Chapter 17

EMPIRICAL STUDIES OF INDUSTRIES WITH MARKET POWER

TIMOTHY F. BRESNAHAN*

Stanford University

Contents

*I owe a tremendous debt to the editors of this handbook for their encouragement and comments. Their contribution is so large that it would be diminished if I named any of the other very helpful friends and colleagues who contributed ideas and discussion. Though Richard E. Quandt's name does not appear in the reference list, his influence on the field shows in dozens of papers. The remaining errors are mine.

Handbook of Industrial Organization, Volume II, Edited by R. Schmalensee and R.D. Willig

1. Introduction

This chapter treats econometric studies of market power in single markets and in groups of related markets. The recent increase in the number of such studies and substantial advances in the methods for carrying them out constitute a dramatic shift in the focus of empirical work in the industrial organization (IO) field. The new literature treated here is based largely on time series data from single industries, or on data from closely related markets. It has taken a markedly different view of what can be observed, and how economic quantities are to be measured, than earlier work did. At some risk of oversimplifying a growing and varied literature, I summarize the new approach as having these central ideas:

- Firms' price–cost margins are not taken to be observables; economic marginal cost (MC) cannot be directly or straightforwardly observed. The analyst infers MC from firm behavior, uses differences between closely related markets to trace the effects of changes in MC, or comes to a quantification of market power without measuring cost at all.
- Individual industries are taken to have important idiosyncracies. It is likely that institutional detail at the industry level will affect firms' conduct, and even more likely that it will affect the analyst's measurement strategy. Thus, practitioners in this literature are skeptical of using the comparative statics of variations across industries or markets as revealing anything, except when the markets are closely related.
- Firm and industry conduct are viewed as unknown parameters to be estimated. The behavioral equations by which firms set price and quantity will be estimated, and parameters of those equations can be directly linked to analytical notions of firm and industry conduct.
- As a result, the nature of the inference of market power is made clear, since the set of alternative hypotheses which are considered is explicit. The alternative hypothesis of no strategic interaction, typically a perfectly competitive hypothesis, is clearly articulated and is one of the alternatives among which the data can choose.

This "new empirical industrial organization" (NEIO) is clearly somewhat different than the previously dominant empirical method in the field, the structure–conduct–performance paradigm (SCPP).

For the quarter century following the pioneering work of Bain (1951), the focus of the SCPP was the cross-section study of many industries. Industry and firm profits were predicted from various structural measures. The NEIO is partly motivated by dissatisfactions over three maintained hypotheses in the SCPP: (i)

economic price–cost margins (performance) could be directly observed in accounting data, (ii) cross-section variation in industry structure could be captured by a small number of observable measures, and (iii) empirical work should be aimed at estimating the reduced-form relationship between structure and performance.[1] Furthermore, the SCPP has been caught in a kind of gridlock over the question of whether high accounting profits are to be interpreted as a sign of good or of bad performance.[2] The SCPP did, however, introduce something into the field of tremendous value: systematic statistical evidence. The NEIO is an attempt to continue the use of such evidence while returning to the study of single (or related) industries. On its more optimistic days, the NEIO therefore sees itself as taking the best from the two great empirical IO traditions: SCPP and industry case studies.

A typical NEIO paper is first and foremost an econometric model of an industry. Thus the new literature has been able to draw closely on economic theory to guide specification and inference in the empirical models. Quite a bit of this chapter will treat method: it will attempt to provide a review of the manner in which theory has been applied, and the way in which economic inferences have been drawn from the empirical work. Much of the work in the literature has been focused on one set of issues, those surrounding price and quantity determination in oligopoly. The major subtopics covered include the formation and enforcement of tacitly collusive arrangements, the nature of noncooperative oligopoly interaction in the world, the degree of single-firm market power under product differentiation, and the size and determinants of the industry price–cost margin. The next two sections take up the question of measuring market power in concentrated single product industries; Section 2 reviews the various empirical models of monopoly power and of oligopoly interaction, and Section 3 covers the theoretical and empirical arguments for why it is monopoly power that is being measured. Section 4 takes up the question of measuring market power in the presence of product differentiation. These three sections form the bulk of the chapter, as the material they cover forms the bulk of the literature. The chapter has two conclusions. The first is a review of what the NEIO has learned about the extent of market power, the second a biased view of where the literature should go to learn more about the sources of market power.

[1]This is the extreme "structuralist" view associated primarily with Bain. [See Scherer (1980, ch. 1) for the label.] Other positions which take more of a view that conduct is sometimes observable are not importantly different for my purposes. The SCPP and NEIO remain very distinct on how performance and conduct are to be measured in any view.

[2]See Demsetz (1974) and Schmalensee, Chapter 16 in this Handbook. The conventional story is that high profit measures poor performance, i.e. measures the Lerner index ($\mathscr{L}$). Demsetz' alternative interpretation was that high profits measure good performance, i.e. low costs, an argument he buttressed by the observation that much of the profit in concentrated industries goes to large firms. These firms might therefore be large because of cost advantages. Bain (1951) had already provided a "poor performance" interpretation of this observation, however.

2. Oligopoly theory and the measurement of market power

Many of the advances in methodology for measuring market power can be seen most clearly in a stylized econometric model of oligopoly interaction in a single-product industry. The central inference in the stylized model is about firm and industry conduct: the goal is the estimation of parameters measuring the degree of competition. In parallel to laying out the stylized model, I will follow a single specific treatment, namely Porter's (1984) study of strategic interaction among nineteenth-century railroads. This organization is slightly repetitive, but permits the simultaneous treatment of two topics: the relationship of the empirical inferences to oligopoly theory, and their relationship to the data.

The stylized model has three sets of unknown parameters: costs, demand, and firm conduct. The observable variables that are endogenous to the industry equilibrium include industry price and each firm's quantity (sometimes only industry quantity) in time series; price–cost margins are not taken to be directly observable. This focus was reflected early in the rhetoric of the literature: in his title, Rosse (1970) cast the econometrician's problem as "Estimating Cost Function Parameters Without Using Cost Data". The observables are also taken to include variables that shift cost and demand functions.[3] Oligopoly theory is used to specify the equations of the model to be estimated. In this section, the use of theory to specify the model will be emphasized over inference. Inferences about market power will be identified only through refutable implications of the theory contained in the comparative statics or comparative dynamics of oligopoly equilibrium. Another paper title is symptomatic of the form of departure from tradition: Panzar and Rosse (1977a) "Structure, Conduct and Comparative Statics". The question of exactly how the comparative statics identify the conduct parameters is sufficiently important to deserve discussion in a separate section.

The stylized model's specification of demand functions and of cost functions tends to follow fairly standard applied econometrics treatments, so I will be terse in describing them. Specification of conduct parameters is more novel, less standardized, and less well understood, so I will treat it at somewhat greater length.

2.1. Notation, cost, and demand

The dependent variables of the stylized model are market price P_t, and each firm's quantity Q_{it}. Throughout, i will index firms and t will index observations,

[3]A few studies have also included variables which can be interpreted as directly shifting firm conduct: changes in regulation, entry (or the number of firms), mergers, and so on. These kinds of variation is much more prevalent in the studies of closely related markets than in single-industry time series.

normally taken to be in time series. Since we are treating the single-product case, $Q_t = \Sigma_i Q_{it}$ is well defined. For clarity, it is also useful to assume that the demand function for the product contains no intertemporal linkages from durability, habits, learning or other sources. It is convenient to write the demand function in inverse form:

$$P_t = D(Q_t, Y_t, \delta, \varepsilon_{dt}), \tag{1}$$

where Y_t are all variables shifting demand, δ are unknown parameters of the demand function. The demand equation econometric error terms ε_{dt} are written as entering in a potentially nonlinear way and not necessarily treated as a scalar.

In Porter's (1984) study of an 1880s railroad cartel, Q_t is grain shipped by rail from Chicago to the East Coast, measured in tons.[4] The time index t refers to a week between the first week of 1880 and the sixteenth week of 1886. The price data, P_t, are based on a weekly poll taken by the cartel of its members; given the possibility of secret price cutting, P_t is probably to be interpreted as if it were a weighted average of list prices. The demand function takes the constant elasticity form:[5]

$$\log P_t = \delta_0 + \delta_1 \log Q_t + \delta_2 L_t + \varepsilon_{dt}, \tag{1'}$$

where L_t, the only demand-side exogenous variable, is a dummy variable equal to 1 if the Great Lakes were open to navigation. (Shipment by water is a seasonal competitor to the railroads.)

Returning to the stylized model, the treatment of costs similarly takes a familiar form. For reasons of later convenience, we start from the total cost function:

$$C_{it} = C(Q_{it}, W_{it}, Z_{it}, \Gamma, \varepsilon_{cit}), \tag{2}$$

where W_{it} is the vector of factor prices paid by firm i at observation t, Z_{it} are other variables that shift cost, Γ are unknown parameters, and ε_{cit} are econometric error terms, treated as in (1). The distinction between factor prices and other cost shifters is maintained because many important developments in the literature concern the comparative statics of market equilibrium in W. I have put an i subscript on Z and W, since in some applications the comparative statics of equilibrium in the costs of a single firm or subgroup of firms are emphasized.

[4]Porter provides an argument for why this industry should be treated as a single-product one on pp. 302–303. He considers the aggregation of all grain, the dropping of all nongrain shipments, and the dropping of all westbound shipments.

[5]I shall mention a consistent notation throughout, rather than adopting the notation of individual papers. (1′) is Porter's (1) in inverse form.

More commonly, however, Z and W will not have an i subscript, since they will be measured at the industry level. In applications where the cost function being treated is a short-run cost function, Z will include the (SR) fixed factors. The definition of marginal cost follows from (2):

$$MC = C_1(Q_{it}, W_{it}, Z_{it}, \Gamma, \varepsilon_{cit}), \tag{3}$$

where the nonlinearity of cost in the econometric error has been exploited for the first time. $C_1(\cdot)$ is written as random, just as $C(\cdot)$. The error term in (3) is (harmlessly) written the same as (2), though obviously an additive error in the cost function will not appear in MC.

Outside the perfectly competitive model, firms do not have supply curves.[6] Instead, price- or quantity-setting conduct follows more general supply relations:

$$P_t = C_1(Q_{it}, W_{it}, Z_{it}, \Gamma, \varepsilon_{cit}) - D_1(Q_t, Y_t, \delta, \varepsilon_{dt})Q_{it}\theta_{it}. \tag{4}$$

Since $P - D_1Q$ is monopoly MR, (4) has the interpretation of MC = "perceived" MR for oligopoly models. The parameters θ index the competitiveness of oligopoly conduct. As θ_{it}, a positive unknown parameter, moves farther from 0, the conduct of firm i moves farther from that of a perfect competitor. At first glance, (4) appears to contain MC, even though I have said the approach assumes that MC is unobservable. The contradiction is not real, however, only apparent. Marginal cost, $C_1(\cdot)$, in (4) appears as a function of unknown parameters, Γ, not as an accounting datum. Only after Γ has been somehow estimated can MC be calculated; most of the methods described below draw inferences about Γ and θ in the same econometric step.

The next section will take up the conduct interpretation of θ at some length. Obviously letting θ vary both by firm and observation results in an overparameterized model. It is written here in such generality to permit nesting all of the known theories of oligopoly. Any empirical study will put some structure on the way θ varies across time and firms.

In Porter's railroad study, the constant-elasticity demand assumption implies that (4) can be written with the explicit Q_{it} terms substituted out. Furthermore, (4′) is not estimated for the individual firms, but rather aggregated to industry-wide data:

$$\log(P_t(1 + \delta_1\theta_t)) = MC = \Gamma_0 + \Gamma_1 \log Q_t. \tag{4′}$$

It is assumed that firms have heterogeneous marginal costs arising from a log-linear cost function, but that there are no exogenous shifts in the level of

[6]By a supply curve, I mean a solution for Q as a function of P of the equation $P = MC(Q)$.

costs over time. Thus, MC in (4′) is interpreted as the marginal cost of the average firm in the industry at time t. The overparameterization of θ_{it} is solved by aggregation: the interpretation of θ_t is as the average of the conduct parameters of the firms in the industry. Since there is some entry and some acquisition activity during the sample period, the average firm's marginal cost can be expected to shift over time: this is captured by a series of structural dummies S_t, which enter as $\langle \Gamma_2, S_t \rangle$.[7] Similarly, these changes in industry structure may have changed conduct: this is captured in the same dummies. The lack of other cost shifters (factor prices) or demand shifters (the price of lake shipping) is dealt with not by new data but by a close reading of the estimation results in light of the likely omitted variable biases.

To Porter, a crucial question is whether θ_t varies over time because of changes in industry conduct. High θ_t periods are to be interpreted as successful cartel cooperation, low θ_t times as price wars or similar breakdowns in cooperation. Furthermore, it is assumed that the probability of successful collusion is $1 - \pi$. After a transformation, the supply relation ultimately estimated takes the form:

$$\begin{array}{ll} \textit{Supply relation} & \textit{Probability} \\ \log P_t = \Gamma_0 + \alpha^{\mathrm{a}} + \Gamma_1 \log Q_t + \langle \Gamma_2, S_t \rangle + \varepsilon_{cit} & \pi \\ \log P_t = \Gamma_0 + \alpha^{\mathrm{a}} + \alpha^{\mathrm{b}} + \Gamma_1 \log Q_t + \langle \Gamma_2, S_t \rangle + \varepsilon_{cit} & 1 - \pi \end{array} \tag{4''}$$

where α^{a} is a transformation of the conduct parameter in periods of successful collusion, and α^{b} measures the change in conduct when collusion breaks down.[8] Of course, it is clear that α^{a} cannot be separately estimated from Γ_0 on the basis of estimating these equations, but there is considerable interest in estimating α^{b}, the percentage amount by which a breakdown in the cartel changes prices. As the form of (4″) suggests, Porter estimates the system (1′, 4″) by "switching equations" methods, in which the probability π as well as the regular parameters are estimated from the data.

In general, the endogenous variables of the stylized model are P_t and Q_{it}. Many empirical studies, like Porter's, use only industry-wide data and thus have endogenous variables P_t and Q_t. In either event, the core econometric methods are those for simultaneous equations. Some empirical studies proceed by estimating (1) and (4) directly as structural equations: this is attractive, as the supply parameters θ are of primary interest. Other studies, however, may lack data on price or quantity or on firm-specific quantity. In the first case, some reduced form

[7]The four structural dummies capture the entry by the Grand Trunk railway and the entry by the Chicago and Atlantic, an addition to the New York Central and the departure of the Chicago and Atlantic from the Cartel. The $\langle , \rangle$ notation is the inner product.

[8]Let θ_t be able to take on one of two values, θ^{a} and θ^{b}. Then $\alpha^{\mathrm{a}} = -\log(1 + \theta^{\mathrm{a}} \delta_1)$ and $\alpha^{\mathrm{a}} + \alpha^{\mathrm{b}} = -\log(1 + \theta^{\mathrm{b}} \delta_1)$.

will be estimated. In the second, aggregate data will be used. For both purposes, it is useful to briefly define a few related functions. If (1) and (4) are solved simultaneously for all firms, they yield the reduced forms for price and each firm's quantity:

$$P_t = P^*(W_t, Z_t, Y_t, \Omega, \varepsilon_t), \tag{5}$$

$$Q_{it} = Q_i^*(W_t, Z_t, Y_t, \Omega, \varepsilon_t), \tag{5'}$$

where $\Omega = (\delta, \Gamma, \theta)$ is the vector of all parameters, ε_t is the vector of all structural error terms, W_t is the superset of all the W_{it},[9] and Z_t and Y_t are similarly defined. Similarly, we have the reduced form equation for industry quantity, $Q^*(\cdot)$, and so on.

Discussion of the results of these analyses in general must be delayed until after the discussion of the interpretation of the parameters θ. Porter's results, however, will be discussed briefly here. The value of α^b estimated in (4″) implies that price was raised about 40 percent by successful collusion in the industry. The implicit estimates of θ are consistent with collusive behavior approximately as anticompetitive as that implied by the Cournot model.[10] The implicit fraction of the time that the cartel broke down was 28 percent; both the amount and timing of cartel breakdown activity differed somewhat from the patterns detected by earlier scholars and by contemporaneous sources.

The advantages of the modelling technique embodied in (1) and (4) seem to me to be threefold. The first is that the econometric approach is structural. Each parameter has an economic interpretation, and substantial departures of the estimated parameters from expected values can serve as clues to difficulties with or shortcomings of the analysis. For example, Porter estimates a negative Γ_1; he provides an argument for why MC might be downward-sloping, an argument that is more convincing in the context of railroads than it might be in many others. The second advantage of the approach I take up in the next section: if the interpretation of θ is correct, the relationship of the estimates of conduct to theory will be clear. A third advantage is in the section after that: given the structural nature of the econometrics, the reason why the data identify the conduct parameters can be made clear.

[9]If there are only market factor prices, $W_t = W_{1t}$ without any loss of information. If all firms buy factors at different prices, then $W_t = (W_{1t}, W_{2t}, \ldots)$.

[10]The assertion about Cournot depends on the (unidentified) inference that noncooperative behavior is approximately price-taking.

2.2. Alternative treatments of firm conduct and of θ

Supply relations are more general than supply equations because they permit the possibility of nonprice taking conduct, captured in the strategic interaction parameters (θ). Clearly the form in which this nonprice taking conduct is modelled will be central to the inferences about market power drawn in any particular study. The approach covered in Subsection 2.2.1 takes the specification of θ directly from a theory or group of theories. The logical extension of that approach, testing a small number of distinct theories of oligopoly interaction, is covered in Subsection 2.2.2. Another group of papers, covered in Subsection 2.2.3, has had a looser connection to oligopoly theory. A typical paper in it reports its inferences as "estimating oligopoly conjectural variations". There has been enormous confusion about the interpretation of this work, however, primarily because of a language gap. To resolve the confusion, the phrase "conjectural variations" has to be understood in two ways: it means something different in the theoretical literature than the object which has been estimated in the empirical papers. A brief subsection, 2.2.4, discusses the interpretation of estimates when only industry-wide data are used, another area of some confusion. Throughout, individual papers will be used to illustrate how the analytical ideas are actually carried out in the data.

2.2.1. Supply relations from single, specific theories

The first approach to writing (4) and specifying θ uses explicitly theoretical language. Important examples include Rosse (1970), in which the supply relation is derived from the theory of monopoly. Bresnahan (1981) has product differentiation with Bertrand pricing. Porter (1984) uses the Green and Porter (1984) version of Stigler's (1964) theory of collusive oligopoly over time to specify (4); since that theory implies that both collusive and price-war periods will be seen in the data, θ varies over time in the empirical model. After the specification of models based on a single oligopoly theory, I discuss the specification of alternative models based upon a small number of different solution concepts (Bertrand, collusion, etc.) each of which leads to a different version of (4), as in Bresnahan (1980). Geroski (1983) applies this approach to the coffee industry data of Gollop and Roberts (1979), specifying theories in which conduct (and therefore θ) varies across firms (such as Stackelberg leader/follower models). All of the studies which take this approach to specifying the supply relation impose dramatically more structure on the way θ enters (4) than the way I have written it, the structure implied by the theory or theories used. More generally this approach focuses on the estimation of specific models of strategic conduct. The flavor of this analysis can be gotten by considering a few examples. The first of these is

monopoly and the Cournot model, or (in a higher or at least higherfaulting theoretical language) one-shot Nash equilibrium with quantities as the strategic choice variables. This takes the form $\theta_{it} = 1$ for all i, t:

$$P_t = C_1(Q_{it}, W_{it}, Z_{it}, \Gamma, \varepsilon_{cit}) - D_1(Q_t, Y_t, \delta, \varepsilon_{dt})Q_{it}. \tag{6}$$

When there are several firms in the industry, (6) provides an estimating equation for each firm. When there is only one firm in the industry, Cournot behavior (among others) is the same as monopoly behavior. Thus, $\theta = 1$ for monopoly, and (6) holds.

Rosse's (1970) study of American newspapers estimated an equation like (6) simultaneously with one like (1). In his work, Q_{it} is taken to be a three-vector: column inches of advertising, Q_t^{a}; circulation, Q_t^{c}; and column inches of "editorial" (nonadvertising) material, Q_t^{a}. There are two prices associated with these variables: the price of circulation is measured as the average price per subscription copy, and the price of advertising is the average price charged per inch.[11] It is reasonable to expect that the amount of circulation affects advertisers' demand; and that the amount of editorial material and the amount of advertising affect subscribers' demand for the newspaper itself. Using superscripts on prices for advertising and for circulation, the demand equation (1) takes the form:

$$\begin{aligned} P_t^{\mathrm{a}} &= D^{\mathrm{a}}(Q_t^{\mathrm{a}}, Q_t^{\mathrm{c}}, Y_t, \delta, \varepsilon_{at}) && \text{(advertising demand)}, \\ P_t^{\mathrm{c}} &= D^{\mathrm{c}}(Q_t^{\mathrm{a}}, Q_t^{\mathrm{c}}, Q_t^{\mathrm{e}}, Y_t, \delta, \varepsilon_{ct}) && \text{(circulation demand)}. \end{aligned} \tag{1'}$$

The treatment of MR needs also to reflect this interdependence. The appropriate definition of MR is the derivative of the firms' entire revenue with respect to a single product's quantity. For example, the form of (6) for circulation is:

$$P_t = MC^{\mathrm{c}} - Q_t^{\mathrm{c}} D_2^{\mathrm{c}}(Q_t^{\mathrm{a}}, Q_t^{\mathrm{c}}, Q_t^{\mathrm{e}}, \ldots) - Q_t^{\mathrm{a}} D_2^{\mathrm{a}}(Q_t^{\mathrm{a}}, Q_t^{\mathrm{c}}, \ldots), \tag{6^{c}}$$

where the last term on the right is the unusual MR term: the advertising-price effect of higher circulation. Simultaneous estimation of cost function and supply relation exploits the cross-equation restrictions between demand and MR. If demand can be estimated, and MR thereby inferred, an estimate of monopoly MC can be obtained from $MC = MR$. This permits Rosse to take a very sophisticated view of the MC function. For example, his treatment permits transitory shocks to the sizes of firms in his sample to drive a wedge between

[11] Clearly, this three-product approach to the newspaper involves some aggregation of products, such as the different sizes of advertisement that can be purchased. Also, in the interest of uniform notation, I have suppressed some of the details of Rosse's treatment, and have written the demand equations in inverse form.

LRMC and *SRMC*. Obviously, the degree to which such an analysis is convincing depends on the quality of the demand estimates and on the reliability of the *MR* inference; a linear demand specification (like Rosse's) may fit well, yet provide poor estimates of *MR*. As (6^c) suggests, the ability to support this inference against criticism turns on a sophisticated use of the institutional detail of the industry, in this case newspapers.

The other static, noncooperative, symmetric oligopoly model to receive attention is that of Bertrand (one-shot Nash equilibrium with prices as the strategic variables). Since this model is not interestingly distinct from perfect competition in the case of single product industries with flat marginal costs, the greatest attention has been focused on the product-differentiated case. Bresnahan (1981) estimates such a model on 1977 and 1978 cross-sections of automobiles by type. The demand system comes from a spatial treatment of the demand for automobiles by type. Automobiles of different types are assumed to lie in a one-dimensional space, and consumers are assumed to be distributed according to a one-dimensional parameter describing differences in their demands.[12] The own-price and cross-price elasticities are determined by how close products are in this space. Letting X_i be the quality of good i, the demand for a typical good is given by

$$Q_i = \delta_0\big((P_j - P_i)/(X_j - X_i) - (P_i - P_h)/(X_i - X_h)\big) + \varepsilon_{di}, \tag{7}$$

where products h, i, j are adjacent in quality space. Here the price of automobile i is the manufacturer to dealer list price, which is identical to the transactions price for the vast majority of sales. Quantity is model–year production, and the unit of observation (i) is now not the firm but the product. Quality is taken to be ex ante unknown, but to be a hedonic function of observable characteristics:

$$X_i = f(Z, \delta_1) + \varepsilon_{xi}, \tag{8}$$

where Z is a vector of the typical hedonic characteristics: length, weight, horsepower, etc. Under Bertrand competition, the equivalent of (6) is given by

$$\begin{aligned} Q_i = {} & (P_i - MC_i)\,\delta_0\big(1/(X_j - X_i) - 1/(X_i - X_h)\big) \\ & + d_{ij}P_j\,\delta_0/(X_j - X_i) + d_{ih}P_h\,\delta_0/(X_i - X_h) + \varepsilon_{si}, \end{aligned} \tag{6''}$$

where d_{ij} is a dummy for whether the firm that produces product i also produces product j. As in the Rosse treatment, the definition of *MR* is taken on a

[12]The demand system is that of "vertical product differentiation"; see Prescott and Visscher (1977).

whole-firm basis. In this cross-section work, marginal cost is a function only of product type Z, not of factor prices. It is clear from (6″) that the closeness of competitive products is the key determinant of market power in such a model. Bresnahan (1981) reports two related findings: the large price–cost margins appear to be on the larger vehicles, and the larger vehicles appear to be much farther apart in product–quality space.[13] The obvious problem with this kind of modelling is the highly restrictive assumptions made about the form of the demand system. Bresnahan (1980, 1981), like Rosse (1970), attempts to deal with these primarily by analysis of the residuals; the devices include introducing firm dummies, adding ε_{xi} to (7), and so on.

A second class of models of some importance have separate leaders and followers. Of these, the first is the Stackelberg leader model. I adopt the notational convention that there are I firms in the industry, so $i = 1, \ldots, I$. The Stackelberg model writes (6) for firms $2, \ldots I$, but for the "leader," firm 1 writes:

$$P_t = C_1(Q_{1t}, W_{1t}, Z_{1t}, \Gamma, \varepsilon_{c1t}) - D_1(Q_t, Y_t, \delta, \varepsilon_{dt})Q_{1t}(1 + \theta_S), \tag{7}$$

where θ_S is obtained by first solving (6) simultaneously for all firms except firm 1; this yields functions $Q_i(Q_1, \ldots)$. Then θ_S is the derivative of the sum of these with respect to Q_{1t}.

A closely related model, the dominant firm model, has been estimated by Suslow (1986). In her treatment, there is a fringe of firms that are price-takers, the producers of "scrap" or recycled aluminum. The dominant firm is Alcoa, and the sample refers to the interwar period in which Alcoa had a monopoly on the production of new aluminum. In a general dominant firm model, the fringe firms have supply relations which are supply curves. Without loss of generality, the supply curve of the entire fringe is determined by its collective MC:

$$P_t = C_1(Q_{ft}, W_{ft}, Z_{ft}, \Gamma, \varepsilon_{cft}), \tag{10}$$

$$Q_{ft} = S(P_t, W_{ft}, Z_{ft}, \Gamma, \varepsilon_{cft}), \tag{10′}$$

where $Q_{ft} = \sum_{i=2,I} Q_{it}$, and W and Z are taken to be common to all firms in the fringe: the function in (10′) is the inverse of that in (10). Suslow (1986) permits the dominant firm and the fringe to sell somewhat differentiated products. For clarity I will assume they sell the same product. In her paper, Q_{ft} is the physical quantity of "secondary" aluminum recycled, and P_{ft} the recycled price. The demand shifters Z_{ft} include an estimate of the stock of aluminum available for recycling. The leader in a dominant firm model is taken to be sophisticated, and

[13] This confirms the view of many industry observers, based on accounting profits data. It received further confirmation from a study of auto dealers' (as well as manufacturers') prices in Bresnahan and Reiss (1985).

to set:

$$P_t = C_1(\cdot) + D_1(\cdot)Q_{1t}(1 + S_D), \tag{11}$$

$$S_D = D_1(\cdot)S_1(\cdot)/(1 - D_1(\cdot)S_1(\cdot)). \tag{11'}$$

In Suslow's paper, Alcoa's MC is shifted by cost variables, such as the firm's accounting average variable cost, which clearly do not enter the supply function of the fringe. Price is measured by realized average revenue, quantity by sales in physical units.[14] She finds that Alcoa enjoyed considerable monopoly power, despite the competition of the fringe, since the quantity S_D is estimated to be small. This approach has some of the same interpretational difficulties as the earlier one: MR still needs to be carefully selected. There is also the need to establish or test that the selected firm is in fact the leader; Suslow accomplishes this by arguing from the industry structure of new aluminum (monopoly) and of recycling (many firms).

Leaving static noncooperative theories, leads us to dynamic theories. The problem of repeated oligopoly interaction has received a great deal of theoretical attention, since it is reasonable to presume that long-run considerations might reduce the competitiveness of oligopoly conduct. Oligopolists might go along with a collusive arrangement, even though deviations are profitable, if they recognize that deviations will lead to a general breakout of competition. A pathbreaking model is that of Stigler (1964): if it is purely this self-interest which holds collusive arrangements together, they should be expected to sometimes break down as the parties grow suspicious of each others' motives and behavior. In an uncertain environment, firms will not always know whether secret price-cutting has occurred, and this will lead to some price wars even if there is no actual secret price-cutting. Thus, one should expect cartels to break up and to reform: data on a cartelized or tacitly collusive industry should show both periods of successful cooperation and periods of outright competition. Empirically, of course, this will show up as time-varying θ.

Several recent formalizations have put considerably more structure on the theory, and on the pattern of industry equilibrium over time. The first of these is Green and Porter (1984) (GP). In their theory, price wars actually break out because of shocks that firms cannot observe. When profit drops in the SR, firms cannot tell with certainty whether this is because of chiselling on the cartel agreement or because the aggregate situation has worsened. Bad enough shocks will trigger price wars. GP, and a later paper by Porter (1983) treat the problem of "design" of a cartel arrangement. In their theory, firms decide (collectively)

[14]Suslow's treatment of the marginal cost of a monopolistic that can produce to store as inventory is somewhat shortchanged by this discussion.

how high to raise prices in collusive periods, how trigger happy to be, and how long price wars should be. The degree of competition in a price war is taken as exogenous (it is assumed to take the Cournot form). The theory predicts that there will be alternating periods of price war and of successful collusion. The length of the price wars and the size of a shock needed to trigger a price war are picked to maximize industry profit. The intuition of these and of other Stigler-esque theories comes from incentive economics. Why should collusive firms not raise price all the way to the monopoly level? If they did, it would give too much of an incentive to deviate from the cartel arrangement. Why should price wars last a while before the cartel is reformed? Otherwise, the possibility of a price war could not deter any opportunistic behavior.

Abreu et al. (1986) find cartel designs that are even more profitable for firms than the GP ones, by permitting more complex arrangements among firms. These designs still have alternating periods of successful collusion and of price wars, but now there are "triggers" both for beginning a price war and for ending one. Thus, the length of price wars is random. Furthermore, the amount of competition in a price war is endogenous to the model: it, too, is optimized to maximize the returns to the colluders. Rotemberg and Saloner (1986) take a somewhat different tack. In their model, the environment in which firms operate shifts over time. As a result, the optimal cartel price shifts as well. Suppose (as in the analysis they provide) that current demand is not a predictor of future demand. Then in boom periods, the gain from defecting from a cartel is unusually large at any given price. Therefore the cartel must set an unusually low price to reduce the incentive to defect. The reverse line of argument holds in demand busts.

These various theories have in common the idea that in an imperfectly informed world, "successfully" collusive industries will have periods of cartel pricing and periods of competition.[15] In general, they imply models with θ_{it} not necessarily equal to $\theta_{i\tau}$. The theories differ somewhat in the expected time-series behavior of these two regimes, as the exact equations determining passage from one regime to the other vary between theories. Green and Porter theories, for example, seem to suggest that θ changes from the collusive to the competitive value when there is an unanticipated shock to demand, and that returns to collusion will follow with a fixed lag. Abreu et al. have θ following a time-series process driven by demand shocks as well, but the process is Markov. The Rotemberg and Saloner theory suggests endogenous strategic variation in conduct within the collusive regime. It is easy to imagine other theories of success or failure in tacit collusion which predict different patterns; taking all of these theories at once would lead to even more complex potential time-series behavior for θ.

[15]Rotemberg and Saloner do not write their theory in this way, because they assume that there is no imperfect information. This is clearly an assumption of convenience (irrelevant to the point they are making) rather than a central feature of the model.

I have already discussed the details of Porter's (1983) approach to time-varying θ, and the discussion of the previous paragraph begins the discussion of the relationship of that approach to theory. For clarity, let me reprint (4″) here. The supply relation estimated in Porter (1983) [see also Lee and Porter (1984)] is

$$\begin{array}{ll} \textit{Supply relation} & \textit{Probability} \\ \log P_t = \Gamma_0 + \alpha^{\mathrm{a}} + \Gamma_1 \log Q_t + \langle \Gamma_2, S_t \rangle + \varepsilon_{cit} & \pi \\ \log P_t = \Gamma_0 + \alpha^{\mathrm{a}} + \alpha^{\mathrm{b}} + \Gamma_1 \log Q_t + \langle \Gamma_2, S_t \rangle + \varepsilon_{cit} & 1 - \pi \end{array} \tag{4''}$$

By estimating only a single probability of colluding or competing, π, Porter puts less structure on the problem than the theories suggest. Estimation of π in (4″) can tell us how frequently cartels break down, but cannot tell us what predicts breakdowns and reformations – the area in which the theories disagree. At a minimum, one would like π to be state dependent. Alternatively, whether changes in markup from regime b to regime a appear statistically to be the same event as shocks to demand could be investigated. An initial attempt to investigate the time-series behavior of the regimes is contained in Porter (1985). Rotemberg and Saloner (1986) provide a somewhat less structural analysis of the same question, using Porter's reported regime shifts as well as the time-varying pattern of automobile industry competition and collusion reported in Bresnahan (1987). I see little in these investigations that differentiates among the various theories; the time-series behavior of conduct is harder to estimate than the average level of conduct.

It will be surprising (to me, anyway) if further similar investigations succeed in strongly differentiating among the theories. Investigations of the question, Do there seem to be price wars?, can essentially take advantage of the data in all periods. The investigator is trying to find out that there are two distinct regimes. On the other hand, an investigation of the question, What sets price wars off and what stops them?, necessarily will have a much smaller sample size. Instead of having the number of observations equal to the number of sample periods, it has the number of observations equal to the number of price wars. This low assessment of the success probability of further testing is not particularly troubling (to me, anyway) since the various theories are identical for practical purposes.

A final remark on this subsection goes better here than anywhere else. Industrial organization economists have frequently felt that their field was data-starved, or at least starved of appropriate data. The studies reviewed here show this to be false. The dissertation of work of Rosse, Bresnahan and Suslow each involved collection of data from industry sources in the public domain, with no reliance on government sources for endogenous variables. Porter's work, too, is based on industry sources.

2.2.2. *Supply relations from a small set of theories*

The papers treated in the previous subsection took a single (in the last case, complex!) theory as a starting point for the specification of the supply relation. A closely related development has been attempts to estimate the supply relations of a small number of different theories and to test among them. This is the approach of Bresnahan (1980, 1987). The data used are again cross-sections of automobiles, this time from the mid-1950s. The demand equations are exactly as in (7). But the supply relations for joint monopoly pricing as well as those for Bertrand pricing are estimated. These differ from (6″) only in that $d_{ij} = 1$ whether the neighboring products are produced by the same firm or not. (MR for a joint monopoly of all firms is the derivative of industry revenue with respect to quantity.) The estimates, otherwise much like those described above, show collusive behavior in some years, but competitive (Bertrand) behavior in 1955. This provides a strategic explanation of part of the large expansion of auto production in that model year. In related work, Geroski (1983) and Roberts (1983) put the structure implied by a small number of leader–follower type theories on the data for coffee roasting firms in the United States, finding that the smaller firms in the industry are price-taking followers. The leading firms do not appear to joint profit maximize (even given the constraint implied by the fringe's supply curve) but do behave less competitively than Cournot firms.

Since each of the theories reviewed up to now in this section is associated with different values of the parameters in θ, one might decide to treat θ as a continuous-valued parameter and estimate it. This approach is the one discussed in the next section. It risks the possibility that values of θ which are "in between" existing theories will be estimated, but that is hardly a disaster. The distinction between the continuous-valued θ and the distinct θ's from several theories is purely econometric. The researcher who has estimated θ from a continuum will test theories by nested methods. The other researcher will use non-nested tests to distinguish among the few theories entertained, as I did in the work described in the previous paragraph.[16]

2.2.3. *Supply equations in "conjectural variations" language*

The second approach to specification of (4) uses "conjectural variations" (CV) language and treats conduct as a continuous-valued parameter to be estimated. In this language, the parameters describing firms' conduct are not written in the form of θ in (4). Instead, the parameters are described in terms of firms' conjectural variations, that is, their "expectations" about the reaction of other firms to an increase in quantity. These parameters are typically allowed to take

[16]See Geroski, Phlips and Ulph (1985) for a different position on this issue.

on any values in a broad range. An important early paper is Iwata (1974), whose title "Measurement of Conjectural Variations in Oligopoly" is illustrative of the thrust of the literature.[17] He saw the question as inferring where, in the continuum between perfect competition and monopoly, the Japanese plate glass industry lay. Another important early paper was Gollop and Roberts (1979), which permitted conjectures to vary across firms. Later work by Spiller and Favaro (1984) and Gelfand and Spiller (1987) on banking continued this emphasis on heterogeneity of firm conduct.

In the CV language, the empirical supply relationship is written in the form:

$$P_t = C_1(\cdot) - D_1(\cdot)Q_{it}\big(1 + r_i(Q_{it}, Q_{jt}, Z_{it}, \Phi)\big). \tag{12}$$

Here Q_{jt} is the vector of all other firms' quantities, and the dependence of cost and demand on exogenous variables and parameters can be temporarily suppressed. Note first that the only difference between (4) and (12) is that the term θ_{it} in (4) has been replaced by the term $1 + r(\cdot)$. This does suggest some practical differences between the papers discussed in the previous subsection and the CV papers. The CV papers tend to permit all values of r_i, not just those associated with particular theories. There is clearly nothing fundamental about this: as discussed above, one could easily treat θ in (4) as lying in a continuum. Second, the CV papers have tended to emphasize the relationship between firm size and conduct: hence the explicit dependence of $r_i(\cdot)$ on quantities in (12). There is also an implicit dependence: different values of Φ are often permitted firms in different size classes.

As a matter of logic, there is absolutely no difference between (4) and (12) in general, since the identity $\theta_{it} - 1 = r_i(Q_{it}, Q_{jt}, Z_{it}, \Phi)$ implies that the two specifications can nest the same models. (Nothing in the previous subsection implied that θ_{it} needed to be a constant, though many theoretical models have constant θ_{it}.) Therefore I will use θ to mean $1 + r$, and vice versa according to convenience, in the rest of this chapter. I cannot overemphasize this point: there is *no difference* between saying what the "conjectural variation" is and saying what theory of oligopoly holds in the data. Misunderstandings of the phrase "conjectural variations" to mean something other than it does mean in the empirical papers has been rife, however.

There are some cases where no misunderstandings arise: for example, the Cournot model is labelled "zero conjectural variations". It is usually innocent to think of Cournot firms as ones that "expect" other firms' quantities to be

[17]Iwata (1974) differs from many of the papers surveyed here in that it assumes that accounting data reveal PCMs. Its role in using the data to try to draw inferences about conduct was very influential, however.

constants. Similarly, $r(\cdot)$ of -1 is perfect competition, and it is completely correct to say of a competitive firm that it "expects" price to be a constant.

The linguistic difficulties arise in other cases because allusion to underlying theoretical models is typically made in the same "expectations" or "conjectures" language. Suppose we think of (12) as the derivative of a single-firm profit function:

$$\max_{Qi} D\left(\sum_j Q_{jt}(Q_i), \dots\right) Qi - C(Q_i, \dots) \tag{13}$$

Equation (13) has every other firm's quantity as a function of Q_i. Then we read $1 + r_i$ in (12) by

$$1 + r_i(\cdot) = \mathrm{d}Q_i/\mathrm{d}Q_i + \sum_j \mathrm{d}Q_j/\mathrm{d}Q_i, \tag{14}$$

where the sum over firms j is $j \neq i$. Some minor variations in language occur, but the typical understanding is that the $\mathrm{d}Q_j/\mathrm{d}Q_i$ terms measure the way firm i "expects" firm j to "react" to an increase in quantity.[18]

It is when the estimated "conjectures" are ones which lead to prices close to the collusive level that the simple "expectations" interpretation is suspect. The point can be seen under the assumption that all firms have the same cost curves and "conjectures". Let there be I firms in the industry. Suppose we get the conjectures associated with the collusive level of output, $(1 + r_i) = I$:

$$P + C_1(\cdot) - D_1(\cdot) I Q_i \tag{15}$$

for each firm, since that particular value is the solution to the problem "maximize the profits of all I firms". How, then, are we to interpret $1 + r_i = I$? Taken literally, the coefficient says that the firm picks its output to maximize industry profits because it "expects" the other firms in the industry to match its output: $r_i > 0$ is an expectation of matching behavior, and the r_i of (15) imply an expectation of proportional matching: $\mathrm{d}Q_j/\mathrm{d}Q_i = 1$.

In a great many treatments of oligopoly as a repeated game, firms produce output in most periods according to (15), but the reason they do is that they *expect* deviations from that level of quantity to lead to a general breakdown in restraint. (The exact form this would take varies among the theories: see Subsec-

[18]Some papers describe estimates of (11) as "estimating conjectural variations", other as "estimating firms' first-order conditions". If (11) is solved for Q_i it is "estimating firms' reaction functions". Adding up the first-order conditions and interpreting the result in light of Cournot theory even leads to the language "estimating the equivalent number of firms". Obviously, there are no important distinctions between these languages.

tion 2.2.1 above, and see Shapiro, Chapter 6 in this Handbook, for a much fuller treatment.) Thus, the matching behavior is unobserved; firms expect that if they deviate from the collusive arrangement, others will too. This expectation deters them from departing from their share of the collusive output.

The crucial distinction here is between (i) what firms believe will happen if they deviate from the tacitly collusive arrangements and (ii) what firms do as a result of those expectations. In the "conjectural variations" language for how supply relations are specified, it is clearly (ii) that is estimated. Thus, the estimated parameters tell us about price- and quantity-setting behavior; if the estimated "conjectures" are constant over time, and if breakdowns in the collusive arrangements are infrequent, we can safely interpret the parameters as measuring the average collusiveness of conduct.[19] The "conjectures" do *not* tell us what will happen if a firm autonomously increases output (and thereby departs from the cartel agreement), the normal sense in which theoretical papers would use "conjectural variations".

A second set of interpretive questions arises when the variation across firms in r_i is modelled by estimating Φ in $r_i(Q_{it}, Q_{jt}, Z_{it}, \Phi)$, as in Gollop and Roberts (1979), Spiller and Favaro (1984), and Gelfand and Spiller (1987). A speculative interpretation is that the dependence of the $r_i(\cdot)$ on own quantity tells us something about "mutual forbearance". The notion here is that one can read the derivative of $r_i(\cdot)$ with respect to own quantity as revealing something about how competition would change if firms were to deviate from agreed production. This seems to trip over the distinction just raised between what the conjectures enforce and what they are. I would therefore not use the "mutual forbearance" language.

The use of different strategic parameters for firms of different sizes suggests that the CV models may provide a strategic explanation of the size distribution of firms, since their endogenous variables include the quantity produced or market share of each firm. By permitting the firms to have different conduct, such models permit ex ante identical firms to be of different sizes in equilibrium. For example, the three papers just listed all permit different conjectural variations for different size classes of firms. When economies of scale are permitted (as in these papers) this can provide information about the details of large-firm–small-firm interaction. It was this which motivated Geroski (1983) and Roberts (1983) to test specific theories in this context: they were particularly interested in questions

[19]All of these theories of going along with a collusive arrangement suggest that it is not necessarily the fully collusive outcome that will arise. Somewhat smaller r_i's than in (13), and therefore somewhat larger equilibrium output, can also arise (when worse information makes collusive arrangements harder to enforce, for example.) In these circumstances, firms' "expectations" about what would happen if they deviated from planned output might be exactly the same, but their production levels would be different. That is to say, there is *no information* about firms' expectations contained in the estimates of the $r_i(\cdot)$. The $r_i(\cdot)$ tell us how close to a completely collusive outcome the expectations induce. The only thing they tell us about expectations is that the expectations are sufficient to deter departure from the normal arrangement.

like: "Are large firms, taken as a group, leaders and small firms followers?" I think this very interesting line of research is still incomplete, since alternative explanations of the size distribution of firms have yet to be introduced. In particular, it would seem important to let different firms have different cost functions. Then it would be possible to test the alternative hypothesis that the size distribution is driven by relative costs despite identical conduct parameters.[20]

2.2.4. *Work with aggregate industry data and aggregate conduct*

In many circumstances, the lack of single-firm data will prohibit estimation of supply relations for individual firms. Instead aggregate industry data must be used. One approach is to simply rewrite (4) in aggregate form:

$$P_t = C_1(Q_{it}, W_{it}, Z_{it}, \Gamma, \varepsilon_{cit}) - D_1(Q_t, Y_t, \delta, \varepsilon_{dt})Q_{it}\theta_{it} \tag{4}$$

becomes

$$P_t = C_1(Q_t, W_t, Z_t, \Gamma, \varepsilon_{ct}) - D_1(Q_t, Y_t, \delta, \varepsilon_{dt})Q_t\theta_t. \tag{16}$$

This approach is familiar from as different works as Appelbaum (1979) and Porter (1984). Under the null hypothesis of no market power, $\theta = 0$, the interpretation of C_1 as industry marginal cost is clear. When there is market power, however, different firms will almost certainly have different marginal costs in equilibrium. Analysis like that of Cowling and Waterson (1976) shows that noncooperative oligopoly will tend to have variation in price–cost margins across firms unless they have identical, constant MC. In these circumstances, a stable industry marginal cost curve need not exist, and interpretation of (16) may be clouded.

There are clearly some circumstances in which industry-wide marginal costs are equal to each firm's MC; consider the example of a cartel that succeeds in rationalizing production. Generally, however, (16) will need to be interpreted as some sort of average. In this context, an error in the interpretation of (16) has crept into the literature: the assertion that $\theta_{it} = \theta_{jt}$ is an implication of theory.[21] This is clearly incorrect, as there is nothing in the logic of oligopoly theory to force all firms to have the same conduct. It is better to follow Cowling and Waterson (1976) and interpret the aggregate θ_t in (16) as industry average conduct, and $(P - C_1)/P$ as the industry average markup.

[20]This is not a trivial task. The existing specifications of a common cost function with scale economies do permit heterogeneity in the level of MC across firms. The interesting hypothesis turns on whether firms of different sizes would have similar MC at the same output, a tricky measurement problem.

[21]Appelbaum (1982) made this argument; it has been picked up by Lopez (1984).

2.2.5. Final thoughts on θ and r_i

Both the work closely based on economic theory and the conjectural variations work has overwhelmingly cast its (logical) tests of theories of strategic interaction as (statistical) hypothesis tests about θ. These studies tend to state their problem as one of measuring conduct or strategic interaction rather than as of measuring market power. Thus, they focus on hypothesis tests about θ or r_i. In this connection, an important observation was made by Appelbaum (1979): setting the entire vector θ to zero in (4) or (equivalently) setting all of the r_i to -1 imposes the restriction of price-taking conduct. Thus all approaches to specifying (4), even those which do not use explicitly theoretical language, can be thought of as "Testing Price-Taking Behavior", Appelbaum's title. This would not be particularly interesting, except that the particular alternative hypothesis against which price-taking has been rejected is one with market power.

The next section treats the question of what constitutes an adequately rich specification of cost and demand so as to permit a reasonably convincing case that a strategic interaction hypothesis is in fact being tested. The section will show that the hypothesis of market power is in fact identified on reasonable data. This is an important step: if it were merely true that perfect competition were rejected, and that no positive indicia of power over price were found, the observation that the results might be a statistical artifact would be compelling. For now, however, let me point out an extremely important advance implicit in this approach. The alternative hypothesis includes price-taking behavior: when it is rejected, it is rejected against specifications based on theories in which firms succeed in raising prices above *MC*. Only econometric problems, not fundamental problems of interpretation, cloud this inference about what has been determined empirically.

3. How the data identify market power

An advantage of the use of structural econometric models and explicit theories of industry equilibrium is that the class of models the data are allowed to treat is made explicit. Thus, the class of alternatives within which the inference of monopoly power has been drawn can be clearly stated. This in turn limits the number of alternative explanations which can be reasonably advanced. More importantly, however, this procedure permits an explicit answer to a central question: Why is the economic inference of monopoly power identified by the data? What implication of the theory of perfect competition has been found to be false when market power is measured by these methods? These questions do not arise for SCPP methods, of course, since SCPP takes price–cost margins to be observable. In the NEIO, PCMs are to be estimated, and it is therefore of

immediate interest what observable feature of the data, and what natural experiments, reveal them to the analyst. To date, there are four new classes of identification arguments: (i) comparative statics in demand, (ii) comparative statics in cost, (iii) supply shocks, and (iv) econometric estimation of MC. This section takes up these arguments in turn. There is a fifth area, the comparative statics in industry structure, which is familiar in its logic; price is predicted as a function of the number of firms or of other concentration measures. This area is, I believe, awaiting its identification arguments, for reasons I lay out in Subsection 3.5.

3.1. Comparative statics in demand

A natural empirical procedure is to write out a system consisting of (1) and of one (4) for each firm, under one of the parameterizations of θ from the previous section. Some appropriate econometric method yields estimates of δ, Γ and especially θ, under the assumption that they are all separately identified. Thus, the same data, and inferences based on the comparative statics of equilibrium, provide estimates of cost, demand and conduct parameters. This natural procedure, however, should make clear why it is that it works. What idiosyncratic implication of perfect competition has been rejected, what idiosyncratic implication of market power or oligopoly interaction has been observed in the data? The first approach to this question was to ask whether the comparative statics of monopoly, oligopoly, and perfect competition models are logically distinct, and if so how.

The encouraging answer to this question can be seen in a very simple model, in which only an aggregate supply relation is estimated and in which θ is taken to be a constant over time.[22] To further simplify, assume that the econometric errors enter both demand and supply in an additive way, and that the slope of the demand curve does not depend on Q_t. Then we write (1) and (4) as:

$$P_t = D(Q_t, Y_t, \delta) + \varepsilon_{dt}, \tag{17}$$

$$P_t = C_1(Q_t, W_t, Z_t, \Gamma) + \varepsilon_{ct} - D_1(Y_t, \delta)Q_t\theta. \tag{18}$$

Obviously, δ can be estimated: (17) has only quantity as an included endogenous variable, and instruments are available from the cost function. Call any estimate which has been obtained δ^*. Then one could calculate the "datum" $D_{it}^* =$

[22]The analysis of this section is based on Bresnahan (1982) and Lau (1982), which make the argument presented here in a more precise way.

$D_1(Y_t, \delta^*)$, and consider estimation of the equation:[23]

$$P_t = C_1(Q_t, W_t, Z_t, \Gamma) + \varepsilon_{ct} - D_{it}^* Q_t \theta. \tag{19}$$

This has two endogenous variables: Q_t, which occurs in cost, and $D_{it}^* Q_t$, the variable whose coefficient is θ. When can θ be estimated? Econometrically, two conditions must hold. First, the two endogenous variables must not be perfectly correlated. The definition of D_{it}^* makes clear that they will in fact be perfectly correlated unless D_1, the slope of the demand curve, depends on Y_t. Second, instruments must be available for both endogenous variables. This will obviously be the case if Y_t is a two-vector, with one element of Y_t entering D_1, the other not. More generally, Lau (1982) has shown that a sufficient condition for identification is that the inverse demand function $D(\cdot)$ cannot be written in a way such that Y_t is separable from Q_t; since Q_t is a scalar, this clearly requires that Y_t be a two- (or more) vector.

The economics of this argument can be stated very simply. The comparative statics of models with market power have a particular role for changes in the slope of the demand curve. Suppose that the exogenous variables entering demand can (in principle) perform a particular natural experiment: they can rotate the demand curve around a given point, say the industry equilibrium point. Under perfect competition, this will have no effect: supply and demand intersect at the same point before and after the rotation. Under any oligopoly or monopoly theory, however, changes in the elasticity of demand will shift the perceived marginal revenue of firms. Equilibrium price and quantity will respond. Thus, the comparative statics of models with monopoly power do have idiosyncratic predictions, and the market comparative statics of perfect competition are distinct from those of monopoly.

The papers that have relied on this identification principle are those that have had a good natural experiment shifting the demand equations in an appropriate way. In Just and Chern (1980), it is the buying side which has the market power: a concentrated manufacturing industry buys tomatoes from atomistic farmers. The crucial exogenous variable was a change in harvesting technology, one which they argued changed the elasticity of supply.[24] In Bresnahan (1981, 1987) the firms possibly having market power are sellers of automobiles: demand elasticities depend on how close to one another firms' products are in a product space. More generally, we might think about the most attractive applications of this identification argument. The two elements of Y_t might be something measuring the size of the economy, such as national income, and a second variable

[23]This is describing an econometric procedure so ugly that no one would ever undertake it: it does however, show that and why more powerful techniques can identify.

[24]Just and Chern treat the case of oligopsony, so it is the supply elasticity which is shifted by the technical change. This has no effect on the logic of the argument.

measuring the price of a substitute or substitutes.[25] Use of this method of identification obviously turns on the ability to estimate the demand elasticities in a reliable way. The analyst will need to answer such criticisms in any particular case with standard econometric techniques for investigation of the robustness of results. Many observers have noted that alternative interpretations of the Just and Chern technology shift are available. The dependence of the Bresnahan automobile results on the exact ordering of the products in quality space is frequently pointed out. Equation (8), above, introduces an unobserved error into product quality, thereby relaxing the assumption that the ordering of products can be determined solely on the basis of observable proxies for quality.

A further refinement of this line of reasoning is available if one is prepared to assume that marginal cost is homogeneous of degree one in observed factor prices.[26] Then the θ in (17) or the more general (4) is clearly identified. The analyst interprets the coefficients of Q_t that are interacted with cost shifters to be part of MC, and those that are not to be part of the perceived marginal revenue term with coefficient θ. (Since MC is homogeneous of degree one in factor prices, Q_t cannot enter MC in a way such that it is not interacted with one or more elements of W.) Implicitly, this is how early papers like Appelbaum (1979, 1982) obtained identification. The homogeneity property is guaranteed if MC is obtained by differentiation of a total cost function, possibly one of the forms (translog, generalized Leontief, etc.) commonly used in factor demand system estimation. Then all coefficients in the supply relation which are *not* functions of factor prices are interpreted as indicators of market power. This line of argument obviously leans very hard on the assumption that the functional form of MC is correct and that all of the true marginal prices of the inputs can be observed. The true marginal price of capital is a potential problem for such studies. The way for such a study to rebut alternative interpretations of the results is to explore the robustness of the results to alternative treatments of MC: alternative functional forms, alternative treatments of the quasi-fixity of capital and labor, nonaccounting definitions of the cost of capital, etc.

3.2. Comparative statics in cost

An alternative comparative statics analysis is that of Panzar and Rosse (1977a, 1977b, 1987) (PR). PR propose two separate ideas: first, that an appropriate method for analysis is estimation of the reduced form, with particular attention given to the coefficients of factor prices W_t. The second idea is that reduced-form

[25]Clearly, this idea leans on an older line of thought, especially in connection with cartelization in the markets for primary commodities. See Scherer (1980, pp. 229ff).

[26]See the discussion of duality and cost below, and in Panzar and Rosse (1987).

revenue equations are likely to be estimable in many circumstances, since revenue is likely observable even where price and quantity are not. In light of this, let us begin with the reduced-form revenue equation, called $R^*(\cdot)$. This is the total revenue for a single firm. R^* is equal to equilibrium quantity (which depends on cost, demand and conduct) times equilibrium price (which has the same determinants). The observable shifters of cost and demand – Z, W, and Y – all enter this function. In the case of monopoly, solve the single firm's (1) and (4) for the price and quantity as a function of exogenous variables, parameters, and error terms, and then continue by calculating revenue. This yields a reduced-form revenue function of the form:

$$R_{it} = R^*(W_{it}, Z_{it}, Y_t, \text{params}, \varepsilon_t). \tag{20}$$

Equation (20) will be written in the same form when there is more than one firm in the market, and (1) and (4) have been simultaneously been solved for several firms. It will depend on the exogenous variables for all firms, of course. Let $R_W(W_{it}, Z_{it}, Y_t, \text{params}, \varepsilon_t)$ be the vector of derivatives with respect to all inputs, and let $\langle , \rangle$ be the inner product. The PR statistic is

$$H_R = \langle W_{it}, R_W(\cdot)\rangle / R^*(\cdot), \tag{21}$$

the sum of the elasticities of the reduced-form revenues with respect to all factor prices.[27] The PR statistic requires little data on endogenous variables in the system, although it does need *all* of the variables which shift demand or cost. The analyst proceeds by estimating the reduced-form revenue equation, R^* including all available information on W, Z and Y. Then H_R is calculated.[28] A particular advantage of estimating only a revenue equation is that no quality correction need be made to define a true price for the industry. The product may be better in some markets, so that its price per pound overstates its true price there. Yet this tricky data problem does not affect the reduced-form revenue equation. More generally, H_R can obviously be calculated whenever the structural system (1), (4) has been estimated. But $R^*(\cdot)$ can also be estimated in many circumstances when the structural equations, especially the supply relations, cannot.

The PR statistic has a clear economic interpretation in several cases. First, suppose that the market studied is a monopoly. Then $H_R < 0$. A very general proof is available in PR: the intuition, however, can be seen here, for the case of

[27]Existing applications have either parameterized the reduced form so that this statistic is a constant or have reported estimates near the center of the sample in some sense.

[28]The idea that W are the only exogenous variables needed to estimate H_R has been somewhat oversold. In some circumstances, the estimating equation for revenue is misspecified when only cost variables are included as exogenous variables. As the results discussed below imply, it is appropriate to omit demand shifters only when the hypothesis being tested is perfect competition. The test for monopoly requires a revenue function with all exogenous variables.

"elementary" monopoly. By elementary monopoly I mean simply the case in which a single firm picks only a single quantity. I will suppress the econometric errors and the parameters. Let $R(Q, Y)$ be monopolist's revenue function in the usual sense, so that R_Q is MR.[29] Then the monopolist solves:

$$R^*(W, Z, Y) = \max_Q R(Q, Y) - C(Q, W, Z)$$
$$= QD(Q, Y) - C(Q, W, Z). \tag{22}$$

Let R_k^* be the derivative of equilibrium revenue with respect to the kth factor price, and W^k be the kth factor price. A comparative statics analysis of (22) implies:

$$H_R = \sum_k R_k^* W^k / R = (R_Q)^2 (R_{QQ} - C_{11})^{-1} / R \leq 0. \tag{23}$$

Thus, the statistic H_R is signed for the elementary monopolist. The intuition of the simple result is straightforward. The H_R statistic gives the percentage change in equilibrium revenues that would follow from a 1 percent increase in all of the firm's factor prices. A 1 percent increase in all factor prices must lead to a 1 percent upward shift in MC. Thus H_R reveals the percentage change in equilibrium revenue that would follow from a 1 percent change in cost. Elementary monopoly theory tells us that a monopolist's optimal revenue will always fall when costs rise: otherwise, the monopolist's quantity was too large before the cost rise.

PR show, in a powerful result, that this finding generalizes to the case of a monopolist that has many choice variables, including both the case where the variables are the outputs of many products and the case where the variables include variables such as quality, advertising, etc.

Even this straightforward implication of monopoly theory has important uses. Suppose we have a sample of "monopolists" that face competition from sellers of other related products. A natural question is whether they are in fact monopolists, or whether the competition from other firms means that they should be treated as in a larger, more competitive market. The PR statistic speaks directly to this question; if they are monopolists, H_R should be negative. Unfortunately, it is not necessarily true that H_R has to be positive if the firms are not monopolists. PR show that in some specific models of oligopoly and of monopolistic competition, H_R must be positive. Thus, it is appropriate to see H_R as a statistic which has some ability to discriminate among alternative competitive hypotheses. There can, however, easily be a false finding of monopoly, since $H_R < 0$ can occur for reasons other than monopoly.

[29] $R(Q, Y)$ is distinct from $R^*(\cdot)$, the reduced-form equation for revenue.

A second economic hypothesis that can be cast as a test on H_R comes when the markets studied are in LR perfectly competitive equilibrium in the strong sense: free entry has driven out inefficient firms, and remaining firms produce at the bottom of their *LRAC* curves. Then $H_R = 1$. Let *MAC* be minimum average cost, and *QMAC* be the quantity which minimizes *AC*. A proportional shift in all factor prices will raise *MAC* by the same proportion without changing *QMAC*. The estimation of H_R proceeds using data on the revenue for single firms. At the firm level, revenue will shift proportionately to cost in LR equilibrium. At the industry level, revenue will expand less than proportionately to cost, as the increase in price will lead to lower quantity demanded: in this LR theory, the supply adjustment comes through entry and exit. Furthermore, the use of single-firm data is warranted, since the only determinants of price and of firm revenue in LR equilibrium are *MAC* and *QMAC*.[30] On the same argument, $R^*(\cdot)$ should not be a function of demand variables in a test of this hypothesis. Thus, the reduced-form revenue equation, estimated on firm data, has two distinct testable restrictions under the hypothesis of LR perfect competition.

The LR flavor of the comparative statics analysis in the PR analysis, both of monopoly and of competition, is reflected in the existing applications of the PR statistic, which are on cross-section data in similar local markets. Panzar and Rosse (1977b) treat the case of newspaper firms in local media markets. An observation is a newspaper, with its revenue as the dependent variable (of course, the majority of revenue is from advertising). If newspapers are monopolies, it is because they do not face intense competition from other media. They are able to reject the hypothesis that newspapers are monopolies even when they are the only newspaper in the market: the interpretation goes to the importance of competition from other media. Shaffer (1982) applies the PR analysis to a cross-section of banking firms in New York State in 1979, finding that the hypothesis of monopoly as well as the hypothesis of LR perfect competition could be rejected. How convincing these studies are depends on two areas: whether estimates on the cross-section of local areas reveals differences in LR equilibrium, and whether all of the variables which shift cost have been identified and correctly entered. The first of these points has been thought through: see Rosse (1970) on the "permanent plant hypothesis". At a minimum, it is clearly important to treat the cases of markets with rising demand separately from those with falling demand. The second point is very similar to one discussed in the previous section.

Recent work by Sullivan (1985) and Ashenfelter and Sullivan (1987) has extended the PR comparative statics in *W* idea to circumstances where variables other than revenue are observable: the results are based in the comparative

[30]Thus, it is appropriate to use only cost shifter exogenous variables in a test of LR perfect competition, and the reduced-form equations are not misspecified under the null if all demand variables are omitted.

statics of market price and quantity in factor prices. As a result of the additional observables, they can treat the oligopoly estimation problem of attempting to draw inferences about conduct. It will be most convenient to write the supply relation for a typical oligopolist in the conjectural variations form:

$$P_t = C_1(Q_t, W_t, Z_t, \Gamma, \varepsilon_{ct}) + D_1(\cdot)Q_i(1 + r_i(\cdot)). \tag{24}$$

In Sullivan's treatment, only market-wide data on price and quantity, and exogenous variables are available. Solving (24) for all firms simultaneously with the demand curve will yield reduced-form equations for P and Q, call these P^* and Q^*. Following PR, estimation of these reduced-form equations could yield H_Q and H_p, the elasticities of Q and of P with respect to marginal cost, possibly measured by a comparative statics exercise involving a proportional increase in all factor prices. Assume that all firms have common marginal cost, and let $\mathscr{L} = (P - C_1)/P$ be the Lerner index. In our notation, Sullivan shows:

$$\sum_i (1 + r_i)^{-1} \geq -H_P/(H_Q\mathscr{L}) \geq -H_P/H_Q, \tag{25}$$

where the last inequality follows because $\mathscr{L}$ must be less than or equal to unity with non-negative MC.

The left-hand inequality in (25) relates one unobservable quantity to another, because only H_P and H_Q are estimated by the technique: no estimates of marginal costs or of the price–cost margin are formed. However, under assumptions that MC is no less than zero, the right-hand inequality of (25) does imply a bound on the competitiveness of conduct. The statistic on the far right can be estimated. The larger is the statistic on the far left, the closer is conduct to competition. Thus, (25) can permit rejection of the hypothesis of successful collusion, though not of competition.

In his empirical work, Sullivan (1985) uses a cross-section (states of the United States) time series (years) on the cigarette industry. The crucial exogenous variable is state taxes, which clearly proxy for MC; all other exogenous variables are captured in an ANOVA procedure. The analysis obtains a slightly tighter bound than (25) by assuming costs are at least as large as taxes (paid by the seller), and is able to reject the hypothesis that cigarette prices are set as if by a cartel. Ashenfelter and Sullivan (1987) take a nonparametric approach to the same data, using year-to-year changes in tax rates and in the endogenous variables in the same state to estimate H_P and H_Q. In thinking about this approach, it is clear that its main potential problems in application are similar to those of PR: Has it been established that the variables which shift MC are not acting as proxies for any other variables? Is the quantitative relationship between these variables and MC certain? The use of tax data is obviously particularly

strong on the second point. I suspect that the first point will usually turn on a detailed argument from the institutional detail of the particular industry at hand, from econometric investigations of robustness, and from ancillary data.

Why is it that the comparative statics in cost can only lead to inequality restrictions on oligopoly conduct, while comparative statics in demand variables provide an estimate of the degree of oligopoly power? [Compare (25) and (18).] The answer follows directly from the nature of the econometric exercise in each case. Consider the two-equation system determining industry price and quantity: there is a demand equation, and a supply relation. The conduct parameters we are particularly interested in are in the supply relation. When demand is shifted by some exogenous variable, it tends to trace out the supply relation, which is after all what we are trying to estimate. The statistic based on the comparative statics in cost could very easily identify the demand equation. They can only cast indirect light on parameters in the supply relation.

3.3. Estimation of marginal cost more directly

The two methods discussed in the previous subsections have in common that they treat the comparative statics of the industry or market equilibrium in isolation. Price and quantity are the only endogenous observables. To the extent that price–cost margins are estimated, the inference is based on the supply behavior of firms. I now turn to an alternative approach, which attempts to econometrically estimate MC from cost data or from factor demand data. This approach uses the methods of cost and factor demand function estimation using flexible functional forms. It relies heavily on the economic theory of cost as dual to production.[31]

The pioneering work in this area was done by Gollop and Roberts (1979) and Appelbaum (1979, 1982). Their approaches start from the total cost function, $C(\cdot)$ [see (2)]. To the observables of the stylized model they add quantity demanded of factors of production: typically broken down only into labor, capital and materials (sometimes energy is separate from other materials inputs). I label the demand for a particular factor of production x_k, that for all factors taken together as X. The key to the approach is to note that MC is the derivative of $C(\cdot)$ with respect to quantity, $C_1(\cdot)$ and that (using standard duality results) the factor demand equations are the derivatives of $C(\cdot)$ with respect to factor prices.[32] Then the approach estimates the demand equation, the supply relation,

[31] Obviously, this approach and the ones described in Subsection 3.1 are complements rather than substitutes.

[32] This follows from Shephard's lemma: see Diewert (1971).

and appends to that system the factor demand equations:

$$P_t = D(Q_t, Y_t, \delta, \varepsilon_{dt}), \tag{26}$$

$$P_t = C_1(Q_{it}, W_{it}, Z_{it}, \Gamma, \varepsilon_{cit}) - D_1(Q_t, Y_t, \delta, \varepsilon_{dt})Q_{it}\theta_{it}, \tag{27}$$

$$X_t = C_W(Q_{it}, W_{it}, Z_{it}, \Gamma, \varepsilon_{cit}). \tag{28}$$

Clearly, appending equations (28) to the system offers at least the possibility of substantial increases in the precision with which MC can be estimated, since there will be cross-equation restrictions between the factor demand equations and the supply relations; the cost parameters Γ appear in both. It is reasonable to expect these restrictions to be quite powerful. Since $C(\cdot)$ is necessarily homogeneous of degree 1 in W, its derivatives $C_W(\cdot)$ will (taken together) depend on all of the parameters of $C(\cdot)$. Thus, all of the parameters in MC also appear in the factor demand equations.

Clearly, the important questions about the utility of this technique in practice turn on the success in estimating MC. Questions of the appropriate functional form for $C(\cdot)$ can probably be addressed by trying several alternatives, or by using prior information about the industry at hand to specify the technology. To the extent that (28) includes a demand equation for capital, users of this approach must face the problem of valuing the capital assets of the firm: capital needs to be decomposed into the price of capital services and their quantity. Thus, the body of criticisms of the SCPP which centered on the accounting treatment of capital will likely reappear as criticisms of the cost function approach. Furthermore, if all factors are treated as SR variable in (28), the price–cost margins will need to be interpreted as price relative to $LRMC$.

Another approach to using factor demand information has recently been introduced by Hall (1986). He starts from the attractive notion that MC could be directly observable by the conceptual experiment of changing quantity produced, holding everything else constant, and measuring by how much expenditures on inputs changed. As the empirical analog of this, Hall works with data on the rates of change of output and of the labor input. One way to think of this is that average incremental cost is revealed by the data: the discrete changes in outputs that occur between sample periods lead to discrete changes in inputs, and the resulting empirical AIC is taken to be the estimate of MC. The second notion in Hall is that changes in the labor input alone can reveal MC. Under the assumption of (LR) constant returns, the wage rate times the change in labor demand divided by labor's share in cost should be AIC. To date, this approach has been largely implemented on aggregate data.[33]

[33] Hall only attempts to estimate $\mathscr{L}$; Shapiro (1987) extends the same logic to estimate θ as well.

This approach has clearly closely related to the previous one, in much the same way that index number approaches to cost and productivity are related to econometric cost and production functions, and therefore shares many of the same advantages and disadvantages. Some of the problems of interpretation have been overcome: the use of the labor demand only helps somewhat with the problem of capital valuation, though labor's share in cost still needs to be calculated. The index-number flavor adds another potential difficulty: if MC is not flat, AIC can be a poor proxy for it. Since the MC curve of interest is $SRMC$, it is unlikely to be flat in applications.

Although I said earlier that the methods described in this section were a complement to, rather than a substitute for, other methods, empirical practice done not yet reflect this. There are two regularities in scholarly practice to note. First, none of the papers cited in this subsection uses industry detail to provide a defense of its maintained hypothesis. Second, all of the papers in the literature can be divided into two classes: those cited in this subsection, which argue identification argument only from the restrictions between MC and factor demands, and all other papers described in this chapter, none of which tried to use factor demands to get better estimates of MC.

3.4. Supply shocks

I argued above that a core implication of modern theories of cartels, as well as an ancient empirical assertion about them, is that their conduct is not constant over time. I would now like to return to Porter's (1983) [Lee and Porter (1984)] switching regressions method for determining this. The question will be: What is it in the data that identifies the inference that cartels break up and reform? I will write (4″) slightly differently. Under the assumption that there are two kinds of conduct, which I think of as "price wars" and "collusion", there are two supply relations that hold in the data:

$$P = C_1(\ ) - D_1(Y_t, \delta)Q\theta_{\mathrm{r}}, \quad \text{during price wars,} \tag{29}$$

$$P = C_1(\ ) - D_1(Y_t, \delta)Q\theta_{\mathrm{c}}, \quad \text{during periods of collusion.} \tag{30}$$

Porter completes the model by specifying the (constant) probability π that (29) holds vs. (30). In price-war periods, prices and quantities are determined by the intersection of (29) and of the demand curve (1), while in periods of successful collusion, these are determined by (30) and (1). The analyst does not know whether there are in fact these two different regimes in the data: that inference is to be drawn from the pattern of prices, quantities and exogenous variables.

A natural question to ask is why this inference can in fact be drawn. I believe that the inference comes from a particular shape of the joint distribution of P

and Q conditional on Z, W and Y. Let us hold all of those exogenous variables fixed at some arbitrary levels. Let the P and Q which solve (29) and (1) be called P_r and Q_r, and those which solve (30) and (1) be called P_c and Q_c. These are random variables, since all of (29), (30), and (1) have econometric error terms. But the two different random variables have two different centers of distribution. In the r-regime, the mean of price will be lower and the mean of quantity will be higher: the regimes differ only in that the r-regime has a lower supply relation than the c-regime. If θ_c is much larger than θ_r, i.e. of collusion is successful at all, we should expect these two means to be far apart.

Now consider the entire distribution of P and Q conditional on the exogenous variables. It has two local modes: one each at (P_r, Q_r) and (P_c, Q_c). Empirical techniques for dealing with bimodal distributions, of which the switching regressions method is a leading example, will be able to detect the presence of the two modes. Thus, the Stigler-esque theories do have an idiosyncratic implication about the shape of the distribution of prices and of quantities. This line of inference departs somewhat from those described above, where the emphasis was on the comparative statics in observable exogenous variables. Here, the variable describing which regime the industry is in was taken to be unobservable. The implications of the theory for the data were drawn by making a simple assumption about that unobservable: that it took on two distinct values. Thus, the nature of the inference here comes from identifying a specific component in the error term: a component that enters the system as a supply shock.

The main potential difficulty with this inference is this: some unobserved shock other than changes in conduct may be moving in the supply equation. Since the inference is based on an error component, there is nothing in the procedure itself to guarantee that the conduct interpretation is the right one. If there are uncaptured changes in factor prices (recall Porter has no factor-price data) or shocks to technology, these could shift the supply relation as well. The size of the effect in the railroad data – the shocks lead to changes in both price and quantity on the order or 50 percent – makes it appear likely that the conduct interpretation is the right one, particularly in light of the extensive contemporaneous discussion of cartel adherence.[34]

3.5. Comparative statics in industry structure

The methods for identification of market power described in the previous four subsections yield estimates of the degree of power over price of a particular

[34] The alternative approach is to decide a priori on the sample split, and then attempt to separately estimate θ_c and θ_r, as in Bresnahan (1987). This has the substantial disadvantage of requiring prior information, but the advantage of being able to more directly assess whether the apparent shocks to supply are due to changes in conduct.

industry in its particular setting. The industry's structure, in the SCPP sense, is a given of the analysis. Substantial time-series changes in industry structure are rare events; thus, the single-industry case study method only rarely, and only on some bodies of data, permits the question of how changes in market structure affect conduct and performance. Methods based on cross-sections of similar markets have also cast some new light on the relationship between industry structure and market power. Unlike earlier mainstream work, which used accounting profit as the dependent variable, many recent studies use price or a price index as the dependent variable. The goal of the investigation is to see how concentration affects prices. Let me briefly outline the work in this area before discussing its interpretation.

The cross-section study of similar markets has been focused on businesses that are geographically local. The dependent variable is price, either a price index or one of the prices of a multi-product firm. The estimating equation is typically a reduced form for price. The industries studied include banking, for which Rhoades (1982) lists dozens of studies, retail food [Cotterill (1986), Lamm (1981)], gasoline suppliers [Marvel (1978)], airline city-pair markets [Graham, Kaplan and Sibley (1983)], cement [Koller and Weiss (1986)] and no doubt others. These studies typically take concentration to be exogenous.[35] The equation they estimate is therefore close to the reduced-form equation for price, departing from it only in that quantity (as transformed into a concentration measure) is included as exogenous. These studies confirm the existence of a relationship between price and concentration, which is at least suggestive of market power increasing with concentration. An interesting variant uses time-series changes in industry structure: see Barton and Sherman (1984) on the effects of a merger in the microfiche film industry on prices and profits.

Most of these studies offer the interpretation that the empirically estimated relationship can be interpreted to cast light on the prediction of almost all theories of oligopoly that higher concentration causes higher price–cost margins by changing conduct. I have seen no careful defense of this interpretation, and I am troubled by it; I offer a series of interpretational difficulties here not because I believe they are true but because they have not yet been rebutted.

If markets are less concentrated when they are larger, and more firms will "fit", then what relationship are we seeing in the data? Take the stark case of free entry as soon as entrants' profits are positive. We interpret the relationship in the data as being about concentration and $P - MC$, yet there is another equation in the model: $P = AC$ for entrants. In the larger markets, firms are also larger, have

[35]An important exception is Graham et al. (1983) which tests for the exogeneity of its concentration measures. Exogeneity is not rejected, even though the coefficients of the concentration measures change substantially when they are treated as endogenous. This suggests that exogeneity cannot be rejected because the test has little power, rather than because the assumption is substantively innocuous.

lower (average) costs, but everywhere firms break even. This reinterpretation is not necessarily a hostile one, but the welfare economics are somewhat different: price and concentration are related in a way that has no obvious bad effects, and does not imply entry barriers.

A somewhat different endogeneity problem arises within industries with the same number of firms. If the firms are selling the same products, then the more concentrated industry likely has more heterogeneity in costs. Greater heterogeneity in costs might interact with conduct in a way that increases prices or it might not.[36]

Even given the exogeneity of concentration, if firms are heterogeneous in their cost functions, markets with more firms allow more statistical "draws" on the lowest-cost firm. One should expect, on average, that the lowest-cost firm out of five has lower costs than the lowest out of three. If the lowest-cost firm is particularly important in the determination of price, as in some competitive models as well as in some oligopoly models, then this purely statistical effect will result in lower P in less concentrated markets. Since estimates that link price to concentration are necessarily on market-wide data, we are in the world of Subsection 2.2.4. The crucial equation is:

$$P = \text{Average}[MC] + D_1(\cdot)Q\,\text{Average}[\theta], \tag{17'}$$

where the notation Average$[\cdot]$ means share-weighted average taken over firms in the market. In general, we do not know whether it is Average$[MC]$ or Average$[\theta]$ or both that is lower in the less concentrated markets. It is the latter interpretation most authors have in mind. Furthermore, there can be links between these two: a firm with a substantial cost advantage may have less competitive conduct than it would facing more equal competition.

That last point can be fleshed out with some observations of the way actual heterogeneity has been sometimes explicitly measured. Consider the Graham et al. (1983) finding that not only concentration affects price in airline city-pair markets, but also that who the competitors are matters. Markets in which one or some competitors are new entrants into the airline business overall have lower prices than other markets, all else equal. It is extremely likely (see Graham et al., section 5) that these entrants have lower MC. It is also possible that their presence changes the conditions of competition. Which is it? The analysis of the paper cannot say. Exactly the same question applies to the interpretation of the industry structure dummies used in Porter (1983) [see (4″) above].

These questions of interpretation are not unanswerable; the previous four subsections discussed explicitly methods of telling MC from θ. The questions are, however, unanswered.

[36]Spiller and Favaro (1984) escape the problem of endogeneity in their time-series study of Uruguayan banking. Their sample period includes a change in the regulation of entry into their industry. They find that freer entry shifts the supply relation downward.

3.6. On the identification of market power

This section has reviewed a large body of method, all developed in the recent past, for empirical investigation of the hypothesis of market power. Several distinct lines of argument have been advanced, each of which relies on a distinct implication of market power for identification.[37] This variety reflects the variety in the data available in different industry studies. In any particular industry, the available information and institutional detail allows different kinds of analysis and different defenses of different analyses. We can therefore expect some continuing variation in desired method. It also reflects the great many implications of the comparative statics of equilibrium in markets with market power which are not found in competitive markets.

4. Market power in product-differentiated industries

Product differentiation raises two kinds of empirical questions, loosely divisible into the SR and LR. In the SR, the stock of products offered by firms and the attributes of these products is fixed. In the SR the measurement question of interest is how much monopoly power firms have because of existing product differentiation. This is (at least) a two-part question. First, as firms' products grow more distinct, each firm's profits will depend less on the policies of other firms. This first part of the question can be adequately answered by investigation of the elasticities of demand, including the cross-elasticities. Second, as products grow more distinct, each firm will respond less to competitive moves by rivals. Understanding this part of the question requires an empirical model of competitive interaction. The measurement of this SR market power has seen tremendous progress in recent times, and that forms the subject of this section.

The market power conceptual issues associated with the SR product differentiation questions are not all that distinct from those in the single-product case. The measurement problems are more severe, however. There are more demand parameters to be estimated: even under the assumption of constant elasticity (or slope) and symmetry, an N-product industry has N own-price elasticities, N income elasticities, and $(N-1)N/2$ cross-price elasticities. There is almost no industry for which the position that there are more than 100 products is untenable: without putting more structure on the problem, the analyst could need to estimate literally thousands of elasticities. On the cost side, the fact that firms produce multiple products suggests the existence of economies of scope: the cost

[37]At least one false identification argument has been proposed, as well. Koutsoyiannis (1982) argues that sales maximization by oligopolists can be empirically distinguished from entry-deterring behavior and from static profit-maximizing behavior. His model assumes monopoly rather than oligopoly: see his equation (26).

function may have some new complexities as well. The use of prior information to guide the specification of the model becomes crucial in such circumstances. Fortunately, in many contexts, prior information will be available from sources like industry trade journals, marketing studies, and so on. As a result of the industry specificity of this prior information, there is considerable variation across industries in the way one would like to proceed with the analysis.[38]

Far and away the most common technique for apparently product-differentiated industries is to assume that the products in the industry are basically fairly close substitutes, use an index of several products' prices as the observable price, and proceed.[39] This procedure is not inherently wrong. It may, however, result in the attribution of market power to noncompetitive conduct when in fact the source of the market power is differentiated products.

When the analyst wishes to study the product-differentiation issues directly, some procedure to reduce the complexity of the analysis from its full size must be employed. There is some experience with three general forms: modelling the product choice part of the demand system, aggregating similar products until there are only a few left in the system, and estimating only a few functions of the parameters of interest.

Tools for the product choice elements of demand have been a major topic of econometric theory and practice in recent times. The work of McFadden (1982) and others on discrete choice has provided a framework for modelling *individuals'* choices of products. These techniques, such as nested logit models, are clearly appropriate in circumstances where there is prior information about groupings of products, such as when industry sources emphasize the existence of distinct product segments within which competition is much more direct than without. A parallel literature, in the theory of "spatial" product differentiation, has concentrated on the relationship between heterogeneity in consumers' tastes and the demand curves facing differentiated oligopolists. It is more appropriate to industries in which there are no clear segment boundaries, i.e. where the fact that products A and B are both important parts of the competitive environment of C need not imply that A is an important part of B's environment. The spatial models thus emphasize the localization of competition as a way to reduce the number of demand parameters, while the discrete choice models emphasize grouping.[40] Both modelling approaches treat product quality similarly. Not

[38]Some other approaches have been attempted as well. Haining (1984) uses the spatial autocorrelation of prices of retail gasoline stations to attempt to infer something about the pattern of interaction among them. I could see not relationship between his statistical hypotheses – "pure competition" is the name of one and "supply and demand" is another – and any economic hypothesis.

[39]See, for example, Gollop and Roberts (1979), Roberts (1983), and Geroski (1983) on roast coffee; Appelbaum (1982) on tobacco and textiles; Sullivan (1985) and Ashenfelter and Sullivan (1987) on cigarettes.

[40]In the limit, models like multinomial logit (without any nests) have the entire industry forming the market segment. Then competition is completely symmetric.

surprisingly, there has been considerable interest in these models in the marketing field; Schmalensee and Thisse (1986) provide an overview of both the relevant economics and marketing literatures.

Empirical examples of this approach can be found in Bresnahan (1981, 1987), which use a spatial model of the demand for automobiles by type as the demand system. The flavor of this approach is that explicit functional form assumptions are made about the distribution of demands across individuals. These distributions, in turn, determine the form of the aggregate demand system. As in econometric work in discrete choice, typical distributional assumptions lead to empirical models with many fewer parameters than the unstructured approach described above. The degree to which such an analysis is convincing turns critically on the quality of the information used to specify the demand system. The best procedure for this is undoubtedly a close reading of the industry trade journals and of typical marketing practice.

Nonetheless, any approach which begins with a highly structured demand system naturally raises questions about the appropriateness of the particular structure. The distinction between localized competition and more systematic or segmented models is particularly important in this regard. Schmalensee (1985) devises test statistics for competitive localization that uses only the measurable movements of endogenous variables, the prices and quantities of particular brands. If exogenous shocks to the system are either market wide (i.e. shift the demand or supply of all products together) or are product-specific, then the extent to which particular products' prices and quantities tend to move together are an indicator of competitive localization. In an application to the ready-to-eat breakfast cereal industry, Schmalensee is able to decisively reject the symmetric model in which all products compete equally. The particular pattern of localization implied by his estimates leads him to doubt the (covariance) restrictions that identify the degree of localization, however.

The approach of aggregating products until there are only a few elasticities to be estimated was taken up by Gelfand and Spiller (1987), Suslow (1986), and Slade (1987). Gelfand and Spiller use data on banking firms competing to make loans of a great many different types. They aggregate the loans until only two types are left, and investigate the demand elasticities in the resulting two-by-two system. An important advance in their work is a model of interrelated oligopoly in the multiple markets, as firms' profits in each market are affected by strategies of other firms in both markets, or even possibly strategies that link the two markets.[41] Presumably such effects can only be studied when the number of markets has been reduced to a reasonably small level. Slade's (1987) treatment of gasoline station "majors" and "independents" is quite similar. The work of

[41] Gelfand and Spiller cast this intermarket interdependence in CV form: each firm has "conjectures" about how other firms will "react" in each of the two markets.

Suslow (1986) aggregates outputs into two: all of those produced by the dominant firm (Alcoa in the aluminum industry before the Second World War) and those produced by the fringe. The dominant firm's *MR* is a function of the degree of substitutability between its product and the product sold by the fringe, as well as by the usual determinants, the market demand elasticity and the fringe supply elasticity.

A third approach to the problem of multiple products has been taken by Baker and Bresnahan (1983, 1985). In their approach, the problem of estimating all of the cross-elasticities of demand is avoided by estimating only the interesting summary statistics of the demand for that product. To estimate the market power associated with a particular product, it is unnecessary to estimate all of the effects of all of the other products' prices in the market. Instead, only the total effect of competition from other products as a brake on the pricing power of the firm owning a particular product is of interest. Consider the seller of product 1, facing the demand system:[42]

$$P_1 = D(Q_1, P_2, \ldots, P_N, Y, \delta), \tag{31}$$

where $P_2, \ldots, P_N$ are the prices of products of competitors and econometric errors are suppressed for convenience.

The measurement problem is that the demand parameter vector δ can be very long, containing the cross-elasticity of demand with each of the $2, \ldots, N$ products. Baker–Bresnahan substitute out $P_2, \ldots, P_N$ in (31) by solving the supply and demand equations for each of them. Suppose that there are $(N-1)$ more equations like (31), one for each of the other products. Also, there is a supply relation for each of products $2, \ldots, N$:

$$P_i = C_1(Q_i, W_i, Z_i, \Gamma_i) - D_1(Q, Y, \delta)Q_i\theta_i, \quad i = 2, \ldots, N, \tag{4^{d}}$$

where Q is the vector of all firms' (and equivalently, all products') quantities. The $2*(N-1)$ equations (31), (4^{d}) can be solved for the prices and quantities of products $2, \ldots, N$. Call the solution for P:

$$P_i = P_i^*(Q_1, W_N, Z_N, \Gamma_N, \theta_N, Y, \delta),$$

where the dependence on Q_1 arises because *only* products $2, \ldots, N$ have been solved out, and the subscript I refers to the superset of all the subscripts i. The equation to be estimated is:

$$\begin{aligned} P_1 &= D(Q_1, P_2^*(\cdot), \ldots, P_N^*(\cdot), Y, \delta) \\ &= D^{\mathrm{R}}(Q_1, W_n, Z_N, \Gamma_N, \theta_N, Y, \delta), \end{aligned} \tag{32}$$

[42]This is slightly unfamiliar notation. It would be more familiar to write quantity for this product as a function of the prices of all products: the function presented is simply the inverse of that.

the residual demand curve for product 1. There are two immediate observations here. First, an enormous amount of information has been lost here by substituting out the prices of all the other products; it will be impossible to estimate all of the separate elements of δ from (32), much less all of the other parameters in it. But this is of little importance. The elasticity of D^{R} with respect to Q_1 tells us how much power the firm has over product 1's price, taking into account the adjustment of all other firms' prices and quantities.

A somewhat similar example may clarify the technique. Suppose that (4^{d}) takes the form $P_i = MC_i(W_i)$ for all the other firms: they are price-takers. Equation (32) then predicts the price of product 1 as a function of its quantity, and variables shifting the costs of all other products. If firm 1 has no market power in this product, the prices (and in this example, the costs) of other products will determine its price. In the no-market-power case, the elasticity of D^{R} with respect to Q_1 will be zero.

Thus, the Baker–Bresnahan approach estimates the demand elasticity facing the single firm or product, taking into account the competitive reaction of all other firms in the market. This demand elasticity summarizes the market power of the firm: knowing it is insufficient to determine the sources of that market power.[43]

The relationship between the Gelfand–Spiller or Suslow approach and the Baker–Bresnahan approach is this: in the first approach, all of the elasticities of demand, supply, and competitive interaction are estimated. From them the market power of any particular firm could be calculated. Of course, for practical implementation this approach requires specifying a relatively small number of different products. The second approach works when there are a large number of products, but does not yield estimates of all of the elasticities, only of the summary statistics relating to each firm's market power.

4.1. "Market definition", policy analysis, and product differentiation

Some of the techniques for assessing market power have been applied to the problem of "market definition" in antitrust analysis. In antitrust applications, it is frequently of some importance to determine whether a group of firms *would* have any market power if they chose to act in concert, or in other contexts whether a single firm in fact has any market power.[44] The latter question can be

[44]These questions are well-posed, even where the usual method of answering them, defining a "relevant market" and calculating market shares in it, is senseless.

[43]All of this presumes that the residual demand curve can be estimated. The condition for that is that firm 1's costs have moved independently of all other firms' costs. An obvious application of the technique, therefore, is in the international context. One would ask how steep the demand curve facing producers in a single country was; the natural experiment for estimating that quantity would be good if, for example, exchange rate movements had moved real relative costs in different countries.

directly answered by the Baker–Bresnahan technique. Scheffman and Spiller (1987) extend the Baker–Bresnahan technique to estimate the elasticity of demand facing a group of firms, thereby providing an answer to the former question. Baker and Bresnahan (1985) ask how much steeper the demand curve facing two firms would be post-merger than the pre-merger level. If two firms sell products that are very close substitutes, then each likely provides an important part of the competitive brake on the other, unless there are several other firms providing similar products as well. The increase in the steepness of the residual demand curve measures this effect.

Methods based on the Panzar–Rosse statistic have also been used in this connection.[45] Panzar and Rosse (1987) give a new interpretation to their monopoly test which is directly relevant here. Suppose a particular firm (firms) has been studied by PR methods. A rejection of "monopoly" for this firm (group of firms) implies that it (they) cannot be treated as acting in isolation. Other firms must be interacting with the firm (firms) at hand.

In a slightly different context, Schmalensee and Golub (1983) examine the spatial product differentiation of firms in the electricity market. They use models of the demand for electricity, the costs of transmission, and of competition to assess the likely impact of deregulation.

4.2. Product differentiation in the LR

In the LR, firms can add products, change their attributes (either physically or in consumers' perceptions) or new firms can enter with new products or imitations. This is a very complex area, full of hypotheses. Strategic interaction effects of many kinds are possible: preemption by establishment of a reputation for product superiority, preemption by filling out the product space, coordination of investment in distinct product types so as to reduce competition, and so on. Essentially nothing empirical is known about any of these hypotheses. Furthermore, the welfare implications of SR market power in a product differentiated industry are not transparent. In the Chamberlinian tangency of monopolistic competition, every firm has market power in the sense of this section. Yet that does not establish that there is any inefficiency, once the need to cover the fixed costs of product design are taken into account. The crucial issue here is an adequate empirical treatment of the supply curve of new firms and of new, different products. Empirical work on this area is likely to be forthcoming soon, but little exists now.[46]

[45]See Shaffer (1983) and Slade (1983).

[46]Many of the relevant analytical issues to support empirical work can be found in Panzar and Rosse (1981). Bresnahan and Reiss (1986) estimate an equation for the entry of a second product differentiated firm into a monopoly.

5. What has been learned about market power?

However useful its methodological contributions, the industry case-study nature of much work in the NEIO has raised questions of interpretation. How general are the results? What do these studies, taken together, reveal about market power in the economy as a whole? What have we learned about the conditions under which market power tends to arise? What is known about the easily measured correlates of market power, such as concentration? In short, a single industry case study cannot paint a broad picture; it can only reveal the nature of industry conduct and performance in the industry studied. The original idea of the SCPP was that empirical research would estimate a function mapping structural characteristics into measures of conduct (where that is possible) and of performance (more commonly). Empirical knowledge of this map is obviously valuable. It contains information about the sources as well as the location of market power. It could be used to guide policy in those areas, such as merger policy, where it could influence structure rather than conduct directly. An industry case study, whether done by the methods of the 1930s or the 1980s, can hope to reveal at most one point on the function. The integration of different case studies to give a unified picture of the whole map is an obviously attractive prospect. It can only be partially carried out now, even though the empirical papers described in this chapter have treated well over a dozen industries. In Table 17.1 I reproduce the estimated price cost margins ($\mathscr{L}$) from several different NEIO studies reviewed

Table 17.1
Summary of existing empirical work

Author	Industry	$\mathscr{L}$
Lopez (1984)	Food processing	0.504
Roberts (1984)	Coffee roasting	0.055/0.025[a]
Appelbaum (1982)	Rubber	0.049[c]
Appelbaum (1982)	Textile	0.072[c]
Appelbaum (1982)	Electrical machinery	0.198[c]
Appelbaum (1982)	Tobacco	0.648[c]
Porter (1983)	Railroads	0.40[b]
Slade (1987)	Retail gasoline	0.10
Bresnahan (1981)	Automobiles (1970s)	0.1/0.34[d]
Suslow (1986)	Aluminum (interwar)	0.59
Spiller–Favaro (1984)	Banks "before"[e]	0.88/0.21[f]
Spiller–Favaro (1984)	Banks "after"[e]	0.40/0.16[f]

[a] Largest and second largest firm, respectively.
[b] When cartel was succeeding: 0 in reversionary periods.
[c] At sample midpoint.
[d] Varies by type of car; larger in standard, luxury segment.
[e] Uruguayan banks before and after entry deregulation.
[f] Large firms/small firms (see their table 2).

here.[47] Different scholars will undoubtedly differ on the extent to which it offers answers to the questions of the last paragraph. A few preliminary conclusions, however, are available. These are cast in somewhat guarded language primarily because of the limited coverage of the available studies.

Conclusion A

There is a great deal of market power, in the sense of price-cost margins, in some concentrated industries.

The conclusion seems almost forced by the last column of Table 17.1. Several studies have found substantial power over price. Available data do not permit a systematic assignment of concentration indexes to the industries listed in Table 17.1, since they are not drawn from the economic censuses; several are based on the primary data gathering of different scholars. A glance down the list of industries, however, is sufficient to demonstrate that they are overwhelmingly drawn from the highly concentrated end of the industrial spectrum.

Finding A, I think, cannot be controversial, particularly with its qualification that it refers to "some" concentrated industries. The finding would be less controversial without the "some", and I think this is right. There are at least two reasons to suspect the generality of the findings in the papers reviewed in this chapter. First, authors who invent methods for the detection of market power are likely to first apply them in industries where they expect to find it. Thus, the existing studies have largely treated quite concentrated industries, industries where there were known or suspected cartels, industries where a solid old-style case study suggested anticompetitive conduct, and so on. The field is now ripe for revisionism! Or at least for continued expansion of the set of industries in which conduct and performance are well measured.

Second, the list of industries studied to date is special in another sense. Since the data are often drawn from trade journals, regulatory bodies, court proceedings or similar sources, the industries covered may be unrepresentative. Consider the industries with excellent trade journals; they are ones in which information about what competitors are doing is quite good. Thus the repeated finding of successful collusive arrangements might reflect the particular information structure of these industries. Similarly, Suslow's ability to mine the trial transcript in

[47]A few papers were left out of the table because their estimates were not given in such a form as to permit calculation of the Lerner index. (These papers heavily emphasize conduct over performance, of course.) Panzar–Rosse methods and Sullivan methods are excluded because they do not provide an estimate of the Lerner index. Instead, they provide an estimate of conduct parameters or a test of certain hypotheses about conduct. (Comments on a draft of this chapter suggest that this is not well understood: see Subsection 3.2, above.) Baker–Bresnahan methods are excluded because the demand elasticity estimates they provide correspond exactly to the Lerner index only in certain circumstances. (See Section 4.)

Alcoa for data arose *because* there was reason to suspect that firm had substantial market power. More new data sources are needed!

Conclusion B

One significant cause of high price-cost margins is anticompetitive conduct.

The studies under review distinguish between conduct, in the sense of firms' behavioral rules for price-fixing, and performance, in the price–cost margin sense. It is not the case that, systematically, we see tiny departures of conduct from price-taking plus very steep demand curves leading to large departures of performance from the competitive standard. Instead, some of the studies appear to be finding conduct well toward the collusive end of the spectrum. For example, Porter (1983) and Bresnahan (1987) both find explicitly collusive behavior. I should emphasize that conduct is not uniform. Roberts (1984), for example, finds that most of the firms in his industry should be classified in an (essentially) price-taking fringe. The largest firms have much less competitive conduct, but have not succeeded in raising prices to the profit maximizing level given the fringe's behavior. The variety of conduct across industries, as well as the variety of performance, suggest the importance of the continued study of market power as a phenomenon.

Conclusion C

Only a very little has been learned from the new methods about the relationship between market power and industrial structure.

There are two points here, one implicit in the discussion below Conclusion A. Table 17.1 is drawn mostly from the highly concentrated end of the industry spectrum. We therefore have new information about the map from structure to conduct and performance over only a very limited range. The second point is that the causes of market power have not been addressed by very many of these studies. One presumes, for example, that long-surviving market power is an indicator of some failure of the entry process to discipline conduct. Yet entry has hardly been discussed in the papers. That leads, naturally enough, to my final section.

6. The future: The sources of market power

This should properly be a short section: although the NEIO has had a great deal to say about measuring market power, it has had very little, as yet, to say about

the causes of market power. In particular, the topics of entry, predation, entry deterrence, and strategic competition in the LR generally have not yet been extensively taken up in empirical work with explicit theoretical foundations. These topics, then, remain primarily for the future. There are a few scattered contributions.

6.1. Predation

Burns (1986) casts considerable light on the possibility and profitability of predatory pricing in a study of the tobacco trust. He finds that acquisitions of competitors made by the trust became cheaper after predatory incidents. Since predation has long been believed not to be an equilibrium phenomenon, this is a useful and important contribution. It relies on nineteenth-century data, perhaps the unique data available for its method. The tobacco trust often thought it was predating; in that era before antitrust law prevented such self-revelation, the trust left a solid documentary record of when the predatory incidents took place. Investigations of the trust's behavior by the Federal government and the courts yielded a rich data source for Burns' study. More circumspect modern firms will be less well documented. We have not yet taken the methodological step of discovering the empirical implications of predatory conduct when acts of predation must be inferred, not observed.

6.2. Entry

The problem of entry has received some useful methodological contributions have been made in working papers. Panzar and Rosse (1977b) work out an empirical model of LR entry, treating the number of firms in the industry as a continuous variable. They argue that the comparative statics of monopolistically competitive industry equilibrium have distinctive, testable implications. The key to their argument is the addition of a third set of equations to (1), demand, and (4), pricing. The new equations take the form:

$$P = LRAC(W, Y) \tag{33}$$

for the marginal firm. The application of this model, or similar ones, in monopolistic competition contexts is obviously an attractive prospect.

Bresnahan and Reiss (1986b) take up the problem of econometric models of entry with an integer number of firms. They provide an empirical application to monopolies and duopolies on a sample of automobile dealers in small, isolated towns. They model the entry decision of each of the first two firms into a market,

and exploit the comparative statics of the level of firm profits in the size of the market to draw inferences about monopoly and duopoly conduct and entry behavior. Like Panzar and Rosse, they argue strongly that the right kind of sample for the study of entry is a cross-section of closely related markets. The particular characteristics of those markets in which more firms have entered will reveal the determinants of firm profitability.

Clearly much more work is needed on the determinants of, and the effects of, entry. In Subsection 3.4, above, I discussed the existing literature on the effects of entry and concentration on price in cross-sections of related markets. A careful working out of the analytics of the number of firms in an industry, their sizes, and so on, is a crucial step in the successful analysis of the effects of entry. This is one area in which we do not lack for data; many studies treating entry as exogenous have already been carried out.

6.3. Final remark

In stating the need for further study so strongly, I do not mean to suggest that the accomplishments to date are small. By departing from the tradition of treating performance as observable in accounting cost data, the *NEIO* has provided a new form of evidence that there is substantial market power in the economy, a form of evidence that is not susceptible to the standard criticisms of earlier approaches. Furthermore, the individual studies of particular industries are specific and detailed enough that alternative explanations of the findings can be rebutted. The current state of affairs is quite encouraging: we know that there is market power out there, and need to know a lot more about exactly where. We know essentially nothing about the causes, or even the systematic predictors, of market power, but have come a long way in working out how to measure them.

References

Abreu, D., Pearce, D. and Stacchetti, E. (1986) 'Optimal cartel equilibria with imperfect monitoring', *Journal of Economic Theory*, 39:251–269.

Anderson, J. (1984) 'Identification of interactive behavior in air service markets: 1973–76', *Journal of Industrial Economics*, 32:489–507.

Anderson, J. and Kraus, M. (1984) 'An econometric model of regulated airline flight rivalry', *Research in the Economics of Transportation*, 2.

Appelbaum, E. (1979) 'Testing price taking behavior', *Journal of Econometrics*, 9:283–294.

Appelbaum, E. (1982) 'The estimation of the degree of oligopoly power', *Journal of Econometrics*, 19:287–299.

Ashenfelter, O. and Sullivan, D. (1987) 'Nonparametric tests of market structure: An application to the cigarette industry', *Journal of Industrial Economics*, forthcoming.

Bain, J.S. (1951) 'Relation of profit rate to industry concentration: American manufacturing, 1936–40', *Quarterly Journal of Economics*, 65:293–324.

Baker, J. and Bresnahan, T. (1983) 'Estimating the elasticity of demand facing a single firm: Evidence for three brewing firms', mimeo.
Baker, J. and Bresnahan, T. (1985) 'The gains from merger and collusion in product differentiated industries', *Journal of Industrial Economics*, 33:427–444.
Barton, D. and Sherman, R. (1984) 'The price and profit effects of horizontal merger: A case study', *Journal of Industrial Economics*, 33:165–177.
Bresnahan, T.F. (1980) 'Three essays in the American automobile oligopoly', dissertation, Princeton University.
Bresnahan, T.F. (1981) 'Departures from marginal-cost pricing in the American automobile industry: Estimates for 1977–1978', *Journal of Econometrics*, 11:201–227.
Bresnahan, T.F. (1982) 'The oligopoly solution concept is identified', *Economics Letters*, 10:87–92.
Bresnahan, T.F. (1987) 'Competition and collusion in the American automobile oligopoly: The 1955 price war', *Journal of Industrial Economics*, forthcoming.
Bresnahan, T.F. and Reiss, P. (1985) 'Dealer and manufacturer margins', *Rand Journal of Economics*, 16:253–268.
Bresnahan, T.F. and Reiss, P. (1986) 'Entry in monopoly markets', mimeo.
Bulow, J. and Pfleiderer, P. (1983) 'A note on the effect of cost changes on prices', *Journal of Political Economy*, 91:182–185.
Burns, M. (1986) 'Predatory pricing and acquisition cost of competitors', *Journal of Political Economy*, 94:286–296.
Chamberlin, E.H. (1933) *The theory of monopolistic competition*. Mass.: Harvard University Press.
Cotterill (1986) 'Market power in the retail food industry: Evidence from Vermont', *Review of Economics and Statistics*, 68:379–386.
Cowling, K. and Waterson, M. (1976) 'Price cost margins and market structure', *Economica*, 43:267–274.
Demsetz, H. (1974) 'Two systems of belief about monopoly', in: H.J. Goldschmid, H.M. Mann and J.F. Weston, eds., *Industrial concentration: The new learning*. Mass.: Little, Brown.
Diewert, E. (1971) 'Application of Shephard's duality theorem', *Journal of Political Economy*, 79:481–507.
Gelfand, M.J. and Spiller, P. (1987) 'Entry barriers and multi-product oligopoly', *International Journal of Industrial Organization*, forthcoming.
Geroski, P. (1983) 'The empirical analysis of conjectural variations in oligopoly', mimeo.
Geroski, P., Phlips, L. and Ulph, A. (1985) 'Oligopoly, competition and welfare: Some recent developments', *Journal of Industrial Economics*, 33:369–386.
Gilbert, R. (1984) 'Bank market structure and competition – a survey', *Journal of Money, Credit and Banking*, 16:617–645.
Gollop, F. and Roberts, M. (1979) 'Firm interdependence in oligopolistic markets', *Journal of Econometrics*, 10:313–331.
Graham, D.R., Kaplan, D.P. and Sibley, D.S. (1983) 'Efficiency and competition in the airline industry', *Bell Journal of Economics*, 14:118–138.
Green, E.J. and Porter, R.H. (1984) 'Noncooperative collusion under imperfect price information', *Econometrica*, 52:87–100.
Haining, R. (1984) 'Testing a spatial interacting-markets hypothesis', *Review of Economics and Statistics*, 66:576–583.
Hall, R. (1986) 'The relationship between price and cost in U.S. industry', mimeo.
Iwata, G. (1974) 'Measurement of conjectural variations in oligopoly', *Econometrica*, 42:947–966.
Just, R. and Chern, W. (1980) 'Tomatoes, technology and oligopsony', *Bell Journal of Economics*, 11:584–602.
Koller, R. and Weiss, L. (1986) 'Price levels and seller concentration: The case of Portland Cement', mimeo.
Koutsoyiannis, A. (1982) 'Goals of oligopolistic firms: An empirical test of competing hypotheses', *Southern Economic Journal*, 51:540–567.
Lamm, R. (1981) 'Prices and concentration in the food retailing industry', *Journal of Industrial Economics*, 30:67–78.
Lau, L. (1982) 'On identifying the degree of competitiveness from industry price and output data', *Economics Letters*, 10:93–99.

Lee, L.F. and Porter, R.H. (1984) 'Switching regression models with imperfect sample separation information – with an application on cartel stability', *Econometrica*, 52:391–418.
Lerner, A. (1934) 'The concept of monopoly and the measurement of monopoly power', *Review of Economic Studies*, 11:157–175.
Lopez, R.E. (1984) 'Measuring oligopoly power and production responses of the Canadian food processing industry', *Journal of Agricultural Economics*, 35:219–230.
Marvel, H. (1978) 'Competition and price levels in the retail gasoline market', *Review of Economics and Statistics*, 60:252–258.
McFadden, D. (1982) 'Econometric models of probabilistic choice', in: C. Manski and D. McFadden, eds., *Structural analysis of discrete data with econometric application*. Cambridge, Mass.: MIT.
Panzar, J. and Rosse, J. (1977a) 'Structure, conduct and comparative statics', mimeo.
Panzar, J. and Rosse, J. (1977b) 'Chamberlin vs. Robinson: An empirical test for monopoly rents', mimeo.
Panzar, J. and Rosse, J. (1987) 'Testing for 'monopoly' equilibrium', *Journal of Industrial Economics*, forthcoming.
Porter, R.H. (1983) 'A study of cartel stability: The joint executive committee, 1880–1886', *Bell Journal of Economics*, 14:301–314.
Porter, R.H. (1984) 'Optimal cartel trigger price strategies', 29:313–338.
Porter, R.H. (1985) 'On the incidence and duration of price wars', *Journal of Industrial Economics*, 33:415–426.
Prescott, E.C. and Visscher, M. (1977) 'Sequential location among firms with foresight', *Bell Journal of Economics*, 8:378–393.
Roberts, J. (1984) 'Testing oligopolistic behavior', *International Journal of Industrial Organization*, 2:367–383.
Roberts, M. (1983) 'Testing oligopolistic behavior: An application of the variable profit function', *International Journal of Industrial Organization*, 2:367–383.
Rhoades, S. (1982) 'Structure–performance studies in banking: An updated summary and evaluation', FRB staff studies no. 119, see also Federal Reserve Bulletin.
Rosse, J.N. (1970) 'Estimating cost function parameters without using cost data illustrated methodology', *Econometrica*, 38:256–275.
Rotemberg, J. and Saloner, G. (1986) 'A supergame-theoretic model of business cycles and price wars during booms', *American Economic Review*, 76:390–407.
Scheffman, D. and Spiller, P. (1987) 'Geographic market definition under the DOJ guidelines', *Journal of Law and Economics*, forthcoming.
Scherer, F.M. (1980) *Industrial market structure and economic performance*, 2nd edn. Chicago: Rand McNally.
Schmalensee, R. (1985) 'Econometric diagnosis of competitive localization', *International Journal of Industrial Organization*, 3:57–70.
Schmalensee, R. and Golub, B. (1983) 'Estimating effective concentration in deregulated wholesale electricity markets', *Rand Journal of Economics*, 5:12–26.
Schmalensee, R. and Thisse, J.F. (1986) 'Perceptual maps and the optimal location of new products', Center of Operations Research and Econometrics, core reprint no. 8620.
Shaffer, S. (1982) 'A nonstructural test for competition in financial markets', in: *Proceedings of a conference in bank structure and competition*. Chicago: Federal Reserve Bank of Chicago.
Shapiro, M. (1987) 'How (monopolistically) competitive is U.S. industry', mimeo.
Slade, M. (1987) 'Conjectures, firm characteristics and market structure: An analysis of Vancouver's gasoline-price wars', mimeo.
Spiller, P. and Favaro, E. (1984) 'The effects of entry regulation on oligopolistic interaction: The Uruguayan banking sector', *Rand Journal of Economics*, 15:244–254.
Stigler, G.J. (1964) 'A theory of oligopoly', *Journal of Political Economy*, 72:44–61.
Sullivan, D. (1985) 'Testing hypotheses about firm behavior in the cigarette industry', *Journal of Political Economy*, 93:586–598.
Sumner, D. (1981) 'Measurement of monopoly behavior: An application to the cigarette industry', *Journal of Political Economy*, 89:1010–1019.
Suslow, V. (1986) 'Estimating monopoly behavior with competitive recycling: An application to Alcoa', *Rand Journal of Economics*, 17:389–403.