trace or identify supply or demand functions respectively (see also the discussion in chapter 2 on identification). By including variables in the regression that are present in one of the structural equations but not in the other, we allow one of the structural functions to shift while holding the other one fixed.

6.2.2 Conduct Parameters

Bresnahan (1982) elegantly provides the conditions under which conduct can be identified using a structural supply-and-demand system (where by the former we mean a pricing function). More precisely, he shows the conditions under which we can use data to tell apart three classic economic models of firm conduct, namely Bertrand price competition, Cournot quantity competition, and collusion. We begin by following Bresnahan’s classic paper to illustrate the technique. We will see that successful estimation of a structural demand-and-supply system is typically not enough to identify the nature of the conduct of firms in the market.

---

24 The technical conditions are presented in Lau (1982).
25 We do so while noting that Perloff and Shen (2001) argue that the model has better properties if we use a log-linear demand curve instead of the linear model we use for clarity of exposition here. The extension to the log-linear model only involves some easy algebra. Those authors attribute the original model to Just and Chern (1980). In their article, Just and Chern use an exogenous shock to supply (mechanization of tomato harvesting) to test the competitiveness of demand.
In all three of the competitive settings that Bresnahan (1982) considers, firms that maximize static profits do so by equating marginal revenue to marginal costs. However, under each of the three different models, the firms’ marginal revenue functions are different. As a result, firms are predicted to respond to a change in market conditions that affect prices in a manner that is specific to each model. Under certain conditions, Bresnahan shows these different responses can distinguish the models and thus identify the nature of firm conduct in an industry.

To illustrate, consider, for example, perfect competition with zero fixed costs. In that case, a firm’s pricing equation is simply its marginal cost curve and hence movements or rotations of demand will not affect the shape of the supply (pricing) curve since it is entirely determined by costs. In contrast, under oligopolistic or collusive conduct, the markup over costs—and hence the pricing equation—will depend on the character of the demand curve.

### 6.2.2.1 Marginal Revenue by Market Structure

Following Bresnahan (1982), we first establish that in the homogeneous product context we can nest the competitive, Cournot oligopoly and the monopoly models into one general structure with the marginal revenue function expressed in the general form:

$$
MR(Q) = \lambda Q P'(Q) + P(Q),
$$

where the parameter $\lambda$ takes different values under different competitive regimes. Particularly,

$$
\lambda = \begin{cases} 
0 & \text{under price-taking competition,} \\
1/N & \text{under symmetric Cournot,} \\
1 & \text{under monopoly or cartel.}
\end{cases}
$$

Consider the following market demand function:

$$
Q_t = \alpha_0 - \alpha_1 P_t + \alpha_2 X_t + u^D_t,
$$

where $X_t$ is a set of exogenous variables determining demand. The inverse demand function can be written as

$$
P_t = \frac{\alpha_0}{\alpha_1} - \frac{1}{\alpha_1} Q_t + \frac{\alpha_2}{\alpha_1} X_t + \frac{1}{\alpha_1} u^D_t.
$$

The firms’ total revenue $TR$ will be the price times its own sales. This will be equal to

(i) $TR = q_i P(Q(q_i))$ for the Cournot case,

(ii) $TR = Q P(Q)$ for the monopoly or cartel case,

(iii) $TR = q_i P(Q)$ for the price-taking competition case,
6.2. Directly Identifying the Nature of Competition

where \( Q \) is total market production and \( q_i \) is the firm’s production with \( q_i = Q/N \) in the symmetric Cournot model. Given these revenue functions marginal revenues can respectively be calculated as

(i) \( MR = q_i P'(Q) + P(Q) \) for the Cournot case,
(ii) \( MR = Q P'(Q) + P(Q) \) for the monopoly or cartel case,
(iii) \( MR = P(Q) \) for the price-taking competition case.

All these expressions are nested in the following form:

\[
MR = \lambda Q P'(Q) + P(Q).
\]

6.2.2.2 Pricing Equations

Profit maximization implies firms will equate marginal revenue to marginal costs. Using the marginal revenue expression we obtain the first-order condition characterizing profit maximization in each of the three models:

\[
\lambda Q P'(Q) + P(Q) = MC(Q).
\]

Under one interpretation, the parameter \( \lambda \) provides an indicator of the extent to which firms can increase prices by restricting output. If so then the parameter \( \lambda \) might be interpreted as an indication of how close the price is to the perceived marginal revenue of the firm (see Bresnahan 1981). If so, then \( \lambda \) is an indicator of the market power of the firm and a higher \( \lambda \) would indicate a higher degree of market power while \( \lambda = 0 \) would indicate that firms operate in a price-taking environment where the marginal revenue is equal to the market price. This interpretation was popular in the early 1980s but has disadvantages that has led the field to view such an interpretation skeptically (see Makowski 1987; Bresnahan 1989). More conventionally, provided we can identify the parameter \( \lambda \), we will see that we can consider the problem of distinguishing conduct as an entirely standard statistical testing problem of distinguishing between three nested models.

The pricing equation or supply relation indicates the price at which the firms will sell a given quantity of output and it is determined in each of these three models by the condition that firms will expand output until the relevant variant of marginal revenues equals the marginal costs of production. The pricing equation encompassing these three models will depend on both the quantity and the cost variables. Its parameters are determined by the parameters of the demand function \((\alpha_0, \alpha_1, \alpha_2)\), the parameters of the cost function, and the conduct parameter, \( \lambda \).

Assuming a linear inverse demand function and marginal cost curve, the (supply) pricing equation can be written in the form:

\[
P(Q_t) = \beta_0 + \gamma Q_t + \beta_2 W_t + u_{2t}.
\]

where \( \gamma \) is a function of the cost parameters, the demand parameters, and the conduct parameter, and \( W \) are the determinants of costs.
Given the inverse linear demand function,

\[
P_t = \frac{\alpha_0}{\alpha_1} - \frac{1}{\alpha_1} Q_t + \frac{\alpha_2}{\alpha_1} X_t + \frac{1}{\alpha_1} u_{D1t}^P
\]

and the following linear marginal costs curve:

\[
MC(Q) = \beta_0 + \beta_1 Q + \beta_2 W_t + u_{2t}^S,
\]

where \(W\) are the determinants of costs, then the first-order condition that encompasses all three models, \(\lambda QP'(Q) + P(Q) = MC(Q)\), can be written as

\[
\frac{\lambda}{\alpha_1} Q_t + P(Q_t) = \beta_0 + \beta_1 Q_t + \beta_2 W_t + u_{2t}^S.
\]

By rearranging we obtain the firm’s pricing equation:

\[
P(Q_t) = \beta_0 - \frac{\lambda}{\alpha_1} Q_t + \beta_1 Q_t + \beta_2 W_t + u_{2t}^S,
\]

which can be written in the form that will be estimated:

\[
P(Q_t) = \beta_0 + \gamma Q_t + \beta_2 W_t + u_{2t}^S,
\]

where \(\gamma = \beta_1 - \lambda / \alpha_1\).

We wish to examine the system of two linear equations consisting of (i) the inverse demand function and (ii) the pricing (supply) equation. We have seen in chapter 2 and the earlier discussion in this chapter that we can identify the parameters in the pricing equation provided we have a demand shifter which is excluded from it. Similarly, we can identify the demand curve provided we have a cost shifter which moves the pricing equation without moving the demand equation. In that case, we can identify the parameter \(\gamma\) from the pricing equation and also the parameter \(\alpha_1\) from the demand curve. Unfortunately, but importantly, this is not enough to learn about the conduct parameter, \(\lambda\), the parameter which allows us to distinguish these three standard models of firm conduct. Given \((\gamma, \alpha_1)\) we cannot identify \(\beta_1\) and \(\lambda\) individually.

In the next section we examine the conditions which will allow us to identify conduct, \(\lambda\).

### 6.2.2.3 Identifying Conduct when Cost Information Is Available

There are cases in which the analyst will be able to make assumptions about costs that will allow identification of the conduct parameter. First note that if marginal costs are constant in quantity (so that we know the true value of \(\beta_1\), in this example \(\beta_1 = 0\)), then if we can estimate the demand parameter \(\alpha_1\) and the regression parameter \(\gamma\), we can then identify the conduct parameter, \(\lambda\) since \(\gamma = \beta_1 - \lambda / \alpha_1 = -\lambda / \alpha_1\). Then we can statistically check whether \(\lambda\) is close to 0 indicating a price-taking environment or closer to 1 indicating a monopoly or a cartelized industry. In that special case,
6.2. Directly Identifying the Nature of Competition

The conditions for identification of both the pricing and demand equations and the conduct parameter remains that we can find (i) a supply shifter that allows us to identify the demand curve, the parameter α₁, and (ii) a demand shifter that identifies the pricing curve and hence γ.

Alternatively, if we are confident of our cost data, then we could estimate a cost function, perhaps using the techniques described in chapter 3, or a marginal cost function and then we could equally potentially estimate β₁ directly. This together with estimates of α₁ and γ will again allow us to recover the conduct parameter, λ.

6.2.2.4 Identifying Conduct when Cost Information Is Not Available: Demand Shifts

There are many cases in which there will not be satisfactory cost information available to estimate or make assumptions about the form of firm-level marginal cost functions. An important question is whether it remains possible to identify conduct. Without information about costs, the only market events that one could use for identification are changes in demand. In this section and the next we consider respectively demand shifts and demand rotations and in particular whether such data variation will allow us to recover both estimates of the marginal cost function and also estimates of the demand function. Demand shifts arise, for instance, because of an increase in disposable income available to consumers for consumption. Demand rotations on the other hand must be factors which affect the price sensitivity of consumers. There are many examples, including, for example, the price sensitivity of the demand for umbrellas, which probably falls when it is raining, while the demand for electricity to run air conditioners will be highly price insensitive when the weather is very hot.

First consider demand shifts. We have already established that demand shifters provide useful data variation, helping to identify the supply (pricing) equation. We have also algebraically already shown that such demand shifters are not generally useful for identifying the nature of conduct in the market. In this section our first aim is to build intuition first for the reason demand shifters do not generally suffice to identify conduct. We will go on to argue in the next section that demand rotators will usually suffice.

Suppose that we observe variation in market demand because of changes in disposable income. Such variation in demand will trace out the pricing curve, i.e., the optimal prices of suppliers at different quantity levels. The situation is illustrated in figure 6.2, which shows the changes in price and quantity in a market following a shift in demand from D₁ to D₂. Notice in particular that demand shifts trace out the pricing equation to give data points such as (Q₁, P₁) and (Q₂, P₂), but that such a pricing equation is consistent with different forms of competition in the market. First it is consistent with the firm setting P = MC in a case where marginal costs are increasing in quantity, in which case the “pricing equation” is simply a marginal cost
6. Identification of Conduct

Figure 6.2. Demand shifts do not identify conduct.
Source: Authors’ rendition of figure 1 in Bresnahan (1982).

curve. Second, the same pricing curve could be generated by a more efficient firm that exercises market power by restricting output so that marginal revenue is equal to marginal cost but where marginal revenue is not equal to price. If the pricing curve is the marginal cost curve, then we are in a price-taking environment. If the firm faces a lower marginal cost curve and is setting MR = MC and then charging a markup, the firm has market power. The two ways of rationalizing the same observed price and quantity data are shown in figure 6.2. The aim of the figure is to demonstrate that the demand shift provides no power to tell the two potential underlying models apart (unless we have additional information on the level of costs) even though demand shifts do successfully trace out the pricing equation for us.

6.2.2.5 Identifying Conduct when Cost Information Is Not Available: Demand Rotations

The underlying behavioral assumption in each of the three models considered is that firms maximize profits and to do so they equate marginal revenue and marginal costs. Each of the three models (competitive, Cournot, and monopoly) differs only because they suggest a different calculation of marginal revenue and this has direct implications for the determinants of the pricing curve. Each model places a differential importance on the slope of (inverse) demand for the pricing curve. This can be seen directly from the first term in the first-order condition which describes the pricing equation, \( \lambda Q P'(Q) + P(Q) = MC(Q) \). Alternatively, we can rearrange this equation to emphasize that prices are marginal cost plus a markup which depends
6.2. Directly Identifying the Nature of Competition

on the slope of demand, \( P(Q) = MC(Q) + \lambda Q|P'(Q)| \), differentially across the models.

This equation suggests a route toward achieving identification. Specifically, if a variable affects the slope of demand, then each of the three models will make very different predictions for what should happen to prices at any given marginal cost. For the clearest example, note that in the competitive case absolutely nothing should happen to markups while a monopolist will take advantage of any decrease in demand elasticity to increase prices. Given this intuition, we next consider whether conduct can be identified when the demand curve rotates.

Rotation of the demand curve changes the marginal revenue of oligopolistic firms. Flatter demand and marginal revenue curves will cause firms with market power to lower their prices. On the other hand, price-taking firms will keep the price unchanged since lowering the price would cause them to price below marginal cost and make losses. Figure 6.3 illustrates this point graphically by considering a demand rotation around the initial equilibrium point, \( E_1 \). In particular the figure allows us to compare the lack of reaction of a price-taking firm, which starts and finishes with prices and quantities described by \( E_1 \), with the response of the monopolist who begins at \( E_1 \) but finishes with different price and quantities, those at \( E_2 \), after the demand rotation.

Intuitively, demand rotations allow us to identify conduct even when we have no information about costs because such changes should not cause any response in a perfectly competitive environment, there should be some response in a Cournot market and a much larger response in a fully collusive environment. If demand

![Figure 6.3. Reactions of competitive firm and monopolist to a demand rotation. Source: Authors’ rendition of figure 2 in Bresnahan (1982).](image-url)
bears more elastic, prices will decrease and quantity will increase in a market with a high degree of market power. If, on the other hand, demand becomes more inelastic and consumers are less willing to adjust their quantities consumed in response to changes in prices, then prices will increase in oligopolistic or cartelized markets. Prices will remain unchanged in both scenarios if the market is perfectly competitive and firms are pricing close to their marginal costs.

While intuitive, a simple graph cannot show that given an arbitrarily large amount of data a demand rotator is sufficient to tell apart the three models, which is the statement that we would like to establish for identification. We therefore examine the algebra of demand rotations.

Let us look at the algebra of identification using the demand rotation. Formally, we can specify a demand function to include a set of variables \( Z \) that will affect the slope (and potentially the level) of demand:

\[
Q_t = \alpha_0 + \alpha_1 P_t + \alpha_2 X_t + \alpha_3 P_t Z_t + \alpha_4 Z_t + u_{D1t}.
\]

For our three models the encompassing pricing equation becomes

\[
P_t = \left(\frac{-\lambda}{\alpha_1 + \alpha_3 Z_t}\right) Q_t + \beta_0 + \beta_1 Q_t + \beta_2 W_t + u_{2t}^S.
\]

To consider identification note that if we can estimate demand and retrieve the true parameters \( \alpha_1 \) and \( \alpha_3 \), then we can construct the variable \( Q^* = -Q / (\alpha_1 + \alpha_3 Z) \). In that case, the conduct parameter will be the coefficient of \( Q^* \) when estimating the following equation:

\[
P_t = \beta_0 + \lambda Q^*_t + \beta_1 Q_t + \beta_2 W_t + u_{2t}^S.
\]

An important challenge in the demand rotation methodology is to identify a situation where we can be confident that we have a variable which resulted in a change in the sensitivity of demand to prices. On the other hand, a nice feature of the demand estimation method is that when estimating the demand curve we can test whether a variable actually does rotate the demand curve or whether it merely shifts the curve. Events that may change the price elasticity of a product at a particular price include the appearance of a new substitute for a good or a change in the price of the main substitutes. For instance, the popularization of the downloading of music through the internet may have increased the elasticity of the demand for physical CD players because consumers may have become more price sensitive and more willing to decrease their purchases of music CDs in the case of a price increase. In the case of digital music, one might expect that there has been both a demand rotation and a demand shift so that at given prices, the demand for physical CDs has dropped. Only the demand rotation will help us identify conduct. Similarly, weather may affect both the level of demand for umbrellas and also demand may be less elastic
6.2. *Directly Identifying the Nature of Competition*  

When it is raining. While there is no theoretical difficulty if the same variable affects both the level and the slope of demand, we may run into the practical difficulties associated with multicollinearity, which may make telling apart the demand shift and the demand rotation rather hard empirically. Empirical work is challenging and also requires creativity.

A second important practical issue is the difficulty of explaining a somewhat technical issue to a nontechnical legal audience. However, this can be overcome by understanding the principles and explaining them correctly in plain language. By using demand rotators, we are trying to use the fact that firms with market power will adjust to changes in the level of their market power while firms with no market power will price close to marginal cost and will not react to changes in the level of demand elasticities. Firms pricing close to marginal cost will not react to changes in the price sensitivity of demand while firms with some degree of market power will adjust their prices to such changes, according to these models.

A third issue is whether to estimate $\lambda$ or test models with particular values against one another. If we estimate $\lambda$, we will rarely (or never) get values of 0 or 1 but most likely something between the two. In practice, we would get an estimate of, say, $\lambda = 0.234352$ and we could then test the hypotheses that $\lambda = 0$ or $\lambda = 1$ or $\lambda = 1/N$, where $N$ is the number of firms, since we know that these correspond to competition, perfect collusion, and the Cournot model. For example, we could test whether the data suggest that the parameter value is more likely to be one or another value of the parameters using, for example, a likelihood ratio test (see, in particular, Vuong 1989). Such an approach allows us to tell whether the data are consistent with one of the three models given enough data.

The reason to prefer the specific values of $\lambda$ is that we are usually really trying to test which of the three specific models best fit the data since it can be difficult to draw a specific conclusion on a value of $\lambda$ between 0 and 1 that does not equal any of the values predicted by the theory models we have outlined. Specifically, we do not usually have a model which corresponds directly to an estimated value of a number like $\lambda = 0.234352$. For that reason most researchers prefer to test between the perfect competition, the perfect cartel model, and the symmetric Cournot model rather than over-interpreting intermediate values of $\lambda$. That said, in a challenge to that practice, Kalai and Stanford (1985) do present a model which may rationalize a continuum of equilibrium solutions between the competitive and monopoly outcomes.

Finally, we note the difficulties researchers face when identifying marginal costs using first-order conditions derived from theoretical models, particularly when the theoretical model involves some level of market power. The estimation approach we described implies that a researcher is able to identify both demand and supply equations, and subsequently marginal costs. There are some mixed assessments of our ability to identify marginal costs using first-order conditions derived from theory. Genovese and Mullin (1998) test this methodology by comparing costs implied by
the estimated conduct and demand structure with the actual cost data in the cane sugar refining industry in the late nineteenth century and early twentieth century in the United States. They first find that the estimated conduct parameter using no cost data is not too different from the one derived using actual cost information. The estimated costs will nevertheless be very sensitive to the imposition of a particular static model of competition. The authors defend the usefulness of defining a “loose” conduct parameter in the specification of the pricing equation. Corts (1999) and Kim and Knittel (2006) have less enthusiastic assessments of the accuracy of the estimated costs when a particular competitive setting is imposed. The estimated marginal costs, those consistent with the estimated demand elasticities and price levels, will sometimes be negative. The reason is clear: if demand is estimated to be inelastic but observed prices are actually fairly low, then margins can be predicted to be so high that the only marginal costs that can rationalize the high margins would be negative. In a recent paper Kim and Knittel (2006) find that the conduct parameter technique poorly estimates markups and markup adjustments to cost shocks in the California electricity market.

Corts argues that the estimation of conduct parameters in the above methodology will often fail to measure market power accurately not least because the model of perfect collusion Bresnahan emphasizes is not motivated from a specific dynamic pricing model of collusion and moreover it is only one of many potential models of collusion (other models of collusion may have features such as price rigidity making such exercises likely to be problematic). Salvo (2007) argues that unobserved constraints faced by firms can limit their pricing levels resulting in an underestimation of their ability to react to price changes following changes in demand conditions. Concretely, he shows that threat of entry kept the prices of a cement industry cartel in Brazil lower than would have been predicted by its documented market power. The conduct parameter technique miscalculates the costs and underestimates the degree of market power in that particular case. On the other hand, Salvo provides a potential solution to the threat of entry difficulty while Puller (2006) and Kim (2005) each suggest a solution to at least one element of the Corts critique.

In summary, the objective of this branch of the industrial organization literature is to facilitate our ability to test between the various models of firm behavior to see which best matches the data. In order to test one model against the other we must have some appropriate sources of identifying data variation. In the case we examined the sources of the required data variation were isolated as (1) demand shifters, (2) cost shifters, and (3) demand rotator(s). In all but very special circumstances all three were required.

More generally, the main theoretical and practical challenge to such an approach is to understand the kind of data variation that will help distinguish one economic model from another and then find an actual variable or set of variables which provide that source of data variation in the particular case at hand. While the homogeneous product Bertrand, Cournot, and perfect collusion cases studied by Bresnahan are
now well-understood, the challenge to develop a raft of identification results for standard industrial organization models has not been widely taken up by the industrial organization academic community and there are numerous important examples of identification results which remain to be explored and tested. For example, one case that regulators and competition authorities should certainly like to understand would involve identification results for the difference between Ramsey and monopoly prices. Identification results exist for only a relatively small subset of standard industrial organization models. For that reason a major and important topic for future research in industrial organization involves the study of identification.

6.2.3 Identifying Tacit Collusion

Collusion occurs when firms in an industry coordinate to maximize (or at least increase) joint industry profits as opposed to individual profits. In standard models of oligopolistic competition, firms maximize their own profits and ignore the consequence of their actions on competitor’s profitability. As a result of this fundamental horizontal externality, whereby a firm takes actions (e.g., increases output or cuts prices) without any consideration of the negative impact on its competitors’ profits, total industry profits are not maximized and firms will end up producing more and at lower prices than if they were acting together in a concerted fashion. Thus economic theory argues that selfish actions by individual firms are (i) ultimately self-defeating and (ii) ultimately generate great benefits for consumers in the form of lower prices and higher output.

In any discussion of collusion, it is useful to distinguish between a cartel or explicit collusion and tacit collusion. In an explicit cartel, firms will directly communicate with each other about their expected behavior and reactions and will jointly decide on the market outcome. In contrast, under tacit collusion, there will be no explicit communication, but firms will nonetheless understand their rivals’ likely reactions when setting output and prices. If a sufficiently large fraction of the players in an industry understand that selfish behavior will ultimately be self-defeating and they also understand that their rivals understand that, we may find that coordinated behavior emerges even without the need for explicit communication. Under such tacit collusion, the expected reaction of competitors to moves in prices or output

---

26 One area where this line of research—the development of identification results—has been more active is the auction literature (see, for example, Athey and Haile 2002).

27 For an extensive discussion of the determinants of the success of cartels, see the edited volume by Grossman (2004). For a detailed discussion of three prominent U.S. cases during the 1990s (the lysine, vitamins, and citric acid cartels), see the account by Connor (2001). The title comes from an infamous quote by James Randall, President of Archer-Daniels-Midland of the United States during a meeting with fellow lysine cartel members Anjinimoto Co. of Japan in 1993. Mr. Randall was captured secretly on tape by another ADM employee (who had signed an agreement with the FBI to be an informant in their investigation). A fuller version reads (see Eichenwald 1997, 1998): “We have a saying at this company,” said Mr. Randall. “Our competitors are our friends and our customers are our enemies.”
will be to follow these moves. Firms may succeed in tacitly coordinating using signaling of strategies through media, suppliers, or customers and perhaps also engage in occasional punishments so that, without needing explicit communication, firms end up pricing in ways that increases margins and total industry profits. Informal evidence of both tacit and explicit collusion can emerge from company pricing or strategy documents.

Legally, the treatment of the two forms of collusion is radically different as cartels are per se illegal and even criminalized in many jurisdictions (including the United States, the United Kingdom, Israel, Korea, and Australia) while tacit collusion is not typically criminalized and yet would, at least in principle, be subject to antitrust enforcement. For example, in the European Union, some forms of tacit collusion could be covered by Article 81, which prohibits “concerted practices.” In addition, tacit collusion would be included in the concept of “collective dominance,” which has been interpreted by the courts as a particular form of “dominance” and abuse of dominance is, for example, prohibited under Article 82. In addition, mergers that are thought to result in an increase of “collective dominance” are forbidden in EU law. Furthermore, sector inquiries (in the EU) and in particular market inquiries (in the United Kingdom) can be used to target industries where such behavior is suspected.

The legal distinctions between tacit and explicit collusion may reflect economic reality since explicit and tacit collusion differ in the sense that the form and nature of collusion are typically explicitly agreed between the players in a cartel, so that it may be more effective at raising prices or restricting output than a collection of firms that are only tacitly colluding. Specifically, tacit colluders must find ways to convey sufficient information to each other indirectly, and they must overcome uncertainty about the extent to which rivals are “playing along” since the kind of direct—perhaps face-to-face or even evidenced with independent accounting reviews—reassurance possible in a cartel will not generally be possible for tacit colluders. Such communication difficulties may diminish either the effectiveness of the collusive arrangement or its longevity. The lack of direct communication may in particular reduce a tacitly colluding set of firms’ ability to react optimally to changes in market conditions.

Both cartels and tacitly collusive accommodations can be unstable. Successful coordinated behavior will generate high prices, high margins, and low output and

---

28 See, in particular, *Laurent Piau v. Commission* T-193/02, which confirms that collective dominance can be a form of dominance for Article 82, a view already existent in the EC merger regime following Airtours. On the other hand, a tacit collusion case has not arisen yet and indeed it would be an unusually difficult case since it would simultaneously be both a (i) “collective dominance” case and (ii) an “exploitative abuse” case (i.e., prices are high). Each form of case is rare. Specifically, *Laurent Piau v. Commission* involved a football industry association, FIFA, which introduced structural links between companies, whereas a tacit collusion case would not involve direct linkages. Furthermore, exploitative abuse cases against (single) dominant firms are rare in comparison to “exclusionary abuse” cases such as those involving predatory pricing. Thus it seems a pure tacit collusion case could in principle now be developed, but would need to overcome two potentially very difficult hurdles.
as a result every firm will have a private short-run incentive to increase its sales to
take advantage of the higher margin. But it must do so undetected so that there are
no reactions by competitors to eliminate the benefits of the deviation. If competitors
respond by increasing their own output and causing prices to drop to competitive
levels, the benefits of the deviation and thereby the incentives to deviate disappear.
The potential lack of stability of a collusive agreement is therefore related to the
likelihood that firms can carry out deviations that are both significant and undetected
or a detectable deviation that brings enough profits to more than compensate for the
losses of the cartel benefits. On the other hand, game theorists since the 1970s
have demonstrated that there do exist credible punishment mechanisms that can
eliminate incentives to deviate from a collusive agreement and result in stable tacitly
collusive equilibria. Furthermore, some “stable” agreements are of rather complex
appearance. For example, some will involve recurrent periods of apparent “price
wars” but in fact these are just one part of the stable agreement designed to deal
with episodic periods of low demand resulting in low prices (Green and Porter 1984).

Either form of collusion in an industry harms consumers because it drives market
prices up (and output down) toward monopoly outcomes where firms can extract
much of the value generated by market activity to the detriment of consumers. It is,
however, difficult to detect collusion when evidence of explicit collusion is missing
or does not exist. How do we identify cartelized behavior from price competition?
How do we distinguish tacit collusion from legitimate oligopolistic competition?

6.2.3.1 Difficulties in Directly Identifying Tacit Collusion

Identifying tacit collusion or the likelihood of tacit collusion is notoriously difficult.
One direct approach to showing the existence of collective dominance is to attempt to
establish the extent to which any firm’s price is based on market demand sensitivity
to price changes as distinct from the firm’s own demand sensitivity to price changes.

To understand the logic of this direct approach, consider first that an indication
from company documents that a firm’s prices are being set with the reaction of
consumers in mind is an indication of market power (although every firm has some
degree of market power and not every firm is involved in pricing behavior of concern
to competition authorities). If the prices of an individual firm are found to be set
taking into account the anticipated full extent of the reaction of market demand
distinct from their own firm’s demand, then we may have an indication of a
collusive industry. Indeed, on the face of it, if the firm monitors and takes into
account the effect of its actions on other market participants profitability, then we
potentially have direct evidence for tacit or explicit collusion. In practice, such
evidence must be interpreted carefully as many firms will engage in monitoring
of rivals’ behavior and this may be normal strategic behavior as distinct from the
kind of dynamic strategic behavior that results in collusive outcomes. Evidence of

29 This is formalized in Friedman (1971) and Abreu (1986).
monitoring rivals is certainly not in itself evidence of tacit collusion. Rather we must find evidence that the firm is taking, or attempting to take, decisions which actively accommodate its rivals’ needs and in particular their likely profitability. Such direct evidence may be available from company documents or testimony, but even apparently direct documentary evidence can appear ambiguous given the intervention of skilled legal professionals. Evidence may also be available from econometric analysis (following the approach to identifying collusion outlined in the first part of this chapter which emphasized the power of “demand rotators” for identification in simple models) but again such evidence is rarely unambiguous. The difficulty in making these distinctions in practice should not be understated.

To further understand the difficulties in establishing tacit collusion directly, note that firms may tacitly collude with varying degrees of success. First, if firms are heterogeneous, they may not gain much directly from the optimal tacitly collusive action. For example, consider that a two-plant monopolist may sometimes minimize costs by using only its most efficient plant and not its inefficient plant. A tacitly collusive arrangement between two single-plant firms in which one firm produced nothing would probably be difficult to sell to the owner of the unused plant, at least without some form of (possibly indirect) side-payments between players, perhaps through industry associations, shared industry-level advertising, or commercial activities in other markets. Second, the world changes and tacitly colluding firms must have a strategy for dealing with change. For example, demand or costs may be high or low and, in a standard model of firm behavior, collusive prices would change with costs and demand conditions. If so, then tacitly colluding firms may need to re-establish a new tacit agreement about the level of collusive prices fairly frequently. However, if change threatens stability, then collusive arrangements may well involve only very infrequent changes in pricing or market territories. For each of these reasons the outcomes of a tacitly collusive arrangement can be somewhat or greatly distinct from either competitive or perfectly collusive outcomes.

We have already mentioned the critique of the econometric attempts to measure market power provided in Corts (1999). However, the critique in large part also applies to noneconometric evidence. Fundamentally, the problem is that dynamic game theory has only succeeded in showing that tacit collusion may be a sustainable market outcome and then provided us with a wide variety of examples of (potentially complex) pricing strategies that could result. The theory has not then yet provided a comprehensive “identification” strategy for distinguishing general classes of models of collusion from models of competition. Numerous market histories appear consistent with collusion and yet also appear consistent with other competitive environments. For example, collusion can produce stable prices or a succession of price wars depending on the level of uncertainty or the nature of the punishments. Collusion may also produce procyclical or countercyclical prices depending on, for example, capacity utilization levels or whether we are at turning points of business.
cycles or not. Some consensus has emerged on the conditions that are more likely to promote collusion: small numbers of players, stability of demand, and firm symmetry. But these characteristics are mostly indicative as collusion is still possible when these characteristics are absent. For example, symmetry will rarely be the case in differentiated product markets and, we shall see, firm asymmetry makes collusion harder in at least one important sense, but on the other hand does not typically rule out situations arising when collusion can nonetheless be sustained.

Because of the apparently weak predictive power of economic theory with regards to the exact manifestation of collusion, most empirical casework to detect collusion has centered on showing that the very basic conditions that are necessary for collusion to exist can be found in a given market. The presumption is that if these necessary conditions exist (so that firms have both the ability and incentive to collude), then collusion is likely. The analysis of coordination in antitrust settings currently tends to consist of analysis of the three essential points introduced by Stigler (1964) nearly fifty years ago, which we present below.

6.2.3.2 Assessing the Conditions for Agreement, Monitoring, and Enforcement

Stigler (1964) provided a general framework for evaluating the features of a market which are likely to facilitate the movement toward coordination. Subsequently this framework has largely been adopted in most jurisdictions, although the exact terminology varies from guidance to guidance. It relies on the conclusion that for collusion to be viable, it must be feasible for participants to reach an agreement on the terms of coordination; it must also be possible to monitor that this agreement is being respected by the colluding firms; and deviating firms must be punished and, in the case of tacit collusion, it is the credibility of this punishment mechanism that holds the collusive agreement together, i.e., enforces it. The framework is equally applicable for explicit or tacit collusion, but the form of each element can differ. In the case of cartels for example, agreement may be arrived at by discussion, monitoring may occur by exchange of information, perhaps even independent reports by accounting firms and/or trade associations, while enforcement may in some cases remain via similar mechanisms to those emerging from tacit collusion. In others the mechanism may be quite different. For example, in the extreme case of legal cartels, enforcement may result from contract enforcement via the courts. It is worth noting that export cartels remain legal in a number of jurisdictions. We next discuss each element of Stigler’s framework in turn.

---

30 See, in particular, Rotemberg and Saloner (1986) and Haltiwanger and Harrington (1991). See also García et al. (2009) for a brief review of the subsequent collusion literature.

31 For a summary of the literature, see Ivaldi et al. (2003).

32 For example, the categories Agreement, Monitoring, and Enforcement are sometimes replaced with the terms Consensus, Detection, and Punishment.

33 For example, in the lysine case, sales were reported to a trade association and each year a firm of accountants audited the sales numbers in both London and Decatur, IL.
Agreement. Colluders must reach some form of understanding about what exactly it means to coordinate. This means that there must be an understanding of the dimensions on which coordination is taking place as well as an indication of the expected behavior. In tacit collusion the agreement will not be explicit but will have to be inferred by market players from the information available to them. Firms can publicize their price lists and make public announcements to provide the market with an indication of a potential focal point around which behavior will be coordinated. These signaling practices are normally frowned upon by market authorities when they suspect collusive behavior, but on the other hand publishing price information is not uncommon and in other circumstances is actively encouraged by competition authorities, for example, to facilitate consumer search. Focal points may also be inferred from past behavior or historical prices and in such cases markets may tend to exhibit stronger degrees of price rigidity. A market with complex transactions or with customized transactions will be less susceptible to firms being able to find a mutually acceptable understanding of what it means to tacitly collude. Similarly, a market with very diverse products such as different brands and different versions of a particular product will be more difficult to coordinate. Since complexity makes agreements about what it would mean to collude difficult to achieve, sometimes we see firms adopting practices that “simplify their prices for consumers” or harmonize the conditions for a transaction. For an example of a pricing structure which might be considered by some authorities to potentially facilitate collusion, recall that at one stage some U.S. airlines proposed using per-mile pricing so that every route between every city would be easy to price by all parties. Such initiatives may have the ultimate purpose of facilitating a collusive outcome since coordination largely reduces to tacitly agreeing on a single number, the per-mile price. Finally, when firms have very different incentives, perhaps because of differences in scale or efficiency, it will be harder to get everyone to agree to a particular market outcome. It may be easier to evolve toward agreements in industries where change occurs only slowly as it is not always obvious for firms to understand or agree on a coordinated response to change.

In a coordinated effects merger case it is desirable but probably should not be necessary to say exactly what the form of a coordinated agreement might look like, since it is unlikely that a competition authority will put the same effort into finding an ingenious solution to a difficult problem as the companies involved, should they have a sufficiently strong incentive to cooperate. For this reason, most competition authorities do not give quite the same weight to the agreement element of Stigler’s framework in their guidelines as they do to the monitoring and enforcement areas. Even explicit agreements can be incredibly difficult to uncover. In the famous “phases of the moon” cartel case, twenty-nine colluding firms in the market

---

34 See, for example, O’Brian (1992). To see that such proposals may not succeed, see, for example, McDowell (1992).
6.2. Directly Identifying the Nature of Competition

for electrical equipment led by the two giants General Electric and Westinghouse literally devised a codebook of lists of numbers which determined how much each company in the cartel would bid on a particular contract. The price spread was geared toward giving an impression of competition and the fact that the price spreads across companies were cyclical led to the cartel being known as the “phases of the moon” cartel. That particular cartel lasted seven years and rigged bids estimated to be worth a total of $7 billion.35

Monitoring. Dynamic oligopoly theory suggests that for coordination firms must be aware of the behavior of their competitors. They must be able to observe it or at least to infer it with certain degree of confidence. In particular they must be able to spot deviations from prevailing behavior in order that “cheaters” from the coordinated prices can be spotted. Monitoring will be harder in markets where prices and/or quantity choices are difficult to observe, demand or cost shocks are large, or when orders are lumpy and as a result both prices and quantities tend to be volatile. But it has been argued in the economics literature that tacit collusion can certainly occur without full transparency. Specifically, the literature emanating from Green and Porter (1984) has shown that tacit collusion is possible even without full monitoring of firms’ prices and quantities. For example, a strategy that would temporarily revert to a price war every time market prices fell below a threshold can sustain tacit collusion.36 In this case, tacit collusion would take the form of alternating phases of price stability and price wars.

In spite of these contributions, the issues of transparency, complexity, and the ability to monitor competitors’ actions and prices are usually considered very important for a finding of collusion or coordinated effects. It is possible to look at the extent of monitoring and the extent of both complexity and transparency of information both through interview evidence and documentary evidence. Price lists, price announcements, and industry association publications are clear ways of announcing one’s behavior but more may be needed to detect small-scale deviations. List prices or “price books” can sometimes facilitate coordination because they can dramatically improve the amount of information available to rivals. If customers mainly pay list prices, or list prices are highly correlated with transaction prices (in the extreme, transaction prices may be some fixed discount from list prices), then such price lists may help firms find their way toward coordination. Price lists need not be paper price lists and in some famous examples the price lists have been electronic. For example, in the U.S. Airline Tariff Pricing case, participating U.S. airlines could post nonbinding ticket prices for particular routes that were for an initial period unavailable to customers. In fact, they used features of the electronic fare system

---

35 For a wonderful description of what has become known as the great electrical conspiracy, see “The great conspiracy,” Time Magazine, February 17, 1961.
36 For the first test of the Green and Porter model, see Porter (1983).
6. Identification of Conduct

as signaling devices. Baker (1996) provides an interesting commentary on information exchange in cyberspace. However, before condemning price lists, one must keep in mind that, at their best, price lists can hugely improve the information available to consumers which in turn can save consumer search costs, increase the price sensitivity of demand, and encourage firms to charge lower prices than their rivals.

Information flows between customers and suppliers in the case of stable customer–supplier relationships can be an important way of getting exact market information particularly when customers shop from different suppliers. The visibility of contracts and of changes in market shares is useful to detect potential deviations. Investigators should certainly invest in assessing the level of transparency and monitoring mechanisms that may imply that a coordinated outcome is viable.

Scheffman and Coleman (2003) provide a nice summary of the kinds of empirical work that may be undertaken to assess coordination. Those authors emphasize that coordination can happen in a number of ways and may involve coordination on prices, quantities, capacities, or some form of market division, say, by territory or type of customer. As a result many of the following remarks while phrased in terms of prices are equally relevant to other potential dimensions of coordination. Scheffman and Coleman suggest, for example, that we may wish to look empirically at the following:

1. Differences or patterns in the relationship between list and transaction prices.

Figure 6.4 provides an example where list prices have little predictive power for actual transaction prices. In this case, list prices do not carry enough information about actual market prices and cannot be used as a monitoring device.

Figure 6.4. List prices versus actual prices.

Source: Scheffman and Coleman (2003, figure 4).


---

6.2. **Directly Identifying the Nature of Competition**

<table>
<thead>
<tr>
<th></th>
<th>Competitor Y</th>
<th>Competitor Z</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of customers that company X identifies as supplying</td>
<td>55</td>
<td>46</td>
</tr>
<tr>
<td>identifies as supplying when did not</td>
<td>22</td>
<td>12</td>
</tr>
<tr>
<td>does not identify as supplying when did</td>
<td>12</td>
<td>8</td>
</tr>
<tr>
<td>Percentage of customers for whom company X’s volume estimate was off by more than 20%</td>
<td>75%</td>
<td>82%</td>
</tr>
<tr>
<td>volume estimate was off by more than 60%</td>
<td>39%</td>
<td>47%</td>
</tr>
</tbody>
</table>

*Source: Scheffman and Coleman (2003, figure 5).*

2. Variation in prices across consumers, controlling for observable differences in the type of customer or order behavior in terms of volume or location. We can look at the coefficient of variation and range of prices paid by various customer types. To that end a transaction-level regression of price on volume, location, and customer characteristics may be run in order to understand and evaluate the extent of variation in prices across customers or customer groups.

3. Variation in transaction prices within customer for the same product across different suppliers. We may also want to look at the percentage of instances where prices to the same customer by different suppliers differ by, say, more than 5%. We might, for example, want to break that down by customer type.

4. Variation in changes in transaction prices across customers again controlling for observable differences.

As with all such studies it is vital to bear in mind that the mere existence of co-movement in, say, list and transaction prices does not prove coordination since we would expect co-movement to result for innocent reasons such as cost variation. However, the basic intuition that such analysis relies on is that if significant variation in a firm’s price changes is found, we might expect that coordinated interaction is likely to be more difficult. We examine this approach further (see section 6.2.3.4) by looking at the European Commission’s empirical evidence in the Sony–BMG merger case.

We may also want to look at transparency directly by comparing one company’s estimates of competitors’ volumes versus their competitors’ actual volumes. Such an analysis is provided in table 6.5, which shows that competitor X’s estimates were quite considerably different from the truth.

**Enforcement.** In the theory of tacit collusion, enforcement action involving members of the cartel (internal enforcement) takes the form of the threat of a credible punishment directed at either a deviating firm or in a nontargeted fashion at all firms if they move away from the tacitly collusive outcome when a deviation is detected.
A successful punishment regime will eliminate the potential gains from cheating on other participants. When cheating on a collusive agreement is easily detected and a credible punishment exists for such behavior, tacitly collusive environments are predicted to be stable. Moreover, in some (at least theoretical) environments, no actual punishments need ever be observed which may make detection by competition authorities rather difficult.

On the other hand, while many theoretical models generate tacit collusion rather easily, it does seem that even explicit cartels, where direct communication is possible, do certainly break down. In a review of a large set of known cartels, Suslow and Levenstein (1997) find that the average longevity of an explicit cartel is about five years but that the distribution is bimodal: while some cartels last for decades, many others last for less than a year.

In addition to a mechanism that enforces internal stability of a collusive arrangement, there must be some form of mechanism for enforcing “external” stability. In particular, all else equal, high profits will soon attract new entrants so that it will be necessary to have either actual barriers to entry or an ability to punish entrants so as to deny them returns (in the sense of profit) following entry. For example, in the lysine case, a cartel member, Archer-Daniels-Midland, quickly built a new plant as part of strategy to deter a new entrant (Connor 2001). For tacit collusion to be an antitrust problem an industry must be able to benefit from both internal and externality stability.

In addition to suggesting that a credible punishment mechanism is important, economic theory does make some suggestions regarding the nature of such punishments. One particularly simple punishment involves the reversion to static competition. The theory suggests that the threat of a permanent or even temporary price war can be an effective punishment provided cartel participants are sufficiently patient and such punishments may sometimes involve “harsher” punishments than reversion to the competitive price. Such theoretical results suggest that a key variable linked to the effectiveness of punishment is the ability of the punishing firms to rapidly expand output so that prices fall sharply enough to generate the losses that will deter opportunistic deviation. As a result there is an important literature on the role of excess capacity both on the incentives to cheat and the ability to punish. Excess capacity is generally considered to facilitate tacit coordination (see, for instance, Brock and Scheinkman 1985; Davidson and Deneckere 1990). Highly asymmetric holdings of capacity on the other hand probably, but not necessarily, hinder collusion (Compte et al. 2002; Vasconcellos 2005).

Other forms of punishment can exist particularly in multiproduct markets, although Bernheim and Whinston (1990) suggest that multimarket contact is actively helpful to sustaining collusion in the presence of firm or market asymmetry (Bernheim and Whinston 1990). Such asymmetry seems likely to arise fairly generically in

---

38 See Abreu et al. (1990). Harsher punishments can involve prices below the competitive levels and stability can sometimes be maintained by using harsh but fairly short punishments.
6.2. Directly Identifying the Nature of Competition

real world markets making multimarket contact potentially a relevant consideration. Intuitively, under perfect firm and market symmetry, the incentive to collude and the incentive to cheat for all firms in all markets will be identical so that multimarket contact adds little. However, with firm and/or market asymmetry, the incentives for collusion and cheating will generally differ across firms in multimarket contexts. Within market, firm asymmetry means that different firms must each find collusion attractive. Multimarket contact means that incentive constraints will be evaluated in total across markets rather than within any individual market. As a result, punishments, for example, might be targeted to greatest effect.

Punishment mechanisms should be effective not only at deterring participating firms in an industry from cheating (internal stability) but also at deterring potential entrants in the market (external stability). Because it is difficult to discipline a very large number of firms that could enter at any time in an industry, tacit collusion will be more effective in markets that exhibit some barriers to entry. Indeed, in their review of the case history, Suslow and Levenstein (1997) find that, while cartels do sometimes break up occasionally because of cheating by incumbents, entry and an ability to react to changes in market positions pose a greater problem. Relatedly, not all firms in an industry will necessarily be involved in a particular cartel and if customers of those which are in a cartel can react by switching to nonparticipating suppliers, then that will help destabilize a collusive equilibrium.

While Stigler (1964) introduces the agreement, monitoring, and enforcement framework we have described, there is an important question as to the extent of analysis necessary about the form of the likely agreement. In particular, the summary of the European Court of First Instance judgment in the Airtours case reads:39

Three conditions are necessary for the creation of a collective dominant position significantly impeding effective competition in the common market or a substantial part of it:

- first, each member of the dominant oligopoly must have the ability to know how the other members are behaving in order to monitor whether or not they are adopting the common policy. In that regard, it is not enough for each member of the dominant oligopoly to be aware that interdependent market conduct is profitable for all of them but each member must also have a means of knowing whether the other operators are adopting the same strategy and whether they are maintaining it. There must, therefore, be sufficient market transparency for all members of the dominant oligopoly to be aware, sufficiently precisely and quickly, of the way in which the other members’ market conduct is evolving;

- second, the situation of tacit coordination must be sustainable over time, that is to say, there must be an incentive not to depart from the common policy on the market. It is only if all the members of the dominant oligopoly maintain the parallel conduct that all can benefit. The notion of retaliation in respect of conduct deviating from the common policy is thus inherent in this condition.

---

In that context, the Commission must not necessarily prove that there is a specific retaliation mechanism involving a degree of severity, but it must none the less establish that deterrents exist, which are such that it is not worth the while of any member of the dominant oligopoly to depart from the common course of conduct to the detriment of the other oligopolists. For a situation of collective dominance to be viable, there must be adequate deterrents to ensure that there is a long-term incentive in not departing from the common policy, which means that each member of the dominant oligopoly must be aware that highly competitive action on its part designed to increase its market share would provoke identical action by the others, so that it would derive no benefit from its initiative;

– third, it must also be established that the foreseeable reaction of current and future competitors, as well as of consumers, would not jeopardise the results expected from the common policy.

Broadly, the first condition relates directly to monitoring, while the second and third relate directly to internal and external enforcement. Thus, the agreement element of Stigler’s framework is played down in the current EU legal environment presumably for reasons we have discussed earlier in this section.

In establishing these conditions, the competition case handler will need to examine carefully the specific facts about an industry, understanding the nature of multimarket contact, the extent of asymmetry, the lumpiness or orders, and so forth. An analyst would also go on to attempt to understand at least qualitatively the incentives of firms in an industry to sustain collusion and hence their ability to do so before she is able to conclude whether tacit collusion is likely or unlikely to be viable.

6.2.3.3 Other Evidence Potentially Relevant to an Inference of the Presence of Tacit Coordination

The issue of whether mergers are likely to increase the likelihood of tacit collusion will most certainly consist of an assessment of the evidence regarding the three elements discussed above, in particular in Europe as determined by the Court of First Instance’s Airtours decision of 2002. Regarding the assessment of existing tacit collusion, the Court for First Instance in its Impala judgment said that:

…in the context of the assessment of the existence of a collective dominance position, although the three conditions defined by the CFI in Airtours v. Commission … are indeed also necessary, they may, however, in the appropriate circumstances, be established indirectly on the basis of what may be a series of indicia and items of evidence relating to the signs and manifestations and phenomena inherent in the presence of a collective dominant position. (§251 Impala v. Commission)\(^{40}\)

The European Court of Justice, in its annulment of the CFI decision, upheld the right of the court to freely assess different items of evidence. It also argued against the mechanical application of the so-called Airtours conditions detailed above but

rather asked for these criteria to be related to an “overall economic mechanism of a hypothetical tacit coordination.” So that any evidence pointing to tacit collusion is admissible but a realistic mechanism of collusion consistent with the economic theory of collusion must also be laid out.

This can be understood as an invitation to use available evidence to directly identify a collusive outcome as distinct from the outcome generated by a competitive oligopoly. We have already seen that this is very difficult to do due in part to the lack of a wide variety of predictions that emerge from the theoretical framework for tacit collusion. It is particularly important to keep two factors in mind. First, coordination need not be complete in the sense of implementing the perfectly collusive outcome in a market. Second, information need not be perfect to sustain collusion. Most realistic scenarios of tacit collusion assume some degree of incomplete information which may then be reflected in some inefficiency in the reaction of the coordinated firms.

Still, one can certainly pay attention and give proper weight to such things as the existence of facilitating practices: observed industry practices which seem to have no other purpose than to allow information to flow or to facilitate an agreement. For instance, Kühn (2001) proposes that, given the intrinsic difficulty in inferring whether prices are the result of competitive oligopoly or of tacit collusion, it is more desirable to focus on suppressing certain forms of communication between firms, which do not bring efficiency and are likely to sustain a collusive equilibrium. His paper contains a review of the experimental evidence of the positive role of communication in collusion. See also the more recent experimental evidence reported in Cooper and Kühn (2009).

The extent of price rigidity may be relevant to such an evaluation of tacit collusion, and/or the presence of unexplained price wars in a market, where legitimate explanations for such outcomes can potentially be excluded. If prices sometime oscillate widely when there are no obvious demand or cost causes, competition authorities will want to consider alternative potential explanations, one of which is tacit collusion.

Since all actual instances of tacit collusion are likely to occur in a world of imperfect information, it is likely that agreements will not always work smoothly all the time. Firms will also rarely be completely symmetric and agreements, once reached, may not satisfy the ambitions of all players robustly. Some firms will probably have more incentives than others to cheat and to do so they will be more likely to take advantage of sudden fluctuations in demand or costs to lower prices and sell more than their agreed share. Competitors, unable to distinguish between the consequences of demand changes and cheating may retaliate and all this instability may become apparent in the data. It is possible that an examination of price series

---

41 ECJ ruling of 10/07/08 in case C-413/06P in particular paragraphs 117–34.
shows periods of price stability alternating with periods of price drops and output expansion. If such sequences of price stability and price fluctuations cannot be explained by exogenous changes in demand, costs, or institutional environment, one may consider the possibility of a change in the competition regime in the industry with collusion alternating with competition. For example, Suslow and Levenstein (2006) find that problems in resolving the bargaining game played by cartel participants following changes in market conditions have frequently played an important role in cartel breakdowns.

Porter (1983), Bresnahan (1987), and Baker (1989) each suggest that examination of the way conduct varies over time can provide a useful source of information about the likelihood of tacit collusion. For example, Baker (1989) empirically identifies changes in competition regime that occur after unpredicted negative demand shocks trigger cheating from a cartel and causes temporary reversion to competition in the U.S. cartelized steel industry between 1933 and 1939.

In general, a tacitly collusive arrangement that is working well will sometimes, perhaps even often, tend to stabilize prices and/or quantities particularly when the terms of the agreement are complex and the transaction costs of renegotiating the targeted outcome are high. Also, when there is a lot of uncertainty about the evolution of the market, communicating and agreeing on new collective behavior can be difficult. For this reason, colluding industries can sometimes be less reactive to observed changes in costs or demand compared with a competitive market as they may tend to keep doing what they know until they succeed (or not) in collectively adjusting to change. Excessive price rigidity in the face of changing market conditions can therefore also be the sign of a colluding industry, especially if no particular efficiency gains can be attributed to the high stability of prices.

Abrantes-Metz et al. (2006) examine the effect of the collapse of a bid-rigging conspiracy in the frozen seafood industry on price levels and dispersion. The collapse of the cartel caused a decrease in price by 16% while the price variance more than doubled. Based on these results, Abrantes-Metz et al. designed a test that they applied to the market for retail gasoline in Louisville, Kentucky, between 1996 and 2002, and did not find a pattern of particularly low variance in the data. Connor (2005) provides a review of the empirical evidence and the theory underlying the relation between collusion and price dispersion. In principle, a reduction in price dispersion will sometimes be expected in a collusive setting for reasons including: production may be allocated more efficiently; shocks may provoke a coordinated response; the effect of differences in buyer’s search costs on the price they ultimately get is diminished. In practice, most of the few studies that have analyzed price dispersion during collusion periods have found a reduction of the variance during the coordination. Bolotova et al. (2008) fail to detect this reduction in price variance during the citric acid cartel although those authors do find it in the lysine cartel. Such findings, however, relate mostly to explicit cartels where communication is likely to be better than in a tacitly collusive environment.
6.2. Directly Identifying the Nature of Competition

An alternative and significantly more involved approach to assessing the likelihood of tacit collusion involves empirically estimating the incentives and ability to collude in a way that is explicitly motivated by theoretical models. Kovacic et al. (2006) explicitly propose calculating the payoff to collusion between various subsets of firms in order to evaluate the incentive to collude. They propose empirically evaluating the profits that would be available from various firms getting together to collude with the Coasian view that firms are good at solving coordination problems when there are sufficient incentives in place. This approach requires only an application of the framework used for unilateral effects merger simulation, which we cover in detail in chapter 8.

Davis and Sabbatini (2009) go further. Those authors propose building on the contributions of Friedman (1971) and the unilateral effects mergers simulation literature (see chapter 8). Specifically, they propose calculating not only the incentive to collude should collusion be successful but also (i) evaluate the other potentially relevant incentives such as the incentive to “cheat” and (ii) evaluate the ability of a given group of firms to sustain coordination. To do so they note that a standard dynamic oligopoly game suggests that firm will only be able to sustain collusion if the net present value of payoffs to collusion are greater than the net present value of the payoff to cheating (defecting) subtracting the consequences of whatever punishments rivals impose.

Following Friedman (1971), let the one-period payoff to collusion, defection, and competition be respectively $\pi^\text{Collusion}_f$, $\pi^\text{Defection}_f$, and $\pi^\text{NE}_f$. The net present value (NPV) incentive compatibility constraint can be written:

$$V^\text{Collusion}_f(\delta_f) > V^\text{Defection}_f(\delta_f) \iff \frac{\pi^\text{Collusion}_f}{1 - \delta_f} > \frac{\pi^\text{Defection}_f}{1 - \delta_f} + \delta_f \pi^\text{NE}_f,$$

where $\delta_f$ represents firm $f$’s discount factor and the punishment is assumed to be a reversion to Nash competition.

Davis and Sabbatini (2009) follow Friedman’s (1971) model of tacit collusion while allowing for firm heterogeneity and differentiated products. Primarily, their modest contribution is to propose actually empirically implementing that model in differentiated product (and multimarket) contexts where previously authorities have only used a checklist of factors likely to facilitate collusion to arrive at a view where collusion is more or less likely. Given the development of the theory described in the technical box, they show (and we will see in chapter 8) that $\pi^\text{Collusion}_f$ and $\pi^\text{NE}_f$ are available from unilateral effects merger simulation models while the theory suggests we can also fairly trivially calculate $\pi^\text{Defection}_f$ as the payoff to cheating against cooperating rival firms that are choosing the cooperative price and this can

---

42 The joint paper combines and extends two earlier discussion papers, one from each author: the first was by Sabbatini (2006) and the second by Davis (2006f).
also be calculated using the techniques developed by the unilateral effects merger simulation literature. Davis and Huse (2009) implement the approach using data from the network computer server market and evaluate the incentives to coordinate in the HP–Compaq merger (see chapter 8).

6.2.3.4 Empirical Assessment of Collective Dominance: The Sony–BMG Merger

The Sony–BMG merger provides an important recent example of an empirical assessment of the likelihood of collective dominance in a market. The assessment was undertaken by the European Commission following a notified joint venture between SONY Music and BMG that would bring together the worldwide recording businesses of both music majors. The merger was cleared by the Commission in 2004 but that decision was subsequently annulled in July 2006 by the Court of First Instance, which decided that the Commission had made manifest errors of assessment when considering the case. The merger was renotified in 2007 and subsequently cleared a second time by the European Commission that same year after an in-depth investigation. The European Court of Justice eventually overturned the Court of First Instance decision in 2008, validating the first decision.43

Before the merger the industry was dominated by five music majors: Universal, Sony, BMG, EMI, and Warner Music. There were also significant “independent” labels but there was a concern that these labels were not in a position or did not have the incentives to challenge a potential coordination on recorded music CD prices by the majors. The assessment of the merger therefore centered on establishing whether conditions in the market were sufficiently conducive to tacit coordination. Stigler’s three conditions (and hence the three Airtours conditions)—agreement, monitoring, and enforcement—were examined but the core of the assessment centered on whether there was sufficient transparency in the market for recorded music for an agreement to be monitored and therefore enforced. To analyze this question, the European Commission gathered the most extensive database it had ever collected. Specifically, it requested from merging parties and third parties transaction-level data that indicated what CD title was sold to what customer at what price on a particular day or week between January 2002 and June 2006.44 In addition, titles were categorized according to whether they were in the charts at the time of the transaction and how long they had been released on the market. The data provided information on both the list price and the net price of the transactions. This extensive data set allowed the Commission to perform a comprehensive data analysis of the stability and therefore the predictability of the discounts.

---

43 ECJ of 10/07/08 C-413/06P.
44 Customers are retailers of various kinds (specialized music stores, nonspecialized stores like electronics chains, mass merchants, and supermarkets) or intermediate distributors such as rack jobbers.
6.2. Directly Identifying the Nature of Competition

Table 6.6. Methodology used by the European Commission in the assessment of the SONY-BMG merger.

<table>
<thead>
<tr>
<th>Customer</th>
<th>No. of CDs</th>
<th>Published price to dealers (PPD)</th>
<th>Chart</th>
<th>First week?</th>
<th>Total discount</th>
<th>Average standard deviation</th>
<th>Low dispersion?</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>500</td>
<td>€12.5</td>
<td>Yes</td>
<td>Yes</td>
<td>15%</td>
<td>1 pp</td>
<td>Yes</td>
</tr>
<tr>
<td>A</td>
<td>100</td>
<td>€12.5</td>
<td>Yes</td>
<td>No</td>
<td>10%</td>
<td>3 pp</td>
<td>No</td>
</tr>
<tr>
<td>A</td>
<td>200</td>
<td>€12.5</td>
<td>No</td>
<td>No</td>
<td>15%</td>
<td>5 pp</td>
<td>No</td>
</tr>
<tr>
<td>A</td>
<td>200</td>
<td>€10.0</td>
<td>Yes</td>
<td>No</td>
<td>8%</td>
<td>1 pp</td>
<td>Yes</td>
</tr>
</tbody>
</table>

Source: European Commission.

At any point in time there are thousands of CDs actively sold in the market representing a large variety of artists, styles, and degrees of popularity. Still, as they are sold in a similar format and distributed in similar ways, the Commission acknowledged the possibility that one might think of a CD as a fairly generic item. It also considered the fact that most sales in the music business are actually generated by a very limited number of charted CDs so that collusion could be profitable by only coordinating the prices of these “few” high sales CDs. Moreover, because most of the sales of an album are generated in the first few weeks after its release, the Commission focused on whether there was coordination on prices at and shortly after the release date.

The fundamental question that the Commission asked to evaluate the extent of price transparency was: can a knowledgeable market participant infer the net price of a transaction from observable transaction characteristics? The observable characteristics of a transaction were the title, the customer, the time, whether the title was in the chart, whether the title was in its first week of release, and the list price of the title. What was not observed was the discount granted by the major to the retailer (customer) and the question was whether these discounts were sufficiently systematic to be sufficiently predictable.

To answer this question the Commission separated titles of each major into groups according to the observable characteristics mentioned above: list price (published price to dealers), whether it is the first week of release, whether charted, and customer identity. For every such group, the average discount and weighted within-group average standard deviation of the discount level was calculated. Groups of titles were then separated in two categories: those for which the discount variation is deemed respectively to be large and small. Specifically, large (small) was defined as those that exhibit a weighted standard deviation of more than (less than) two percentage points (pp). A summary of the results for a particular combination of major and customer for four groups of titles might look like table 6.6.
The total number of units in groups with low dispersion is then related to the total units sold by that major. In our example, and assuming there was only this customer, the number of units sold by that major under a “regime” of low dispersion of discounts would represent 70% of the sales.\footnote{(Sales in rows 1 and 4)/(Total sales in rows 1 to 4) = (500 + 200)/(500 + 100 + 200 + 200) = 700/1,000 = 0.7, i.e., 70% of sales are low dispersion.}

Commission results showed that before the merger, sales under a regime of “discount stability” represented less than 60% of all sales for most majors. This was sufficient evidence for the Commission to decide that it was highly improbable that there was coordination on the prices of CDs based on whether they were charted or newly released. A higher degree of complexity in the segmentation of products for the purpose of coordination was also deemed unlikely, since it would have involved a more complex definition of CD groupings.

\subsection*{6.2.3.5 Bid-Rigging: Collective Dominance in Auctions}

Bidding markets are often the subject of coordination investigations. When a market is such that contracts are awarded by customers through auctions or bids from suppliers, procurement auctions, suppliers may agree to coordinate in order to keep prices high. In such cases, firms decide in advance who will win which auction and those not selected commit to offer higher prices on that particular auction, usually in exchange for winning in other auction processes. In such collusion, competitors’ prices or bids are sometimes not directly observable by competitors but the outcome usually is.\footnote{Sometimes, particularly in procurement auctions by government agencies, there is a tendency to tell all participants all of the bids in an auction. While good practice in terms of transparency of government, on balance such a practice can lead to serious problems breaking coordination and as a result procurement will only occur at sometimes very high prices. Governments can be very inelastic demanders in some procurement areas, e.g., military equipment.}

A firm can then monitor whether any competitor cheated by lowering its particular bid when it was not supposed to.

It can be difficult to detect collusion in such markets because transactions are sometimes less frequent and the goods involved may even be unique, making it difficult to compare prices or even establish a market price. Consider, for example, how many aircraft carriers governments buy: very few. Nonetheless, several strategies have been proposed to detect coordination in auction markets. For fairly recent surveys, see Bajari and Summers (2002), Porter (2005), or Harrington (2008).

Efforts to identify collusion using empirical applications of auctions now have a long and distinguished history, though the techniques are generally not for the technicality shy. Authors in this empirical tradition include Porter and Zona (1993, 1999), Baldwin et al. (1997), and, more recently, Bajari and Ye (2001). The latter, for example, argue that identification is best achieved noting that bids should fulfill two conditions, which, if violated, would exclude competitive behavior.\footnote{For a nontechnical exposition, see Bajari and Summers (2002).} First, that...
6.2. Directly Identifying the Nature of Competition

once we take into account all publicly observable cost information that determines a conditional expected value of the bid amount, the deviations from that expected value of the bid should not be correlated among bidders. That is, we should not see many firms bidding particularly highly at any time. Their second condition is exchangeability: the bid of a firm should not be affected by the identity of the firm with the next lower costs. A competing firm will always bid such as to cover its costs and beat firms with the next highest cost and it should not matter who their closest competitor is beyond the rival’s cost when there is competition. In contrast, when the closest competitor is a member of a cartel, then the firm will be able to price above this competitor’s costs and still win the bid. In this case, the identity of the closest competitors, whether it is a cartel member or not, will affect the amount of the firm’s bid. Bajari and Ye propose statistical ways to test these conditions which each form the basis of a method of identifying whether data are generated by one model or another. In particular, they argue that violations of these conditions are not consistent with a competitive market. If markets are such that bids seem uncorrelated and there is “exchangeability” of competitors, there is no proof that competition is taking place but competition cannot be rejected. We leave the reader with this highly incomplete introduction and a route to those authors’ papers for further information but note that their proponents suggest that the power of these kinds of tests are demonstrable since, for example, both Porter and Zona (1993, 1999) and Pesendorfer (2000) analyze data sets where collusion is known to have taken place and they find that (1) cartel members tend to bid less aggressively than noncartel members and (2) the bids of cartel members tend to be more correlated with each other than with the bids of noncartel members.

6.2.4 Single Dominance: Market Power with Differentiated Products

Collusion often arises in markets where the products are relatively homogeneous so that, under competitive conduct, firms impose a significant competitive constraint on each other. Because firms can exploit little market power individually, they have an incentive to coordinate to extract rent collectively. With differentiated products, firms have some degree of market power due to the fact that consumers have preferences for particular goods offered by the different firms and are often willing to pay more for the specific product they like compared with other similar products. Firms do exert pricing constraints on each other because customers will eventually switch if their preferred product becomes relatively too expensive, but the degree of the constraint can vary greatly. The exercise of market power obtained through product differentiation is not the subject of competition policy scrutiny in itself in many jurisdictions, most notably the United States, but such exploitative abuses by

48 Such an identification strategy appears to rely heavily on the bidding firms knowing no more “public” knowledge than the investigator.
dominant firms are potentially actionable in other jurisdictions, including Europe, so that sufficient differentiation could generate concerns. In addition, actions undertaken by a firm for the purpose of eliminating significant competitive constraints and thereby maintaining or increasing its own market power are not allowed when they seriously harm consumers (exclusionary abuses). While the appropriate extent of activity to combat exploitative abuses is a matter of debate, many jurisdictions examine the consequences of firms’ decisions to acquire or merge with rivals because such actions have the potential to generate substantial increases in market power that would be detrimental to consumers. In this section we begin our analysis of the pricing power of firms in a market for differentiated products in which firms compete on prices. The framework we describe is discussed in further detail in chapter 8, where we present a general analysis of merger simulation models testing for unilateral effects. We then discuss the conditions for the identification of coordination in a differentiated product pricing setting. Doing so allows us to examine how to extend Bresnahan’s (1982) identification results for the homogeneous goods context (described above) to a differentiated product context. Nevo (1998) provides a numerical example while we follow the more formal identification results provided in Davis (2006d). We then illustrate with an empirical example drawing on Bresnahan (1987).

6.2.4.1 Pricing Equations

In order to identify the competitive behavior of firms in differentiated product markets, we must understand the pricing decisions made by the different market players and also how they interact. We will see that (static) economic theory suggests that firms will react to the presence of close substitutes owned by rivals by pricing more aggressively. Identifying the pricing equations of firms can help us measure the level of market power faced by an individual firm in a particular market.

Consider a simple theoretical example involving two differentiated but substitute products whose prices must be determined. We contrast the incentives to set the prices of (i) two firms competing in prices in a standard differentiated products’ Bertrand model with (ii) a single firm owning both products. (For background, see the introductory discussion in chapters 1 and 5.) When comparing the two sets of first-order conditions generated by these two different models, the firm maximizing joint profits takes into account the effect of the change in price of good \( j \) on the quantity of all goods and not only on the quantity of good \( j \). If an increase of the price of good 1 causes the demand for good 2 to increase, this increase in the revenues coming from the sales of good 2 will mitigate the impact of the lower sales of good 1. Therefore, the firms maximizing joint profits will have more of an incentive to increase prices compared with the single-product firm.
Generally, we can write down first-order conditions which encompass both models as follows:

for good 1: \[ (p_1 - c_1) \frac{\partial Q_1(p)}{\partial p_1} + Q_1(p) + \Delta_{12}(p_2 - c_2) \frac{\partial Q_2(p)}{\partial p_1} = 0, \]

for good 2: \[ \Delta_{21}(p_1 - c_1) \frac{\partial Q_1(p)}{\partial p_2} + Q_2(p) + (p_2 - c_2) \frac{\partial Q_2(p)}{\partial p_2} = 0, \]

where \( \Delta_{ij} \) indicates whether changes in the quantity demanded of product \( j \) will affect the pricing of product \( i \) (see box). In the single-product firm’s case, \( \Delta_{12} = \Delta_{21} = 0 \). In the case where one firm produced both products, we set \( \Delta_{12} = \Delta_{21} = 1 \). In an industry with several firms producing different ranges of products we would have a pricing equation for each product and a \( \Delta \) matrix indicating the ownership structure of the industry. We will consider the general version of this game in chapter 8.

Suppose the demand function for each of the firms are linear in parameters and prices, so that the demand for product \( j \) is given by

\[ Q_j = \alpha_j + \alpha_{j1} p_1 + \alpha_{j2} p_2. \]

If two single-product firms play a Bertrand–Nash pricing game, they will maximize profits with respect to prices

\[ \max_{p_j} \pi_j(p_1, p_2) = \max_{p_j} (p_j - c_j) Q_j(p), \]

where \( p = (p_1, p_2) \) denotes the vector of prices and this will result in a set of optimal pricing equations:

for firm 1: \[ (p_1 - c_1) \frac{\partial Q_1(p)}{\partial p_1} + Q_1(p) = 0, \]

for firm 2: \[ (p_2 - c_2) \frac{\partial Q_2(p)}{\partial p_2} + Q_2(p) = 0. \]

Next let us assume that one firm now produces both of the two goods. It will maximize the joint profits from both goods:

\[ \max_{p_1, p_2} \pi_1(p_1, p_2) + \pi_2(p_1, p_2) = \max_{p_1, p_2} (p_1 - c_1) Q_1(p) + (p_2 - c_2) Q_2(p). \]

The resulting optimal pricing equations become

for good 1: \[ (p_1 - c_1) \frac{\partial Q_1(p)}{\partial p_1} + Q_1(p) + (p_2 - c_2) \frac{\partial Q_2(p)}{\partial p_1} = 0, \]

for good 2: \[ (p_1 - c_1) \frac{\partial Q_1(p)}{\partial p_2} + Q_2(p) + (p_2 - c_2) \frac{\partial Q_2(p)}{\partial p_2} = 0. \]

In the case of linear demands, each of the derivative terms will be parameter values, \( \alpha_{j1}, \alpha_{j2} \).
Note that different ownership structures or different competition models will have different implications for the equilibrium prices. In this constant marginal cost example, shocks to cost and demand will affect prices differently depending on the value of $\Delta_{ij}$, which indicates the products that enter a given firm’s profit-maximization function.

If we can estimate the demand equations, then we will have estimates of the demand parameters, $\alpha$. From a traditional analysis of estimation of linear equations, we know we can do this for a demand equation if we have as many excluded cost variables (or, more generally, supply (pricing) equation shifters) as we have endogenous variables in the demand equation. In the case where marginal costs are constant in quantity we can therefore retrieve the conduct parameters $\Delta_{ij}$ in much the same way as was done in the homogeneous product case. Demand and cost shifters will be needed for identification and given enough of them will also be sufficient to test our model of collusion against the analogous model of competition. In this case, instead of a single demand and a single pricing equation, we will have a system of $J$ demand and $J$ pricing equations. Much like in the homogeneous goods example, we can substitute the demand function for the quantities in the pricing equation and reduce the system to one which involves “only” $J$ equations. The estimated parameters capturing the effect of demand and cost shifters on other products will provide us with information about the extent to which these products are constraining the price of the product being analyzed.

We illustrate with a very simple example using our equation for equilibrium prices. Assuming the products’ linear demand functions,

$$Q_j = \alpha_{j0} + \alpha_{j1} p_1 + \alpha_{j2} p_2,$$

the pricing equations become

$$\begin{align*}
(p_1 - c_1)\alpha_{11} + (p_2 - c_2)\Delta_{12}\alpha_{21} + Q_1(p) &= 0, \\
(p_1 - c_1)\Delta_{21}\alpha_{12} + (p_2 - c_2)\alpha_{22} + Q_2(p) &= 0,
\end{align*}$$

which can be written in matrix form as

$$\begin{bmatrix}
\alpha_{11} & \Delta_{12}\alpha_{21} \\
\Delta_{21}\alpha_{12} & \alpha_{22}
\end{bmatrix}
\begin{bmatrix}
p_1 - c_1 \\
p_2 - c_2
\end{bmatrix}
+ \begin{bmatrix}
Q_1(p) \\
Q_2(p)
\end{bmatrix} = 0.$$

The differentiated product setting moves us from our usual demand and supply (pricing) equations in a homogeneous product setting, where we analyze two simultaneous equations, to a situation with a total of $J$ demand and $J$ supply curves, where $J$ is the number of products being sold. In this case, such an approach would leave us with four equations to solve. The $2J$ equations form the “structural form” of the differentiated product model. Alternatively, we need solve only a two-equation
system if we substitute in the demand system
\[
\begin{pmatrix}
\alpha_{11} & \alpha_{12} \\
\alpha_{21} & \alpha_{22}
\end{pmatrix}
\begin{pmatrix}
p_1 \\
p_2
\end{pmatrix}
+ \begin{pmatrix}
\alpha_{01} \\
\alpha_{02}
\end{pmatrix}
\]
to give
\[
\begin{pmatrix}
\alpha_{11} & \Delta_1 \alpha_{21} \\
\Delta_2 \alpha_{12} & \alpha_{22}
\end{pmatrix}
\begin{pmatrix}
p_1 - c_1 \\
p_2 - c_2
\end{pmatrix}
+ \begin{pmatrix}
\alpha_{11} & \alpha_{12} \\
\alpha_{21} & \alpha_{22}
\end{pmatrix}
\begin{pmatrix}
p_1 \\
p_2
\end{pmatrix}
+ \begin{pmatrix}
\alpha_{01} \\
\alpha_{02}
\end{pmatrix}
= 0.
\]
A small amount of matrix algebra allows us to solve for equilibrium prices:
\[
\begin{pmatrix}
p_1 \\
p_2
\end{pmatrix}
= - \begin{pmatrix}
2\alpha_{11} & \Delta_1 \alpha_{21} + \alpha_{12} \\
\Delta_2 \alpha_{12} + \alpha_{21} & 2\alpha_{22}
\end{pmatrix}^{-1}
\begin{pmatrix}
\alpha_{01} \\
\alpha_{02}
\end{pmatrix}
+ \begin{pmatrix}
2\alpha_{11} & \Delta_1 \alpha_{21} + \alpha_{12} \\
\Delta_2 \alpha_{12} + \alpha_{21} & 2\alpha_{22}
\end{pmatrix}^{-1}
\begin{pmatrix}
\alpha_{11} & \Delta_1 \alpha_{21} \\
\Delta_2 \alpha_{12} & \alpha_{22}
\end{pmatrix}
\begin{pmatrix}
c_1 \\
c_2
\end{pmatrix}.
\]
A numerical example may be useful. Suppose, for example, that \(\alpha_{11} = \alpha_{22} = -2\), \(\alpha_{12} = 2\), and \(\alpha_{21} = 1\). Then two single-product firms in a competitive scenario would produce the following prices:
\[
\begin{pmatrix}
p_{1t} \\
p_{2t}
\end{pmatrix}
= \begin{pmatrix}
-1 & -4 \\
14 & -1
\end{pmatrix}
\begin{pmatrix}
\alpha_{01t} \\
\alpha_{02t}
\end{pmatrix}
+ \begin{pmatrix}
1 & 4 \\
14 & 2
\end{pmatrix}
\begin{pmatrix}
c_{1t} \\
c_{2t}
\end{pmatrix}.
\]
While under collusion the prices will be determined by
\[
\begin{pmatrix}
p_{1t} \\
p_{2t}
\end{pmatrix}
= \begin{pmatrix}
-1 & -3 \\
7 & -3
\end{pmatrix}
\begin{pmatrix}
\alpha_{01t} \\
\alpha_{02t}
\end{pmatrix}
+ \begin{pmatrix}
1 & 2 \\
7 & -2
\end{pmatrix}
\begin{pmatrix}
c_{1t} \\
c_{2t}
\end{pmatrix}.
\]
Note that under collusion the firms have a weight of \(\frac{3}{7}\) on the impact of other products’ demand shifters when setting the price for good 1. Under perfect competition, however, the firms only put a weight of \(\frac{2}{14} = \frac{1}{7}\) on product 2’s demand shifter.

As our numerical example illustrates, in this model common ownership or coordinated behavior puts greater weight on what happens to the demand of the other product setting its price. As a result, the differentiated products Bertrand model suggests that movements in the rival product’s demand will affect a product’s price far more under collusion than under perfect competition. Such a result is perhaps intuitive since collusive arrangements “internalize” effects across products. Given demand estimates and an ownership structure, we can measure how these transmissions occur. In doing so we can compare how much weight is genuinely given to rivals’ demand or cost shifters. This difference in transmission reactions to demand or cost shifts can be enough to identify whether firms are setting the prices of products independently or not. This can be considered intuition for identification in an
econometric model, but it can also be helpful when collecting other evidence in a
given case (e.g., documentary evidence). On the other hand, such an observation
may concern us since we noted earlier that on occasion cartels have often resulted
in relatively less variation in prices, perhaps because of stability concerns. As Corts
(1999) noted, a different model of collusion would have different implications for
observed collusive prices.

6.2.4.2 Identification of Pricing and Demand Equations in Differentiated Markets

In a fashion entirely analogous to the homogeneous products case, the identification
of conduct generally requires that the parameters of the demand and pricing equa-
tions are identified. Even if demand rotation can also be used to identify conduct
in differentiated industries in the same way as is done for homogeneous products,
demand does need to be estimated to confirm or validate assumptions. This presents
a challenge because a differentiated product industry has one demand curve and one
pricing function for each of the products being sold. In contrast, in the homogeneous
product case, there is only one market demand and one market supply curve that
need to be estimated. Now we will need to estimate as many demand functions as
there are products and also as many pricing equations as there are products. Iden-
tification naturally becomes more difficult in this case and some restrictions will
have to be imposed in order to make the analysis tractable. We discuss differentiated
product demand estimation extensively in chapter 9.

A general principle for identification of any linear system of equations is that the
number of parameter restrictions on each equation should be equal to, or greater
than, the number of endogenous variables included in the equation. A normalization
restriction is always imposed in the specification of any equation so in practice
the number of additional restrictions must equal or be more than the number of
endogenous variables less one. This is equivalent to saying that the restrictions
must be equal to or more than the number of endogenous variables on the “right-
hand side” of any given equation. The total number of endogenous variables is also
the number of equations in the structural model. This general principle is known as
the “order condition” and is a necessary condition for identification in systems of
linear equations. It may, however, not be sufficient in some cases. Previously, we
encountered the basic supply-and-demand two-equation system, where we had two
structural equations with two endogenous variables: price and quantity. In that case
we needed the normalization restrictions and then at least one parameter restriction
for each equation for identification. We obtained the parameter restrictions from
theory: variables that shifted supply but not demand were needed in the equations
to identify the demand equation and vice versa (these exclusion restrictions are
imposed by restricting values of the parameters to zero). A more technical discussion

49 The normalization restriction is usually imposed implicitly by not placing a parameter on whichever
one of the endogenous variables is placed on the left-hand side of an equation.
of identification of demand and pricing equations in markets with differentiated products is provided in the annex to this chapter (section 6.4), which follows Davis (2006d).

### 6.2.4.3 Identification of Conduct: An Empirical Example

When conduct is unknown, we will want to assess the extent to which firms take into account the consequences of pricing decisions on other products when they price one particular good. In this case, one strategy is to estimate the reduced form of the structural equations and retrieve the unknown structural parameters by using the correspondence between reduced-form and structural parameters derived from the general structural specification. Assuming that the demand parameters are identified and marginal costs are constant, we will need enough demand shifters excluded from a pricing equation to be able to identify the conduct parameters (see Nevo 1998). In particular, we will need as many exogenous demand shifters in the demand equation as there are products produced by the firm. Although identification of conduct is therefore technically possible, in practice it may well be difficult to come up with a sufficient number of exogenous demand and cost shifters.

An early and important example of an attempt to identify empirically the nature of competition in a differentiated product market is provided by Bresnahan’s (1987) study of the U.S. car industry in the years 1953–57. Bresnahan considers the prices and number of cars sold in the United States during those years and attempts to explain why in 1955 prices dropped significantly and sales rose sharply. In particular, he tests whether this episode marks a temporary change of conduct by the firms from a coordinated industry to a competitive one. The data that Bresnahan (1987) is trying to explain are presented in table 6.7. The important feature of the data to notice is that it is apparent that 1955 was an atypical year with low prices and high quantities. Real prices fell by 5%, quantity increased by 38%, and revenues increased by 32%.

To begin to build a model we must specify demand. Bresnahan specifies demand functions where each product’s demand depends on the two neighboring products in

<table>
<thead>
<tr>
<th>Year</th>
<th>Auto production (units)</th>
<th>Real auto price/CPI</th>
<th>% in quality-adjusted prices</th>
<th>Sales revenues ($)</th>
<th>Quantity index</th>
</tr>
</thead>
<tbody>
<tr>
<td>1953</td>
<td>6.13</td>
<td>1.01</td>
<td>—</td>
<td>14.5</td>
<td>86.8</td>
</tr>
<tr>
<td>1954</td>
<td>5.51</td>
<td>0.99</td>
<td>—</td>
<td>13.9</td>
<td>84.9</td>
</tr>
<tr>
<td>1955</td>
<td>7.94</td>
<td>0.95</td>
<td>2.5</td>
<td>18.4</td>
<td>117.2</td>
</tr>
<tr>
<td>1956</td>
<td>5.80</td>
<td>0.97</td>
<td>6.3</td>
<td>15.7</td>
<td>97.9</td>
</tr>
<tr>
<td>1957</td>
<td>6.12</td>
<td>0.98</td>
<td>6.1</td>
<td>16.2</td>
<td>100.0</td>
</tr>
</tbody>
</table>

*Source: Bresnahan (1987).*
6. Identification of Conduct

terms of quality: the immediately lower-quality and the immediately higher-quality product. He motivates his demand equation using a particular underlying discrete choice model of demand but ultimately his demand function takes the form,

\[ q_i = \delta \left[ \frac{P_j - P_i}{x_j - x_i} - \frac{P_i - P_h}{x_i - x_h} \right], \]

where \( P \) and \( x \) stand for price and quality of the product and \( h, i, \) and \( j \) are indicators for products of increasing quality. Quality is one dimensional in the model, but captures effects such as horsepower, number of cylinders, and weight. Note that, all else equal, demand is linear in the prices of the goods \( h, i, \) and \( j \) and that given a price differential the cross-price slopes will increase with a decrease in the difference in quality, \( x \). In this rather restrictive demand model there is only a single parameter to estimate, \( \delta \).

To build the pricing equations, he assumes a cost function where marginal costs are constant in quantity produced but increasing in the quality of the products so that \( x_j \geq x_i \geq x_h \) for products \( j, i, \) and \( h \). These assumptions imply that the whole structure can be considered as a particular example of a model where demand is linear in price and marginal costs are constant in output. By writing a linear-in-parameters demand equation, where \( q_i = \alpha_{i0} + \alpha_{ii} p_i + \alpha_{ij} p_j + \alpha_{ih} p_h \), we can see that for fixed values of the quality indices, \( x_i, x_j, \) and \( x_h \), the analysis of a pricing game using Bresnahan’s demand model can be incorporated into the theoretical structure we developed above for the linear demand model where the parameters in the equation are in fact functions of data and a single underlying parameter. (More precisely, we studied the linear demand model with two products above and we will study the general model in chapter 8.) Specifically, the linear demand parameters are of the form,

\[ \alpha_{ii} = -\delta \left( \frac{1}{x_j - x_i} + \frac{1}{x_i - x_h} \right), \]
\[ \alpha_{ij} = \delta \left( \frac{1}{x_j - x_i} \right), \]
\[ \alpha_{ih} = \delta \left( \frac{1}{x_i - x_h} \right). \]

Bresnahan estimates the system of equations by assuming first that there is Nash competition so that the matrix \( \Delta \) describes the actual ownership structure of products (i.e., there is no collusion). Subsequently, he estimates the same model for a cartel by setting all the elements of the \( \Delta \) matrix to 1 so that profits are maximized for the entire industry. He can then use a well-known model comparison test called the Cox test to test the relative explanatory power of the two specifications.\(^{50}\) Bresnahan
6.3 Conclusions

Concludes that the cartel specification explains the years 1954 and 1956 while Nash competition model explains the data from 1955 best. From this, he concludes that 1955 amounted to a temporary breakdown of coordination in the industry.

Intuitively, Bresnahan is testing the extent to which close substitutes are constraining each other. If the firm maximizes profits of the two products jointly, there will be less competitive pressure than in the case where the firm wants to maximize profits on one of the products only and therefore ignores the negative consequences of lower prices on the sales of the close substitute product. Thus, in figure 6.5, if close substitute products 2 and 3 are owned by rivals, then they will have a low markup under competition but far higher markups under collusion.

Given his assumptions about costs and the nature of demand, Bresnahan finds that the explanation for the drop in price during 1955 is the increase in the level of competition of close substitutes in the car market.

The demand shifters that helped identify the parameter estimates are presented in table 6.8 as well as the accounting profits of the industry. The accounting profits, however, are not consistent with Bresnahan’s theory, as he notes. If firms are coordinating in the years 1954 and 1956, industry profits should be higher than in 1955 when they revert to competition. Bresnahan’s response is that accounting profits are not representative of economic profits and are not to be relied upon. We must therefore make a decision in this case about whether to believe the accounting measures of profitability or the econometric analysis. In other cases, one might hope each type of evidence allows us to build toward a coherent single story.

6.3 Conclusions

- Structural indicators such as market shares and concentration levels are still commonly used for a first assessment of industry conduct and performance,
Table 6.8. Demand and cost shifters of the car market in the United States 1953–57.

<table>
<thead>
<tr>
<th>Year</th>
<th>Per capita disposable income Level</th>
<th>Growth</th>
<th>Interest rates</th>
<th>Durable expenditures (nonauto)</th>
<th>Accounting profits ($)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1953</td>
<td>1,623</td>
<td>1.9</td>
<td>14.5</td>
<td>2.58</td>
<td></td>
</tr>
<tr>
<td>1954</td>
<td>1,609</td>
<td>0.9%</td>
<td>14.5</td>
<td>2.25</td>
<td></td>
</tr>
<tr>
<td>1955</td>
<td>1,659</td>
<td>3.0%</td>
<td>16.1</td>
<td>3.91</td>
<td></td>
</tr>
<tr>
<td>1956</td>
<td>1,717</td>
<td>3.5%</td>
<td>17.1</td>
<td>2.21</td>
<td></td>
</tr>
<tr>
<td>1957</td>
<td>1,732</td>
<td>0.9%</td>
<td>17.0</td>
<td>2.38</td>
<td></td>
</tr>
</tbody>
</table>

Source: Bresnahan (1987).

although they are not usually determinative in a regime applying an effects-based analysis of a competition question. The fact that they are not determinative does not mean market shares are irrelevant, however, for a competition assessment and many authors consider they should carry some evidential weight.

- Developments in static economic theory and the availability of data have shown that causality between market concentration and industry profitability cannot be easily inferred. However, economic theories built on dynamic models do frequently have a flavor of considerable commonality with the older SCP literature. For example, Sutton (1991, 1998) emphasizes that prices are indeed expected to be a function of market structure in two-stage games where entry decisions are made at the first stage and then active firms compete in some way (on prices or quantities) or collude at a second stage.

- The broad lesson of game theory is that quite detailed elements of the competitive environment can matter for a substantial competition analysis. The general approach of undertaking a detailed market analysis aims at directly identifying the nature of competition on the ground and therefore the likely effects of any merger or alleged anticompetitive behavior.

- Technically, the question of identification involves asking the question of whether two models of behavior can be told apart from one another on the basis of data. The hard question in identification is to establish exactly which data variation will be helpful in moving us to a position where we are able to tell apart some of our various models. The academic analysis of identification tends to take place within the context of econometric models, but the lessons of such exercises typically move directly across to inform the kinds of evidence that competition authorities should look for more generally such as evidence from company documents.
The degree to which firms are reactive to changes in demand conditions in the market can provide direct evidence of the extent of a firm’s market power. Formal econometric models can use the methods involving the estimation of conduct parameters in structural models to determine whether the reactions of firms to changes in prices are consistent with competitive, competing oligopoly, or collusive settings. However, the more general lesson is that changes in the demand elasticity can provide useful data variation to identify conduct. For example, we might (at least conceivably) find documentary evidence suggesting that firms’ pricing reactions accommodate prices in a fashion consistent with a firm’s internal estimates of market demand sensitivities (rather than firm demand sensitivities).

We examined identification results for both homogeneous product markets and also subsequently differentiated products markets. Analysis of identification in the former case suggests that demand rotators are the key to identification. In the differentiated product case, the results suggest that (i) examining the markups of close-substitute but competing products may be useful and (ii) examining the intensity with which demand and cost shocks to neighboring products are accommodated may sometimes be helpful when understanding the extent of coordination in a market.

In examining the likelihood of collusion, one must assess whether the necessary conditions for collusion exist. Following Stigler (1964), those are agreement, monitoring, and enforcement. The assessment of each of these conditions will typically involve a considerable amount of qualitative evidence although a considerable amount of quantitative evidence can be brought to bear to answer subquestions within each of the three conditions. For example, the European Commission examined the extent to which transaction prices were predictable given list prices to examine market transparency in the Sony–BMG case.

In addition to qualitative analysis of the factors which can affect the likelihood of collusion, it is sometimes possible and certainly desirable to develop an understanding of the incentives to compete, collude, and also to defect from collusive environments.

6.4 Annex: Identification of Conduct in Differentiated Markets

In this annex we follow Davis (2006d), who provides a technical discussion of identification of (i) pricing and demand equations in differentiated product markets and (ii) firm conduct in such markets. In particular, we specify in more detail our example of a market with two firms and two differentiated products. Define the
marginal costs of production which depend on variables $w$ such as input costs to be independent of output so that
\[
\begin{bmatrix}
c_{1t} \\
c_{2t}
\end{bmatrix} = \begin{bmatrix}
\gamma'_1 & 0 \\
0 & \gamma'_2
\end{bmatrix} \begin{bmatrix}
w^1_t \\
w^2_t
\end{bmatrix} + \begin{bmatrix}
u_{1t} \\
u_{2t}
\end{bmatrix}.
\]

Similarly, suppose that demand shifters depend on some variables $x$ such as income or population size which affect the level of demand for each of the products:
\[
\begin{bmatrix}
\alpha_{01t} \\
\alpha_{02t}
\end{bmatrix} = \begin{bmatrix}
\beta'_1 & 0 \\
0 & \beta'_2
\end{bmatrix} \begin{bmatrix}
x^1_t \\
x^2_t
\end{bmatrix} + \begin{bmatrix}
\epsilon_{1t} \\
\epsilon_{2t}
\end{bmatrix}.
\]

Then linear demand functions for the two products can be written as
\[
\begin{bmatrix}
q_1 \\
q_2
\end{bmatrix} = \begin{bmatrix}
\alpha_{01} \\
\alpha_{02}
\end{bmatrix} + \begin{bmatrix}
\alpha_{11} & \alpha_{12} \\
\alpha_{21} & \alpha_{22}
\end{bmatrix} \begin{bmatrix}
p_1 \\
p_2
\end{bmatrix},
\]
while the pricing equations derived from the first-order conditions are
\[
\begin{bmatrix}
\alpha_{11} & \Delta_{12} \alpha_{21} \\
\Delta_{21} \alpha_{12} & \alpha_{22}
\end{bmatrix} \begin{bmatrix}
p_1 - c_1 \\
p_2 - c_2
\end{bmatrix} + \begin{bmatrix}
q_1 \\
q_2
\end{bmatrix} = 0.
\]

The full structural form of the system of equations is
\[
\begin{bmatrix}
\alpha_{11} & \Delta_{12} \alpha_{21} & 1 & 0 \\
\Delta_{21} \alpha_{12} & \alpha_{22} & 0 & 1 \\
-a_{11} & -a_{12} & 1 & 0 \\
-a_{21} & -a_{22} & 0 & 1
\end{bmatrix} \begin{bmatrix}
p_1 \\
p_2 \\
q_1 \\
q_2
\end{bmatrix} = \begin{bmatrix}
u_{1t} \\
u_{2t} \\
v_{3t} \\
v_{4t}
\end{bmatrix},
\]
or, more compactly in matrix form,
\[
Ay_t + C x_t = v_t,
\]
where the vector of error terms is in fact a combination of the cost and demand shocks of the different products,
\[
\begin{bmatrix}
u_{1t} \\
u_{2t} \\
v_{3t} \\
v_{4t}
\end{bmatrix} = \begin{bmatrix}
\alpha_{11} & \Delta_{12} \alpha_{21} & 1 & 0 \\
\Delta_{21} \alpha_{12} & \alpha_{22} & 0 & 1 \\
0 & 0 & \beta'_1 & 0 \\
0 & 0 & 0 & \beta'_2
\end{bmatrix} \begin{bmatrix}
w^1_t \\
w^2_t \\
x^1_t \\
x^2_t
\end{bmatrix} = \begin{bmatrix}
u_{1t} \\
u_{2t} \\
v_{3t} \\
v_{4t}
\end{bmatrix}.
\]

Following our usual approach, this structural model can also be written as a reduced-form model:
\[
y_t = -A^{-1} C x_t + v_t = \Pi x_t + v_t.
\]
The normalization restrictions are reflected in the fact that every equation has a 1 for one of the endogenous variables. This sets the scale of the parameters in the reduced form so that the solution is unique. If we did not have any normalization restrictions, the parameter matrix \( \Pi \) could be equal to \(-A^{-1}C\) or equivalently (in terms of observables) equal to \(-(2A)^{-1}2C\).

In our structural system we have four equations and four endogenous variables. Our necessary condition for identification is therefore that we have at least three parameter restrictions per equation besides the normalization restriction. In general, in a system of demand and pricing equations with \( J \) products, we have \( 2J \) endogenous variables. This means that we will need least \( 2J - 1 \) restrictions in each equation besides the normalization restriction imposed by design.

There are exclusion restrictions that are imposed on the parameters that come from the specification of the model. First, we have exclusions in the matrix \( A \) which are derived from the first-order conditions. Any row of matrix \( A \) will have \( 2J \) elements, where \( J \) is the total number of goods. There will be an element for each price and one for each quantity of all goods. But each pricing equation will have at most one quantity variable in, so that for every equation we get \( J - 1 \) exclusion restrictions immediately from setting the coefficients on other good’s quantities to 0.

Second, the ownership structure will provide exclusion restrictions for many models. Specifically, in the pricing equations, there will only be \( J_i \) parameters in the row, where \( J_i = \sum_{j=1}^{J} \Delta_{ij} \) is the total number of products owned by firm \( i \) (or, under the collusive model, the total number of products taken into account in firm \( i \)’s profit-maximization decision). The implication is that we will have \( J_i \) restrictions.

Third, in each of the demand equations in matrix \( A \), we also have \( J - 1 \) exclusion restrictions as only one quantity enters each demand equation (together with all \( J \) prices); the parameters for the other \( J - 1 \) quantities can be set to 0.

Fourth, we have exclusion restrictions in matrix \( C \) which come from the existence of demand and cost shifters. Demand shifters only affect prices through a change in the quantities demanded and do not independently affect the pricing equation. Similarly, cost shifters play no direct role in determining a consumer’s demand for a product; they would only affect quantity demanded through their effect on prices. Those cost and demand restrictions are represented by the zeros in the \( C \) matrix. Define \( k^D \) as the total number of demand shifters and \( k^C \) as the total number of cost shifters. For each of the pricing equations in \( C \) we have \( k^D \) exclusion restrictions because none of the demand shifters affect the pricing equation directly. Similarly, for each of the demand equations we have \( k^C \) exclusion restrictions since none of the cost shifters enter the demand equations.

Additionally, even though any row in matrix \( C \) will have as many elements as there are exogenous cost variables and demand shifters, there will only be as many new parameters in a pricing equation as there are cost shifters in that product’s pricing
equation. Similarly, there will only be as many new parameters in the demand equation as there are demand shifters in that product’s demand equation.

In addition to the exclusion restrictions we have just described, there are also cross-equation restrictions that could be imposed on the model. Cross-equation restrictions arise, for example, when we have several products produced by a firm. In that case, since prices are set to maximize joint profits for the firm, their pricing equations will be interdependent for that reason. Theory predicts that the way the demand of product \( j \) affects product \( i \)’s pricing equation is not independent of the way the demand of product \( i \) affects product \( j \)’s pricing equation. This gives rise to potential cross-equation restrictions. For example, the matrix \( A \) we wrote down has a total of sixteen elements but in fact it has only four structural parameters. We could impose that the reduced-form parameters satisfy some of the underlying structural (theoretical) relations. For instance, the first elements of rows 1 and 3 are the same parameter with opposite signs. This could be imposed when determining whether the structural parameters are in fact identified from estimates of the reduced-form parameters. The more concentrated the ownership of the products in the market the more cross-equation restrictions we will have, but the fewer exclusion restrictions we will have since we will have fewer zero elements of \( \Delta \). In addition, we will need more exclusion restrictions in each pricing equation to identify all the demand parameters that will be included.