GAME-THEORETIC MODELS OF MARKET STRUCTURE*

by

John Sutton
London School of Economics and Political Science

The Toyota Centre
Suntory and Toyota International Centres for
Economics and Related Disciplines
London School of Economics and Political Science
Houghton Street
London WC2A 2AE
Tel.: 020-7955 6674

Discussion Paper
No. EI/15
October 1996

* This paper was prepared for the World Congress of the Econometric Society, Tokyo, 1995. I would like to thank Alison Hole for her helpful comments on an earlier draft. My thanks are also due to my discussant, Jacques Cremer, for his very helpful suggestions. The financial support of the Economic and Social Research Council and the Leverhulme Trust is gratefully acknowledged.

Abstract

This paper reviews the recent literature on game-theoretic models of market structure and their empirical implementation.

Keywords: Game-theoretic models, market structure, industrial organization, framework, rationalise, equilibrium outcomes, theory.

Contact address: Professor John Sutton, STICERD, London School of Economics and Political Science, Houghton Street, London WC2A 2AE. email: j.sutton@lse.ac.uk

1. Game Theory and Industrial Organization: an ABC.

It has become a familiar observation in recent years that the literature on game-theoretic models in Industrial Organization faces a serious dilemma. The richness and flexibility of this class of models provide a framework within which we can 'rationalise' a huge range of possible 'equilibrium outcomes'. Whatever the phenomenon, we seem to have a model for it. Should we see this as a success, or as an embarrassment? Does this body of theory allow *any* outcome to be rationalised? After all, the content of a theory lies in the set of outcomes which it excludes. Judged on these terms, is the enterprise empty[2]?

The huge range of outcomes that can be rationalized can be traced to two features of these models. First, many of the models in this literature have multiple equilibria. Second, the

---

[2]For comments on this dilemma, see Shaked and Sutton (1987), Fisher (1989), Sutton (1990) and Pelzman (1991). The dilemma is not special to Industrial Organization, although much recent comment suggests that it is. The dilemma is in fact as old as economics: Edgeworth called it the 'problem of indeterminacy'. The issue is whether the operation of the market mechanism pins things down so tightly that we can model its operation using a set of assumptions which lead to a unique equilibrium outcome. For a discussion of the Edgeworth-Marshall debate and later developments, see Sutton (1993).

appropriate specification of the model is rarely obvious. (Is competition to be à la Bertrand, or à la Cournot? Should entry be modelled as simultaneous, or sequential?) Sometimes it is possible, by referring to the observable features of some particular market, to decide in favour of one model specification over another. In other cases, however, the features that distinguish candidate models must be treated as unobservables, at least from the point of view of the modeller. Both these features tend to widen the set of outcomes that may be rationalized as equilibria (Figure 1).

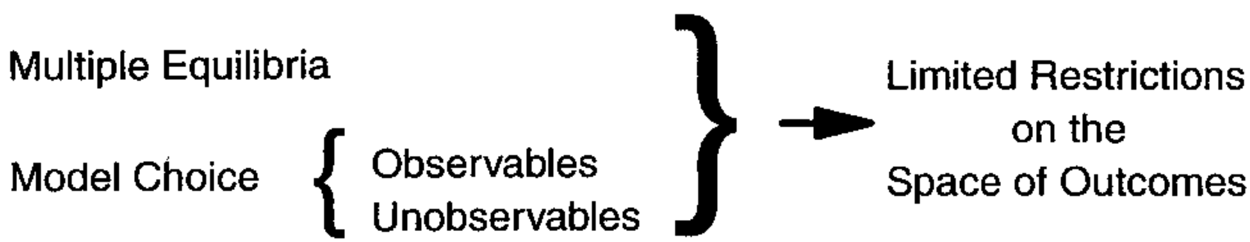

Figure 1. The Dilemma: Game theoretic models are flexible, but do they have content?

These observations lead to two conclusions,

(a) It will not usually be possible, by reference to observable market characteristics, to specify a unique outcome as 'the equilibrium outcome'. We may have to be content with placing limited restrictions on the space of outcomes, partitioning outcomes into those that can be supported as equilibria of some admissible model, and those that can not.

(b) Insofar as part of the problem arises as a result of the influence of observable characteristics that vary across industries, it follows that the range of candidate models may be narrowed by restricting attention to one industry, or set of cognate industries. In other words, a trade-off may appear between the breadth of application of a theory, and the tightness of the restrictions that it places upon the set of outcomes.

The trade-off between breadth of application and tightness of restrictions motivates the currently popular literature on 'single industry studies'. Here the aim is to analyse a specific market, relying on special (institutional or other) features of that market to motivate assumptions. In Figure 2, A is for Auctions. Here, the strategy of focusing on a highly specific context comes into its own. The institution of the auction specifies explicitly the rules of the game. We know the actions available to players, and specifying the strategy space poses few problems. Here, we avoid almost all the unpleasant arbitrariness in specifying the game: the rules of the auction (almost) specify the game itself. Moreover, in some settings the model gives precise, non-trivial, and pleasing predictions. At its most impressive, the theory delivers convincing explanations of patterns in the data that would be hard to account for in terms of some alternative model (see, for example, Hendricks and Porter (1988))[3].

---

[3]Unfortunately, this is not always the case. In many settings, the outcome is driven by unobservable distributions of buyers' valuations, and the theory does not constrain the data to any useful degree. For a recent review of these issues, see Laffont (1996).

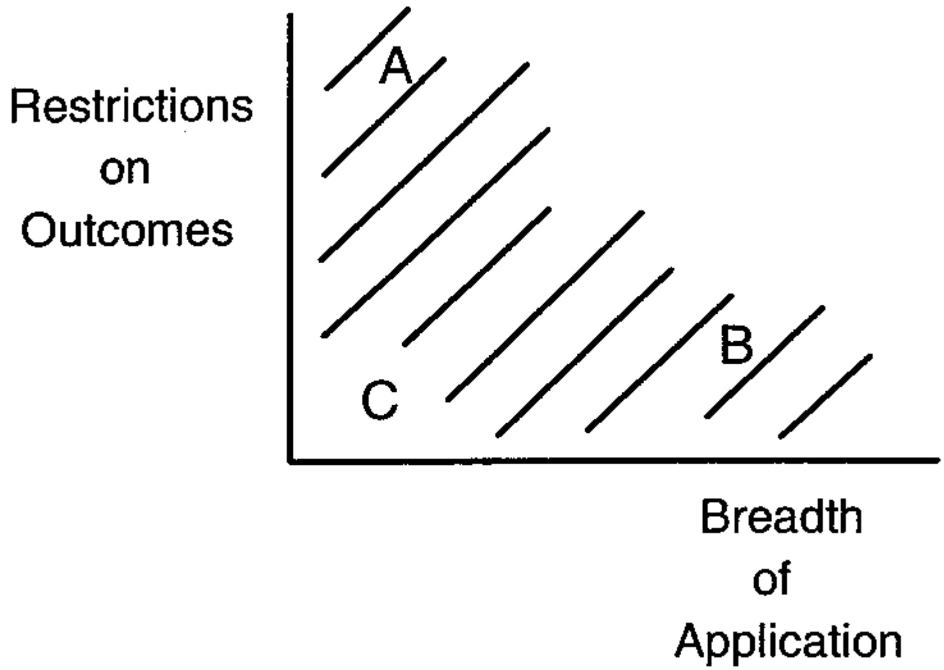

Figure 2. A Trade-off.

But a narrowing of the domain does not always lead to such happy results. The classic problem area in this respect is that of dynamic oligopoly models and in particular the analysis of cartel stability. Here we are at point C in Figure 2. Many quite different cartel models are available. Case studies of cartels show that different cartels do indeed behave in quite different ways[4]. but even if we narrow the domain to a specific cartel over a specific period, we still fall short of any precision of predictions. The best we can hope for here is a 'model selection' exercise. This problem arises more generally throughout the whole area of 'dynamic oligopoly models'. (See for example, Gasmi, Laffont and Vuong (1990)).

The opposite end of the trade-off arises in the Bounds approach to market structure (Sutton (1991)). Here, at point B in the figure, the idea is to turn away from the now-dominant emphasis on single industry studies and to return to the traditional emphasis on mechanisms that appear to be relevant across the general run of industries. The price we pay for widening the domain of application is that the set of candidate models that we must admit is now wider, and the constraints on outcomes that hold good for all these candidate models are correspondingly weaker. The aim is not to identify some unique 'equilibrium outcome' but rather to place some bounds on the set of outcomes that can be supported as equilibria.

---

[4]Contrast, for example the (various) mechanisms that have been considered in discussing the JEC cartel (Porter (1983), Ellison (1994)) with the quite different type of story appropriate to the Bromine cartel (Levenstein (1993)).

This paper looks at one area of the recent I.O. literature, the part concerned with 'Explaining Market Structure', from the perspective displayed in Figure 2. No attempt is made to be comprehensive; rather, the aim is to discuss some examples of current research from this perspective. With that in mind, we begin with the quest for 'general' properties (point B), before turning to studies which narrow the domain in order to sharpen the constraints on outcomes (B → A). In the final section, we look at the inherent limits of this kind of approach. Here, part of the problem lies in the fact that, however narrow the scope of inquiry, the presence of multiple equilibria and the problem of unobservables place serious limits on the extent to which we can impose constraints on the space of outcomes (point C).

2. Strong Mechanisms I: Price Competition and the Market Size-Market Structure Relationship

Much of the recent I.O. literature on market structure has been formulated within the framework of multi-stage games. Over a series of stages, firms make choices that involve the expenditure of fixed and sunk costs, whether by entering a market by candstructing a plant, by introducing new products or building additional plant capacity, or by carrying out advertising or R&D. In a final stage subgame, all the results of such prior actions are summarized in terms of some 'space of outcomes', i.e. the final configuration of plants and/or products that emerges at the penultimate stage of the game. A description of this 'outcome' enters as a set of parameters in the payoff function of the final stage ('price competition') subgame, and so the 'outcome' of the entry process, together with a specification of the nature of price competition, determines the vector of final stage profits, and of market shares.

In order to circumvent the problems posed by multiple equilibria, and by the role of unobservables, it is of interest to develop propositions that hold good across some suitably defined *class* of models. This class should encompass a range of models between which we cannot hope to distinguish empirically. We might, for example, want to look at propositions that hold good independently of the nature of price competition (Bertrand, Cournot), the entry process (simultaneous, sequential) and so on. It is possible to identify several 'mechanisms'[5] that operate in a fairly robust way across

---

[5]The word 'mechanism' is used loosely here; it is possible to formalize this notion, and the related idea of a 'Natural Experiment', but to do so requires that we first redefine equilibrium directly on the space of 'outcomes'; see Sutton (1997)).

a wide class of models, and in what follows, we focus attention on these 'strong' mechanisms.

The most elementary mechanism examined in the recent literature rests upon the assumption that equilibrium price falls (strictly, is non-increasing) with entry. This assumption is well founded, both theoretically and empirically. At the theoretical level, it holds good in a wide class of standard models, including elementary one-shot oligopoly models and various product differentiation models. In the dynamic oligopoly (cartel stability) literature, where multiple equilibria are the norm, some corresponding statements can be made regarding the maximum price, or profit per firm, that can be sustained at equilibrium. Such a robust result invites attempts to construct counterexamples, and these are indeed available. It is a measure of the robustness of the result that such examples are rather contrived, involving for example a carefully constructed distribution of consumer tastes over the space of product characteristics (Rosenthal (1980)). At the empirical level, too, direct evidence on entry and price is strongly supportive of the assumption; the most important body of evidence is the volume edited by Len Weiss (1989).

What concerns us here are the implications of this assumption for equilibrium market structure. These implications are non-trivial, and they throw some interesting light on certain arguments regarding competition policy[6]. The most important implications relate to the relationship between the size of a market and equilibrium market structure. In describing these implications, we confine attention, in this section, to the class of 'symmetric' product differentiation models. These models share the property that each product variety enters into the consumers' utility function(s) in the same way, so that the firms' profits depend only upon the number of product varieties offered by each firm[7].

We confine attention in this section to models in which the cost of entering the market, or of introducing a new product, is fixed exogenously. (There is no advertising, or R&D.) Consider, first, a setting in which N firms enter, each with one product, at an entry cost of $\epsilon > 0$[8]. Symmetry ensures that all prices are equal

---

[6]They suggest, for example, that attempts to reduce market concentration in order to increase the intensity of price competition may be ineffectual.

[7]These models include the 'linear demand model' (Shubik and Levitan (1980)) and the model of Dixit and Stiglitz (1977). They exclude 'Hotelling-type' location models.

[8]Matters are much more complicated once endogenous sunk costs (such as advertising or R&D) are introduced (Symeonidis

at equilibrium. We summarize the properties of the final stage subgame by expressing equilibrium price as a function $p(N|\theta)$, where p is price, N is the number of firms, and θ is a shift parameter indexing the 'toughness of price competition'. It is assumed that firms operate with constant marginal cost, and that increases in market size occur by way of successive replications of the population of consumers, so that the distribution of tastes remains constant. Under these circumstances, the vector of equilibrium prices is independent of market size, S, and equilibrium profit can be written in the form $S\pi(N|\theta)$, where the function $\pi(N|\theta)$ is the 'solved out' profit function of the final stage ('price competition') subgame. The function $\pi(N|\theta)$ is decreasing in N. A rise in θ, by definition, leads to a fall in equilibrium price, for any given N. If we assume that the profit per firm is also decreasing in price for all prices below the monopoly level, then an increase in θ implies a fall in profit per firm.

(1995)).

If we define the equilibrium number of firms, N^*, as the largest integer satisfying

$$S\pi(N^*|\theta) \geq \varepsilon$$

then N^* rises with S, leading to a fall in concentration, measured by 1/N, as market size increases.

Once this argument is extended to a multiproduct firm setting, in which each firm is free to enter any number of distinct product varieties, at a cost of ϵ per product, this functional relationship is replaced by a lower bound relation. At a given S, we may have a large number of single product firms, or a smaller number of firms, each with several products (Sutton (1991), Chapter 2).

The parameter θ captures the effect of exogenous influences, such as legal restraints on competition, or changes in transport costs that intensify competition between remote firms. Changes in such factors lead to a shift in the functional relationship between equilibrium price, and profit, for any *given* market structure. The phrase 'toughness of price competition' refers to this functional relationship. For any fixed value of S, an increase in the toughness of price competition shifts the lower bound to concentration upwards (Figure 3).

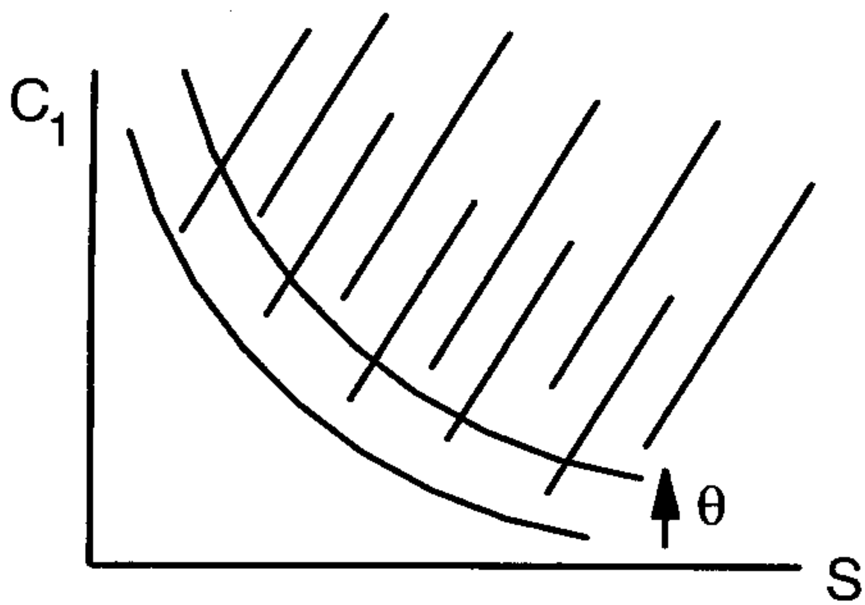

Figure 3. Increasing the toughness of price competition.

The 'price competition mechanism can be observed empirically by looking at certain 'natural experiments' in which the institutional factors affecting the toughness of price competition underwent a substantial change at some point. For example:

- In Sutton (1991), the histories of the salt and sugar industries are examined by reference to changes in transport costs, and to shifts in competition policy, both across countries, and over time. The structural shifts and cross-country differences in structure appear to be closely in line with the operation of this mechanism.

- A major shift in competition policy occurred in the U.K. in the 1960s, with the outlawing of various restrictive agreements between firms. This allows a comparison of structural shifts between the group of industries so affected, and a control group of industries in which no such agreements had existed prior to the legal changes. It has been known for some time that concentration appeared to have increased in those industries that were affected by this increase in the toughness of price competition. A recent detailed comparison of the two groups of industries by Symeonidis (1995) offers strong support for this view.

Narrowing the Domain

The preceding results turned on the assumption that profit-per-firm was decreasing in N. In the special setting in which firms offer homogenous products, and in which all firms earn equal profit at equilibrium, a stronger assumption can be justified: that total industry profit is decreasing in N. In other words, we replace our earlier assumption that $\pi(N|\theta)$ is decreasing in N, by the stronger assumption that $N\pi(N|\theta)$ is decreasing in N. This assumption is not quite so restrictive as might appear to be the case at first glance. If, for example, products are differentiated, and if each firm offers a single variety, this assumption remains valid so long as total industry sales respond only weakly to the introduction of new products, prices being held constant, i.e. the 'market expansion' effect is weak, in the terminology of Shaked and Sutton (1990). In this form, the assumption can be justified for an interestingly wide range of markets. One important example is where each firm operates a single retail outlet within a small town. Since no customer is very far from any store, the market expansion effect from adding stores is small.

This stronger property implies that, as we increase the size of the market (by way of successive replications of the population of consumers), *the equilibrium number of firms increases less than proportionally with the size of the market*. To see this, define the minimum ('threshold') market that can support N sellers as S_N, via the equation

$$S_N \pi(N|\theta) = \varepsilon$$

whence $$\frac{S_N}{N} = \frac{\varepsilon}{N\pi(N|\theta)}$$

The assumption that $N\pi(N|\theta)$ is decreasing in N now implies that the threshold size S_N rises more than proportionally with N. In other words, an increase in market size leads to a less than proportionate increase in the number of sellers.

Bresnahan and Reiss (1987,1990) analyse this effect by reference to a set of 'isolated towns' across the United States. They look at the number of retail establishments of a specific kind, such as gas stations, as a function of the size of the market, measured by population (together with some ancillary variables whose influence is minor).

A central focus of interest lies in comparing the threshold size of market at which entry by a monopolist becomes profitable, with the threshold size that suffices to support a duopoly. Given data on town population and the number of sellers present, the authors proceed to estimate these threshold sizes.

The basic specification used by the authors is as follows: Firms move simultaneously. Entrants have the same profit function.

Fixed cost is unobservable, and are different for each firm, the realizations of fixed cost being independent draws from the same normal distribution. This allows the model to be estimated using an ordered probit model[9].

This analysis is extended in Bresnahan and Reiss (1987) to various types of retailers (or sellers of professional services) and to larger towns supporting several outlets or sellers. The most striking result to emerge is that the price competition effect is exhausted once 3-5 sellers are present; thereafter, an increase in market size leads to a proportionate increase in the number of sellers.

It seems then, that the predicted 'less-than-proportional' increase in the number of sellers is indeed borne out in the data. But this observation begs an obvious question: could this less-than-proportionate increase be explained by reference to a simple alternative story, quite independently of the price competition effect? Suppose some sellers are more efficient than others, so

---

[9]In specifying an appropriate econometric model, some assumption is needed on the appropriate error specification. Various forms are experimented with in Bresnahan and Reiss (1990), but the choice makes little difference to the estimates. Alternative assumptions were also tried regarding the appropriate form of the entry game. The main results were not sensitive to these changes.

that a pool of potential entrants of varying efficiency levels is available. Markets of small size attract the most efficient entrant. If efficiency levels differ greatly, a large increase in size is needed before the second firm will enter, even if there is no fall in price after entry.

How can this 'heterogeneity of firms' interpretation of the 'less-than-proportional' increase in the number of sellers be distinguished from the 'competition effect' interpretation? Berry (1992) shows how this problem can be tackled by reference to a set of markets across which the same group of firms (potential entrants) are active. His study relates to airlines servicing 1,219 routes between fifty U.S. cities. It is known that substantial efficiency differences exist between different airlines, and that their relative efficiency levels may differ across markets (routes). This context is a natural one in which to tackle the 'heterogeneity of firms' issue.

The unit of analysis (individual market) in this case is a city pair. Each airline is either 'active' or 'inactive' in any market. This allows the same form of 'single product' model to be used, as was used in the Bresnahan-Reiss study.

Bresnahan and Reiss modelled firms' fixed costs (or profits) as independent draws from some underlying distribution, and proceeded to estimate entry thresholds for 1, 2, 3 ... firms, without restricting the form of the relation between firm numbers and profit levels. Berry, on the other hand, posits a particular (logarithmic) form for the relationship linking profits to the number of firms, but he goes beyond the Bresnahan-Reiss specification by introducing a firm-specific contribution to profits. The profit of firm i in market k is written as

$$\pi_{i,k}(N) = X_i\beta - \delta\, lnN + Z_{i,k}\alpha + \rho u_{i,0} + \sigma u_{i,k}$$

Here X_i is a vector of market characteristics, N is the number of firms, $Z_{i,k}$ is a vector of firm characteristics, while β, δ, α, ρ, and σ are parameters to be estimated.

The unobserved component $\rho u_{i,0} + \sigma u_{i,k} = \varepsilon_{i,k}$ is a combination of a market specific contribution $u_{i,o}$ and a firm-specific term $u_{i,k}$ that is unobserved by the econometrician, but is known to the firm.

One could, in principle, proceed by partitioning the space of $\varepsilon_{i,k}$ into regions corresponding to different equilibrium outcomes, and writing down a likelihood function. However, once we have a large number of entrants that differ in their observed characteristics, this partitioning involves a very large number of irregularly-shaped zones. Writing down an explicit representation of the likelihood function is infeasible, and the author uses simulation estimators to get around this difficulty. The estimates are then compared to those obtained by ignoring firm heterogeneity and applying an ordered probit model. The results obtained in the two cases differ substantially[10].

The preferred model, which allows for firm heterogeneity, indicates a substantial price competition effect, and this is consistent with the Bresnahan-Reiss interpretation of the observed 'less-than-proportional' increase in the number of sellers[11].

---

[10]It turns out that allowing for the presence of heterogeneity has a major effect on the estimated parameters of the model. A specification that assumes homogeneity of the firms does not, in this setting, lead to the predicted form of the relation between firm numbers and market size, in the sense that the coefficient δ is not significantly different from zero.

[11]The restriction introduced by Berry on the functional form of the profit/numbers relation appears not to be unduly restrictive. To explore the robustness of results in this regard, the model was estimated with separate intercept terms for $N = 1, 2, 3, 4$ and with profit declining linearly with N for $N > 4$. The results were qualitatively similar to those obtained with the restricted form, and the restricted specification could not be rejected against the more general model.

3. Strong Mechanisms II: Escalation and Nonconvergence

Once we turn to those industries where advertising and R&D play a significant role, a second type of mechanism appears, which shares the 'robust' features of the price competition mechanism. The basic theorem is again stated relative to the class of multistage games, in which each firm incurs a fixed and sunk cost F in some early stage(s), and thereafter earns ('gross' or 'post-entry') profit $S\pi$ in some final stage ('price competition') subgame, where S denotes market size and π is a function of the pattern of products entered (and so of the fixed costs incurred) by all firms in earlier stages.

The main theorem is as follows (Shaked and Sutton (1987)):

Suppose: for some constants $a > 0$ and $K > 1$, a firm that spends K times as much as any rival on fixed outlays will earn a final stage payoff no less than aS;

Then: there is a lower bound to concentration (as measured by the maximal market share of the largest firm), which is independent of the size of the market.

The idea is this: as market size increases, the incentives to escalate spending on fixed outlays rises. Increases in market size will be associated with a rise in fixed outlays by at least some firms, and this effect will be sufficiently strong to exclude an indefinite decline in the level of concentration.

The lower bound to concentration depends on the degree to which an escalation of fixed outlays results in profits at the final stage, and so on the constants a and K via the ratio $a/(1+K)$. If we choose the pair (a,K) which maximises this ratio, and write the maximal value of the ratio as α, then each industry can simply be labelled by the scalar index α.

The main empirical problem lies in the fact that there is no direct way of measuring α and so predicting the value of the lower bound to concentration. One way forward lies in making the bold hypothesis that for some group of (advertising - or R&D - intensive) industries, α lies above some minimal positive level, so that for the pooled sample of industries, the empirically estimated lower bound to concentration lies above some strictly positive value.

How can such a prediction be tested? One route would be to look at the same industry across a series of countries. A potential problem arises where some firms operate across several countries. In the case of advertising-intensive industries, this may not be a serious problem, since the firm must spend fixed outlays to create its 'brand image' in each country. For R&D-intensive industries, the problem is fatal, for products need only be invented once. In R&D intensive industries, we need to think in terms of a single global market for each product.

If cross-country studies are ruled out, what of studies based on a comparison of different industries within a single country? Here, the first problem is to control for the 'exogenous' element of fixed outlays, which may crudely be identified with the cost of constructing a production plant. Measuring market size in units equal to the cost of constructing a single m.e.s. plant offers a crude way of controlling for this. This line of attack is clearly more attractive at very low levels of aggregation, where we are dealing with industries in which all firms produce the same range of products using similar production methods.

Both these lines of attack have been pursued in the recent literature. Sutton (1991) presented evidence for twenty food and drink industries across six countries, splitting the sample into a (very) 'low advertising' group and a 'high advertising' group. Robinson (1993), using the PIMS dataset for the U.S., examined 1,880 observations on businesses, classifying the industries in which the businesses operated into Advertising-intensive, R&D-intensive and others. Most recently, a consortium of European economists have assembled a large dataset for 3-digit industries across four European countries, and have looked at both advertising intensive and R&D intensive industries (Matraves (1992), Lyons et al. (1995)). All these studies indicate that the 'non-convergence' property appears to hold good for advertising intensive and R&D intensive industries[12].

The non-convergence property might seem at first glance to represent a fairly weak constraint on the data. It is interesting, therefore to ask what this relationship implies for the older 'regression analyses' of the determinants of concentration. In

---

[12]These studies all involve the notion that α lies above some minimal level for all industries in the advertising - or R&D - intensive group. Observed levels of the advertising-sales ratio or the R&D/sales ratio are used to partition the sample. This is a crude assumption, and in the case of R&D intensive industries is problematic. Sutton (1996a) notes that the value of alpha may be arbitrarily close to zero for some types of technology, even though the equilibrium R&D/sales ratio is high; and describes a method of attack which circumvents this problem.

these studies it was assumed that observed concentration levels might be 'explained' by certain 'Barriers to entry' that might be proxied by measures of scale economies, advertising-intensity, and R&D intensity. Regressions of concentration on these variables indicated that scale economies and advertising intensity were associated with higher levels of concentration[13]. An interesting implication of the lower bound property is that the presence of this bound is sufficient to *imply* that the elementary regressions of concentration on scale economies and advertising intensity, which under this theory are a mis-specification, would indeed yield the positive correlations that were observed by Hart and Clarke (1980); in other words, the bounds results encompass these basic regression results (Sutton (1991), pages 124, 127-8).

One important feature of the game-theoretic approach is that its claims can be tested in a direct and powerful way by turning to case-history evidence. In contrast to 'Walrasian' models based on a fictitious 'auctioneer', any model based on a Nash Equilibrium concept makes a claim about *how* disequilibrium situations are resolved. If an outcome cannot be supported as a Nash equilibrium, then it follows by definition that some

[13]The results regarding R&D-intensity are more complex. (Sutton (1996a)).

'profitable deviation' is available to some firm. The empirical content of any game-theoretic model lies in a claim that certain outcomes will not be observed, and the model of necessity provides us with a qualitative description of the kind of deviation which will be profitable in such a configuration. This feature of the game-theoretic approach greatly enhances the scope for testing: if the theory is correct, then it should be possible in case studies to show such 'profitable deviations' at work. The escalation mechanism carries a fingerprint that should be observed in case histories under certain well defined circumstances. The fingerprint of an escalation process involves a combination of rising advertising/sales or R&D/sales ratios, together with declining profit/sales ratios, leading in due course to the shakeout of all but a small group of leading firms. Examples of this process at work have been documented, for example in Sutton (1991), Chapters 8, 10, 12 and 13.

Narrowing the Domain

The nonconvergence property places a lower bound on concentration. In other words, a small number of firms will dominate the market at any time. Will this group be stable over time, or will its composition change from one product generation to the next? To proceed further, we must narrow the domain of analysis, and focus attention on features of the market that are highly industry-specific.

In general, it is extremely difficult to say anything about patterns of industry leadership across successive product generations. Extreme patterns have attracted much attention in the literature ('persistence of dominance' versus 'leapfrogging'). Yet the factors determining whether such patterns will emerge are notoriously sensitive to the beliefs of agents, the nature of price competition, and other factors which are notoriously difficult to measure[14] (Beath, Katsoulacos and Ulph (1987), Vickers (1986), Budd, Harris and Vickers (1993)). Only in quite special circumstances can we hope to make any progress on this issue.

One interesting set of circumstances is that in which learning effects are large, and the spillover of benefits to rival firms is relatively small, so that the influence of learning effects on the evolution of market structure is important. Games which feature learning effects have been widely studied in the

[14]They are also very sensitive to factors that are usually 'assumed away' in the I.O. literature, relating to inter-divisional conflicts within firms (Foster (1986)).

literature (see for example, Spence (1981), Cabral and Riordan (1994)). When this feature is combined with a strong carryover of the private benefits of learning from one product generation to the next, then a leadership position today generates an advantage which consolidates that leadership position tomorrow. In this setting, some conclusions can be drawn for the evolution of leadership over time (Gruber (1994)).

An industry that has been much studied in recent years is the market for semiconductor memory chips during the 1970s and '80s[15]. Here, the role played by learning effects is known to be very large. Learning is measured by changes in the proportion of chips that are 'satisfactory', initial wastage rates being as high as 80% and final rates being as low as 20%, the latter figure being achieved within a couple of years - a time period which is large compared with a product generation. Moreover, the carryover of learning benefits across successive generations of chips appears to be very substantial.

An unusual 'natural experiment' is afforded by a comparison of

[15]The evolution of market structure in the semiconductor industry has been widely studied, notably by Flaherty (1984) and Dorfman (1987). Recent studies of learning effects include Irwin and Klenow (1994).

the evolution of the markets for two types of memory chips during the 1970s and '80s, as reported in Gruber (1994). Chips of the EPROM type differ from those of the DRAM type in two relevant respects. As Gruber remarks:

> Production engineers maintain that the learning curve at the firm level is very similar for single generations of DRAMs and EPROMs. Because of the larger market for DRAMs, a DRAM producer can learn faster than an EPROM producer. On the other hand, DRAM producers have to defend their market share within a given generation for a longer time before the availability of a new generations gives them scope for (vertical) product differentiation. In other words, DRAM producers have to compete for a long period while sitting on the flat part of the learning curve. Cost advantages would be possible if there were economies of scale to exploit. EPROM producers, on the other hand, require much more time to complete learning because of the smaller market. Moreover, once a firm has moved down the learning curve, then after not too long the next generation is already coming out. Because of the relatively slow learning, firms tend to differ in the position on the learning

curve. Competition in the EPROM market is therefore more likely to be driven by learning curve effects.

Now if learning effects matter, as in the case of EPROM chips, it follows in Gruber's model that an equilibrium pattern of market shares may persist over time, in which one firm operates as a 'leader' over successive generations, entering each generation early, and exiting early also. This firm enjoys high profits as the sole producer of the new generation for a short period during which price is very high. It then switches its production to the next generation as entry occurs and prices fall. Other firms may follow one or other of two strategies. They may choose to enter later than the leader, incurring lower fixed costs, as a result of learning spillovers or slower R&D programs, but remain active longer than the leader within each generation. On the other hand, a firm may choose to spend even less on fixed outlays, and enter late, eventually becoming the sole remaining supplier of 'old generation' chips (for which some residual market will remain, after all rivals have switched their production capacity forward to new generations). Within any given generation of chips, the 'leader' firm will have a market share that declines over time, while a firm following the 'late entry' strategy will have a market share that rises over time.

Firms following an intermediate strategy will have a share that rises and then falls.

The evolution of market shares in the EPROM market follows this pattern, with Intel as 'leader', and with Texas Instruments as one of a number of 'second tier' suppliers, while AMD plays the third 'late entrant' strategy. The leadership pattern is not completely stable (for example, Texas led Intel in 64K chips, and its market share for this category declined rapidly over time). Nonetheless there appears to be a strong pattern in the market share profiles, with Intel's share in each segment falling rapidly, while AMD's starts late and rises over time. That this pattern can indeed be traced to the role of learning effects seems to be well established by direct evidence on differences in learning effects as between EPROMs and DRAMs, and by the fact that this stable pattern of market shares over successive generations did not appear in the DRAM market (Gruber (1994), p.67).

But is this a 'test of theory'? As with most attempts to narrow the domain of analysis with a view to obtaining tighter restrictions on outcomes, this exercise falls short of providing a test of the kind we have been discussing in earlier sections. The model incorporates the market-specific features of learning effects which carry over across product generations. But it is not the case that any model with these features will necessarily generate the 'three-strategy' pattern of Gruber's equilibrium. The emergence of this 'realistic' feature of the market turns on the exact design of the model, and on parameter values that can not be estimated directly. In other words, the model illustrates a possible pattern of events that meshes well with what we see in practice. Even the best of single-industry studies may be able to progress no further than this. The problems we face in such single-industry studies are typical of the more general class of problems posed by 'unobservables', to which we turn in the next section.

4. The Limitations of Game-Theoretic Models

Unobservables and History

If a process of narrowing the domain by reference to observable industry characteristics could be extended indefinitely, then we might move up the frontier shown in Figure 2, arriving eventually at a point where we had 'one true model' which specified a single 'equilibrium structure' for each industry. If this were so, then empirical studies of market structure could be forced back into the traditional ('regression analysis') mode, in which observable market characteristics are assumed to determine a unique equilibrium structure, up to some 'random error' term.

But a central message of the game-theoretic I.O. literature is that such a programme is infeasible. The presence of multiple equilibria, and - more importantly - the role played by unobservable features of the market rules out any such goal. What kinds of feature must be regarded as 'unobservables'? It is useful to distinguish between two types of candidate. The first is a feature that is simply hard to identify, measure of proxy within available datasets. Consider, for example, the kind of 'strategic asymmetry' which we model as a 'first-mover advantage'. This kind of asymmetry is subtle. Even if we have detailed historical information for a particular industry, it may be difficult to decide whether a firm chose its plant capacity on the basis of a correct belief that some rival firm would adjust its planned capacity accordingly. And yet we can, in rare and special circumstances, be lucky. Sometimes the world throws up natural experiments in which accidents in the timing of market entry are such that we can confidently assert that a 'strategic asymmetry' was present. Better still, we may be able to find examples where such an asymmetry was present in some countries, but absent in others. The infamous 'margarine laws', which inhibited the sale of retail margarine in the United States up to the 1950s, afford an unusual and instructive natural experiment of this kind (Sutton (1991), Chapter 9). Notwithstanding such happy accidents, however, it would be a hopeless business to try to incorporate the influence of such subtle but important influences on structure into a cross-industry study, except by way of exploring particular industry histories in the hope of uncovering occasional natural experiments.

The second kind of feature that must be treated as unobservable relates to the beliefs held by agents. Here, we are dealing with an aspect of the market that is not merely difficult to measure, but one which is intrinsically unobservable as far as the researcher is concerned. Yet the influence of such beliefs can be far-reaching, as the game-theoretic models insist.

The nature of the difficulty is well illustrated by the events surrounding a sudden and substantial jump in concentration in the U.K. Bread and Flour industries in the 1960s, which are documented in Sutton (1991), p. 166-168. In these industries, a wave of acquisitions in both industries was set off by a shared belief among ('upstream') flour millers and ('downstream') bread bakers that other firms were going to engage in acquisitions -

and this stimulated others to move quickly in order to avoid 'foreclosure'. The interesting thing about this incident is that it was precipitated by a quarrel between two firms, and the actions that followed hinged on the fears of each firm that if it failed to acquire, someone else would. What is remarkable about the incident is that its effects on structure were far-reaching, and have persisted for three decades. Moreover, these events were peculiar to the U.K. market, and appear to be without parallel in other countries. It would seem that any attempt to 'explain' this shift in concentration by reference to the pattern of technology and tastes in the industry must be implausible.

As this example makes clear, the roles of 'unobservables' is closely intertwined with the claim that 'history matters'. Business historians continually emphasise the role of accident and personality in shaping the evolution of firms and industries. What the game-theoretic approach does is to push us into a middle ground, in which dialogue becomes easier. It tells us that economic mechanisms related to observable industry characteristics place important but limited constraints on outcomes, while leaving ample room for the accidents of history to influence what happens within such bounds. The economist is free to extend the list of relevant economic influences on outcomes, but only at the cost of introducing more subtle influences than we can hope to control for, by reference to any objective measures of 'industry characteristics'.

Independence Effects

While the problem posed by unobservables is intrinsic and unavoidable there is a second problem which, though central, is more tractable. This relates to the presence of 'independence effects'.

Any industry will contain clusters of products or plants that compete closely. But an industry, as conventionally defined in official statistics, will usually contain more than one such cluster; it will be possible to identify pairs, or sets, of products that do not compete directly. What is at issue here is that a firm's profit function may be additively separable into contributions deriving from a number of 'remote' products[16]. Any real market in which products are spread either over some geographic space, or some space of attributes, will tend to

---

[16]Consider, for example, the standard Hotelling model where products are placed along a line. A firm offering a set of non-neighbouring products has, at equilibrium, a profit function which is additively separable into contributions from each product.

exhibit this feature. In other words, most conventionally defined industries exhibit both some strategic interdependence, and some degree of independence across submarkets.

The game-theoretic literature has been concerned with exploring strategic interdependence, and this program involves characterizing the full set of 'equilibria' for the corresponding model. Once separate submarkets are present, however, it is natural to ask whether some combinations of outcomes are more or less 'likely' to occur[17].

The reason for emphasising the importance of this issue, in the

---

[17]A serious theoretical issue arises in this setting. Game-theoretic models of markets containing independent submarkets will usually have many equilibria. Some of these equilibria will involve strategies in which actions taken in one market are conditioned on earlier actions taken in another. Indeed, equilibria of this kind do seem to be empirically relevant in some cases, as, for example, in the 'chain-store' paradox literature. Yet in practice, this kind of 'strategic interdependence' is probably not very common across the general run of markets. A focus on 'listing all the equilibria', if used as a general device, may lead us to overstate the scope of strategic interdependence, and to ignore the role played by independence effects. The introduction of 'independence effects' within game-theoretic models demands that certain restrictions be placed on the strategy space of the game, a move which runs counter to current practice in this area.

present context, is because of the ubiquity in the standard game-theoretic models of 'least concentrated' outcomes in which N firms each have the same minimal (unit) size. These 'symmetric' outcomes play a central role in the theoretical literature, especially in relation to the definition of lower bounds to concentration. Such 'symmetric' outcomes are rarely, if ever, encountered in practice; rather, it is well known that the size distribution of firms in an industry is normally rather skewed. One way of seeing why this is so, and thereby bringing game-theoretic models into a closer mesh with empirical evidence, lies in building game-theoretic models of markets that consist of separate sub-markets, in which the roles of strategic interdependence, and of independence effects, can be combined. By doing so, we might build a bridge between the modern game-theoretic literature, in which strategic interactions are the sole focus of attention, and the older I.O. literature on the 'Growth of Firms', which appealed to independence effects in order to account for the skewed nature of the size distribution of firms. This point is developed in (Sutton 1996b)).

5. Concluding Remarks

Five years ago, it was widely claimed that the game-theoretic approach was 'empty', because everything depended on the details, and no useful constraints were placed upon the data. Nowadays, such criticisms are becoming rare. In following the logic of the game-theoretic approach, we have been led in a natural way to a new set of theories, and there seems to be some basis for the claim that these theories 'work'. This has in turn led to a new kind of criticism. Since we emphasise the primacy of a few strong mechanisms, the huge scaffolding of the game-theoretic literature appears to collapse into a simply articulated theory, which captures the first-order effects in the data, together with a rich menu of models corresponding to 'special cases'. A criticism we now hear is: why bother with the game-theoretic structure at all? Why not just write down the simple (general) models directly?

This argument is unpersuasive. For thirty years, empirical research focused heavily on the study of cross-sectional regularities. In looking at such regularities, researchers could have turned to the kind of structure used, for example, in the Bounds approach. In fact, however, they turned to quite different structures. Moreover, the regression relationships they looked for, and reported, are not the ones to which we are led by the game-theoretic approach.

I have, however, a deeper reason for being uneasy about the claim that the simple theory could have been 'written down directly'. My unease comes from the fact that the lesson we were forced to learn from a decade of game-theoretic models was a painful one, to which there was a long history of resistance in the profession (Sutton (1993)). A tradition stretching from Marshall to Samuelson emphasised the value of attacking the data 'as if' it was generated by some 'true model'. A minority view, which can be traced to Edgeworth, questioned whether this approach was always justified. But so long as our list of reasonable 'candidate models' remained manageably small, it was easy to dismiss such questions; if several candidate models were available, we could 'let the data decide' by carrying out some kind of model-selection exercise. What the game-theoretic literature did was to make this response sound hollow: the sheer unmanagability of the class of 'reasonable models' forced a change of tack. The only way to obtain any empirically useful constraints on the data was either to confine attention to some very narrow domain where the number of candidate models was small, or else to look at strong mechanisms that held good over some very broad class of candidate models. These approaches have indeed led us back to theories whose structure is pleasingly simple; but these theories are not only different in their detail, but in their form, from what went before. Had some researcher written down these theories in 1970, without reference to game-theoretic models, this would have invited an obvious line of criticism: why stop at these few restrictions? Why these mechanisms rather than others? Why not add more mechanisms, or more structure, or additional assumptions in order to make the model 'richer', or more 'realistic', or to get 'more interesting (tighter) predictions'? And this would simply bring us back to the issues with which I began.

REFERENCES

Beath, J., Katsoulacos, Y., and Ulph, D., (1987), "Sequential Product Innovation and Industry Evolution," Economic Journal (supplement), vol. 97, pp. 32-43.

Berry, S., (1992), 'Estimation of a Model of Entry in the Airline Industry,' Econometrica, vol. 60, pp. 889-917.

Bresnahan, T., and Reiss, P., (1987), 'Do Entry Conditions Vary Across Markets?', Brookings Papers on Economic Activity, vol. 3, pp. 833-881.

Bresnahan, T., and Reiss, P., (1990), 'Entry in Monopoly Markets,' Review of Economic Studies, vol. 57, pp. 531-553.

Bresnahan, T., and Reiss, P., (1991), 'Empirical Models of Discrete Games,' Journal of Econometrics, vol. 48, pp. 57-81.

Budd, C., Harris, C., and Vickers, J.S., (1993), "A Model of the Evolution of Duopoly: Does the Asymmetry Between Firms Tend to Increase of Decrease?", Review of Economic Studies, vol. 60, pp. 543-574.

Cabral, L.M.B. and Riordan, M.H., (1994), "The Learning Curve, Market Dominance and Predatory Pricing," Econometrica, vol. 62, pp. 1115-1140.

Dixit, A.K. and Stiglitz, J.E., (1977), "Monopolistic Competition and Optimum Product Diversity," American Economic Review, vol. 67, pp .297-308.

Dorfman, N.S., (1987), Innovation and Market Structure: Lessons from the Computer and Semiconductor Industries, Cambridge, MA: Ballinger Publishing Company.

Ellison, G., (1994), 'Theories of Cartel Stability and the Joint Executive Committee,' Rand Journal of Economics, vol. 25, pp. 37-57.

Fisher, F.M., (1989), 'Games Economists Play: A Noncooperative View, Rand Journal of Economics, vol. 20, pp. 113-24.

Flaherty, M.T. (1984), "Market Share Determination in International Semiconductor Markets," Harvard Business School Working Paper.

Foster, R.N., (1986), Innovation: The Attacker's Advantage, New York: Summit Books.

Gasmi, F., Laffont, J.J. and Vuong, Q.H., (1990), 'A Structural Approach to Empirical Analysis of Collusive Behavior,' European Economic Review, vol. 34, pp. 513-523.

Gruber, H. (1992), "Persistence of Leadership in Product Innovation, Journal of Industrial Economics, vol. 40, pp. 359-375.

Gruber, H. (1994), 'Learning by Doing Spillovers in the Semiconductor Industry,' Journal of Political Economy, vol. 102, pp. 1200-1227.

Gruber, H. (1994), Learning and Strategic Product Innovation: Theory and Evidence for the Semiconductor Industry, Amsterdam: North Holland.

Gruber, H. (1995), 'Market Structure, learning and Product Innovation: Evidence for the EPROM Market,' International Journal of the Economics of Business, vol. 2, pp. 87-101.

Hart, P.E. and Clarke, R., (1980), Concentration in British Industry, 1935-75, Cambridge: Cambridge University Press.

Hendricks, K. and Porter, R.H. (1988), An Empirical Study of an Auction with Asymmetrical Information,' American Economic Review, vol. 14, pp.301-314.

Irwin, D.A. and Klenow, P.J., (1994), "Learning-by-Doing Spillovers in the Semiconductor Industry," Journal of Political Economy, vol. 102, pp. 1200-1227.

Laffont, J. J., (1996), "Game Theory and Structural Economics: the Case of Auction Data," (Marshall Lecture, European Economic Association 1995), European Economic Review, forthcoming.

Levenstein, M., (1993), "Price Wars and the Stability of Collusion: A Study of the Pre-World War I Bromine Industry," NBER Working Paper No. 50, Cambridge, MA.

Lyons, B., and Matraves, C., (1995), "Industrial Concentration and Endogenous Sunk Costs in the European Union," Economics Research Centre Discussion Paper No. 9505, University of East Anglia.

Pelzman, S. (1991), 'The Handbook of Industrial Organization: A Review Article,' Journal of Political Economy, vol. 99, pp. 201-

217.

Porter, R., (1983), 'A Study of Cartel Stability: the Joint Executive Committee 1880 - 1886, Rand Journal of Economics, vol. 14, pp. 301-314.

Roberts, M.J. and Samuelson, L., (1988), 'An Empirical Analysis of Dynamic Nonprice Competition in an Oligopolistic Industry,' Rand Journal of Economics, vol 19, pp. 200-220.

Robinson, W., (1993), "Are Sutton's Predictions Robust?: Empirical Evidence on Price Competition, Advertising and the Evolution of Concentration," University of Michigan, unpublished.

Rosenthal, R.J., (1980), "A Model in Which an Increase in the Number of Sellers Leads to a Higher Price", Econometrica, vol 48, pp. 1575-9.

Shaked, A., and Sutton, J., (1987), 'Product Differentiation and Industrial Structure,' Journal of Industrial Economics, vol. 36, pp. 131-146.

Shaked, A., and Sutton, J., (1990), "Multiproduct Firms and

Market Structure," Rand Journal of Economics, vol. 21, pp. 45-62.

Shubik, M. and Levitan, R., (1980), Market Structure and Behaviour, Cambridge, MA: Harvard University Press.

Spence, M.A., (1981), "The Learning Curve and Competition," Bell Journal of Economics, vol.12, pp. 49-70.

Sutton, J., (1990), Explaining Everything, Explaining Nothing? Game Theoretic Models in Industrial Organization, European Economic Review, vol. 34, pp. 505-512.

Sutton, J., (1991), Sunk Costs and Market Structure, Cambridge, MA: MIT Press.

Sutton, J., (1993), "Echoes of Edgeworth: The Problem of Indeterminacy", European Economic Review, vol. 37, pp. 491-499.

Sutton, J., (1995), "One Smart Agent," STICERD Discussion Paper No. EI/8, London School of Economics.

Sutton, J., (1996a), "Technology and Market Structure," (Schumpeter Lecture, European Economic Association, September 1995), European Economic Review, forthcoming.

Sutton, J., (1996b), "Gibrat's Legacy," Journal of Economic Literature, forthcoming.

Symeonidis, G., (1995), "Competition Law and Market Structure: A Study of the Effect of Restrictive Practices Legislation in the U.K., 1960-1975," unpublished.

Trajtenberg, M., (1990), Economic Analysis of Product Innovation: The Case of CT Scanners, Cambridge, MA: Harvard University Press.

Vickers, J.S., (1986), The Evolution of Market Structure When There is a Sequence of Innovations, Journal of Industrial Economics, vol. 35, pp. 1-12.

Weiss, L. (ed.), (1989), Concentration and Price, Cambridge, MA: MIT Press.