PRICE-COST MARGINS AND MARKET STRUCTURE

by

Michael John Waterson B.A., M.Sc.,(Econ.).

submitted for the degree of Ph.D. in Economics

at

The University of Warwick

Research conducted in the Department of Economics

Submitted: September 1976

M.J. Waterson

Price-Cost Margins and Market Structure

Summary

In this thesis, the relationship between some aspects of industrial market structure and industry price-cost margins or profit-revenue ratios is investigated. This is done mainly by building mathematical models based upon the tenet of profit maximisation. Empirical tests of the hypotheses developed are carried out using regression analysis on recent UK data.

After an introductory chapter, the arguments are developed by successively taking structural features into account. Thus initially, problems involved in relating the structure of established firms in an industry to price-cost margins are considered. Then the possibilities and problems of potential entry into an industry are opened up. After that, the power of buyers from and sellers to the industry are brought into partial account. Additional potentially relevant structural factors receive a more cursory treatment before the analysis passes to empirical testing. At every stage, the relevant established literature is reviewed. It is found theoretically that the price-cost margin may be related to two main aspects of market structure, the "Herfindahl" index and a bilateral power index developed here. However, the commonly included "entry barrier" variables need not, under reasonable assumptions, be considered relevant. The empirical results lend support to the theoretical conclusions regarding the Herfindahl and bilateral power indices.

The contribution to knowledge in the subject area achieved herein is (hopefully) mainly in the rigorous development and application of

models which have, in general, previously been rather vaguely based upon commonsense extensions of the fundamentals of economic theory. In fact, the thesis consists to a large extent in the belief that industrial economic problems often considered as having theoretically indeterminate solutions may be profitably examined and "solved" rigorously, with the judicious use of restrictive assumptions.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS

I should like to thank the following people for their helpful comments on one or more points in the thesis: Keith Cowling, John Cubbin, Steve Davies, Avinash Dixit, Tim Hazeldine, Norman Ireland, John Kay, Alan Kirman, Dave Morris, Geoff Renshaw, Charles Rowley and George Yarrow. Their assistance is as a rule acknowledged at the relevant place in the thesis. The main exception to this rule is Keith Cowling, my supervisor, who has given much help and encouragement at all stages and has been very conscientious in discussing and reading through my work. I should also like to thank Dianne Ellwood for some assistance with data preparation in the early stages. Finally, I should like to thank Ann Bateman, and especially Christine Garner, for their excellent work in typing the thesis.

DECLARATION

Some of the material of chapter 2 on models of oligopoly was joint work with Keith Cowling and has been published in Cowling and Waterson (1976), (see bibliography for the reference). The models involved are those on pp. 13-14 and pp. 32-33 of the thesis. I consider that these were truly joint work in that Keith Cowling gave the initial impetus and ideas, while I was substantially responsible for their development.

(The use of the first person plural in the thesis is merely a stylistic device, it should not be taken to indicate joint work).

LIST OF DIAGRAMS

LIST OF TABLES

LIST OF SYMBOLS

The following general conventions on symbolism have been adopted in the thesis; detailed definitions are given at every stage where alterations and additions are encountered:

Π	profits
p	prices
q,Q	quantities (outputs)
R	revenue (sales)
c	variable costs
MC	marginal cost
AVC	average variable cost
F	fixed costs
η	own price elasticity of demand
H	the "Herfindahl" index of industry concentration

Upper case letter symbols do not necessarily bear any relationship to lower case letters.

Chapter 1 INTRODUCTION

I The General Area:

The purpose of this thesis is to investigate, in both theoretical and empirical terms, parts of the relationship between industrial structure and profit performance. This is by no means a novel aim. The originality hopefully lies here in the amount of attention directed towards making more explicit the often inchoate theoretical basis of such structure-performance models.

In fact the estimation of structure-profit equations has quite a long history, at least in the US dating back to the 1950's. It lies at the heart of one side of empirical Industrial Economics, the cross-industry estimation of relationships suggested by "economic theory", (the other side being studies of individual industries). The basic premise behind this empirical work is that various factors characterised as structural features of an industry relate, via the conduct of those in (and to some extent outside) the industry, to performance. These relationships, mainly flowing from the structural aspects to performance,[1] form what is known as the Structure-Conduct-Performance paradigm.

To give a very simple example, relevant to our purpose, consider the firm i with profits defined as:

$$\Pi_i = p_i q_i - c_i(q_i) - F_i$$

(Π is profit, p price, q quantity, c(q) variable costs and F fixed costs.) If that firm maximises profit by choosing optimal output then we may say that:

$$\frac{p_i - c'_i(q_i)}{p_i} = -\frac{q_i}{p_i} \cdot \frac{dp_i}{dq_i} = \frac{1}{|\eta_i|} \qquad (1)$$

Footnotes are gathered at the end of each chapter, references at the end of the thesis.

Suppose the firm i is a monopolist, then from (1) we find its "price-cost margin" is equal to the reciprocal of the industry elasticity of demand for the good i sells. The relevant structural feature is the elasticity of demand facing i (at optimal output), and this feeds through, via the assumption that output and price are set to maximise profits (conduct) to the monopolist's price-cost margin, a measure of performance (also known as Lerner's measure of monopoly power). The theory, profit maximisation, has provided, via the paradigm, a hypothesis which is in principle testable cross-sectionally (say) by comparing the price-cost margins of monopolists faced by differing industry elasticities of demand.

Now of course the sample of monopolists on which we could test such a hypothesis would be small. On the other hand if we are able to say something, for an industry of several firms, about the relationship between the elasticity of demand facing a firm and that facing the industry then we could expand the potential sample greatly. This is one aspect of the task facing oligopoly theory, which is concerned with divining likely pricing and output outcomes when numbers in the industry are such that each firm has a perceptible influence on the fortunes of others. Almost by definition, because of the interdependence, there is no unique solution, but there is a "general presumption" among economists that the greater the numbers in an industry, the more likely it is that prices and outputs approach the competitive level. In fact, as we show in the following chapter, despite the variety of solutions offered to the oligopoly problem, there are quite useful ways to characterise a wide range of these.

Wrapped up also with the concept of oligopoly theory is the problem of potential entry. For if firms look as if they are going to enter an industry, thereby lowering profits for those firms already established, we might expect some reaction from established firms to the potential threat. It is normally presumed that there exist partial barriers to new entry into an industry, either pre-existing or erected by established firms, but that nevertheless prices have to be set with some regard to the possibility of entry.[2]

The two paragraphs above have outlined the main considerations which those who estimate structure-performance relationships wish to capture. Thus, a common activity in the UK is to attempt to explain the price-cost margin by means of the five-firm concentration ratio, some measure of plant scale economies, the advertising-sales ratio and so on.[3] Perhaps (on a charitable interpretation) because of space pressures, empirical studies on these lines have tended to neglect any wide-ranging discussion of the theoretical underpinnings, and have often been content to refer back to previous work (which may be no more explicit), or to use as their justification a reference to the general corpus of economic theory. For example, Khalilzadeh-Shirazi (1974) considers that "the theoretical justification for the inclusion of the foregoing variables (seller concentration, measures of entry barriers, the rate of growth of demand) has been widely discussed in the literature (so) we will only briefly touch upon them." (p.67). Yet in a similar study Holterman (1973) is quite brazen:

"The hypotheses to be tested cannot be derived from economic theory without making many unacceptably restrictive assumptions ... economic theory is about an individual firm producing a single good .. Nevertheless they are all intuitively reasonable and there is nothing in economic theory to suggest that they are false." (p.121)

This is obviously an understatement of the status of economic modelling in the area. But the fact that the foundations for the hypotheses appear obscure makes progress by refutation of untrue postulates rather difficult.

There is a methodological point here. Suppose that existing largely ad hoc formulations work reasonably well. A Friedmanite would probably then argue that we are doing very little if we achieve a similar (or little better) degree of explanation based upon more rigorous theorising. We question this argument by means of an example: Assume a variable Y is in truth related to X^2. Then for certain numerical values of X (around unity), regressing Y on X will provide a good fit to the data. But obviously if some new observations come up in which X is negative (say), the correspondence between X and X^2 will be much less close and our empirical relationship may appear even to have broken down. We would not wish to take this example too far. However, prediction would appear to be less hazardous if we decide to introduce all variables ensuing from our theoretical discussion in as "correct" a form as possible.[4]

Thus our thesis is almost diametrically opposed to that of Holterman. We feel strongly that empirical work in the field should start by being based fairly closely on a particular theory or theories. For this reason we wish to develop mathematical models embodying precise assumptions about firm behaviour following from the basic tenet of profit maximisation. From these we derive formulae relating structural variables to our performance measure. As will be clear when we proceed in later chapters to a more detailed consideration of the theoretical art in the various areas to be covered, this involves substantial new work in extending existing models or even in developing models more or less from first principles. In performing this exercise, we do not of course believe that firms will act entirely accordingly to our formulae, but we do consider that the development of such equations will enable the empirical work to be much more soundly based.[5]

II A Plan of the Thesis

As we have said, we concentrate here on developing oligopoly theory (in directions most useful for subsequent empirical work) prior to empirical estimation of our model. Accordingly we allot the three chapters following this to a fairly extensive discussion and elaboration of the effects important structural variables might have on performance. The general basis, that entrepreneurs attempt (in some sense) to maximise profits, is relaxed, and then only briefly, in chapter 5. We feel we have enough to cover without venturing too far from the traditional.

Chapter 2 is concerned with oligopoly theory proper, without the possibility of entry. As we note in that chapter, there are a number of traditional solutions to the problem, starting with that given by Cournot (1927).[6] There is also a more modern type of approach through game theory. We attempt to review both types of work before going on to develop our particular synthesis which, while holding under fairly general conditions, still leads to interesting predictions.[7] Basically we find that the price-cost margin can be expressed as a function of the Herfindahl index of industry concentration.

The area covered by chapter 2 is probably most developed in terms of previous mathematical modelling of the topics we cover. For we find when we move to consider the problem of potential competition in chapter 3 that work on the static limit pricing model has progressed very little since Modigliani's (1958) masterly analysis. We discuss this at length, together with a recent alternative, and build a synthesis of the two which we then extend to cover, among others, a case where there are many established non-collusive firms. We also attempt an assessment of the predictive content of the dynamic models concerned with entry retardation. However we finally reject the limit pricing model in favour of the capacity creation approach pioneered by Spence (1974) which will normally dominate in terms of profit available to established firms. This does not mean the earlier work is wasted, because the problem of the amount of excess capacity to create is conceptually the same as that of finding the limit price.

Chapter 4 extends our model into the rather neglected area of the influence of bilateral power, that is the power of sellers to and buyers from the industry in question.[8] We consider this to be one of the most novel and important parts of the thesis. As we see in that chapter, theoretical work in this direction has reached only as far as the indication of solutions to bilateral monopoly situations, though we are not alone in believing the concept of bilateral power to be potentially important. It is unfortunately a difficult topic to deal with comprehensively. Accordingly we develop a model (based on that given by Cournot) emanating from quite specific and fairly tractable assumptions. Having done so we consider its relationship to other possible formulations; given the paucity of previous work we do not feel able to do more. Again we derive a formula which can be used (approximately) for empirical purposes, moreover one which has slightly surprising implications. We find that under certain assumptions the "bilateral power" part may be separated from other effects. This finding is used (implicitly) to facilitate some of our exposition in chapter 5.

Chapter 5 is more of a mixture than those listed above. The basic approach having been developed, we feel qualified to discuss previous empirical work. We first assess that work performed on UK data, then consider the question of an alternative performance measure widely used on US data. At the same time these studies throw up additional empirically used explanatory variables which we argue in the main are of rather doubtful relevance as used, within our framework. We do however consider that cyclical fluctuations have a place in our model and discuss a proxy for their effect. The final portions of the chapter very briefly extend our remit somewhat wider in thinking about utility maximisation and the question of lags and possible simultaneity problems.

The final chapter, apart from some isolated points of theory, is given over to detail on the data and sample and then estimation of the model. The data mainly emanates from two published sources, the Census of Production series and the series of Input-Output tables for the U.K. Our tests show, briefly, a reasonable measure of empirical support for our model, with both main structural variables achieving statistical significance, and indicate an apparent absence of serious econometric problems.

There are two appendices to the thesis. The first discusses two assumptions which will be fairly extensively used in developing the models of later chapters. These are that firms have constant marginal cost schedules (equal to average variable cost), and that they face demand curves of the constant elasticity form, within the relevant range. The former is necessary in order to relate the price-cost margin to the profit-revenue ratio and is used in developing the models of chapter 2. The latter is mainly used in chapters 3 and 4 but is also advantageous when we wish to discuss what happens to the margin as the number of firms in an industry changes. We provide brief justifications for the use of both these simplifying assumptions mainly by reference to empirical studies, where possible on UK data.

The second appendix contains some basic facts about the data we use, some of which are presupposed in later chapters (particularly chapter 5). It also includes data on variables we generate which do not come directly from published sources, namely the Herfindahl index and our bilateral power index. One point on the data which has widespread ramifications for our developed functional forms should however be mentioned here. We do this in the section below.

III The "Ratio" formulation

As we saw from our very simple model of the first section, one of the major determinants of price-cost margins in theory is the industry price elasticity of demand. This carries through to the more complex models of later chapters. Unfortunately, values for this have not been tabulated by anyone on any consistent basis. The most detailed work in recent years has been that performed by Deaton. He (1975 b) estimated the elasticity of demand for 37 commodities using various functional forms on post-war time-series data for Britain. Not unnaturally, he concentrated on consumer good industries which, again not unnaturally, are not a predominant feature of the Census. The upshot of this is that we estimate we could use only five of his values directly if we wanted to include figures for the elasticity of demand.

As we shall see in chapter 5, the approach followed by other studies in this field[9] has been to neglect this variable entirely. In doing so they implicitly assume that the elasticity of demand is the same for all industries. An alternative, followed here, is to take the basic theoretically generated structure-performance relationship at two time periods and form a ratio. If we then assume that the ratio of the demand elasticities in the two periods is a constant across industries, a postulate which seems intuitively more reasonable than the alternative, we may neglect that variable altogether. In fact, as we discuss in chapter 5, we modify this assumption by briefly considering possible determinants of differential effects on the elasticity of demand ratio between industries.

Moreover, it turns out that by taking the ratio form of the basic equation we also hopefully cancel out some other elements of inter-industry variation. For example, as we see in Chapter 2, we might expect individual industry's performance at a certain level of concentration to vary while in ratio form a change in concentration seems more likely to produce similar effects across industries. A similar argument might be made about barriers to entry. It is mainly for this reason that we decide to retain the ratio form despite making quite severe assumptions about constancy of elasticities when we come to use the bilateral power measure of chapter 4 empirically.

Despite saying that the ratio form of the basic equation should cancel out some sources of interindustry variance in performance, we do not believe that we are bound to obtain superior results by the use of that technique; perhaps the contrary. Our reasoning here would be that we might expect the profit-revenue ratio across industries not to exhibit proportionately as much variance as the ratio (at two time periods) of profit-revenue ratios across industries. This is fact turns out to be true for our sample. The coefficient of variation in the former case for 1968 is 0.232, while for the ratio of 1968 on 1963 values it is 2.247, nearly ten times as large.[10] This being so, our method of estimation perhaps provides a stricter test of the model to be developed and, in that we are attempting to explain what happens as structure changes, it possibly also provides more policy-relevant conclusions.

FOOTNOTES

1. Though we cannot think of the phenomena as totally unidirectional, as we note briefly in chapter 5.
2. This should not be taken as a full summary of our arguments. See the following section.
3. We go into detail on these studies in chapter 5.
4. This does not imply we would be perfectly happy to indulge in prediction of course.
5. It could be that even after such theoretical work the empirical results would be poor. That at least should enable us to say something about behaviour though.
6. The book was of course written well before 1927, in fact in 1838, but we refer throughout the thesis to the 1927 edition.
7. That approach emanates partially from some work done jointly with Keith Cowling, as we mention where relevant.
8. It turns out in our particular model of the situation, sellers have no effect on the margin though.
9. An exception is Cowling and Waterson (1976), the ratio approach of which we utilise here.
10. The precise samples used will be discussed in Chapter 6. These figures are the ratio of sample standard deviation to mean of the logarithms of the relevant variables, over a sample of 50 industries. A similar pattern emerges for our alternative sample of 51 industries, and also if we use 1963 'level' figures instead of those for 1968.

Chapter 2: Oligopoly - A Basic Structure-Performance Model

I Introduction:

There has been a tremendous volume of writing on the theory of oligopoly in the economics literature, which indicates that it has presented a sizeable problem to the economics profession. While the cases of perfect competition and monopoly have been solved to the satisfaction of the vast majority, oligopoly contains many features which make simple solutions unlikely.

A definition of what actually constitutes an oligopoly is hard to discern, but we shall consider that it is what is probably the general state of affairs prevailing in the economy, where the actions of some or all of the individual firms in an industry are not wholly unaffected by other firms in the industry. It is this interdependence which gives rise to many of the theoretical problems in modelling oligopoly behaviour.

Saying this of course begs the question of what an industry is. For most purposes it is not sensible to define an industry as a group of firms producing a perfectly homogenous product. Although some products appear almost perfectly homogenous (e.g. pure sulphur, cement, 1" wire nails) and many are reasonably homogenous (e.g. tea, lavatory paper, motor oil), a great number of goods, while being of the same type, differ greatly in detail (e.g. shoes, paint and motor cars). It would seem reasonable to consider that even the last of the above sets of examples should constitute industries, since each one satisfies the same type of wants among consumers and is produced by similar processes: that is there is a market for motor cars, for example. However, the presence of heterogeneity does cause problems in formulating theories of oligopoly. Of course, those engaged in empirical work usually have to use statistics calculated on the basis of much wider categories than these, and this may also cause difficulty. Further, the fact that firms often provide a wide variety of products and that consumption of certain goods may enhance or discourage consumption of related goods causes difficulties when using pure theory to explain actual observations.

Allied with the fact that many goods are of a heterogenous nature is the problem that firms may use different mixes of strategies in selling their products in the market. The pure strategy of price/ quantity setting is usually augmented by differing levels of advertising, salesmanship, service, and so forth.[1] Any theory attempting a close explanation of reality needs to take this into account, though of necessity those that do, incorporate such factors in a fairly simplistic manner.

Referring back to the concept of the industry, we should note that even if we can specify the exact number and size of firms solely producing one undifferentiated product forconsumption we still have to consider the possibility that other firms may consider entry into the industry feasible at certain prices. Thus, the behaviour of the individual firms presently in the industry may be affected not only by their anticipations of how their rivals may react, but also by the possibility of the number of rivals increasing (or decreasing). Again, behaviour may be affected by the presence of powerful buyers from, or powerful sellers to, the industry. Such factors will have to be left to later chapters, for here we only intend to discuss the theory of oligopoly in pure form.

In dealing with solutions to the problem of oligopoly pricing, we shall first consider the classical theories using calculus methods and then go on to enquire into the contribution of game theory. Of necessity neither of these sections pretend to be complete reviews of the extant literature, they merely provide a flavour of the type of models which have been used in attempting to describe oligopoly behaviour.

II The Classical Theories:

As stated earlier, interdependence among producers gives rise to a multitude of possible solutions to the oligopoly problem. This is clearly seen if we consider a fairly conventional model,[2] where there are N sellers of a standardised product with a single selling price in a market with no possibility of entry, inputs being purchased at given prices and outputs sold to price takers. Each firm will have a profit function.

$$\Pi_i = pq_i - c(q_i) - F_i \qquad i = 1,2,\ldots n.$$

where q_i is the i^{th} firm's output ($\Sigma q_i = Q$), p is market price, each firm has common variable cost function $c(q_i)$ and fixed costs F_i. Equilibrium is reached by quantity variation, so that for maximum profits we require:

$$\frac{d\Pi_i}{dq_i} = 0 \; ; \qquad \frac{d^2\Pi_i}{dq_i^2} < 0$$

Thus: $$\frac{d\Pi_i}{dq_i} = p + q_i \frac{dp}{dQ} \cdot \frac{dQ}{dq_i} - c'(q_i) = 0 \qquad (1)$$

and $$\frac{d^2\Pi_i}{dq_i^2} = 2\frac{dp}{dQ} \cdot \frac{dQ}{dq_i} + q_i \frac{dp}{dQ} \cdot \frac{d^2Q}{dq_i^2} + q_i \frac{d^2p}{dQ^2} \cdot \left(\frac{dQ}{dq_i}\right)^2 - c''(q_i) < 0$$

The second-order condition implies that the firm's assumed marginal revenue function cuts its marginal cost function from above. Assuming that this holds, we can easily see that the firm's conjecture about interdependence, that is the value of $\frac{dQ}{dq_i} = \frac{d(q_1 + q_2 + \dots + q_n)}{dq_i}$ is of crucial importance to the value of price and quantity at equilibrium. We may alternatively write the above term as:

$$\frac{dQ}{dq_i} = 1 + \frac{d \sum_{j \neq i} q_j}{dq_i} = 1 + \lambda_i, \text{ say.}$$

therefore, summing (1) over N firms yields:

$$Np + \frac{dp}{dQ} \Sigma q_i \frac{dQ}{dq_i} - \Sigma c'(q_i) = 0$$

Dividing by p and writing $\lambda = \Sigma q_i \lambda_i / Q$, $|n| = -\frac{p}{Q} \cdot \frac{dQ}{dp}$ we have:

$$N - \frac{(1 + \lambda)}{|n|} - \frac{\Sigma c'(q_i)}{p} = 0 \quad ; \qquad \text{or}$$

$$\frac{p - \Sigma c'(q_i)/N}{p} = \frac{1 + \lambda}{N|n|} \qquad (2)$$

Cournot's (1927) theory of oligopoly assumes that each firm considers that the other firm's output will not change as a result of the firm in question changing his output. This means that $\lambda_i = 0$ for all i, so $\lambda = 0$; the "conjectural variations" term is zero. From (1) we have:

$$p + q_i \frac{dp}{dQ} - c'(q_i) = 0 \qquad (3)$$

and from (2): $\frac{p - MC}{P} = \frac{1}{N|n|}$ where MC = c' (q_i), all q_i being equal.

(3) yields a set of "reaction functions" which should all intersect at positive outputs to give the equilibrium solution. They can be considered as defining a path to equilibrium.

These reaction functions are utilised by the leader in Stackleberg's theory of oligopoly. Consider the case where all but one firm act as "followers". For them we have that $dQ/dq_i = 1$. However the leader, knowing their reaction functions, uses them to obtain a more favourable position for himself. If he be the m^{th} firm, then he has:

$$\frac{dQ}{dq_m} = \frac{d}{dq_m}(q_1 + q_2 + \dots + q_n)$$

$$= 1 + \frac{\sum_{j \neq m} d\Psi_j(q_m)}{dq_m}$$, where the $\Psi_j(q_m)$ are the followers' "reduced form" reactions to m's output changes. This is obviously a special case of the general result given earlier, where not all the λ_i are equal.

Other special cases which can be developed from the general model include the limiting case of collusion. Here, each firm knows that, if he raises or lowers output, the others will do likewise. Thus $dQ/dq_i = N$ and we have from (1):

$$p + Qdp/dQ - c'(q_i) = 0$$

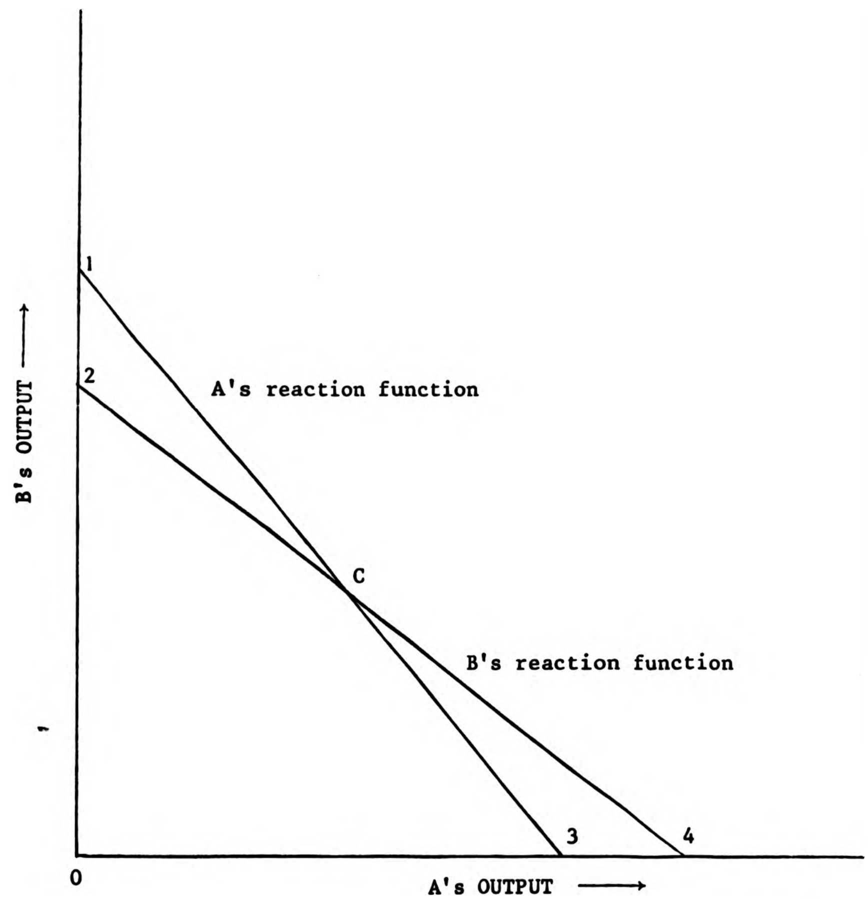
while from (2): $\frac{p - MC}{P} = \frac{1}{|\eta|}$

A natural problem posed by all models involving conjectural variations is that of the stability of the solution. If for simplicity we choose to take a Cournot duopoly then we may represent the reaction functions in a diagram as overleaf (Diagram 1). The idea of the approach to equilibrium may be stated thus:

Suppose A is initially a monopolist producing at point 3. B's initial reaction on entry is read off by tracing a vertical line upward to his reaction function, which gives a new output point at which A finds his position non-optimal. We trace a horizontal line to A's reaction function to obtain his preferred position and the process continues until, hopefully, point C is reached. This is then a stable equilibrium. Obviously, we may not reach a stable point for a number of reasons. In talking about the Cournot model, Fellner(1949) states that "the reaction functions will intersect, and at least one intersection will have to be stable, although there may be several intersections

DIAGRAM 2.1*

Cournot reaction functions



This figure is substantially a redrawing of Fellner's (1949) figure 1 p. 59.

* i.e. diagram 1 of chapter 2. Where it does not cause ambiguity we reference diagrams in the text simply as "diagram 1" etc.

including stable ones. Cournot, of course, actually implied that points 1, 2, 3 and 4 were arranged in this manner ... " (p.63). Again, when talking about conjectural variations on output more generally, he says: "It is generally realised, of course, that the F functions in this extended sense need not intersect for positive outputs, and that if they intersect, the point of intersection need not be stable even in the sense in which the Cournot intersection is stable" (p.73). Further problems could be caused by increases in numbers and by the fact that the firms may realise that their assumptions are being proven wrong.

There are two general answers to such difficulties. The first is to assert that although the dynamic stability may be questionable, the static equilibrium points may exist independent of this. This point will be considered later.[3] The other is of course slavishly to set up Cournot-type models under various assumptions and to consider their dynamic properties. Amongst those who have done this are Quandt (1967), Quandt and McManus (1961), and Hadar (1966).[4] For example, Quandt and McManus set up a Cournot model with (in one case) linear demand and cost functions. Lags are discrete and single-period so that a first-order difference equation results. Stability of this is then easily discerned in the general solution. Quandt summarises the conclusions on the Cournot model as depending heavily on whether the dynamic adjustment is formulated as discrete or continuous: with discrete models, increasing marginal cost is stabilising, and for given demand and cost functions, an increase in the number of firms tends to be destabilising. In contrast, continuous dynamic adjustment processes are stable for all a priori admissable values of the parameter. Quandt and McManus say: "As is usual in dynamics, a lot depends upon just what dynamic assumptions are made". We shall not pursue this particular aspect of stability any further here.

We might also consider that we can represent Sweezy's (1939) Kinked Demand Curve theory using the concept of conjectural variations on output. Although that theory was in fact set up in terms of price reactions, the main points of the model may be brought out by taking it that the firms assume:

$$dq_j/dq_i = 0 \quad \text{for } dq_i < 0$$

and $$dq_j/dq_i = 1 \quad \text{for } dq_i > 0$$

As is well known, the fact that the conjecture about price falls (or quantity increases) is different from that about price rises (quantity falls) gives rise to a kink in the imagined demand curve. The upper section of the curve is very much more elastic than the lower portion, as others follow downwards, yet not upwards. This means of course that the marginal revenue curve is discontinuous over a certain range of prices at the going quantity. However, the supposed main advantage of the kinked demand curve theory, that is its explanation of infrequent price changes as due to marginal cost changes within the discontinuous range not affecting price, is at the same time its major disadvantage. For the corollary of this is that the theory can attempt little explanation of the extent to which price is above marginal cost. Obviously limits may be prescribed, and these will be affected by the industry elasticity of demand and the number of firms in the industry, but within these limits the theory is silent. There is also the question (considered by Efroymson (1943)) of whether the assumptions made are valid in boom conditions and, if so, when a switch is made.

III The Game Theoretic Approach:

So far, we have been considering conventional maximisation solutions to the oligopoly problem; we have encountered a number of these yet there are many more. This plethora of models has lead other economists to look at the situation afresh. As Shubik (1959, p.viii) puts it " ... in an oligopoly ... no maximum problem exists; indeed the notion of a maximum has no meaning. It is necessary to erect a new conceptual and, by necessity, mathematical edifice. This is precisely what has been accomplished by the theory of games." It is our aim to take at least a cursory look at this alternative approach, and to consider its achievements both from a theoretical standpoint and as guides to the direction of empirical work. A full-scale review is not intended, nor will a comprehensive list of the terminology and concepts used be given. We shall be content to give a flavour of the usefulness of game theory by discussing in fair detail two of the major works in this area (Shubik 1959 and Telser 1972) and a closely allied work (that of Nicholson 1972) without wading into their mathematical complexities.

Shubik's book is in two parts. The first half is taken up with setting out most of the conventional models of oligopoly in a game theoretic framework. In doing this some conceivable plays not covered by traditional analysis are exposed, for example the idea of a mixed strategy. There is much discussion of the difficulty of defining the appropriate firm's demand curve, given the industry demand curve, in cases where price is the policy variable; this difficulty also occurs where quantity is the policy variable in models with product differentiation. The problem is solved, theoretically at least, by the introduction and clarification of the concept of "contingent demand". In essence, the difficulty arises because while a firm on setting quantity is (implicitly) willing to accept market price, a price-setter must be able to produce the output required by the market which may vary widely. If, for example, the firm in question sets a price slightly higher than others, he may sell very little or nothing, while a price slightly lower would oblige him to supply virtually the whole of the market. If he is unable or unwilling to do the latter then others may be able to sell at higher prices, giving rise to the concept of contingent demand, which helps to explain the difference between Bertrand's and Edgeworth's theories of oligopoly. Unfortunately, " ... the computation of families of contingent demand functions is complicated and requires more information than is usually available. However, for the purposes of the theory of price-variation duopoly the important feature of contingent demand functions is that they are rarely convex" causing secondary maxima, (p.87).

This analysis of contingent demand enables us to say that with the Bertrand solution "price is not determinate but can fluctuate over a range" (p.126). Shubik's diagram (redrawn overleaf, diagram 2), gives an illustration of this and the line "B' B_1' shows that when the number of competitors increases it becomes progressively less and less desirable to step out of line with one's price policy. Most of the odds are concentrated on prices lying close to EE_1 with only a small possibility left for any major price fluctuation." Thus average price in the Bertrand solution depends on the number of firms in the industry as it does with the Cournot solution, aside from the possibility of massive inventories.[5]

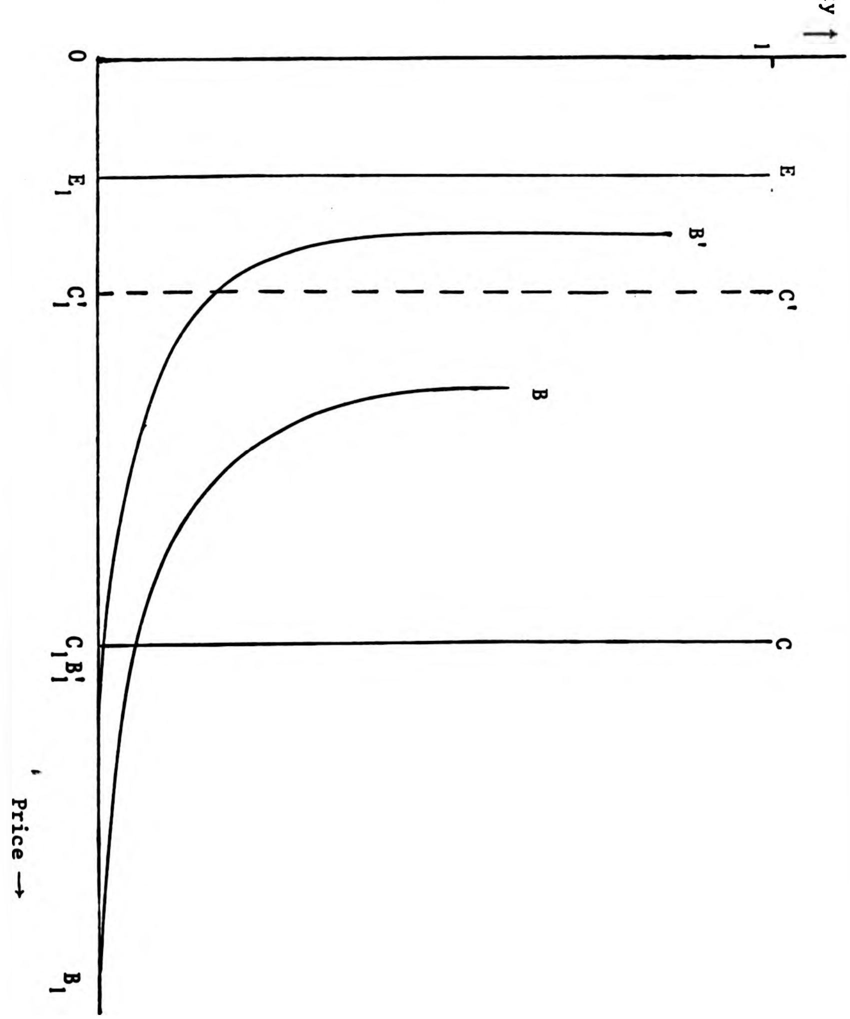
The efficient point comes from another aspect arising out of the game theoretic treatment. This is the "minimal surface or threat curve," the threat being that of flooding the market with a large quantity of output.[6]

Shubik's main original contribution which takes the second half of his book, is his work on games of economic survival. However, we leave discussion of this for a while as it is important to consider a

DIAGRAM 2.2

Shubik's Price and Quantity Variation Models

This diagram is a redrawing of Shubik's Figure 27 p. 126.



E E_1 is the efficient point and limiting strategy curve; firms acting as if with no market control.

C C_1 is the Cournot duopoly curve, C' C_1' the Cournot n firm strategy curve ($\infty > n > 2$)

B B_1 is the Bertrand (price-strategy) duopoly curve, B'B_1' the Bertrand n firm strategy curve (Bertrand solutions incorporate consideration of contingent demand).

concept central to all three works mentioned, that of the non-cooperative game and its equilibrium point. The solution comes from Nash (1950), and Shubik states that "An n-tuple S is an equilibrium point if and only if for every i:

$$G_i(S) \geq \max G_i\ (S;\ \varepsilon_i) \text{ for all } \varepsilon_i$$

Thus an equilibrium point is an n-tuple such that each player's strategy maximises his expected payoff if the strategies of the others are held fixed. Hence each player's strategy is optimal against those of the others" (p. 62, the notation has been changed so that it does not conflict with that used elsewhere in the chapter).

It has since been realised that the Cournot equilibrium is a special case of the above, with strategy being output (or the Cournot-type equilibrium if the strategy is price). Hence the emphasis on Cournot in Telser, for example. We shall discuss the idea in detail, and to facilitate this we take a very simple example:

Suppose a duopoly with demand conditions given by:

$$p = 200 - q_1 - q_2 \quad \text{(p is price, } q_1 \text{ and } q_2 \text{ outputs),}$$

and costs: $c_1 = 20q_1$, $c_2 = 20q_2$

For clarity we consider only three strategies for each firm, to produce the joint monopoly output (45 units each), the Cournot output (60 units each) or the output which sets price equal to marginal cost (90 units each). Given these outputs we may solve for profit in each case and so obtain a profit payoff matrix:

Firm B's action[7]

Firm A's action	Monopoly	Cournot	p = MC
Monopoly	(4050,4050)	(3375,4500)	(2025,4050)
Cournot	(4500,3375)	(3600,3600)	(1800,2700)
p = MC	(4050,2025)	(2700,1800)	(0,0)

Firm A's profit is put first in brackets, Firm B's profit second

We have to decide in what sense, and subject to what conditions, the Cournot output is an equilibrium point. We can usefully split the discussion into two - firstly stability from below, then from above. Consider then a firm allowed a choice between outputting either 90 units or 60 units. There is no conceivable reason why 60 should not be chosen as it assures him at least some profit and given that firm A chooses 60, firm B cannot do better than producing at most 60 units. In Nicholson's words the Cournot point is the one "point in the matrix which is stable without reference to any other point, for the values to A lower in the column are less ... and the values to B to the right of it are similarly less ... " (p. 196).

This important result has only recently achieved the recognition it deserves for (given that firms are agreed on the policy variable) it implies that the Cournot point, not the point where price is equal to marginal cost, yields a lower-bound to the set of reasonable outcomes.[8] (See also Cubbin (1973)).

Consider now the type of stability involved in attempting to achieve profits above the Cournot point. The choice of strategy here is more difficult and is known as "prisoner's dilemma". If the firms may collude, they would obviously agree on the monopoly output. This is the co-operative solution. However, they are unlikely to choose it independently. For if firm B were to choose the monopoly output, firm A could do better for himself by setting output at the Cournot level (and vice versa). If B then retaliates by setting Cournot output, firm A becomes worse off, the Cournot outcome is dominant.[9] Obviously there are circumstances in which we might expect, even without specific collusion, that the firms might reach (or approximate) the monopoly output, and others in which near-Cournot output would be the norm. All three authors spend some time exploring such situations.

For Nicholson the important factor is time, and in this context retaliation time and its effect on discounted profit flows. If firm B were utilising his monopoly strategy and firm A switched from this to his Cournot strategy, B would lose and A would gain. If B were able to retaliate quickly then it would be more likely that they would maintain their (seemingly) collusive behaviour than if B could perforce only retaliate slowly. Because Nicholson believes that (for example) an advertising policy is less flexible he considers that firms are more likely to price fairly collusively and be competitive with regard to their advertising policies.[10] Thus, it is impossible to say anything very coherent about the behaviour of oligopolies unless we can say something independently about the state of information in a market. Shubik comes to a very similar conclusion regarding the state of information and notes that in longer games "the threat of reprisal ... creates equilibria which do not exist in the subgames of finite duration" (p. 225). Telser examines this possibility extensively and states that "the Cournot-Nash theory retains its validity in a T-period model for finite T despite the fact that now implicit collusion between the two sellers is feasible" (p. 139), while "we see that if there is a sufficiently high probability of continuing then the two firms will collude in order to obtain the joint profit maximum. Otherwise despite the infinite horizon, they will find it to be more profitable to compete" (p. 145).[11]

So far it may well seem that the Game-Theorists have contributed little that is new to theories of oligopoly, an impression which should be dispelled. As noted earlier Shubik's main novelty is in his development of "games of economic survival" or in general in the field of what he calls "Mathematical Institutional economics". Essentially this comes about by limiting "duopoly models to situations in which firms (are) assumed unable to borrow outside money" (p. 256), that is by relaxing the assumption of "perfect" capital markets. This means that such factors as the differing kinds of management structure inherent in a capitalist or mixed economy bring about different solutions - the owner of an unlimited liability firm may well be expected to act differently to the owner of a limited liability company, or one which is management controlled. Also of importance, of course, is the asset condition of the various firms involved for "The financial dominance of one firm may be enough to entitle it to the lion's share of a peacefully divided market" (p. 212). In general " ... a two person game of economic survival can be completely characterised by the corporate assets X,Y, a discount rate ρ, the numbers A_1, A_2, representing the value of the market to a surviving firm at the time of exit of its competitor, and the market matrices ... " (p. 245).

In discussing the "solution" to such a game we should note that the outcome is not normally a single determinate point, and that the asset structure is of great importance in determining the likely pay-offs. If the asset structure is"not too" dissimilar, then it may be neglected but in general it will matter. While Shubik discusses in general terms the Tobacco and Automobile industries, the information requirements are tremendous and foreboding for any cross-sectional study. Perhaps it is as well to limit ourselves to the assumption of "perfect" capital markets or "similar" managements and asset structures as an approximation when considering empirical work.

Telser develops his theoretical treatment from the concept of the "core" of a game. Without dwelling overtly on the properties of the core, we may observe that "an outcome is in the core of a game if no subset (coalition) of players can collectively do better for its members and thus ' ... no-one can make himself better off by trade.'

When applied to market exchange this implies maximising behaviour analogous to that in traditional theory, but there are contrasts" (Clarke (1973) p.250 who is quoting Telser).

"The theory of the core ... forces a rigorous examination of several of the neglected aspects of oligopoly. For example with core theory it is necessary to prove in every case whether or not there will be group rationality (Pareto-optimality) and whether or not there will be price discrimination." (Telser (1972) p. 119). Thus it is a theory which defines competition before deducing its implications. The essence of Telser's approach then, is that it takes as given a much wider range of situations than traditional maximising theory, and can yield a much wider conception of the "solution" to the oligopoly problem.

Having said this, it remains true that the traditional theories, particularly the Cournot and collusion theories, crop up extensively and form the basis for much of the discussion. Indeed, one might be struck by the similarity between Telser's book and that of Fellner (1949). Each discusses the traditional theories within a wider canvas than is usual and ends up dwelling on the likelihood of collusion. Further, the empirical work performed by Telser, while painstakingly and thoughtfully done, could with minor exceptions have been equally performed by someone with little or no background in game theory. All this is not meant to imply that the theory is useless and the empirics mundane, but simply to provide a partial justification for neglecting an extensive treatment of game theory in the present work.

IV Stigler's "Theory of Oligopoly" (1964)

Viewed in the context of the above discussion of non-cooperative equilibria it can be seen that in many senses Stigler's theory of oligopoly contributes to our understanding of behaviour in much the same way as the work on game theory does, that is in setting the solution between limits based not on assumption but on deduction. Indeed, Telser at one point (especially p. 196) develops a rather similar model to Stigler's.

The essence of Stigler's approach is to posit a situation where overt collusion is illegal but profitable, thus "collusion takes the form of joint determination of outputs and prices by ostensibly independent firms ... " (p.45). As has been explained above, each partner to a collusion has an interest in cutting price (secretly) to earn more profit, given that the others maintain their collusive position. This will cause sales by the chiseller to rise; the question posed for the other n-1 firms who do not break ranks is at what point a transfer of sales to one firm becomes large enough to be suspicious. "We move then to the world of circumstantial evidence, or, as it is sometimes called, of probability", as Stigler puts it. (p. 48) As to the retaliation envisaged, or the possibilities when more than one firm "cheats", the theory is silent, yet it remains an interesting insight into oligopoly behaviour.

Three interesting predictions which follow directly from the setting of the model are that collusion will often be effective against small buyers, also against buyers (like the U.S. government) who report fully the prices quoted them, and is limited where the significant buyers constantly change identity. However, none of these are particularly important cases, so that the evidence on the extent of collusion must in general be obtained by considering in more detail the probabilities of detection involved, and it is with this area that the paper is mainly concerned.

Essentially the individual seller has three main areas in which he can look for information on chiselling: the behaviour of his own old customers, the rival's attraction of the old customers of other firms and the behaviour of new customers. It turns out that the first and third are the most important.[12] Within these, factors such as the number of sellers, the number of buyers, the probability of repeat purchase and the rate of entry of new buyers all have important effects on the maximum additional gains in custom without detection. For example, the larger the number of buyers, the less one firm may gain at the expense of others without detection so that the more likely is adherence to agreed pricing. One interesting result thrown up by the theory is that when firms are unequally sized the Herfindahl index of concentration becomes the relevant measure of inequality "if we wish concentration to measure likelihood of effective collusion" (p. 55).

Actually, as McKinnon (1966) points out, Stigler's exposition is in some senses statistically naive as he chooses fixed decision rules which are not easily interpreted with reference to each other. Rather the approach should consist in balancing the cost of unjustly accusing a rival of price-cutting against the losses from undetected price-cutting, and a statistically efficient method of pooling (sample) information on both the behaviour of old and of new customersshould be used, bearing in mind the costs of collecting such information. While McKinnon's alterations lead to minor differences in the conclusions (for example the relative efficacy of observing old rather than new customers), the basic findings of Stigler's novel approach remain untouched.

Turning to the place of Stigler's theory within the traditional framework, we see it as providing some illustrative factors affecting the position of price between collusion and Cournot. Foremost among these factors are the number of sellers and of buyers (buyers will be discussed in a later chapter), and empirical work performed by Stigler shows that not only the number of sellers but also the extent of market knowledge by buyers alters the actual price away from list price. His work should be seen not as a substitute for the traditional theories of oligopoly but rather as a complement adding meat to the bare bones of traditional theory's crude assumptions; a weighting scheme indicating the extent to which various postulates are likely to be more nearly true than others.

V Experimental Gaming in Oligopoly

The basic idea behind experimental oligopoly games is that, given the number of forces which affect firms' actions in the real world, and given the plethora of oligopoly theories, to design a test of a particular oligopoly theory in the real world would verge on the impossible. The experiment thus assists by modelling a simplified version of the real world in the laboratory, in order that pointers to the factors of relevance to manufacturers in making their decisions can be discovered. We shall briefly discuss some of these experiments in terms of the insight they offer: the work we consider is that by Dolbear et al. (1968), Fouraker and Siegel (1963), Friedman (1963), Hoggatt (1967)[13] and Murphy (1966).

There are several factors common to all work of this type:- the cost and demand functions take fairly simple algebraic forms, each firm has one decision variable, the number of firms in the market is usually very small, subjects play in "games" lasting for many bids or "periods", and no direct communication is allowed between players (see Friedman (1969)). Among the many factors differing between experiments, we shall focus on three; the number of players, the state of information and the decision variable used. Experimenters generally did most of their work with 2, 3 or 4 person games, though Dolbear et al. also has an experiment involving 16 players. We should note that in the Dolbear and Friedman experiments the structural effects of changes in the number of players were minimised by making the market size increase as the numbers increase. This has the effect of isolating behavioural characteristics for investigation (see Sherman (1971)).

The information given can be characterised either as complete (where payoffs to both parties given the respective decision variable values are specified) or incomplete (where payoffs only to the particular individual concerned are divulged). The decision variable used is either price or quantity; where price is the relevant variable the experimenter has specified differentiated products except in some experiments by Fouraker and Siegel and Murphy.[14]

Fouraker and Siegel, whose work is the most comprehensive and completely documented, perform two types of oligopoly experiment, using both information states. In their "quantity adjuster" models, their participants tend to arrive at Cournot contracts when they have incomplete information, although there are outcomes below as well as above this position and, compared with duopoly experiments, there tends to be more dispersion around the Cournot level in their triopoly tests. With complete information the contracts are, rather intriguingly, more diverse than in the incomplete information experiments. Their "price adjuster" models follow similar patterns, though this time of course the incomplete information contract tends to be at the Bertrand price.

In an interesting re-run of Fouraker and Siegel's price adjustment experiment with incomplete information, Murphy amends and extends the profit tables to allow bids below the Bertrand price level (which involve the players in lossess). He finds that "the change in the profit table only has greatly increased the amount of cooperation at the expense of competitiveness. The new results are more like those from (their) complete information experiment ... " He also notes that "as more trials were run, the tendency toward cooperative ruling prices became more pronounced" (both quotes p. 301).

Friedman's work is based on price adjustment with product differentiation, and the aim is to discover how much cooperation takes place in complete information models. In fact, however, he considers that his experiments are unable to yield any real evidence on the matter.[15] He does note though that "In comparing the F-S complete information games to Friedman's, one difference that stands out is the frequency of joint maximum games"[16] (p. 410). This, he feels, is a result of the experimental design whereby his participants play in several games each, so learning more about the technique of play. He finds that the more players there are, the less likely is joint profit maximisation to be reached.

Hoggatt's experiments are somewhat novel in that the participants were not paid and they compete in duopoly against a robot. Despite this element of unreality, the robot opponent permits greater control and his second experimental series brings the interesting result that the more co-operative the robot, the more co-operative the human player.

Dolbear et al. use a similar basic model to that of Friedman but include both states of information. They find that the equilibrium market price tends to be between the Cournot and joint profit maximisation positions while being inversely related to the number of firms. The other hypotheses they propose do not reach full statistical significance, though there is a presumption that full information raises profits and the dispersion of equilibrium profits. Thus "Information seems to induce bargaining attempts that tend to result in price war or collusion" (p.259).

Summarising the conclusions of the experimenters mentioned above, we find the range of outcomes tends to be between Cournot and joint profit maximisation and that while "the Cournot solution characterises behaviour in incomplete information situations" (Friedman 1969 p.414) with complete information there is a greater diversity of outcome, partly because a remarkably long time is needed to get the feel for the game and partly because this information state provides more opportunity for individual preferences to be revealed in play. Given this Sherman (1971) feels that the evidence of simple prisoner's dilemma experiments which attempt to control for such personality factors as sex, isolationism and risk attitude is relevant, although unfortunately the results here are not clear-cut. Thus, while in common with the Stiglerian prediction a tendency towards more cooperative outcomes, even when abstracting from structural factors, is apparent, we must expect individual attitudes to play a part.

VI Extensions to the Traditional Models:

From our brief review of Game Theory, Stigler's model of oligopoly and Gaming Experiments, we find that the simple model presented earlier in this chapter, while obviously not incorporating all factors deemed to be relevant, still provides a basis for discussion and straightforward empirical work. With this in mind we now turn to generalising the simple model of the first section in some fairly obvious directions.

Let us first consider the case[17] where not all firms are of equal size, specifically because their costs are different:

Profits are given by:

$$\Pi_i = pq_i - c_i(q_i) - F_i$$

so that the first-order condition for profit maximisation becomes:

$$\frac{d\Pi_i}{dq_i} = p + q_i \frac{dp}{dQ} \cdot \frac{dQ}{dq_i} - c'_i(q_i) = 0 \qquad (4)$$

We again assume the second-order condition holds.

From (4), multiplying by q_i and summing over the N firms yields:

$$\Sigma p q_i + \frac{dp}{dQ} \Sigma q_i^2 (1 + \lambda_i) - \Sigma q_i c'_i(q_i) = 0$$

so that $$\frac{\Sigma p q_i - \Sigma q_i c'_i(q_i)}{pQ} = - \frac{Q}{p}\frac{dp}{dQ} \Sigma \frac{q_i^2}{Q^2} (1 + u) \qquad (5)$$ [18]

where $u = \Sigma \lambda_i q_i^2 / \Sigma q_i^2$, a weighted sum of conjectural variation terms.

If we can now assume that marginal cost is equal to average variable cost then $\Sigma q_i c'_i(q_i)$ is equal to industry total variable cost. Given this, we may rewrite (5) as:

$$\frac{\Pi + F}{R} = \frac{H}{|\eta|} (1 + u) \qquad (6)$$ where $|\eta|$ is defined as before

and H is the herfindahl index of concentration.

We should now consider the consequences for our theory of the products not being perfectly homogenous. We first develop the case where output is again the firm's decision variable.[19] The profit function for the ith firm should now be written:

$$\Pi_i = p_i q_i - c_i(q_i) - F$$

where $p_i = f_i(q_1, q_2, \ldots, q_N)$

First order conditions for profit maximisation require that:

$$\frac{d\Pi_i}{dq_i} = p_i + q_i \frac{dp_i}{dq_i} - c'_i(q_i) = 0 \qquad (7) \qquad \text{for all } i$$

We shall assume that second-order conditions hold.

Multiplying (7) by q_i yields:

$$p_i q_i + \frac{q_i^2}{Q^2} \cdot \frac{dp_i}{dq_i} Q^2 - c'_i(q_i) q_i = 0$$

Now, in order to establish a link between the firm's demand function and the market demand function we have to take a pragmatic approach and talk about the market elasticity of demand for a homogenous product. To do this we rewrite the above equation as:

$$p_i q_i - q_i c'_i(q_i) = -\frac{q_i^2}{Q^2} \cdot \frac{dp_i}{dp} \cdot \frac{dp}{dQ} \cdot \frac{dQ}{dq_i} \qquad (8)$$ [19a]

where p is defined as: $\Sigma p_i q_i / \Sigma q_i \; = \; \Sigma p_i q_i / Q$

Summing (8) over i and dividing through by industry revenue gives:

$$\frac{\Sigma p_i q_i - \Sigma q_i c'_i(q_i)}{pQ} = -\frac{dp}{dQ} \cdot \frac{Q}{p} \sum \frac{q_i^2}{Q^2} \cdot \frac{dp_i}{dp} \cdot \frac{dQ}{dq_i}$$

$$\text{or} \quad \frac{\Pi + F}{R} = -\frac{dp}{dQ} \cdot \frac{Q}{p} \sum \frac{q_i^2}{Q^2} \cdot \frac{dp_i}{dp} \cdot \frac{dQ}{dq_i} \qquad (9)$$

Comparing this equation with the equivalents for the homogenous product case, (5) and (6), we see that the difference lies in the term $^{d}p_i/_{d}p$. Now the reciprocal of this differential is obviously less than unity, and in a limiting case is equal to the market share of the ith firm. This indicates that the price-cost margin in a differentiated product industry is greater than that in the homogenous case, all other things equal.[20]

Now, it could be objected that quantity is not the relevant decision variable in the heterogeneous product case, for firms are more concerned with setting prices than outputs. This would not be a valid objection in a homogeneous product industry, the case Cournot considered, because normally only one price may exist in such a market at any one time. Even in the situation where products are heterogeneous it is not necessarily true that price is the decision variable; as an empirical matter price, quantity or a mixture of the two may be most important in a given market. Having said this it makes the analysis more complete if we consider the results obtained when price is the decision variable.[21] To this end, let us discuss the fairly general model postulated by Cubbin (1974):

Given an industry of N firms selling a differentiated product, yet one which is sold at a common market price and produced at common costs, he finds that the ith firm's price-cost margin (equal to industry price-cost margin) is given by:

$$\frac{p_i - dC_i/dq_i}{p_i} = \frac{-1}{\alpha\eta^I + (1-\alpha)\eta_i^f} \qquad (10)$$

where p is price, dC_i/dq_i is marginal costs, α is a measure of apparent collusion and the industry and individual firm's elasticities of demand (η^I and η_i^f respectively) are related according to:

$$\sum_{j \neq i} \frac{\partial q_i}{\partial p_j} = \frac{q_i}{p_i}\left(\eta^I - \eta_i^f\right) \qquad (11)^{21a}$$

where q_i is, as before, the quantity produced by the ith firm.

Thus, when $\alpha = 0$, (10) becomes:

$$\frac{p_i - MC_i}{p_i} = \frac{-1}{\eta_i^f}$$

and with $\alpha = 1$ we have: $\dfrac{p_i - MC_i}{p_i} = \dfrac{-1}{\eta^I}$

The second of these is the case of collusion, while the first is the case equivalent to Cournot's. (It is not necessarily the Chamberlinian large numbers case, which requires in addition that the cost curve has a downward sloping segment, see Harrod (1967) p. 74). The relationship between these two special cases obviously depends in general on the size of $\Sigma\partial q_i/\partial p_j$, that is the degree of homogeneity.[22] One of the most important questions to be answered is obviously that of the role of numbers in the industry, for in contrast to the previous models, they do not appear explicitly.

Consider then, substituting (11) into (10) to yield:

$$\frac{p - MC_i}{p} = \frac{-1}{\eta^I - (1-\alpha)\,\frac{p_i}{q_i}\,\sum\limits_{j\neq i}\frac{\partial q_i}{\partial p_j}} \qquad (12)$$

Let us assume for simplicity that $\partial q_i/\partial p_j$ is the same for all j, so that we have from (12):

$$\frac{p_i - MC_i}{p_i} = \frac{-1}{\eta^I - (1-\alpha)\frac{p_i}{q_i}(N-1)\frac{\partial q_i}{\partial p_j}} = -\frac{1}{A}, \text{ say.}$$

Differentiating partially with respect to N:

$$\frac{\partial \frac{(p-MC)_i}{p}}{\partial N} = -A^{-2}(1-\alpha).\left[\frac{p_j}{q_i}\cdot\frac{\partial q_i}{\partial p_j} + (N-1)\frac{\partial\left(\frac{p_j}{q_i}\frac{\partial q_i}{\partial p_j}\right)}{\partial N}\right] \qquad (13)$$

Assuming for ease of interpretation (and as originally specified in the model) that $p_i = p_j$, and that α does not vary with N.[23]

We know that the first term inside the square brackets is positive and we should normally expect that the larger the number of firms (and so products) in the industry, the closer the individual firm's products become to each other, so that the last term should also be greater than (or in a limiting case equal to) zero. Thus an increase in the number of firms brings about a fall in the firm's and so the industry's price-cost margin.[24]

At a certain level of homogeneity we can achieve a result equivalent to the Cournot quantity solution; that is when $\alpha = 0$.

When $\eta_i^f = N\eta^I$

we have that: $\eta^I - \eta_i^f = (1-N)\eta^I$

Thus from (11): $(1-N)\eta^I = \frac{p_i}{q_i}\sum\limits_{j\neq i}\frac{\partial q_i}{\partial p_j}$ (14)

Again if we assume that $\frac{\partial q_i}{\partial p_j}$ is the same for all j (14) becomes:

$$(1-N)\eta^I = \frac{p_i}{q_i}(N-1)\frac{\partial q_i}{\partial p_j}$$

$$\therefore \frac{\partial q_i}{\partial p_j} = -\frac{1}{N}\frac{dQ}{dp_i} = -\frac{1}{N}\frac{dQ}{dp_j} = -\frac{dq_i}{dp_j} = -\frac{1}{N}\frac{\partial q_j}{\partial p_j} \qquad (15)$$

This means that when firm j raises his price without the others following suit, the fall in his output is N times as much as the rise in i's output consequent on j's action. Also that when j raises his price with the others doing likewise, the fall in his output and i's output is equal to the rise in i's output when j raises his price without the others following. While this is a delicate balance - that is if the others follow, their output will change by an equal amount but in the opposite direction to the effect on them if they don't follow, it by no means represents an obviously significant degree of homogeneity within the industry.

The question then arises as to the position of the barely-differentiated product industry. In this situation the firm's elasticity of demand will be very large (negatively). Thus the Cournot rule on price ($\alpha = 0$) which gives the Nash non-cooperative equilibrium will yield a very low margin to all firms in the industry. In contrast, if instead the firms were to use output as the competitive weapon then the non-cooperative equilibrium would quite conceivably yield a higher price.[25] In this situation, how are we to assess the likely outcome when apparent collusion is very low?

The most sensible solution for the firms involved would be to have an implicit agreement that output, rather than price, should be the competitive variable, so that they can at least assure themselves Cournot profit. At the same time they should try and increase differentiation of the product, to raise fallback profit. In fact it would seem that such an agreement would be fairly easy to enforce, as it requires minimal quasi-cooperation and to compete on output rather than price would appear a fairly conventional wisdom for making profits. This policy would also be likely considering that price variation with barely differentiated products would most probably be extremely unstable. In any case the game theoretic treatments suggest that strategies , such as price-cutting, to which others can quickly retaliate and which cause large changes in business are unlikely to be profitable in the long-term, and so are unlikely to be used. Much more likely is the strategy of gradually trying to build up and maintain custom through availability, salesmanship and product differentiation, in the process lowering cross elasticities of demand. The whole system of exclusive dealerships could be considered as an attempt by the manufacturer to make the relevant policy weapon output. Perhaps one of the prime examples is the "solus" system of petrol retailing in the UK,[26] price being used as a weapon only by new distributors in general. Having said this, the same industry has recently shown signs of price competition in the face of stagnant demand, although interestingly enough this competition is mainly by retailers not distributors. There also seems to be a feeling that this is "not in the long term interests of the industry".

To summarise, we would argue that in barely-differentiated product industries quantity competition seems to be a more logical weapon for powerful firms to sue, with price competition as an exceptional and drastic step.[27] Given this assumption our previous models can be taken as a reasonable basis for a testable theory of oligopoly.

VII Conclusion:

In conclusion, we have argued that a model of the form:

$$\frac{\Pi + F}{R} = \frac{p - AVC}{p} = \frac{H}{|\eta|}(1 + u) \qquad \text{(equation (6) repeated)}$$

comes fairly directly from simple models of oligopoly behaviour and fits reasonably well a wider class of oligopoly theories, assuming that we may take marginal cost as equal to average variable cost.[28] In addition we should note that Stigler's theory suggests that u=g(H) so that we may write:

$$\frac{\Pi + F}{R} = \frac{L(H)}{|\eta|} \qquad (16) \quad \text{where } 1 > L(H) > 0,\ L' > 0$$ [29]

This equation is relevant for a single industry; when it comes to comparing industries cross-sectionally we must realise that L(H) is likely to vary between industries due to a number of more or less measurable factors such as the ease of collusion and reprisal and the state of knowledge in that industry. We turn to testing procedure some chapters later after we have considered modifications of the theory due to the relaxation of some assumptions made here. However, perhaps it should be pointed out at this stage that our purpose in empirical work is not to test between theories of oligopoly but rather to consider whether the relatively simple models of oligopoly developed above yield reasonable results when applied fairly directly to the data.[30] In doing this we attempt some recognition of the fact that L(H) is industry specific by taking ratios of formula (16) at two time periods for each industry.

Pages 40 and 41 have been deleted

NOTES

1. These other factors could also be considered formally as quantity differences.
2. See, for example, Cowling and Waterson (1976).
3. See the discussion of Nicholson (1972), but see also Telser (1972) p.134, who considers that criticism of dynamic assumptions in Cournot is somewhat misplaced.
4. Hadar's model has a slightly different approach. See also Friedman (1968) for an extension to a more general type of behaviour where discounted profits are maximised and the firm assures the others' reaction function, not output, fixed.
5. Of course, this discussion assumes that arbitrage is not possible. Such action would tend to raise the average price paid by final buyers as arbitrageurs attempt to make money out of the imperfections in the market. The result that average price depends on the number of firms should still hold though, for the larger the number of suppliers the less the opportunity for arbitrage.
6. As such it is similar to and includes the perfectly competitive output point.
7. A rather similar example can be found in Nicholson p. 196.
8. We have taken a model in which output was the operational variable but this proposition can be seen to be true also in Scherer's model of price variation duopoly (with Cournot-type reaction) in a model with product differentiation (Scherer (1970) p. 133 fig 5.1). Construct a horizontal and a vertical from the "Cournot" equilibrium point in his diagram and notice that the square enclosed by these lines and the axes contains no value of profits more than that obtaining at the Cournot point for either firm.

comes fairly directly from simple models of oligopoly behaviour and fits reasonably well a wider class of oligopoly theories, assuming that we may take marginal cost as equal to average variable cost.[28] In addition we should note that Stigler's theory suggests that u=g(H) so that we may write:

$$\frac{\Pi + F}{R} = \frac{L(H)}{|\eta|} \qquad (16) \quad \text{where } 1 > L(H) > 0,\ L' > 0$$ [29]

This equation is relevant for a single industry; when it comes to comparing industries cross-sectionally we must realise that L(H) is likely to vary between industries due to a number of more or less measurable factors such as the ease of collusion and reprisal and the state of knowledge in that industry. We turn to testing procedure some chapters later after we have considered modifications of the theory due to the relaxation of some assumptions made here. However, perhaps it should be pointed out at this stage that our purpose in empirical work is not to test between theories of oligopoly but rather to consider whether the relatively simple models of oligopoly developed above yield reasonable results when applied fairly directly to the data.[30] In doing this we attempt some recognition of the fact that L(H) is industry specific by taking ratios of formula (16) at two time periods for each industry.

Pages 40 and 41 have been deleted

NOTES

1. These other factors could also be considered formally as quantity differences.
2. See, for example, Cowling and Waterson (1976).
3. See the discussion of Nicholson (1972), but see also Telser (1972) p.134, who considers that criticism of dynamic assumptions in Cournot is somewhat misplaced.
4. Hadar's model has a slightly different approach. See also Friedman (1968) for an extension to a more general type of behaviour where discounted profits are maximised and the firm assures the others' reaction function, not output, fixed.
5. Of course, this discussion assumes that arbitrage is not possible. Such action would tend to raise the average price paid by final buyers as arbitrageurs attempt to make money out of the imperfections in the market. The result that average price depends on the number of firms should still hold though, for the larger the number of suppliers the less the opportunity for arbitrage.
6. As such it is similar to and includes the perfectly competitive output point.
7. A rather similar example can be found in Nicholson p. 196.
8. We have taken a model in which output was the operational variable but this proposition can be seen to be true also in Scherer's model of price variation duopoly (with Cournot-type reaction) in a model with product differentiation (Scherer (1970) p. 133 fig 5.1). Construct a horizontal and a vertical from the "Cournot" equilibrium point in his diagram and notice that the square enclosed by these lines and the axes contains no value of profits more than that obtaining at the Cournot point for either firm.

9. In Shubik's proposed terminology, the difference, in an n-person game, between the three situations is as follows: "A market equilibrium is defined to be $\{k_i\} - \{r_j\}$ stable ... if no joint action by the r players yields them more than maintaining the steady state, on the assumption that each of the k players is committed to a threat strategy, and that each of the n-k-r players is playing a steady-state strategy" (p.276-7) r and k are specific subsets of the n players. Thus the Cournot equilibrium is $\{0\} - \{1\}$ stable, in that it needs no policing, the absolutely competitive equilibrium is $\{0\} - \{n\}$ stable, while the collusive equilibrium is $\{n-1\} - \{1\}$ stable; it needs all to police and carry out the threat on the chiseller(s).

10. It should perhaps be mentioned that "competition" to the game-theorist is not the same as perfect, pure, absolute or maximum competition but is nearer to being a synonym for non-cooperation.

11. This discussion of Telser's appears to have something in common with Friedman's (1971) concept of non-cooperative equilibria in "supergames", which consist of an infinite number of ordinary games. Here Pareto optimal pay-off vectors are among the equilibria, in contrast to the ordinary non-cooperative games.

12. A price-cutter is harder to detect by attraction of old customers of other firms on Stigler's criteria. The individual seller also has the possibility of pooling information with the other firms with the attendant risk of pooling with the price-cutter.

13. We do not consider his previous paper.

14. We shall say more about price as a decision variable in the next section; with identical products price is not the natural weapon.

15. In his second article, Friedman (1969) includes a remarkably impersonal review of his 1963 paper.

16. He considers that Fouraker and Siegel misclassify outcomes by making the wrong assumptions about dynamic adjustments; the effect of this reclassification (which does not alter the point stated) is to place more of their firms in the Cournot outcome category.

17. Developed also in Cowling and Waterson (1976).

18. See Rader (1972) pp. 271-2. I am grateful to D. Morris for pointing this out.

19. This model was originally intended to appear as an appendix to Cowling and Waterson (1976), but was omitted for reasons of space.

19a. In this equation dQ/dq_i should again be considered as firm i's conjecture about the output reactions of (him and) the other firms when he changes his output. dp_i/dp is of course not a proper derivative in the sense that p is not an independent variable; hence our comments about a "pragmatic approach". It is probably best considered as representing a certain degree of heterogeneity in the industry, indicating the extent to which prices have to move together.

20. We should note that the definition of the Herfindahl in terms of quantities in the differentiated product case is not strictly accurate unless all prices are identical.

21. Shubik points out that to specify price necessarily implies a willingness either to produce a large enough output to satisfy the market or to specify in addition a production rate. Thus he develops the concept of "contingent demand" for price variation models and quantity variation models with product differentiation, on which see our earlier comments. Cubbin's model ignores such complications by assuming adequate inventories, a great simplification to the analysis.

21a. This equation represents the actual relation between industry and firm elasticities, dependent upon the degree of heterogeneity. Firm i's <u>conjecture</u> about the relationship is bound up purely in α, which is a weighted average of the individual firms' perceptions about the values of dp_j/dp_i (see Cubbin, pp. 1-3).

22. See Triffen (1949) p. 104 and elsewhere for explanations of the use of this concept.

23. Any variations of α with N would on usual assumptions strengthen our result.

24. Allowing for "contingent demand" considerations would similarly strengthen the conclusions.

25. This of course depends upon the degree of heterogeneity, as is evident from (15).

26. See Shaw (1974) for a discussion of this industry.

27. The formula developed by Weitzman (1974 p.484), though based on a different situation, would tend to reinforce these arguments. If a firm has <u>some</u> uncertainty about marginal costs in the area of maximum profit and the marginal revenue function is of steeper (negative) slope than the marginal cost functions (positive) slope, as would be the case under our assumptions (see Appendix I), his formula indicates that price is an inferior instrument to quantity. I am grateful to Avinash Dixit for suggesting the reference.

28. An assumption we should have in any case to make for empirical test. See also Appendix I.

29. We expect that the slope of L(H) is initially greater than one, but slackens off so that $L'' < 0$. This is dealt with in more detail later. Those oligopoly gaming experiments which attempt to abstract from structural effects of numbers in the industry suggest a similar type of result to that given by Stigler's theory.

30. This is to some extent novel for Joskow (1975, p. 273) has argued that "Virtually none (of the previous empirical work in this area) appeals to particular formal models of oligopoly markets".

Chapter 3: A Consideration of the Problem of Potential Entry

I Introduction:

In the previous chapter we made the assumption that the number of firms in the industry was fixed, that is no entry or exit was allowed. Our purpose in the present chapter is to explore the implications of relaxing this assumption. Consider initially then, allowing perfectly free entry to and exit from an industry, so moving to a general equilibrium framework.

Specifically, let us take the case where one industry is monopolised and all others contain a very large number of firms, so that they can be considered as perfectly competitive industries and the marginal firm in each earns zero or "normal" profits. The monopolist earns above-normal profits so that there is an incentive for firms from other industries to enter his industry until returns to the marginal firm in that industry are zero in common with the rest of the economy. Thus, if the monopolist wishes to retain his position as sole supplier of his particular product under perfectly free entry he also must price so as to obtain only normal profits.

Further, Fama and Laffer (1972) have shown that even if there are only two firms in each industry, and perfectly free entry and exit for each industry, then each firm is in effect a perfect competitor in that he earns solely normal profits and his output decisions have no effect on price. They illustrate this by supposing that one firm expands output. The other must then contract output to <u>exactly the same extent</u>[1] as this expansion for otherwise returns in the industry are below normal and exit occurs (after which presumably the survivor produces output sufficient to ensure normal returns). Similarly, if one firm were to contract output, the other must expand to make up the deficit or entry will occur. Thus the actual, as opposed to the conjectural, variation is such that output decisions do not affect price.

Now in the case where the possibility of entry or exit does not exist the optimal reaction will be different from that which is required to prevent entry or exit in the previous case. For example if one firm contracts output from the Cournot level to the monopoly level then the optimal reaction for the other firm is not to expand output but rather to contract output to that same level, as we saw earlier; each firm has only to consider the others in his industry rather than the whole economy. Thus we may say that a necessary (but not sufficient) condition for the number of firms in an industry to affect the level of profits in that industry is that entry is not completely free, costless and quick. In the remainder of this chapter we consider cases where entry conditions affect the established firms decision processes to a greater or lesser extent. Exit conditions will receive a more perfunctory treatment.

Those factors which prevent entry being perfectly free are known as Barriers to Entry and Bain (1962) has discussed the types of barriers which may exist in an industry in some detail. Basically he considers that there are three major categories of barriers, which are: absolute cost advantages, product differentiation advantages and economies of scale. We leave further consideration of the different types of barriers to entry until we have pursued some theoretical development.

However the way that Bain has defined and detailed barriers to entry does give rise to one problem which we ought to discuss before proceeding further. This is the problem of the time interval we are considering. It is fairly obviously not the theoretical short-run, for in that case there would not be any possibility of entry. It cannot easily be the theoretical long-run either, for that would mean no firm having, for example, any product differentiation advantage over another, and all firms both potential and actual would have access to capital on the same terms. Nevertheless, Bain feels that such structural features as entry barriers are not ephemeral short-lived advantages, which leads Williamson (1963) to consider that we are talking about an "intermediate-run" situation (see, for example, p. 113, n. 6). This may in fact cause difficulties for the theory.

Bain then defines the "condition of Entry" to be the extent to which the established firms may raise price above costs without inducing entry, and categorises four general areas in which the barriers to entry may place an industry:

(i) Blockaded entry, where barriers are such that established firms could price even at the monopoly level yet still not incur entry,

(ii) Easy entry, where barriers are so small that pricing even very slightly above costs allows entry,

(iii) Ineffectively impeded entry, where pricing at a level at which no entry will occur is less profitable than maximising short-run profits and allowing entry,

(iv) Effectively impeded entry, pricing at the level at which no entry will occur is more profitable than maximising short-run profit and allowing entry.

The level of price, below which no firm, even the most favourably placed, will find it profitable to enter is called the "Limit

Price." We now consider its formation, or the "Theory of Limit Pricing".[2]

II The Formation of Limit Price - Modigliani and Osborne:

In formulating the limit price, we straightaway run into consideration of the expected reaction by the established firms to a potential entrant. In fact, the problem from the established firms' point of view is that of deciding what the entrant thinks they will do regarding their output if he decides to enter. This situation of potential indeterminacy is normally solved by making a particularly useful and straightforward assumption known as "the Sylos' postulate".[3] The entrant assumes that the established firms will maintain their output in the face of potential entry, and the established firms know this to be the case. Making this assumption fixes the portion of the industry demand curve along which the entrant may operate. A further common assumption made for simplicity is that the established industry consists of a monopolist or a tightly knit group of collusive oligopolists. We initially consider models utilising both these assumptions.

The classic article in this area is of course Modigliani's (1958); his arguments may be briefly summarised by discussion of the comparative statics of his simple case, while we leave fuller consideration until later.

If both the established and potential entrant firms have cost curves which are sharply discontinuous at minimum optimal scale, so that outputs lower than that are infinitely costly to produce, then we may say that the entry limiting output is given by $q_o = q_c - \bar{q}$, where q_c is "competitive" output,[4] and $\bar{q}$ is the minimum optimal scale. Defining the size of the market as $s = q_c/\bar{q}$ then we may write:

$$q_o = q_c \left(1 - \frac{1}{s}\right) \qquad (1)$$

We want to say something about limit price in relation to costs, thus we have to transform the above equation into one in terms of prices. This in general will involve knowing the form of the demand curve, in order to solve the equation:

$$p_L = f(q_o) = f\left[q_c\left(1 - \frac{1}{s}\right)\right] \text{ explicitly } (p_L \text{ is limit price}).$$

To take a particularly simple example, where the demand curve is of linear form, $p = \alpha - \beta q$ then

$$q_o = \frac{\alpha - p_L}{\beta}, \quad q_c = \frac{\alpha - p_c}{\beta} \quad \text{where } p_c \text{ is "competitive" price.}$$ [5]

Thus from (1):

$$p_L = \alpha - (\alpha - p_c)\left(1 - \frac{1}{s}\right)$$

$$= p_c\left(1 + \frac{\alpha}{p_c s} - \frac{1}{s}\right)$$

$$= p_c\left(1 + \frac{\beta q_c}{p_c s}\right)$$

$$\therefore p_L = p_c\left(1 + \frac{1}{|n|_c s}\right) \quad (2), \quad \text{where } |n|_c = \frac{p_c}{q_c}\left|\frac{dq_c}{dp_c}\right|,$$

the modulus of the industry elasticity of demand at q_c.

Other forms for the demand curve give different results; the constant elasticity case mentioned by Modigliani involves other terms, since 1 and $\frac{1}{|n|s}$ are only the first two terms in the expansion of

$$\left(1 - \frac{1}{s}\right)^{-1/|n|}.$$

He therefore writes (2) as an approximate equality. Following from this, it is easy to show that:

$$\left(\frac{\Pi}{R}\right)_L \equiv \frac{p_L - p_c}{p_L} = \frac{1}{|\eta|.s+1} \qquad (3)$$

This of course implies that:

$$\frac{\partial(\frac{\Pi}{R})_L}{\partial|\eta|} < 0, \quad \frac{\partial(\frac{\Pi}{R})_L}{\partial s} < 0$$

and if we define the importance of economies of scale[6] as $\bar{q}/q_c$ or $1/s$ then

$$\frac{\partial(\frac{\Pi}{R})_L}{\partial(\bar{q}/q_c)} > 0$$

In the words of his famous quote:

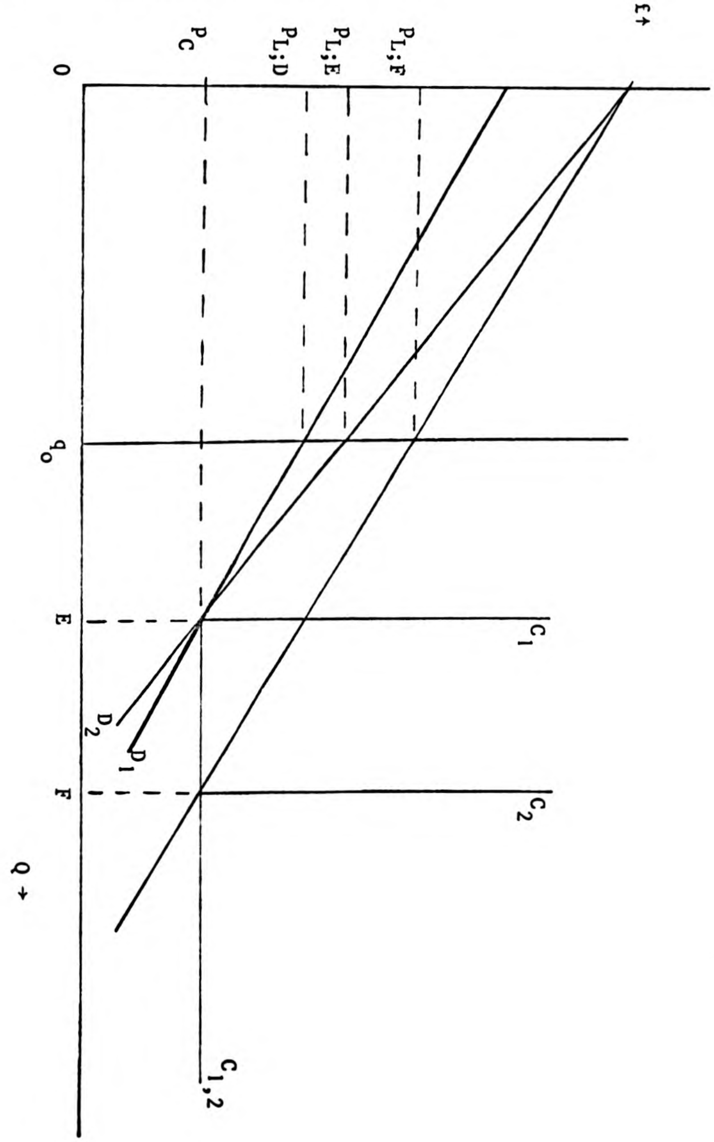
"In summary, under Sylos' postulate there is a well-defined maximum premium that the oligopolists can command over the competitive price, and this premium tends to increase with the importance of economies of scale and to decrease with the size of the market and the elasticity of demand." (p.220).

It might be illuminating to illustrate these comparative static results diagrammatically (see diagram 1 overleaf). The first is indicated by the two demand curves, $D_{1,2}$ at point (p_c, E). The steeper demand curve is the less elastic and gives rise to a higher limit price $p_{L;E}$, scale constant. The second (and third) are illustrated by demand curves $D_{2,3}$ at points $(p_{c,} E)$ and $(p_{c,} F)$ respectively. They have equal elasticities at these points by construction. In the former case size of the market s is given by $\frac{OE}{q_o E}$ and in the latter by $\frac{OF}{q_o F}$; thus s is larger in the first case but gives rise to a lower limit price $p_{L;E}$.

Now, Modigliani's more complicated cost curve structures give rise to similar qualitative results; essentially we now allow that

DIAGRAM 3.1

Modigliani comparative static results



$C_{1,2}$ are two different cost curve assumptions

$D_{1,2,3}$ are three different demand curve assumptions

$P_{L;D,E,F}$ are three different limit prices to which correspond total outputs

OE, OE, OF and limit outputs q_oE, q_oE, q_oF.

the cost curve may slope less sharply to optimal scale. However, given his definition of the size of the market (and optimal scale), he is only able to consider a certain class of cost curves. Specifically, there must be a discontinuity in the curve at optimal scale for his definition to remain meaningful. Also, Modigliani does not consider cases where the cost curves differ between firms. In this sense his model is applicable to the true long-run, yet intermediate-run factors may be relevant to particular industries' pricing policies. Lastly, it is not easy to generalise his model to an oligopoly situation.

Osborne's (1973) model of limit pricing similarly suffers from lack of generality, though of a rather different nature. This is what gives rise to his quite different results regarding comparative statics. Osborne considers that we have a Stackleberg situation. The firm proposing to enter the industry in question believes that the existing firm will honour the Sylos postulate; that is, that output of the existing firm will remain at pre-entry level. Further, the established firm knows that the entrant will act in his best interest under these assumptions. This means that the existing firm may take advantage of this and act as a leader, the entrant being a follower, in the Stackleberg system. We take the second firm's (entrant's) profits to be:

$$\Pi_2 = q_2 p - c_2(q_2); \qquad p = f(q_1 + q_2)$$

These he maximises, assuming q_1 constant:

$$\frac{\partial \Pi_2}{\partial q_2} = q_2 \frac{dp}{dq} + p - c_2'(q_2) = 0 \qquad (4)$$

Limit price, Osborne says, is achieved where the entrant's optimal output is zero. Substituting this into the first-order condition for optimality ((4) above):

$$p_L = c_2'(0)$$

This means that limit price is set at the level at which the entrant's marginal costs will be at zero output, so that the existing firm's limit profit revenue ratio may be written as:

$$\left(\frac{\Pi + F}{R}\right)_L \equiv \frac{p_L - c_1'}{p_L} = \frac{c_2'(0) - c_1'(q_1)}{c_2'(0)},$$

(assuming marginal cost = average variable cost for firm 1).

Notice that in this equation we have no elasticity of demand term and also that "It is the behaviour of marginal cost, and its rate of change, in the neighbourhood of the origin that, given demand, governs the rationality of limit pricing. The behaviour of average cost ... is germane only to the extent that it is implied by the marginal quantities." (p.78). Osborne does allow the case where the entrant has a different (possibly higher) cost curve than the established firm, a case Modigliani does not discuss.

The main problem with this analysis is revealed by a consideration of the second order condition associated with (4). We have:

$$\frac{\partial^2 \Pi_2}{\partial q_2^2} = q_2 \frac{d^2 p}{dq^2} + 2\frac{dp}{dq} - c_2''(q_2) < 0$$

i.e. $\dfrac{dMR_2}{dq_2} < \dfrac{dMC_2}{dq_2}$

or, where $q_2 = 0$:

$$2\left.\frac{dp}{dq}\right|_{q_2 = 0} < c_2''(0) \qquad (5)$$

- the slope of the marginal cost curve at zero output must be greater (or less negative) than twice the slope of the demand curve. In fact, if we define average variable costs as AVC = $B(q_2)$,

then Total variable costs = $B(q_2).q_2$,

$$\text{Marginal cost} = B(q_2) + q_2B'(q_2) \ .$$

$$\frac{dAVC}{dq_2} = B'(q_2) \ ; \quad \left.\frac{dAVC}{dq_2}\right|_{q_2=0} = B'(0) \ ,$$

$$\frac{dMC}{dq_2} = 2B'(q_2) + q_2B''(q_2); \quad \left.\frac{dMC}{dq_2}\right|_{q_2=0} = 2B'(0) \ .$$

Another way that we may state (5) then, is as:

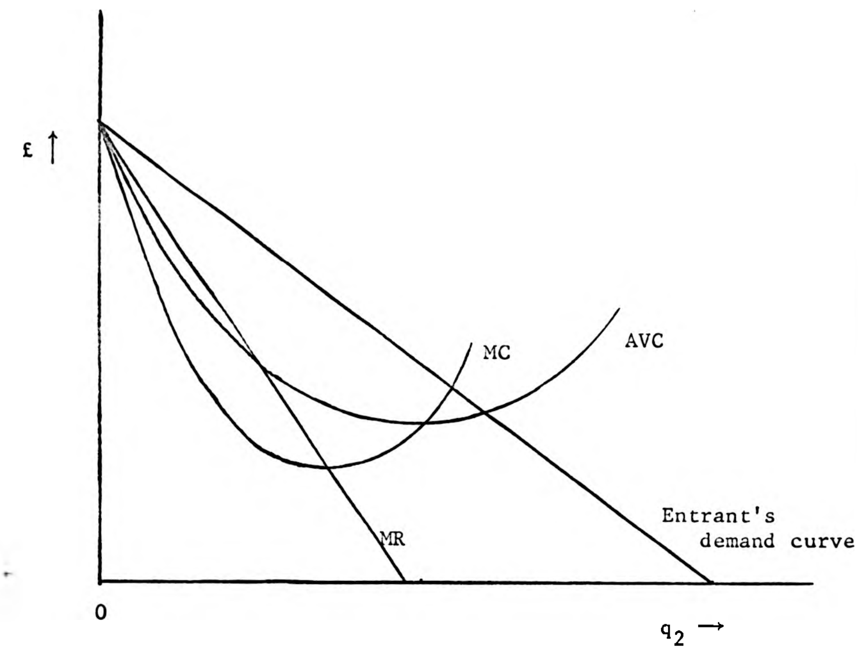
$$\left.\frac{dp}{dq}\right|_{q_2=0} < \left.\frac{dAVC}{dq_2}\right|_{q_2=0}$$

Essentially, this rules out cases where marginal and average variable cost curves bow steeply enough to fall below marginal and average revenue curves respectively, for example in the diagram overleaf (diagram 2), the first order condition is satisfied, yet entry might be profitable. Two points occur here:

(i) Although entry may be profitable in the case immediately above, a firm would surely not enter if fixed costs were such that he were unable to cover them. While, for the existing firm in an industry, fixed costs are irrelevant as long as variable costs are being recovered; for the entrant, looking around among industries for possible entry, fixed costs in a single industry become a decision variable - they are avoidable. That is, given a set of industries each with identical demand and marginal cost conditions, the entrant would

DIAGRAM 3.2

A case not covered by Osborne's second-order condition



MC and AVC are the entrant's cost curves

Osborne's limit price is at the origin of all curves, yet the entrant can make no profits at a price substantially higher than that.

choose that industry where net profit was highest (some presumably yielding a positive return). This is not taken into account by Osborne and perhaps indicates why others have not, unlike Osborne, considered marginal costs more relevant than average costs.

(ii) We have to reformulate the problem so that the second order conditions are not as restrictive as they appear to be in his case. The actual extent of restriction imposed by his model (as with those restrictions imposed by Modigliani), is of course an empirical matter.

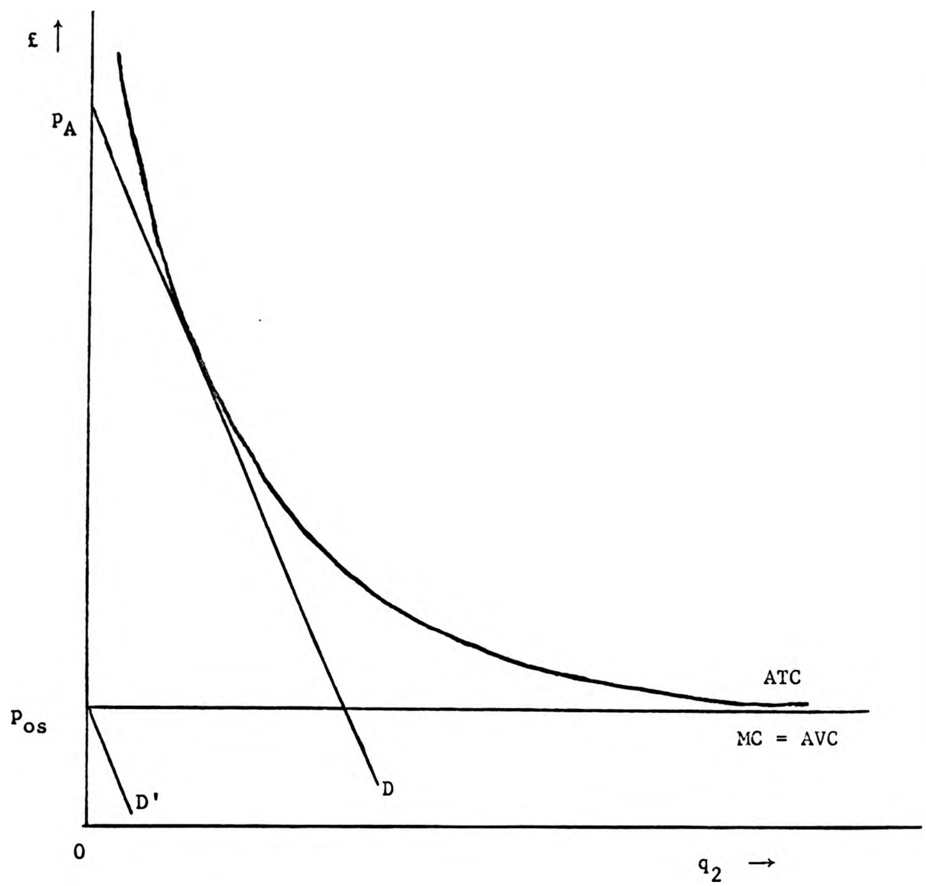
Specifically however, and as a counter-example to diagram 2, if we do include fixed costs in the analysis then in an industry with constant marginal and average variable costs the same for both firms, the Osborne limit price would be unnecessarily low. In diagram 3, drawn under these conditions, p_A is the actual price, given the demand curve drawn, which would halt entry and is much above the Osborne limit price.

It should be noticed that this case of constant marginal costs with fixed costs cannot be dealt with by the Modigliani approach either.

What the Osborne method in fact does is to combine two operations into one, in order to simplify the problem in hand. For an entrant to be on the margin of entry, we require that his maximum profit (defined as above to include fixed costs) should be zero. This need not necessarily occur at zero output, and in fact in general will not do so. Having found the price-output combination at which maximum profit is zero, we now (conceptually) have to move back along the demand curve to solve for the price at which $q_2 = 0$. This is the limit price. Since Osborne considers only those cases where maximum profit is zero at zero output, then his method is limited. At the same time, we shall want to obtain comparative static predictions in a form

DIAGRAM 3.3.

A case not covered by either Osborne or Modigliani



p_A is the actual limit price, and D the positioning of the entrant's demand function under that case. p_{os} is the limit price, and D' the position of the entrant's demand curve under Osborne's assumptions. MC, AVC and ATC are the entrant's marginal, average variable and average total cost functions respectively.

suitable for estimation purposes, and where the definitions are meaningful, unlike the Modigliani examples.

The problem may be stated fairly easily, and is solved using parts of the analysis of both Osborne and Modigliani. The potential entrant's problem is:

Maximise: $\Pi_2 = pq_2 - c_2(q_2) - F_2; \quad p = f(q_1 + q_2)$

This yields the first order condition:

$$(g =) \quad p + q_2 \frac{dp}{dQ} - c_2'(q_2) = 0^7 \qquad (6)$$

Equation (6) defines firm 2's reaction function in implicit form, i.e. $q_2 = \phi(q_1)$.

Meanwhile, the established firm's problem is:

Maximise: $\Pi_1 = pq_1 - c_1(q_1) - F_1; \quad p = f(q_1 + \phi(q_1))$

subject to: $\Pi_2 = \Pi_2(q_1) \leq 0.$

$q_1 \geq 0$

Assuming $q_i \neq 0$, the constrained maximisation problem yields the Kuhn-Tucker conditions:

$$\left.\begin{aligned} \frac{d\Pi_1(q_1)}{dq_1} - \lambda \frac{d\Pi_2(q_1)}{dq_1} &= 0 \\ \Pi_2(q_1) &\leq 0 \\ \lambda\left[-\Pi_2(q_1)\right] &= 0 \\ \lambda &\geq 0 \end{aligned}\right\}$$

(λ is the Lagrangean multiplier for the problem).

Now, if $\lambda > 0$:

$$\Pi_2 = pq_2 - c_2(q_2) - F_2 = 0 \qquad (7)$$

and $\dfrac{d\Pi_1(q_i)}{dq_1} - \lambda \dfrac{d\Pi_2(q_i)}{dq_1} = 0$

Alternatively, if $\lambda = 0$, we have:

$$\Pi_2 = pq_2 - c_2(q_2) - F_2 \leqslant 0 \tag{7a}$$

and $$\frac{d\Pi_1(q_1)}{dq_1} = p + q_1\frac{dp}{dQ} + q_1\frac{dp}{dQ}\cdot\frac{dq_2}{dq_1} - c_1'(q_1) = 0$$

i.e.: $$p + q_1\frac{dp}{dQ}(1 + \phi') - c_1'(q_1) = 0 \tag{8}$$

Here $\phi' = \dfrac{dq_2}{dq_1} = \dfrac{-\partial g/\partial q_1}{\partial g/\partial q_2}$ from (6),[8]

thus: $$\phi' = \frac{-\left(\frac{dp}{dQ} + q_2\frac{d^2p}{dQ^2}\right)}{q_2\frac{d^2p}{dQ_2} + 2\frac{dp}{dQ} - c_2''(q_2)} \tag{9}$$

In the case where $\lambda > 0$, pricing according to the Sylos postulate would fail to prevent entry. The established firm, if he wishes to prevent entry, is therefore bound to set a lower price and take output beyond the profit-maximising point.[9] The output chosen will satisfy the conditions that the potential entrant, maximising perceived profits, finds that it is just not worthwhile his entering in that he makes at most zero profits on entry. Here then equations (6) and (7) can in principle be solved simultaneously by the established firm for q_1, q_2 and so p. In order for this to be done explicitly here, we would need to know the functional forms of the demand curve and both cost curves. We shall not want to assume all these in general, though we see that without assuming a form for the demand curve the analysis can proceed very little further.

Following Modigliani for the moment, we may write:

$$q_{1L} = Q - q_{2M}$$

where q_{2M} is the output which maximises firm 2's profits and q_{1L} the limiting output for firm 1.

Thus:

$$q_{1L} = Q\,(1 - \frac{q_{2M}}{Q})$$
$$= Q\,(1 - \frac{1}{s})$$

However this is not very useful since s is no longer a constant. As stated earlier, our goal is to solve for that price which the established firm may set without incurring entry. Using solely the information available above we are unable to derive such a price in general, since our equations are in terms of output.[10] For concreteness we take the constant elasticity of demand case so that:

$$q_{1L} = kp_L^{-|\eta|} \quad \text{and} \quad Q = kp^{-|\eta|}$$

Thus $$p_L = p\,(1 - \frac{1}{s})^{-1/|\eta|} = p\,(\frac{q_{1L}}{Q})^{-1/|\eta|} \qquad (10)$$

From (7): $$p = \frac{c_2(q_2)}{q_2} + \frac{F_2}{q_2}$$

which we may write as:

$$p = AVC_2 + AFC_2 \qquad (11)$$

where AVC_2 is the average variable cost of the second firm at output q_2 and AFC_2 the average fixed cost for firm 2 at that output.

Now $$(\frac{\Pi + F}{R})_L \equiv \frac{p_L - AVC_1}{p_L} = 1 - \frac{AVC_1}{p_L}$$

(AVC_1 being defined analogously at output q_1),

so that from (10) and (11):

$$(\frac{\Pi + F}{R})_L = 1 - \frac{AVC_1}{AVC_2 + AFC_2}\left[\frac{q_{1L}}{Q}\right]^{1/|\eta|} \qquad (12)$$

It is not particularly enlightening to attempt to solve this equation any further. However, we might note that, to a first-order approximation, it may be written:

$$(\frac{\Pi + F}{R})_L \simeq 1 - \frac{AVC_1}{AVC_2 + AFC_2}\;(\frac{MC_2}{AVC_2 + AFC_2}) \qquad (12a),$$

MC_2 being the second firm's marginal cost at q_2 (= q_{2M}).[11]

We should also consider the situation obtaining where $\lambda = 0$, that is pricing with regard to the entrant (i.e. according to the Stackleberg postulate) prevents entry. As such, Osborne (1973, p. 74) calls it the situation where "the theory will be logically consistent". It should be distinguished from the position of blockaded entry, for that is where setting profit-maximising price, <u>without</u> regard for the entrant, fails to make entry attractive.

Of particular interest here is the possibility (since we have not specified ccst functions) of defining a point at which Osborne's "consistency condition" is just satisfied, in that pricing according to the Stackleberg postulate makes it just not worthwhile to enter. Inequality (7a) then takes on the form (7), and (6), (7) and (8) are all true at a single point. This point represents the highest possible price (and lowest possible output) at which entry may be prevented by using the Stackleberg policy; or, to look at it another way, the minimum level at which entrants' costs have to be in order that consistent following of the Sylos postulate will prevent entry. To solve partially for this limiting consistent point, we write from (6), (7) and (8):

$$q_{1LC} = \frac{AVC_2 + AFC_2 - MC_1}{\left|\frac{dp}{dQ}\right|(1 + \phi')} ,$$

MC_1 being the established firm's marginal costs at q_{1LC}, the limiting consistent output.

$$q_{2M} = \frac{AVC_2 + AFC_2 - MC_2}{\left|\frac{dp}{dQ}\right|} .$$

This means that: $$\frac{q_{1LC}}{Q} = \frac{AVC_2 + AFC_2 - MC_1}{AVC_2 + AFC_2 - MC_1 + (AVC_2 + AFC_2 - MC_2)(1 + \phi')}$$

and we have, as an alternative to (12), the limiting consistent margin:

$$\left(\frac{\Pi+F}{R}\right)_{LC} = 1 - \frac{AVC_1}{AVC_2 + AFC_2}\left[\frac{AVC_2 + AFC_2 - MC_1}{AVC_2 + AFC_2 - MC_1 + (AVC_2 + AFC_2 - MC_2)(1+\varphi')}\right]^{1/|n|} \quad (13)$$

Equations (12) and (13) can be considered as more general formulations than the models of Modigliani and Osborne, though they do not appear to be in a form particularly useful for prediction.

We can in fact verify both Osborne's and Modigliani's results as special cases of (12). In the case of Osborne we had:

$$p_L = c_2'(0),$$

which was a special case among the Kuhn-Tucker conditions, and that:

$$\left(\frac{\Pi + F}{R}\right)_L = \frac{p_L - AVC_1}{p_L} \qquad \left[= \frac{p_L - c_1'(q_1)}{p_L} \quad \text{if } MC_1 = AVC_1\right]$$

Recall that for the entrant's optimal output of zero to yield the limit price then $\Pi_2 = 0$ at $q_2 = 0$. This means of course that the entrant has no fixed costs. But in this case $c_2'(0) = AVC_2(0)$, and we may rewrite the formula for the profit-revenue ratio as:

$$\left(\frac{\Pi + F}{R}\right)_L = 1 - \frac{AVC_1}{AVC_2(0)}$$

This is precisely the formula obtained from (12) when $q_{2M} = 0$, with $AFC_2 = 0$.

Modigliani simplifies the problem (when dealing with his more complicated cost structures) by assuming that both firms have the same cost curves, there are no fixed costs and (at least in his diagrams) that the established firm has always attained his lowest unit cost position. We have from (12)

$$\left(\frac{\Pi + F}{R}\right)_L = \left(\frac{\Pi}{R}\right)_L = 1 - \frac{AVC_1}{AVC_2}\left[\frac{q_{1L}}{Q}\right]^{1/|n|}$$

where AVC_1 is a constant.

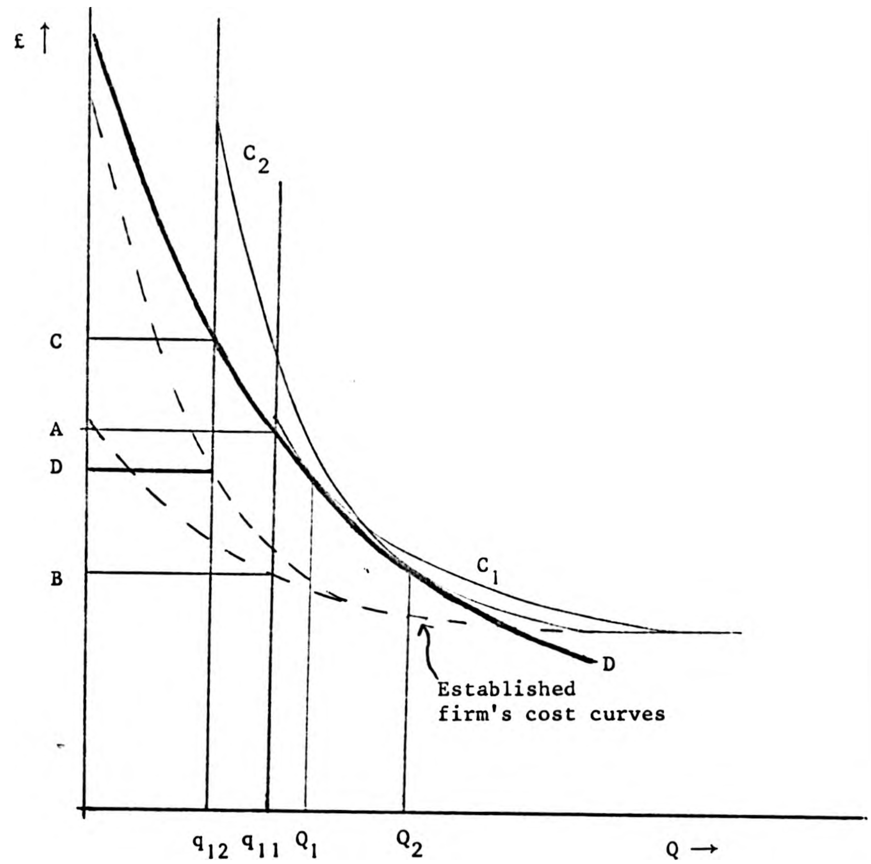
Unfortunately it does not appear possible to establish Modigliani's results rigorously with the apparatus that we have developed, we must be content to see their applicability in many cases. Taking, for example, his prediction regarding increased economies of scale, we note that this is defined by him as represented by a cost curve which is everywhere steeper than one showing lesser scale economies. Now in certain cases this would not affect $(\Pi/R)_L$, specifically where all cost curves are steeper than the demand curve at the relevant point, so that a corner solution is reached.

However, we may take a more straightforward case, where the demand curve touches firm 2's cost curve from below. Here the cost curve is more bowed than the demand curve and its slope is initially greater (negatively) but finally smaller than the demand curve in the relevant region. Thus if the cost curve becomes steeper due to greater economies of scale then the second firm's output increases as the cost curve and demand curve have the same slope further along the cost curve (nearer to the point at which costs cease to fall). Since the "size of the market" is fixed, so fixing the demand curve, and because Modigliani's economies of scale always last for the same interval of output then the steeper cost curve must be nearer the origin. Thus although the second firm's output and total output rise, firm one's output will fall (see Diagram 4 neglecting the dotted lines). The import of this fall in the established firm's output is that he can allow limit price to rise. In terms of the equation for $(\Pi/R)_L$, q_{1L}/Q falls and this fall outweighs the rise in AVC_1/AVC_2 as economies of scale increase.

Despite this, it is not completely clear that if the market is such that the established firm is not producing at his minimum cost then the result that increasing economies of scale increases the limit price cost margin necessarily follows through.[12] This possibility is indicated, though not demonstrated, by diagram 4 where the dotted lines refer to the

DIAGRAM 3.4

Modigliani's Economies of Scale effect



D is the industry demand curve, fixed by "market size". C_2 is the entrant's steeper cost curve ("greater economies of scale") yielding output for firm 1 of q_{12}, total output Q_2 and limit price of C. Under cost curve C_1, output for firm 1 is q_{11}, total output is Q_1 and limit price is A. In each case at the respective limit prices, firm 2 is on the margin between entry and exit since zero profits are being made.

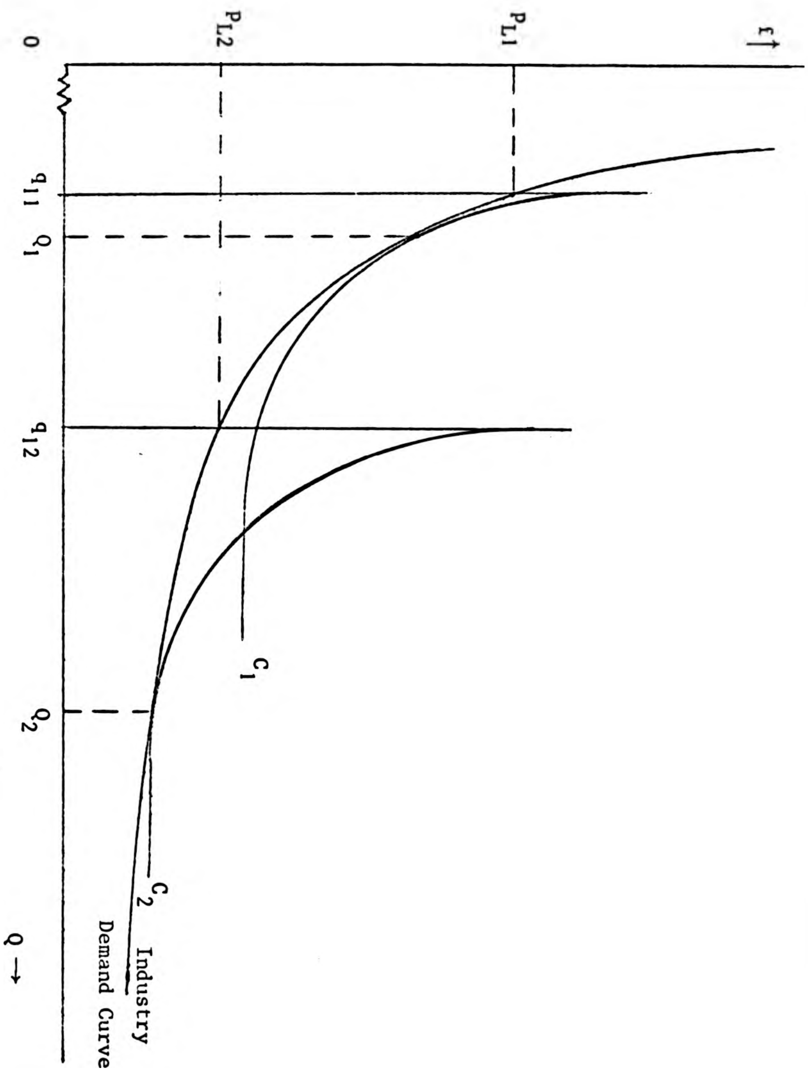
established firm's cost functions under the two assumptions about economies of scale. Firm 1's price-cost margin is greater where economies of scale are smaller, the difference between price and cost being AB in the first case and CD in the second. (Both output points appear to be blockaded entry positions though, the requirement that the first firm's output be beyond the point at which marginal revenue cuts marginal cost from above is not satisfied).

Modigliani's definition of the size of the market, "competitive output" divided by minimum cost output while, as he admits, having drawbacks for empirical estimation[13] does lead to fairly unambiguous results. For as the size of the market increases, elasticity of demand and scale economies constant, the output of the established firm if he is to prevent entry must similarly rise. In fact the output of the potential entrant will usually also rise slightly but not, it would appear, enough to be greater than the increase in output of the first firm (see diagram 5). Thus in the equation from (12) we have that q_{1L}/Q rises; AVC_1 divided by AVC_2 probably rises also. That is as the size of the market increases, the limit profit-revenue ratio falls. It is again possible that this conclusion may be upset when we allow the market, before and after the size increase, to be such that the first firm may not attain full scale economies, though as the market size increases this situation becomes less likely.

Finally, we ought to examine Modigliani's prediction regarding the elasticity of demand. Superficially this looks fairly straightforward since $q_{1L}/Q < 1$, thus in (12) raising the elasticity of demand $|\eta|$ means reducing the exponent on q_{1L}/Q which raises the second term and therefore reduces $(\Pi/R)_L$. Unfortunately the actual outcome is rather more complex because when elasticity of demand changes ordinarily so will both q_1 and q_2.

DIAGRAM 3.5

Modigliani's "size of market" effect



The industry demand curve here is constantly elastic. As market size increases, scale economies constant, the entrant's cost curve shifts from position C_1 to C_2. P_{L1}, q_{11}, and Q_1 are limit price, firm 1's output, and total output initially; P_{L2}, q_{12} and Q_2 ditto after the increase in market size.

One prediction which follows fairly directly in our model of equation (13) is that, if fixed costs faced by the second firm on entry rise, then the established firm acting according to the postulates is more likely to prevent entry. The upshot of this is that the limiting consistent margin should rise consequent upon an increase in AFC_2 in equation (13). Equation (12a) also indicates that higher fixed costs for the entrant produce a higher limit margin. Notice that an increase in the variable costs of the second firm, with those of the first firm remaining unchanged can be dealt with by the Osborne or Modigliani models, using the former if a horizontal cost curve is elevated. In the Modigliani case, a <u>cet. par</u>. increase in the second firm's variable costs can be thought of as an increase in economies of scale where the established firm has exhausted his scale economies (i.e. in producing at minimum cost). Thus the prediction that an increase in the second firm's variable cost will raise the limit price-cost margin is unambiguous.

III Some Measurement Problems

The basic model developed above suggests that there are two main types of factors which affect the limit price. These are scale economies and cost advantages. We now consider some of the measurement problems associated with them.

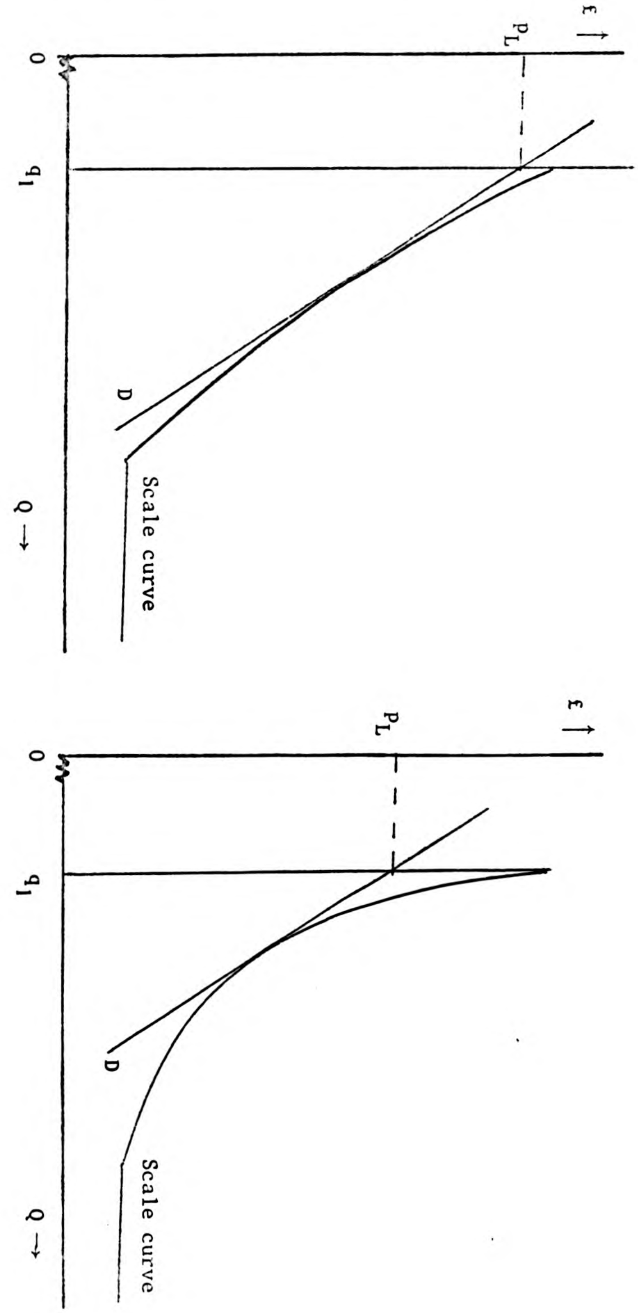
For analytical convenience, Modigliani separates what we commonly call scale economies into two different effects. The first is what might be called the extent of scale economies, that is the range of output or the "size of the market" in his terminology, over which they operate. Having abstracted this he is then free to consider a standardised scale economy of length one unit and use the drop of this curve, or what might be called the degree of the scale economy, as his economies of scale effect. While this is elegant, it is not particularly helpful from an empirical point of view, due to his definition of the size of the market. Furthermore, we should expect some interaction since they are not separable in practice.

In point of fact many empirical studies use a scale economy measure which is at best a proxy for only the first of these effects. Comanor and Wilson in their classic (1967) article for example, use the average size of plants accounting for that half of output produced in larger plants. Caves et al. (1975) attempt to do better by using a composite measure based both on a value similar (or identical in their empirical work on the U.S.) to Comanor and Wilson's, called MEPS, and on a "cost-disadvantage measure". The latter is taken as the ratio of the average value added by the smaller plants divided by average value added by larger plants where again the distribution of plants is cut at the 50th percentile of output. Their composite measures are of two sorts, the first being the product of MEPS and a zero-one dummy cut-off variable for extent of cost disadvantage, while the second is the ratio of MEPS and the cost disadvantage measure. Although these measures are ingenious, and probably an improvement on the original, nevertheless there is no real theoretical basis for that particular type of combination. Their proxies would appear not to take account of curvature in the scale curve which may greatly affect the level of the limit price as the diagram overleaf shows (diagram 6). They may also be measuring fixed rather than variable cost advantages. Unfortunately scale economies are very difficult to categorise by means of a summary statistic, and even if it could be done the way in which such a statistic should enter a multivariate relationship is not clear.

Superficially, since absolute cost disadvantages (or advantages) enter more directly into a relationship such as (13) they might appear to be easier to measure than scale economies. This is not the case. An entrant is at a cost disadvantage relative to the established firm if he cannot reach the level of that firm's cost curve on the same

DIAGRAM 3.6

Curvature of the scale curve affects limit price



Both demand curves, D above, are of the same slope, both scale curves are of the same "range" of output and have the same "drop", but curvature is different so making for a different limit price, P_L, between the diagrams.

basis that the first firm did. As Sherman (1974) puts it:

"A cost of hiring resources is not in itself an entry barrier. A prejudicial circumstance, however, that places potential entrants at a disadvantage can easily mar the prompt adjustment of resources to meet consumer demand." (p. 244)

Theoretically, then, large capital requirements are no barrier to entry unless the entrant does not have access to that capital on the same basis. Thus the entrant is at a disadvantage and his cost curve is raised above that of the established firm by the discounted value of the extra payment on capital demanded from the entrant firm; not an easy thing to measure even in the simplest case of the capital required to set up plant.

A similar argument would appear to hold for other types of "capital", for example, the capital provided to the established firm in the form of goodwill towards his particular product, perhaps due to advertising. In the case of advertising, we in fact know even less about how it is supposed to affect the limit price. The theoretical gap here has been noted by Williamson:

"the study of the entry question can hardly be considered complete until selling expense has been introduced formally (rather than disconnectedly - or not at all) into the analysis" (p.112, 1963).

Williamson attempts to plug this gap by setting up a model where "selling expense" is treated as the product differentiation barrier, and its increasing use raises the limit price in a sinusoidal manner. Revenue is also increased by selling expense. Profits are maximised subject to price being less than or equal to the relevant point on the limit price curve. Using this model, except where entry is blockaded, he finds that:

$$\frac{\partial R}{\partial q_1} = MC_1 + \lambda \frac{\partial p}{\partial Q}$$

$$\frac{\partial R}{\partial S} = 1 + \lambda \left(\frac{\partial p}{\partial S} - \frac{dp_L}{\partial S} \right)$$

(R is revenue, S selling expense and λ (> o) the lagrangean multiplier. Other notation is ours; since entry does not take place, $q_1 = Q$).

From these equations we may derive the following expression for the limit price-cost margin:

$$\left(\frac{p - MC}{p} \right)_L = - \frac{1}{p} \frac{\partial p}{\partial Q} \left[Q - \lambda \right]$$

$$= \frac{1}{|\eta|} \left[1 - \frac{1}{Q} \left(\frac{\frac{\partial R}{\partial S} - 1}{\frac{\partial p}{\partial S} - \frac{dp_L}{dS}} \right) \right]$$

Obviously, unless we are willing to quantify $\partial p/\partial S$ and dp_L/dS then his model does nothing to fill the gap in the theory, though it makes it more obvious and indicates the types of factors we need to know about. The fact is that equations such as (12) & (13) provide very little guide to profitable empirical work.

IV Some Extensions:

Despite the rather negative conclusions of the previous section we ought to consider the possible directions in which the basic model ought to be extended to render it nearer to reality. We briefly outline some of these before moving to consideration of a selection in more detail:

1. Our assumptions about the established industry have been extremely naive, either there is a monopolist or a tightly-knit collusive oligopolistic group in residence. As Stigler says "The theory of Oligopoly has been solved by murder" (1968 p.21). In order partially to remedy this

situation we later take the case where the established firms form a Cournot oligopoly but collectively act as leader to the entrant's followers as an example of the extensions possible.

2. Perhaps less importantly, we have also assumed that there is only one potential entrant to the industry. While there are quite likely to be several potential entrants, it is unlikely that many will be equally favourably placed to enter. However we do consider this situation briefly later in the present section.

3. The model presented above is purely deterministic. If the established industry prices even slightly above p_L then a firm is bound to enter, while a price below p_L will shut out any possibility of entry. Such a situation has tempted some, for example Williamson, to ask whether a more probabilistic setting for the model might not be superior. While we will not investigate this suggestion in detail, it will be mentioned again in the following section.

4. As was pointed out at the beginning of this chapter, Bain considered that there are two cases where the limit price would be important, those of effectively impeded and ineffectively impeded entry. According to his scenario, the firm either prices at p_L and keeps out all entry or prices at the short-run profit-maximising price and allows entry over time. This dichotomy prompted Hicks,[(1954)] in an exploratory paper, to point out that it might be optimal instead to price at some value between the two. That is, long run profits may be maximised by retarding rather than preventing or passively allowing entry. To this end, Hicks developed the concept of "stickers" and "snatchers" who tend to put more weight on long-run considerations and quick profits respectively. Since then there have been many papers which have considered the possibility more fully, and we attempt an evaluation of these papers in the next section.

5. Lastly, we might reflect on the validity of the key assumption of limit pricing theory, the Sylos postulate. An interesting alternative due to Spence is discussed in the penultimate section of this chapter.

Let us now assume that there are N established firms each with a profit function of the type:

$$\Pi_{1i} = pq_{1i} - c_{1i}(q_{1i}) - F_{1i}$$

First order conditions for maximum profit require that:

$$\frac{d\Pi_{1i}}{dq_{1i}} = p + q_{1i}\frac{dp}{dQ}\cdot\frac{dq}{dq_{1i}} + q_{1i}\frac{dp}{dQ}\cdot\frac{dq_2}{dq_1}\cdot\frac{dq_1}{dq_{1i}} - c'_{1i}(q_{1i}) = 0$$

Assuming for simplicity that these firms act towards each other (though not the entrant[14]) in a pure Cournot manner, we may set each firm's perceived derivative regarding established firm reactions, $dq_1/dq_{1i} = 1$. Multiplying the above equation by q_{1i} throughout:

$$pq_{1i} + q_{1i}^2\frac{dp}{dQ}(1 + \phi') - q_{1i}c'_{1i}(q_{1i}) = 0 \quad (\phi' \text{ defined as before})$$

Summing over the N firms:

$$\frac{\Sigma pq_{1i} - \Sigma q_{1i}c'_{1i}(q_{1i})}{pq_1} = -\frac{q_1}{p}\frac{dp}{dQ}(1 + \phi')\frac{\Sigma q_{1i}^2}{q_1^2} = \frac{q_1}{Q}\frac{H_1(1+\phi')}{|\eta|} \qquad (14)$$

Equation (14) holds at the total output point under the postulates.

Let us make the simplifying assumption (which may or may not be reasonable) that the established firms are able to co-ordinate in some manner in order to set limit price if they so wish. If this be the case then there seems no particular reason why the limit price should necessarily change if there are N, rather than 1, firms in the established industry.[15] The limit price, after all, may be obtained by the simultaneous solution of equations (6) and (7) alone. Our attention then focusses naturally upon the question of consistency, and we here move towards the N firm equivalent to equation (13) by partially solving (6), (7) and (14).

Rearranging (14):

$$p\ \Sigma q_{1i} - \Sigma q_{1i} c'_{1i}(q_{1i}) + \frac{dp}{dQ}(1+\phi')\ H_1 q_1^2 = 0$$

$$\therefore \quad q_1 = \frac{p - \Sigma q_{1i} c'_{1i}(q_{1i})/q_1}{H_1 \left|\frac{dp}{dQ}\right| (1+\phi')} = \frac{p - AMC_1}{H_1 \left|\frac{dp}{dQ}\right| (1+\phi')}$$ [16]

From (6) we may substitute for p to yield:

$$q_{1LC} = \frac{(AVC_2 + AFC_2 - AMC_1)}{H_1 \left|\frac{dp}{dQ}\right| (1+\phi')}$$

and we have still:

$$q_{2M} = \frac{AVC_2 + AFC_2 - MC_2}{\left|\frac{dp}{dQ}\right|}$$

Thus since $Q = q_{1LC} + q_{2M}$ at the limiting consistent point:

$$\frac{q_{1LC}}{Q} = \frac{AVC_2 + AFC_2 - AMC_1}{AVC_2 + AFC_2 - AMC_1 + H_1(AVC_2 + AFC_1 - AMC_1)(1+\phi')} \tag{15}$$

Limit profits in the established industry are given by:

$$\left(\frac{\Pi + F}{R}\right)_L = 1 - \frac{AVC_1}{p_L}$$

where AVC_1 is the average variable cost of producing output q_1. Again assuming that we have constant elasticity of demand so that (10) continues to hold, we may write the limiting consistent margin from (12) as:

$$\left(\frac{\Pi + F}{R}\right)_{LC} = 1 - \frac{AVC_1}{AVC_2 + AFC_2}\left[\frac{q_{1LC}}{Q}\right]^{1/|n|},$$

but this time for consistency q_{1LC}/Q is given by (15) and involves the concentration in the established industry measured by H_1. In the particularly simple case where $MC_{1i} = MC_2$ (for all i), so that the marginal cost of the potential entrant is equal to the marginal cost of all the established firms, then all established firms are of the same size and $H_1 = \frac{1}{N}$. This yields the following expression for consistency:

$$\left(\frac{\Pi + F}{R}\right)_{LC} = 1 - \frac{AVC_1}{AVC_2 + AFC_2} \left(\frac{1}{1 + \frac{1}{N}(1+\phi')} \right)^{1/|\eta|}$$

As the number of firms increases it would seem more likely that consistency occurs. A similar sort of result would presumably obtain in the more complicated case where H_1 falls (as in (15)). For as the number of firms increases it is optimal (under the Sylos and Cournot assumptions and in a short-run sense) for the sum of outputs of the established firms to increase so making it less likely that the entrant can make profits in the residual demand. Since this increase in output is bound up in the determination of the limiting consistent profit revenue ratio the latter also should fall. (However this assumes that $AVC_2 + AFC_2$ does not rise sufficiently to offset the rise in q_{1LC}/Q).

Sherman and Willett (1967) consider a further elaboration of the basic model where there is more than one firm considering entry into an industry. They utilise a game-theoretic framework, but it is easy enough to place their problem within the context of the present model. If we take first the case where each firm considering entry has no knowledge of other potential entrants, then the analysis follows the pattern we have already specified in every respect. For if the entrant knows nothing of others then the established industry must treat him in this light and attempt to maximise industry profits, subject to the constraint that the maximum profit the entrant can obtain is zero. It is as if that firm were the only potential entrant.

As an example of a more complex case we take, with Sherman and Willett, the situation where each entrant is completely cognisant of the other aspirants to that industry. The established firm's problem[17] is to maximise Π_1, subject to Π_{2j} (maximum) ≤ 0 for all j (unless entry is blockaded).

Now: $\Pi_{2j} = pq_{2j} - c_2(q_{2j}) - F_{2j}$[18] $j = 1, 2, \ldots, E$.

For maximum profits:

$$\frac{\partial \Pi_{2j}}{\partial q_{2j}} = p + q_{2j}\frac{dp}{dQ} \cdot \frac{dq_2}{dq_{2j}} - c_2'(q_{2j}) = 0$$

assuming the potential entrants do not recognise that the established firm will react to them so they consider $dq_1/dq_{2j} = 0$.

Also, under our assumptions regarding potential entrants' reactions to other potential entrants, they perceive that $dq_2/dq_{2j} = E$, so the above equation becomes:

$$p + q_2\frac{dp}{dQ} - c_2'(q_{2j}) = 0 \qquad (16)$$

since all firms are of equal sizes.

Forming predictions in terms of the limit profit-revenue ratio we have:

$$p = AVC_{2j} + AFC_{2j} \qquad (17)$$

Thus in place of (12):

$$\left(\frac{\Pi + F}{R}\right)_L = 1 - \frac{AVC_1}{AVC_{2j} + AFC_{2j}} \cdot \left[\frac{q_{1L}}{Q}\right]^{-1/|\eta|}$$

where $Q = q_{1L} + \sum_E q_{2jM}$.

To the same first order approximation as in (12a):

$$\left(\frac{\Pi + F}{R}\right)_L \simeq 1 - \frac{AVC_1}{AVC_{2j} + AFC_{2j}} \left[\frac{MC_{2j}}{AVC_{2j} + AFC_{2j}}\right]$$

Comparing these expressions with (12) and (12a) in the basic model, we note that unless the cost curves for the entrants are horizontal we would expect $[(\Pi + F)/R]_L$ to be larger the greater the number of potential entrants. This is mainly because of the very much higher incidence of fixed costs for each entrant firm.

Of course we would normally expect that each potential entrant would be only partially aware of his rivals, in which case the solution would lie between the two extremes presented above.[19] In general, we find, with Sherman and Willett, that "Potential entrants discourage entry".

V Dynamic Entry Models:

Recently several papers concerned with the optimal dynamic strategy of an industry faced with potential entry have been written. Among these are papers by Pashigian, (1968), Gaskins (1971), Kamien and Schwartz (1971), Pyatt (1971), Ireland (1972 (a),(b)), Jacquemin and Thisse (1972) and Schupack (1972). We here attempt a brief evaluation of these, considering the Gaskins and Kamien and Schwartz papers in more detail.

The general approach of these papers is first to postulate a fixed entry limiting price, below which no new firms will enter. As Jacquemin and Thisse, overstating the determinacy of static entry-barrier theory put it: "The work of Bain, Sylos-Labini and Modigliani allows us to determine the exact size of the discrepancy between the limit price and the competitive price" (p. 69). To the extent that this assumption is simplistic, the models lack economic content. Secondly, the established industry (which is normally a monopolist, possibly with a competitive fringe, or collusive group) is considered to maximise long-run profits (at some relevant rate of discount) using a particular control variable, usually price. Thirdly, the assumption which makes these models novel is the reaction that potential entrants are assumed to take with respect to some industry performance characteristic (prices, profits or "profit opportunities"), known as the state equation. The inherently dynamic problem posed in this manner is then solved using calculus of variations or control theory.

Attention naturally focusses on the path followed by industry price over time and any comparative static predictions which can be gleaned from the models. We proceed here by stating Gaskin's basic model, discussing how it relates to the other papers, and then returning to our consideration of Gaskin's and Kamien and Schwartz' dynamic results.

Gaskin's maximand is the function:[20]

$$V = \int_0^{\infty} \left[p(t) - c\right] q_1 \, (p(t),t) \, e^{-rt}.dt \qquad (18)$$

where $q_1(p(t),t) = Q - q_2(t)$

$$= f(p(t),t) - q_2(t) \qquad (19)$$

c is average (or marginal) cost, t is time and r is the discount rate. The established firm wishes to maximise long run profits, the control variable being price. The rate of entry is given by a linear function:

$$q_2'(t) = k\left[p(t) - p_L\right] \qquad q_2(0) = q_{20} \quad p_L \geq c \qquad (20)$$

If industry price is above the limit price this causes entry by a passive competitive fringe.

Using optimal control theory, Gaskins forms the Hamiltonian for the problem and solves with the help of Pontryagin's maximum principle and a phase diagram. He finds that the optimal price path involves the established firm cutting price over time from a point below the myopic profit-maximising price to the limit price, with the change in price slackening off as time passes.[21]

Compared to Gaskins' model, Pashigian's earlier paper has a more limited aim. He considers only the two classical policies of short-run profit maximisation and limit pricing and simply derives the optimal point at which to change from one to the other. Pyatt's paper is rather similar to that of Gaskins, though he places it within the context of a Harrodian type model (e.g. Harrod (1967)). However his reaction function is linear in profits not prices, and the outcome "that in many, but not in all situations, firms may [illegible]ise profits in the short run ..." (p.254) is rather different. It is interesting to consider why this is the case. Ireland (1972(b)) has explored this result and finds that it occurs because Pyatt's reaction function is linear in profits (see pp 7-8). Although Gaskin's reaction function is also linear, profits are not normally assumed to be linear in prices so that Gaskin's reaction function is not in fact linear in profits and his conclusions are in some sense more general.

Actually, one of the problems of modelling in this area appears to be that what seem to be fairly subtle changes in specification can lead to widely differing results. As an example of this, one of Gaskin's results is that if the dominant firm has no cost advantage then in the long-run its market share will tend to zero. However, when Jacquemin and Thisse consider an extention to Gaskins' model whereby the dominant firm is allowed to take-over entering rivals (if the benefits of doing so outweigh the costs) they find that the dominant firm with no cost advantage will achieve a positive long run market share. On a similar point, Gaskins finds that when he introduces growth into his model the above-mentioned result does not hold and even with no cost advantage the dominant firm does not eventually price itself out of the market. Now his dominant firm demand function (equation (19)) is altered to :

$$q_1\ (p(t),t) \ = \ f(p(t))e^{gt} - q_2(t)$$

in the growth case. But, as Ireland (1972(a)) points out, this essentially assumes that only the dominant firm grows, fringe firms derive no benefits from industry growth. If we adjust the above equation to allow fringe firms to grow at the same rate as the dominant firm then Gaskins' conclusion regarding the growth case is upset. As a final example Schupack introduces advertising as an additional control variable in his models. Depending upon the way it is introduced, long-run equilibrium price will either be raised while short-run price is lowered, or short run price is raised, long run price remaining unchanged.

Kamien and Schwartz build a model rather different to those we have so far considered. Their purpose is not to derive an optimal path as entry proceeds but rather to discover the pricing policy prevailing until entry occurs with the help of a simple two-period analysis.[22] An interesting departure is that they view entry probabilistically.

Profit before entry takes place is given by: $e^{gt}\,\Pi_I(p(t))$ where the industry grows over time at rate g and profits are a concave function of price. After entry, since this is purely a two-period model, profits are maximised in the industry. Also because industry attraction increases with the market growth rate they assume that the magnitude of the entering firm may also vary directly with the market growth rate. Expected profit after entry is then

$$e^{gt}\Pi_{II}(g) \quad \text{with } \Pi'_{II}(g) \leqslant 0; \quad 0 \leqslant \Pi_{II}(g) < \Pi_I \text{ max.}$$

Π_{II} is not a function of p in the context of their model because profits are always maximised (with respect to price) in the second period.

Entry is a probabilistic function. They assume that the amount of entry is <u>not</u> a function of first period price but in fact it is the timing or occurrence of entry which is stochastically governed by the established firms pricing policy. A higher price makes entry more attractive. The (instantaneous) conditional probability of entry at time t is then given by:

$h(p(t);g) \geq 0$ where $\partial h/\partial p \geq 0$, $\partial^2 h/\partial p^2 \geq 0$, $(p \geq 0)$, $\partial h/\partial g \geq 0$

We may write $h(p(t);g) = F_2'(t)/(1 - F_2(t))$ where $F_2(t)$ is the probability that entry has occurred by time t.

Drawing these various strands together, the problem is to maximise:

$$V = \int_0^\infty e^{-(r-g)t} \left[\Pi_1(p(t))(1-F_2(t)) + \Pi_2 F_2(t)\right] dt$$

subject to $F_2'(t) = h(p(t))(1 - F_2(t))$. $F_2(0) = 0$.

(With Kamien and Schwartz, we have suppressed the arguments relating to g in this equation). Again the problem is one in control theory and is solved for the optimal price p*. They find that p* is in fact a constant lying between the short-run profit maximising price and the limit price.

Recently Kamien and Schwartz (1975) have built a similar model to that laid out above but concerned with a Cournot established industry rather than a monopolist. They find that either the firm realises the possibility of entry but considers it independent of its own actions and thus prices in a short-run profit maximising manner, or there is a Chamberlinian zero-profit equilibrium, or finally that output exceeds and price falls short of the myopic profit-maximising level so that the firm's marginal revenue is less than its marginal cost due to its taking account of rivals. In each case, unless marginal and average costs are equal, a Chamberlinian equilibrium eventually ensues, as we might expect.

We now turn to a comparison and discussion of the comparative static (and in Gaskins' case comparative dynamic) predictions of these two models. As regards the predictions of his first model, Gaskins obtains the following comparative static predictions:

$\frac{d\,\hat{q}_2}{d(p_L - c)} < 0$ - the dominant firm prices to allow less entry as his cost advantage increases.

$\frac{d\,\hat{q}_2}{dr} \geq 0$ - As the dominant firm's discount rate rises, he sacrifices long-run market share.

$\frac{d\,\hat{q}_2}{dk} \leq 0$ - the more rapidly rivals respond to price signals, the larger is his long-run market share.

$\frac{d\,\hat{q}_2}{dq_{20}} = 0$ - the initial size of the competitive fringe does not affect its final size.

($\hat{q}_2$ is the equilibrium market share of the competitive fringe).

The first, second and fourth of these predictions seem eminently reasonable. The prediction regarding the response coefficient k is rather more surprising, Gaskins does not explain it fully but it may occur because the earlier periods are weighted more heavily and this is when most entry will occur. He obtains very similar predictions to these using his growth model.

Much of the interest focusses on predictions regarding price. These are at two levels, equilibrium price and price path results. From his growth model, Gaskins obtains the following signs for equilibrium price level changes:

$$\frac{d\hat{p}}{dg} > 0$$

$$\frac{d\hat{p}}{dk_o} < 0$$

$$\frac{d\hat{p}}{dc} > 0$$

$$\frac{d\hat{p}}{dr} > 0$$

$$1 > \frac{d\hat{p}}{dp_L} > 0$$

Recalling our earlier comments on Gaskins' specification of the growth model, we should probably treat at least the first of these predictions with caution. The third and last of the predictions are as would be expected, whereas the others are less intuitively obvious.

Finally turning to his predictions on optimal price, we should note that Gaskins has obtained these by means of inspection of changes in the phase plane portrait. While this can show us changes in the position of the optimal trajectory, it says nothing about velocity along the price path. He therefore adds the caveat that "Each of these conditions indicates the <u>short term</u> effect of variation of a particular model parameter." (p. 314) To that extent he establishes the following results which hold in both static and growth models:

$$\frac{dp^*}{dk}(t) < 0$$

$$\frac{dp^*}{dr}(t) > 0$$

$$\frac{dp^*}{dc}(t) > 0 \quad \text{if } f''(p) \approx 0$$

$$\frac{dp^*}{dq_{20}}(t) < 0$$

The first of these is the counterpart to the earlier result on $\hat{q}_2$, while the second, third and fourth are intuitively fairly obvious. However he does state a further important result for certain classes of demand curve, including linear and constant unitary elasticity types. This is that

$$\frac{dp^*}{dp_L}(t) < 0$$

which together with the third of the results immediately above implies that:

$$\frac{dp^*}{d(p_L - c)}(t) < 0 \ .$$

We shall comment on this prediction shortly.

Kamien and Schwartz also obtain some comparative static results regarding their (constant) price policy. They are:

$$\frac{dp^*}{dr} > 0$$

$\frac{dp^*}{d\Pi_{II}} > 0$ where Π_{II} is affected by some exogenous change.

$$\frac{dp^*}{dg} < 0$$

$$\frac{dp^*}{d(\text{barrier})} > 0$$

Of these predictions, the first two seem unexceptional; our attention naturally focusses on the third and fourth results. As Kamien and Schwartz explain, an increase in the growth rate causes current profits to be larger so that entry is more attractive. Since future profits will similarly be larger, firms are more likely to want to enter the industry and possibly will enter in a bigger way. However in considering this result we must remember the asymetric nature of their model; entry only occurs once. If there were the prospect of future entry or other firms interest in entry at the same time, then a potential entrant would be likely to take this into account when deciding upon the profitability of entry. It may be the case that such effects would weigh heavily against their prediction.

Kamien and Schwartz do not have limit price explicitly in their model, it is somehow implicit in their $h(p(t);g)$ function. Thus to derive their prediction regarding barriers they split this hazard function into separate components and an increased level of barriers to entry is given by a fall in m in the function:

$$h(p;g) = mk(p)\,J(g)$$

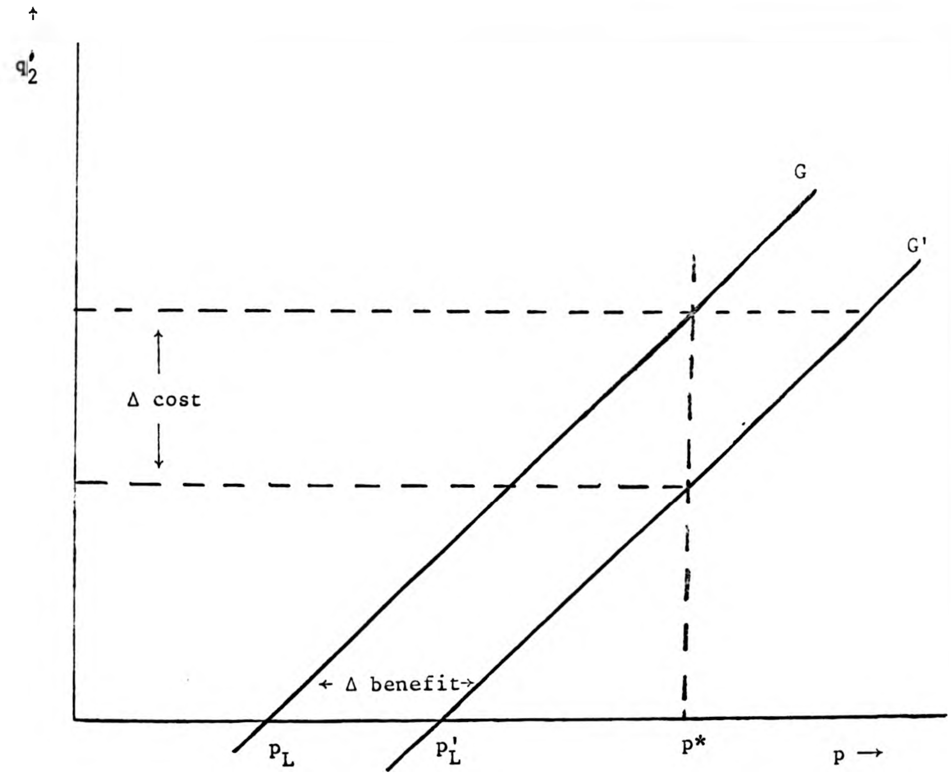
This is in fact not the only way in which higher barriers to entry might be considered to enter their hazard function, and other formulations can alter this prediction.[23]

We should also consider Gaskins' result that, for certain demand functions a rise of p_L with respect to c can cause p* to fall. At the same p* before and after a rise in p_L, both the benefit, in terms of extra profit, of maintaining this price rather than p_L, and the cost in terms of future entry fall (see diagram 7). It is not intuitively obvious that one will definitely change more than the other, this will depend on the shape of the profit function (and the reaction function if we allow it to vary), as his limited result implies.

Having said this, it remains true that Gaskins' result is extremely annoying to those who wish to perform empirical work explaining price-cost margins while including as explanatory variables measures of barriers to entry. For we are not assured that an increase in our measures of entry barriers will necessarily raise price and so price-cost margin. In fact, it is probably not unfair to say that dynamic entry models have rather negative conclusions for cross-sectional empirical work on explaining the behaviour of profits. If we recall the predictions on price that Gaskins and Kamien and Schwartz have given us we find that, as we would expect, actual price tends to be between myopic profit-maximising price and limit price. Apart from this we find that the discount rate affects price, yet there is no particular reason why it should vary between industries; that entrants' responsiveness affects price, where the same comment applies; that post-entry profit opportunities affect price, while there is no real reason why these should suddenly change; that the size of the initial

DIAGRAM 3.7

The effect of a rise in limit price on Gaskins' reaction function



q_2' is the rate of entry

G is Gaskin's reaction function before change in p_L, G' that for the new p_L'. (equation (20)).

p* is the pre-change optimal price at a certain point in time. Benefits and costs are not, of course, measured on the same scales in the diagram which contains insufficient information to enable us to perform measurement.

fringe, whatever that is, affects price; and that growth rate and barriers to entry may or may not affect price. This is not meant to imply that those building dynamic models have been wasting their time, just that both their assumptions, which lack economic richness, and their predictions do not particularly lend themselves to use in empirical work.

VI Spence's Contribution

As we have noted at the beginning of this chapter, the key assumption of limit pricing theory is the Sylos postulate, an assumption which permits great simplification in the treatment of potential entry. But this does not mean there is anything sacred about supposing that established firms maintain output in the face of entry; they might conceivably contract or expand output though the former would appear unlikely. In fact among alternatives, Andrews (1949) assumed that the entrant would normally obtain only his share of the increase in output (see Bhagwati (1970 p. 302)) and consideration of an active output expansion policy on entry has received attention in the literature on predatory pricing and elsewhere (see particularly Yamey's (1972b) discussion of the concept of predatory pricing). However it was left to Spence (1974) to relate the idea of output expansion, together with its implications, formally to limit pricing theory.

Such a policy requires capacity to be in excess of current output except when entry occurs. Spence's key postulate is therefore that "capacity will be maintained at a level where, if entry is threatened, the existing industry can expand output and lower price <u>within the time required for the entrant to enter</u>, to a point where profits cannot be made by the entrant in the residual demand" (p.2, our emphasis). Any potential entrant must of course be fully aware that this is established industry policy. We proceed by first elaborating Spence's simplest model:

In the long run variable costs are a function of both output and capacity, but capacity does not affect marginal costs. Therefore in the short run variable costs are purely a function of output:

$$c_1 (q_1, k) \quad = \quad c_1 (q_1) + rk$$

where k, capacity, is measured in output units and r, the cost of capacity, is chosen to allow this. Since, if entry occurs, the firm must expand output to capacity then capacity must be at least equal to the entry limiting output, q_{1L}, which we will here write k_L to avoid confusion. Thus capacity, but not necessarily output, must be at the level indicated by the Sylos postulate, and as such k_L is determined as in our previous models. The established firm's profits are of course given by $\Pi_1(q_1,k) = pq_1 - c_1(q_1) - rk$, making the not implausible assumption that capacity costs are the only (potentially) fixed costs. Profits are maximised subject to the obvious constraints that output cannot exceed capacity but that to prevent entry capacity cannot be less than k_L. Formally, maximise:

$$\left.\begin{aligned} \Pi_1 &= pq_1 - c_1(q_1) - rk \\ \text{subject to } q_1 &\leqslant k \\ k_L &\leqslant k \end{aligned}\right\} \qquad (21)$$

Assuming that q_1, $k \neq 0$ (and after some slight simplification), the Kuhn-Tucker conditions for this problem are:

$$\left.\begin{aligned} & p + q_1 \frac{dp}{dQ} - c_1' (q_1) = \lambda \\ & r = \lambda + \mu \\ & \lambda(k - q_1) = 0, \ \mu(k - k_L) = 0 \\ & \lambda, \mu \geqslant 0, \text{ where } \lambda, \mu \text{ are the} \\ & \quad \text{Lagrangean multipliers} \end{aligned}\right\} \qquad (22)$$

There are three cases of economic interest to be derived from (22): Firstly, if $\mu = 0$ but $\lambda \neq 0$, then $\lambda = r$ and

$$p + q_1 \frac{dp}{dQ} = c_1'(q_1) + r \qquad [q_1 = Q]$$

This is the long-run blockaded entry position.

Secondly, if $\lambda \neq 0$, $\mu \neq 0$, we have that $q_1 = k_L$ and

$$p + q_1 \frac{dp}{dQ} < c_1'(q_1)$$

That is output is taken beyond the short-run profit maximising level because of the threat of entry. This case is identical to that which we have previously considered at length, since q_1 is set equal to q_{1L} to forestall entry.

Finally, Spence's model allows us to consider a further case of extreme interest where $\lambda = 0$, $\mu \neq 0$. Here we have $\mu = r$, $k = k_L$, but $q_1 < k$. Also:

$$p + q_1 \frac{dp}{dQ} = c_1'(q_1)$$

"the short run pricing game is to some extent, strategically independent of entry" (Spence p.7). In other words, the determination of price and of capacity to deter entry are conceptually separate. As long as no entry occurs, $q_1 = Q$ and we have:

$$\frac{p - MC}{p} = \frac{1}{|\eta|} \quad (23)$$ for the established firm (MC is marginal cost)

This case extends fairly easily to the situation where we have more than one established firm. Essentially we have that marginal revenue is equated to marginal costs for the ith established firm. Capacity to deter entry may be determined along the lines of our previous model of the N firm case by solving for q_{1L} which gives k_L and so the capacities for the individual firms. In the short run, assuming that marginal cost is equal to average variable cost for established firms, we have for a monopolist from (21):

$$\frac{\Pi + F}{R} = \frac{1}{|\eta|} ,$$

and for an established oligopolistic industry we would argue from the previous chapter on oligopoly we have:

$$\frac{\Pi + F}{R} = \frac{H(1 + \alpha)}{|\eta|}$$

We comment further on the implications of Spence's results for the measurement of the profit revenue ratio in Chapter 5.

From his basic model, Spence goes on to consider several elaborations. He first takes the case where capacity affects marginal costs and finds results similar to those presented above. Marginal revenue is now equal to the partial derivative of variable costs with respect to output. He also relates his model to the Averch-Johnson (1962) result and includes intertemporal considerations along Kamien and Schwartz lines. He finally extends it to cases where control variables other than capacity are manipulated by the firm in a similar manner. However the main novelties of his approach are illustrated by the above discussion.

In evaluating Spence's contribution, we should note that a monopolist wishing to prevent entry without charging limit price for his product must ensure that his intentions, should entry occur, are completely clear to any would-be entrants. If this is reasonable, and if capacity does not affect marginal cost, then the Spencian particular solution is dominant, in that it offers more profit to the established firm. For fixed costs are no higher than under a static limit pricing policy yet output is at a more profitable level.[27] Indeed his particular solution is dominant even over a dynamic limit pricing policy <u>under these</u>

<u>assumptions</u>. For what retards entry is not price, but the expectation of the post-entry situation. If this is more effectively signalled by capacity than price, then price may be maintained at a short-run profit-maximising level while capacity does the work of retarding entry, as the threat of the established firm flooding the market causing new entrants' losses becomes more real. Throughout the sequence, fixed costs are no higher, but output is more optimally regulated, than under the dynamic limit pricing policy.

When we move to considering an oligopolistic industry, we must bear in mind that capacity is a double-edged sword, for a policy of price-chiselling by an established firm is relatively easy to implement. Here the firms have to allocate excess capacity amongst themselves; this is conceptually exactly the same problem as that already covered of oligopolists deciding upon the limit price. Now it would seem that a policy of price-chiselling by any established firm is relatively easy to implement, so the limiting quantity will be quickly produced. Yet, for the very reason that it is easy for each firm to expand output, we would not expect such an expansion to take place. The argument here is exactly equivalent to that Nicholson (1972) makes (see chapter 2). Reaction times are short, shorter than when there is no spare capacity, so the transitional profit gains to any firm seeking to sell at capacity are small. This makes price-chiselling an unlikely strategy to pursue.

On the other hand of course, if capacity affects marginal costs deleteriously, if capacity has to be kept needlessly high to dissuade entry or if a combination of low pricing and spare capacity is required, then Spence's novel result is not necessarily superior to some limit pricing policy. The costs of the above effects have to be traded off against the benefits of persuing his policy.

VII Conclusions:

Many of the sections of this chapter have been ended on a rather negative or even pessimistic note. It would seem that the commonly categorised barriers to entry do not affect performance in a particularly clear-cut manner. The possibility of retarding rather than preventing entry adds to these complications. With this in mind it would perhaps seem that when turning to empirical work, as a first approximation it might be better to take a Spencian view of short-run pricing behaviour. The alternative is to include statistics purporting to measure barriers to entry as determinants of industrial performance; many of these variables could only be included in an ad hoc manner. On this point we note that the problems of ignoring barrier to entry variables may well be less severe when, as is intended, ratios of the performance variable at different points in time, not the level of profits, are to be "explained".

NOTES

1 Constant returns to scale exist everywhere in their economy.

2 Osborne (1964) considers that the theory of limit pricing (p.398) "is not a theory of price, but a proposition in Welfare Economics". We disagree, considering it to be a theory suggesting the level at which price will be set in order to prevent entry under certain assumptions. The firm is then free to decide on its actual pricing strategy while bearing the limit price in mind, as will be discussed later.

3 After P. Sylos-Labini (1962), one of the early contributors. An alternative, predating Sylos, was provided by Andrews (1949) and refined by Edwards (1955), on which see later.

4 That output at which price is equal to marginal (and average) cost for all firms.

5 We shall subsequently alter this notation slightly further away from that of Modigliani.

6 This is not precisely Modigliani's definition.

7 Assuming the second firm follows the Sylos postulate, and so believes $\frac{dq_1}{dq_2} = 0$.

8 Notice from (8) that we should have $-1 < \phi' < 0$ (see also Osborne).

9 That is, unless $\frac{d\Pi_2(q_1)}{dq_1} = 0$, which seems unlikely.

10 In fact, with certain types of demand curve the result may be indeterminate.

11 From (6), substituting in for elasticity of demand we have:

$$p\left(1 - \frac{q_2}{|\eta|Q}\right) = c_2'(q_2) = MC_2$$

Substituting for p from (7) and rearranging:

$$\left(1 - \frac{q_{2M}}{|\eta|Q}\right) = \frac{MC_2}{AVC_2 + AFC_2}$$

at the total output point.

Now, to a first order approximation:

$$\left(\frac{q_{1L}}{Q}\right)^{1/|\eta|} = \left(1 - \frac{q_{2M}}{Q}\right)^{1/|\eta|} \simeq 1 - \frac{q_{2M}}{|\eta|Q}$$

Together with the equation above this demonstrates result (12a).

11a This gives a limiting consistent margin in that a margin below this value would imply negative rather than zero profits to the potential entrant. A margin above this value would fail to prevent entry, so being inconsistent. In an actual situation, of course, (13) would only be true by accident.

12 Modigliani neglects this possible situation.

13 These drawbacks are especially severe when we relax his assumptions about the shape of the cost curve.

14 As Fisher (1959) points out, if the Cournot producers also act towards entrants in a Cournot manner, then entry proceeds freely. Here dq_2/dq_1 is the entrant's actual reaction to established firm output changes.

15 The fact that there is only one established firm may imply that scale economies exist up to a very substantial proportion of industry output. But this is not necessarily the case.

16 AMC_1 is the weighted average of the established firm's marginal costs.

17 We take a monopolist for simplicity.

18 This formulation implies that all j firms have the same cost function. Unless they do we run into problems with their simultaneous entry.

19 Further theorising in this direction would seem rather spurious unless we could contemplate identifying prospective entrants and the likelihood of their entry in empirical work.

20 The notation is based on that earlier in this chapter.

21 Jacquemin and Thisse point out that this is a fairly typical result of the control theory approach.

22 To avoid conflict without previous symbolism we shall call the periods

I and II and it will be understood that the profits accrue to firm I, the established firm.

23 I am grateful to George Yarrow for pointing this out to me.

24 That is unless by chance entry is blockaded at limit price.

Chapter 4:

Bilateral Power; the consideration of mutual relations between producers

I Introduction:

We have so far been implicitly assuming that each industry in our system has no connexions with any other or, to be rather more precise, that each industry sells in markets where it alone has market power and buys all inputs from competitive industries. While this might be empirically apt, it is theoretically illogical in a model concentrating on using market power variables to explain industry performance. In fact the area of bilateral or multilateral power has received little attention in either the theoretical or the empirical literature of Industrial Economics.[1] For this reason we shall start our discussion of the topic of bilateral power, leading up to the development of a particular model, very much from first principles.

At first sight, our task might seem hopeless; we have Hicks (1935) who wrote:

"Bilateral Monopoly is a phrase which has been applied to two different problems ... . The first is the case of isolated exchange, or exchange between a group of buyers and a group of sellers, each acting in combination ... . Now so far as this problem is concerned I think one may say that there is complete agreement among economists ... that ....the problem is indeterminate ... .

"The second problem is a more complex one. It arises when the commodity sold is a raw material or factor of production; so that we have to take into account the relation of the buyer of the raw material to another market - that in which he sells the finished product. For this problem ... there is indeterminateness also ... " (pp. 16-17)

While thirty five years later, Scherer considered that:

"The theory of Bilateral Monopoly is indeterminate with a vengeance. It embodies all the problems met in oligopoly theory ... i.e. do the parties attempt to maximise their individual profits, ignoring their interdependence, or do they co-operate to maximise joint profits? And, unlike oligopoly, even if the (two) parties do collaborate to establish the joint profit-maximising output the price is indeterminate within a potentially wide range." (p.242)

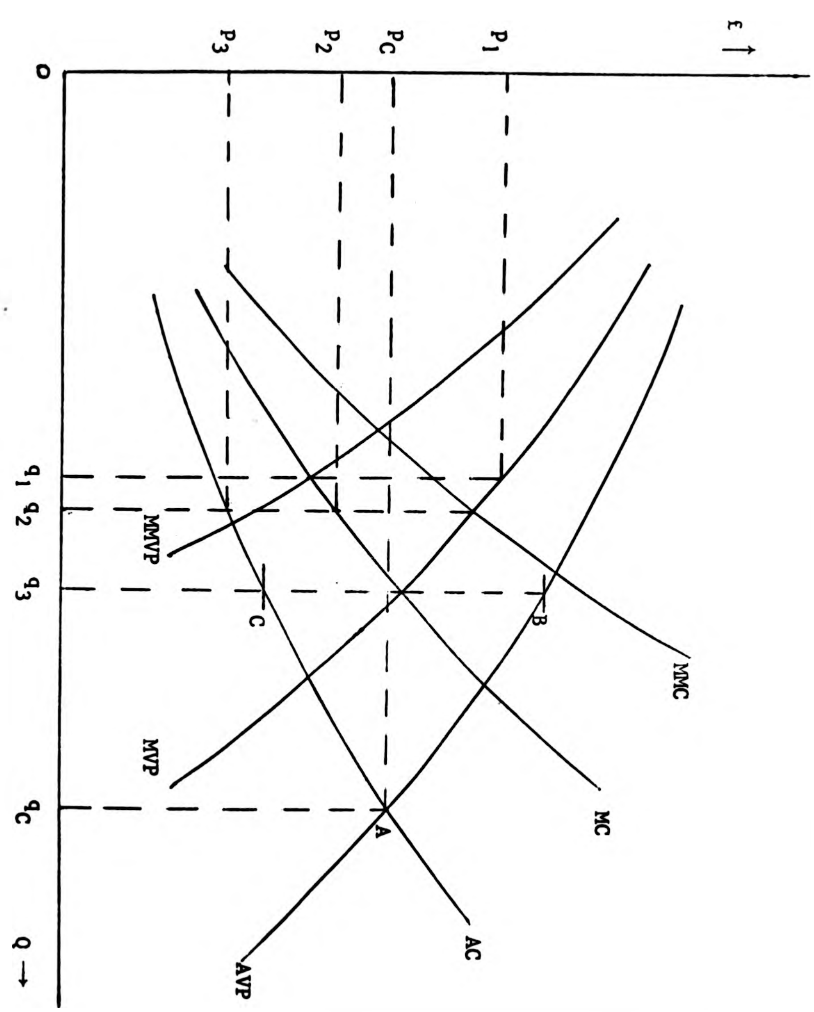
What then of bilateral oligopoly?

II The Nature of the Indeterminacy under Bilateral Monopoly:[2]

We start by considering Hicks' first case. In doing so we follow Morgan (1949) in his clear exposition of the problem; he gives a full list of the assumptions required. Let us consider that there are two industries, the iron ore industry and the steel industry. There are very many firms in both industries, yet trading is performed on an industry-wide basis as if by a monopolist and a monopsonist. Everywhere else in the economy perfect competition obtains. Expansion of the ore and steel industries leads to rising prices for their factors of production (other than the ore exchange price) and falling prices for steel, but there are no technical economies or diseconomies present. Thus the average and marginal value products of ore (to the steel industry) curves are downward sloping whilst the average and marginal cost curves of ore production are upward sloping (see Diagram 1). Of course, as Morgan points out, AVP is not the true average value product curve since, as the price of ore changes, the proportions in which inputs are used changes so that average physical productivity changes. Given the ambiguity, it might possibly be better if we thought of the inputs as being used in constant proportions so that average physical productivity may remain constant,[3] though then the

DIAGRAM 4.1

Possible outcomes of trade between the iron-ore and steel industries



This diagram is substantially a redrawing of Morgan's (1949 p.374) where he discusses the same problem. Labelling is discussed in the text.

marginal physical productivity curve disappears. We discuss such a model later in this section.

Leaving aside this consideration, we turn to the diagram to discover the various possibilities. If we first allow that both sides of the market are atomistic then although it might seem that the industry would set marginal cost equal to marginal value product, this is not in fact the long run equilibrium. For firms may enter the industry on either side. As a consequence the supply curve for ore becomes the average cost curve, since firms will enter until output is driven up to the point at which both average and marginal cost for the firm equal price paid. Similarly, the average value product curve becomes the demand curve and industry equilibrium takes place at A.

Now if the ore monopoly (selling organisation) could set price while the steel monopsonist would accept this then the steel monopsonist is a price taker. Given a certain ore price, he would buy up to the point at which price paid is equal to marginal value product. His demand curve is therefore the marginal value product curve. The ore monopoly of course sets marginal revenue equal to marginal cost and thus equates the curve marginal to the marginal value product curve (MMVP) to marginal cost, with exchange of p_1q_1. In like manner, if the ore monopoly was a price taker, so that the steel monopsonist could set price, then the supply curve of the ore monopoly becomes MC. Thus the steel industry equates the curve marginal to this, MMC, to its marginal revenue which accords to the marginal value product (marginal physical productivity times marginal revenue which is price). In this case the choice would be point p_2q_2. (It is clear that q_1 is not necessarily less than q_2 in all cases.)

Finally, it is possible that the joint profit maximising position could be achieved. Here marginal revenue to the steel industry would be equated to marginal cost for the ore industry, and q_3 units would be exchanged. However, the price is not fixed, since both firms may make profits anywhere between points B and C on the diagram; the firms must bargain between themselves as to price, in some way not specified by the theory.

In summary on this simple case, we may note that the industries have opportunities similar to those of duopoly open to them. They may either co-operate and maximise joint profits (in which case though, we agree with Scherer that price remains to be determined) or they may attempt to act partly independently, and somehow reach an equilibrium with more or less co-operation. Notice that, in contrast to the duopoly case, independent action tends to result in lower output than maximising joint profits. However, there seems no reason why, for certain categories of quasi-independent action, the solution need be indeterminate. The position at which we normally start the analysis of a Cournot duopoly path to equilibrium is one where neither price nor output are equal.

Of course the situation where a monopoly buyer has no product market power is somewhat unrealistic. However we may perform the same type of analysis assuming that the monopsonist is also a monopolist, that is in Joan Robinson's terminology, the buying firm is a monemporist. We have now to relabel the average and marginal value product curves as the average and marginal revenue products respectively, and of course average value product will lie above these. With this difference, the range of outcomes is much as before. One point should be noted: the maximum profit that the monemporist can obtain is given by exchange at point C on the (appropriately relabelled) diagram 1. He cannot therefore gain extra profit by equating the curve marginal to the ore monopolist's "supply" curve, MMC, to his demand curve while paying p_2 (or even p_3). The total maximum area of profit available is the same whether or not the buying monopolist has monopsony power also.

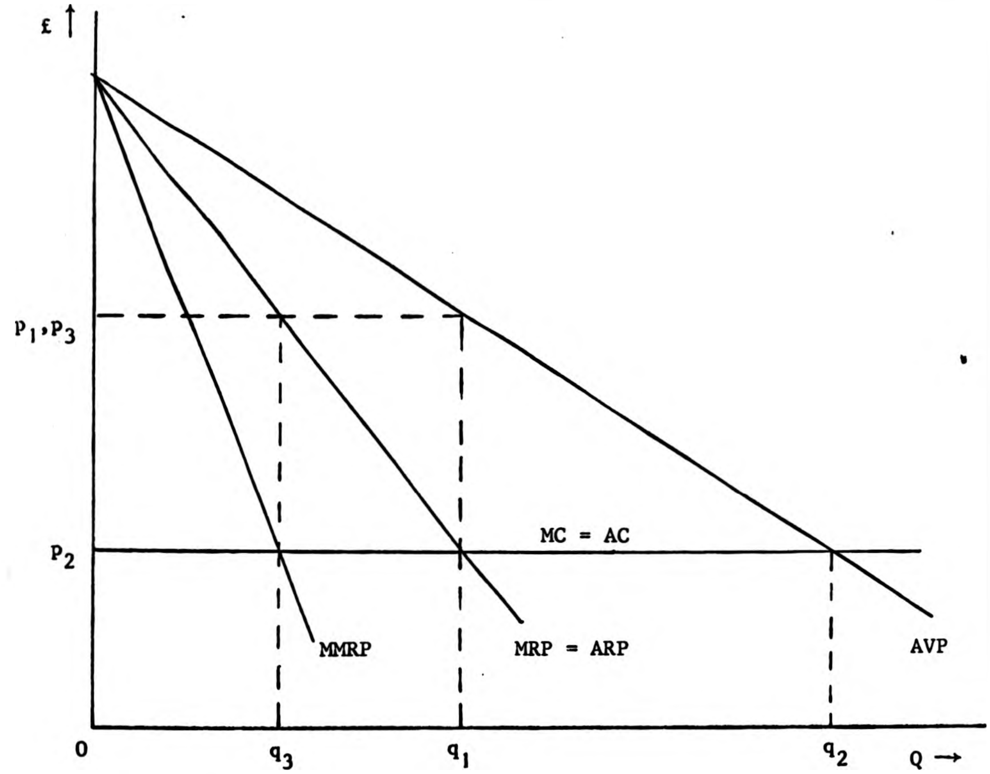
A similar situation exists even in the case where average costs are invariant with output, so that no opportunity for the exercise of monopsony power exists. This is the case discussed by Scherer (1970) and illustrated in our diagram 2; the additional assumption that input is converted in fixed proportions into output is utilised. A competitive industry's demand would be along AVP so that the factor monopolist would wish to sell q_1 units at price p_1. A product market monopolist would prefer to buy q_1 units at a price of only p_2, while a factor market monopolist selling to a monopolistic industry taking price as given would like trade of q_3 units at a price p_3. The joint profit maximising solution is exchange of q_1 units at some price between p_2 and p_1. Again, individualistic action seems to dictate an exchange of output units less than or equal to the joint profit maximising exchange, and under joint profit maximisation itself, price is indeterminate.

III Approaches to Solutions:

There have been a large number of authors eager to solve the previously noted indeterminacy problem by splitting the proceeds of bilateral monopoly power between the participants. Many of these studies have had as their prime concern the union-management bargaining process. Because of this they tend to be mainly concerned about institutional factors.[4] While we shall touch obliquely upon such studies, our main concern in this brief review will be the more

DIAGRAM 4.2

A simpler bilateral power situation



Here average cost is invariant with output, so no opportunity for the exercise of monopsony power exists. AVP is a competitive industry's demand for the input, which is used in fixed proportions. MRP is the curve marginal to that, and MMRP the curve marginal to that. Output q_1, q_2, q_3 correspond to prices p_1, p_2, p_3 and are explained in the text.

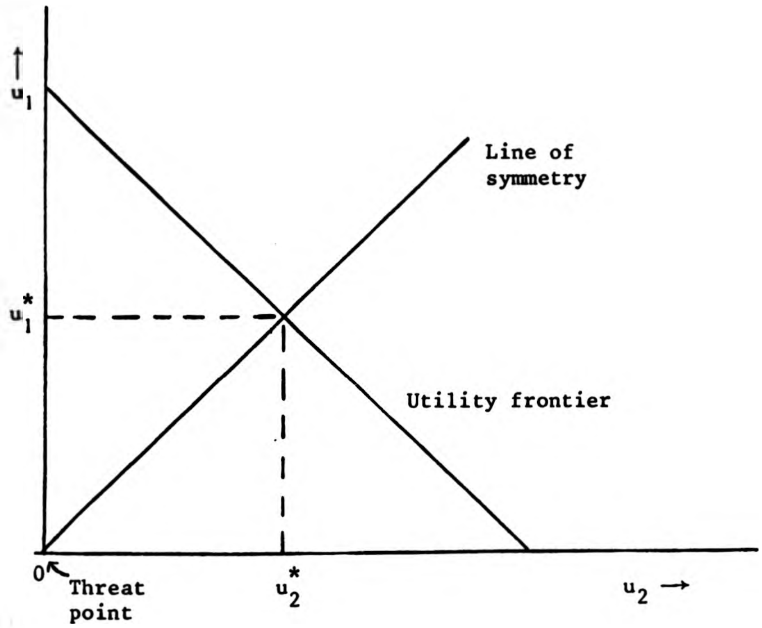
analytical studies with somewhat wider relevance. Most of the work in this (narrower) field has been concerned with solving the indeterminacy under joint-profit maximisation bilateral monopoly. One notable but largely neglected [5] exception is Cournot's approach (1927 Chapter IX), slightly refined by Zeuthen (1930 Chapter III). This will receive detailed consideration in later sections; at present we shall discuss the joint profit maximisation "solutions".

Both Bishop (1963) and de Menil (1971) review parts of this literature and, as they imply, much of it is game theoretic in approach. In the language of the Theory of Games, Bilateral Monopoly is normally considered to be a two-person co-operative non-zero sum game, with fixed-threat bargaining. It is non-zero sum because there are gains from trade and is fixed threat since the most obvious and compelling threat is to refuse trade. As von-Neumann and Morgenstern (1953) analyse the game, there is a solution given by the contract curve, or the line BC in our diagram 1. This is of course not a unique solution since any point on the contract curve will do as well as any other, so that we have come no further than orthodox theory will take us. However, many other game theorists are willing to add extra assumptions in order to prescribe a unique outcome and we shall describe some of their models in outline - detailed consideration and critique would seem unnecessary given the work of Bishop and de Menil.

The most famous of solutions is undoubtedly that of Nash (1950). At the joint-profit-maximising output we may define an objective-payoff frontier in profit space, a straight line of unit downward slope defines each player's profit. From this we map to utility space[6] to derive the utility frontier using the players' utility of profit functions. Discussion then proceeds on the basis of choosing a point on this utility frontier (drawn for a special case in diagram 3),

DIAGRAM 4.3

The Nash Bilateral Power solution



The solution maximises the product of the utilities of the two players at points u_1^*, u_2^*.

which is scaled to set the threat point at the origin for both players. Besides assuming efficiency, Nash uses three other assumptions to provide a solution, now known as the Symmetry, Transformation Invariance and Independence of Irrelevant Alternatives axioms. The first asserts that if the utility frontier is symmetrical with respect to u_1 and u_2, so that we can change the axes without altering the functional form, then the solution gives equal utility (measured from the threat point) to both players. Transformation Invariance allows for order-preserving linear transformations of utility to occur without altering the solution. If the utility frontier is now unfavourably altered at any but the above given solution point, this does not alter the outcome by the third axiom. These axioms are sufficient to establish the Nash solution which involves maximising the product of the two players' utilities.

Now if we are willing to restrict ourselves to utility functions which are linear in profit for both parties, we may dispense with the third axiom, since the utility frontier is always a straight line. We might well want to make this restriction when talking about pure profit-maximising protagonists. If a firm wishes to maximise profit then naturally profit is the only argument in its utility function. Further, a pure profit maximiser should be indifferent between receiving £A from source a plus £B from source b as compared with £A + B from source c, or again between receiving £A with certainty and a 50% chance of receiving £2A. Thus a pure profit-maximiser must have a utility function linear in profit, so that if both firms are pure profit-maximisers the utility frontier becomes linear (as in diagram 3). In this case, the first two axioms are sufficient to establish the solution where equal utility increments and equal profits are received by each player. We shall continue by considering only pure profit-maximising firms.

Given this assumption, we turn to a consideration of Raiffa's four solutions (1957). Of these, one is equivalent to that of Nash, and two others are also equivalent unless for some reason the utility frontier does not continue uninterruptedly from one axis to the other. In the latter two theories Raiffa effectively rejects Nash's third axiom. This makes no difference unless we consider that, while being pure profit-maximisers, one or both of the firms is subject to some external constraints; Bishop gives the example of minimum wage legislation, yet if such legislation were general it would surely be built into the underlying cost curves. Raiffa's final scheme departs more radically from the Nash solution, since for this solution he rejects the transformation invariance axiom; once we have done this we can say very little about the actual outcome.

A further approach along similar lines to the above is the Zeuthen (1930 chapter IV) theory, taken up more recently and put into a game-theoretic framework by Harsanyi (1956). While the basis of Zeuthen's theory is to be found in his desire to explain the bargaining process in terms of successive concessions rather than a wish to develop a determinate theory from a small number of axioms, Harsanyi in fact shows that (at least as he interprets Zeuthen) the solution point is identical to the Nash point - that is as given by the product of the utilities of the two parties measured from the threat point.

Foldes (1964) has a rather different approach to finding a determinate outcome under joint-profit-maximisation bilateral monopoly. His model has as its analytical basis a Hicksian-type theory of bargaining where determinacy is obtained by taking into account time preferences, or threats of delay before trade can take place. Such threats can be used to extract concessions. Thus, his bargainers have utility functions with arguments of profit share obtained and time before agreement is reached. He shows that the point at which agreement is reached is that point where:

$$\frac{\partial u/\partial t}{\partial v/\partial t} = -\frac{\partial u/\partial \Pi_1}{\partial u/\partial \Pi_2}$$

u and v being the first and second bargainers' utility functions with arguments t and Π_1 and t and Π_2 respectively. In a simple example where utility is a linear function of discounted profit flows, Foldes finds that

$$\frac{\Pi_1}{\Pi_2} = \frac{s}{r},$$

r being the first bargainer's time rate of discount, and s being the discount rate for the second firm. That is, the smaller the firm's discount rate, the larger its profit share.

In fact Bishop (1964), in deriving a "Zeuthen-Hicks" theory of bargaining, obtains precisely the same result though by different means as in the above example of the prediction from Foldes' theory. Bishop shows that his solution involves maximising $u(v)^{r/s}$; in the special case where r = s the outcome is of course identical to that which Nash prescribes. Actually, there seems in general no reason why the discount rates of the two actors involved should differ systematically in the absence of any possibility of entry. If this is indeed acceptable then both the immediately preceding theories should lead to a result identical to the half-way rule obtained from the simple Nash case considered earlier.[7] In order for the discount rates to differ we would presumably have to introduce liquidity constraints into the analysis.

A more recent paper by Spindler (1974) which is also concerned with finding a determinate solution under bilateral monopoly, reaches the conclusion that joint profit maximisation with each firm taking half the profit available is the most reasonable outcome. His argument is that the bargaining power of each firm can be directly measured by the degree of economic dependence of the other firm, which dependence is linearly related to the amount of profit to be received, since each extra unit of profit has the same value. Both firms are in a position to offer a share of the same total profit area initially, thus both have equal bargaining power initially. Whenever one or the other firm has a greater relative bargaining power it is less dependent on the other firm and so should be able to extract concessions. The natural outcome of such a process is that bargaining power and economic dependence are equalised when both firms receive a half share of the profits available.

While discussing joint-profit-maximisation bilateral power models, we should also mention that in Stigler's (1964) theory of oligopoly, buyers enter the picture explicitly. However, Stigler specifically does not wish to discuss situations where the number of buyers are few,[8] so that we shall not consider his paper further in the present context.

We could of course criticise the assumptions underlying the above models especially those based on the axiomatic approach; Bishop (1963) has an extensive critique along these lines. It is probably more useful though, to accept that despite the widely differing starting points of the theories of determinacy the conclusion that two monopoly firms facing one another should share equally the profits accruing by joint maximisation would find widespread approval. This solution appears eminently reasonable unless we accept that there are systematic factors which would tend to make either one or the other party predominant in any bargain. Having said this, we should perhaps note that such a solution could be considered as either a normative or a positive conclusion. Again Bishop considers this question in some detail, and while the axiomatic approach implies some positive theoretic value in the solution, Raiffa for example believes that his solutions are normative in the sense that they are alternatives which a "fair arbitrator" might suggest. To the extent that such a fair arbitrator seems to occupy a similar post to that of the Walrasian auctioneer we may well consider that the solution outlined has positive content.

IV Extensions to the Joint-profit-maximisation solution:

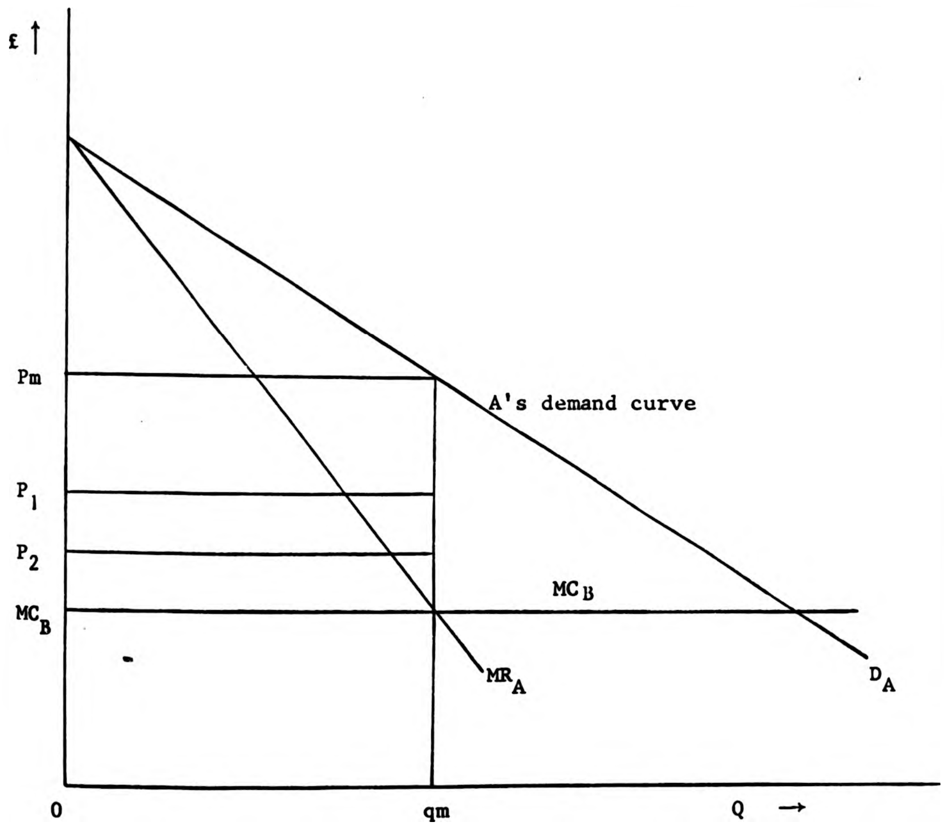
It is when we move from considering bilateral monopoly under joint profit maximisation to situations more commonly obtaining in the "real world" that we get into some difficulty in extending the solution discussed above. In fact work in this area is almost non-existent, so that when de Menil, for example, moves towards his empirical work he states that:

"since the Nash model refers to bargaining between one employer and one union the equation had to be tested on highly unionised industries." (p.51)

We should first consider extensions to the fixed proportions model described above when there is more than one firm on each "side" of the bargain.[9] Assume for example that we have a monopoly buyer A, who is also a monopoly seller in a market where customers have no buying power. A purchases inputs from level B and transforms them costlessly and linearly into his output. At level B there are two firms who buy all their inputs from price-takers (see diagram 4). We can immediately isolate polar cases. It might be the case that A is able to play the two firms at B off one against the other for the contract to supply q_m units of their product to him. Trade will then

DIAGRAM 4.4

Polar cases in a "two against one" situation under joint profit maximisation



A is a monopoly buyer and monopoly seller who purchases inputs from level B, transofrming them costlessly and linearly into his own output. Thus final (and intermediate) product output is at level qm while bargaining goes on our price paid. MC_B is the cost curve for both B firms.

take place at a price equal, in our diagram, to MC_B. Alternatively the same two firms might tacitly be able to collude and act as a monopolist in which case, as we saw earlier, trade should take place at p_1 (halfway between p_m and MC_B), with each firm supplying half of the required output. Further, given that collusion is possible,[10] neither of the firms at level B should be persuaded to sign an exclusive contract at any price below p_2 (half-way between p_1 and MC_B) since more profit can be obtained by collusion with the other B firm.

Even with this very simple example it can be seen that once we move to considering the effects of having more than one firm on either side of the bargain then we have to solve an oligopoly problem. While this is not the normal type of oligopoly problem in that total output is determinate we have similar polar cases and presumably a solution akin, for example, to the Cournot solution. It would seem legitimate that there can be this range of possible outcomes, since all that we need for joint profit maximisation is (tacit) agreement on the total traded quantity and the ability of the A sector to impose the profit maximising price on the final consumers. Collusion within the A sector about final price need not imply collusion regarding the price paid for inputs (or vice versa). One point, discussed earlier in the chapter on oligopoly models, arises again here: the range of outcomes in bargaining between sectors A and B is circumscribed by the ability of any one firm or small group to tender for an exclusive contract, so that again inequality of firm size is important. Some type of numbers equivalent measure of concentration would seem relevant here for this reason. Thus we might say, in a rather ad hoc manner, that the level of the traded price (and so the level of price-cost margins on both sides) will depend on the equivalent numbers of firms on both sides of the market and the extent of their realised interdependence.[11]

We now leave this two sector model at its (limited) stage of development and proceed to further considerations which are required for our model to provide even a reasonable mirror on the real world. First let us take the situation where A is supplied by two monopolists providing products which are perfect complements in the production of A's product (say labour and capital used in fixed proportions, for example). In this case any of the three parties can threaten to halt the operations of the other, assuming no other sources of factors or outlets. Thus the only effective coalition involves all three parties and under the Nashian scheme the product of their utilities becomes the effective maximand. By symmetry the solution involves each player obtaining a third of the profit maximising surplus even if, for example, one input is a relatively unimportant part of the product. In extending the model in this manner, we run into one of Bishop's criticisms of the model. He notes (p.579) that if a previous monopoly union were to split itself into n constituent skill unions then the Nashian arbitrator would award each party $1/(n+1)$ of the total profit, whereas previously Labour as a whole obtained only one half of this surplus. The most obvious, though not fully satisfactory, answer to this point is to say that if indeed each of the separate skills is indispensable then the monopoly union obtaining initially was not an optimal structure for that particular industry.

At the other end of the spectrum, if the two factors are perfect substitutes then of course the bargaining power of the input sellers is only that of a duopoly. This indicates that in the case where factors are imperfect substitutes so that we have the possibility of variable proportions some solution between these polar extremes should be achieved, though it must be admitted that the limits are potentially extremely wide.

Another complication which should be introduced into the scheme is to allow sellers of factors to supply several sectors of industry with intermediate products, rather than just supplying sector A. It is of course possible that the firm(s) at B can obtain a large share of the profits of each of these sectors. This is however not necessarily the case unless B is able to practice discriminatory pricing as between the different types of uses to which each such sector puts its product and can prevent significant resales. It may therefore be the case that B is forced to sell at a common price to all users dependent on its general market power vis a vis the power of its buyers.

One rather important feature of our economy which causes difficulties for the joint profit-maximising approach (and maybe also others) is that many of its products pass through, not just two, but a whole chain of manufacturing stages on their way to the final consumer. This means that we ought to consider whether the joint-profit-maximising type of bargaining and share-out approach is applicable to a "chain of monopolies" situation where the total share is that area between final price and cost levels at source, or whether some prices are necessarily parametric. In other words we should open up the possibility that the margin in industry B is determined not only by the power of sector(s) A but also sellers to B and sellers to these industries. A related problem is the extent to which strong intermediate-product suppliers can push up prices despite a weak final market sector. When we bring such matters into consideration then a partial equilibrium approach becomes increasingly difficult to maintain, yet any general equilibrium framework for the type of economy visualised here would necessarily be extremely complex.

Rather than developing the model outlined above in any of the directions indicated we shall turn to the alternative formulation proposed by Cournot (1927, Chapter IX) and discuss extensions to his model in some detail. In doing so we shall find that the considerations noted above also enter into such an enlarged Cournot model, and we shall be content to provide fuller discussion of them in that context. Once the Cournot model has been formulated and its ramifications explored, we shall be in a position to compare the predictions it provides with those of the joint-profit-maximising model and thereby reach some conclusions on bilateral power models in general. Unfortunately, unlike the pure oligopoly case discussed in an earlier chapter, it proves very difficult to say anything very concrete about the possible performance of a general bilateral power model where joint-profit-maximisation and Cournot behaviour are to some extent polar cases.

V Cournot's Model[12]

Cournot begins his analysis with a copper monopolist and a zinc monopolist who must both sell to a perfectly competitive brass industry, for there are no other uses for copper and zinc. The brass industry is able to convert these two inputs costlessly into brass and we assume one unit of each input produces one unit of output.[13] Further, entry into the copper or zinc industries is impossible. In fact, the problem of potential entry is not dealt with in this chapter, in the sense that we either assume barriers to entry high enough or assume that capacity is maintained (in a Spencian manner)[14] high enough to remove any threat to established firms' short run profit-maximising behaviour.

Profits in the brass industry are given by:

$$\Pi_{Bi} = p_B q_i - p_c q_i - p_z q_i \qquad \text{for each firm } i = 1,2,\ldots, n$$

First order conditions for profit maximisation then imply for each firm:

$$\frac{d\Pi_{Bi}}{dq_i} = p_B - p_c - p_z = 0 \qquad (1)$$

Thus, the (identical) derived demands for copper and zinc are given by:

$$q = f(p_B) = f(p_c + p_z)$$

The copper monopolist's profits are given by:

$$\Pi_c = p_c q - c_c(q) - F_c$$

Now, Cournot assumes the copper monopolist maximises profits with respect to price not quantity, since his output must be identically equal to that of the zinc firm and the brass industry. This means that we may write

$$\frac{d\Pi_c}{dp_c} = q + p_c \frac{dq}{dp_c} - c'_c \frac{dq}{dp_c} = 0^{15} \qquad (2)$$

Here, $\dfrac{dq}{dp_c} = \dfrac{dq}{dp_B} \cdot \dfrac{dp_B}{dp_c} = \dfrac{dq}{dp_B}\left(1 + \dfrac{dp_z}{dp_c}\right)$

But Cournot assumes

$$\frac{dp_z}{dp_c} = 0$$

from the Copper man's point of view, that is the Copper monopolist assumes that changing his price will have no effect on the price set in the zinc industry.

Similarly, the zinc monopolist assumes:

$$\frac{dp_c}{dp_z} = 0$$

We have then, from (2)

$$(p_c - c'_c)\frac{dq}{dp_B} = -q \qquad (3)$$

or

$$\frac{p_c - c'_c}{p_c} = \frac{-q}{p_c}\cdot\frac{dq}{dp_B} \qquad (4)$$

and similarly for the zinc man.

From (3) and its equivalent for zinc, with the aid of (1):

$$\frac{p_B - c'_c - c'_z}{p_B} = \frac{-2q}{p_B}\cdot\frac{dp_B}{dq} = \frac{2}{|\eta_B|} \qquad (5)$$

where $|\eta_B|$ is the elasticity of demand for brass. (5) is Cournot's result, but we shall instead follow the direction in which (4) takes us, and attempt to obtain price-cost margins for copper and zinc in this manner. Note that (4) may be rewritten as

$$\frac{p_c - c'_c}{p_c} = \frac{1}{|\eta_B|}\frac{p_B}{p_c}$$

and what we would want to do is to interpret the right hand side in terms of the elasticity of demand for the copper monopolist's product.

Now, this is not a model of Bilateral Monopoly in the modern sense; previously the term had two meanings.[16] But Zeuthen (1930) uses the model interchangeably in both the case of a copper-zinc-brass situation, and the situation where a monopoly seller faces a monopoly buyer, and states that "If we consider actual examples, we have in all essentials the same case (as the former) when two monopolistic concerns face each other as buyer and seller" (p.64). The easiest way to see this is if we consider that the copper monopolist is integrated vertically with all the brass producers, but that now brass competes with other products in a perfectly competitive market. The copper owner-brass producer now has monopoly power in that he provides the only use for zinc (and has a monopoly in copper). Even if brass had no substitute so that he provided both the only use for zinc and held a monopoly in the manufacture of brass and kindred products, the situation should be no different, since a monopolist may take his profit only once. He may charge himself a monopoly price for copper, <u>or</u> charge himself cost-price for copper and charge a monopoly price for brass: to do both would reduce output to such an extent that profits would be reduced.[17] We proceed to develop the Cournot model of bilateral monopoly assuming in effect that the Copper monopolist is now vertically integrated with the brass industry.

VI <u>Our Basic Cournot Model</u>:

We have, for the brass and zinc firms respectively:

$$\Pi_B = p_B q - p_z q - c_c(q) - F_B \qquad (6)$$

$$\Pi_z = p_z q - c_z(q) - F_z \qquad (7)$$

Again, profits are maximised with respect to price, but the Cournot assumptions have to be modified slightly. We take it that B assumes that

$$\frac{dp_z}{dp_B} = 0;$$

if B alters his price then this will not, in his opinion, affect z's price.[18]

The equivalent assumption for z is that

$$\frac{dp_B}{dp_z} = 1.$$

This is again in the spirit of Cournot's "quasi-competitive" assumptions; his firms assume that their actions will have an effect on the others equivalent to the actual effect if the firms were all in perfectly competitive markets. For, if z were to raise his price to the firms in a perfectly competitive industry, then their marginal costs would rise by that amount and so too their prices. Incorporating these assumptions in the maximisation of (6) and (7) yields the following equations:

$$\frac{d\Pi_B}{dp_B} = q + p_B \frac{dq}{dp_B} - p_z \frac{dq}{dp_B} - c'_c \frac{dq}{dp_B} = 0 \qquad (8)$$

and
$$\frac{d\Pi_z}{dp_z} = q + p_z \frac{dq}{dp_B} - c'_z \frac{dq}{dp_B} = 0 \qquad (9)$$

Adding these two equations above gives:

$$2q + (p_B - c'_c - c'_z) \frac{dq}{dp_B} = 0$$

or
$$\frac{p_B - c'_c - c'_z}{p_B} = \frac{2}{|\eta_B|} \qquad (10) \text{ (identical to (5))}$$

Forming predictions in terms of B's and z's individual margins, we note from (8) that:

$$\frac{p_B - p_z - c'_c}{p_B} = \frac{1}{|\eta_B|} \qquad (11)$$

This means that B's price-cost margin remains unchanged when faced by a monopoly seller, because he treats that firm's prices as given in making his decisions. This feature of the Cournot model becomes quite important when we turn to the empirical tests.

Equation (9) is more difficult to interpret initially. We have:

$$\frac{p_z - c'_z}{p_z} = \frac{-q}{p_z} \cdot \frac{dp_B}{dq} = \frac{1}{|\eta_B|} \cdot \frac{p_B}{p_z} \qquad (12)$$

Now, we know that the demand for z's product by B is given by B's net marginal revenue curve. Thus,

$$p_z = MR_B - c'_c = p_B\left(1 - \frac{1}{|\eta_B|}\right) - c'_c$$

Therefore,

$$\frac{1}{\eta_z} = \frac{q}{p_z} \cdot \frac{dp_z}{dq} = \frac{\left(\frac{d\,MR_B}{dq} - c''_c\right) \cdot q}{MR_B - c'_c} = \frac{MR_B}{p_z}\left(\frac{q}{MR_B} \cdot \frac{dMR_B}{dq} - \frac{c''_c \cdot q}{MR_B}\right) .$$

If we make the assumptions that the marginal cost of production, c'_c, is constant, and that B faces constant elasticity of demand (both of which appear empirically plausible and are defended in Appendix I), then we may simplify this relationship greatly.[19] The latter implies the particularly simple result that the elasticity of B's marginal revenue curve is equal to the elasticity of his demand curve.

Thus the above relationship becomes:

$$\frac{1}{|\eta_z|} = \frac{p_B\left(1 - \frac{1}{|\eta_B|}\right)}{|\eta_B| \left[p_B\left(1 - \frac{1}{|\eta_B|}\right) - c'_c\right]} \qquad (12a)$$

Then from (12) and (12a), and the definition of p_z, z's margin is:

$$\frac{p_z - c'_z}{p_z} = \frac{1}{|\eta_z|\left(1 - \frac{1}{|\eta_B|}\right)} \qquad (13)$$ [20]

This is interesting, for it implies that, contrary to what might be thought, it is possible that if a firm faces a monopolist, then this may allow it to have a higher price-cost margin than if it were to face a large number of buyers. As can be seen from equations (10) and (11), what is actually happening here is that price rises very high and output is much reduced, for output is lower than it would be if there were only one monopolist in the market or if joint profit maximisation was the agreed policy. The situation in fact has a nice symmetry with Cournot's duopoly model, as can be seen in diagram 5 (overleaf).

VII Some Extensions to the "Cournot" model:

Having built the basic model, we seek to extend it. This is done in the following ways:

(i) We allow B to use other than one unit of z's output to make one of his own.

(ii) We consider the case where there is more than one firm of type B, and more than one of type z in the brass and zinc industries.

(iii) We allow firms B to buy from other industries than z.

(iv) We allow firms z to sell to other industries than B.

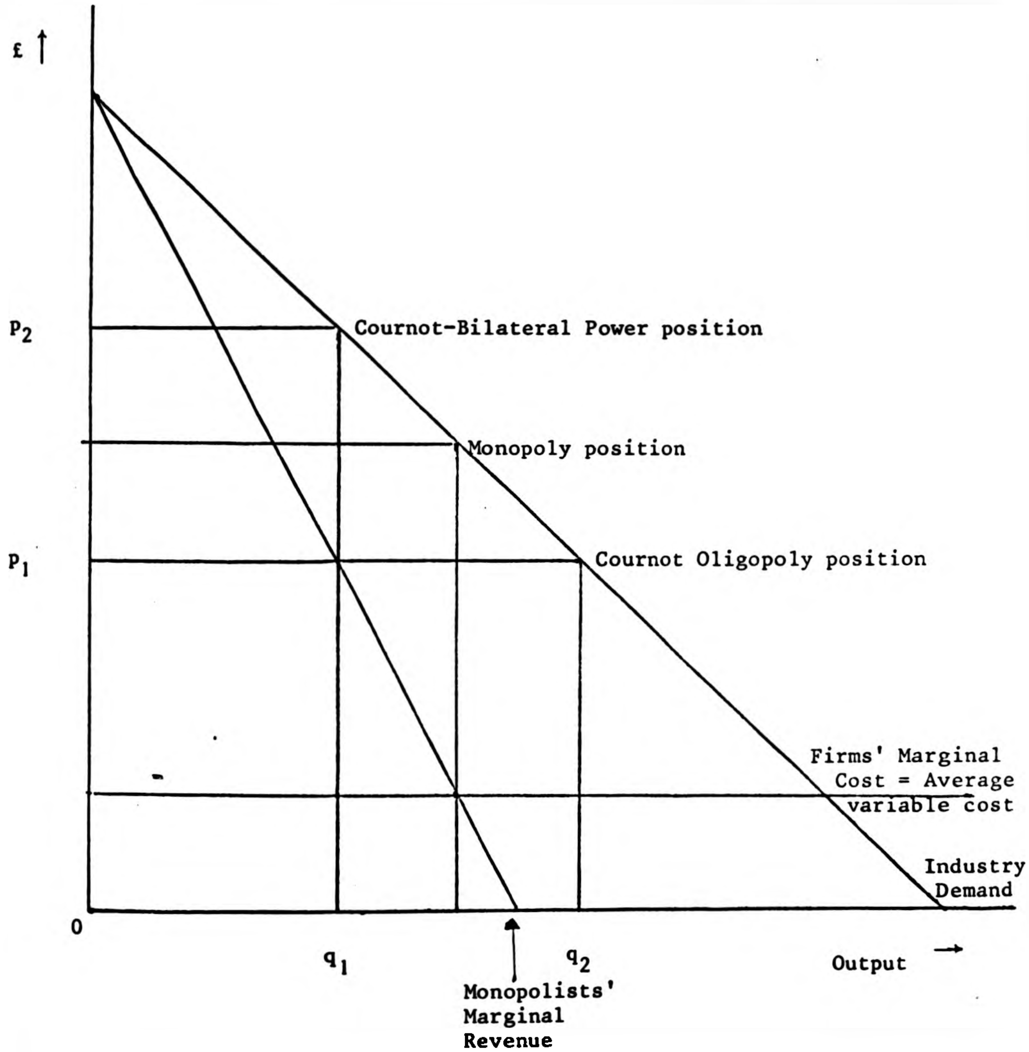
Some of these extensions require more theoretical development than others; we take them in the above order:

(i) Firstly consider the case where B buys say two units of z's product in the course of making one of his own; or more generally, $q_B = s_z q_z$. We then have: $\frac{dq_B}{dq_z} = s_z$ ($s_z = \frac{1}{2}$, for example), (for efficient production).

We also have to alter our assumptions about changing prices as far as z is concerned. If z raises his prices, B will have to raise his by $\frac{1}{s_z}$ (or twice) as much in order to leave profits unchanged. Thus, z's quasi-competitive assumption is now $\frac{dp_B}{dp_z} = \frac{1}{s_z}$.

DIAGRAM 4.5

The Cournot Bilateral Monopoly solution



For ease of draftsmanship we have used linear demand and marginal cost curves. The Cournot Oligopolists produce output $0q_1$ and q_1q_2 at price p_1. The Cournot Bilateral Monopolists produce output $0q_1$ sold by the first to the second at price p_1 and by the second at price p_2.

We have $\frac{d\Pi_z}{dp_z} = q_z + (p_z - c'_z) \cdot \frac{dq_z}{dq_B} \cdot \frac{dq_B}{dp_B} \cdot \frac{dp_B}{dp_z} = 0$

$$\text{or} \quad p_z - c'_z = -q_z s_z^2 \frac{dp_B}{dq_B} \tag{14}$$

The demand for z's product is now according to its net marginal revenue product, that is:

$$p_z = \left[(p_B + q_B \frac{dp_B}{dq_B}) - c'_c\right]\frac{dq_B}{dq_z} = s_z\left[p_B(1 - \frac{1}{|\eta_B|}) - c'_c\right] \tag{15}$$

Note that the elasticity of this marginal revenue product curve is still $|\eta_z|$ as in (12a). We again work in terms of $|\eta_z|$ in obtaining the price-cost margin equation.

From (14)

$$\frac{p_z - c'_z}{p_z} = -\frac{q_B s_z}{p_z}\frac{dp_B}{dq_B} = -\frac{q_B}{p_B}\frac{dp_B}{dq_B}\frac{s_z p_B}{p_z}$$

and substituting (15) and (12a) into this:

$$\frac{p_z - c'_z}{p_z} = \frac{1}{|\eta_z|(1 - \frac{1}{|\eta_B|})} \tag{16}$$

Thus, the extension of the model to a simple proportional relationship between B's and z's output, rather than a one-to-one relationship, does not alter z's price cost margin from its former value (and neither will it affect B's price-cost margin). This is to be expected as such a proportional relationship amounts really only to a change in units. We could continue by developing a model where there is a more flexible relationship between the two outputs; however this would yield somewhat intractable results, and in any case the data to be used (UK input-output tables) assumes constant proportions, so that there seems little point in developing the more complex model.

(ii) Assume now that there are m, rather than one, purchasing firms in the B industry, and n selling firms in the z industry. We then have, for individual firms:

$$\Pi_{Bi} = p_B q_{Bi} - p_z q_{zi} - c_{ci}(q_{Bi}) - F_{Bi} \qquad (17) \qquad i = 1,2,..,m$$

$$\Pi_{zj} = p_z q_{zj} - c_{zj}(q_j) - F_{zj} \qquad (18) \qquad j = 1,2,..,n$$

We take it, with the support of result (16), that we may simplify by making $q_{zi} = q_{Bi}$ in (17), and that the results would be identical if $q_{Bi} = s_z q_{zi}$. We further assume that each of the i firms in the B industry act as Cournot oligopolists towards each other, and that the j firms in the z industry do likewise, and we retain the assumptions of the earlier model regarding relations between the B and z industries. A question however arises as to the appropriate control variable in the firms' maximising process; we assume it to be quantity.[21]

From (17) then (with $q_{zi} = q_{Bi}$):

$$\frac{d\Pi_{Bi}}{dq_{Bi}} = p_B + q_{Bi}\frac{dp_B}{dq_B}\cdot\frac{dq_B}{dq_{Bi}} - p_z - q_{zi}\frac{dp_z}{dp_B}\cdot\frac{dp_B}{dq_{Bi}} - c'_{ci} = 0$$

Under our assumptions, each B firm takes it that:

$$\frac{dq_B}{dq_{Bi}} = 1; \qquad \frac{dp_z}{dp_B} = 0$$

so that we have:

$$p_B - p_z - c'_{ci} = -q_{Bi}\frac{dp_B}{dq_B}$$

Multiplying by q_i ($\equiv q_{Bi}$):

$$p_B q_i - p_z q_i - c'_{ci}\cdot q_i = -q_{Bi}^2\frac{dp_B}{dq_B}$$

and summing over the m firms:

$$\Sigma p_B q_i - \Sigma p_z q_i - \Sigma c'_{ci} \cdot q_i = -\frac{dp_B}{dq_B} \sum q_{Bi}^2$$

$$\text{or } \frac{\Pi + F}{R}_B = \frac{\Sigma p_B q_i - \Sigma p_z q_i - \Sigma c'_{ci} \cdot q_i}{p_B q} = -\frac{dp_B}{dq_B} \cdot \frac{q_B}{p_B} \cdot \sum \frac{q_{Bi}^2}{q_B^2} = \frac{H_B}{|\eta_B|} \tag{19}$$

That is, the industry's price-cost margin becomes equal to the ratio of the Herfindahl index of industry concentration to the elasticity of demand as in the ordinary Cournot oligopolistic market. We write $\frac{\Pi + F}{R}$ as the margin for brevity.

Taking (18):

$$\frac{d\Pi_{zj}}{dq_{zj}} = p_z + q_{zj} \frac{dp_z}{dq_{zj}} - c'_{zj} = 0$$

so that:

$$p_z - c'_{zj} = -q_{zj} \frac{dp_z}{dp_B} \cdot \frac{dp_B}{dq_B} \cdot \frac{dq_B}{dq_z} \cdot \frac{dq_z}{dq_{zj}}$$

and making the assumptions for the z firms that:[21a]

$$q_z = q_B, \quad \frac{dq_z}{dq_{zj}} = 1, \quad \frac{dp_B}{dp_z} = 1 \quad (\text{so that } 1 / \frac{dp_B}{dp_z} = 1),$$

we have:

$$p_z - c'_{zj} = -q_{Bj} \frac{dp_B}{dq_B}$$

Multiplying both sides by q_{Bj} ($= q_{zj}$), and summing:

$$\Sigma p_z q_{Bj} - \Sigma c'_{zj} q_{Bj} = -\sum q_{Bj}^2 \cdot \frac{dp_B}{dq_B}$$

$$\text{or } \frac{\Pi + F}{R}_z = -H_z \frac{dp_B}{dq_B} \cdot \frac{q_B}{p_z} = \frac{H_z}{|\eta_B|} \cdot \frac{p_B}{p_z} \tag{20}$$

The demand for z is again according to its net marginal revenue product, i.e.

$$p_z = p_B\left(1 - \frac{H_B}{|\eta_B|}\right) - c'_c \,.^{22}$$

Thus: $\eta_z = p_z \eta_B / p_B \left(1 - \frac{H_B}{|\eta_B|}\right)$, so that in (20), with similar manipulations to our previous cases:

$$\frac{\Pi + F}{R}_z = \frac{H_z}{|\eta_z| \; (1-\frac{H_B}{|\eta_B|})} \tag{21}$$

Thus, for z both the number of firms in the z industry, and the number of firms in the B, or buying, industry matter, in contrast to the situation in the B industry.

(iii) Allowing B firms to buy from other industries than z does not affect either B's margin or z's margin. To see this, assume B buys also from an industry y, and replace (17) by:

$$\Pi_{Bi} = p_B q_{Bi} - p_z q_{zi} - p_y q_{yi} - c_{ci}(q_{Bi}).$$

- No essential change occurs in either (19) or (21), the price-marginal cost margin remains as it was.

(iv) Despite this, the situation where z sells to more than one industry does have some effect on z's margin, though not on B's (since B takes z's price as given in making his calculations). The effect on z's margin comes about because he now takes into account that other industry as well. The formula yielded is in fact a fairly straightforward generalisation of (21).

We take only the case where price does not differ between the two (or more) markets z sells to, so that discrimination is not possible. The model we build only has two buying industries A and B, but the generalisation to many markets is obvious.

We have then, for z, assuming he is a monopolist:

$$\Pi_z = p_z \cdot (q_A + q_B) - c_z(q_A + q_B) - F_z, \quad \text{where } q_z = q_A + q_B$$

He maximises profits with respect to q_A and q_B, sales to the A and B industries. Thus:

$$\frac{\partial \Pi_z}{\partial q_A} = p_z + q_z \frac{\partial p_z}{\partial q_A} - c'_z = 0 \qquad (22)$$

$$\frac{\partial \Pi_z}{\partial q_B} = p_z + q_z \frac{\partial p_z}{\partial q_B} - c'_z = 0 \qquad (23)$$

From (22) and (23) we immediately see that

$$\frac{\partial p_z}{\partial q_A} = \frac{\partial p_z}{\partial q_B},$$

as a condition of profit maximisation.

Multiplying (22) by q_A and (23) by q_B, and adding yields:

$$p_z.(q_A + q_B) + q_z(q_A \frac{\partial p_z}{\partial q_A} + q_B \frac{\partial p_z}{\partial q_B}) - c'_z.(q_A + q_B) = 0 \qquad (24)$$

We consider that q_A, z's sales to the A industry, produce q'_A units of A's output, so that $q'_A = s_A q_A$ for efficient use,

and similarly $q'_B = s_B q_B$.

Firm z then assumes for sales to A:

$$\frac{dp_A}{dp_z} = 1 / \frac{dq'_A}{dq_A} = \frac{1}{s_A}$$

and similarly for sales to B.

Now:

$$\frac{\partial p_z}{\partial q_A} = \frac{\partial p_z}{\partial p_A} \cdot \frac{dp_A}{dq_A} = \frac{dp_z}{dp_A} \cdot \frac{dp_A}{d\,q_A} = \left(1 / \frac{dp_A}{dp_z}\right) \cdot \frac{dp_A}{dq_A} = s_A \cdot \frac{dp_A}{dq_A}$$

as far as z's assumptions about A are concerned, and similarly:

$$\frac{\partial p_z}{\partial q_B} = s_B \frac{dp_B}{dq_B} \quad \text{from z's point of view.}$$

(z assumes A and B to act independently)

Thus in (24):

$$q_z p_z - c'_z q_z = -q_z \left(q_A s_A \cdot \frac{dp_A}{dq_A} + q_B s_B \cdot \frac{dp_B}{dq_B}\right)$$

or: $$\frac{\Pi + F}{R} z = -\left(\frac{q_A s_A}{p_z} \cdot \frac{dp_A}{dq_A} + \frac{q_B s_B}{p_z} \cdot \frac{dp_B}{dq_B}\right) \qquad (25)$$

Now, the demand for z's product from A is:

$$p_z = NMRP_A = \left[p_A \left(1 - \frac{H_A}{|\eta_A|}\right) - c'_A\right] s_A$$

(where c'_A are A's other marginal costs),

and from B:

$$p_z = NMRP_B = \left[p_B \left(1 - \frac{H_B}{|\eta_B|}\right) - c'_c\right] s_B .$$

$$\therefore \frac{\partial p_z}{\partial q_A} = \frac{dp_A}{dq_A} \left(1 - \frac{H_A}{|\eta_A|}\right) s_A$$ and similarly for B, in actual fact.

Substituting these relations into (25) yields:

$$\frac{\Pi + F}{R} z = -\left[\frac{q_A}{p_z} \cdot \frac{\partial p_z}{\partial q_A} \frac{1}{\left(1 - \frac{H_A}{|\eta_A|}\right)} + \frac{q_B}{p_z} \cdot \frac{\partial p_z}{\partial q_B} \frac{1}{\left(1 - \frac{H_B}{|\eta_B|}\right)}\right]$$

$$= -\left[\frac{q_A}{q_z} \cdot \frac{q_z}{p_z} \cdot \frac{\partial p_z}{\partial q_A} \frac{1}{\left(1 - \frac{H_A}{|\eta_A|}\right)} + \frac{q_B}{q_z} \cdot \frac{q_z}{p_z} \frac{\partial p_z}{\partial q_B} \cdot \frac{1}{\left(1 - \frac{H_B}{|\eta_B|}\right)}\right] \qquad (26).$$

Recall that

$$\frac{q_z}{p_z} \cdot \frac{\partial p_z}{\partial q_A} = \frac{q_z}{p_z} \cdot \frac{\partial p_z}{\partial q_B} .$$

We shall write these $= \frac{q_z}{p_z} \cdot \frac{dp_z}{dq_z} = \frac{1}{\eta_z}$, taking it that q_z can be considered an independent variable.

(i.e. $\frac{dp_z}{dq_z} = \frac{\partial p_z}{\partial q_A} \cdot \frac{dq_A}{dq_z} + \frac{\partial p_z}{\partial q_B} \cdot \frac{dq_B}{dq_z}$, but $\frac{dq_A}{dq_z} + \frac{dq_B}{dq_z} = 1$)

Then from (26):

$$\frac{\Pi + F}{R} z = \frac{1}{|\eta_z|} \left[\frac{q_A}{q_z(1 - \frac{H_A}{|\eta_A|})} + \frac{q_B}{q_z(1 - \frac{H_B}{|\eta_B|})} \right] \qquad (27)$$

This result is obviously easily capable of generalisation to the case where z sells to a great number of industries rather than two. Note that, since we have assumed that the price to every industry is the same, then

$$\frac{q_A}{q_z} = \frac{q_A p_z}{q_z p_z}$$

which is the proportion of revenue gained from industry A by z, and similarly for B.[23]

Equation (27) can further be generalised, without making any particularly severe restrictions, to the case where there are many firms of unequal sizes at level z. If we assume that for all i (firms):

$$\frac{q_{Ai}}{q_{zi}} = \frac{q_A}{q_z} \; ; \quad \frac{q_{Bi}}{q_{zi}} = \frac{q_B}{q_z} \qquad (28)$$

then the entire expression in square brackets in (27) becomes invariant with respect to the structure of the z industry. Now considering that there are n firms in the z industry, not all of equal size, we may for each firm write:

$$\frac{p_z - c'_{zi}}{p_z} = \frac{1}{|\eta_{zi}|} \cdot C$$

where C is the (constant) value in square brackets.

Thus:

$$\frac{p_z q_{zi} - c'_{zi} \cdot q_{zi}}{p_z} = - \frac{q_{zi}^2}{p_z} \frac{dp_z}{dq_{zi}} \cdot C$$

$$= -\frac{q_z^2}{p_z}\ \frac{q_{zi}^2}{q_z^2}\ \frac{dp_z}{dq_z}\ (1 + \lambda_i).C, \qquad (29)$$

where $\lambda_i = \frac{d \sum_{i \neq s} q_{zs}}{dq_{zi}}$, allowing for the firms in the z industry to entertain other reactions besides the pure Cournot reaction between themselves; if they are strict Cournot firms of course $\lambda_i = 0$. Equation (19) would appear to hold true as long as $\frac{dq_A}{dq_{Ai}} = \frac{dq_B}{dq_{Bi}}$ is true, but this is implied by the relations (28) anyway.

From (29) summing over the n firms:

$$\frac{\Sigma p_z q_{zi} - \Sigma c'_z q_{zi}}{p_z q_z} = -\frac{q_z}{p_z}\ \frac{dp_z}{dq_z}\ C \sum \frac{q_{zi}^2}{q_z^2}\ (1 + \lambda_i)$$

$$= \frac{C}{|n_z|}\ H_z\ (1 + u_z),$$

where $u_z = \sum \frac{q_{zi}^2}{q_z^2} \lambda_i$, [24]

and H_z is the Herfindahl index of concentration for industry z. Therefore:

$$\frac{\Pi + F}{R}\, z = \frac{H_z(1+u_z)}{|n_z|} \left[\frac{q_A}{q_z(1-\frac{H_A}{|n_A|})} + \frac{q_B}{q_z(1-\frac{H_B}{|n_B|})} \right] \qquad (30)$$

assuming that marginal costs are equal to average variable costs in industry z.

Equation (30), straightforwardly extended to the case where there are more than two buying industries is a fairly general Cournot-type formulation of the typical short-run profit-maximising industry. As such it is the basis for later empirical work. It is obviously not completely general, for other extensions do suggest themselves, but we consider these to be of second-order importance. For example, although z does not affect B's margin, and so does not affect any firms buying from B, B does himself affect the z industry margin, so that those firms which buy from B will affect B's margin and, through it, z's.

VIII Comparisons with the Joint-Profit Maximising model:

As was discussed earlier in the present chapter, this Cournot model is by no means the only theoretical model of bilateral power possible, but it is one model which yields quite firm conclusions with fair ease. These conclusions are of a rather surprising nature, as a moment's consideration of equation (30) will show. For example, as H_A increases there will a *ceteris paribus* increase in $\frac{\Pi + F}{R}z$. In general then, the suggestion is that it would be better to be faced by buyers with market power than by a competitive industry. It may well be that the model developed here is idiosyncratic in this respect.

For as Hicks (1935 p.18) points out, the determinacy of the Cournot model is assured by the assumption of the buyer that a change in his price does not affect the prices charged for inputs. This belief effectively disallows the possibility that the buyer has any monopsony power, for he treats input prices as parametric. Under certain circumstances such a model would not be unreasonable. For example there is no monopsony power independent of monopoly power available to a buyer who faces horizontal supply functions for all factors; the case mainly discussed above where the buyer uses inputs in fixed proportions and has average cost equal to marginal variable cost (as with Scherer's model) would fall into this category. However as has been suggested to me by John Kay, it remains true that the assumptions made in the course of developing the model may favour the suggested conclusion.[25]

One problem with disallowing monopsony power if the circumstances above do not hold is that we might normally expect monopoly power to be positively associated with monopsony power.[26] This causes difficulties in a more general model than Cournot's because the increase in z's margin predicted because of an increase in B's monopoly power may well be at least partially offset by a concommitant increase in B's monopsony power.

Unfortunately then, unlike the pure oligopoly case, it is by no means easy to generalise from the Cournot model to other plausible models of firm behaviour. For in general, B's margin could well be affected by the z industry whereas here it is not; z's margin is thus also affected by the firms from which it in turn buys. Whilst noting this caveat, and the subsequent limitations placed upon the model, let us turn to the important question of the extent to which the model developed above is representative of a general class of bilateral power models by considering its relationship to our ideas on the joint-profit maximisation model as outlined earlier. In performing this exercise we shall note that empirical work with an equation of the form of (30) above will use ratios of that equation at different times in the cross-section. Thus while the actual margin may well be different under different behavioural assumptions, the ratio of two margins at different time periods may vary much less with the particular behavioural assumption involved.

To facilitate comparisons, let us take the simpler form for the Cournot equation as given in (21). This may be rewritten:

$$\frac{\Pi + F}{R} z = \frac{H_z}{|\eta_B| - H_B} \cdot \frac{|\eta_B|}{|\eta_z|} \qquad (21a)$$

As was said earlier in this section, we have the somewhat surprising conclusion that

$$\frac{\partial(\frac{\Pi+F}{R}z)}{\partial H_B} = H_z\,(|\eta_B| - H_B)^{-2} . \frac{|\eta_B|}{|\eta_z|} > 0$$

However, if we consider a change in H_z:

$$\frac{\partial(\frac{\Pi+\ F}{R}z)}{\partial H_z} = \frac{1}{|\eta_B| - H_B} \cdot \frac{|\eta_B|}{|\eta_z|}$$

then we see that this will be greater than the above, since we must have $(|\eta_B| - H_B) > H_z$.[27] Further if we consider an increase in both H_B and H_z:

$$\frac{\partial(\frac{\Pi+F}{R}z)}{\partial H} = \frac{|\eta_B|}{|\eta_z|}\left[(|\eta_B|-H_B)^{-1} + H_z(|\eta_B| - H_B)^{-2}\right] > \frac{\partial(\frac{\Pi+F}{R}z)}{\partial H_B} > 0$$

This means that although an increase in the power of the industry to which z sells will increase the margin for the z industry, greater increases in the margin will result from an increase in the power of both z and its purchaser(s), because both the difference between final price and costs and share of that profit area increase in the latter case, instead of merely the height of the total profit area.

In fact if B sells to a further industry C we have:

$$\frac{\Pi\ +\ F}{R}B = \frac{H_B}{|\eta_C|-H_C} \cdot \frac{|\eta_C|}{|\eta_B|}$$

then
$$\frac{\partial(\frac{\Pi+F}{R})B}{\partial H_B} = \frac{1}{|\eta_C| - H_C} \cdot \frac{|\eta_C|}{|\eta_B|}$$

which will be greater than

$$\frac{\partial(\frac{\Pi+F}{R}z)}{\partial H_B}$$

under quite plausible circumstances.[28] Thus it is quite likely that the z industry's <u>share</u> of total profit accruing to the B and z industries drops when H_B rises, although its profit-revenue ratio rises. The point is that when H_B rises, so that the number of firms in that industry falls, then output is necessarily cut back, thus final price and so the total price-cost margin to the chain of industries rises. We have seen that this causes the margin for the z industry to increase; however it is a debatable question as to whether its share of the total margin increases because the total price-cost margin is likely to increase proportionately more than that margin for z.

Put this way, the conclusion of our Cournot theory that an increase in H_B increases the margin for the z industry seems much less surprising than it does at first sight. Furthermore, the types of effect that changes in the market power of the B or z industries provoke seem rather similar between the two theories. In saying this it must be remembered that a decision to act in a joint-profit-maximising manner over splitting the proceeds of market power in the B sector need not imply that the firms in B are able to collude to produce output equal to that a monopoly firm B would produce in their place.

To elucidate, consider that there are say two firms at level B and they act towards the final buyers as a Cournot duopoly. Given this, total output, final price and the margin between price and costs is determined. Suppose now that they are faced by a few powerful sellers of their inputs and decide to act with these sellers jointly rather than, as with the Cournot bilateral monopoly model, treating their prices as parametric. What this means is that total output and so input quantities are determined, but that bargaining may take place over the prices at which these inputs are exchanged. As far as the input producing z firms are concerned, we would expect their share of the total profits to rise if their number diminished. If instead the number of B firms dropped (to one) then we would expect the share of the z firms to diminish but in our example the profit area has risen greatly so that the margin actually obtained in the z industry may well be larger than before. This would of course depend on the precise formula used for splitting the proceeds. Finally if the number of firms in both B and z industries fell then we would expect the margin for the z industry to rise, and to do so more than in the immediately previous case.

In other words, at least in the qualitative sense there seems no reason why our Cournot theory of bilateral power should not act as a proxy for alternative theories of bilateral power also, with the caveat that it does not really incorporate the influence of monopsony power nor power among suppliers lower down in the production chain. Having said this, it is unfortunately true that we cannot generalise the Cournot theory in the same elegant manner as was done with the pure theory of oligopoly.

IX Empirical Work:

As we said earlier, very little empirical work has been done using variables purporting to measure bilateral power, although several writers have mentioned it as a possible influence on profits (e.g. Shepherd 1972). In this final section we first comment briefly on two recent papers which have developed bilateral power measures[29] and then move on to a few points regarding the use of bilateral power measure in empirical work.

Guth et.al. (1973) developed some "buyer concentration ratios" for factors of production sold as intermediate products. These were not actually based on any clearly defined theory,[30] and in fact they did not use them to test any structure-performance hypotheses. Their measure can be symbolised as:

$$BCR_i = \sum_j a_{ij} CR4_j \qquad (31)$$

where BCR_i is the buyer concentration in the ith industry, $CR4_j$ is the (four firm) concentration ratio in the jth industry and a_{ij} is the proportion of ith industry output sold to the jth industry as indicated by 1963 input-output tables for the U.S. Having calculated these measures, they found that $BCR_i < CR4_i$ in general. They then used their measures to test the "countervailing power hypothesis" that high "buyer power" is associated with high seller power, but the relationship was barely statistically significant.

More relevant to our present purpose is Lustgarten's (1975) paper where he does include "buyer power" as an explanation of price-cost margins in the U.S. His theoretical discussion of their place in such a relationship rests largely on an analysis similar to that in Stigler (1964) regarding the costs of collusion among buyers, though he considers that "collusive agreements are less likely to be successful for buyers than for sellers, because firms are typically sellers in only one market but buyers in many markets" (p.126), so that the gains from collusion are smaller for buyers. The measure of bilateral power that he uses, which is loosely based on the above argument, is identical to that given in equation (31), although he also briefly adduces arguments for, and measures three other aspects of buyer power - "relative buyer firm size", "average annual firm purchase" and "sector dispersion of buyers".[31] Lustgarten's BCR measure is then used to help explain the 1963 level of price-cost margins, the other explanatory variables being the four-firm concentration ratio and the capital-output ratio. It proves to have a negative and significant though not powerful effect on price-cost margins. He performs the further experiment of splitting his sample into "low seller concentration" and "high seller concentration" industries on the hypothesis that buyer power matters little unless sellers have the opportunity of gaining prices significantly above marginal cost. For each of these samples price-cost margins are regressed on the capital-output ratio and his BCR variable, though surprisingly not on seller concentration, an omission which probably helps to explain the dramatic fall in overall significance in both samples. Buyer power now has a stronger negative effect on margins in concentrated industries but an insignificant effect in unconcentrated industries; these results might change if seller concentration was included as an explanatory variable. He also uses his BCR measure to explain advertising expenditures.

We may rewrite our own measure of buyer concentration for the z industry from (30) as:

$$H_{BUY_z} = \sum_j \frac{a_{zj}}{(1-\frac{H_j}{|n_j|})} \qquad (32)$$ [32]

where $a_{zj} = \frac{q_j}{q_z}$ and A and B were examples of the j industries who buy from z. We expect H_{BUY} to have a positive effect on price-cost margins, since (30) becomes:

$$\frac{\Pi + F}{R}_z = \frac{H_z(1+u_z)}{|n_z|} \cdot H_{BUY_z} \qquad (33)$$

Immediately we notice that while from Lustgarten's results, a cet. par. increase in the market power of one of the j industries will cause an increase in BCR and so a decrease in the seller's margin, in our model an increase in a buyer's market power causes an increase in H_{BUY} and our prediction would be for a rise in the margin. It remains to be seen whether this will be borne out in our results. Notice however that our prediction can be argued to follow more directly from a well-defined theoretical model.

Moving more closely towards the equation to be estimated empirically,[33] we take ratios of (33) between two time periods to yield:

$$\frac{\left(\frac{\Pi + F}{R}\right)_t}{\left(\frac{\Pi + F}{R}\right)_{(t-1)}} = \frac{H_t(1+u_t)}{H_{t-1}(1+u_{(t-1)})} \cdot \frac{|n_{(t-1)}|}{|n_t|} \cdot \frac{H_{BUY\ t}}{H_{BUY(t-1)}}$$

dropping the z subscript for simplicity. Let us assume following chapter 2 that we may write[34]

$$\frac{H_t(1+u_t)}{H_{t-1}(1+u_{t-1})} \quad \text{as} \quad \frac{L(H_t)}{L(H_{t-1})}$$

and that we may approximate this function by the form

$$\frac{L(H_t)}{L(H_{t-1})} = \left(\frac{H_t}{H_{t-1}}\right)^{\alpha}, \quad \alpha > 0$$

which appears to have some advantages both in simplicity and flexibility since it allows the reaction to a change in concentration to be both greater or less than according to Cournot (see chapter 5 for more detail on this point). Also assuming that

$$\frac{|n_{t-1}|}{|n_t|} = 1,$$

and that some factors might affect changes in price-cost margins across all industries to the same extent we have:

$$\frac{\left(\frac{\Pi + F}{R}\right)t}{\left(\frac{\Pi + F}{R}\right)t-1} = A\left(\frac{H_t}{H_{t-1}}\right)^{\alpha} \frac{H_{BUY\ t}}{H_{BUY\ t-1}} \qquad (34)$$

Now, reactions to changing buyer power need not be according to the pure Cournot bilateral power model. To allow for this in a fairly simplistic manner we could alter (34) to read:

$$\frac{\frac{\Pi+F}{R}_t}{\frac{\Pi+F}{R}_{t-1}} = A\left(\frac{H_t}{H_{t-1}}\right)^{\alpha} \left(\frac{H_{BUY\ t}}{H_{BUY\ t-1}}\right)^{\beta} \tag{35}$$

which may be written more concisely as:

$$\Delta\left(\frac{\Pi+F}{R}\right) = A(\Delta H)^{\alpha}\ (\Delta H_{BUY})^{\beta} \tag{35a}$$

By introducing the coefficient β we allow for changes in buyer power over the period to affect margins (across all industries) to a greater or a lesser extent than implied by the Cournot bilateral power model which seems a sensible generalisation.

Further elaborations which are fairly easily introduced are to allow the elasticity of demand for z's product to vary in some determinate manner and to allow some lag in reaction to structural changes. These points will be discussed in more detail in later chapters, as will problems encountered in attempting to obtain values for H_{BUY} using equation (32). Apart from such considerations, (35) becomes the basis for empirical work.

FOOTNOTES

1. Indeed a recent 600 page text in Industrial Economics (Devine et al. 1974) allocates the subject a bare half page! Two empirical studies will be commented upon briefly in the final section of this chapter.

2. Machlup and Taber (1960) also consider this question to some extent.

3. The problem is that noted by Vernon and Graham (1971) which can lead to vertical integration between a monopolist and a perfectly competitive industry being profitable.

4. Many of these works have as their origin or inspiration Hicks' (1964 ch.7) original strike threat model. An example is provided by Johnston (1972 b).

5. Sonneschein (1968) has a few comments.

6. In the sense of von-Neumann and Morgenstern.

7. Except that in Foldes' more general formulation, the transformation invariance property need not necessarily be assumed.

8. See p. 48 for a statement to this effect.

9. If, for example, there is more than one supplier of an input which may be used in variable proportions with another (or many other) input(s) then the situation becomes more complex as explained later. It is for this reason that we choose to indicate extensions to the model from this more limited point of departure.

10. and has negligible costs associated with it.

11. For example, one formula which would fit into this scheme and has other desirable properties is to share total profits between A and B sectors by giving the B sector $\frac{n_A}{n_A + n_B}$ of the total, where n_A = number of

firm equivalents in sector A, n_B = similar for B. This has the desirable property of allocating half shares whenever $n_A = n_B$, and more generally of yielding the same shares whenever the ratio n_A/n_B is the same.

12. A. Cournot (1927 chap. IX).

13. Cournot assumed proportions $m_1 : m_2$, but it makes little difference.

14. See the conclusion to the previous chapter on this point.

15. We shall not consider second order conditions explicitly in this chapter; in general we shall require that the slope of the firm's marginal revenue curve is steeper than its marginal cost curve.

16. See Hicks (1935) also Machlup and Taber (1960).

17. See Scherer (1970) pp. 242-5 and Needham (1970) pp. 117-122. This assumes fixed proportions in manufacture, see Vernon and Graham (1971).

18. We comment on this assumption in the penultimate section of the present chapter. As a shorthand method of referring to the actors involved, we shall henceforth use B to refer to the (firms in the) brass industry, or the brass monopolist, and similarly z for the zinc industry.

19. Since $MR = p + q.\frac{dp}{dq}$, then $\frac{dMR}{dq} = 2\frac{dp}{dq} + \frac{qd^2p}{dq^2}$. This means that:

$$\frac{q}{MR} \cdot \frac{dMR}{dq} = \frac{2q\frac{dp}{dq} + q^2\frac{d^2p}{dq^2}}{p + q.\frac{dp}{dq}}$$

which bears no simple relationship to n_B as a general rule.

20. I am grateful to John Cubbin for pointing out an error in my original formulation at this stage. Also to Avinash Dixit for pointing out another!

21. See our comments on this choice in chapter 2.

21a The first of these assumptions is purely for simplicity, and is made with the support of result (16), while the second is the normal Cournot oligopoly assumption. The final assumption is that of the z firms regarding the assumed relation between their demand curve and that of the B industry which we have made earlier. As such it describes the behaviour of z rather than B and occurs in the equation above in the text in inverted form. It should of course be distinguished from the actual relation enshrined in the equation of the demand for z at the bottom of the page.

(P.T.O.)

22. The meaning of c'_c is $\Sigma c'_{ci}q_i/q$. It is of course necessary at least that $|\eta_B| > H_B$ for net marginal revenue product to be positive and so for (21) to represent a point of maximum profit.

23. If we allow z to discriminate between buying industries in charging them different prices, we obtain as an alternative to (27) the following:

$$\frac{\Pi + F}{R}z = \frac{R_A}{R_z|\eta_{zA}|(1 - \frac{H_A}{|\eta_A|})} + \frac{R_B}{R_z|\eta_{zB}|(1 - \frac{H_B}{|\eta_B|})} .$$

(R_A is the proportion of revenue gained from industry A, similarly for R_B, and R_z is total revenue. $|\eta_{zA}|$ is the elasticity of demand for z's sales to A, similarly $|\eta_{zB}|$ is the elasticity of B's demand for z's product).

It should perhaps be pointed out that, as we approximate the generalisation of (27) for empirical testing purposes, we are essentially not making any distinction between it and the generalisation of the above formula.

24. This definition is the same as that of equation (5) chapter 2.

25. It must be admitted that the prediction that when H_B increases, $[(\Pi + F)/R]z$ increases cet.par. does not necessarily follow if we allow for inputs to be used in variable proportions or even if the demand curve is of a different type to that assumed here. On the latter point, compare our model with Zeuthen's examples using a Cournot model with a linear demand curve of slope -1. (Zeuthen 1930 p. 85).

26. Though see our comments on Guth et al. (1973) in the next section.

27. This is a simplified argument, assuming both $|\eta_B|$ and $|\eta_z|$ fixed. In fact of course, assuming $|\eta_B|$ fixed in this context is reasonable, but we should remember that $|\eta_z|$ is a function of H_B when differentiating

with respect to that variable. However, the conclusions are not changed when we do this. After some manipulation we find that while:

$$\frac{\partial(\frac{\Pi+F}{R})z}{\partial H_z} = \frac{1}{H_z} \quad \frac{\Pi+F}{R}z \qquad \text{(as in the text)},$$

$$\frac{\partial(\frac{\Pi+F}{R})z}{\partial H_B} = \frac{1}{H_z}\left(\frac{\Pi+F}{R}z\right)^2, \quad \text{smaller than the above.}$$

28. The precise condition is that:

$$\frac{\Pi + F}{R} B \;/\; H_B \quad > \quad \left(\frac{\Pi+F}{R}z\right)^2 \;/\; H_z$$

As mentioned earlier, in this case the formula for $\frac{\Pi + F}{R}z$ will not be entirely accurate.

29. Which, like ours, are not actually measures of monopsony power, but measures of seller concentration.

30. Though they do make a number of points regarding the way that such a variable should be constructed.

31. We will not discuss these alternatives further.

32. We call our measure one of buyer power on the grounds that Guth et al. and Lustgarten use the name. Actually though, it is rather a misleading title. As we have seen in previous sections, oligopsony considerations are not taken into account in forming the measure. Rather, the sellers are affected, in the pure Cournot model, by their perceived version of the demand curve facing them, which perceived curve is shifted according to the net marginal revenue product of the buyers. It is thus the selling power of the buyers (B), which determines the net marginal revenue product and so the effect on the sellers (z). H_j in the formula is the selling power of a particular industry of these buyers.

33. neglecting any consideration of the problem of potential entry.

34. See equation (16) chapter 2, and the arguments for it.

Chapter 5: A CONSIDERATION OF SOME ADDITIONAL POTENTIALLY RELEVANT FACTORS

I Introduction:

In the previous three chapters we have developed a basic theoretical framework for consideration of structure-performance relationships. Having done this, we now attempt an assessment of the empirical literature in the field and then go on to discuss some generalisations and wider theoretical considerations. This will lead us to our empirical estimations in chapter 6.

Drawing together some of the results of previous chapters, we argued in chapters 2 and 4 that allowing for quite general reactions among firms within an industry but imposing a strict Cournot-type reaction to other connected industries gave a relationship:

$$\frac{\Pi + F}{R}\, z = \frac{H_z(1+u_z)}{|\eta_z|} \cdot H_{BUY_z} \qquad (1)$$

(equation 33 chapter 4)

Further, we suggested in chapter 2 that $H_z(1+u_z) = L_z(H_z)$ where $1 > L_z(H_z) > 0$, $L_z' > 0$ so as mentioned in chapter 4 a simple and flexible form for the ratio of the function over time which obeys these restrictions might be

$$\frac{L(H_t)}{L(H_{t-1})} = \left(\frac{H_t}{H_{t-1}}\right)^{\alpha}, \ \alpha > 0$$

(dropping the z subscript for simplicity).

Hence, supposing for the moment that $|\eta|$ and H_{BUY} remain constant between two time periods,

$$\frac{\left(\frac{\Pi+F}{R}\right) t}{\left(\frac{\Pi+F}{R}\right) t-1} = \left(\frac{H_t}{H_{t-1}}\right)^{\alpha}$$

Therefore values of $\alpha = 1$ would indicate a Cournot-type reaction while values greater than one imply a more than proportional increase in profit on revenue, and values less than one a less than proportional increase. We cannot say with certainty however that if α were greater than unity then firms are acting more collusively. For example, if two firms merged to one then the more collusively they were acting beforehand the less the rise in profit-revenue ratio would be. This is rather a special case though, in general an increase in the degree of collusion when the number of firms falls is indicated by a value of α greater than one.

Similar, though much less definite considerations would lead us to treat the H_{BUY} term in like manner to allow reactions other than the strict Cournot one towards changes in bilateral power.[1] Then assuming that the elasticity of demand does not change between the two periods in question leads us to the formulation:

$$\Delta\left(\frac{\Pi+F}{R}\right) = A.(\Delta H)^{\alpha}\ (\Delta H_{BUY})^{\beta} \qquad (2)^{2}$$

(equation (35a) chapter 4)

If the elasticity of demand is liable to vary differentially across industries within the time period, then this relationship has to be amended.[3] The obvious form is:

$$\Delta\left(\frac{\Pi+F}{R}\right) = A.(\Delta H)^{\alpha}\ (\Delta H_{BUY})^{\beta}\ (\Delta|\eta|)^{-1}, \qquad (3)$$

which is derived from (1) utilising the approximations discussed above to obtain (2) except that elasticity of demand is allowed to vary between periods. ($\Delta|\eta| \equiv |\eta_t|/|\eta_{t-1}|$). In a later section of this chapter we consider some arguments bearing on the determination of the elasticity of demand in relation to its likely changes between time periods, that is in the business cycle.

As we said, this model neglects the possibility of entry. However, we concluded in chapter 3 that the relationship between (those factors, for example the level of advertising expenditure, commonly categorised as) barriers to entry and industry performance, especially as regards the likely effect of changes in their magnitude upon changes in the profit-revenue ratio, is by no means clear-cut. We also noted that the model formulated by Spence had the implication that short-run pricing behaviour could be considered as more or less independent of entry considerations provided capacity is maintained at a level sufficient to prevent entry. Such action would normally appear more profitable than a limit pricing strategy since fixed costs should be little or no higher than under the limit pricing policy yet price and variable costs may be set at more favourable levels. For these reasons we feel disinclined to consider altering the formulation of the model to incorporate variables purporting to measure barriers to entry, despite the fact that they are extensively used in other UK studies (see sectionII). Having said this, it remains true that a Spencian outlook does have implications for the estimation of the model, and we comment on these when we discuss the dependent variable specification in section III.

Other areas which receive some discussion later in this chapter are the problems involved when firms are assumed not to be purely profit maximisers, and the question of lags in the relationship with associated possible problems over simultaneity. Of necessity these topics will not be treated with the same degree of formality and rigour as other aspects of the model.

First, however, we turn to the empirical literature. Because of the quantity of work in this area we feel disinclined to present a formal review, particularly since there are at least two recent papers (Weiss, 1971 and Yamey, 1972) which have dealt at some length with this general area. In contrast, similar studies for the UK have only recently been forthcoming, and their numbers are small as yet; in the next section we discuss them individually, albeit briefly. Despite the fact that we do not wish to discuss individual studies for the US, we feel we should comment on what is probably the main point of contrast between UK and US studies. Nearly all the latter use the profit-capital ratio as the dependent variable, whereas we (among others) have exclusively used the profit-revenue ratio. This important divergence is discussed in the third section of the chapter.

II Previous work in the area performed on UK data:

Both Yamey (1972) and Weiss (1971), in their reviews of the structure-profit relationship for the US, consider that there is a significant, though not powerful, influence exerted by concentration on performance (however defined, see the next section). Thus from Weiss:

"Almost all of the 32 concentration profit studies except Stigler's have yielded significant positive relationships for years of prosperity or recession, though they have depended on a wide variety of data and methods" (p.371).

This conclusion mainly emanates from studies which do not include important structural variables apart from concentration. There are two main methods people have used in adding variables representative of barriers to entry; the dummy variables approach and the continuous entry variables approach. The former, characterised by Bain and Mann, gives a significant empirical place to concentration in combination with barriers (see Weiss p. 376 for references and results). An example of the latter is provided by Comanor and Wilson (1967) who find concentration fairly collinear with their "minimum efficient scale" variable (a barrier to entry) and, probably for this reason, unable to exert a significant independent influence. From the results of such multiple regression analysis, Yamey still feels able to conclude that "the concentration ratio emerges as a more or less powerful partial 'explanation' of differences in profitability ... " (p.302) and Weiss similarly is "reluctant to reject the independent importance of concentration ..." (p.378) The pattern to be gleaned from previous work in this area performed on UK data is somewhat ambivalent. We refer here mainly to eight studies, those by Phillips (1972), Shepherd (1972), Holterman (1973), Khalilzadeh-Shirazi (1974), Caves et. al. (1975); also Dutton (1976), Hart (1975) (both working papers), and Cowling and Kelly (1976) (listed separately). Apart from Cowling and Waterson (1976), which has a similar approach to the present study and is not discussed in this section, the only other published study at time of writing (as far as we are aware) is an early paper by Hart (1968), mentioned by Yamey. This differs from those catalogued above in using a profit upon capital surrogate instead of the profit-revenue ratio, and he finds a positive but insignificant effect of concentration upon the dependent variable. In addition there are probably more unpublished studies but we cannot undertake comprehensive coverage of these. We proceed by listing what we hope can be taken as a typical (good) equation from each of the five authors, then commenting upon some salient points.

Phillips:[4] 71 observations; data : centred on 1951.

$$\frac{\Pi}{R} = \underset{}{0.147} + \underset{(2.22)}{0.0010\,CR3} - \underset{(2.22)}{0.1127\,APS} + \underset{(2.76)}{5.609\,\frac{A}{R}}$$

$$- \underset{(1.51)}{0.0730\,CR3.\frac{A}{R}} - \underset{(1.50)}{0.0118\,G} + \underset{(1.27)}{0.002\,EPF} + \underset{(1.01)}{0.0155\,PROCON}$$

$$\bar{R}^2 = 0.260, \quad r = 0.45 \text{ (defined below)}$$

t statistics in parentheses.

Shepherd: 113 observations; data: 1963, 1958 average figures.

$$\frac{\Pi}{R} = 14.5 + \underset{(0.035)}{0.122\,CR} - \underset{(13.232)}{6.797\,\frac{K}{Y}} - \underset{(0.003)}{0.005\,R} + \underset{(2.029)}{4.788\,G}$$

standard errors in parentheses, $R^2 = 0.166$

Holterman:[5] 113 observations; data: 1963

$$\frac{\Pi}{R} = \underset{(-1.09)}{-0.026\ CR} - \underset{(-0.96)}{0.00\ EB} + \underset{(5.88)}{0.123} \frac{A}{R} + \underset{(3.64)}{0.061\ G}$$

$$+ \underset{(2.10)}{4.291} \frac{KA}{O} + \underset{(4.14)}{1.437} \frac{K}{R} \qquad \bar{R}^2 = 0.4544$$

$$r = 0.425$$

t statistics in parentheses.

Khalilzadeh- Shirazi: 60 observations; data: 1963.

$$\frac{\Pi}{R} = 6.34 + \underset{(0.34)}{0.009\ CR} + \underset{(3.86)}{0.833\ MEPS} + \underset{(2.34)}{0.081} \frac{KA}{O} + \underset{(1.46)}{0.038\ G}$$

$$+ \underset{(2.95)}{2.97\ PDD} - \underset{(-1.34)}{0.082\ IMP} + \underset{(2.09)}{0.103\ EXP}$$

$$R^2 = 0.598$$

$$r = 0.60$$

t statistics in parentheses.

Caves et.al.: 60 observations; data: 1963

$$\frac{\Pi}{R} = 5.50 + \underset{(1.88)}{0.044\ CR} + \underset{(1.99)}{0.071} \frac{KA}{O} + \underset{(1.94)}{0.050\ G} + \underset{(2.87)}{2.97\ PDD}$$

$$- \underset{(-0.81)}{0.051\ IMP} + \underset{(2.16)}{0.109\ EXP} + \underset{(3.41)}{0.640\ MEPS!O}$$

$$R^2 = 0.578$$

$$r = 0.35$$

t statistics in parentheses.

Dutton: 34 observations; data: 1963.

$$\log \frac{\Pi}{R} = \underset{(2.21)}{1.693} + \underset{(2.06)}{0.082} \log (CR4.A) - \underset{(-0.12)}{0.016\ t_1} - \underset{(-2.23)}{0.395\ t_2}$$

$$- \underset{(-1.77)}{0.184\ M_1} - \underset{(-2.53)}{0.410\ M_2} \qquad \bar{R}^2 = 0.490$$

t statistics in parentheses.

Hart: 112 observations; data: 1968.

$$\log \frac{\Pi}{Y} = \underset{}{-0.487} + \underset{(0.1816)}{0.0056} \log CR5 + \underset{(6.1927)}{0.1560} \log \frac{K}{L} + \underset{(0.5147)}{0.0346} \log G$$

$$+ \underset{(4.2514)}{0.0717} \log \frac{A}{R} - \underset{(0.7664)}{0.0217} \log MEPS + \underset{(0.2221)}{0.0026} \log IMP$$

$$R^2 = 0.4207$$

t statistics in parentheses

Definitions:

$\frac{\Pi}{R}$ = $\frac{\text{Net output - wages and salaries}}{\text{sales}}$ (estimated value)

For Hart, the denominator is net output.

CR = average of 4 digit , 5 firm, sales concentration ratios

CR3 = 3 firm concentration ratios (source: Evely and Little, 1960)

CR4 = 4 firm employment concentration ratios (source: Sawyer, 1971)

CR5 = 5 firm employment concentration ratios (source: National Institute)

APS = average plant size measured as thousands of employees per plant (source: Evely and Little)

EB = average size of the largest plants (ranked by employment) accounting for 50% of industry employment

MEPS = size of the mid-point plant as a percentage of net output (as a percentage of employment in Hart's work)

MEPS10 = MEPS if the costs of operating above this level are less than 90% of those of a smaller operation, zero otherwise.

K/Y = capital expenditure/net output

KA/O = net assets/gross output (order level). Holterman and Khalilzadeh use different definitions of net assets

K/L = capital expenditure/labour

R = sales

K/R = capital expenditure/sales

G = change in sales 1963/1958 (for Phillips, $\frac{\text{output 1954} - \text{output 1948}}{\text{output 1948}}$; for Hart, proportionate change in sales 1968/1963)

A = advertising expenditure

A/R = advertising/sales. (based on a more aggregative level than the majority of the data)

PDD = 1 where A/R > 1%, zero otherwise (some data from US sources)

EXP = exports as a percentage of industry output

IMP = imports as a percentage of industry output (or sales) (Source: input-output tables, disaggregated)

M_1 = IMP > 5% } (Source: Annual State of Trade for UK)

M_2 = IMP < 1% }

t_1 = tariffs/imports > 5% (source: see Dutton)

t_2 = tariffs/imports < 1%

EPF = "effectiveness of price fixing"; see Phillips for details.

PROCON = 1 where industry is a "producer good" industry, zero otherwise. Again see Phillips

r = correlation between "concentration" and "scale barrier" variables

$R^2, \bar{R}^2$ = coefficient of determination and corrected ditto.

Except where otherwise stated, data comes from the Censuses of Production and is (roughly) at Minimum List Heading (MLH) level.

As we can see from the results all these studies are similar in that they use multiple regression analysis on a Comanor-Wilson type of structure-profit equation.[6] They all include a concentration variable and (apart from Shepherd), at least one continuous entry variable. We have argued that the most relevant concentration variable is not the simple five-firm (or three-firm) ratio, and have argued against the inclusion of barrier variables.[7] All studies but Phillips include some version of a capital-output ratio, we explain why this might be relevant in the next section. They also mostly include a growth term. The theoretical justification behind entering such a variable is not always clear though. For example:

"In a growing market is it easier to make profits than in a stagnating one. The demand curve is continually shifting to the right, thus raising the equilibrium price. It takes time for supply to increase through entry of new firms or the expansion of capacity of existing firms, and during that time higher profits are made by existing firms in the industry" (Holterman p.121).

The problem we have with this argument is that if demand is growing at a known constant rate than there seems to be no particular reason to expect any lags before output expands. Again, if demand is falling at a rate which then suddenly decreases, firms might come up against capacity constraints having planned for a larger decrease in output so that profits might rise despite a negative growth rate. Unplanned growth or decline would appear more germane. This perhaps helps to explain the erratic significance of that term.

Moreover, these studies not only have several variables in common, those for 1963 all use very similar data. The main source is the 1963 Census of Production. One might perhaps consider that they all wished to test very much the same theoretical relationship between concentration, entry barriers and profits but in the face of imperfect data amended their models according to their own particular judgements: Shepherd decides to remove some random noise by averaging over two periods, he decides against including additional structural variables on the grounds of lack of data. Both he and Holterman are of the opinion that the whole sample should be used whilst Khalilzadeh-Shirazi would rather select a more homogeneous subgroup. In including variables to represent barriers to entry, Holterman is in favour of one measure of minimum efficient plant size while Khalilzadeh chooses another; she favours a continuous advertising variable while he plumps for a dichotomous classification. Dutton in contrast considers that, since concentration is only effective (in the longer term) in the presence of barriers, her concentration and (advertising) barrier variables should be multiplicative rather than having separate influences. The main novelties (apart from the cost disadvantage measure of Caves et.al.) are probably Khalilzadeh's import and export variables and Dutton's tariff variables.[8]

Unfortunately the various estimates produced do not,as we might hope, perhaps even expect, yield similar conclusions. Some of the differences are fairly straightforward; it is quite possible that Shepherd's concentration ratio is significant while others are not because of multicolinearity problems between it and the plant economy measure. Caves et.al. consider this to be reasonable when they compare alternatives for the latter (see their p. 137).[9] But could multicolinearity really be responsible for making both of Holterman's measures of these variables insignificant and of wrong sign? Comparing the coefficients and significance of the capital-output ratio and especially the growth variables also yields some rather unexpected divergences.

Looking at the 1963 studies alongside Phillips earlier work and Hart's work on a later period, we find some measure of agreement in that the advertising variable is significant in all studies where it is included. Since however the numerator of the dependent variable contains advertising expenditure (among other things) as well as profits, then this result should be treated with some caution. Apart from this, Phillips' results differ from those of Khalilzadeh and Holterman in that the concentration variable is significant and the plant economies variable is (oddly) negatively significant. In the case of Hart's work, comparison is made slightly difficult by reason of the different dependent variables he uses. Notice though that his factor proportions variable is significant, in common with most other studies where it is included. Also we see that his plant economies variable is insignificant, unlike Khalilzadeh's, and his growth variable is insignificant, unlike Holterman's. A further interesting feature is that when he includes the number of firms rather than a concentration index, this variable achieves (negative) significance in an otherwise identical formulation.

In short, confident conclusions on the effects of changes in structure on profit performance in the UK are not easily drawn from published work. There are two main morals that might be taken from this. Either the data are not sufficiently good to indicate the nature of the true relationship that rules in our economy, or the theory lying behind the equations does not hold good. It was for the latter reason that we decided to attempt a much fuller examination of the relevant theory before proceeding to tests, taking little from other empirical studies for granted.[10] Even then, given the unexpected divergences between, say, Holterman and Khalilzadeh-Shirazi, it could be the case that the selection of the sample might play some part.

Cowling and Kelly's work differs from the above studies in being at the firm level across the food industry. As such, much of the data comes from non-census sources. Their basic approach is to relate price-cost margins to concentration advertising and firm size variables. After a correction for heteroskedasticity, they find that both advertising variables (of which they have two types) and concentration affect margins, probably in a non-linear manner.

III On the choice of the performance measure:

We now turn to the somewhat vexed question of the choice of performance measure in structure-profit studies. As we have said, most studies taking the United States as their data base use the ratio of profit to capital, sometimes just called profitability which we shall write Π/E, whereas those UK studies so far produced use (Π+F)/R, the profit-sales ratio, otherwise called the profit-revenue ratio or price-cost margin.[11] We attempt to discover the basis for the existence of these competing measures, and also to suggest the superiority of the latter for our purposes.

Collins and Preston (1966 and elsewhere), who are almost alone in using the price-cost margin for a study of the US, are fairly explicit about the hypothesis they are testing when they state, for example, that "the closer an industry is to a monopolistic structure, the higher will be the price-cost margins, after account is taken of difference in capital requirements ... Of course, the optimal price-cost margin under monopoly is a function of the price elasticity of demand" (p.228).

Yet Benishay (1967) is equally explicit in his criticism:

"But the price-cost margin, or profit margin on sales, is not the relevant rate of profit ... . Competitive forces tend to equalize rates of return on equity or on assets not profit margins on sales, for investment will flow from poorer to better returns on investment. The use of a sales profit margin to represent a rate of return on assets or on equity is simply inappropriate" (p.73 , his emphasis).

However, if we revert to Bain (1951), we find that "Average excess profits on sales should thus be higher with than without monopoly or effective oligopolistic collusion. This prediction evolves into one that there will be larger profit rates with higher seller concentration than with moderate or low seller concentration..." (p. 295). He goes on to make his position even clearer:

"As the hypothesis is developed to this point, the predicted profit-rate differences are explicitly differences in ratios of excess profit to sales. Because data on profit rates on equity are more readily available (for the U.S.), let us enquire whether the predicted relationship should also hold for the ratios of profit to equity ... . For comparisons of individual cases the answer is no, since the ratio of equity to sales will vary among cases. However, so far as there is on average among groups of firms or industries being compared about the same ratio of equity to sales, their average equity rates should stand in about the same relation as their sales ratios" (pp. 296-7).

Yet, despite profit-sales figures being more easily available in the UK, the debate has relevance for our work when we find in Khalilzadeh-Shirazi (1973)[12] that:

"Most US studies of structure-performance relationships have employed rates of return on owner equity as their dependent variable. Conceptually, this is the proper variable, since we are testing against the null hypothesis generated by the competitive model that rates of return are equal (in the long run) across firms and industries" (p.2) (though in fact he finds some theoretical place for the price-cost margin and uses it in his empirical work). The thesis of this section will be that both null and alternative hypotheses need to be considered, and that this has implications for the choice of measure.

Now the null hypothesis Khalilzadeh-Shirazi mentions is not unique to profit on capital; in pure economic profit terms profit on anything would be equated (at zero) across industries under competition. It is only when we move to considering recorded profit that difficulties arise. Recall that under monopoly:

$$\frac{p - MC}{p} = \frac{1}{|\eta|} \qquad \text{whereas under competition} \qquad \frac{p - MC}{p} = 0$$

p is price, MC marginal cost and η the elasticity of demand. Thus, making the assumption that marginal cost is equal to average variable cost[13] and multiplying numerator and denominator by output we obtain the profit-revenue ratio $\frac{\Pi + F}{R}$ as $\frac{R - C}{R} = \frac{1}{|\eta|}$ for monopoly, and for competition $\frac{\Pi + F}{R} = 0$. More generally we have found that under a given market structure in a particular industry we may write:

$$\frac{\Pi + F}{R} = \frac{H(1 + u)}{|\eta|}$$ (equation (6) chapter 2; this n bilateral power considerations)

Without changing the nature of the argument in this section, we drop the (1 + u) term and rewrite the above equation more simply as:

$$\frac{\Pi + F}{R} = \frac{H}{|\eta|} \qquad (4)$$

Multiplying both sides of (4) by R/E then <u>indicates</u> that

$$\frac{\Pi + F}{E} = \frac{H}{|\eta|} \cdot \frac{R}{E} \qquad (5)$$ [14]

is an equivalent relationship concerning profit on equity. Again under competition $(\Pi + F)/E = 0$.

The null hypothesis is the same; why then should the proponents of the profit on equity measure want to take the alternative hypothesis in what looks to be a more complicated form (and a form which generates some further difficulties, as we see later)? The answer presumably lies in the fact that we cannot measure profits in the manner in which economists are accustomed to speak of them. Specifically, let us take it that variable costs may be split up into labour costs (wL) and capital costs (mK)[15] (others can of course be added).

Thus though profit is given by:

$$\Pi = R - wL - mK - F,$$

what we actually measure is:

$$(\Pi + F)_M = R - wL,$$

since we cannot easily measure the cost of capital.[16] For the measured profit-revenue ratio then we have:

$$\left(\frac{\Pi + F}{R}\right)_M = \frac{H}{|\eta|} + \frac{mK}{R} \quad (6),$$

and by the same token:

$$\left(\frac{\Pi + F}{E}\right)_M = \frac{H}{|\eta|} \cdot \frac{R}{E} + \frac{mK}{E} \quad (7)$$

Now, if we could assume mK/E to be constant across industries in that K = iE say (where i is the rate of interest in units of the reciprocal of time and E equity in pounds sterling), we would have:

$$\left(\frac{\Pi + F}{R}\right)_M = \frac{H}{|\eta|} + \frac{miE}{R} \quad (8) \quad \text{from (6)}$$

and $$\left(\frac{\Pi + F}{E}\right)_M = \frac{H}{|\eta|} \cdot \frac{R}{E} + mi \quad (9) \quad \text{from (7).}$$

The null hypothesis in the latter case would then be simpler than that in (8), due to its yielding a constant value in the cross-section.[17]

In fact Stigler (1963, ch.3), in tackling the problem of the null hypothesis, implies that we would expect the rate of return on capital in a set of competitive industries to approach equality.[18] This equality comes about because firms are assumed to enter an industry until entrepreneurs providing capital services earn only a normal return on their investment.

However, for an important reason we may consider that an approach using $\left(\frac{\Pi + F}{R}\right)_M$ as the dependent variable provides a superior estimating equation. This arises because if the entrepreneur supplies capital plus his services in a monopoly industry then he will obtain an extra return, being the monopoly profit involved. As a consequence, his capital stock due to its being in that monopoly industry will be worth more and should be revalued in the capital market to reflect that increased (monopoly) return. The import of this is that we should expect if H increases, and with it the monopoly return, then E also will increase so eroding that return.[19] There are two important consequences flowing from this possibility. Firstly, it it illegitimate to derive (8) and (9) from (6) and (7) in general, because we cannot assume mK/E to be constant. Secondly, in using an approach along the lines of (7), the effect of monopoly in the product market is ameliorated in the equation by the effect of the capital market in revaluing assets. When doing such work on the structure performance nexus we should wish to test the gross effect of product market power not the net effect once the capital market has revalued the assets concerned to some extent. That is, if the sales/asset ratio falls as H rises then (6) clearly becomes the better form to use. In this context it is interesting to note that Qualls (1972), who reworks the studies by Bain and Mann using the profit-revenue ratio rather than their profit equity ratio, finds that "In general, the results confirm their basic conclusions concerning the importance of concentration and the height of entry barriers in influencing market performance. In some instances, however, minor modifications are called for, and in others considerable strengthening is provided" (p. 151). One of Collins and Preston's studies

(continued on p. 160

- p. 159 has been omitted)

also finds the price-cost margin to be better explained than profitability.

Given that the profit revenue ratio seems superior to profit on equity in this pure oligopoly model, there appear to be two defences for using the $\frac{\Pi}{E}$ approach. One is that entry barrier theory may indicate the relevance of profit on equity rather than profit on sales, since what is relevant to the entrant is the return on the capital he invests upon entry. Another is that entrepreneurs acting in the best interests of their shareholders should in fact attempt to maximise profit on equity. However the latter argument would seem to be a red herring, for no formal model has been given by the proponents of this approach (as far as we are aware), and it is hard to see how one yielding different but still sensible results could be produced. In any case, the fact that profit on equity is the maximand provides no assurance that a simple relationship linking profitability to structural features will result.

The first argument is perhaps best illustrated by the position of Weiss (1971) who says that the profit-revenue "ratio would be a correct statement of the optimal margin if entry were blockaded so that margin depended on demand elasticity". Yet he believes that "conditions of entry rather than elasticity is the main determinant of optimal margin in most markets. The entry-inducing price yields a normal return on total investment (including entry costs) to the most likely entrant. The optimal price is a function of the entry-inducing price. It should yield a correspondingly higher return on equity to insiders the higher the barriers to entry and the greater the insiders' ability to collude" (p. 367 footnote 6).

However there are other ways of looking at the established firm's position having regard to entrants. From Bain (1959) for example we have that "the condition of entry may be measured by the degree to which established firms can persistently elevate their prices above minimal average or competitive costs before making it attractive for new firms to enter." (p.33) Again we have Modigliani's (1958) famous quote:

"In summary, under Sylos' postulate there is a well-defined maximum premium that oligopolists can command over the competitive price and this premium tends to increase with the importance of economies of scale and to decrease with the size of the market and the elasticity of demand" (p.220).

As we have seen, both these statements can be easily translated into the form of a measure of limit price-cost margins, and in fact we developed our models of chapter 3 purely on the basis of predictions about the size of the price-cost margin. For example, with the simple Modigliani model we found:

$$P_o = P_L \left(1 + \frac{1}{|\eta| s}\right)^{20} \qquad \text{(equation (2) chapter 3).}$$

Then, making the same assumptions as we did with regard to measurement in the oligopoly model above, we have that:

$$\left(\frac{\Pi + F}{R}\right)_{LM} = \frac{1}{|\eta| s + 1} + \frac{mK}{R} \qquad (10)$$

$$\left(\frac{\Pi + F}{E}\right)_{LM} = \frac{1}{|\eta| s + 1} \cdot \frac{R}{E} + \frac{mK}{E} \qquad (11)$$

There would seem to be little to choose between these two limit pricing formulations as regards ease of estimation.

Further, the main message of the dynamic limit-pricing models[21] is probably that, in general, industry price will be set with regard to both limit price and short-run profit-maximising price,[22] so as to retard rather than prevent new entry, since this mixed strategy is more profitable in the long term than either limit pricing or myopic profit maximisation. In this case, an amalgum of the two relationships discussed, that is some combination of (10) and (6), or alternatively of (11) and (7), is probably required and given our earlier arguments the profit-revenue ratio retains its superiority over the profit-equity ratio in empirical work.

Suppose now that capital costs are all capacity costs. Spence (1977) then suggests that capacity is fixed by the desire to prevent entry at a level k_L. As such it is determined outside of the short-run profit-maximising framework. We now have fixed $mK = rk_L$ so that profit plus all fixed costs:

$$\Pi + F + rk_L = R - wL$$

which is equal to price minus average variable costs multiplied by output. In this case the measured quantity $R - wL$ is equated with the theoretical quantity, price minus marginal and so (given our assumptions of appendix I) average variable cost, and hence from equation (23) chapter 3:

$$\frac{\Pi + F + rk_L}{R} \equiv \left(\frac{\Pi + F}{R}\right)_M = \frac{1}{|\eta|} \qquad (12)$$

Here there is no problem over the capital output ratio.

To the extent to which capital costs are not all fixed in this sense:

$$\left(\frac{\Pi + F}{R}\right)_M - \frac{\delta mK}{R} = \frac{1}{|\eta|}, \qquad (13)$$

where δ is the proportion of capital costs not relating to the creation of capacity, so not fixed. A certain proportion of capital costs may not be fixed if, for example, the Spencian solution does not hold fully in that the extent of overcapacity to dissuade entry varies with the height of the actual price above limit price. Another example is the case discussed by Spence where capacity costs affect marginal costs. If we now generalise (13) to the level of equation (1) this chapter:

$$\left(\frac{\Pi + F}{R}\right)_M - \delta \frac{mK}{R} = \frac{H(1+u).H_{BUY}}{|\eta|} \qquad (14)$$ [23]

we see that a Spencian outlook has largely separated the effect of potential entry out from the effect of established numbers, the former being fixed in the short term in his framework.

Now as we said earlier, the strategy of retaining capacity at the limit-pricing level would appear more profitable than limit-pricing itself. Thus, given this model, our plan will not be to include as explanatory variables either variables attempting to measure barriers to entry or the capital-output ratio (or some proxy) but rather use only determinants of actual market power (horizontal, bilateral and proxies for the change in elasticity of demand). However on the left hand side of the equation we shall experiment with subtracting certain proportions of (a proxy for) the capital-output ratio to see whether this improves the fit of the relationship. If it does not, this would appear to be evidence in favour of the Spencian arguments.

We should contrast quite strongly the separation achieved above with the situation obtaining under the preceding assumptions when the explanatory variable is the profit equity ratio, or "profitability". From (13) we may obtain:

$$\left(\frac{\Pi + F}{E}\right)_M - \delta \frac{mK}{E} = \frac{R}{E|\eta|}$$

while from (14):

$$\left(\frac{\Pi + F}{E}\right)_M - \delta \frac{mK}{E} = \frac{H(1+u).H_{BUY}}{|\eta|} \cdot \frac{R}{E} \qquad (15)$$

The variable on the right hand side is now a mixture of two types of effect. We have as before the effect of established industry numbers as reflected in H. But we also have E. If the height of entry barriers change then deterrent capacity will have to change. In the profit-revenue ratio case (as in (14)) this effect is subsumed in the dependent variable. By comparison in the present case a fall in entry barriers necessitating an increase in capacity means that equity may well rise initially, the firm now requiring increased capital in plant and so on. Thus when using profitability as the dependent variable it seems that its explanation should be sought both from market power and barriers to entry variables in the reasonably short run.

The fact remains that observers using profit-revenue ratios have obtained statistically significant results on barriers to entry variables, as we saw in the previous section. If the argument of the present section is accepted there would seem to be at least two possible explanations of this.[24] The first is that in actual practice (as Bain (1951) hoped) profit-equity ratios may vary rather similarly across industries to profit-revenue ratios. A second is that of the variables used, some are similar, or identical, to components of the dependent variable and others may well be quite collinear with concentration variables. Examples of the former are the advertising-sales ratio and the capital-output ratio.[25] On the latter the proxies used to represent minimum efficient size of plant or firm are based, like concentration measures, on statistics measuring the size distribution (of firms or plants) within the industry.

IV Elasticity of Demand and the Trade Cycle

Perhaps coincidentally, in 1936 two articles appeared concerning the effect of the trade cycle on the elasticity of demand and so on profits in oligopolistic industries. We have Harrod who said that in a slump:

"(A man) is loth to relinquish enjoyments to which he has become accustomed (in the boom), and immediately begins to cast about for means of meeting adversity with the least inconvenience to himself. That same force of habit which in times of improvement tends to make him an imperfect buyer, reinforces his activity when it is a question of economising ... (therefore) once the slump has set in, demand becomes suddenly much more elastic." (1936 p.87).

In contrast, Galbraith argues that:

"Where the decrease in demand is the result of depression, an increase in elasticity may be considered improbable. People with decreased money incomes and increased concern for their economic security are less rather than more responsive to lower prices. Producers and consumers alike tend to postpone purchases of durable equipment. The market comes to be composed more and more of very able and very needy buyers. Demand is less rather than more elastic" (1936 pp. 463-4).

Undoubtedly, part of the reason why they reach different conclusions is due to their respective desires to explain different "received facts" of the depression they had recently experienced. Galbraith's concern was the observation that prices in oligopolistic industries appeared sticky in the downswing whereas Harrod wished to explain a fall in profits. Both arguments seem to hold some water, for they appeal to different facets of the slump.

Harrod's picture is of a satisficing individual who is more diligent in searching for his purchases in the slump, he moves nearer to choosing the consumption bundle which maximises his utility.[26] However, a similar point has been made by Cowling and Cubbin (1970) of a maximising individual who faces search costs:

"At high levels of demand market share is determined more by considerations of availability than fine price differences. On the other hand in a recession the time and trouble involved in finding the 'best buy' have a lower value placed upon them"(p. 20).

As Stigler (1966 p.2) says: "the buyer searches for additional prices until the expected saving from the purchase equals the cost of visiting one more dealer" so that if earnings are lower and/or hours are shorter in the slump, the opportunity cost of search drops, more search takes place, fewer people pay or are willing to pay high prices, and the elasticity of demand increases.

Galbraith's point is much more of a comment on the propensity to save. This is clear in a similar argument made by Sylos-Labini, who quotes Schumpeter with approval:

"People who in a depression worry about their future are not likely to buy a new car even if the price were reduced by 25% especially if the purchase is easily postponable, and if the reduction induces expectations of further reductions" (Sylos-Labini 1962 p. 70).

As such it is more difficult to formalise within the present context than the Harrodian argument, though it does appear to relate more obviously to durable goods. A simplistic justification might lie in the idea that those consumers who would only buy the good at relatively low prices tend to be more affected by the desire to increase their savings in the slump; that is their ratio of the marginal utility of savings to the marginal utility of consumption of that good rises relatively more than that of those previously willing to buy the good at relatively high prices. If this were true then the demand curve would become more steeply inclined in the slump.

The alternative argument, especially in the form stated by Cowling and Cubbin, is easier to symbolise. Suppose an individual's demand for a commodity is related to the true price the consumer pays (that is the manufacturer's price plus opportunity costs of search, see Mincer (1963)) in constant elasticity form. Then:

$$q = a(p + c_o)^{-|\alpha|}$$

where c_o is the opportunity cost element and α the individual's elasticity of demand.

Thus: $$\frac{dq}{dp} = -|\alpha| \, a \, (p + c_o)^{-|\alpha|-1}$$

and so: $$|\eta| = -\frac{p}{q} \cdot \frac{dq}{dp} = |\alpha|(p + c_o)^{-1} p$$

This last quantity is the elasticity of demand as faced by the firm.

Now:

$$\frac{\partial|\eta|}{\partial c_o} = -|\alpha|(p + c_o)^{-2} p < 0,$$

so that if the opportunity cost rises, elasticity facing the firm falls. If we consider that a drop in the individual's income or wage rate causes the opportunity cost of search to fall then we would expect elasticity of demand facing the firm to rise in the slump.

The problems arise in this case when we move to considering measurement of the relevant variable having this effect on opportunity costs and so the firm's elasticity of demand. Fairly obviously for the final consumer it is income per unit of time. But what of the buyer which is a firm? It would seem most sensible to consider that the variable is again income (or revenue) per unit of time. As an *ad hoc* justification let us note that in the case of the consumer his opportunity cost is assumed to fall because the price per unit of output he produces falls. Therefore by analogy the price per unit of the firm's output drops (in a depression) which means that the opportunity cost of searching for inputs falls and more search takes place; staff are switched towards searching for cheaper inputs so that the manufacturers that the firm buys from are faced with more elastic demands for their product. In summary, the elasticity of demand for a particular industry's products then increases with a fall in the per unit incomes of its purchasers.

However, we have no real theoretical guidance as to how to aggregate these income falls across the various purchasers of that particular product because we do not know the functional form of the relationship between opportunity cost and demand. If it were of constant elasticity form then presumably aggregation would proceed by calculating geometric means (as with log-linear engel curves), but this is infeasible. Given such a situation, aggregation has to proceed in some arbitrary manner, and we might as well choose the method that seems most straightforward. One candidate which then presents itself is to take the downturn in the industry in question as a measure of the downturns in incomes of those industries and consumers which purchase from it.

Unfortunately, using the ratio of revenues between the two periods in that industry as a proxy for this effect brings difficulties in its path. There is first the problem of spurious correlation since revenue would appear both on the left and right hand sides of the amended equation (3) and there is a second attendant problem that bias and inconsistency result when one of the right hand variables is related to the error term in the statistical version of the equation. For this reason a better proxy, which is also related to the slump but not as directly connected to the dependent variable, might be the percentage unemployed in that industry. The higher the unemployment rate, the greater the depression in that industry so hopefully the lower the opportunity cost of search of buyers from that industry and the more elastic is industry demand. Replacing elasticity of demand by a function of the percentage unemployed in (3) we have:

$$\Delta(\frac{\Pi+F}{R}) = A\ (\Delta H)^{\alpha}\ (\Delta H_{BUY})^{\beta}\ (\Delta U)^{-\gamma};\ \alpha,\beta,\delta \text{ positive;}$$

or

$$\log \Delta(\frac{\Pi+F}{R}) = \log A + \alpha\log(\Delta H) + \beta\log(\Delta H_{BUY}) - \gamma\log(\Delta U) \quad (16)$$

Since unemployment is only a proxy, we might consider as a simple alternative formulation:

$$\log \ \Delta\left(\frac{\Pi+F}{R}\right) = \log A + \alpha\log(\Delta H) + \beta\log(\Delta H_{BUY}) - \gamma(\Delta U) \qquad (17)$$

We leave the problem of the time period over which ΔU is measured until the final section of this chapter.

Such a proxy has the property that it might be able to take into partial account other factors not as yet mentioned which might be considered to affect the model. For instance it is possible that over the cycle the relationship between marginal and average variable cost may change slightly[27] and such a cyclical variable may pick up effects like these to some extent. The other side of this coin is that a negative coefficient on ΔU need not be evidence in favour of Harrodian-type arguments regarding cyclical behaviour of the elasticity of demand.

Although this formulation arose from Harrod's arguments rather than Galbraith's, this does not imply either that a positive coefficient on ΔU need be evidence in favour of the latter. For as we implied earlier, the two positions are not necessarily mutually inconsistent as they look at the problem from rather different standpoints. We have said little about the Galbraithian argument and its implications though we shall experiment with the idea that durable goods (to which his arguments mainly seemed to refer) may be differentially affected by cyclical fluctuations. There are in addition alternative reasons why we might expect such a differential effect between durables and non-durables; these are mentioned in Cowling and Waterson (1976).

V <u>Managerial Discretion - A few comments</u>:

Our approach to the literature in this area will be particularly cavalier, for a discussion of any length would take us well outside the main areas of the thesis. We shall be content merely to point out that whilst the rest of the thesis has incorporated the assumption that firms wish to maximise profits, by no means all of our conclusions are negated if we take account, in a very naive manner, of the possibility that managers might have objectives in addition to profit. To this end we utilise a managerial utility function for a monopolist, involving preference for a factor of production, and then move to calculating the price-cost margin. Some extensions to an oligopolistic industry are suggested. The approach then is similar to that of Peel's (1973) generalisation of the "Dorfman-Steiner" result though our conclusions are not as strong.[28]

In this section we shall not concern ourselves explicitly with the power of shareholders or owners of the firm. Thus management's utility function always contains profit as an argument but we do not specify reasons for this. The constraints on the maximand are then purely definitional; we give a full set below.

$$\text{Profit } \Pi \equiv p.q. - c(q) - F$$

We utilise the assumption of earlier chapters in specifying the variable cost function $c(q) = c.q$, where c is marginal and average variable cost. However when we wish to talk about preference for a particular factor of production, we let $c(q) = wL + mK$, that is we assume there are two variable factors of production. Quantity supplied is then a function of the two variable factors $q_s \equiv q_s(L,K)$. Revenue $R = p.q$, and quantity demanded is specified as $q_d = q(p,z)$. Quantity supplied will equal quantity demanded at equilibria. (z is simply an additional demand shift variable and like F does not play any real part.)

The procedure we adopt is to form a Lagrangean function from the maximand and relevant constraints. We then derive the first order conditions for a maximum; second order conditions are neglected.

Suppose a managerial utility function of the type:

$$U = U(\Pi, L)^{29}$$

Form the Lagrangean:

$$G = U(\Pi,L) + \lambda_1 \left[q - q_s(K,L)\right] + \lambda_2 \left[q - q(p,z)\right] + \lambda_3 \left[\Pi - pq + wL + mK + F\right]$$

Some relevant first order conditions are:

$$\frac{\partial G}{\partial \Pi} = \frac{\partial U}{\partial \Pi} + \lambda_3 = 0 \tag{18}$$

$$\frac{\partial G}{\partial L} = \frac{\partial U}{\partial L} - \lambda_1 \frac{\partial q_s}{\partial L} + \lambda_3 w = 0 \tag{19}$$

$$\frac{\partial G}{\partial K} = -\lambda_1 \frac{\partial q_s}{\partial K} + \lambda_3 m = 0 \tag{20}$$

$$\frac{\partial G}{\partial q} = \lambda_1 + \lambda_2 - \lambda_3 p = 0 \tag{21}$$

$$\frac{\partial G}{\partial p} = -\lambda_2 \frac{\partial q}{\partial p} - \lambda_3 q = 0 \tag{22}$$

Now we have that costs are:

$$C = wL + mK + F,$$

so that:

$$dC = wdL + mdK,$$

assuming that w and m are considered by the firm as constants, in addition to F being constant.

Also:

$$dq_s = \frac{\partial q_s}{\partial K} \cdot dK + \frac{\partial q_s}{\partial L} \cdot dL \tag{23}$$

so:

$$MC = \frac{dC}{dq_s} = \frac{wdL + mdK}{\frac{\partial q_s}{\partial K} \cdot dK + \frac{\partial q_s}{\partial L} \cdot dL} \tag{24}$$

Now from (19):
$$w = \frac{\lambda_1}{\lambda_3}\frac{\partial q_s}{\partial L} - \frac{\partial U}{\partial L}\cdot\frac{1}{\lambda_3}$$

$$= \frac{\lambda_1}{\lambda_3}\frac{\partial q_s}{\partial L} + \frac{\partial U/\partial L}{\partial U/\partial \Pi} \qquad \text{from (18)}$$

(20) yields
$$m = \frac{\lambda_1}{\lambda_3}\frac{\partial q_s}{\partial K}$$

Substituting these values into (24) we obtain:

$$MC = \frac{\frac{\lambda_1}{\lambda_3}\frac{\partial q_s}{\partial L}\cdot dL + \frac{\lambda_1}{\lambda_3}\frac{\partial q_s}{\partial K}\cdot dK + \frac{\partial U/\partial L}{\partial U/\partial \Pi}\cdot dL}{\frac{\partial q_s}{\partial L}\cdot dL + \frac{\partial q_s}{\partial K}\cdot dK}$$

$$= \frac{\lambda_1}{\lambda_3} + \frac{\partial U/\partial L}{\partial U/\partial \Pi}\cdot\frac{dL}{\frac{\partial q_s}{\partial K}\cdot dK + \frac{\partial q_s}{\partial L}\cdot dL}$$

or $$MC = \frac{\lambda_1}{\lambda_3} + B \qquad \text{(say)} \tag{25}$$

Now from (21):
$$p = \frac{\lambda_1 + \lambda_2}{\lambda_3}$$

$$\therefore \frac{p - MC}{p} = \frac{\lambda_2 - \lambda_3 B}{\lambda_1 + \lambda_2}$$

Since $\lambda_2 = \dfrac{-\lambda_3 q}{\partial q/\partial p}$ from (22),

then from (21):
$$\frac{\lambda_2}{\lambda_1+\lambda_2} = -\frac{q}{p\,\frac{\partial q}{\partial p}} = -\frac{1}{\eta}$$

Thus:
$$\frac{p - MC}{p} = \frac{1}{|\eta|} - \frac{\lambda_3 B}{\lambda_1+\lambda_2}$$

$$= \frac{1}{|\eta|} - \frac{B}{p}$$

$\therefore$ substituting back for B:

$$\frac{p - MC}{p} = \frac{1}{|\eta|} - \frac{1}{p}\left[\frac{\partial U/\partial L}{\partial U/\partial \Pi} \cdot \frac{1}{\frac{\partial q_s}{\partial K} \cdot \frac{dK}{dL} + \frac{\partial q_s}{\partial L}}\right]$$

But from (23), $\frac{\partial q_s}{\partial K} \cdot \frac{dK}{dL} + \frac{\partial q_s}{\partial L} = \frac{dq_s}{dL}$

Finally then:

$$\frac{p - MC}{p} = \frac{1}{|\eta|} - \left[\frac{\partial U/\partial L}{\partial U/\partial \Pi} \cdot \frac{1}{p\,\frac{dq_s}{dL}}\right] < \frac{1}{|\eta|} \tag{26}$$

We should note here that we still have technical efficiency, for we are on the production surface. Preference for a particular factor can simply be thought of as changing the price relatives between factors of production as viewed by management. A more general approach (along the lines, say, of Moreland (1972)) would allow for technical inefficiency, but once we incorporate such considerations we would expect to be able to say little about the margin. However, a simpler model, with margin unchanged from the profit maximising case, would result if the managerial preference factor (e.g. staff expenditures) was assumed to be a fixed expenditure rather than affecting marginal costs.[30]

In order to evaluate qualitatively the price-cost margin indicated by equation (26) we have to consider the relative magnitudes of the two first-order partial derivatives with respect to utility and the nature of influences on their size. Obviously we could incorporate a plethora of such influences, particularly factors like the nature and extent of shareholder power and so on. However, our concern is more specifically with possible cross- sectional differences between the relative magnitudes mentioned, assuming managers and shareholders to be fairly homogeneous groups so that we may talk more nearly in ceteris paribus terms.

From this point of view the costs of non-profit-maximising behaviour presumably depend inversely upon the earnings under profit-maximising behaviour, so that we would expect movement furthest away from profit orientation when the maximum profits obtainable are largest. On this we note that the less elastic is demand, the higher is the margin and so the more opportunity there is of moving away from a profit-maximising position. The suggestion is then, that the ratio of marginal utilities varies, amongst other things, with the inverse of the elasticity of demand.[31] If this is so, then from (26) we have:

$$\frac{p - MC}{p} = \frac{1}{|\eta|} - g\left(\frac{1}{|\eta|}, \ldots\right) = h\left(\frac{1}{|\eta|}, \ldots\right)$$

$$\frac{\partial h}{\partial|\eta|} < 0^{32} \qquad (27)$$

If we can accept this then we may say that even under the more general situation of the alternative behavioural assumptions postulated above, the margin will vary with numbers in the industry. For, from (27), if that firm is simply the ith in an n firm industry (each firm having identical marginal costs) we have:

$$\frac{\Pi + F}{R} = \frac{p - MC}{p}\, i = h\left(\frac{1}{|\eta_i|}, \ldots\right) = h\left(\frac{1}{n|\eta|}, \ldots\right) \qquad (28)$$

under pure Cournot assumptions. Explicit generalisation beyond this stage is somewhat difficult unless we are willing to specify the functional form, but in principle could be accomplished. It is of course most likely that with a utility maximisation model of this type, reaction to a fall in numbers is attenuated in comparison with the reaction under profit maximisation. This is because, when potential profit rises, as long as profit claimants become no more powerful, there is a further chance for management to indulge in non-profit oriented discretionary behaviour.[33] We shall not explore this avenue further, so we neglect the possible influence of (bilateral) shareholder power on our model. To do otherwise would involve a very substantial data collection problem in order to perform tests.

VI Lags and Simultaneity

One assumption implicitly underlying equations (2) and (3) of this chapter is that changes in an industry's structure immediately effect changes in the margin. Our initial purpose in the present section is to consider this assumption in more detail, from whence we shall go on to discuss some related points. Actually, if we remind ourselves of the data to be used,[34] the assumption we have made need not be as strict as that. Since the change in structure is that taking place between the 1963 and 1968 censuses of production, then it is possible to maintain equation (2) while allowing structure to affect performance up to almost five years later. For it might be the case that, say, concentration increases the day after the census form is submitted and that the margin gradually moves upward towards a new equilibrium over the period so that performance in the year reported for the next census reflects a large part or even the whole of this change in the margin. This is obviously a completely atypical case though and we can say little about the average lag, if any, between changes in actual structure and actual performance in the data.

As far as factors which can theoretically affect structure in equation (2) are concerned, we should note that an increase in concentration (measured by H) can come about either by merger[35] or by reallocation of sales among the existing firms. H_{BUY} is increased by the same factors in the buying firms, but in addition by reallocation among purchasing industries. Unfortunately, very little work has been done on the sort of lags to be expected from such structural changes, which is a manifestation of the fact that industrial economists have not particularly concerned themselves with conduct as the link between structure and performance.[36] In the case where there is, for one reason or another (e.g. a merger or a successful advertising campaign), a reallocation of sales in favour of the larger firms in an industry, then the Herfindahl index rises. This increase in power might mean for example that a loose collusion among large firms can be strengthened at low cost or that price leadership becomes more straightforward; the mechanics of an ensuing eventual increase in margins are fairly clear but the path to the new equilibrium is ill-defined. In a time of incomes policy it is probably difficult to move in any way but with caution; if there are worries about reference to the Monopolies Commission caution is again required; in the case of a merger it may take time to move to a profitable organisational structure. The benefits of an immediate increase in margins via higher prices are obvious, what is lacking is any comprehensive succinct evaluation of the mass of factors which go to make up the costs of moving quickly to a new equilibrium position and so indicate the likely time-scale.

Given this situation, how are we to treat the timing question in empirical work? One piece of evidence which assists us to some extent is Singh's (1971) belief that it can take up to five years between a merger and its effects on performance. This gives a rough outside limit, though not one which circumscribes us greatly since we cannot obtain data after 1972 in any case. Changes in structure between 1963 and 1968 then are expected to influence performance sometime between 1963 and 1972. But we cannot take the whole of this period, for if there is a lag in the relationship, performance changes between 1958 (the previous census year) and 1963 will still be feeding through in the early years. As we have no data on performance between 1963 and 1968, we are left with two favoured but rather imperfect proxies after taking timing into account. If we believe the lag to be short and/or that the majority of structural changes occurred early in the period then the 1963 to 1968 change in performance should be most affected. On the other hand, if the lag is long or structure changed mainly towards the end of the period, 1968 to 1972 performance changes should be used. The latter is unfortunately a poor proxy even on those assumptions, covering only four years and not including quite as full responses to the questionnaire as previous years (at the time of writing). It is quite obvious that the empirical results we obtain can throw little light on the timing question given the very limited data at our disposal; we shall simply seek a reasonable degree of explanation.

The question of lags also crops up in connection with the use of unemployment changes for changes in the elasticity of demand as in equations (16) and (17). There are two points here. The first is that even in the contemporaneous form (with the 1968 on 1963 performance variable) we should not necessarily use unemployment changes between those two years. This is because it is well known that unemployment lags behind many other cyclical variables. For this reason the change between 1969 and 1964, say, may be more relevant. Secondly in the lagged form (with 1972 on 1968 as the performance variable) it is not clear that the lag on unemployment should be of the same length as that on other structural variables. It is unlikely to take longer for firms to react to changed demand conditions than to changed industry structure; rather it may well not take as much time, although again there is little previous work to guide us. This point of course holds in addition to the first for the lagged relationship. For these reasons we indulge in some experimentation on the lagging of the unemployment proxy in empirical work.

Moving very briefly to a somewhat wider consideration of an equation such as (16), a further fundamental assumption we have made is that the performance variable is the dependent variable, while those on the right hand side are independent. As some writers (e.g. Phillips 1970) have argued, this is not necessarily the whole story. Indeed, in discussing dynamic models of entry behaviour we have implicitly accepted that conduct and performance may feed back to structure, since those models incorporate, albeit simply, fringe response to established industry performance. Thus Phillips, among others,[37] considers that we should take account of the possibility that the equation we have been discussing is merely one among a simultaneous system of relationships. Obviously this opens up a very wide avenue which we cannot do justice to here.

Phillips' point is relevant in the present context of our discussion of lags as, if performance only reacts to structural changes with some delay, then despite the potential simultaneity of the system, we need take no account of this in practice. On the other hand, if the lag is short enough for us to estimate the relationship as a contemporaneous one, we might consider that our problems are greater. One response to the Phillips' type of argument might be to say that, although we estimate the equation as contemporaneous, it is in fact not so and as such need not be treated as part of a system. A more convincing response would of course be to estimate our equation by a simultaneous equation method. The requirements for doing so, however, would be that we know at least the variables present in the system which are not incorporated in our structure-performance equation. Unfortunately, theoretical discussion of possible simultaneous systems along the lines outlined above is at such a rudimentary (or even <u>ad hoc</u>) level that the information required cannot really be gleaned from them.[38] In the absence of reliable assumptions about the variables to use as instruments, simultaneous estimates are impossible. Again, to develop a simultaneous model of our own is rather beyond the scope of the present work. Thus we must simply resign ourselves to the fact that our empirical work in the contemporaneous case may be clouded by problems of simultaneity, with consequent biases which are hard to determine.

In chapter 6 in our empirical work we attempt to take some account of the additional factors considered relevant in the sections of this chapter.

FOOTNOTES

1. See Chapter 4 on some comparisons between the Cournot and joint profit maximisation bilateral power models.

2. This equation receives some further discussion in Appendix I.

3. A general movement affecting all industries to the same extent would be taken up in the constant term, as pointed out in the previous chapter.

4. This study is also mentioned in the final section of this chapter.

5. It is not clear whether or not Holterman includes a constant term, she does not report it.

6. The Khalilzadeh-Shirazi study and Caves' et al. work are particularly close for the sample is identical and only one variable is not shared. We discussed the construction of MEPS 10 and similar statistics briefly in chapter 3.

7. We discuss the position of barrier variables again later in the chapter.

8. We do not consider, except marginally, the place of either imports or exports in this thesis. Our reason for excluding consideration of exports is that we feel the government provides incentives sufficient to offset the risks (trade fairs, low cost insurance etc,). Concerning imports, the most obvious place is as a deflator of the market power variable. But this requires much more sophisticated data than is available. Dutton points out that the effect of imports on margins may not be unambiguous in the cross-section except in long-run equilibria. She also in some equations includes additional dummy variables concerning government regulation and monopolies commission and restrictive practice court findings. We cannot do justice to this work here; see her paper for details.

9. Dutton similarly finds that when she splits up her composite market power variable, significance is lost.

10. As we have said, we consider, particularly in the light of our decision to take a ratio form of the relationship, that "barrier" variables need not be included. Also, as we have indicated, we are not satisfied with the arguments adduced for the "growth" terms. In later sections of this chapter we do however examine the possible place of other variables.

11. For brevity we denoted this as Π/R when discussing the UK studies in the previous section.

12. This passage is deleted in the published article (1974).

13. We discuss this assumption in appendix I.

14. Our manipulations here are, we notice, vaguely similar to those in Ornstein's (1975) recent paper. However we believe his fundamental assumption, that return on capital is a function of the concentration ratio, to have a rather doubtful pedigree.

15. We are here considering capital analogous to labour in that it is measured as a flow of services K (units/time) and receives a "wage" in (£/unit). m is not the rate of interest (see e.g. Friedman 1970 p.244). In fact under certain definitional conditions we may dispense with m, as we mention later.

16. This is not quite fair to those who favour profit on equity, since they normally are able to subtract depreciation. But that is not the full (longer run) cost of capital since it is only a payment to maintain the service capability and does not include a return to the suppliers of that service. We neglect this distinction for simplicity. There is an additional point here, being a difference between US and UK studies. We are assuming the model to be short-run and so include fixed costs in

our definition of profit, distinct from variable capital costs. This is more applicable to the UK studies than those for the US where experimenters normally are able to consider the longer run by averaging most variables over a number of years. In their applications then, although fixed costs are not actually added to profits as measured, they are implicitly assumed to be small and, again for simplicity, we neglect the refinement.

17 It is of course easier and more usual to define K such that m = 1 (K being in £/unit of time rather than physical units of capital services /unit of time). This is perfectly possible both definitionally and dimensionally in the above scheme but has not been done in order to spell out the relationship, and the pitfalls, in more detail. Bain, (1959 p. 369) for example, immediately uses iE as the cost.

18 That is, subject to several qualifications concerning factors such as risk, non-monetary supplements to returns, and so on.

19 Many people have made the point that monopoly profits may be capitalised, for example Stigler (1963).

20 For a full definition of the terms, see the original article, pp. 217-8.

21 See for example, Jacquemin and Thisse (1972).

22 Where weights depend upon such factors as the discount rate and probability of entry.

continued on p.183

23. Subscript z has been omitted for simplicity. The generalisation can obviously be made since all we are doing is rewriting the left hand side in terms of (potentially) measurable quantities.

24. There are obviously other reasons if the argument is not acceptable as is clear from the limit pricing models of chapter 3.

25. These might also act as proxies for, or determinants of, the elasticity of demand.

26. In 1938, Kalecki commented briefly on Harrod's argument with the assistance of some calculations, and expressed disapproval of it on balance.

27. On which point, see appendix I.

28. Obviously we could consider preferences for other things than factors of production, but it would seem less reasonable that management should hold such preferences. For example, Baumol (1959) suggested managerial preference for sales, though it is difficult to see why sales should be taken beyond the profit-maximising level unless management gain utility because salaries or employment of staff are thereby increased. But in that case it seems more direct to consider they have a preference for staff employment, say.

29. An alternative, more complicated problem would be to include the wage bill along with profits in the utility function.

30. See also Cowling (1975) and references cited there for discussion on this topic.

31. Thus cyclical factors should influence the adherence to profit maximisation. Such effects might well be picked up by our proxy for cyclical behaviour discussed earlier.

32. It is barely conceivable that the reaction in $g(\frac{1}{|\eta|}, \ldots)$ to a drop in $|\eta|$ would be such as to overcompensate for the fall in $|\eta|$ itself.

33. That is if, for example, management obtains utility from profit only as a security against dismissal, but shareholders do not fully appreciate that a fall in numbers implies a greater profit potential.

34. This draws on the preliminary discussions about data in chapter 1, and the notes on data in Appendix II.

35. The Herfindahl index incorporates such changes in structure in a particularly straightforward manner, see Stigler (1968, Chapter 4).

36. Amongst others, Joskow (1975) bemoans this lack of concern over conduct.

37. For example, Williamson (1965).

38. For example, agreement even as to how many equations the system should contain would be unlikely. Phillips (1972), whose estimations we touched upon earlier, does attempt a simultaneous equation estimation of two alternative two-equation models. Besides the conventional structural effects on performance he also incorporates equations hoping either to explain the "effectiveness of price fixing" or the "propensity to attempt price-fixing agreements". It is not altogether clear how he chooses the variables to explain these nor why he focusses only on these particular aspects of conduct, although the latter may be partly because of some (confidential) data he was able to obtain. The results are not particularly interesting, except that in the "structure-performance" equation concentration and the advertising/sales variable (and their interaction) mostly maintain their significance of the ordinary least squares estimations. (The other equation of the system contains at most one significant variable.)

Chapter Six: EMPIRICAL ESTIMATION OF THE MODEL

I Introduction:

In this chapter we estimate an approximate version of the theoretical model developed earlier. We proceed by discussing data problems and introducing experiments as they occur naturally.

To reiterate what we said in chapter 1, the basic data used for empirical testing of the model emanates from the U.K. Censuses of Production (hereinafter, Census). The bilateral power index that we use requires knowledge of interrelationships between industries so the level of aggregation is that of the industry input-output tables for 1963 (with one exception).[1] The 1968 tables are less aggregative, but can be compressed to the 1963 level except that "Hosiery and Lace" and "Other Textiles" (1963) have to be amalgamated in order to produce comparable tables between the two years. This gives us a maximum of 58 "industries" covering the whole of manufacturing plus "Other Mining" and "Construction". There seems no reason why these last two should not be included. Problems of compatability in the Census industries between 1963 and 1968 are fairly minor as the 1963 figures were recalculated alongside those for 1968 in the 1968 Census, and provide no real basis for excluding any observations. Figures for later years than 1968 are according to the 1968 Standard Industrial Classification also.

Seven industries have however been excluded from the sample. The reason is that either the commodity is not produced mainly by the firms in that industry, or that the firms assigned to that industry do not produce mainly that commodity. Our definition of "mainly" here is of necessity somewhat arbitrary, but it provides a purely objective criterion. We decided to exclude those industries whose "specialisation" and/or "exclusiveness" ratios were less than 80% in either of the two years.[2] Any higher figure would have reduced the sample appreciably more. The industries we reject on these grounds, together with their Minimum List Heading (MLH) numbers are: Grain milling (211), agricultural machinery (331), engineers small tools (390), industrial engines (334), office machinery (338), other mechanical engineering (342,349), insulated wires (362). Our basic sample on which the experiments are performed is then of 51 industries, though since the dependent variable data are not available for the construction industry beyond 1968, we sometimes use a sample of 50.

Our main dependent variable, the ratio of profits plus fixed costs to revenue, comes fairly directly from census data. The numerator is net output minus wages and salaries, net output basically being defined as gross output minus materials (adjusted for stock changes), transport costs, etc.[3] As we said in chapter 5, the cost of capital services is not subtracted and in a later section we experiment with a slightly different form, hoping to exclude these costs if relevant. Aside from this, one problem with the dependent variable is whether enough cost elements are excluded. For 1963 and 1968 we are able, making certain straightforward though not necessarily correct assumptions regarding allocation, to subtract some additional cost elements including for example advertising expenditure. However, it turns out that the ratios calculated after extracting such cost elements from the numerator are so closely correlated with the original ratios that the experimental results are not worth reporting separately, and we do not discuss this measure further.[4] A second problem is that, unlike the theoretical model, firms in practice may increase or decrease stocks of the finished product over a year, so gross output need not equal sales. It is difficult to see what simple adjustment to the data would cater for this problem, but the difference between sales and gross output is of the order of one per cent, so hopefully the effects are not serious.

The independent variables require rather more discussion and are awarded separate sections.

II The Herfindahl Measure:

In this section, we first describe the measure calculated and then go on to our reasons for rejecting simple alternatives.

The Herfindahl index, as we have developed it, should theoretically be calculated for each industry according to the formula:

$$H = \frac{\Sigma q_i^2}{Q^2}; \qquad i = 1, 2, \ldots n \text{ firms}$$

Unfortunately neither volume outputs nor even sales data are available to perform this calculation. We have to use figures on employment size distributions to derive an approximation of the formula:

$$H_L = \frac{\Sigma L_i^2}{L^2}$$, which we might call the employment herfindahl.

Even if employment data exactly corresponding to this were used, there might be consequences for the estimation of the model, since it is well known that the distribution of firms by employment size has a smaller range than the distribution by size of sales, larger firms being more capital intensive in general.[5] In addition, data on individual firms' employment shares are not available in the census of production; we have to use the distribution of firm size by employment tables. This distribution combines enterprises into groups of three or more by size categories. The size categories, although following a general pattern, differ across industries and census years in order to avoid disclosure problems.

The most straightforward way to calculate the Herfindahl index is to allocate employment within each size class equally between the enterprises in that class, then to apply the formula given above. That is basically what has been done here. As a consequence, a value for our index greater than 0.33 would be impossible, as even if one firm employed 99% of the industry total, it would have to be grouped with at least two other firms in the tables. In effect then, such a calculation provides a minimum value for the Herfindahl, since any size variation within classes would increase the size of the statistic.

Unlike the case of the four-firm concentration ratio (on which see later), no maximum value for the Herfindahl can reliably be calculated. A calculation along these lines would involve allocating employment in each size class among hypothetical firms in such a way as to maximise the within-class variance, but the computational problems would be severe. For this reason we only employed the minimum value.

Of course, we must remember that the index we calculate is to be used in ratio form, one year upon another, so that the problem of the level of the Herfindahl is not as important to us as the ratio of the two levels as calculated. This may alleviate some of the difficulties referred to above. On the other hand, it becomes important to ensure that the ratio of 1968 over 1963 values bears some significance across industries. In particular, changes in the number of size groups for an industry between years can effect quite spurious changes in the ratio. For this reason, the size groupings we used were as nearly as possible identical between years for each industry. This involved a fair amount of consolidation in classification.

Three final points on calculation: Total industry employment was taken as the total including firms with less than 25 employees for each industry. The level of aggregation of the data is MLH or lower: In calculating indices for our "industries" where more than one size distribution table is involved, we used all of this information together, rather than taking arithmetic averages of the values obtained from each table separately. We should like to thank Malcolm Sawyer for supplying his calculated Herfindahl figures for 1963.[6]

Given the divergences between the actual measures calculated and the theoretical Herfindahls, it is probably as well to consider alternatives. Concerning the problem of employment as against output indices, we note that the only reasonable alternative available is the figures for five-firm concentration ratios calculated in the Census on a sales basis. While these might have a fairly high rank correlation with Herfindahl figures, there are major problems. Because they are at sub-MLH level, a considerable degree of aggregation would be needed, and it is not easy to see how this could reasonably be accomplished. Simple weighted averaging would not necessarily produce a meaningful statistic. In any case, there are gaps in these measures in that the sub-MLH categories for which they are supplied do not together normally cover the whole of the MLH industries of which they are part. In addition, some figures have to be omitted from the census on disclosure grounds. For these reasons, and also because the concentration ratio is less theoretically relevant, given our model, we rejected a measure based on this data. We also rejected the method of calculating concentration ratios from the employment size distribution tables, where both maximum and minimum values are obtainable and the average taken.[7]

Another plausible alternative would be to assume that firm sizes follow some useful statistical distribution. The prime candidate , on theoretical and empirical grounds, is probably the lognormal distribution. Assuming for the moment that this distribution does fairly describe the data then knowledge of only the mean and median size of firm for each industry is sufficient to give a unique value to the Herfindahl.[8] The median has of course to be interpolated so that again the Herfindahl obtained is not sacrosanct, yet it would probably be superior to our calculated value. We therefore turn briefly to evidence on lognormality.

For the UK, Hart and Prais (1956) chose a sample of "those companies engaged in mining, manufacturing or distribution which are quoted on the London Stock Exchange" (p. 154) for various years between 1885 and 1950. Their measure of size was market valuation, and the test of satisfactory fit to the observed distribution was based on the third and fourth moments of the distribution which have an expected value of zero. The fit is found to be "satisfactory" but unfortunately appears to get progressively worse through time. The main problem as far as we are concerned is that the results are for the whole of industry. It would appear to be statistically dubious to assume that our industries could be random drawings from that population, and there would be severe difficulties with small number industries. For this reason it is also interesting to refer to Silberman's (1967) comprehensive tests of the thesis on US data. His work is on 90 four-digit industries, and he obtained mean and variance of (logarithmic) size data from the census authorities directly. To determine the goodness of fit, he compared expected and actual 4,8, 20 and 50 firm concentration ratios over various size measures and census years. There was no significant difference between these four actual and expected measures in at most 42 out of his 90 industries. He concludes that "though the lognormal hypothesis cannot be rejected for specific industries, it is inappropriate to consider the function as a generalised statistical summary measure of the industry size distributions in manufacturing" (p. 809). Again unfortunately, he was "unable to discover a consistent relationship between the occurrence of lognormality and specified attributed of the industries which were well described." (p.831).

Finally, we chose one industry (Bacon Curing, Fish and Meat Products) at random from the (UK) Census for 1963 and calculated its Herfindahl based on the lognormality assumption. The value was lower than the (minimum) value obtained by the method outlined earlier. Such results, if common, would pose a problem if we decided to use the lognormality assumption, as they obviously lead to misleading values. Thus, despite its limitations, we use the Herfindahl calculated in the straight-forward manner we described. A modest experiment attempting to judge some effects of the problems with our measure is noted below; it does not lead to particularly interesting results though.

III The Bilateral Power Index

Recall that the formula for calculating the bilateral power index is:

$$H_{BUY\,i} = \sum_j \left[\frac{a_{ij}}{1 - \frac{H_j}{|n_j|}} \right]$$

(notation adapted from equation (32) chapter 4) for each industry i.[9] A number of problems arise in approximating this formula.

One difficulty is that firms sell to industries outside the remit of our sample, for example to the sectors: Agriculture, electricity, distributive trades and final buyers. For none of these do we have Herfindahl figures (H_j in the formula) from the Census, so that approximations have to be calculated by more or less ad hoc means. We here briefly describe their sources:

Agriculture: According to the EEC Yearbook of Agricultural Statistics (1967), there are 393,000 holdings of one hectare or more in the UK. We assumed in both years that this industry was so diverse that it warranted a Herfindahl of zero.

Forestry and fishing: We assumed that the main concentration of sellers was the nationalised sector of the forestry industry, which we took as a single seller. The proportion of total forestry owned by the Commission in 1962/64 and 1967/8 was applied (figures from the Annual Abstract of Statistics, AAS), and this figure further weighted by forestry as against forestry and fishing total output.

Coal mining, gas, electricity, communication: For all these, Herfindahls of 0.9 were assumed in both these years, thereby taking it that approximately 95% of all output in these industries is produced by the nationalised industry concerned (assumed to act as a single seller).

Water: Census figures as in our main sample, are available for this industry.

Road and rail transport: Each of the various nationalised and local government bodies in this area was entered as nearly as possible as a separate enterprise, size being measured by receipts (AAS figures). The remainder was considered as too diverse to affect the Herfindahl value, but total revenue was used in the denominator.

Other transport: This is of course mainly sea and air transport. We consider only BOAC and BEA as separate firms, with the remainder assumed diverse. Again we used receipts for the two airlines, compared with total industry revenue (source: annual reports).

Distributive trades: Rough values were calculated here from the Census of Distribution for 1961 and 1966, and these were used as 1963 and 1968 figures respectively.

Miscellaneous services: This is an extremely diverse catch-all category and we assumed the value of zero for the Herfindahl.

Consumers and export purchasers: Also assumed to have Herfindahls of zero in both years.

Public administration and defence, health and education: We considered that the important purchasers were the defence industry, the

National Health Service and the Local Authorities; we took it that the first two were single purchasers. The Herfindahls were calculated on the basis of these bodies' expenditure compared to the total for the years 1963 and 1968 (source: input-output table, e.g. table P 1968).

A second problem with the input-output tabulation is that while current purchasers are allocated fully, capital formation is aggregated to a single column. In the absence of any simple alternative, we decided to apportion this figure (normally small) among buyers in the proportions in which they purchased current output. The input-output tables are then used to calculate the a_{ij} values.

Undoubtedly the main difficulty with applying our formula for the bilateral power index directly to the data is that we are required to know the elasticity of demand of each purchaser for each industry's product. We cannot deny that this is a substantial problem to which there is no satisfactory solution. As we saw in chapter 1, relevant empirically determined elasticities for our sample are not available.[10] Without the benefit of specific knowledge we have to make some assumptions about the elasticities in order to perform the experiment at all. Our favoured solution is to give demand elasticities the same numeric value across all purchasers, industries and both time periods. This solution, it must be admitted, is rather against the spirit of the thesis, but we see no real alternative. We of course expect that different numbers would yield different results, so we experiment with three values.

Before discussing the outcome, there are two side issues which should be mentioned. Firstly, given that we are assuming values for the elasticity in the bilateral power index, it is fair to ask whether the further assumption that elasticities are constant across all industries at a given value might be made. The consequence of this would be to render the model amenable to regression as a "level" rather than a "ratio" equation. There are two points here. The first is that, as will be clear from our earlier work of chapters 2 and 3, a ratio form allows much more flexibility in the interpretation of the theoretical background to the model as more factors are held constant. The second is that the assumption of equal elasticities in the bilateral power measure does not restrict the model as strongly in ratio form as in level form. Assuming identical elasticity values imposes an erroneous weighting on the various purchasing industries. All other things equal, if the concentration in a purchasing industry rises, the effect on the calculated ratio will be in the desired direction, albeit not of the right magnitude. In level form, an erroneous weighting might mean that one industry has a higher bilateral power index as calculated, yet a lower true value for this index. For these reasons we reject the possibility of performing experiments on a "levels" equation as being less general than the "ratio" form.

Secondly, in experimenting at the present stage of development with the various bilateral power indices provided by assuming alternative numbers for elasticities, we accept that there is an omitted variables problem. A superior procedure would be to include all potentially relevant variables. We feel unwilling to adopt this approach at this stage, because there are already a number of variants involving lag structures and so on that we have in mind, and to use every permutation at every stage would make the experimental series exceedingly long. If, as it turns out, none of the bilateral power indices we try is at all highly correlated with other independent variables, then the biases introduced by following our preferred procedure should not be severe.

Turning now to the tests, note first that assumed elasticities across all industries much less than unity are unacceptable as they cause difficulties with the formula in connection with nationalised industry purchasers. We therefore arbitrarily chose three values of 1.0, 1.5 and 2.0 for all elasticities of demand. The results of estimating the model:

$$\log \Delta\left(\frac{\Pi+F}{R}\right) = \log A + \alpha \log \Delta H + \beta \log \Delta H_{BUY} + V$$

on the data thereby generated are listed in table 1. We also provide in table 1A a correlation matrix showing collinearity between the three measures generated under the alternative assumptions.

What these results seem to indicate is at least moderate support for the inclusion of the H_{BUY} measure. It is encouraging to note that it (almost) attains significance at the 5% level under all three assumptions about the elasticity of demand. It is also extremely encouraging, in view of the rather arbitrary assumptions about elasticity of demand made, to see that the three alternative bilateral power measures used are very highly correlated with each other in our sample. Finally, it is fairly clear from this experiment that we should utilise the values for the bilateral power index obtained by setting elasticity of demand equal to unity for further work, since both the significance of that variable, and the overall explanatory power, are better in the first regression. We save further comments on the estimation until we have tested our more complete model.

IV The Basic Estimations

We move now to a discussion of the results obtained in estimating an equation of the form:

$$\log \Delta\left(\frac{\Pi+F}{R}\right) = A + \alpha \log \Delta H + \beta \log \Delta H_{BUY} + f(\Delta U) + V$$

(similar to equations(16) and(17) chapter 5, except for the addition of a disturbance term).

TABLE 6.1

Dependent variable: log ($\frac{\Pi+F}{R}$ 1968 / $\frac{\Pi+F}{R}$ 1963)

Equation	Assumed value for elasticity of demand	Log constant	log ΔH	log ΔH_{BUY}	R^2	Sample size
1	1.0	0.0022 (0.996)	0.2437* (3.412)	0.4541* (2.460)	0.2670	51
2	1.5	0.0220 (1.010)	0.2333* (3.340)	1.9170* (2.167)	0.2547	51
3	2.0	0.0022 (1.006)	0.2349* (3.225)	3.1760 (1.994)	0.2371	51

t statistics in parentheses

* indicates significance at 5% level or better on a two-tailed test

TABLE 6.1A

Correlation between bilateral power indices (log ΔH_{BUY} values)

Assumed value for demand elasticity	1.0	1.5	2.0
1.0	1		
1.5	0.9781	1	
2.0	0.9573	0.9963	1

The additional variable, the percentage unemployed, is defined as Unemployment/(Unemployment + Employment) in that industry. Figures for this variable were obtained from the AAS; they can be consolidated to fit our sample of industries with very little difficulty indeed.[11]

As we said in the previous chapter, we decided to perform the test using both a contemporaneous and lagged form, that is with the dependent variable being based on the ratio of the years 1968 and 1963, and also the ratio of the years 1972 and 1968. We take the former first. Since figures on the price-cost margin are not available for the construction industry beyond 1968, we decided to estimate this equation with and without the inclusion of that industry to facilitate comparisons with results for the lagged relationship. We also experiment by using the unemployment proxy as a normal ratio and as a logarithm of that ratio, both with contemporaneous employment and assuming unemployment to lag behind other cyclical changes. This gives us a series of eight equations which are listed in table 2.

We notice a number of interesting features arising from these estimations. In all equations, both the Herfindahl and the bilateral power measures attain significance at least at the 5% level, confirming our earlier partial estimation. The constant never attains significance at that level, neither does any measure of the unemployment variable. While both of our structural variables are very significant neither has a particularly high coefficient and both are significantly less than unity at the 5% level, suggesting a smaller adjustment than indicated by the pure Cournot model (see Chapter 2) which we considered in some sense a minimum. Therefore, either full adjustment to the structural change has not had time to take place, or other considerations prevent such adjustment, or costs rise when structure becomes more concentrated, or finally our data or estimation procedure is such that it biases the coefficients

TABLE 6.2 Dependent variable: $\log \{\frac{\Pi+F}{R} 1968/\frac{\Pi+F}{R} 1963\}$

Equation	Unemployment variable	Constant	log ΔH	logΔH_{BUY}	Unemployment variable	R^2	Sample size
1	U_{68}/U_{63}	0.1058 (1.672)	0.2198* (3.175)	0.4566* (2.580)	-0.0752 (-1.412)	0.3021	51
2	$\log(U_{68}/U_{63})$	0.0280 (1.285)	0.2170* (3.103)	0.4526* (2.556)	-0.0803 (-1.341)	0.2994	51
3	U_{69}/U_{64}	0.0508 (0.824)	0.2421* (3.431)	0.4400* (2.448)	-0.0184 (-0.505)	0.2765	51
4	$\log U_{69}/U_{64}$	0.0421 (1.190)	0.2413* (3.463)	0.4426* (2.469)	-0.0466 (-0.729)	0.2807	51
5	U_{68}/U_{63}	0.1037 (1.605)	0.2237* (3.099)	0.4655* (2.540)	-0.0734 (-1.348)	0.2939	50
6	$\log(U_{68}/U_{63})$	0.0278 (1.260)	0.2205* (3.018)	0.4607* (2.510)	-0.0782 (-1.271)	0.2910	50
7	U_{69}/U_{64}	0.0486 (0.778)	0.2477* (3.404)	0.4560* (2.448)	-0.0171 (-0.462)	0.2694	50
8	$\log(U_{69}/U_{64})$	0.0408 (1.137)	0.2465* (3.423)	0.4569* (2.459)	-0.0440 (-0.6756)	0.2732	50

(t statistics in parentheses)

* Indicates significance at the 5% level or better on a two-tailed test.

downward. (In general, as we said in the previous chapter, between 1963 and 1968 most industries become more concentrated[12]). The insignificance of the constant is perhaps unsurprising given that the (unweighted) mean value of the ratio of 1968/1963 unemployment is barely over one (1.071) in our sample, indicating quite possibly that there are similar levels of general economic activity in the two years. It remains possible that, using alternative data to proxy differential cyclical changes, we might have obtained a significant coefficient for that cyclical change variable. One slightly surprising feature of the results is that including "construction" as an observation increases significance while adding an extra degree of freedom. Thus, except when making comparisons between this set of results and others where construction figures are unavailable, we choose equation one as our favoured equation. The equation as a whole is significant at better than 1% on an 'F' test.

As we said in chapter 5, one possibility which presents itself is that it takes some time for performance to react to structural changes. For this reason we estimated a similar set to the above equations using the 1972/1968 value of the dependent variable. The results are in table 3. These regressions present a totally different picture to the estimations already discussed, for neither the Herfindahl nor the bilateral power measure are in any case significant, and they are usually of "wrong" sign. Also in contrast the unemployment variable shows significance in many of its guises. There are a number of possible explanations for this turnaround. For example, structural variables may act quickly on performance, perhaps not immediately but maybe with a fairly short lag, though this did not appear to be the case in Cowling and Waterson (1976). We are unable to try any shorter lag here though, due to lack of performance figures between 1963 and 1968.[13] Again, the effect of incomes policies may have

TABLE 6.3 Dependent variable: $\log \{\frac{\Pi+F}{R} 1972 / \frac{\Pi+F}{R} 1968\}$

Equation	Unemployment variable	Constant	log ΔH	$\log \Delta H_{BUY}$	Unemployment variable	R^2	Sample size
1	U_{68}/U_{63}	0.1090 (1.059)	-0.1003 (-0.874)	-0.0045 (-0.015)	-0.0820 (-0.946)	0.0293	50
2	$\log(U_{68}/U_{63})$	0.0229 (0.652)	-0.0995 (-0.855)	-0.0111 (-0.038)	-0.0720 (-0.734)	0.0219	50
3	U_{69}/U_{64}	0.2568* (2.818)	-0.0349 (-0.329)	0.0004 (0.002)	-0.1514* (-2.807)	0.1551	50
4	$\log(U_{69}/U_{64})$	0.1244* (2.333)	-0.0581 (-0.544)	-0.0014 (-0.005)	-0.2442* (-2.528)	0.1312	50
5	U_{72}/U_{68}	0.4435* (3.700)	-0.0462 (-0.460)	0.0142 (0.055)	-0.2061* (-3.677)	0.2351	50
6	$\log(U_{72}/U_{68})$	0.3342* (3.547)	-0.0466 (-0.461)	-0.0084 (-0.032)	-0.4479* (-3.556)	0.2238	50
7	U_{73}/U_{69}	0.0076 (0.056)	-0.0774 (-0.668)	-0.0199 (-0.067)	0.0069 (0.073)	0.0105	50
8	$\log(U_{73}/U_{69})$	0.0072 (0.148)	-0.0722 (-0.622)	-0.0258 (-0.087)	0.0321 (0.289)	0.0122	50

t statistics in parentheses

* indicates a coefficient significantly different from zero at the 5% level or better on a two-tailed test

been to curtail performance changes. Or again, the comparative slump position existing in 1972 as measured by the ratio of 1972/1968 unemployments (unweighted) mean value of 2.096 might have prevented movement towards higher profits.

Our favoured equation, based on goodness of fit, is no. 5. As with the estimations of table 2, contemporaneous unemployment ratios appear to add most to the explanation. Looking in slightly more detail at this equation, we find that the constant is significantly positive, while the unemployment variable is significantly negative. This suggests that performance (as understood here) would have risen greatly but for the rise in unemployment indicating a depression causing a fall in margins, via elasticity of demand changes or otherwise. It is perhaps unwise to speculate on the reason for the significantly positive constant value, it could proxy a structural change after 1968,[14] or a number of other effects. Similarly, as we said in chapter 5, there is certainly more than one explanation of the negative sign on unemployment. We continue our discussion of experiments on the assumption that performance reacts quickly to structural changes so that equation 1, table 2 is our main focus for further work.

We noted in chapter 5 that there is a possibility of durable and non-durable goods being affected differently by cyclical changes;[15] we attempt to test this. One problem which immediately crops up here is that of finding an appropriate definition of a durable good. We decided, <u>faut de mieux</u>, to select statistical criteria for the definition. Using the input-output tables we calculated the proportion of each industry's output which was sold to purchasers for gross domestic capital formation in 1963 and 1968. If this proportion exceeded a certain value over the two years then the good was considered as "durable". We took three proportions: 1%, 5%, 10% and also a fourth criterion: if some output was sold for this purpose in both years.[16] Using this information, we performed a complete covariance analysis for homogeneity of the whole sample. This was done using the specifications of equation 1 table 2 and of equation 1 table 1, the latter being included in case the effect of durability was masked by the presence of the unemployment variable. This procedure, rather than that of using "durability" as a continuous variable, was followed because we did not expect our measure of durability to be very accurate. The covariance analysis basically involves a regression on all the data, a regression on all data with a dummy variable for durability, and separate regressions on "durable" and "nondurable" observations. The residual sums of squares are then compared with the F statistic in the manner described in Johnston (1972a) pp. 192-207, for example. In every case, whatever the definition of durability and whether or not unemployment was included as a variable, we were unable to reject the null hypothesis of homogeneity of the complete sample at the 5% level; both intercept and slope are insignificantly different between subsamples. (For example, testing for homogeneity of the complete relationship and using equation 1, table 2, the theoretical F value, $F(4,43)$ at 5% is 2.61, while the calculated statistic takes on values 0.9390, 1.073, 0.7481 and 0.5855, depending upon the definition of durability). It is of course possible that alternative definitions of durability might indicate the need for separate treatment or an additional coefficient.[17]

We also discussed in the previous chapter the potential problem with our work in that capital costs representing the deterioration of equipment due to use in production are not subtracted from the numerator of our performance measure. We argued that under certain assumptions this might be reasonable, but the assumption should be tested and we suggested subtracting various proportions of (some proxy for) capital costs defined in this manner.[18] This procedure is statistically superior, given our model, to the alternative of using capital costs as an independent variable. Whatever the method used though, we run into the problem of available data. We decided to take the census values for capital expenditure in the year in question as a proxy for capital costs as discussed above; this is only reasonable on the assumption that firms are spending regularly, purely on replacement of equipment worn out in production. As such the figures used could over- (or even under-) estimate the true figures.[19]

Our method is to subtract a proportion of capital expenditure from the numerator of the profit-revenue ratio. We chose to subtract tenths at a time,giving ten regressions of which a selection are listed in table 4. These should be compared with our "favoured equation" (equation 1 table 2). The results are interesting but also slightly surprising. First, as we increase the proportion of capital expenditure subtracted, the overall goodness of fit falls. But at the same time, the Herfindahl increases its significance up to the stage at which 0.3 of capital expenditure is subtracted and its coefficient continues to increase in size throughout the whole series. Meanwhile, the bilateral power variable falls off in significance and drops below the 5% level after 0.3. The unemployment variable always remains insignificant at 5% and its t value falls off. Thus if we are concerned only with overall fit, we should choose our favoured equation of table 2, while if our main purpose is the Herfindahl's significance, equation 3 table 4 (or some value subtracted between 0.3 and 0.4) would be the most interesting. Given our overall approach, we incline towards the former. However we do notice that, in contrast to the results of Ornstein (1975, p. 111), subtraction of capital expenditure as a cost does not remove significance from the monopoly power variable. Of course, since the capital variable used is

TABLE 6.4 Dependent variable: $\log\{(\frac{\Pi+F}{R})\ \frac{1968}{1963}\}$ with subtractions of part of capital expenditure from "Π+F"

Equation	Proportion of Capital Expenditure subtracted	Constant	log ΔH	log ΔH_{BUY}	U_{1968}/U_{1963}	R^2	Sample size
1	0.1	0.1004 (1.548)	0.2307* (3.254)	0.4419* (2.438)	-0.0745 (-1.366)	0.2990	51
2	0.2	0.0946 (1.411)	0.2428* (3.310)	0.4246* (2.264)	-0.0741 (-1.312)	0.2929	51
3	0.3	0.0886 (1.263)	0.2563* (3.340)	0.4039* (2.059)	-0.0738 (-1.250)	0.2835	51
4	0.4	0.0822 (1.105)	0.2715* (3.339)	0.3791 (1.824)	-0.0740 (-1.182)	0.2709	51
5	0.5	0.0753 (0.943)	0.2889* (3.307)	0.3489 (1.562)	-0.0745 (-1.109)	0.2553	51
6	0.6	0.0678 (0.778)	0.3091* (3.242)	0.3114 (1.278)	-0.0757 (-1.032)	0.2373	51
7	0.9	0.0376 (0.293)	0.4030* (2.867)	0.1143 (0.318)	-0.0847 (-0.783)	0.1746	51
8	1.0 (or, the whole)	0.0195 (0.123)	0.4619* (2.676)	-0.0193 (-0.044)	-0.0904 (-0.681)	0.1526	51

t statistics in parentheses

* indicates a coefficient significantly different from zero at 5% or better.

not theoretically correct, we can say little about whether we have generated evidence in favour of the Spencian position[20] (see chapter 5). Because the fit on all variables falls off sharply as large proportions of capital expenditure are subtracted though, we would be inclined to conclude that his simple model contains some element of truth.

We performed a similar experiment to that described above on our favoured equation of table 3 (1972/1968 values for the dependent variable). The overall fit falls throughout, the coefficient on ΔH moves from insignificantly negative to insignificantly positive. We do not consider it worthwhile to detail the results obtained.

V Further Statistical Considerations

Other structure-performance studies for the UK have encountered quite high correlations between independent variables, but multicolinearity does not appear to be a problem in our sample. For example, in our favoured equation of table 2, the correlation between the three independent variables is very low; between the Herfindahl and bilateral power variables the correlation coefficient is 0.0172 and between these and the unemployment variable, - 0.1637 and 0.0789 respectively.

Another statistical problem which Comanor and Wilson (1967) noted in their study of the US was that of heteroskedasticity. They tested for this by splitting the sample into four and calculating the variance of residuals in each quartile. Based on this tabulation they derived an appropriate weighting scheme. We decided to use the more formal parametric test of Goldfeld and Quandt (1965) in considering the problem of heteroskedasticity in connection with our work.[21] We felt that the most likely reason for possible difficulties in this area would be reaction to large changes in industry structure. That is if the Herfindahl say rose sharply between 1963 and 1968 in a certain set of industries, we might well expect the variance of the performance changes resulting to be larger than the variance of performance changes in a set of industries where structure changed little. Once the particular variable which might effect heteroskedastic disturbances has been chosen, the sample is then ordered according to the size of that variable and split into equal segments with a number of observations around the middle omitted. The authors give some guidance on the size of the omitted segment formulated from a "Monte Carlo" study, but no hard and fast rules can be adopted. Based on their work, we chose to rank our observations by the magnitude of the Herfindahl variable and set the size of the omitted middle portion at 11, 13 and 15 observations (samples of sizes 20, 19 and 18 respectively). In each case the two subsamples were estimated independently and a ratio of the sum of squared residuals formed. This sum is compared with the F statistic with appropriate degrees of freedom, the null hypothesis being that the sample is homoskedastic. If the squared residuals are significantly larger in one sub-sample than the other, we reject the null hypothesis. The ratios of the sum of squared residuals in the subsample of larger Herfindahl changes divided by that value for the subsample of smaller Herfindahl changes, together with the appropriate theoretical F statistic at the 5% level are listed below :

Sample sizes	Calculated ratio	F statistic 5%	Degrees of freedom
20	1.83	2.33	16, 16
19	2.40	2.40	15, 15
18	2.28	2.48	14, 14

Based on these results, we have scant evidence that the sample as a whole is not homoskedastic, the calculated ratio being smaller than (or exactly equal to!) the theoretical value and so inside the acceptance region. For, whilst we are not assured that the test will give us an unambiguously clearly defined central omissions region for maximum power, it can be argued that the power will increase and then diminish as the central gap is widened (see e.g. Goldfeld and Quandt p. 541 n. 8). If at a (supposed) maximum we are only on the borderline of rejecting the null hypothesis, then it seems fair to proceed on the assumption that heteroskedasticity is not a problem in our work. Having said this, it is possible that alternative tests, perhaps based on other size orderings, would give different results, but the number of potential alternatives is too large for full exploration.

As we have mentioned in various places previously, all of our independent variables are potentially measured with quite large errors. If there is a persistent tendency towards over-estimation of the Herfindahl ratio (say), then this might be one explanation for the low coefficient obtained. But as is well known, measurement errors even of zero expectation give rise to biased and inconsistent estimation of the model parameters. The bias, moreover, is towards the coefficients being too small as estimated. Thus in some degree the low coefficient on the Herfindahl ratio might be explained by the measurement error problem. If we proceed here on the assumption that there is no persistent over or under estimation, then consistency may be achieved by the use of appropriate instrumental variables.[22] As usual though, candidates do not immediately spring to mind as ideal. Perhaps one of the best is Durbin's suggestion of using the rank of the particular variable in the series as an instrument both highly correlated with the actual variable size and hopefully not related to the error in the limit (see Johnston 1972a pp. 285-6). We performed such an instrumental variable estimation on equation 1 of table 2 (the "favoured equation") using the rank of the variable as an instrument for each series. The following equation was obtained:

$$\log \widehat{\Delta\left(\frac{\Pi+F}{R}\right)} = \underset{(1.604)}{0.1079} + \underset{(2.426)}{0.1737} \log \Delta H + \underset{(2.019)}{0.4092} \log \Delta H_{BUY}$$

$$- \underset{(-1.215)}{0.0692}\, \Delta U, \qquad R^2 = 0.2937 .$$

(t statistics in parentheses, sample size 51, all variables 1968/1963 values).

Comparing this equation with our favoured equation, we see that the overall fit is poorer, the coefficients and their t values smaller! There are several explanations for the fall in the size of the coefficients:[23] We must remember that we do not achieve unbiased results even if the instruments are good, the instruments may only poorly reflect the ranking of the "true"values, and measurement errors involved may not in fact have the desired properties. It is difficult to apportion "blame" for the unexpected result obtained, we must simply accept that what appears to be the best available instrument to deal with the problem encountered has not had the effect we hoped for.

VI Conclusion

We are thus drawn back to our favoured equation of table 2 as representing best what we consider to be a good partial explanation of the underlying relationships between changes in industrial structural variables and changes in performance. It is repeated below:

$$\log \widehat{\Delta\left(\frac{\Pi+F}{R}\right)} = \underset{(1.672)}{0.1058} + \underset{(3.175)}{0.2198} \log \Delta H + \underset{(2.580)}{0.4566} \log \Delta H_{BUY} - \underset{(-1.412)}{0.0752}\, \Delta U$$

$$R^2 = 0.3021$$

(t statistics in parentheses, 51 observation, all ratios 1968/1963).

Taking this equation at face value, we see that a rise in industry concentration has a firm, but fairly weak, positive effect on the price cost margin, as does an increase in the bilateral power index. This is at variance with the conclusions of some of the other UK studies cited earlier, although similar to the main conclusion of Cowling and Waterson (1976).[24] There is obviously a myriad of other influences on the change in the price-cost margin between these two years, as is only to be expected when trying to explain such a volatile variable;[25] it is to be hoped that our omission of many possible effects does not seriously bias the results as estimated.

We resist the temptation to include any discussion of policy implications arising from our work; we would feel much more confidence in doing so if and when similar results are generated using data from other time periods for the UK.[26]

FOOTNOTES

1. There is obviously an aggregation problem here, as many of these categories are fairly broad. We do not pretend to be able to provide a complete answer to this difficulty.

2. The definitions of these ratios are: the "output of 'principal products' of each industry group as a percentage of industry group total output" and the "output produced as 'principal products' as a percentage of total output of the commodity group" (Input-output table 8, 1968), respectively.

3. The full definition of net output, along with definitions of the other components of the dependent variable, are listed in the introduction to each census.

4. The correlation between these two measures is 0.9775. The additional costs subtracted are those listed in Table 12 of the individual industry reports, where given. We decided to report results only for the case where the additional costs are not subtracted to facilitate comparisons between the

contemporaneous and lagged equations.

5. See Smyth et. al. (1975), who discuss the problems of surrogate measures and the conditions under which they will give a correct representation of the true measures.

6. Because Sawyer's purpose and sample size were different only some of his figures have been used here, but they eased the burden of calculation.

7. These are what Sawyer (1971) tabulates in his article. We would expect the distance between maxima and minima to be further apart in general for our sample than his, as ours is the more aggregative. In Cowling and Waterson (1976), concentration ratio figures are used alongside Herfindahls, with "inferior" results.

8. I am grateful to Steve Davis for pointing this out to me.

9. i is the selling industry, j the purchasers; $a_{ij} = q_j / Q_i$, sales to industry j divided by total sales of the ith industry.

10. Deaton's (1975 b) work, heavily based on final consumption, would allow us at most a sample of five industries and seven purcahsers.

11. The only slight problem is that in later years in two cases the employment figures were too aggregative. We apportioned employment in these cases as it was apportioned in the previous year. Employment figures (as opposed to unemployment) change little from year to year and such apportionment is likely to cause very little error.

12. The unweighted mean of the logarithmic ratio is 0.0642, whereas if concentration had not increased we would expect a value of zero.

13. For example we cannot relate a 1970/1965 performance measure to 1968/1963 structural changes because the census taken in 1965 does not contain anywhere near enough information. (See Appendix II)

14. We are unable to obtain figures for the Herfindahl, for example, after 1968.

15. They may also be affected differently by structural changes because of the opportunities open to sellers, see Cowling and Waterson (1976).

16. Higher proportions than those mentioned would have left few observations in the "durable" category. These statistics presumably do not include consumers' capital expenditures. It might then be objected that we are picking up a partial measure of producer (as against consumer) goods. However, casual inspection of the figures indicates that most categories of goods commonly thought of as durables are picked up as such in our measures (eg. domestic electrical appliances).

17. In Cowling and Waterson (1976), using a subjective definition of durability, markedly different results for the two subsamples involved were obtained.

18. These capital costs should be distinguished sharply from costs fixed in the short-run. We refer here to factors such as wear and tear on machinery caused by using it in production.

19. As we have seen in adding the "capital-output ratio" as an independent variable, Shepherd (1972) used a similar measure to ours.

20. This will depend upon the relative sizes of the capital variable actually subtracted and the theoretical values for that variable.

21. They took only the case where the disturbances were heteroskedastic in a particular manner. In fact though, as Glesjer (1969) makes clear, the test is more general than might be implied by their description. Johnston(1972a) p 221 suggests the Goldfeld and Quandt test as the most useful for exploratory purposes.

22. This is of course a very rough experiment. There is no particular reason why the assumption should be true, nor a reason why the error should exhibit constant (asymptotic) variance and be uncorrelated with the general error term.

23. It is not particularly surprising that the overall fit is poorer.

24. One of the main equations from Cowling and Waterson is:

$$\log \widehat{\Delta \left(\frac{\Pi+F}{R}\right) \left(\frac{1968}{1963}\right)} = \underset{(0.728)}{0.0220} + \underset{(2.572)}{0.2501} \log \Delta H \left(\frac{1963}{1958}\right),$$

$R^2 = 0.067$, 94 observations (MIH level), t statistics in parentheses. (notice the lag structure)
The addition of "trade union" and "durable" variables did not add materially to the fit but, as mentioned earlier, splitting the sample into "durablo" and "nondurable" goods

categories gave quite a good fit for the former, very poor for the latter. Apart from the "trade union" variable, bilateral power was not considered in that paper.

25. As we said in chapter 1, though we are able to explain a smaller proportion of the variance in our dependent variable than, say, Khalilzadeh-Shirazi (1974) is in his, we most likely have far more to be explained.

26. The results of Cowling and Waterson (1976) cannot be said either to be from a completely different time period, nor to be entirely in agreement with those presented here.

Appendix I : Two Assumptions discussed

The Assumption of Constant Elasticity of Demand

The theoretical work of the preceding chapters has relied to some extent upon the assumption that demand for every industry's product is constantly elastic. This is true particularly of the work in developing the bilateral power model, but the limit pricing model also involved the same assumption. Having said this, it remains the case that both models could have been developed to similar levels from alternative suppositions about demand curves, though the conclusions would obviously have differed in some respects. We do not of course require for our results that every demand curve is exactly constantly elastic; as long as the double-log form fits observed patterns reasonably faithfully, even if other formulations will do as well, then our conclusions remain more or less intact. Given this position we should consider, albeit briefly, the evidence from demand studies bearing upon the viability of the constant elasticity of demand form.[1]

First though, it should be pointed out that if we do not assume (in the sense described above) constant elasticity of demand then a further problem is created for empirical work. Using equation (1) of chapter 5 (equation 33, chapter 4), assume that there is a change in the concentration of that industry. We have:

$$\frac{\Pi + F}{R}\, z = \frac{H_z(1+\upsilon_z)}{|\eta_z|} \cdot H_{BUY\ z},$$

Thus:
$$\frac{\partial\left(\frac{\Pi+F}{R}\, z\right)}{\partial H_z} = \frac{(1+\upsilon_z)\cdot H_{BUY\ z}}{|\eta_z|} + \frac{H_z \cdot H_{BUY\ z}}{|\eta_z|} \cdot \frac{\partial \upsilon_z}{\partial H_z} - \frac{H_z(1+\upsilon_z)\, H_{BUY\ z}}{\eta_z^2}\, \frac{\partial|\eta_z|}{\partial H_z}$$

As we explained, our particular formulation (equation (2) of chapter 5 is the simplest case) is designed to allow for the first two types of effect present in the equation above. However, it cannot fully allow for the third effect, which effect is zero in the constant elasticity of demand case. The point here may be illustrated with reference to a linear demand curve. If the number of firms drops, H_z will rise, price should rise, and total output falls; there is a movement up the demand curve. But then demand becomes more elastic so that $\partial|\eta_z|/\partial H_z$ is positive and the rise in the margin is less than would be expected from a constant elasticity curve. Problems are caused here because, even in the pure Cournot case ($u_z = 0$), we cannot say with any certainty what the magnitude of the change in margin will be and, a partly connected point, the level of marginal cost will have an effect on the change in margin expected.

One point which might worry some is that if we assume that demand is constantly elastic then we must assume all demand curves to have elasticity greater than unity in order to cater for monopolists. This is an overstatement though. All that is really required is that demand is approximately constantly elastic in the "normal region of observation". If there are at present fifty firms in the industry then the demand curve can be very inelastic as long as we do not expect dramatic changes in the structure of that industry.[2]

It is well known that, despite being consistent in differential form[3] the double-log or constant-elasticity of demand function is inconsistent in many respects with the postulates of consumer theory, especially when we consider extrapolation from given data. Despite this there seems to be no basis for rejecting it out of hand for two main reasons: it is not necessarily true that the postulates hold in practice, and within the region of observation it may provide as good an explanation of reality as alternatives satisfying the postulates.

In fact when we turn to the evidence on the veracity of the properties of a complete demand system, following from the assumptions that consumers maximise utility while spending all their budget, we find that, despite Barten's (1967) early encouraging results with a four commodity system, later work at a less aggregative level has tended not to bear the properties out too well. For example Deaton (1974) in his nine commodity study of the U.K. finds that intercepts are necessary (though not implied by the pure theory) and that homogeneity of degree zero cannot be accepted for all commodities, though "symmetry" is accepted and "negativity" almost so (given homogeneity).[4] He is at rather a loss to explain the rejection of such a straightforward effect as homogeneity.

Interestingly, Deaton also tests for "additivity", a property not required from utility theory but often imposed by those who estimate demand equations. This is rejected at a very high level of confidence. Such a finding is important because probably the most widely used alternative to the "pragmatic" constant elasticity format is the Linear Equation System which imposes additive preferences at the level of the utility function.

In a more recent paper Deaton (1975a) makes a direct comparison between the constant elasticity and linear expenditure models using a more general non-additive framework as a benchmark. Concerning the price elasticities he finds that the constant elasticity format provides a much closer approximation to the simple non-additive model than does the linear expenditure system. Thus:

"The effects of choice of functional form on measurement depend upon the amount of information in the data, but in any case are small relative to the effects of assumptions, such as additivity, which destroy much of the available evidence." (p.272)

As far as the constant elasticity form is concerned he feels that this "model remains a useful means of summarising the evidence over the sample period" (p.262).

Such a conclusion is most reassuring from the point of view of the assumption of constant elasticity of demand in our model. We further note that Stone (1954) who uses this form throughout his extensive work on demand estimation considers that "these simple relationships may be said on the whole to fit the observations reasonably well" (p.278). Also Deaton (1975b) in his study of thirty seven commodities, while mainly favouring the linear expenditure system for his estimates of demand relationships,[5] uses the constant elasticity form for many commodities. His estimates of the log-linear model from a time series of seventeen observations (including a trend term on elasticities) show excellent fits in general. Twenty nine of his thirty seven commodity groups yield an R^2 better than 0.9 (12 of these greater than 0.99). Given this general good fit and lack of any demonstrably superior simple alternatives we feel it not unreasonable to assume that constant elasticity of demand is a useful generalisation.

The Assumption of Constant Marginal Cost in the Short-Run:

This assumption, and its implication that we may write the firm's marginal cost as equal to average variable cost, has been utilised in several places in the preceding work. More specifically, the later oligopoly models of chapter 2 suppose it; and the work of chapter 4 on the Cournot-type model of bilateral power both uses the assumption of constant marginal cost and is strengthened by it as we have noted there in our comments on Hicks (1935). Furthermore, when we move on to talk of a price-cost margin as a ratio of profits plus fixed costs to revenue then in every case that assumption is made. In fact it is probably true to say that without it, there could be very little rigorous work on relating industry structure to a performance measure. This arises because the normal basis for any such model is a maximising entrepreneur or manager so that marginal quantities naturally appear in the first order conditions, yet data availability is such that these cannot normally be identified. Given this situation, it is easiest if one moves to assuming the marginal quantity equal to an average quantity which is what we have done here. It then becomes of some importance to at least briefly assess the evidence for the assumption that short-run marginal cost is constant.

There are basically five areas where we can glean evidence on the shape of cost curves, some more useful for our purpose than others. Two of these, the Survivor Technique and Rate of Return studies, are rather wider in scope than the alternatives because factors additional to costs are included in the assessment. The survivor technique, as used by Stigler (1958) for example, involves looking for patterns in the distributions of (firm or plant) sizes over time. If say it is observed that the largest size category contains an increasing proportion of the observations as time proceeds, then we might conclude that there is evidence of scale economies. On the other hand (and this is of course its advantage for some purposes) it might be evidence of luck, advantages of vertical integration, outside acquisitions and so on. Of its nature, this technique is likely to be affected by the definition of the industry whose observations we are perusing. Also, unless the hypothesis regarding distributional changes is rather sophisticated, we are unable to say anything about the cost disadvantages of firms not at the "optimum" size. The other side of this coin is that we are not necessarily able to conclude that the absence of any discernable pattern in the size distributional changes means that there is an absence of significant scale economies.[6] For the present purpose evidence in this area is ambivalent and will not be considered further save to say that, in the view of Shepherd (1967), taking his own and earlier estimates into account, "there has been no widespread shift at all towards larger plants" and "there remains the inability of the survivor technique to provide normative estimates of scale economies above the plant level." (both p. 122).

The second of the broad approaches is the rate of return type of study like that of Hall and Weiss (1967). Here the exercise is to relate profitability to firm size. As will be obvious from our earlier work there are a large number of factors, besides scale economies, which should be reckoned as additional determinants of profitability.[7] For this reason, the results of Samuels and Smyth (1968), probably the most useful of U.K. studies in this area, should in the present context be treated with a good deal of caution. They found, briefly, that "Profit rates and size are inversely related" (p.139); if the relationship were entirely due to scale economies then this would indicate continually decreasing returns. Such a result conflicts directly with the work of Hall and Weiss for the U.S., but of course there are many other explanations of the divergence which we do not intend to pursue here. The evidence does not bear directly enough on the question of scale economies.

By contrast, a third approach to measuring scale economies, the "Engineering" study, probably has too narrow a compass for present purposes. The idea of such studies is to question those in the industry and ask what would be the cost (or cost advantages) of producing certain output levels based on available technical information. Thus: "Their accuracy is particularly suspect when dealing with some of the non-technical forces determining the effects of scale, for example when estimating the relationship between size and quality of management..." (Pratten 1971 p.20). In fact Pratten provides probably the best example of this type of approach in the U.K. After studying a large number of industries he considered that "there are large technical economies of scale for such ranges of products in many industries" (p.268),[8] and later he concludes that there are also economies on the marketing and managerial sides (see e.g. p. 302), though he has much less information in this area.

The focus of engineering studies then, is very much on the technical side and so at plant level. There are advantages in this, for demand considerations, relative prices and so on are held in abeyance. But as far as we are concerned, Pratten's work concentrates too much on what ought to be rather than what is. As such, the scale curves he derives are essentially planning curves rather than short run schedules holding many factors fixed. Quite a large amount of the decline in average total cost can probably be ascribed to the cost of the physical plant being spread, and may in no way be inconsistent with constant short-run marginal costs. We should also remember that the technical economies of scale he observes, while requiring plants operating up to a significant proportion of total industry output, are in most cases for quite narrowly defined "industries" or economies of producing a "narrow range of products" (p. 268). Our industries, being MLH level or broader, are much more aggregative. Thus we feel that his evidence does not contradict the assumption we wish to make about short-run marginal costs for the firm.

The fourth area in which we can obtain evidence on scale economies is from "Statistical Cost" studies, best exemplified by Johnston's (1960) oft-quoted work. This approach involves the use of empirical data on costs, usually from published sources. Besides estimating cost functions for six industries himself, Johnston also provides a fairly comprehensive review of previous studies in the same area. He finds that "Two major impressions ... stand out clearly. The first is that the various short-run studies more often than not indicate constant marginal cost and declining average cost as the pattern that best seems to describe the data that have been analysed. The second is the preponderance of the L-shaped pattern of long-run average cost that emerges so frequently from the various long-run analyses" (p.168). The conclusion on long-run functions is of course compatible with Pratten's results. However, our main concern is with short-run cost functions which are generally derived from time-series data.

Now there have been many criticisms of the methods and consequences of statistical estimation of cost curves and in a further chapter Johnston comments on the main ones. He considers six relating to the short-run function, the main thrust of these being that statistical estimation renders the rejection of a curvilinear form for the total cost function more likely than in "truth" should be the case. To a greater or lesser extent he is able to qualify or reject all these, so that he considers that"MC may rise at extremely high output rates; but over substantial ranges of output, in cases where divisibility or segmentation of capital equipment is possible, it is probably constant." (p.192). In fairness though, it should perhaps be pointed out that, after his survey of the empirical results and arguments, Walters (1963) reports that "the evidence in favour of constant marginal cost (in the short run) is not overwhelming" (p.51). This does not mean that the assumption is not a reasonable approximation, of course.[9]

Finally, we should note that a further possible source of information on scale economies is the large number of econometric studies which involve estimating production functions. Obviously to move towards cost functions from these we would have to make a number of (perhaps questionable) assumptions about factor prices and so on, not all compatible with the oligopolistic basis of the models developed here. Thus we consider that the connection is not strong enough to allow us to say anything very reliable about short-run cost functions from estimations of production functions.

Given the available relevant evidence, we consider that our assumption of constant short-run marginal cost is a fairly useful approximation over wide ranges of output. It remains true that if output varies greatly the average cost curve may turn upward at its extremes, but we can do little about this in empirical work apart from noting that our proxy for elasticity of demand changes (on which see Chapter 5) may pick up such effects to some extent.[10]

FOOTNOTES

1 Estimation has generally proceeded using the form:

$$\log q_i = \alpha_i + \eta_{yi} \log(y) + \sum_j \eta_{ij} \log p_j,$$

or assuming homogeneity of degree zero in income and all prices:

$$\log q_i = \alpha_i + \eta_{yi} \log (\frac{y}{\bar{p}}) + \sum_j \eta_{ij} \log (p_j/\bar{p}),$$

where y is money income (or expenditure), p_j are the prices of the n goods in the system and q_i the quantity of the ith good purchased, $\bar{p}$ is a general price index. Often many or all of the prices other than the ith are omitted in estimation (as in Deaton (1975a) where only p_i is included). In this notation, the elasticity we are concerned with is η_{ii}

2 Any possibility of "Giffen" goods is naturally very serious for our, or almost any, structure-performance study. We assume they do not exist.

3 That is where the demand for the ith good is of the form

$$d \log q_i = \eta_{yi} \, d\log y + \sum_j \eta_{ij} d \log p_j$$

(where certain relationships hold between the elasticities within this equation and between it and the others in the system).

4 The agg- ation property of the system was imposed by the data. Symmetry and negativity are properties derived from utility theory, the latter being the generalised substitution effect and the former the equivalence of compensated cross-partials. Aggregation and symmetry are not properties of the constant elasticity model (see e.g. Stone 1954 p.278).

5 This was presumably written prior to his last-mentioned (1975a) paper.

6 See also Shepherd (1967) for further critique of the method.

7 Pratten (1971) pp 348-9 has some discussion on this point. Hall and Weiss do include some of these factors.

8 The idea of the "range of products" is defined on the same page in Pratten.

9 We ought to mention a further point on these conclusions. Some of the empirical studies referred to by Johnston and Walters were performed at the plant level while others are at the level of the firm; we are really concerned with the latter. But management can presumably duplicate plants so that constant marginal cost is retained for the firm unless there are substantial economies or diseconomies in running several plants under one umbrella. If we look at the U.K. statistics however, we find that the number of establishments per enterprise in manufacturing is only 1.33 in 1968 (Census of Production figures), which indicates that this possibility should not worry us much empirically.

10 Of course if we believe Spence (1974), whose model was discussed in chapter 3, then we would expect excess capacity to be maintained, making it rather unlikely that the average cost curve would turn upward at high output levels.

APPENDIX II

The Data

Some Introductory Comments

The predominant source of data for studies such as ours is the series of Censuses of production for the UK. Detailed censuses have recently been taken at roughly five-yearly intervals, the last produced at time of writing being for 1968. These series cover all of manufacturing plus part or the whole of mining, construction and the public utilities. The census material is tabulated in most cases at Minimum List Heading (MLH) Level, or in a more detailed form than that. This derives from the Standard Industrial Classification system for ordering industries, mainly by productive activity. The definitions change slightly from time to time, but in the 1968 census the 1963 figures have been reclassified to the same basis and no major revisions have taken place since then. These yield enough data to derive approximations to the main structure and performance measures normally used.

In addition, small censuses are taken nearly every year, the latest published data being an (as yet) incomplete series for 1973. Between 1964 and 1967 the small census was a very minor affair, containing only two main tables and these at a more aggregate level than the MLH categories. Those tables were concerned with stocks and work in progress and fixed capital expenditure only. However from 1970 onwards, much more detail is included. One of the main differences between the small and large censuses is that the former are based not at all on the "enterprise" (which is a census approximation to a firm), but rather on the "establishment" or plant. As such they are not useful for information on many industrial structural features. Post 1970 though, they are sufficient to obtain industry aggregates such as our performance measure. The upshot of this arrangement is that annual series of the data we require are not available long enough even for averaging purposes; we have to take "snapshot" views of the industrial world.

As we said in chapter 1, since we decide to use a "ratio" formulation for our model we need two years from which we can glean more-or-less compatible data on the structural features used. Our bilateral power measure (on which see chapter 4) requires knowledge of interindustry relationships so we choose 1963 and 1968 as our base years, being years for which UK input-output tables have been produced. These luckily have as their basis an aggregated MLH classification system which becomes our sample level. If we decide to have structure affect performance with a lag, this is possible using performance figures from later small censuses.

Sources

We list below for convenient reference main details on the sources of the data used in empirical estimations. In each case figures are for the UK. Much of the data comes fairly directly from published Government tables so that it seems wasteful to provide all the figures used in estimation in addition. However we end the appendix by tabulating data on structural features which did involve some substantial burden in collection. Full references for all statistical sources are given below rather than in the main bibliography.

The profit - revenue ratio or price - cost margin:

This is defined as total net output minus total wages and salaries all divided by total sales and work done (including sales of merchanted goods). Source: 1968 Census, Part 156, Table 1 (for 1963 and 1968 figures); 1973 Census, Table 1 (for 1972 figures).

Capital expenditure:

This is subtracted from the numerator of the above ratio in some estimations, see chapter 6 for details. It is defined as Total (net) capital expenditure. Source: 1968 Census, Part 156, Table 2 (for 1963 and 1968 figures); 1973 Census, Table 2 (for 1972 figures).

The Herfindahl index:

The construction of this variable is described in full in chapter 6, the figures are tabulated below. The basic source is: 1968 Census, Part 158, Table 42A which gives the distribution of enterprises by employment for each industry.

The Bilateral Power index:

This is constructed partly from the Herfindahl index, partly using the Input-Output tables and partly from some miscellaneous sources as noted and described in chapter 6. The Input-Output source is: Table D: Industry by Industry flow matrix 1968, ditto 1963.

The Unemployment variable:

This is defined as unemployed divided by unemployed plus employed. Source: Annual Abstract of Statistics, various years; Tables : Estimated number of employees in employment and numbers unemployed.

The split between durable and non-durable goods is based on the figures in the column "Gross Domestic Capital Formation: Fixed" in table D of the Input-Output tables.

The choice of which industries to exclude based on "specialisation" and "exclusiveness" ratios was made using the table "Analysis of output of principal products", Table S in the 1968 Input-Output tables, table L in those for 1963.

Full references:

Census of Production:

Department of Trade and Industry, Business Statistics Office: "Report on the Census of Production 1968." London, HMSO, 1972.

Department of Industry, Business Statics Office (Business Monitor PA 1000) : "Report on the Census of Production,1973 Provisional Results." London, HMSO, 1974.

Input-Ouput Tables:

Central Statistical Office: "Input-Output Tables for the UK 1963 (Studies in Official Statistics No. 16)." London, HMSO, 1970.

Central Statistical Office: "Input-Output Tables for the UK 1968 (Studies in Official Statistics No. 22)." London, HMSO, 1973.

Annual Abstract of Statistics:

Central Statistical Office: "Annual Abstract of Statistics (various years)." London, HMSO.

Census of Distribution:

Board of Trade: "Report on the Census of Distribution and Other Services, 1961." London, HMSO, 1963. (Table 7)

Board of Trade: "Ditto, 1966." London, HMSO, 1970. (Table 6).

Airline Annual reports:

Ministry of Aviation: "British European Airways Corporation. Annual Report and Accounts 1963 - 4." HC 348.

Ministry of Aviation: "British Overseas Airways Corporation.Ditto" HC 349.

Board of Trade: "British European Airways Corporation. Annual Report and Accounts for the year ended March 31, 1969" HCP 379.

Board of Trade: "British Overseas Airways Corporation.Ditto" HCP 378.

European Economic Communities: "Yearbook of Agricultural Statistics." Brussels, Statistical Office of the European Communities, 1967.

Table A1 - Selected Data Used

MLH (1968) Nos.	Name	Herfindahl Index		Bilateral Power Index[1]	
		1963	1968	1963	1968
102-4, 109	Other mining	0.00963	0.01120	1.10308	1.55155
211	Grain milling	0.05969	0.07366	1.02005	1.02304[2]
212-3, 219	Other cereals	0.04700	0.05796	1.00117	1.00104
216	Sugar	0.07692	0.09867	1.01265	1.01844
217	Cocoa,chocolate	0.05215	0.06001	1.00337	1.00373
214-5, 218,229	Other food	0.01161	0.00869	1.01141	1.01444
221,275	Soap,oils & fats	0.03369	0.03093	1.02542	1.03034
231,232,239	Drink	0.02553	0.03250	1.00053	1.00126
240	Tobacco	0.09817	0.09903	1.00000	1.00024
262-3	Mineral oil refining	0.10944	0.09103	1.73779	1.95168
274	Paint	0.02966	0.03077	1.12002	1.05607
261	Coke ovens	0.04937	0.10736	1.18393	1.49443
272-3	Pharmaceutical & toilet	0.02320	0.02203	1.05256	1.04623
276	Synthetic resins	0.06849	0.05910	1.03516	1.03718
271,277-9	Other chem. & allied	0.01427	0.01160	1.10953	1.11467
311-3	Iron & steel	0.01432	0.02868	1.43795	1.30818
321	Aluminium	0.05754	0.04324	1.03840	1.04687
322-3	Other nonferrous	0.03343	0.04350	1.06134	1.05264
331	Ag. machinery	0.03547	0.04119	1.03747	1.02525[2]
332	Machine tools	0.01882	0.02532	1.11601	1.10643[2]
390	Engineers small tools	0.00831	0.00925	1.63847	2.02065
334	Indust.engines	0.09110	0.17706	1.16048	1.11001[2]
335	Textile machinery	0.04301	0.03777	1.03197	1.03078
336-7	Contractors plant	0.01134	0.01270	1.08783	1.26699
338	Office machinery	0.08989	0.07389	1.13172	1.14406[2]

MLH (1968) Nos.	Name	Herfindahl Index		Bilateral Power Index[1]	
		1963	1968	1963	1968
333,339	Other non-electrical	0.00475	0.00451	1.43484	1.17970
341	Industrial plant	0.01342	0.01110	1.26220	1.30682
342,349	Other mech.eng.	0.01267	0.01823	1.30134	1.30427[2]
351-4	Instruments	0.00984	0.01311	1.09468	1.11478
361	Electrical machinery	0.04689	0.07652	1.83706	1.79560
362	Insulated wires	0.05826	0.07282	2.76038	3.47236[2]
363-7	Radio & telecomm.	0.02054	0.02063	1.39367	1.27749
368-9	Other electrical	0.02124	0.02588	1.25991	1.30587
395	Cans & boxes	0.07893	0.09252	1.02345	1.03727
391-4,396,399	Other metal gds.	0.00186	0.00215	1.12013	1.25468
370	Shipbuilding	0.03744	0.05159	1.10818	1.07592
380-1	Motor vehicles	0.05694	0.08605	1.04910	1.03373
383	Aircraft	0.10901	0.12178	1.14135	1.09044
382,384-5	Other vehicles	0.03024	0.04182	1.19331	1.05514
411	Man made fibres	0.14286	0.20000	1.01157	1.01084
412-3	Spinning & weaving	0.00853	0.01752	1.05252	1.02362
414	Wool	0.00864	0.01097	1.00561	1.00328
415-9,421-2,429	Other textiles	0.00411	0.00679	1.14283	1.00884
423	Textile finishing	0.01991	0.03355	1.00576	1.02394
431-3	Leather	0.00287	0.00349	1.00620	1.00558
441-6,449	Clothing	0.00241	0.00280	1.09225	1.08299
450	Footwear	0.01193	0.01551	1.01217	1.01049

MLH (1968) Nos.	Name	Herfindahl Index		Bilateral Power Index[1]	
		1963	1968	1963	1968
464	Cement	0.18920	0.25084	1.02527	1.12098
461,469	Other building mats.	0.00595	0.00815	1.29118	1.13298
462-3	Pottery & glass	0.01926	0.02729	1.14684	1.02161
472-3	Furniture etc.	0.00349	0.00509	1.03326	1.02211
471,474-5,479	Wood	0.00139	0.00197	1.38501	1.20321
481	Paper & board	0.04227	0.04510	1.10366	1.02448
482-3,484	Paper products	0.00959	0.00962	1.09185	1.11680
485-6,489	Printing & publishing	0.00771	0.00869	1.14261	1.11006
491	Rubber	0.04480	0.05384	1.22591	1.16560
492-6,499	Other	0.00409	0.00467	1.31252	1.03997
500	Construction	0.00059	0.00105	1.31517	1.52540[3]

Notes: 1. This is the bilateral power index most commonly used in estimation, where the elasticity of demand takes on the value unity everywhere (see chapter 6 for details).

2. These figures are provided purely for information, the observations were not used because of insufficient homogeneity in the data (see chapter 6 for details).

3. This industry was sometimes excluded from estimation because of lack of information for years later than 1968 (see chapter 6 for details).

BIBLIOGRAPHY

This bibliography contains full references for all works listed in the text by author and date. In addition, many other works consulted while writing the thesis, but not mentioned explicitly in the text, are included below.

In order to avoid unnecessary duplication, the following abbreviations have been used in the bibliography:

A.E.R.	=	American Economic Review.
Eca	=	Economica.
E.J.	=	Economic Journal. (also used as a combination, eg. Southern E.J. = Southern Economic Journal).
Etrica	=	Econometrica.
J.I.E.	=	Journal of Industrial Economics.
J.L.E.	=	Journal of Law and Economics.
J.P.E.	=	Journal of Political Economy.
O.E.P.	=	Oxford Economic Papers.
Q.J.E.	=	Quarterly Journal of Economics.
R.E. Stats.	=	Review of Economics and Statistics.
R.E. Studs.	=	Review of Economic Studies.
U.P.	=	University Press. (eg. Cambridge U.P. = Cambridge University Press).
W.E.R.P.	=	University of Warwick : Warwick Economic Research Papers.

M.A. Adelman : "A and P, a study in price-cost behaviour and public policy." Cambridge Mass., Harvard U.P., 1966.

M.A. Adelman : "The measurement of industrial concentration." R.E. Stats., vol. 33, 1951, 269-296.

P.W.S. Andrews : "Manufacturing business." London, Macmillan, 1949.

G.C. Archibald : "'Large' and 'small' numbers in the theory of the firm." Manchester School of Economic and Social Studies, vol. 27, 1959, 104-109.

H. Averch and L.L. Johnson : "Behaviour of the firm under regulatory constraint." A.E.R., vol. 52, 1962, 1052-1069.

J.S. Bain : "Barriers to new competition." Cambridge Mass., Harvard U.P., 1962.

J.S. Bain : "The comparative stability of market structure" in J.W. Markham and G.F. Papanek (1970) (q.v.).

J.S. Bain : "Conditions of entry and the emergence of monopoly" in E.H. Chamberlin (ed.): "Monopoly and competition and their regulation." London, Macmillan, 1954.

J.S. Bain : "Industrial organisation." New York, John Wiley, 1959.

J.S. Bain : "Relation of profit rate to industry concentration : American manufacturing 1936-40." Q.J.E., vol. 65, 1951, 293-324.

J.S. Bain : "Workable competition in oligopoly - Theoretical considerations and some empirical evidence." A.E.R., vol 40, 1950, 35-47.

D.P. Baron : "Limit pricing and models of potential entry." Western E.J., vol. 10, 1972, 298-307.

A.P. Barten : "Evidence on the Slutsky conditions for demand equations." R.E. Stats., vol. 49, 1967, 77-84.

W.J. Baumol : "Business behaviour, value and growth." New York, Macmillan, 1959.

G.S. Becker : "Economic theory." New York, Alfred A. Knopf, 1971.

H. Benishay : "Concentration and price-cost margins : A comment." J.I.E., vol. 16, 1967, 73-74.

J.N. Bhagwati : "Oligopoly theory, entry-prevention and growth." O.E.P., vol. 22, 1970, 297-310.

R.L. Bishop : "Elasticities, cross elasticities and market relationships." A.E.R., vol. 42, 1952, 779-803.

R.L. Bishop : "Game theoretic analyses of bargaining." Q.J.E., vol. 77, 1963, 559-602.

R.L. Bishop : "A Zeuthen-Hicks theory of bargaining." Etrica, vol. 32, 1964, 410-417.

G. Bowman _et. al._ : "Collusion in oligopoly : An experiment on the effect of numbers and information." Q.J.E., vol. 82, 1968, 240-259.

J. Cable : "Profit technology and advertising : A note on the specification of advertising-profitability models." W.E.R.P. no. 59, 1974.

R.E. Caves _et. al._ : "Scale economies in statistical analyses of market power." R.E. Stats., vol. 57, 1975, 133-140.

E.H. Chamberlin : "The Theory of Monopolistic Competiton." Cambridge Mass., Harvard U.P., 1933.

C.F. Christ _et. al._ "Measurement in economics : Studies in mathematical economics and econometrics in memory of Yehuda Grunfeld." Stanford U.P., 1963.

J.R. Clarke : Review of L.G. Telser (1972). (q.v.). E.J., vol 83, 1973, 250-251.

K.J. Cohen and R.M. Cyert : "Theory of the firm : Resource allocation in a market economy." Englewood Cliffs N.J., Prentice Hall, 1965.

N.R. Collins and L.E. Preston : "Concentration and price-cost margins in food manufacturing industries." J.I.E., vol. 14, 1966, 226-242.

N.R. Collins and L.E. Preston : "Concentration and price-cost margins in manufacturing industries." Berkeley, University of California press, 1968.

N.R. Collins and L.E. Preston : "Price-cost margins and industry structure." R.E. Stats., vol. 51, 1959, 271-286.

W.S. Comanor and T.A. Wilson : "Advertising, market structure and performance." R.E. Stats., vol. 49, 1967, 423-440.

A.A. Cournot : "Researches into the mathematical principles of the theory of wealth." New York, Macmillan, 1927 (reprint).

K. Cowling (ed.) : "Market structure and corporate behaviour." London, Gray-Mills, 1972.

K. Cowling : "Oligopoly and the distribution of income." Coventry, University of Warwick, August 1975 (mimeo).

K. Cowling and J. Cubbin : "Price formation in the U.K. motor industry : An analysis of oligopolistic behaviour." Coventry, University of Warwick, May 1970, (mimeo.)

K. Cowling and M. Kelly : "Advertising and price-cost margins." A Chapter of K. Cowling et. al. : "Advertising and economic behaviour." London, Macmillan, 1975.

K. Cowling and M. Waterson : "Price-cost margins and market structures." Eca, vol. 43, 1976, 267-274.

J. Cubbin : "A measure of apparent collusion in oligopoly." W.E.R.P. no. 49, 1974.

J. Cubbin : "A short selective survey of oligopoly theory." Coventry, University of Warwick, 1973. (mimeo.).

R.M. Cyert and M.H. DeGroot : "An analysis of cooperation and learning in a duopoly context." A.E.R., vol. 63, 1973, 24-37.

A.S. Deaton : "The analysis of consumer demand in the United Kingdom 1900-1970." Etrica, vol. 42, 1974, 341-368.

A.S. Deaton : (1975a) : "The measurement of income and price elasticities." European Economic Review, vol. 6, 1975, 261-273.

A.S. Deaton : (1975b) : "Models and projections of demand in post-war Britain." London, Chapman and Hall, 1975.

H. Demsetz : "Industry structure, market rivalry and public policy." J.L.E. vol. 16, 1973, 1-9.

P.J. Devine et. al. : "An introduction to industrial economics." London, Allen and Unwin, 1974.

F.T. Dolbear et. al. : "Collusion in oligopoly : An experiment on the effect of numbers and information." Q.J.E., vol. 82, 1968, 240-259.

R. Dorfman and P.O. Steiner : "Optimal advertising and optimal quality." A.E.R., vol. 44, 1954, 826-836.

P.A. Dutton : "The impact of government competition policy, imports and tariffs on price-cost margins for British manufacturing industries." Coventry, University of Warwick, Centre for Industrial, Economic and Business Research 65, 1976.

H.R. Edwards : "Price formation in manufacturing industry and excess capacity." O.E.P., vol. 7, 1955, 94-118.

C.W. Efroymson : "A note on kinked demand curves." A.E.R., vol. 33, 1943, 98-109.

A.S. Eichner : "A theory of the determination of the mark-up under oligopoly." E.J., vol. 83, 1973, 1184-1200.

R. Evely and I.M.D. Little : "Concentration in British industry." Cambridge, Cambridge U.P., 1960.

E.F. Fama and A.B. Laffer : "The number of firms and competition." A.E.R., vol. 62, 1972, 670-674.

W. Fellner : "Competition among the few." New York, A.A. Knopf, 1949.

C.E. Ferguson : "Microeconomic theory." (Revised ed.), Homewood Ill., R.D. Irwin, 1970.

C.E. Ferguson : "The neoclassical theory of production and distribution." Cambridge, Cambridge U.P., 1975 (reprint).

F.M. Fisher : "New developments on the oligopoly front : Cournot and the Bain - Sylos analysis." J.P.E., vol. 67, 1959, 410-413.

L. Foldes : "A determinate model of bilateral monopoly." Eca, vol. 31, 1964, 117-131.

L.E. Fowraker and S. Siegel : "Bargaining behaviour." New York, McGraw-Hill, 1963.

C.R. Frank : "Entry in a Cournot market." R.E. Studs., vol. 32, 1965, 245-250.

J.W. Friedman : "On experimental research in oligopoly." R.E. Studs, vol. 36, 1969, 399-415.

J.W. Friedman : "Individual behaviour in oligopolistic markets : An experimental study." Yale Economic Essays, vol. 3, 1963, 359-417.

J.W. Friedman : "A non-cooperative equilibrium for supergames." R.E. Studs., vol. 38, 1971, 1-12.

J.W. Friedman : "Reaction functions and the theory of duopoly." R.E. Studs., vol. 35, 1968, 257-272.

M. Friedman : "Price theory : A provisional text." (Revised ed.) Chicago, Aldine, 1970 (reprint).

J.K. Galbraith : "Monopoly power and price rigidities." Q.J.E., vol. 50, 1936, 456-475.

D.W. Gaskins : "Dynamic limit pricing : Optimal pricing under threat of entry." Journal of Economic Theory, vol. 3, 1971, 306-322.

H. Glesjer : "A new test for heteroskedasticity." Journal of the American Statistical Association, vol. 64, 1969, 316-323.

S.M. Goldfeld and R.E. Quandt : "Some tests for homoskedasticity." Journal of the American Statistical Association, vol. 60, 1965, 539-547.

I.M. Grossack : "The concept and measurement of permanent industrial concentration." J.P.E., vol. 81. 1973, 745-760.

L.A. Guth *et. al.* : "Buyer concentration ratios." New York University Graduate School of Business Administration, Working paper 73-48, 1973.

J. Hadar : "Stability of oligopoly with product differentiation." R.E. Studs., vol. 33, 1966, 57-60.

M. Hall and L. Weiss : "Firm size and profitability." R.E. Stats., vol. 49, 1967, 319-331.

G.C. Harcourt : "The accountant in a golden age." O.E.P., vol. 17, 1965, 66-80.

M.N. Harris : "Entry and barriers to entry." Federal Reserve Bank of New York, research paper 7330, 1973.

R.F. Harrod : "Imperfect competition and the trade cycle." R.E. Stats., vol. 18, 1936, 84-88.

R.F. Harrod : "Increasing returns." in R.E. Kuenne (ed.) "Monopolistic competition theory : Studies in impact." New York, Wiley & Sons, 1967.

J.C. Harsanyi : "Approaches to the bargaining problem before and after the theory of games : a critical discussion of Zeuthen's, Hicks', and Nash's theories. Etrica, vol. 24, 1956, 144-157.

P.E. Hart : "Competition and rate of return on capital in U.K. industry." Business Ratios,vol. 2, 1968, 3-11.

P.E. Hart : "Entropy and other measures of concentration." Journal of the Royal Statistical Society, series A, vol. 134, 1971, 73-85.

P.E. Hart and E. Morgan : "Market structure and performance in the U.K." University of Reading, discussion paper in economics series A, no. 69, 1975. (Wrongly credited to Hart alone in text).

P.E. Hart and S.J. Prais : "The analysis of business concentration : A statistical approach." University of Reading, discussion paper in economics series A, vol. 119, 1956, 150-181.

J.M. Henderson and R.E. Quandt : "Microeconomic theory : A mathematical approach." Tokyo, McGraw-Hill/Kogakusha, 1958.

J.R. Hicks : "Annual survey of economic theory : theory of monopoly."

Etrica, vol. 3, 1935, 1-20.

J.R. Hicks : "The process of imperfect competition." O.E.P. vol. 6, 1954, 41-54.

J.R. Hicks : "Theory of Wages." (2nd ed.) London, Macmillan, 1964.

A.C. Hoggatt : "Measuring behaviour in quantity variation oligopoly games." Behavioural Science, vol. 12, 1967, 109-121.

S.E. Holterman : "Market structure and economic performance in U.K. manufacturing industry." J.I.E., vol. 22, 1973, 199-140.

N.J. Ireland (1972a) : "Concentration and the growth of market demand : A comment on the Gaskins limit pricing model." Journal of Economic Theory, vol. 5, 1972, 303-305.

N.J. Ireland (1972b) : Harrod, profit maximisation and new entry." W.E.R.P. no. 23, 1972.

A.P. Jacquemin and J. Thisse : "Strategy of the firm and market structure" in K. Cowling (ed.). (1972) (q.v.).

F. Jenny : "Wage rates, wage earnings, concentration and unionisation in French manufacturing industries." (Mimeo.). 1974

J. Johnston (1972a) : "Econometric methods." (2nd ed.), London, McGraw-Hill, 1972.

J. Johnston (1972b) : "A model of wage determination under bilateral monopoly." E.J., vol. 82, 1972, 837-852.

J. Johnston : Statistical cost analysis." New York, McGraw-Hill, 1960.

P.L. Joskow : "Firm decision making processes and oligopoly theory." A.E.R., vol. 65, 1975 (May), 270-279.

M. Kalecki : "The determination of distribution of the national income." Etrica, vol. 6, 1938, 97-112.

M.I. Kamien and N.L. Schwartz : "Cournot oligopoly with uncertain entry." R.E. Studs., vol. 42, 1975, 125-131.

M.I. Kamien and N.L. Schwartz : "Limit pricing and uncertain entry." Etrica, vol. 39, 1971, 441-454.

J. Khalilzadeh - Shirazi : "Market structure and price-cost margins in U.K. manufacturing industries." R.E. Stats, vol. 56, 1974, 67-75.

J. Khalilzadeh - Shirazi : An earlier version of the above is : Harvard, 1973 (mimeo).

S.H. Lustgarten : "The impact of buyer concentration in manufacturing industries." R.E. Stats, vol. 57, 1975, 125-132.

F. Machlup and M. Taber : "Bilateral monopoly, successive monopoly and vertical integration." Eca, vol. 27, 1960, 101-119.

H.M. Mann : "Seller concentration, barriers to entry and rates of return in thirty industries 1950-1960." R.E. Stats., vol. 48, 1966, 296-307.

J.W. Markham and G.F. Papanek : "Industrial organisation and economic development, in honour of E.S. Mason.", Boston, Houghton Mifflin, 1970.

S.C. Maurice and C.E. Ferguson : "Factor demand elasticity under monopoly and monopsony." Eca., vol. 40, 1973, 180-186.

R.I. McKinnon : "Stigler's Theory of Oligopoly : A comment." J.P.E. vol. 74, 1966, 281-285.

G. de Menil : "Bargaining : Monopoly power versus union power." Cambridge Mass., M.I.T. press, 1971.

J. Mincer : "Market prices, opportunity costs and income effects" in C.F. Christ et. al., 1963, (q.v.).

F. Modigliani : "New developments on the oligopoly front." J.P.E. vol. 66, 1958, 215-232.

R.S. Moreland : "Managerial discretion, property rights and the theory of firms." W.E.R.P. no. 22, 1972.

J.N. Morgan : "Bilateral monopoly and competitive output." Q.J.E., vol. 63, 1949, 371-391.

O. Morgernstern : "Demand theory reconsidered." Q.J.E., vol. 62, 1948, 165-201.

B. Morris : "Multi-dimensional aspects of market structure and market performance." W.E.R.P. no. 12, 1970.

J.L. Murphy : "Effect of the threat of losses on duopoly bargaining." Q.J.E. vol. 80, 1966, 296-313.

J. Nash : "The bargaining problem." Etrica, vol. 18, 1950, 155-162.

D. Needham : "Economic analysis and industrial structure." London, Holt, Reinart and Winston, 1970.

M. Nerlove and K. Arrow : "Optimal advertising policy under dynamic conditions." Eca., vol. 29, 1962, 129-142.

M. Nicholson : "Oligopoly and conflict." Liverpool, Liverpool U.P., 1972.

W.D. Nordhaus and W. Godley : "Pricing in the trade cycle." E.J., vol. 82, 1972, 853-882.

S.I. Ornstein : "Empirical uses of the price-cost margin." J.I.E. vol. 24, 1975, 105-117.

D. Orr : "An index of entry barriers and its application to the structure-performance relationship." J.I.E., vol. 23, 1974, 39-49.

D.K. Osborne : "On the rationality of limit pricing." J.I.E., vol. 22, 1973, 71-80.

D.K. Osborne : "The role of entry in oligopoly theory." J.P.E., vol. 72, 1964, 396-402.

B.P. Pashigan : "Limit price and the market share of the leading firm." J.I.E., vol. 16, 1968, 165-177.

D.A. Peel : "The non-uniqueness of the Dorfman-Steiner conditions." Eca, vol. 40, 1973, 208-209.

M. Peston and B. Corry (eds.) : "Essays in honour of Lord Robbins." London, Weidenfeld and Nicolson, 1972.

A. Phillips : "An econometric study of price-fixing, market structure and performance in British industry in the early 1950's" in K. Cowling (ed.). 1972. (q.v.).

A. Phillips : "Structure conduct and performance - and performance conduct and structure?" in Markham and Papanek (eds.) 1970. (q.v.).

A. Phillips : "A theory of interfirm organisation." Q.J.E., vol. 74, 1960, 602-613.

L. Phlips : "Effects of industrial concentration : A cross - sectioned analysis for the Common Market." Amsterdam, North Holland, 1971.

C.F. Pratten : "Economies of scale in manufacturing industry." Cambridge, Cambridge U.P., 1971. (Department of applied economics occasional paper no. 28).

G. Pyatt : "Profit maximisation and new entry." E.J., vol. 81, 1971, 246-255.

D. Qualls : "Concentration, barriers to entry and long-run economic profit margins." J.I.E. vol. 20, 1972, 142-158.

R.E. Quandt : "On the stability of price adjusting oligopoly." Southern Economic Journal, vol. 33, 1967, 332-336.

R.E. Quandt and M. McManus : "Comments on the stability of the Cournot oligopoly model." R.E. Studs., vol. 28, 1961, 136-139.

T. Rader : "Theory of microeconomics." New York, Academic Press, 1972.

Raiffa (1957) in : R.D. Luce and H. Raiffa : "Games and decisions." New York, Wiley, 1957.

R. Reichardt : "Competition through the introduction of new products." Zeitschrift für Nationalöconomie, 1962, 41-84.

G.B. Richardson : "The organisation of industry." E.J., vol. 82, 1972, 883-896.

J. Robinson : "The economics of imperfect competition." London, Macmillan, 1933.

R.J. Ruffin : "Cournot oligopoly and competitive behaviour." R.E. Studs., vol. 38, 1971, 493-502.

J.M. Samuels and D.J. Smyth : "Profits, variability of profits and firm size." Eca, vol. 35, 1968, 127-139.

E. Saradayar : "Zeuthen's theory of bargaining, a note." Etrica, vol. 33, 1965, 802-813.

P.K. Sawhney and B.L. Sawhney : "Capacity - utilisation, concentration and price-cost margins in Indian industries." J.I.E., vol. 21, 1973, 145-153.

M.C. Sawyer : "Concentration in British manufacturing industries." O.E.P., vol. 23, 1971, 352-383.

F.M. Scherer : "Industrial market structure and economic performance." Chicago, Rand McNally, 1970.

M.B. Schupack: "Dynamic Limit pricing with advertising." Brown University, department of economics, (mimeo) 1972.

L. Shapley and M. Shubik : "Concepts and theories of pure competition" in M. Shubik (ed.) : "Essays in mathematical economics." Princeton, Princeton U.P., 1967.

L. Shapley and M. Shubik : "Price strategy oligopoly with product variation." Kyklos, vol. 22, 1969, 30-44.

R.W. Shaw : "Price leadership and the effects of new entry on the U.K. petrol supply market." J.I.E., vol. 23, 1974, 65-79.

W.G. Shepherd : "Structure and behaviour in British industries with U.S. comparisons." J.I.E., vol. 21, 1972, 35-54.

W.G. Shepherd : "Trends of concentration in American manufacturing industries." R.E. Stats, vol. 46, 1964, 200-212.

W.G. Shepherd : "What does the survivor technique show about economies of scale?" Southern E.J., vol. 34, 1967, 113-122.

R. Sherman : "The economics of industry." Boston Mass., Little, Brown, 1974.

R. Sherman : "An experiment on the persistence of price collusion." Southern E.J., vol. 37, 1970, 489-495.

R. Sherman : "Experimental oligopoly." Kyklos, vol. 24, 1971, 30-49.

R. Sherman : "Risk attitude and cost variability in a capacity choice experiment." R.E. Studs., vol. 36, 1969, 453-466.

R. Sherman and T.D. Willett : " Potential entrants discourage entry." J.P.E., vol. 75, 1967, 400-403.

M. Shubik : "Strategy and market structure." New York, John Wiley, 1959.

I.H. Silberman : "On lognormality as a summary measure of concentration." A.E.R., vol. 57, 1967, 807-831.

A. Silberston : "Surveys of applied economics : Price behaviour of firms." E.J., vol. 80, 1970, 511-582.

A. Singh : "Take-overs: Their relevance to the stock market and the theory of the firm." Cambridge, Cambridge U.P., 1971.

D.J. Smyth et. al. : "The measurement of firm size : Theory and evidence for the U.S. and the U.K." R.E. Stats., vol. 57, 1975, 111-114.

H. Sonnenschein : "The dual of duopoly is complementary monopoly : or, two of Cournot's theories are one." J.P.E., vol. 76, 1968, 316-318.

M. Spence : "Entry, capacity, investment and oligopolistic pricing." Stanford, Stanford University Institute for Mathematical Studies in the Social Sciences : Economics series, Technical report 131. 1974.

M. Spence : "Product selection, fixed costs and monopolistic competition." R.E. Studs., vol. 43, 1976, 217-235.

Z.A. Spindler : "A simple determinate solution for bilateral monopoly." Journal of Economic Studies, vol. 9, 1974, 55-64.

G.J. Stigler : "Administered prices and oligopolistic inflation." Journal of Business, vol. 35, 1962, 1-13.

G.J. Stigler : "Capital and rates of return in manufacturing industries." Princeton, Princeton U.P., 1963.

G.J. Stigler : "The economies of scale." J.L.E., vol 1, 1958, 54-71.

G.J. Stigler : "The organisation of industry." Homewood Ill., R.D. Irwin, 1968.

G.J. Stigler : "A theory of oligopoly." J.P.E., vol. 72, 1964, 44-61.

G.J. Stigler : "The theory of price." (3rd ed.) New York, Macmillan, 1966.

R. Stone et. al. : "The measurement of consumers' expenditure and behaviour in the United Kingdom 1920-1938 vol. 1." Cambridge, Cambridge U.P., 1954.

P.M. Sweezy : "Demand under conditions of oligopoly." J.P.E., vol. 47, 1939, 568-573.

P. Sylos-Labini : "Oligopoly and technical progress." Cambridge Mass., Harvard U.P., 1962.

L. Telser : "Competition, collusion and game theory." London, Macmillan, 1972.

R. Triffin : "Monopolistic competition and general equilibrium theory." Cambridge Mass., Harvard U.P., 1949.

J.M. Vernon and D.A. Graham : "Profitability of monopolisation by vertical integration." J.P.E., vol. 79, 1971, 924-925.

J. Von Neumann and O. Morgernstern : "Theory of games and economic behaviour." (3rd ed.) Princeton, Princeton U.P., 1953.

A.A. Walters : "Production and cost functions : An econometric survey." Etrica, vol. 31, 1963, 1-66.

S. Weintraub : "Intermediate price theory." Philadelphia, Chilton, 1964.

L. Weiss : "Quantitative studies of industrial organisation." in : M.D. Intriligator (ed.). "Frontiers of quantitative economics." Amsterdam, North Holland, 1971.

M.L. Weitzman : "Prices vs. quantities." R.E. Studs., vol. 41, 1974, 477-491.

J.T. Wenders : "Entry and monopoly pricing." J.P.E., vol. 75, 1967, 755-760.

M.R. Wickens : A review of de Menil (1971) (q.v.). E.J., vol. 83, 1973, 1340-1341.

K. Wicksell : "Selected papers on economic theory." London, Allen and Unwin, 1958.

O.E. Williamson : "A dynamic theory of interfirm behaviour." Q.J.E., vol. 79, 1965, 579-607.

O.E. Williamson : "Selling expense as a barrier to entry." Q.J.E., vol. 77, 1963, 112-128.

O.E. Williamson : "Wage rates as a barrier to entry : the Pennington case in perspective." Q.J.E., vol. 82, 1968, 85-116.

S.Y. Wu : "The effects of vertical integration on price and output." Western E.J., vol. 2, 1964, 117-133.

B.S. Yamey (1972) : "Do monopoly and near-monopoly matter? A survey of empirical studies." in M. Peston and B. Corry (eds.), 1972 (q.v.).

B.S. Yamey (1972b) : "Predatory price cutting, notes and comments." J.L.E., vol. 15, 1972, 129-142.

G.K. Yarrow : "Managerial utility maximisation under uncertainty." Eca, vol. 40, 1973, 155-173.

F. Zeuthen : "Problems of monopoly and economic warfare." London, Routledge, 1930 (reprint).