Firm Size, Intra Industry Performance and the Business Cycle: Empirical Studies using UK Panel Data

Peter Zoltan Lukács

Thesis Submitted for the Degree of Ph.D. in Economics

Department of Economics
University Of Warwick

April 1996

Firm Size, Intra Industry Performance and the Business Cycle

Contents Page

Acknowledgements

My most important intellectual debt is to Keith Cowling who both in person and in his written work on industrial economics and the development of modern capitalism has provided a constant stimulus for the ideas which have shaped this thesis. Without that stimulus I have no doubt that my understanding of the economic world and the research contained in this thesis would be infinitely poorer.

I would like to thank both Keith Cowling and Mike Waterson for the encouragement and intellectual stimulus they have provided whilst supervising this thesis and for the patience and understanding which they have shown during the process of its completion. I would also like to thank Martin Conyon who as friend, teacher and collaborator over the past few years has greatly aided the process of this work. I would also like to thank Martin for advising me about the collection of data and providing me with some of the data used within the thesis. I would like to thank Dennis Leech, Norman Ireland, Morten Hviid, Malcolm Sawyer and other members of the Economics Departments in Warwick and Leeds for providing stimulating discussions, comments and productive environments for generating research. Of the many fellow PhD students with whom I have had productive discussions I would like to thank in particular George Chouliarakis, Guy Vernon and Robbie Mochrie.

On a personal level my supreme debt is to Caroline Williams who has borne the brunt of the bad times as well as sharing in the good times which occur in the process of conducting a piece of research like this. Finally I would like to thank my mother, Judith Lukács, for provided a financial helping hand when I needed it, but mainly for her patience and support and for having faith in my abilities over the years.

Chapter 1: Cyclical movements in intra industry performance - an introduction

"*The most superficial examination of competition ... shows that, under certain conditions, if the bigger capitalist wants to make more room for himself on the market and expel the smaller capitalists, as in times of crisis, he makes practical use of this advantage and deliberately lowers his profit rate in order to drive smaller ones from the field.*" Marx (1981[1894] p331)

"*The struggle of the average-sized enterprise against big capital cannot be considered a regularly proceeding battle in which the troops of the weaker party continue to melt away directly and quantitatively. It should rather be regarded as a periodic mowing down of the small enterprises which rapidly grow up again, only to be mown down again by large industry.*" Luxemburg (1970[1900] p47)

There has been considerable interest in time varying performance particularly that related to business cycles in recent years. This topic has been a persistent focus from the end of the last century, as evidenced by the quotes above, continuing through much of economics since the 1930's, when the peculiarities of pricing behaviour, particularly in the United States, during the depression years were the driving force behind interest in the topic. Since then, there has been an ongoing debate on the effects of the business cycle upon pricing and profitability. In recent years this debate has intensified having been fuelled both by technical factors and politico-economic developments. On a politico-economic level the ending of the "Golden Age" of capitalism, which had been characterised by steady growth, low unemployment and relatively small cyclical fluctuations, in the 1970's can be seen as a prime reason for a resurgence of interest in such issues. On a technical level within the economics discipline two aspects have promoted increased interest in this issue. Firstly there has been a mushrooming of theoretical approaches to the question driven by the growing dominance of game theory in particular in trying to explain the prevalence of collusion during the business cycle. At the same time, but we would argue not in parallel, there has been an expansion of empirical testing of the question of time varying performance at a microeconomic level. This expansion can be largely attributed to the advent of improved techniques and computing capability for dealing with panel data which facilitates the examination of these issues. This thesis seeks to examine one aspect of this question, the relative performance of large and small firms, a subject which has been largely overlooked within the main body of the literature on time varying performance but which can add to, and point to, useful insights for that literature.

This chapter sets out to examine the focus on firm size within industries, i.e. relative rather than absolute firm size. Section I documents the changes in the size of the small firms sector and discusses some of the reasons for these changes. We will then turn to consider two main areas which are relevant to the general approach adopted but which have not been pursued extensively elsewhere in the thesis. In section II we briefly consider the reasons for differential performance between large and small firms focusing largely on the debate concerning market power and efficiency, which for reasons outlined, are not addressed within the thesis. Section III assesses the contribution of the game theoretic literature to our understanding of cyclical performance and collusion. We argue that the game theoretic approach provides us with neither a solid methodological basis for pursuit of the analysis nor empirically testable hypotheses of sufficient clarity. Section II and III therefore address bodies of material that some may have seen as likely to contribute to a greater degree to a thesis with this focus and explain why that is not the case. Section IV of this chapter presents evidence at an aggregate rather then industry level of the relative performance of large and small firms in UK manufacturing. Sections I and IV of this chapter are designed to describe the background for the theoretical and empirical analyses which follow. Sections II and III place limits upon the scope of the thesis

Chapter 2 considers notions of firm flexibility, one of the main areas of cyclical performance of firms considered within the recent literature. The chapter critically assesses the concept of flexibility and some of the theoretical attempts to model flexibility. In particular the focus is upon the relationship between empirical studies of sales variability and notions of flexibility. It is argued that the current tests for flexibility based upon sales or output variability contain flaws of both method and interpretation and suggests a potential alternative framework of empirical analysis.

Chapter 3 addresses the issue of vertical integration and its implications for the measurement of economic performance. It is argued that many treatments of the issue of vertical integration and intra industry transactions in empirical analysis involve simplistic and/or unstated assumptions concerning the transactional relations between input buyers and suppliers or, at very least, these relations are under-theorised. The chapter attempts to systematically analyse the implications of different forms of supplier/buyer relations for economic performance and provides empirical analysis of these relations at the industry level in UK manufacturing.

Chapter 4 considers intra industry performance in the long run in UK manufacturing. The chapter considers the theories of intra industry performance and the methods used for previous tests of intra industry performance. The chapter exploits the panel data components of data on intra industry performance existing within UK manufacturing to reassess the previous studies of intra industry performance and point to inconsistencies of interpretation.

Chapter 5 examines the cyclical and dynamic behaviour of intra industry performance in UK manufacturing. A detailed discussion of the theoretical predictions concerning the relative cyclical performance of large firms is conducted. A variety of empirical tests aimed at revealing information concerning these predictions are then conducted using a panel data set combining information from the UK census of production and the CBI industrial trends survey.

Chapter 6 briefly offers some concluding remarks on the results obtained, the methodology adopted within the substantive chapters of the thesis and future directions for research.

Before we proceed two initial notes are necessary. The first note concerns the methodological stance of the thesis, the second note concerns use of the term 'business cycle'.

It is perhaps conventional in an introduction to a thesis to set out the methodological framework which is adopted. This is almost certainly less common however in Economics theses than any other discipline within the social sciences. Within this thesis there has been no conscious decision to adopt a specific methodological stance. The primary, but not exclusive focus is on empirical analysis, but it is hoped that the empirical analysis is based upon sound theoretical foundations. Questions of interest have been identified and pursued within an approach dictated by a theoretical pragmatism rather than the positivism that is more commonly associated with Economic analysis. Therefore, an abiding theme of the thesis is to draw out the implicit assumptions of the previous research in this field. It could be argued that the methodological stance adopted within the thesis has itself been implicit and it will be the focus of the concluding chapter of this thesis to consider some of the implicit methodological approaches adopted within the various chapters. The exception to this rule is the direct assessment of the utility of game theoretic analysis that is discussed in section III of this chapter.

Throughout the thesis mention is made of "the business cycle". This is a term which has many connotations, many of which can be fairly unhelpful. It is however a useful shorthand for all the time-varying aspects of economic behaviour which are not secular trends[1]. It is in this respect that the term is employed throughout the thesis. The reader should note that nothing more should be inferred from the use of the term. No specific 'theory' of the business cycle is being adopted and any implicit assumptions concerning the amplitude or periodicity of the cycle are hereby explicitly eschewed. Movements in most economic variables appear to be only rarely regular in time or of uniform size. In addition to this point of clarification, it is probably helpful at this stage to explain the use of the terms pro and counter cyclical. An economic variable is described as procylical (countercylical) if it realises a higher (lower) value, perhaps at an appropriate lag, when the level or rate of growth of demand is high. This is clearly not a precise or all encompassing definition but is one which is designed to ensure the widest possible applicability. At points where a more precise definition is required fuller explanation is given in the text.

[1] We do not presuppose any norm for secular trends. Neo-classical economics regards full employment as the base case but excess capacity is the norm for most alternative (e.g. Marxist or Post-Keynesian) schools of economics.

Specific problems of interpretation arise in attempting to separate trend from cycle. Many recent studies of the business cycle tend to focus on relatively short periods of time that barely cover more than one complete cycle. The focus is usually on short term movements around a trend. A significant proportion of the theory relating to 'cyclical' movements, especially those originating in discussions of the 1930's relates, however, to periods of prolonged slump that cannot be accurately characterised as short run departures from a norm of full employment. Indeed, it may be argued that the important changes to observe may be the changes from periods where full employment is the normal state of affairs to periods of prolonged stagnation, where excess capacity is the norm, and that short run movements around these more fundamental shifts are therefore incidental.

The techniques of economic analysis are not well adapted to considering such shifts. To consider a 'structural break' is to question the ahistorical positivist basis of most economic analysis. In this initial chapter of the thesis, we attempt to outline some of the trends and shifts that may have occurred in recent history. At other points through the thesis we have attempted to set the analysis in context by considering the time period evaluated in relation to studies using different time periods. Despite these attempts we are forced by our continual interaction with conventional economic theory, to focus, unwillingly to some extent, but not entirely so, on the short run portion of the analysis for the substantive elements of the thesis. However we also attempt to be mindful of the historical context that is governing the relations we are investigating.

I. The size of the small firm sector[2]

Generally, within the economics profession small firms have been regarded as of relatively little consequence. The focus has been upon big business and its perceived effects on the economy. This has partly been a result of rising aggregate levels of concentration and rising levels of

[2] A more comprehensive examination of some of the evidence on the size distribution of firms in UK manufacturing for the 1980's can be found in Dunne and Hughes (1992). The discussion here examines both a longer period and focuses on those elements which are of more direct interest to the main subject of the thesis.

concentration within industries. This trend of concentration became particularly marked within the post war period. Since the 1970's and particularly in the 1980's there has been a relative resurgence of interest in small firms as they have grown in importance and the trends towards higher concentration have slowed or been reversed. This has also coincided with a greater focus on small firms outside the economics literature, particularly within forms of industrial sociology; areas which have had relatively little impact on economic theory[3].

Table 1[4] indicates the trends over the last 20 years in the size distribution of firms in UK manufacturing industry. A note of caution must be sounded that the data is based upon the UK Census of Production which changed its definition of unit of observation from establishment to business in the 1984 Census. As a result a large increase in the total number of enterprises is recorded in that year, most of which appear to be very small[5]. There is, therefore, a discontinuity in the figures between 1983 and 1984[6]. Table 1 indicates that the pattern of employment was relatively constant over the 1970's. However, there was a sharp reduction in the proportion of employees in very large firms i.e. those employing more than 5000 employees from 1979-1992 even accounting for the discontinuity in 1984. All other size classes increased their share of employment during the 1980's, the largest increase coming from the smallest firms, i.e. 1-199 employees[7]. As we can see, however, the number of firms in the small category has increased

---

[3] There are a number of key texts on the borders of the economics and industrial sociology literatures which consider the growth of small firms. Some of the more interesting are Piore and Sabel (1984), Hirst and Zeitlin (1988) and Best (1990). There are exceptions within the mainstream economic literature, but not many. An example of an attempt to incorporate some of the features identified within industrial sociology into economic theory is Milgrom and Roberts (1990)

[4] All Tables and Figures in this chapter, and throughout the thesis, are placed at the end of the relevant chapter.

[5] Dunne and Hughes (1992) indicate that most of the increase came from enterprises with less than 10 employees. It is important to also note at this stage that the definition of establishment is a legal one and not a location based one. The establishment is the smallest feasible accounting unit. This may include more than one plant.

[6] This change in definition goes unreported in almost all empirical studies of UK manufacturing in the 1980's. Whether this is through ignorance or an implicit (and in our view invalid) assumption that the change had no material effect on the data is unclear. This omission is especially severe for studies, such as Henley (1994), that attempt to explain changes in concentration over the period.

[7] It may be argued that enterprises with more than 100 employees are not particularly small. There are two reasons for adopting 1-199 employees as a definition of a small firm. Firstly it is consistent with the Bolton report on small firms which suggested in manufacturing that less than 200 employees should be the criteria, Bolton (1971). Secondly this is the criteria used by the CBI in their size class analyses and therefore this definition enables comparability across our two principal data sources.

rapidly. Even accounting for the 1983/4 discontinuity, which adds substantially to the number of smaller firms, there was a 20% increase in the number of small enterprises between 1972 and 1983, the largest portion of which occurred in 1972 and 1973. There was a further 16% increase between 1984 and 1989, before the number of firms fell back slightly in the early 1990's. Meanwhile the number of large firms has fallen consistently over the period apart from a slight increase in numbers at the end of the 1980's.

Figure 1 puts a more sanguine perspective upon this rising importance of small firms because it indicates that, while the share of employment has increased, this has not resulted in substantial absolute gains in employment. In fact the increase in share appears to largely result from declining employment in the larger firms and broadly constant employment amongst the smallest firms. Such a downward trend in industrial employment is the main focus of the ongoing debate on deindustrialisation[8] in the UK. If only because the small firm sector appears to be bucking this trend it is clear that a greater focus on the smaller firms is warranted. Employment is however only one factor and it could be argued that small firms are becoming relatively less productive and therefore are absorbing more employment but not increasing their share of net output. This is not borne out by Table 1 and Figure 2 which indicate that, while the shift is less marked than with respect to employment, an increasing share of net output has been produced by the smallest firms at the expense of the giant firms. Therefore, it would appear that, certainly over the period of the 1980's, but also continuing into the 1990's, small firms have been becoming relatively more important in UK Manufacturing. As most small firms are in the service sector the focus on manufacturing also probably understates the importance of small firms.

Having shown that small firms are becoming more important, in UK manufacturing at least, it is now necessary to assess some of the reasons why this may be the case[9]. One of the main reasons put forward for the growth of the small firm sector, both here and abroad, is a trend on the part of large firms towards vertical disintegration. Large firms may have been more willing to withdraw

[8] See for example Rowthorn and Wells (1987).

[9] For a fuller assessment of the general issues relating to the importance of small firms in the economy, see You (1995).

from direct ownership of organisations performing various tasks. This may take the form of the tendering out of ancillary services or it may involve a withdrawal from certain aspects of the production process itself. The reasons why firms may be more willing to do this now, reversing the trend of the post war period to vertical integration, are many and various. It has been argued that the objective of circumventing the power of organised labour was one factor in this. Small firms have always tended to be less strongly unionised[10], have paid lower wages[11] and have been exempt from certain employment protection legislation. Millward and Stevens (1986) report that the mean union density for establishments with 50-99 employees is 30% while this rises to 72% for establishments employing over 1000 workers.

Vertically disintegrating parts of the production process to smaller units therefore enables firms to bypass unions and hence cut labour costs. Whether this has been the case or not is debatable. The trade union reforms introduced in the 1980's in the UK have certainly weakened the trade unions, but this effect would be expected to benefit the larger, more highly unionised firms to a greater extent. This would be expected, therefore, to lead to a consolidation of large firms rather than the regeneration of small firms. It is an empirical question what the contribution of structural changes in terms of increased presence of small firms has had upon the size of union membership. There is some debate concerning this issue but Green (1992) estimates that 30% of the decline in union density between 1983 and 1989 can be attributed to compositional factors as opposed to the effects of the business cycle or changes in the legal framework. Of these compositional factors the changes in industrial sector and establishment size are found to the most important. The additional factor which must be considered is that the unit of observation, from the point of view of union density, is size of establishment rather than the size of enterprise.

---

[10] There are some notable exceptions to this rule, in particular, the printing unions prior to the reforms of the 1980's were very highly organised despite being predominantly based in relatively small firms. In addition the thriving small firm sector in the "third Italy" is based upon high levels of unionisation and relatively high wages.

[11] This can be largely explained in terms of different levels of skills in the workers together with different capital structure, hence different levels of productivity.

Larger firms may have been able to circumvent union power by reducing plant size but running more plants. Table 2 indicates that while there is a declining average establishment size across the period this largely arises from the increased number of very small businesses. For larger enterprises the average size of establishment was on a downward trend from 1972 through to 1983 then from 1984 onwards establishment size regained an upward trend. The reverse observation can be made for the number of establishments per firm which rose almost continuously from 1972-1983 and thereafter fell. First we have to decide whether this is a quirk of the data redefinition which took place in 1984. It seems unlikely that there would be a consistent rise in the number of "establishments" per firm which is accompanied by a consistent fall in the number of "businesses" per firm, particularly if a Williamsonian account of the need for financial controls in M Division firms is accepted[12]. Even if we can rule data redefinition out, the story is still not immediately clear. Such a path is not consistent with the progressive introduction of flexible technology which requires smaller scale of plant as this process has proceeded throughout the period and the inflow of foreign direct investment is likely to have accelerated rather than decelerated the process. It is possible that large firms sought to reduce establishment size in the late seventies and early eighties because of the strength of trade unions, but, once the reduction in power due to the changed climate and legal environment of industrial relations took effect the pressure to reduce establishment size was relaxed and establishment size grew once again. Alternatively, the effect could be related to the business cycle although no effect is discernible in the recession of the early 90's comparable to the changes from the early eighties. The reasons for these changes therefore remain an open question.

An additional argument for vertical (dis)integration increasing the number of small firms which has been put forward by Milgrom and Roberts (1990) is that the rise in flexible manufacturing technology has meant that firms have invested in less specific machinery. As asset specificity is one of the prime triggers for opportunism in inter firm relations and this opportunism can be circumvented by vertical integration, the decline in the specificity of the technologies being used

[12] Williamson's theory of the multidivisional form of business enterprise (Williamson, 1970) implies that the number of accounting units within a firm should closely follow the number of economic units.

reduces the need for vertical integration. If flexible manufacturing technology also reduces economies of scale, this will be an additional factor favouring the rise of small firms.

A separate explanation for the rise in the number of small firms is that it results from (as opposed to causes) changes in the degree of competition. The relationship between the intensity of competition and the size of the small firm sector is fairly complex. The possession of monopoly power by an oligopolistic group can permit inefficient small firms to coexist in the same industry, what one might call an "uncompetitive fringe". As Cowling (1982) has suggested, the survival of these small firms is dependent upon them not impinging too greatly upon the profits of the oligopolistic group. If the oligopolistic group feels threatened, the small firms will be eliminated through either predation or take-over. This can be seen as a more general version of the constraints implied in Fudenberg and Tirole's (1984) puppy dog ploy. Competition in a dynamic sense of innovation, changes in markets, import competition, may however benefit small firms if small firms are better at adapting to such conditions of uncertainty. Such arguments have been put forward by Carlsson (1989) and Mills and Schumann (1985) and aspects of these arguments are considered in more detail in Chapter 2. If competition of this form has been on the increase then small firms would expect to benefit.

It has been suggested that the size and strength of the small firm sector is affected by the level of unemployment because of the effect of pushing unemployed individuals into self employment or small business ownership. There are two effects to this, one is a pure cyclical effect where some individuals will shift between self employment and employment depending upon the scarcity of employment, i.e. there is a preference for employment but in the absence of such opportunities self employment is better than unemployment. Therefore, if opportunistic small businesses of this sort are widespread, we should expect rising unemployment to be accompanied by rising numbers of businesses, nonetheless as these businesses are unlikely to be more than economically marginal we should not necessarily expect the sector as a whole to grow in strength. The second effect of unemployment, in addition to providing a motive for small business creation, is the creation opportunity in the form of redundancy pay, which may lead to an increased number of small businesses. By the same token, the rate of small business creation is likely to be affected by changes in the housing market as house ownership and values determine potential bank lending given the high collateral levels required for small businesses.

An important point in relation to small firms is that frequently they will not be providing identical goods and services to those of large firms. Instead they will be producing differentiated products and therefore will not necessarily be in direct competition with the larger firms. An example would be in the SIC industry, "motor vehicles" where small bus and coach manufacturers or sports car manufacturers such as Morgan are classified alongside the mass car producers. If the demand for more specialised niche market products is increasing over time with rising incomes, this will tend to increase the number of smaller firms. Again, the attitude of large firms is important as their own strategies will set limits upon the ability of small firms to acquire significant market niches without retaliation. An example of this would be the reaction, set out in Schmalensee (1978), of the major players in the Ready to Eat breakfast cereal market to the development of small independent muesli producers. Small independent producers discovered a substantial hitherto untapped market niche for muesli type cereals but within a few years the major producers such as Kellogg's had managed to dominate the niche through highly advertised branded versions. Without the continual dynamic competition of generating new market niches discussed above the size of the small firm sector is inevitably limited. Shaked and Sutton (1987) and Sutton (1991) consider this aspect in more detail.

II. Cost differentials, market power and small firms

Now that we have examined the size of the small firm sector and some of the determinants we want to look briefly at some of the determinants of the performance of the small firm sector. This task is an ongoing theme of the thesis and is therefore taken up again in more detail, relevant to the particular pieces of empirical analysis, particularly in section II of Chapter 4 and through most of Chapter 2. The object of the discussion at this stage is to give a general flavour as background to the subsequent analysis. To do this we need to focus on the main aspects which differentiate large from small firms. It has been argued, following Marshall's analysis of the representative firm, that in essence, small firms should be treated the same as large firms. The main distinguishing features which we will concentrate upon, and which overlap in areas, particularly with regard to input purchases, are cost differentials between large and small firms and differences in the extent of market power[13]. These factors have been the traditional factors affecting performance in the industrial organisation literature and are therefore the focus of the next section.

Throughout the many models of oligopoly and imperfect competition it is frequently assumed that all firms share the same cost structure. This appears to have two origins. The first is simplification; models become much less complex when all firms have the same cost function, this particularly applies to the game theoretic models. The second is the widespread neo-classical assumption of the uniform availability of technology implying that there is a single unique cost minimising technology; this view of technology is most crucial in the contestability theory of industrial organisation.

For a number of reasons the case of identical cost structures is, in practice, the exception rather than the rule. The existence of specific entrepreneurial and managerial assets will lead to different costs of combining inputs. In addition, factor prices will differ across firms because of the differences in bargaining power and levels of unionisation, and the ability (or lack of it) to achieve economies of bulk transactions. Steindl (1945) has argued that the 'principle of massed reserves' whereby large firms by virtue of a greater number of transactions are less vulnerable to stockouts of finished or intermediate goods hence require lower levels of inventory to absorb such fluctuations. This argument is equally applied to finance where differences in the likelihood and impact of punctual or late payment weigh more heavily on small firms requiring them to maintain larger amounts of credit or financial reserves to cope with these fluctuations. Whereas the latter phenomenon appears to be borne out by anecdotal evidence[14], the evidence, see for example

[13] Steindl (1945) focuses in addition on differences in the cost of borrowing, a factor we subsume under cost differentials.
[14] In February 1996 the British deputy prime minister publicly confessed to exploiting such advantage during his time in control of a large business.

section IV of this chapter, does not necessarily support the former proposition that large firms hold proportionately lower levels of inventory.

Different types of technology are one of the main causes of different cost structures. Small firms are more likely to make use of older or second hand capital. Also in industries where there are economies of scale, small firms may try to compensate for their lack of scale by being more adaptable to demand conditions. This argument has been put forward by Stigler (1939) and is discussed more extensively in Chapter 2 but at this point it is useful simply to note the two strategies firms can adopt to achieve this adaptability. The first is employing where possible fixed plant which is divisible and which therefore minimises the cost penalty of operating at inefficient scale levels, the second is transforming fixed costs into variable costs principally through a shift to more labour intensive services.

Having briefly considered some of the causes of cost differentials we now examine some of the consequences of cost differentials for industrial economic analysis. The most prominent oligopoly model of the past few years is the Cowling Waterson (1976) model based upon short run profit maximisation with given cost structures. Firms produce a homogenous product therefore there is a single uniform price charged within the industry yielding an equilibrium relation.

$$\frac{P - c_i(x_i)}{P} = \frac{s_i(1+\lambda_i)}{\eta} \quad (1)$$

where P is industry price, η is the industry price elasticity of demand, $c_i(x_i)$ is the marginal cost for firm i based upon output level x_i, s is the market share of firm i and λ_i the conjectural term of firm i. Demsetz (1973) argues that the cost differentials between firms are the sole cause of differences in profitability hence the differences in both structure and profitability between industries are solely determined by the differences in their cost structures i.e. differences in levels of efficiency between firms.

This argument is countered by Cowling (1982) who argues that all of the elements on the right hand side of equation (1) are subject to the strategic control of capitalist firms[15]. The elasticity of demand is controlled through advertising policies, market share is controlled through merger and the cost decisions implicit in the strategic choice of technology, and the conjectural term is controlled through tacit or explicit collusion. Such arguments shift the focus onto market power and away from cost differentials in determining performance. The notion of market power is more broadly conceived than simply through the sustainable level of the conjectural term and applies to the scope for, or absence of constraint on, strategic action. It is clear at this point that both market power and efficiency considerations would appear to favour large firms implying a performance advantage.

Returning to the more narrow focus of market power as collusion, if a consensus has emerged at all from recent empirical literature, it is that collusion is centred around price. As Geroski (1988, p 111) points out "the evidence suggests that concentration-price correlations are stronger and more stable than concentration-profits ones". This can fit in with a view of the coexistence of rivalry and collusion, where collusion is centred on price while rivalry is centred upon cost. Taking this point of view reinstates Steindl's (1952) theory of profitability based upon a conventional, but not exclusively, uniform price structure beneath which there is a system of cost differentials which determine the abilities of individual firms to generate profits.

It is well recognised that the existence of cost differences puts pressure on the sustainability of any collusive outcome and, in particular, an agreement on side payments is necessary for the joint profit maximum to be maintained (see Schmalensee, 1987). This then begs the question as to whether any joint profit maximum is practically attainable in this context. In the study of the great salt duopoly, where firms are observed to have different cost functions, Rees (1993) is able to reject competitive behaviour but also is able to reject the hypothesis of a joint profit maximum. This recognises that a joint profit maximum can only be achieved with very close co-ordination

[15] To the extent that Demsetz believes that efficiency is the outcome of strategic decisions taken by firms in pursuit of profit the two arguments are not in complete opposition. Both arguments attempt to reinstate a form of causality to (1) that is rejected by others (e.g. Clarke and Davies (1982)).

which may involve a firm revealing data about its costs which is likely to be strategically damaging and also may also require side payments or uneven profit reductions. Therefore, even in the presence of some form of price co-ordination, joint profit maximisation is unlikely.

The most important qualification of this discussion and the main obstacle to a formal conclusion that there is an ubiquitous performance advantage for large firms is empirical. There have been a number of studies looking at the variation of profitability across size class bands in industries (e.g. Clarke, Davies and Waterson (1984) and Caves and Pugel (1980)). With the assumption of homogeneous conjectures firm size should be directly related to price cost margin through variations in cost. Given homogenous products and a uniform price, the firms with the lower cost will have the highest market shares. Yet all of the studies find examples of industries where there is a negative correlation between price cost margin and firm size. Waterson (1989) has offered three explanations for this finding.

i) there is an aggregation problem when an "industry" is broadly defined. If two distinct industries are combined for the generation of statistics, it is demonstrated that while a positive correlation between profitability and market share may exist in each industry, a negative correlation may be observed in the aggregated industry. This would appear to constitute an important element of the empirical findings as a number of cases are reported in miscellaneous or aggregated industries from the studies including Steindl (1952). This problem would appear to have wider significance, in that the existence of multiproduct firms in a well defined industry could equally create this problem.

ii) the second possibility is that firms do not have a common level of conjecture but have firm specific conjectures. This can be the case even when firms have identical cost functions, with either increasing or decreasing returns to scale but produce different outputs.

iii) the third explanation is that firms produce differentiated products and therefore firms may locate in low volume but high mark-up areas of a market.

III. Game theory and cyclical movements in collusion and profitability

The focus on cyclical movements in collusion has developed in response to the mushrooming of game theoretic analyses. The most prominent of these are the models of Green and Porter (1984) and Rotemberg and Saloner (1986) and the latter model has subsequently been extended in two papers by Haltiwanger and Harrington (1991) and Kandori (1991). In this section we will outline the main thrust of these papers and then attempt to critique them from both within and without. We argue that each of the papers lacks an important element of realism which raises questions about the appropriateness of the choice of game theoretic format. We then go on to argue that these game theoretic approaches do not, and perhaps cannot, offer much help in providing a basis for empirical analysis, especially at the level of analysis adopted within this thesis.

Green and Porter acknowledge that their model is an attempt to refashion the Stigler (1964) model of oligopoly. They analyse the situation where firms are setting output in an attempt to collude. Firms are assumed to be symmetric. The main feature of the model is that demand is subject to exogenous shocks which are i.i.d. and which are unobservable to the participants. This brings about a signal extraction problem that in order to maintain the collusive agreement firms engage in price wars if the price falls below a certain trigger price. At the margin cheating or chiselling is deterred because the expected loss from the increase in the likelihood of a price war is greater than the expected benefit from the chiselling. It is therefore predicted that price wars occur at low points in demand, however this is not due to cheating but is part of the deterrence of cheating. Without the periodic price wars in response to exceptionally low demand then there would be no disincentive for each firm to cheat.

Rotemberg and Saloner point to a different force driving the cyclical movements in prices and profits. They suggest that in the case where demand is perfectly observed but is subject to an i.i.d. disturbance, it will be more difficult to collude at points of high demand. This is largely due to the fact that while the benefits of cheating accrue at periods of high demand hence when greater profits can be made from cheating the losses incurred from any punishment phase are at the average level of demand. Therefore the gains to cheating (relative to the punishments) are greater in periods of high demand. In the Rotemberg and Saloner model this can lead to situations where the price will be lower in periods of high demand compared to the price in periods of low demand. This will be because the maximum sustainable price, i.e. the maximum price at which no firm is willing to defect, will be less than the full joint profit maximising or (in the case of symmetry between the firms) monopoly price. Therefore while there may be no "price war" in the conventional sense in the model there is a clear disposition towards countercyclical pricing.

The framework has been extended by Kandori (1990) and Haltiwanger and Harrington (1991) who, rather than looking at the case of i.i.d. demand shocks, examine serially correlated demand shocks. This more reasonable assumption adds to the analysis of the difficulties of sustaining collusion over the cycle. The findings of Haltiwanger and Harrington are particularly interesting in this respect. They suggest that the incentive to cheat on a collusive agreement is greatest at the start of a recession. The reason for this is that at the start of a recession the level of demand is relatively high whereas the punishment phase will occur during the recession when demand is low. The effect is essentially a generalisation of the expectation approach in terms of the comparison of payoffs. Different expectations of demand will yield different optimal solutions. The important concept is Rotemberg and Saloner's trade-off between current gains from defection and expected losses in a punishment phase.

A criticism which can be levelled at these papers is that despite increases in demand firms never become capacity constrained. Models incorporating capacity constraints, such as Staiger and Wolak (1992), produce qualitatively different results to models where firms are unconstrained. They adapt the i.i.d. demand function of Rotemberg and Saloner to a duopolistic situation where firms operate with capacity constraints. When demand is high, firms will be operating at or close to capacity. Therefore, at times of high demand the opportunities for cheating are decreased because the possible output gain from a price cut is constrained to be small. When demand is low the opportunities for increasing supply are greater and therefore the equilibrium collusive price will be lower. In this way, the results achieved by Rotemberg and Saloner are reversed by the introduction of capacity constraints.

The Staiger and Wolak model builds upon Benoit and Krishna's (1987) multi period version of the Kreps and Scheinkman (1983) model of price competition with capacity constraints. Benoit and Krishna (1987) establish that in an infinitely repeated game excess capacity is a necessary condition for a collusive outcome. Davidson and Deneckere (1990) extend this for a class of equilibria, finding that increases in the *level* of collusion are associated with increases in the *level* of excess capacity. Staiger and Wolak's model reverses this result because in their model the incentive to cheat is increasing in the current level of excess capacity and decreasing in the expected future level of excess capacity, the two being inversely related.

There are a number of problems which arise in considering the applicability of these models to empirical examples of collusion amongst oligopolists. The two features which I wish to focus on in the following discussion are,

i) In none of the models discussed above do the firms actually cheat on the collusive agreement. While price and output choices may vary over time, at each point in time all firms are following the collusive equilibrium path.

ii) The models considered above are games of complete information[16].

We will address the second point before considering the first. In the Green and Porter model, imperfect information arises from the fact that the realisation of demand is not observed before firms make their output decisions. Therefore, the equilibrium price will not be known a priori. Despite this everything else is common knowledge for the participants. In particular the mean realisation of demand and its variance are known by all participants. It is assumed that with symmetry in respect of both cost conditions and the discount rate and with this knowledge of demand conditions each firm can calculate the optimal trigger price strategy. In the Rotemberg and Saloner type models the realisation of demand is observed before price setting and firms also

[16] This is the sense of complete information as defined in Gibbons (1992, p1), that is "each player's playoff function .... is common knowledge among all the players." Therefore Green and Porter (1984) is a game of complete information even though the players have imperfect information on the realisation of demand in any particular subgame prior to moving.

know the distribution of demand fluctuations or, in the case of Haltiwanger and Harrington, know the correct time series path of demand. From this information, again assuming symmetry, firms can calculate the maximum sustainable price. There are multiple equilibria in this situation, the range of which depends upon both the discount rates and the parameters and realisations of the demand process. However because of the perfect knowledge firms can react and choose the Pareto optimal[17] price given these situations.

Invoking bounded rationality as a critique of these models may appear artless. The whole foundations of the game theoretic approach rest upon a superimposed level of rationality through the solution concepts employed. However this criticism would nevertheless appear to have some justification. Only in a relatively mature and stable industry could one expect a sophisticated <u>implicit</u> agreement such as a trigger price strategy to be developed. Similarly, it requires a degree of calculation to find the required monopoly price for each realised level of demand which may not be open to all firms. Clearly, mechanisms to bypass this need for independent calculation can be set up, such as price leadership but these would then need to be incorporated into the pricing framework, with potentially very different results. This goes to the heart of any consideration of collusion. A fundamental aspect of collusion is the monitoring, detection and punishment required to enforce the agreement, a subject treated in great detail by these theories. However, anterior to this we also have the co-ordination problem implicit in setting up a collusive agreement, tacit or otherwise, a topic which is ignored or assumed away in these studies. To remedy this requires an analysis using dynamic games of incomplete information. A model which addresses these issues directly has been developed in Slade (1989).

Slade (1989 p295) explicitly recognises the "seemingly disequilibrium" nature of price wars and also the prevalence of strategies which do not constitute sub game perfect strategies in the complete information context such as tit-for-tat. Slade's model suggests that a change in demand or cost conditions which affects all firms will tend to disturb an existing equilibrium which may

[17] Pareto optimal when assessing the two duopolistic firms as the sole agents deriving welfare from the situation.

or may not constitute a tacitly collusive agreement. This observation of changed conditions will induce price changes on the part of participants. This need not represent cheating of the form prescribed above, i.e. a deliberate attempt to undercut rivals, but could be simply an attempt to adjust price to a new level appropriate to the changed conditions. In Slade's model, firms observe a change in demand and cost conditions has occurred but do not observe the new values of the demand and cost parameters. Successive price movements will signal information about the new demand and cost conditions. Gradually, as firms know more about the new demand (or cost) conditions they can move towards the new equilibrium which may be characterised by a greater or lesser degree of tacit collusion. The reason why the movement is 'gradual' is because the solution concept employed is of a perfect ε-equilibrium[18] as defined in Kalai and Stanford (1985). This reduces the problem of co-ordination as firms approach the equilibrium position using intertemporal reaction functions rather than through simultaneously calculating the optimal price. Analysis using these intertemporal reaction functions therefore relates to the results of Axelrod (1984) and Kreps *et al.* (1982) for as Kalai and Stanford (1985; p146) point out "they possess a kind of fractional strength tit for tat character".

The Slade model therefore has two desirable features absent in the previous models. Firms do cheat and they also have plausible mechanisms for co-ordinating on a price that does not require the imposition of full rationality. Both of these elements are a key feature of the original Stigler model of oligopoly, and the Slade model has a solution to both of them which appears realistic because of its focus on the difficulty in attaining the solution. The feature of the Slade model, which is slightly problematic given the focus of the thesis is, that price wars are a result of changes in cost and demand conditions not of movements in any particular direction in those costs or demand. Therefore, it has relatively little to tell us about the incidence of price wars and the business cycle although one may seek to hazard some informed guesswork on whether it is more likely that downward or upward movements is likely to lead to the breakdown of the collusive agreement.

[18] The intuition behind this solution is akin to the near rationality concept used in neo Keynesian macroeconomics, e.g. Akerlof and Yellen (1985)

The element which is an essential part of the Stigler model, but which receives no attention in any of the previous models, is the existence of asymmetry between firms. It is the smaller firms which are prone to chiselling in the Stigler model because small output changes by these firms are less likely to be detected than changes by the large firms. The existence of asymmetry is also identified more generally as a source of difficulty in the creation and maintenance of collusion elsewhere, e.g. Schmalensee (1987). To examine the effect of asymmetries in the game theoretic analyses we have to look at a separate literature. There are two main types of asymmetries which are considered in this literature, cost structures and discount rates. The conclusions of these studies are unenlightening however, merely confirming formally that which was previously intuition and common knowledge. Harrington (1989) considers the case of different discount rates and concludes, in addition to the standard result from symmetric analyses that a general lowering of discount factors makes collusion more difficult, a conclusion which also carries over to asymmetric analysis, that those firms with lower discount factors, those who are most impatient, will receive a higher output quota to compensate them for not cheating on the agreement. Demange and Ponssard (1985) find that low cost firms have a stronger incentive to price competition than high cost firms. Osborne and Pitchik (1987) find that the firm with a smaller capacity in a collusive duopoly will, apart from the extreme case of total capacity being less than or equal to monopoly output, maintain a higher level of capacity utilisation than the larger firm.

A further literature addresses the issue of asymmetry in the context of imperfect information regarding firms' cost structures. Roberts (1985) notes that imperfect information concerning the costs of rivals is of a different order to imperfect information on the actions of rivals. Green and Porter model a problem of moral hazard where there is a potential incentive to cheat on a collusive agreement. With imperfect information on costs the problem is one of adverse selection, a problem of devising a collusive agreement that will be robust in the sense that firms have the incentive to correctly reveal their private information enabling joint profit maximisation and firms' do not have the incentive to cheat.

Crampton and Palfrey (1990) and Kihlstrom and Vives (1992) extend Roberts' analysis from the duopoly case to the case of multiple rivals, a case that is particularly relevant to participation in agricultural marketing cartels. The key results of these papers depend upon three aspects; the feasibility of making side payments; the timing of potential withdrawal from the cartel; the robustness of the collusive agreement to withdrawal (defection) by a single rival. The main conclusion is that in the large numbers case and with side payments a form of efficient cartel agreement that obtains truthful revelation of private information is sustainable. Therefore Kihlstrom and Vives (1992; p 392) argue "asymmetric information is not a serious deterrent to collusion by a profit maximising cartel."

Does the game theoretic framework represent a potential way forward for empirically examining business cycles and intra industry behaviour? Whilst the literature closely examines the effects of demand fluctuations, it is important, in order to make these models consonant with the empirical investigation, to assess their implications when the symmetry assumption which underlies all of the models is dropped. However, when one tries to extend these model to consider asymmetries between firms, beyond the level of fairly trite analysis, they become very complex.

Sutton (1990) criticises the effectiveness of game theory as a methodological tool for the empirical analysis of industries. There are two strands to the argument, firstly the widespread presence of multiple equilibria leads to difficulty in predicting outcomes. This problem can be reduced but not completely eliminated by imposing a higher degree of rationality, the introduction of equilibrium 'refinements' such as renegotiation proofness, to reject certain equilibria. However it is not clear whether these additional pressures on strategic decision makers are reasonable[19]. Secondly the flexibility of game theoretic modelling ensures that most 'reasonable specifications' can support many outcomes and, by corollary, many outcomes can be supported by a wide variety of reasonable specifications.

[19] Experimental evidence in decision theory regularly provides evidence of "irrationality" in the form of preference reversal in very simple experiments using intelligent, well educated subjects.

The further problem not considered by Sutton is that of observing the relevant variables. Even if the problems identified by Sutton had been resolved, the essential difference between the Green and Porter and Rotemberg and Saloner models concerns the information set available to decision makers prior to the price decision. This information set is not open to us to observe and, even more problematically, neither is the counterfactual decision that would have occurred in the absence of strategic action open to observation.

Empirically successful attempts to implement game theory are few and far between[20]. Bresnahan (1989) summarises a number of industry studies but the notable feature of most of the studies is the absence of an empirical framework derived from game theoretic notions. Instead the major shift in terms of the 'New Empirical Industrial Organization' described by Bresnahan was in terms of the use of price and output data to investigate industry relations rather than profit, sales or other accounting or census based measures of performance. The exceptions have been the string of papers by Slade, e.g. Slade (1986, 1992) and citations therein, and Rees (1993)[21]. Slade exploits data on a price war in the Vancouver gasoline market and attempts to estimate intertemporal reaction functions that may accord with the decisions taken by the gasoline retailers. The relatively close relationship between theory and empirics in this case derives from the very close observation of institutional characteristics. The main conclusion, confirming the conclusion of Slade's theoretical work described above, is that the gasoline firms appear to adopt simple strategic decision rules in their pricing decisions but that the result cannot be accorded generality because of the specific nature of the data set. Rees, by contrast, examines a number of game theoretic models for their consistency with the observed behaviour of the UK salt duopoly. Despite confirming that collusion is potentially sustainable by the punishment threats of a repeated game, Rees finds that no specific model of competition can account for the precise structure of prices, capacities and outputs.

[20] Rotemberg and Saloner (1986, 1989) among others, present some empirical evidence in support of their theories, but the link between the theory and empirical approach is fairly weak and is certainly susceptible to the second criticism made by Sutton.

[21] A further exception, noted in Singh, Utton and Waterson (1995) are empirical studies of auction markets where the "rules of the game" are both closely defined and public knowledge.

The conclusion that arises from this discussion is that the game theoretic revolution that has swept industrial economics has failed to provide us with obvious new avenues for the empirical investigation of the issue of collusion and the business cycle. We may be able to learn something from event studies where theory and empirics can be closely tailored to the characteristics of the event[22]. We are, however, unlikely to be able to address issues that are widespread across industries. The requirement for such analysis is that robust and simple predictions are available from the theory, yet this is precisely the element that is lacking in the game theoretic literature on collusion and the business cycle.

III. Trends, cycles and the performance of large and small firms, 1972-1992

This section will set out at a fairly aggregate level the broad trends in the performance of large and small firms with regard to the economic variables with which we will be primarily concerned during this thesis. This exercise is useful because it sets the background to the empirical work of the subsequent chapters. It also enables a broad assessment of whether the trends in the relative performance of large and small firms fit with some of the key ideas mentioned in the previous discussion of theories of small firms (section II above). The section is predominantly data driven because it makes of the data sources at an aggregate level which can shed light on these issues. In particular, use is made of the size class analyses of the UK Census of Production for 1972-1992 and the size class analyses of the CBI industrial trends survey for the same period. The emphasis is not upon the econometric testing of specific hypotheses which is left to the subsequent chapters but to set out the main characteristics of the data. We will discuss *en passant* the issues of appropriate specification of variables, some of which will be used in the subsequent analyses, and the merits of different data sources. The section will consider three main areas of performance. Firstly, profitability, variously defined, and the use of material inputs, secondly, measures of the business cycle and economic optimism, and thirdly, the management and use of stocks and productive capacity.

---

22 However Sutton's argument presents significant problems for even these types of analysis.

As can be seen from Figure 4 there has been a steady rise in the aggregate price cost margins over the period. It is interesting to note that this trend has been remarkably uninterrupted for the smaller firms but has been more variable for the larger firms. For firms with less than 199 employees only in two years, 1974 and 1976,[23].have price cost margins fallen by more than a fraction of a percent. There are significant falls in price cost margins for the firms employing over 5000 employees in 1974,1975, 1980, 1990 1991 and 1992 all of which were recession years. Apart from these years price cost margins were either stable or rising. It is interesting to note that through the 1980-1982 recession price cost margins were rising for all firms at a fairly rapid rate. This is most likely to be a result of the "productivity miracle" of the early eighties reducing unit labour costs which are not fully passed on in price reductions. The implication is that while the recession was fairly severe, changes made to the economic environment in terms of industrial relations or removal or generation of excess capacity prevented competitive pressures from being heightened in this period. The behaviour of price cost margins in the late eighties and early nineties also poses some questions. For the small firms price cost margins remained static through the eighties boom even registering a slight fall at the peak of the boom in 1988. Their price cost margins rose through the 1990's recession however. The reverse is true of large firms for whom price cost margins display a much more traditionally procyclical pattern registering large rises in the boom and an equally sharp decline in the 90's recession. The net effect is that the aggregate price cost margin has been fairly static from 1987-1992.

An alternative measure of the price cost margin uses net output rather than gross output as the denominator in calculating the margin. This measure has been fairly widely used in empirical studies and the merits or otherwise of its use are examined more closely in Chapter 3. Figure 5 shows the movements in this measure over the period and comparison with Figure 4 demonstrates that they are clearly different in their behaviour.

[23] All price cost margins fell considerably in 1984, this is ascribed to the influx of very small firms in that year's Census rather than to economic factors.

Over the 1970's these price cost margins show a slightly more classical procyclical pattern with co-movements of the price cost margin measures for large and small firms while the downward movements in 1975 and 1980 are more pronounced for the larger firms. This is to be expected for more capital intensive firms and firms which act collusively if collusion is procyclical. For the 1980's the pattern of the alternative price cost margins roughly approximates those of the conventional price cost margins a sharp upwards rise which accelerates into the 1988 boom and falls thereafter aside from a slight upturn in 1992. For the small firm the picture is of broadly constant alternative price cost margins from 1985 onwards with no marked cyclical variation. In order to understand the differences between the two measures it is instructive to look at changes in the importance of material inputs into the production process.

Figure 6 illustrates that the level of bought in inputs as a proportion of output is higher for the larger firms throughout the period indicating a lower level of vertical integration on the part of those firms or alternatively a greater focus on the part of small firms of high value added activities. For all firms since 1977 there has been a steady decline in the use of bought in inputs as a proportion of output. In the case of the largest firms however the downward trend has been reversed since 1988 with the implied level of vertical integration falling as a result since the 1988 boom. For the smallest firms the trend towards increased vertical integration has increased.

In general, apart from the end of the period, the ratios for large and small firms move in tandem exhibiting almost identical co-movements over time. One possible explanation for this could be common price shocks, particularly in terms of fuel prices. There is a marked increase in the ratio in 1974 and a marked decline in 1986, both years with oil price shocks of the same direction. However the effects of the oil price rise in 1979/1980 is not evident as for those years the ratio is in steady decline. It could therefore be argued that Figure 6 actually masks the real situation because of the relative price changes between output and material inputs. We can control for these factors by using the producer price indices for materials and fuel and for output to deflate the nominal figure hence obtain figures in constant 1972 terms. Figure 7 indicates a very different story to Figure 6.

As is very clear from Figure 7 the relative price changes between materials and output dominates the picture. This would appear to indicate that firm are very responsive to relative price changes in forming the composition of inputs. The sharp decline in relative materials volumes resulting from the oil price shocks of 1973/4 and 1979/80 are both apparent as is a corresponding rise in material inputs from the fall in oil prices in 1986 and the decline in prices of raw materials in the early 1990's. The magnitude of these shocks means that it is difficult to discern whether there is any other trend in the real ratios. Through the early eighties and the late eighties when there were no major price shocks the ratios remained roughly constant. These figures would appear to show that production techniques are fairly flexible in order to make such rapid adjustments to relative price movements.

However, a word of caution is necessary, the magnitude of these adjustments could be construed as being unreasonably high, pointing more to data measurement problems than real changes. The data is consistent with a constant relative *value* of material inputs. Chapter 3 identifies firms operating "as if" the value of material inputs is regarded as fixed in the short run as one plausible explanation of the empirical data at the industry level on material input costs. The evidence at an aggregate level is also therefore consistent with this finding. A further difficulty that must be taken into account when comparing small firms against large is that the price series are based upon average prices. If small firms operate in different market segments to large firms the price movements may well be different particularly if there is preponderance of small (or large) firms who provide material inputs. In addition, small and large firms are likely to vary in their intensity of energy usage which means that the fuel component of the material input price index is likely to vary across firms. Unfortunately no data at this detailed level is available.

An important element that we wish to examine at this stage is the relative cyclical position of small and large firms. This we can do making use of the CBI Industrial Trends Survey which partitions the responses to its surveys not only by industry but also, at the aggregate level, by firm size class. To investigate the cyclical position we examine two aspects of the questionnaire:

the existence of below capacity operation and the degree of optimism about the future. Figure 8 displays the proportion of firms reporting below capacity operation in the largest and smallest size classes. Both series display a fairly conventional countercyclical time path as would be expected and there is no clear trend in either series. The proportion reporting excess capacity is high in the early 80's and early 90's while the trough occurs in the boom of the late 1980's. It is of interest to note that large firms appeared to suffer to a greater degree from excess capacity in the early part of the period, particularly the period 1975-1978. However by the mid 1980's the small and large firms were experiencing roughly similar levels of excess capacity and since 1988 the larger firms have been reporting lower levels of excess capacity than the small firms. The size of the trough in 1988 confirms the capacity constraint problem that arose in the late 1980's and indicates that it was most important for large firms.

There are a number of plausible explanations for the change in the position of large and small firms: firstly, the capital scrapping of the early 1980's which is likely to have affected the large firms to a greater extent: secondly, a change in the attitudes of firms to the scheduling of production in terms of reliance upon just-in-time production or inventory control to smooth output fluctuations: thirdly, a greater willingness on the part of large firms to use subcontracting and/or overtime to absorb output fluctuations rather than hold excess capacity. In respect of the final argument the changing power of trade unions may have had an effect on this decision. If firms are stronger relative to trade unions, the costs of overtime operation and/or the extra shifts necessary to meet peak demand may be lower making their use more attractive. Clearly, because we are dealing in broad trends, we cannot provide evidence to show causality but they are all documented trends that may reasonably be expected to contribute to the effect.

Figure 9 indicates the net proportion of firms in each size class that indicate that are optimistic concerning the next three months[24]. The co-movement between the series for the large and small firms is almost exact and both series follow the path of the cycle very closely. The troughs occur

[24] The net proportion is calculated by deducting the proportion reporting pessimism from those reporting optimism. A positive figure indicates on average a balance of optimists a negative figure indicates a balance of pessimists.

as expected in 1974/5, 1980 and 1990/1. This co-movement appears to indicate that there are no significant differences in the incidence of the business cycle across large and small firms. Neither group appears to lead or lag the other in perception of business conditions.

The further element that needs to be considered and has been alluded to in the previous discussions is the use by firms of inventories and how these change over the business cycle. We are able to examine this issue using information both from the CBI survey and the data on stocks and work in progress available in the Census of Production. As is considered in more detail in chapter 5 stocks may be held for both strategic reasons and to smooth either price or output responses to demand fluctuations.

Figure 10 presents the net proportion of firms experiencing an increase or decrease in stocks of finished goods over the past four months. For small firms there is very little change across the whole period. There is an decline in 1981 and a more sustained decline from 1990-1992 but for the rest of the period there is a balance[25]. For large firms there appears to be larger and more systematic changes. For most of the 1980's particularly, 1981-1983 and 1986-7 there are decreases in stocks that appear consistent with a structural change in terms of a decision to hold lower levels of inventory. This is again consistent with the arguments, given above in relation to excess capacity, concerning shifts to just-in-time production and lower costs of changing production. A further element of inventory reduction that relates to the decline in union strength is that stocks of finished goods partially protect firms from strikes, allowing firms to withstand strikes more easily hence raising their bargaining power, albeit at a cost. The reducing need for such devices with the reduction in the incidence of strikes may lead to lower levels of finished goods stocks.

Figures 11 refers to the same information but for stocks of material inputs. In this case there appears to be more pronounced changes over time and a greater degree of co-movement for large

[25] If changes in finished goods stocks are simply responses to unexpected, firm specific, rather than economy wide, shocks or errors in calculating demand we should expect a zero net balance over time.

and small firms. The movement is more conventionally cyclical, although apparently at a year's lag in recessions, with troughs in 1975, 1981 and 1991/2. It seems to be the case that stocks of material inputs follow movements in output quite closely, confirming the relatively constant materials to output ratio over the cycle indicated in Figure 6 above. The apparent lag in response to the cycle may be the result of relatively long term contracts for material inputs which would lead to a slow change in response to output conditions[26].

Finally we look at the ratios of stocks and work in progress to output give by the UK Census of Production. At this level, unfortunately, there is no distinction made in the data between stocks of finished goods and intermediate inputs. The levels are lower for small firms which may be contrasted with the higher materials to output ratio observed in Figure 6. The contrast may be explained as a result of shorter production periods leading to less work in progress being recorded for the smaller, less vertically integrated, firms and that this effect counteracts any tendency for a higher level of stocks of material inputs. The time path for small firms was relatively constant over the 1970's a slight rise at the start of the 1980's followed by a sustained fall through the rest of the period. For the large firms the changes from year to year are greater but do not appear to be systematic. The major fall occurs in 1987 to a level sustained for the rest of the period. We are not aware of any reason for a change in this particular year.

Table 3 presents a correlation matrix for all of the variables taken from the CBI Industrial trends survey. As is to be expected in a sample of variables chosen largely for their evidence of cyclical conditions most of the variables are reasonably highly correlated. The correlations between the observations for large and small firms are .7266 for excess capacity, .9578 for optimism, .5367 for finished goods stocks and .8940 for stocks of material goods. This confirms that, at the aggregate level and with the exceptions that we have discussed in detail above, large and small firms tend to have broadly similar responses to the business cycle on a variety of indicators.

---

[26] Machin and Van Reenen (1993) provide evidence to support this showing the incidence of the cyclical trough to be in 1980 for consumer goods producers and 1981 for producer goods suppliers, many of which will be supplying intermediate material inputs.

We have therefore established in this chapter that there are substantial co-movements in the economic variables relating to large and small firms. There is relatively little evidence that the small firm sector experiences greater volatility of the key variables in response to short run changes in demand. Indeed the large firm sector tends to experience greater degrees of variability in most of the cyclical variables that we have observed. Of course, this may in itself be indicative of one hypotheses that will be considered in chapter 2, that small firms are better at responding to hence smoothing the effects of demand fluctuations;. We have also noted that, in theory, market power may account for different degrees of response to demand fluctuations. The reasons for the relative responses of large and small firms therefore remain an open question, and the attempt to find an answer to this question forms the basis of the investigation in Chapter 2.

Table 1: Relative Importance of Large and Small Enterprises in UK Manufacturing 1972-1992

	Employee Size Class			
	1-199	200-499	500-4999	5000+
	% Share of Total Employment			
1972	21.5	8.0	25.6	44.9
1979	23.1	7.3	24.9	44.7
1983	27.8	8.6	25.7	37.8
1984	30.3	9.6	26.6	33.5
1992	34.4	9.7	28.5	27.4
	% Share of Net Output			
1972	18.4	7.4	25.9	48.4
1979	19.5	6.7	26.7	47.1
1983	22.8	8.1	27.2	41.9
1984	24.7	9.1	28.9	37.4
1992	27.0	9.1	31.7	32.2
	No of Enterprises			
1972	68971	1823	1321	186
1979	86838	1543	1186	174
1983	83482	1361	957	117
1984	116576	1514	978	104
1992	128516	1364	963	93

Notes:

1. Source of data: calculations based upon information from UK Census of Production Table 12 various years. Figures for net output and employment are percentages of the total for manufacturing industries.
2. A change of definition altered the composition of firms between 1983 and 1984 so the data before and after this date are not strictly comparable. In addition a revision of industrial classification effective from 1980 means there are slight differences in the definition of manufacturing. At the current level of aggregation the latter change is of relatively little significance. See text for further comment upon these issues.
3. Totals may not sum to 100 due to rounding.

Table 2: Establishment Size and Number of Establishments in UK Manufacturing 1972-1992

	Employee Size Class (Enterprise)				
	1-199	200-499	500-4999	5000+	All
	Mean Establishment Size (no. employees)				
1972	21	207	346	760	82
1979	16	155	237	556	61
1983	15	141	211	419	48
1984	12	147	236	456	36
1992	11	167	276	554	31
	Mean No. of Establishments per Firm				
1972	1.08	1.50	3.98	22.6	1.20
1979	1.06	1.99	5.74	30.1	1.19
1983	1.06	2.19	6.19	37.5	1.19
1984	1.05	2.07	5.55	34.1	1.12
1992	1.02	1.84	4.65	23.1	1.07

Notes:

1. Source of data, UK Census of Production: Table 12, various years.

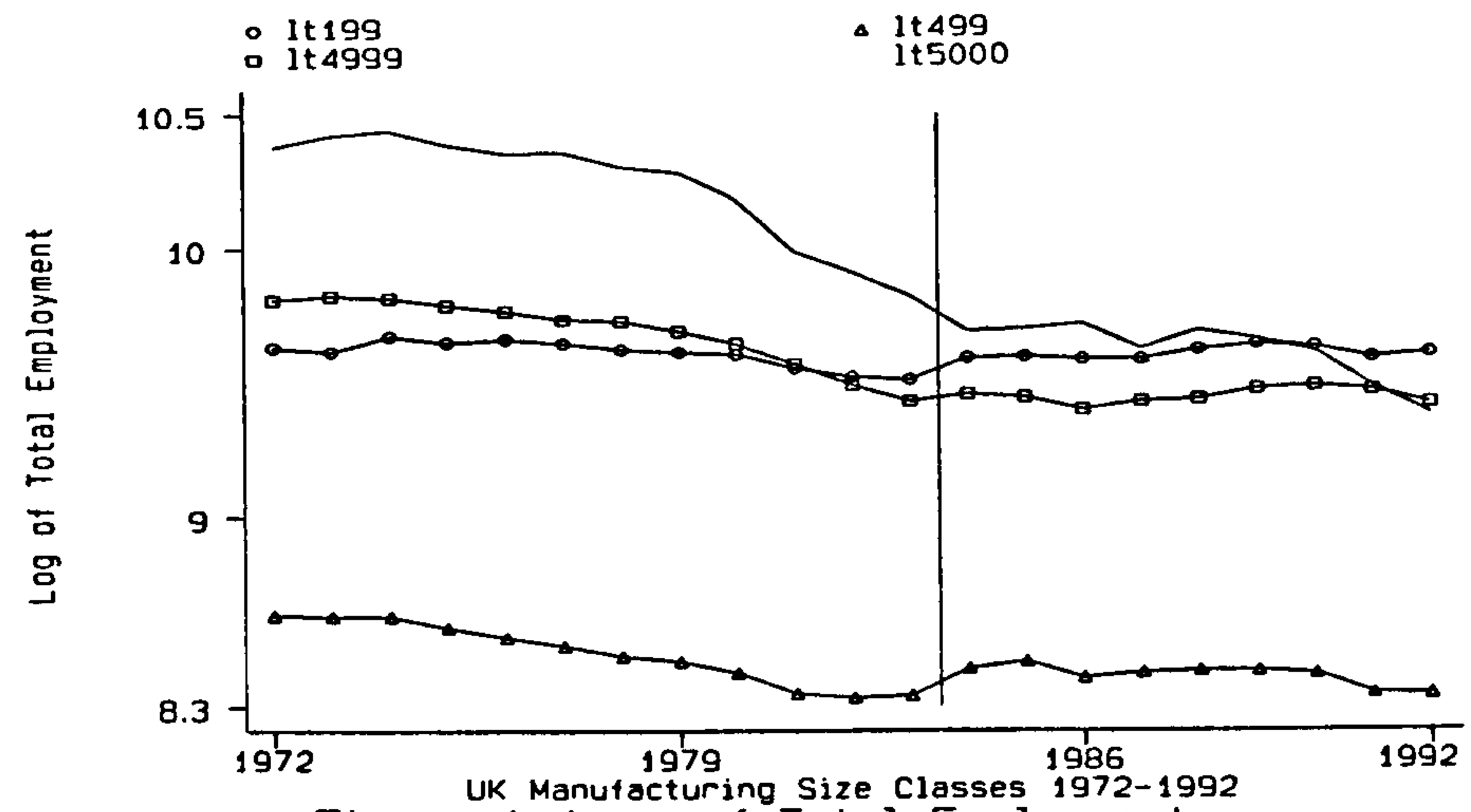


Figure 1: Logs of Total Employment

Notes: lt199, lt499, lt4999, lt5000 are the log of total employment in firms with 1-199 employees, 200-499 employees, 500-4999 employees, 5000+ employees respectively. The vertical line indicates the position of the discontinuity in sampling from pre 1983 and post 1984. The changes between 1983 and 1984 should not be interpreted as having any economic significance.

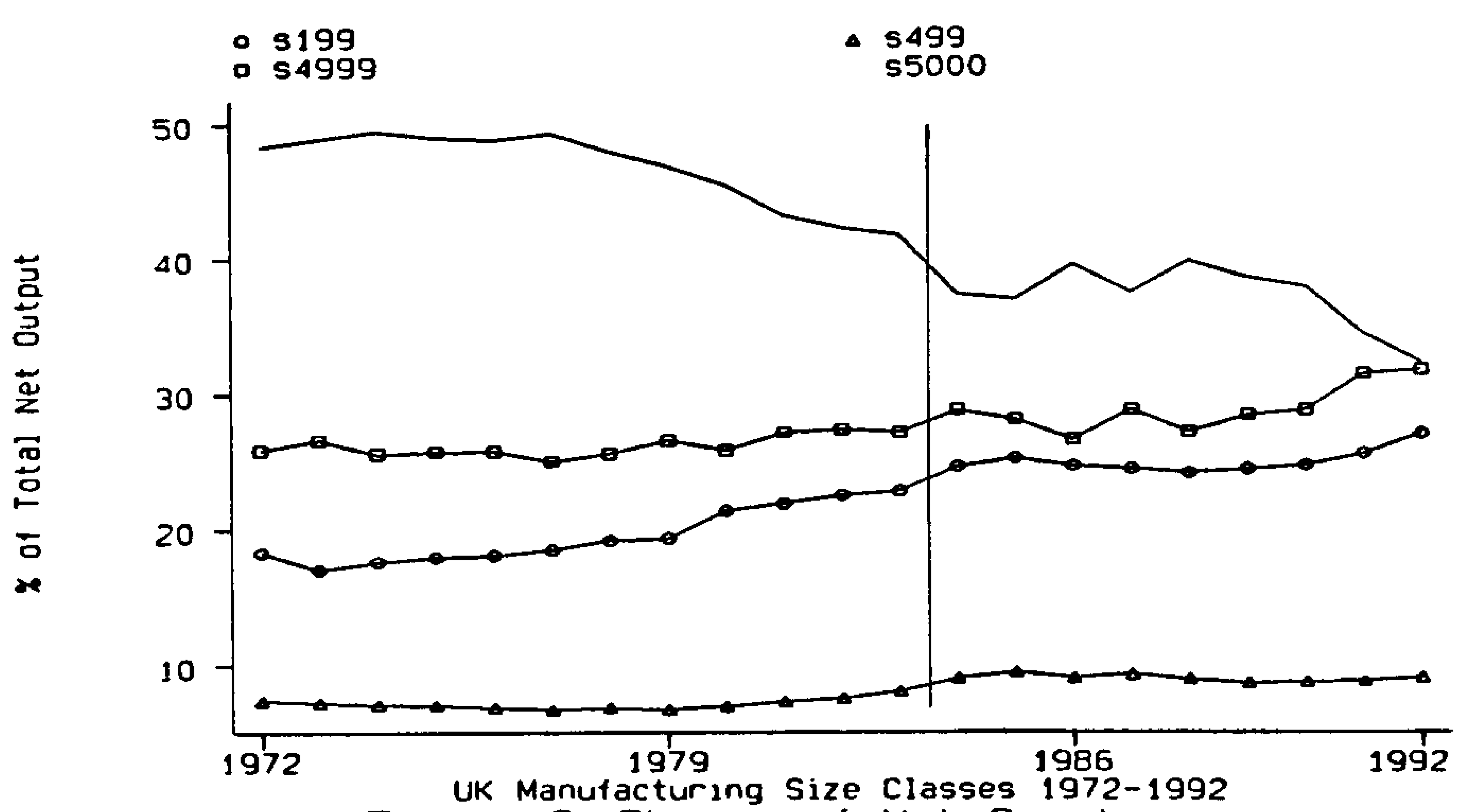


Figure 2: Shares of Net Ouput

Note: s199, s499, s4999, s5000 are the shares of net output for firms with 1-199 employees, 200-499 employees, 500-4999 employees, 5000+ employees respectively. The vertical line indicates the position of the discontinuity in sampling from pre 1983 and post 1984. The changes between 1983 and 1984 should not be interpreted as having any economic significance.

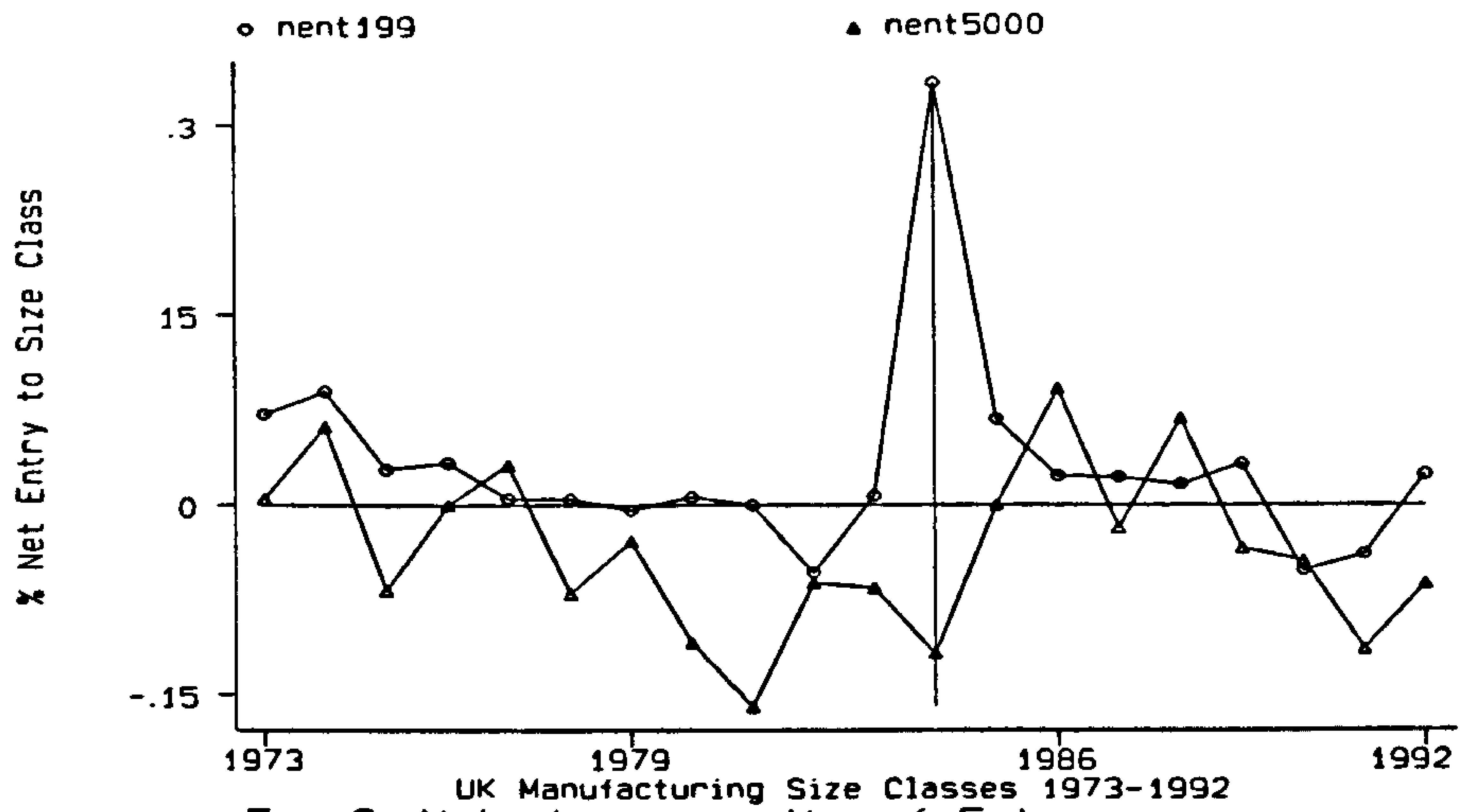


Fig 3: Net change in No. of Enterprises

Note: nent199, nent5000 are the proportionate change in the number of firms in the size classes of firms with 1-199 employees and 5000+ employees respectively. The vertical line indicates the position of the discontinuity in sampling from pre 1983 and post 1984. The changes between 1983 and 1984 should not be interpreted as having any economic significance.

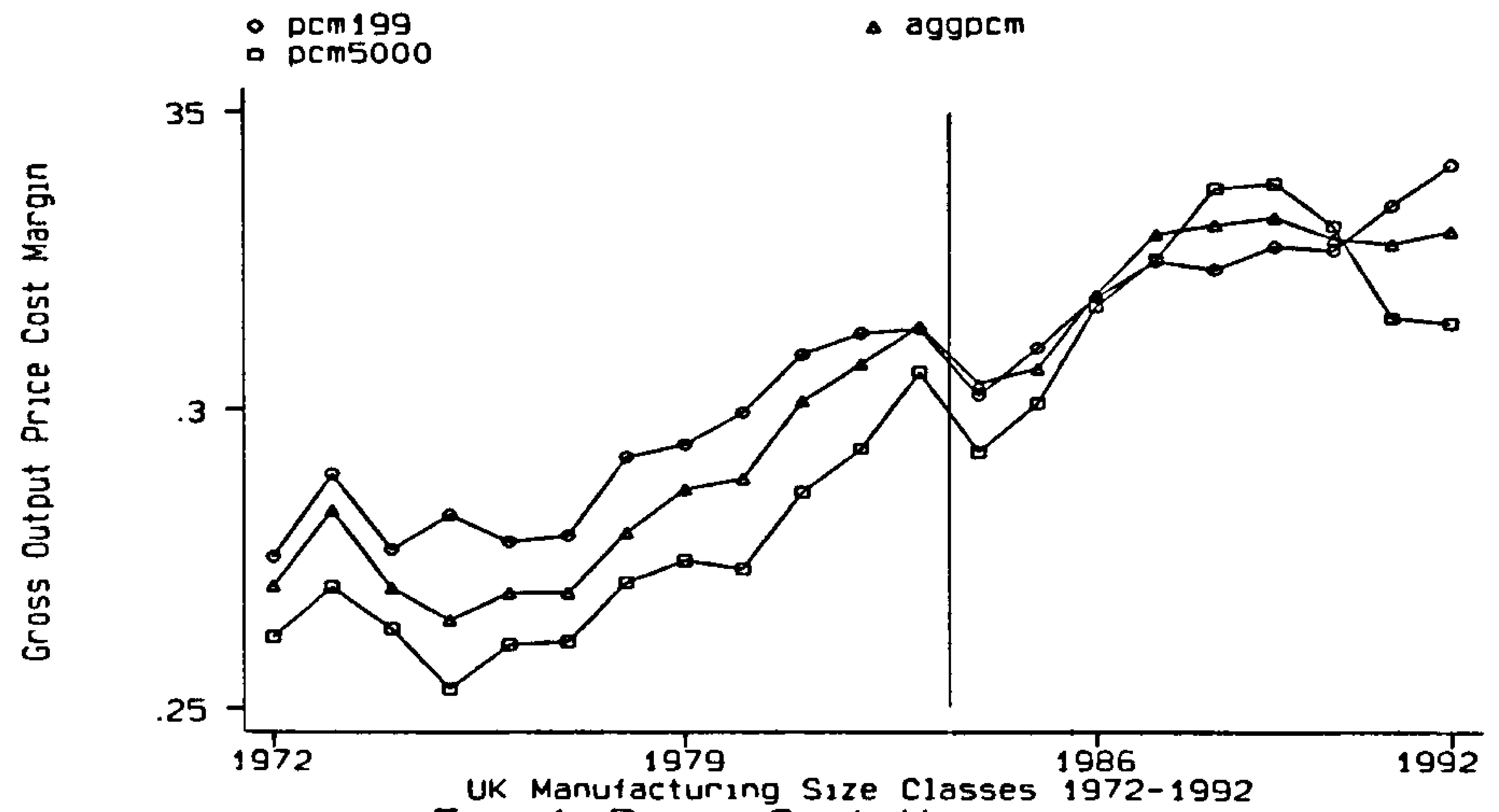


Fig 4: Price Cost Margins

Note: pcm199, pcm5000 and aggpcm are the price cost margins of firms with 1-199 employees, 5000+ employees and all firms in UK manufacturing respectively. The price cost margin is defined as (net output - operative wages)/(gross output). The vertical line indicates the position of the discontinuity in sampling from pre 1983 and post 1984. The changes between 1983 and 1984 should not be interpreted as having any economic significance.

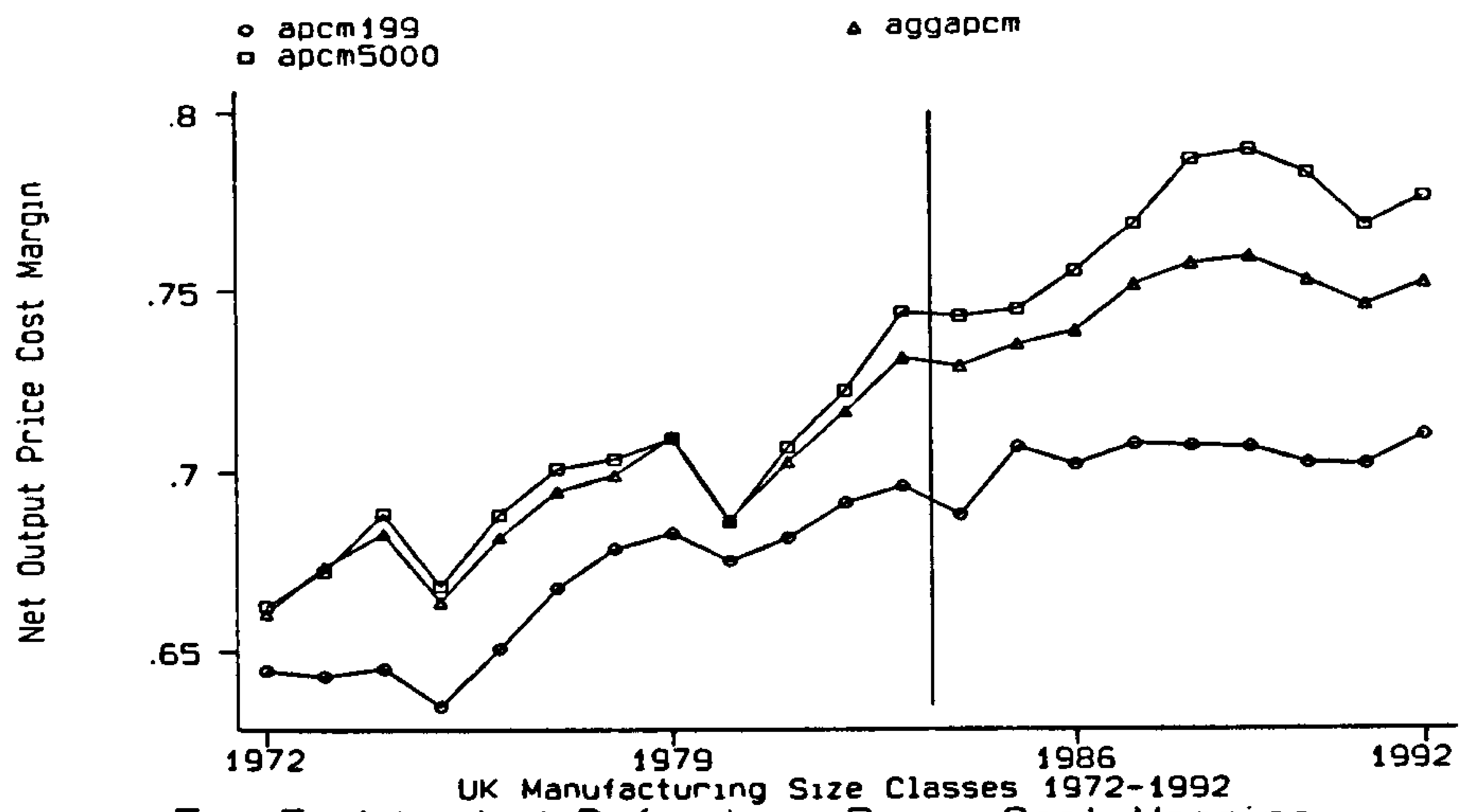


Fig 5: Adjusted Definition Price Cost Margins

Note: apcml99, apcm5000 and aggapcm are the price cost margins of firms with 1-199 employees, 5000+ employees and all firms in UK manufacturing respectively. The price cost margin is defined as (net output - operative wages)/(net output). The vertical line indicates the position of the discontinuity in sampling from pre 1983 and post 1984. The changes between 1983 and 1984 should not be interpreted as having any economic significance.

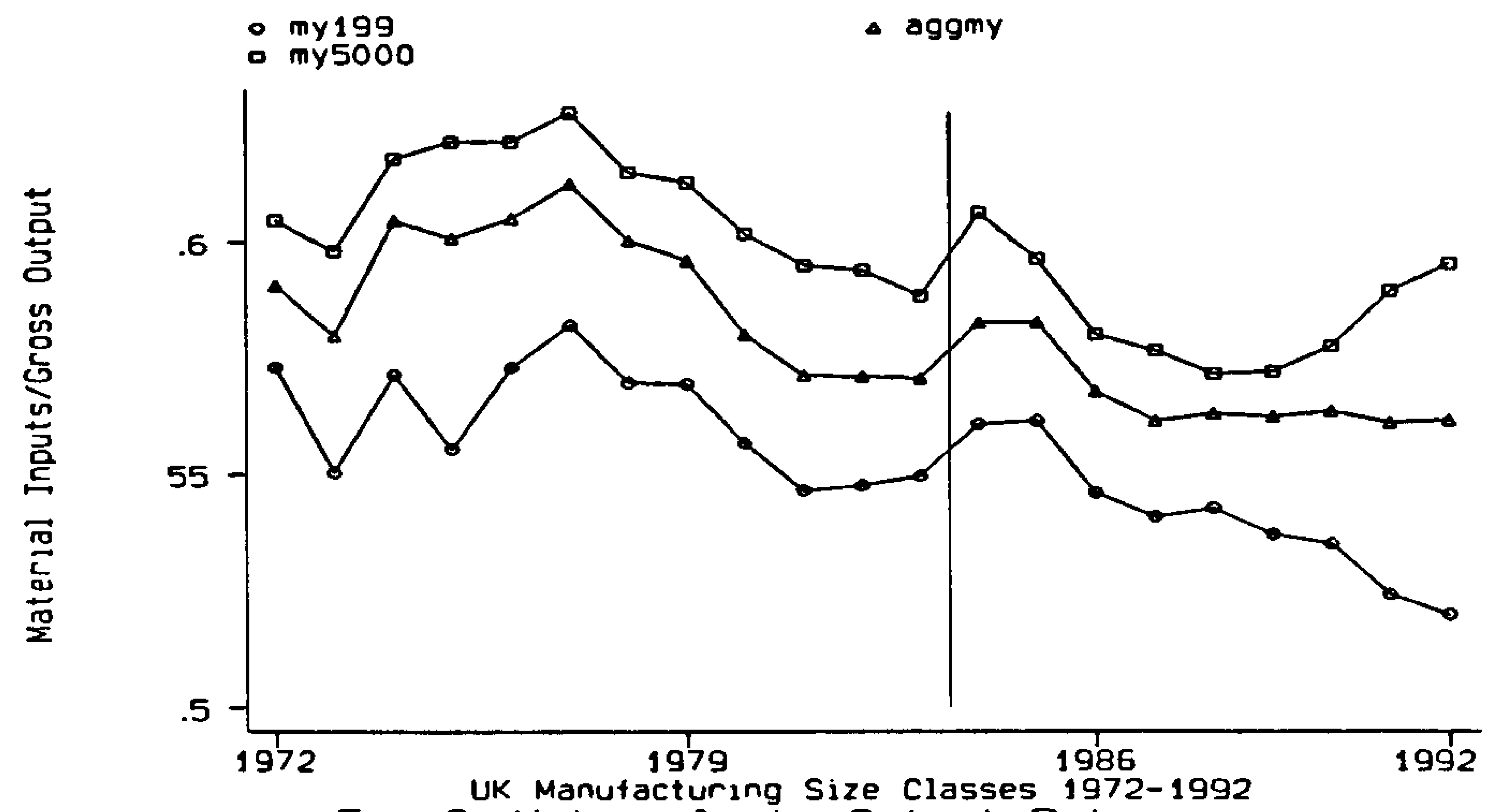


Fig 6: Materials to Output Ratios

Note: my199, my5000 and aggmy are the material inputs to gross output ratio of firms with 1-199 employees, 5000+ employees and all firms in UK manufacturing respectively. The material inputs to gross output ratio is proxied by as (gross output - net output)/(gross output). The vertical line indicates the position of the discontinuity in sampling from pre 1983 and post 1984. The changes between 1983 and 1984 should not be interpreted as having any economic significance.

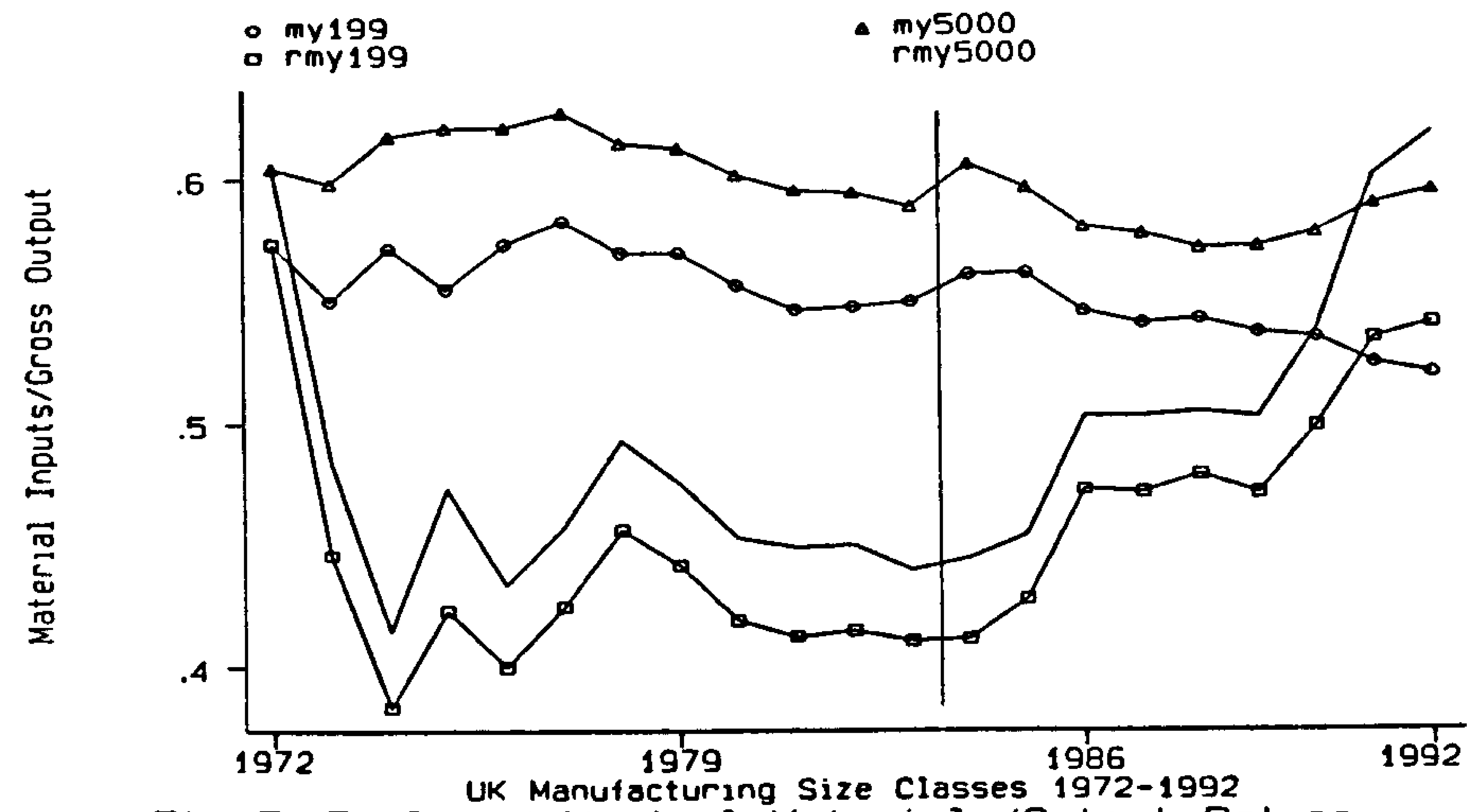


Fig 7: Real and Nominal Materials/Output Ratios

Note: The materials to output ratios in constant 1972 terms are rmy199 for the firms with 1-199 employees and rmy5000 for the firms with 5000+ employees. my199 and my5000 are the nominal counterparts as defined in the note to Figure 6. The vertical line indicates the position of the discontinuity in sampling from pre 1983 and post 1984. The changes between 1983 and 1984 should not be interpreted as having any economic significance.

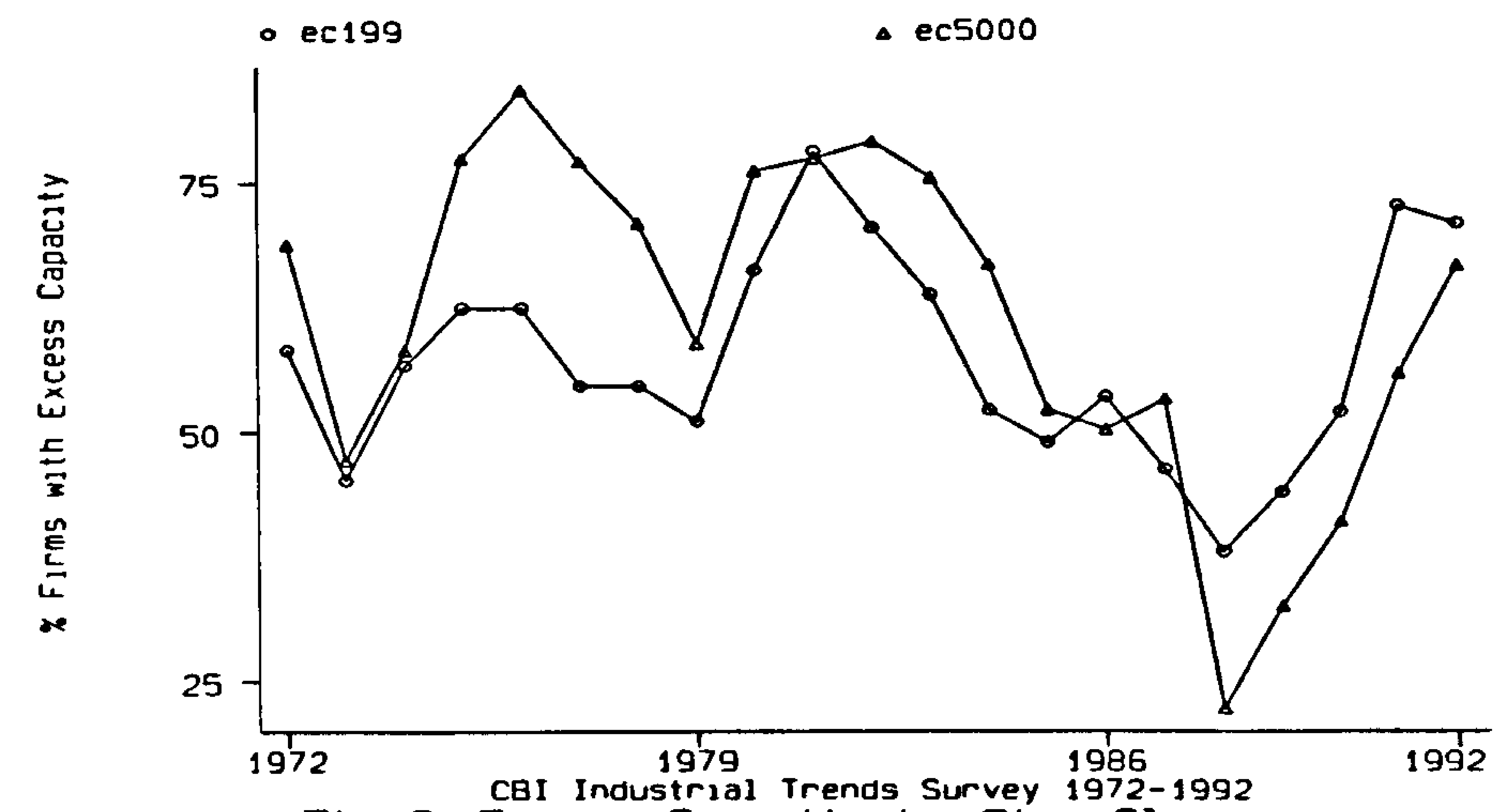


Fig 8: Excess Capacity by Size Class

Note: ec199 and ec5000 are the proportion of firms reporting below capacity operation in the 1-199 employee and 5000+ employee size classes. See Appendix for details of data definitions.

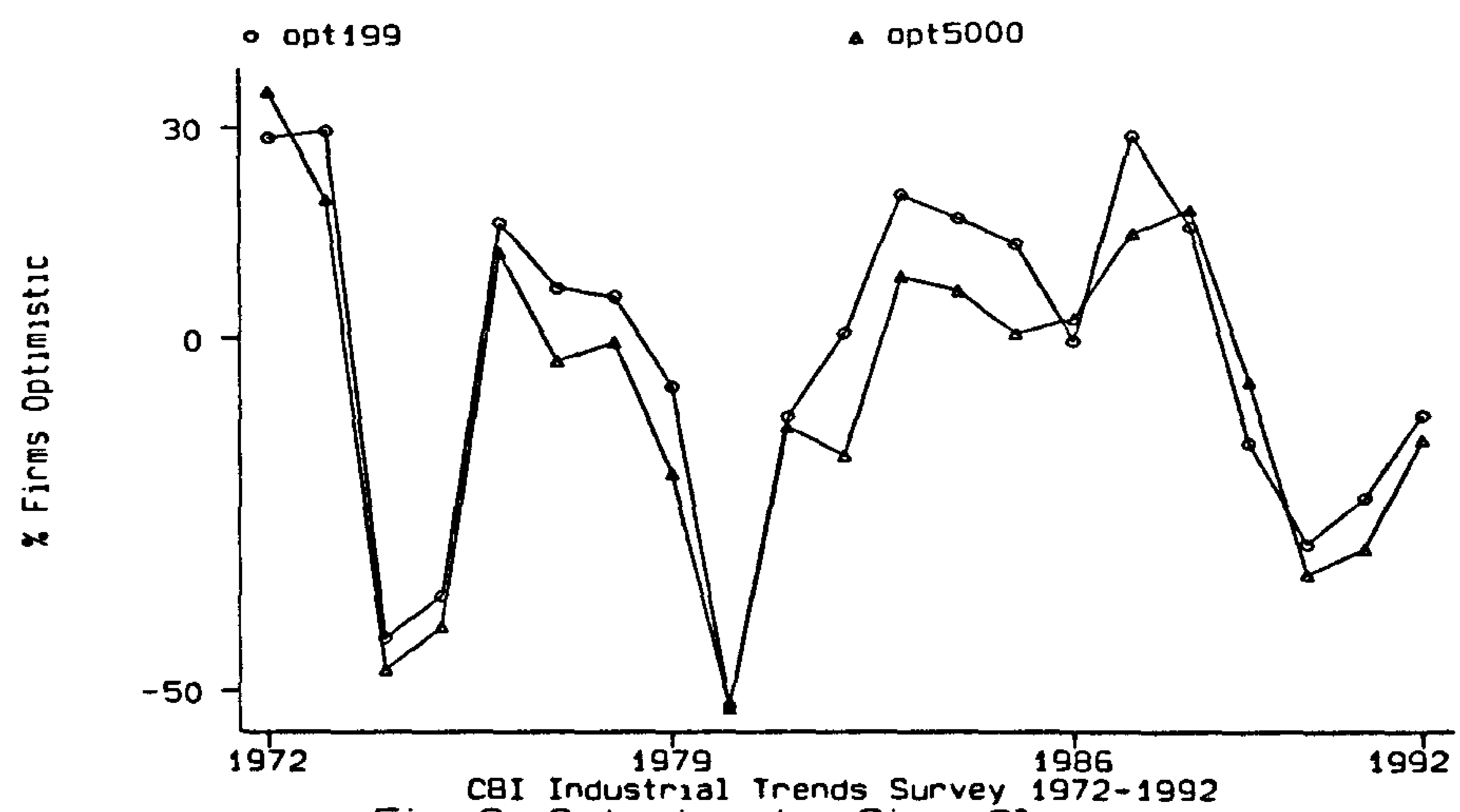


Fig 9: Optimism by Size Class

Note: opt199 and opt5000 are the proportion of firms reporting optimism in the 1-199 employee and 5000+ employee size classes. See Appendix for details of data definition.

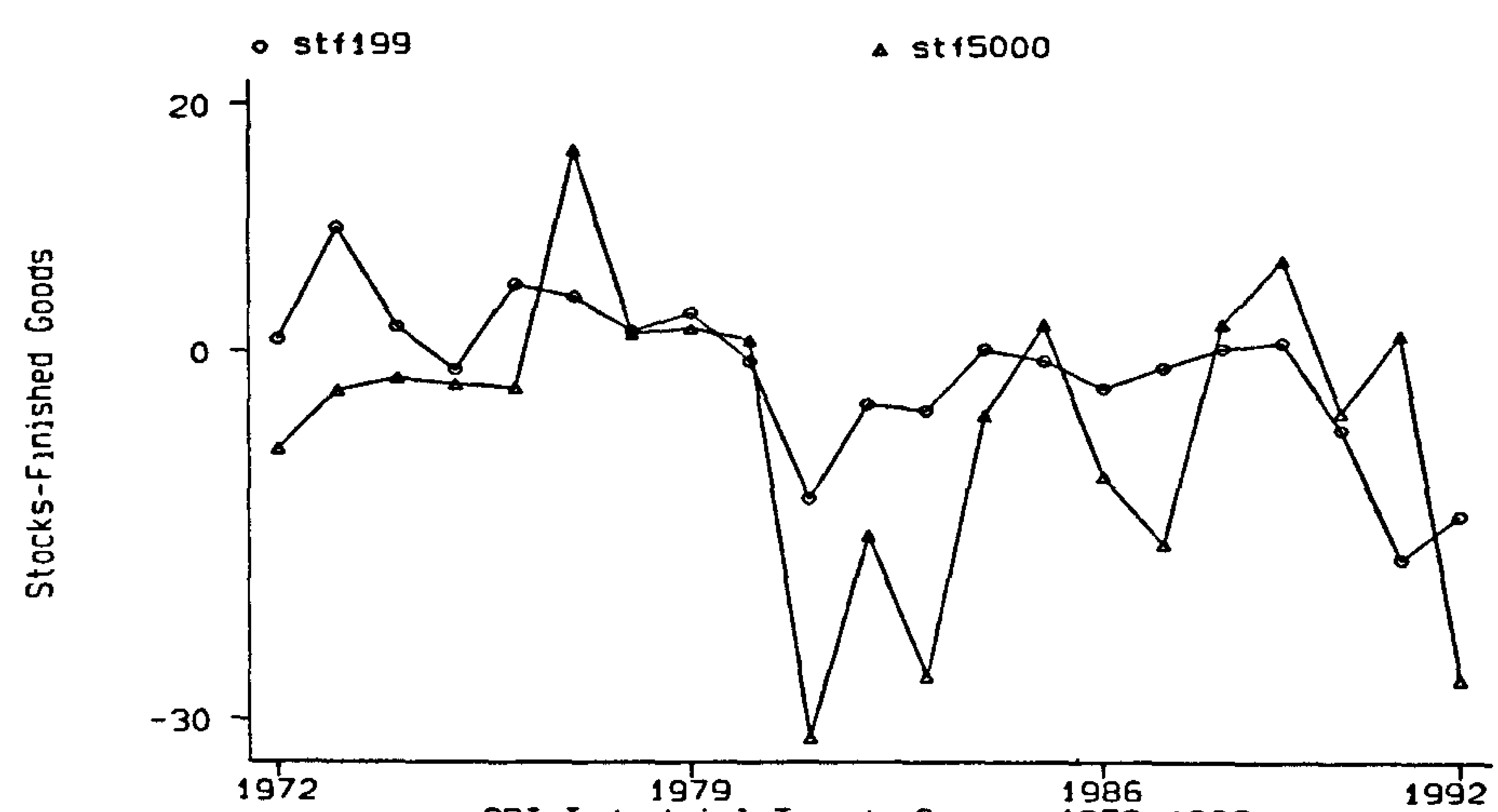


Fig 10: Expected Change in Stocks by Size Class

Note: stf199 and stf5000 are the net proportion of firms reporting an increase in stocks of finished goods in the 1-199 employee and 5000+ employee size classes. See Appendix for details of data definition.

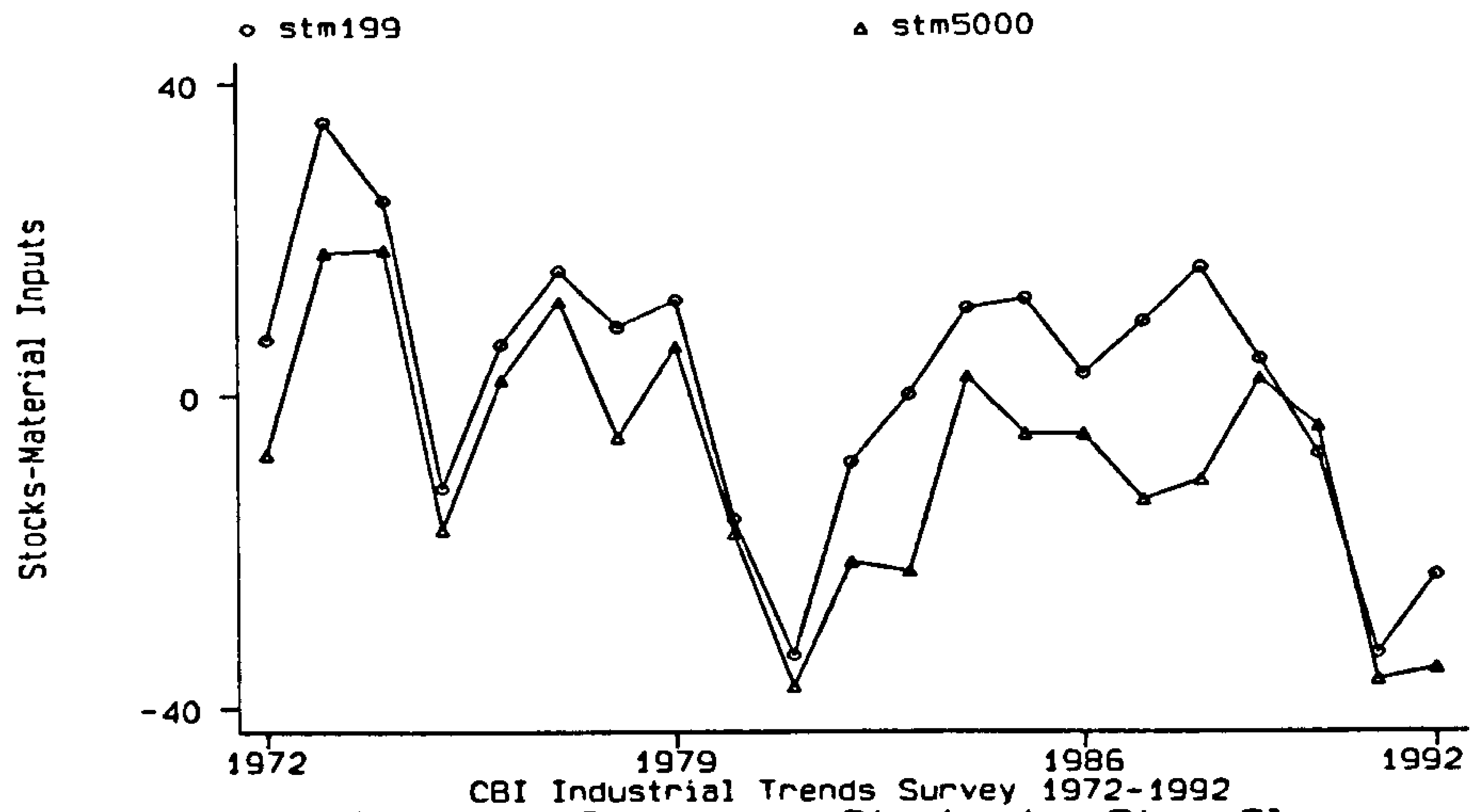


Fig 11: Expected Change in Stocks by Size Class

Note: stm199 and stm5000 are the net proportion of firms reporting an increase in stocks of material inputs in the 1-199 employee and 5000+ employee size classes. See Appendix for details of data definition.

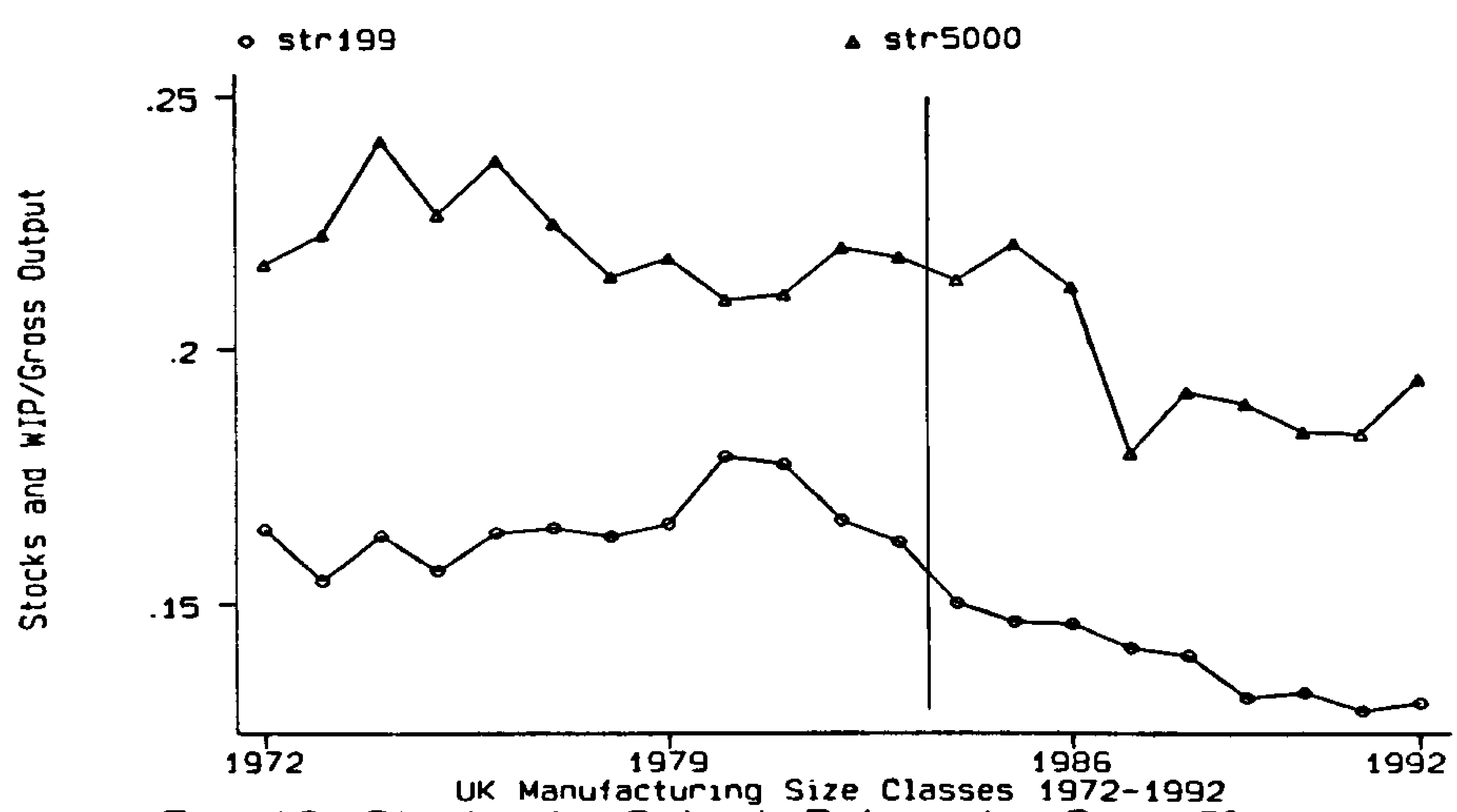


Fig 12: Stocks to Output Ratios by Size Class

Note: str199 and str5000 are the ratio of stocks and work in progress to gross output in the 1-199 employee and 5000+ employee size classes. The vertical line indicates the position of the discontinuity in sampling from pre 1983 and post 1984. The changes between 1983 and 1984 should not be interpreted as having any economic significance.

Table 3: Correlation Matrix of CBI Industrial Trends Survey Variables

	ec199	ec5000	opt199	opt5000	stf199	stf5000	stm199	stm5000
ec199	1.0000							
ec5000	0.7266	1.0000						
opt199	-0.3766	-0.0635	1.0000					
opt5000	-0.4163	-0.1615	0.9578	1.0000				
stf199	-0.6476	-0.0557	0.3485	0.3443	1.0000			
stf5000	-0.5305	-0.2855	-0.1325	-0.1065	0.5367	1.0000		
stm199	-0.7963	-0.3191	0.4773	0.4511	0.8807	0.4587	1.0000	
stm5000	-0.6895	-0.2407	0.1780	0.1738	0.8714	0.6220	0.8940	1.0000

Variable definitions: 199 refers to the firms with between 1-199 employees, 5000 refers to the firms with greater than 5000 employees. ec is the proportion of firms reporting below capacity operation. opt is the net proportion of firms reporting that they are more or less optimistic about the future. Positive values indicate a majority of optimists, negative values indicate a majority of pessimists. stf is the net proportion of firms reporting an increase/decrease in stocks of finished goods. stm is the net proportion of firms reporting an increase/decrease in stocks of material inputs

Data Appendix

The data series from the CBI quarterly Industrial Trends Survey are based upon the following questions

Question 1.
"Are you more, or less, optimistic than you were four months ago about THE GENERAL BUSINESS SITUATION IN YOUR INDUSTRY?" to which firms can respond "More" "Same" "Less". The calculated figure for optimism is the proportion responding "More" minus the proportion reporting "Less"
Question 4.
"Is your present level of output below capacity?(i.e. are you operating below a satisfactory level of operation)"
To which firms can respond "Yes", "No" or "N/A" . The calculated figure for excess capacity is the proportion responding "Yes".
Question 5c.
"Excluding seasonal variations do you consider that in volume terms your present stocks of finished goods are " "More than adequate" "Adequate" "Less than adequate" "N/A"
The measure of excessive inventory is calculated as the difference between the proportion of firms reporting "More than adequate" and the proportion reporting "Less than adequate". A positive figure indicates on balance excessive stocks of finished goods, a negative figure indicates on balance insufficient stocks.
Questions 10 a and c
"Excluding seasonal variations, what has been the trend over the PAST FOUR MONTHS with regard to:"
Q10 a Volume of stocks of raw materials and bought in inputs
Q10 c Volume of stocks of finished goods
The measure of the change in stocks is the proportion reporting "Increase" minus the proportion reporting "Decrease". A positive figure indicates on average an upward trend over the past four months, a negative figure indicates on average a downward trend.

In each case the annual figure of the series is calculated as the mean of the 4 quarterly figures.

Chapter 2. Demand Fluctuations, Firm Heterogeneity and Stiglerian Flexibility

I. Introduction

Mills and Schumann (henceforth MS)(1985) set out an explanation of and empirical evidence for the coexistence of different sized firms within a competitive industry. Their important empirical finding is that there is an inverse relationship between measured output variability and firm size. It is also found that this relationship holds regardless of the degree of concentration in the industry and this is offered as evidence against a market power based explanation for their results. Their result is offered as support for the theory of a trade-off between static and dynamic efficiency as an explanation for the coexistence of large and small firms in the same industry, we will call this trade-off the flexibility hypothesis. This basic empirical result has subsequently been confirmed by Chappell, Das and Shughart (1993) and Zimmermann (1995) although Zimmermann reports some contradictory evidence which requires further explanation and investigation as well as the confirmation of this basic result.

In terms of the focus of this thesis there is also a natural progression from a consideration of cyclical movements in the relative profitability of small and large firms to consider the variation in sales (hence market shares) of firms. A direct connection between the two issues has been drawn through comparisons between the empirical study of Schmalensee (1989b) which examines the cyclical relationship between profitability and firm size *within* industries and the above mentioned studies of the relationship between firm size *within* industries and sales variability. The connection between the results of these studies has been commented upon in Schmalensee (1989b, p356) and Scherer and Ross (1990, p442), who note that Schmalensee's result that the relative profitability advantage of large firms is counter cyclical, i.e. the profitability advantage increases in recessions, is consistent with the finding of Mills and Schumann (1985) that large firms have lower sales variability. Scherer and Ross further suggest that Schmalensee's result is "inconsistent with arguments that large firms are likely to be more burdened by fixed costs in recessions." (ibid. p442). However we would argue that, while the Mills and Schumann and Schmalensee results would appear to be consistent, the underlying explanations for the result are very different. In particular, Mills and Schumann adopt a framework where they assume perfect competition whereas Schmalensee concludes that one explanation for the results obtained in his study, Schmalensee's apparent preferred explanation, is variations in market power and the study is based in a theoretical framework of imperfect competition. We will examine these two results in this chapter and we find that while the consistency exists, strikingly different conclusions must be drawn.

There are three objectives to this chapter. Firstly the chapter seeks to explore the MS empirical result, which would appear to be of wider interest, and question both the theoretical foundations upon which it is based and the interpretation which can be placed upon the results. In particular, with the generalisation of the model to the situation of imperfect competition, it is possible firstly to generate an alternative empirical result while retaining key features of the model and secondly to maintain the original result without the key features of the MS model. The problem of observational equivalence is therefore potentially severe, yet it is a problem that has been largely ignored in the previous discussions. Secondly the chapter seeks to move forward the empirical analysis of these issues by advocating an alternative strategy for empirical analysis of flexibility arguments which it is argued is more closely grounded in appropriate theory and will be potentially more revealing. Thirdly the chapter will consider more closely the relationship between fixed costs/capital structure, sales variability and profitability movements to examine the true degree of consistency between the empirical findings.

The structure of the chapter is therefore as follows. Section II considers the initial conceptualisation of flexibility put forward by Stigler (1939) and the subsequent refinements of the concept discussed in particular by Carlsson (1989) and Del Monte and Esposito (1992). The section then considers other elements of industrial organisation which have largely been ignored by the literature on flexibility but which have a direct bearing on the central empirical issue of this chapter. Section III is an exposition of the theory and assumptions from which MS generate their empirical prediction. Section IV develops the extension of the model to a situation of imperfect competition. Within this alternative framework the main results of the MS framework are reproduced but exceptions are also generated. Section V assesses the empirical specification adopted by MS and others and the implications for the extension for the testing and interpretation of the hypothesis are given. Section VI considers the theoretical determination of the variability of price cost margins an aspect not directly considered by MS but which is related, as noted above, and is an important element of this thesis. Section VII summarises the findings and implications particularly for empirical analysis.

II. Stiglerian and alternative conceptions of flexibility

Stigler's notion of flexibility is derived from his seminal work of 1939. This work has been much referred to but, as with so many seminal articles, the distilled essence of the paper, that which has been appropriated by subsequent theory, has lost much of the original interesting flavour. There are two aspects to Stigler's conception of the use of capital equipment in the short run. The adaptability of the equipment and its divisibility. Oi (1982) identifies the adaptability of labour as an additional element of flexibility. In general, divisibility of fixed plant will enable marginal costs to be maintained at a relatively constant level until the optimum or capacity level of output is reached. Adaptability of the plant will enable marginal costs to be maintained relatively constant in the region in excess of the optimum output level. The degree of flexibility is then a combination of the two. Stigler suggests two strategies that could be adopted by firms to increase the degree of flexibility, the first is to ensure that plant is divisible in order to reduce the cost penalty of sub-optimal output levels, the second is to reduce fixed costs relative to variable costs by a process of substitution of variable for fixed costs[1].

Carlsson surveys the previous literature and suggests three possible levels of flexibility for the firm; operational, tactical and strategic. Operational flexibility relates to flexibility in the day to day organisation of production. It includes that ability to switch between products, vary the length of a production run, and adapt to machinery breakdown etc. Operational flexibility is also

[1] It is interesting to note in passing that Stigler is not rosy eyed with respect to his view of flexibility. He recognises the efficiency and competitive benefits of flexibility, but also notes that an important feature of the second route to flexibility is that it is employed in "the sweatshop industries".

deemed to be one of the determinants of the requirements for inventories, shorter production runs in principle requiring lower levels of inventory and increasing operational flexibility. Tactical flexibility refers to the design of production facilities so that changes in the products mix or level of demand can be met in the medium term. This form of flexibility is probably the closest to the form of flexibility which is described by Stigler and discussed in most of the literature. It is largely concerned with designing a production facility with an appropriate cost structure in the sense intended by Stigler. However the correlation between Carlsson's notions of flexibility and Stigler's is far from exact. To note but one difference, elements of Carlsson's tactical flexibility depends upon the existence of multiple products. Carlsson's notion is more closely related to the economic notion of economies of scope in the short run, rather than short run economies of scale. In this situation of tactical flexibility, there is a potential trade-off noted between inventories and flexibility. What is described as a "functional plant layout", where it is possible to shift between production of one good to another relatively easily and thereby vary the production of individual products at the same time as maintaining high levels of capacity utilisation, leads to higher overall levels of inventory as throughput of semi processed stock is slower. Strategic flexibility relates to the ability to be flexible in the attitude towards strategic decision making. This means that the response to new product developments and design changes must be adequate. It also requires a flexibility of organisational form in terms of information processing so that all parts of a firm may be aware of the new developments in other areas and a strategic overview can be taken.

Alongside these three levels the usefulness of a different partition of flexibility is noted by Carlsson, what Klein(1984) has described as type I and type II flexibility. Type I flexibility refers to the ability of firms to respond to foreseeable outcomes, such as a particular realisation of the level of demand from a predictable range, type II flexibility refers to the ability of firms to respond to unforeseeable outcomes, such as the advent of a radically different technology or product. There is no exact concordance with Carlsson's schematisation, as Carlsson notes, however type II flexibility would largely come within the scope of strategic flexibility and would only be seen in terms of tactical and operational flexibility in relation to the ability of primarily staff and to a lesser extent machines, to trouble shoot without reference to specialist help. Type I flexibility would appear to accord with the general features of operational and tactical flexibility in terms of design of production to deal with the anticipated fluctuations in the medium and short run. In terms of observation, Carlsson notes that aspects of strategic flexibility are very difficult to measure, and the same applies to type II flexibility.

An important aspect of tactical flexibility is the ability to operate at different output levels without significant cost penalties and this, therefore, accords with Stigler's definition. It is also the empirical aspect which has been focused upon within the existing literature. However the additional aspect relevant to empirical analysis which is implicit in the concept of operational flexibility is the notion of output adjustment costs. In part, adjustment costs could come under tactical flexibility but both analytically and empirically the point is different. A firm may be able to operate at different levels of output at the same per unit cost but changing between the levels of output cause the firm to incur significant costs. This necessarily implies a dynamic cost function which is not easily comparable with the static cost functions analysed below. This is a significant weakness of the empirical application of the flexibility approach. To give an example: suppose that a company is faced with an order of either 50 units per month or 100 units per month. A flexible firm in terms of the definition of Mills and Schumann would be able to produce either 50 or 100 at a similar cost. But the real advantage would be if the order is for 50 units in odd months then 100 units in even months. If there are costs of adjusting the level of output the flexible firm is one which can switch between the levels of output easily.

To incorporate the above argument we need a dynamic cost function which contains an allowance for adjustment costs. This point was included in Stigler's consideration but has been lost from subsequent discussions. Stigler notes "The situation in real life is complicated by the fact that the optimum output may be an optimum through time, i.e., the optimum may be based upon anticipated increases or decreases in the output of the firm."(Stigler, 1939, p308) Thus he restricts his theoretical analysis to the static or stationary case which must implicitly assume zero adjustment costs or alternatively subsumes adjustment costs into the definition of adaptability. More precisely, Stigler comes to "the conclusion that time must be an implicit variable which affects the form of the production function. There is not a short run and a long run; rather there are continuous variations in the marginal cost curve from very short periods to full, long run equilibrium."(ibid. p311) The theoretical framework developed by Mills, following from the previous analyses, has proved attractive for its simplicity; it is able to tackle an apparently dynamic issue using a framework which is essentially static in origin. How far this simplification can and should be taken is clearly open to question.

The other explicit acknowledgement made by Stigler is that a one to one plant firm ratio is assumed i.e. there are no multi plant firms, a factor which must be considered when returning from theory to empirical analysis. The existence of multi plant firms clearly implies that there is greater divisibility of output and therefore will lead to greater flexibility. Multi product firms need also to be taken into consideration when considering flexibility. Carlsson assumes for most of his discussion that firms are multi product firms although Stigler's and Mills and Schumann's analysis refers to only monoproduct firms. Del Monte and Esposito (1992) examine the Mills framework but extend it to multiproduct firms and reverse the conventional theoretical finding suggesting that for multiproduct firms large firms will be more flexible. They provide no empirical support for this conclusion, however, and Mills and Schumann do attempt to control for diversification in their empirical analysis but find no evidence of a relationship.

From a different perspective, noted but not considered in detail by Zimmermann (1995), it can be viewed that the small and large firms are producing differentiated products in which case this must be taken into account when considering the intra industry relations. This aspect is difficult to analyse empirically because of the detailed information necessary to examine the relationships. However it is possible to speculate as to the consequences of this. Firstly the relationship may potentially differ between consumer and producer goods industries if consumer goods industries are more susceptible to such segmentation. If small firms are producing different products from the large firms then it can be of two main forms; vertically related products which are potentially supplied to larger companies; or horizontally related products which do not necessarily compete head on but instead use niche markets for their products. If we consider the latter initially, as

Sutton (1991) shows, with differentiated products industries investment in sunk costs particularly in advertising and research and development can lead to the acquisition of large market shares while accompanied by small firms with little investment in sunk costs and smaller market shares. It does not stretch the imagination too greatly to suggest that the nature of advertising will not only tend to increase the demand for a product but will also tend to reduce the variability of demand for the product. Advertisers are keen on generating brand loyalty which would tend to smooth out demand fluctuations as consumers become less responsive to price changes. By contrast for smaller firms niche markets are likely to be more volatile both because tastes will change more quickly and be less manageable and because cyclical fluctuations which affect consumers' incomes lead to greater shifts in consumption of the non staple items than for mass market items. In this way small firms would tend to have more variable sales than large firms if they have differentiated products. For small firms vertically related to large firms, the effect of demand fluctuations will depend on the nature of the technical relationship but a plausible account of subcontracting relations might see small firms to be used as surplus capacity in booms which therefore smoothes out production for large firms but increases the variability of production for small firms.

The other major aspect of Sutton's (1991) sunk costs when considering differentiated products is expenditure on Research and Development[2]. As Caves and Porter (1978) and Geroski (1994), *inter alia*, show, research and development will tend to increase the instability of market shares. It is not clear from the evidence, however, whether the variability will affect large firms and small firms to an equal extent or not. The firm size distribution of innovations is a much debated and complex issue on which the only consensus appears to be that there are no strong results[3]. The

[2] Sutton's model does not differentiate formally between advertising and research and development therefore we might expect them to have qualitatively identical economic effects. However we can consider three potential reasons for a difference; the extent of economies of scale in research and development expenditure may differ from advertising economies of scale; the effects of research and development may be less certain or predictable; the degree of persistence of the effects of research and development may be lower than the persistence of the effects of advertising. All of these factors could reasonably be expected to lead to different structures in the Sutton model.

[3] Geroski (1994, p161) concludes, "the bottom line seems to be that the structure of markets and the size (and internal structure of firms) are not the major determinants of innovative activity"

research and development literature would therefore not appear to be of assistance in explaining a general result relating firm size and flexibility.

III. Mills and Schumann flexibility hypothesis.

This section outlines the basis of the theoretical model developed by Mills (1984,1986) that underpins the empirical analysis of flexibility. The MS explanation for the coexistence of large and small firms is based upon differences in cost structures and, in particular, the idea of a trade-off between flexibility as embodied by the slope of the marginal cost curve and absolute cost advantage as embodied by the level of minimum average cost. Mills (1984) shows that in a competitive industry with free entry and a continuum of available (quadratic) cost structures only one cost structure will be observed in equilibrium. However, Mills (1986) shows that if there is a discrete set of technologies it is possible for two different cost structures to coexist and that the one with the lower minimum average total cost will be less flexible.

Following the derivation of Mills (1986, p205) a cost structure is posited for each firm i in the competitive industry of the form ,

$$c(x_i) = \alpha_i + \beta_i x_i + \frac{x_i^2}{2\gamma_i} \quad (1)$$

where x_i is output of firm i and α, β and γ are positive constants which are all allowed to vary across firms. Industry demand is assumed to be completely inelastic in each period and distributed with mean $\overline{X}$ and variance σ_X^2. The implied endogenous price distribution has mean $\overline{p}$ and variance σ_p^2. Firms are price takers, letting output be determined where price equals marginal cost. Combining these assumptions the profit maximising first order condition yields a supply function,

$$x_i = \gamma_i(p - \beta_i) \quad (2)$$

From (2) we can readily obtain (3)

$$x_i(p) - x_i(\overline{p}) = \gamma_i(p - \overline{p}) \quad (3)$$

which implies, $\sigma_{x_i}^2 = \gamma_i^2 \sigma_p^2$ so that firms with higher γ will have a higher variance of output. However because the empirical test involves a comparison of firms in different industries, while the theoretical focus is variation within industries, we are interested in the relative rather than absolute variability of output[4]. Mills notes that $\eta_i(\bar{p}) = x_i'(\bar{p}).\bar{p}/x_i(\bar{p}) = \dfrac{\bar{p}}{(\bar{p}-\beta_i)}$ is the elasticity of supply for firm i evaluated at mean price. Clearly, this elasticity of supply is greater for firms with higher β_i. Dividing (3) by output evaluated at the mean price $\bar{p}$ yields the coefficient of variation of output,

$$cv_{x_i} = \eta_i(\bar{p})cv_p \quad (4)$$

Where cv_{x_i} is the coefficient of variation of output of firm i and cv_p is the coefficient of variation of price at the industry level. Mills then establishes that in a zero expected profit equilibrium[5] it is possible for firms with two different types of cost structure <u>and only two types</u> to coexist and that the firm type with the higher level of minimum average cost will have a more elastic supply (the measure of flexibility). As this step is crucial in the argument that follows we will draw it out more clearly. The level of average costs according to the cost function in (1) is

$$\frac{c(x_i)}{x_i} = \frac{\alpha_i}{x_i} + \beta_i + \frac{x_i}{2\gamma_i} \quad (5)$$

minimising average costs leads to a point of minimum average costs defined by (6),

$$\frac{c(x_i^*)}{x_i^*} = \beta_i + \sqrt{\frac{2\alpha_i}{\gamma_i}} \quad (6)$$

where x^*_i is the output at the point of minimum average cost. To obtain a point of closure Mills and Schumann then impose a zero expected profit condition. The current period profit function in their framework is

$$\pi_i(p) = \gamma_i(p-\beta_i)^2/2 - \alpha_i \quad (7)$$

taking a Taylor series expansion of this around the mean price yields an expected profit function of the form,

$$E\pi_i(p) = \frac{\gamma_i(\bar{p}-\beta_i)^2}{2} - \alpha_i + \frac{\gamma_i\sigma_p^2}{2} \quad (8)$$

setting this equal to zero and rearranging yields

$$(\bar{p}-\beta_i)^2 + \sigma_p^2 = 2\alpha_i/\gamma_i \quad (9)$$

---

[4] This an important point. The need to normalise on firm size due to the non comparability of different industries is in part responsible for generating the observational equivalence of the two approaches.

[5] Firms are assumed to maximise expected profits within this framework of demand uncertainty under the additional assumption of risk neutrality.

which can be substituted back into (6) to give a value of minimum average cost in terms of β_i and the first two moments of price. This expression can then also be used to show that the crucial two relations as far as Mills and Schumann are concerned.

$\partial(c(x^*)/x^*)/\partial\beta > 0$ (10)

$\partial\eta/\partial\beta > 0$. (11)

Namely that minimum average costs are increasing in β and so is the supply elasticity of output. For Mills and Schumann, this establishes that firms with high minimum average costs will have high output variability. From (11) this is attributed to be due to a trade-off between static and dynamic efficiency, a rise in β must be accompanied by a rise in γ or a fall in α, both of which are consistent with Stigler's flexibility hypothesis in terms of either lower capital intensity or a flatter slope of the marginal cost curve. We will demonstrate in the next section that as a general conclusion this is logically flawed. However, we will continue with the exposition of the Mills and Schumann paper to draw out their results more fully.

This combination of the trade-off between minimum average cost and flexibility and the result in (4) establishes that more flexible firms will have a higher coefficient of variation of output. While there are other results which have been established within the framework, this is the important one for purposes of empirical investigation. MS (1985) then *assume* that economies of scale exist so that lower minimum average cost is associated with larger firms and that therefore the larger firms must be less flexible in the sense of a less elastic supply curve, and have a lower coefficient of variation of output, than small firms if they coexist in a competitive equilibrium with demand fluctuations. One should note that this raises problems of interpretation and falsification of the hypothesis[6]. A rejection of the inverse relationship between output variability and firm size could mean that the model is not a fair description of reality or that economies of scale of the kind specified are absent hence the assumption is inappropriate.

MS do consider the possibility that market power could generate the result. To this end they estimate the relationship for industries of differing levels of concentration and find that it

[6] This is the conventional problem associated with the Duhem-Quine thesis.

continues to hold irrespective of the degree of concentration. The conclusion they draw is that market power is irrelevant to the explanation of sales variability. There are two initial responses to this: the first is that they do not include concentration as an independent regressor which would appear to be a more suitable way of testing for such effects rather than simply truncating the sample. By estimating for different ranges of concentration it is difficult to statistically test for *significant* differences in the sales variability - market share relationship or at least this is not done. The second response is that the implicit assumption would appear to be that the market power and flexibility explanations for sales variability are mutually exclusive, whereas, as we will show below, they are actually from a theoretical standpoint largely observationally equivalent and perfectly consistent with a combination of the two candidate explanations. In the next section we will extend the model to the situation of imperfect competition which allows us to examine more rigorously the precise implications of introducing market power into the explanation of sales variability.

There are three points which need to be made or reiterated in respect of this model of firm flexibility which are important for the economic implications and potential policy directions indicated.

1. The assumption of perfect competition and marginal cost pricing ensure that any equilibrium which is found in the model will be Pareto efficient.

2. The trade-off between the point of minimum average cost and the supply elasticity is a consequence of the zero expected profit condition which is imposed by the assumption of perfect competition.

3. A strong conclusion is that there is no need to resort to either imperfect competition or product differentiation in order to explain the coexistence of heterogeneous firms in industries as demand fluctuations and cost differences are sufficient to explain their existence.

IV. A model of flexibility with imperfect competition.

We now introduce some degree of imperfect competition into the framework and demonstrate the implications of so doing. The most straightforward form for exposition is a Cournot conjectural variation duopoly with two firms labelled i and j, so we will focus on this case but it can be shown the results are robust to the introduction of more than two firms. We assume that industry demand rather than being completely price inelastic is of the form $X = \theta - \omega\, p$[7] where demand fluctuations are generated by variation in θ such that it is distributed with mean $\bar{\theta}$ and variance σ_θ^2. Adopting the same cost function as in (1)

$$c(x_i) = \alpha_i + \beta_i x_i + \frac{x_i^2}{2\gamma_i} \qquad \text{(1a)}$$

Choosing x_i to maximise profits and rearranging the first order condition yields an equilibrium relation analogous to (2) of the form ,

$$x_i = \frac{\rho_i \gamma_i}{(\rho_i + \gamma_i)} . (p(X) - \beta_i) \qquad \text{(2a)}$$

Where $\rho_i = -1 \Big/ \frac{\delta p}{\delta X} \frac{\delta X}{\delta x_i}$ hence in a pure Cournot model $\rho_i = \omega$, under the assumption of competitive conjectures $\rho_i \to \infty$ and 2a collapses to 2. In the converse case of fully collusive conjectures $\rho_i = \omega . s_i$ where s_i is the market share of firm i. It is clear from this that an increase in ρ_i, the competitive stance taken by the firm, and an increase in γ_i, a flattening of the marginal cost curve, are identical in terms of their effect on the relations developed. By defining $\gamma_i^* = \rho_i \gamma_i / (\rho_i + \gamma_i)$ it is equally clear that this then follows through to most of the other relations developed within the Mills and Schumann framework. The reason for this is clear if one considers that the change in optimal output for a firm resulting from a shift in the demand curve will depend on both the slope of the marginal cost curve and the slope of the perceived marginal revenue schedule. Changes in competitive stance are passed through in changes in ρ_i and also imply different slopes of the perceived marginal revenue schedules.

---

[7] Throughout the following exposition the main restrictions on the generality of the results are that we assume that the second derivative of the demand function is zero, implying a linear demand function, and that the third derivative of the cost function is zero, implying a linear marginal cost curve. Both of these restrictions are implicit in the Mills and Schumann exposition although they include the additional restriction that the first derivative of the demand function is zero, an assumption inconsistent with imperfect competition.

It can also be shown (see Appendix 1 for details) that under a more restrictive formulation of demand that output adjustment costs can be incorporated into the value of γ_i^* yielding a further observational equivalence. It should be noted that this requires a specific stationary form of demand fluctuations less general than that adopted here.

We can show that $\sigma_{x_i}^2 = \gamma_i^{*2}\sigma_p^2$ and we can also establish that,

$$cv_{x_i} = \eta_i(\bar{p}).cv_p \qquad (4a)$$

where $\eta_i(\bar{p}) = \dfrac{\bar{p}}{(\bar{p}-\beta_i)}$.

(4a) indicates that the firm with the higher β_i will have a greater coefficient of variation of output and be more flexible in the sense defined by Mills (1986). At this point it should be noted that this measure is not necessarily consistent with the alternative definition of flexibility, $1/\gamma_i$, the slope of the marginal cost curve, which has been used as an inverse measure of flexibility by Stigler (1939), Marschak and Nelson (1962) and Mills (1984). Only if there is a trade-off between β_i and $1/\gamma_i$ are the two interpretations consistent. The slope of the marginal cost curve is the more important aspect of the cost function from the point of view of the flexibility hypothesis characterised by Stigler.

From the above it proves relatively straightforward to demonstrate that the imperfect competition effects and flexibility effects are inseparable and observationally equivalent at this stage. The implications of this equivalence are:

1. It is no longer necessary for the equilibrium to be Pareto efficient and indeed we can expect, given the preponderance of imperfect competition in real markets, that in a large number of cases there will be an inefficient allocation of resources.

2. Because there is no need to impose a zero expected profit constraint there is no necessary trade-off between the slope of the marginal cost curve and the level of minimum average costs.

Wholly inefficient firms in the sense of cost structure may still coexist with more efficient firms. In order to establish that firms with different cost structures can coexist in the same industry, it is sufficient only to assume that there exists some barrier to entry so that the number of firms is quasi fixed[8].

3. The observational equivalence between approaches does not deny the possibility that the flexibility explanation of industry structure is correct, but rather suggests that it is not possible to infer from the relations developed to ascertain whether this is the case or not.

4. A further implication of the imperfect competition framework is that we can relax one of the restrictions on the cost function namely that $\infty>\gamma_i>0$. It is necessary within a competitive framework for the marginal cost function to be upward sloping if the optimal output level for the firm is to be defined. This is no longer the case with imperfect competition. Therefore we can admit constant marginal costs ($1/\gamma_i=0$) or even decreasing marginal costs within a restricted range ($\gamma_i<0$)[9]. In the context of a Cournot conjecture framework this is simply the familiar restriction that marginal cost curve should not be more steeply downward sloping than the demand curve. This point may appear trivial from a theoretical perspective but as will be discussed below in section V it can have important implications for empirical work.

Developing point 2 above gives the most insight into what is actually being measured in the empirical analysis of Mills and Schumann flexibility hypothesis. Their hypothesis depends almost entirely on firms having higher levels of β. We can see this from an examination of (4), (10) and (11). With the imposition of zero profits a trade-off between β and the other parameters can be established; without that imposition no trade-off is necessary but the empirical result will still hold. This is because the empirical result only depends upon β and not upon the identification of the other parameters. Therefore we can turn the Mills and Schumann conclusions on their head.

Firms which are statically inefficient, in the sense of having higher values of β, will have higher output variability. Demonstrating that small firms have higher output variability merely

[8] It should also be noted that within the general Cournot framework more than two different cost structures can be present in the same industry at the same time.

[9] In order for $\gamma_i<0$ to yield a stable outcome the additional restriction $-\gamma_i>\rho_i$ must hold.

demonstrates that small firms are statically inefficient. In order to show that small firms are necessarily dynamically efficient it is necessary to show in addition that they are subject to competitive pressures forcing (expected) profits to zero.

While Mills and Schumann demonstrate the former they merely assume the latter and as we have suggested above the two situations are observationally equivalent using their empirical method so they can make no empirically based conclusion concerning the issue. That the supply elasticity of output is solely determined by β may appear surprising. The intuition behind the fact that statically inefficient firms will have a greater coefficient of variation of output, *regardless of whether they are dynamically more efficient or not,* is that statically inefficient marginal firms will tend to be fairly small but as price rises so output rises fairly rapidly as profitability increases. By contrast efficient firms will tend to produce relatively high levels of output even when prices are low so the proportionate change in output is lower for these (static) efficient firms.

Following on from this, one of the properties of the Mills and Schumann model which has not been exploited previously is that it enables us to establish the market shares of the firms. If we can directly obtain relationships between firm size and flexibility the we can minimise on the assumptions necessary for empirical application of the framework. We shall now explore the conditions under which the more flexible firm, in the empirically observable sense of (4a), will also be the smaller firm. To illustrate this the ratio of the output of firm i to the output of firm j is given from (2a) as,

$$\frac{x_i}{x_j} = \frac{\gamma_i^*(p-\beta_i)}{\gamma_j^*(p-\beta_j)} = \frac{\rho_i\gamma_i(\rho_j+\gamma_j)(p-\beta_i)}{\rho_j\gamma_j(\rho_i+\gamma_i)(p-\beta_j)} \quad (12)$$

From (12) we may obtain two propositions

a) *If there is* <u>*no*</u> *trade-off between the intercept value of the marginal cost curve and its slope then the market share of the firm with the higher value of β will be lower than the market share of the firm with the lower value of β.*

This can be seen by observing that if $\beta_i > \beta_j$ and $1/\gamma_j \leq 1/\gamma_i$ then $x_i < x_j$. This proposition establishes that we may observe a negative correlation between the coefficient of variation of output and market share in situations where flexibility in the Stiglerian sense is positively related or unrelated to market share.

b) *If there exists a trade-off between the intercept value of the marginal cost curve and it slope there is no unambiguous relation to firm size.*

As the trade-off becomes more important then it becomes possible for a positive relationship between output variation and market share to exist. However, if $\beta_i \rangle \beta_j$ and $\gamma_i \rangle \gamma_j$ i.e. there is a trade-off between the intercept value of marginal cost and its slope then firm size cannot be unambiguously determined. In particular, if the trade-off is small then the firm with the lower value of β will still be the smaller firm i.e. a trade-off exists between static efficiency and the slope of the marginal cost curve the firm with the lower β will be the smaller firm. However, if the trade-off between β and γ becomes too great then the situation will be reversed and the more flexible firm will become larger than the less flexible firm.

It also follows from (a) that in the situation that both firm have identical slopes of their marginal cost curves the firm with the higher β_i in the Cournot context[10] will have the lower market share. Therefore it can be demonstrated that within a restricted Cournot model small firms will have a greater coefficient of variation of output than large firms, however, this result is not general. *Ceteris Paribus*, the firm with the flatter marginal cost curve will be the *larger* firm. This goes further than the Mills and Schumann result because they only find that flexible firms have a greater coefficient of variation of output than less flexible firms, they then <u>assume</u>[11] that more

[10] Or any framework where the firms have the same conjectural variation.

[11] This is not to say that the assumption is implausible but simply that the relationship between flexibility and firm size is not derived from within their model. Instead it is generated by a combination of the observed empirical relationship between capital intensity and firm size and the theoretical relationship between capital intensity and flexibility.

flexible firms are small which is then used to support their empirical framework. Again from (12) if all firms have the same β and γ but have different conjectural variation terms, ρ, they will differ in size, the more competitive firm being larger. Firms with the same value of β, however will have the same coefficient of variation of output. This shows that some cost difference is important for establishing the inverse relationship and that market power alone cannot generate the result. However it should be noted that the difference in costs relates to static inefficiency and not to the dynamic cost advantage suggested by Mills and Schumann.

Examining the determinants of firm size and incorporating the features of imperfect competition in this way therefore raises some problems which are indicative of some of the difficulties with this simple theoretical characterisation. A more adequate characterisation of firm market share determination would have to take into account product differentiation, output adjustment costs or, to a lesser extent, discontinuous marginal cost curves in the case of capacity constraints etc. in addition to the current framework, in order to explain some size differentials.

We have established that it is β that matters for establishing the empirical relationship. Therefore looking at price cost margins in this context can potentially yield more insight in trying to understand the relationships developed. In doing so we can further demonstrate the general nature of the empirical result which has been shown by MS. Clearly, in the competitive context of MS the margin between price and marginal cost is zero. However, the empirical price cost margin that is more commonly used is the margin between price and average variable cost. Using (2a) we can see that the price average variable cost margin measured at the mean price is defined by,

$$\overline{PCM_i} = \frac{\bar{p} - \beta_i - x_i(\bar{p}) / 2\gamma_i}{\bar{p}} = \frac{(\bar{p} - \beta_i)(1 - \gamma_i^* / 2\gamma_i)}{\bar{p}} \quad (13)$$

In the case of perfect competition $\gamma^*=\gamma$ so incorporating (4a) the mean price cost margin is equal to $\frac{1}{2.\eta_i(\bar{p})}$. Predictably in relaxing perfect competition the further one moves away from perfect competition the higher will be the price cost margin, i.e. the lower is $\gamma^*/2\gamma$, the higher the price cost margin. From relation (4a) the firms with the higher value of $\eta_i(\bar{p})$ have a greater output variability and from (13) these firms and the more competitive firms will have lower price cost margins. It is also conventionally predicted from theory that the price cost margin is positively related to market share, whatever the reason one might wish to ascribe to this relationship. Combining these results we can therefore show as a general result that as market share is positively related to the price cost margin which is in turn negatively related to output variability and the level of competition. We can therefore establish with more generality than ascribed by Mills and Schumann the empirical relation which shows that market share is negatively related to output variability and we should expect this empirical result to hold with the same regularity that applies to the price cost margin - market share result.

A note of caution must be sounded concerning the true generality of this result. Schmalensee (1989a, p984) produces two "stylised facts" based upon empirical analysis that are applicable to this situation. Stylised fact 4.11 suggests that "In samples of U.S. firms of business units that include many industries, market share is strongly related to profitability." However stylised fact 4.12 also states that "Within particular manufacturing industries, profitability is not generally strongly related to market share." The implication of the theoretical link between sales variability, price cost margins and market share and the empirical facts 4.11 and 4.12 is that we would expect confirmation of the flexibility result with cross industry samples of firms but a weaker relationship looking at individual industries. As we examine in the next section, the empirical analysis of the flexibility hypothesis, that has been broadly affirmed the hypothesis, has exclusively been conducted on the kind of firm level data sets covering many industries that have produced stylised fact 4.11.

V Empirical analysis of the flexibility hypothesis

V.i. Previous empirical studies

In this section we will examine the previous empirical analyses concerning the flexibility hypothesis. These empirical analyses are found to be problematic and in this section suggestions are made for a redefinition of the empirical approach which could usefully shed light on these issues. There have been three main empirical analyses which have been published within the specific flexibility approach put forward by Mills and Schumann (1985), in addition to the initial work two published studies aimed at replicating the results are Chappell, Das and Shughart (CDS) (1993) Zimmermann (1995). In addition to these studies it is possible to draw on previous empirical work on the nature of short and long run cost functions which clearly have a bearing on the questions raised within the flexibility framework.

In their empirical analysis Mills and Schumann use a basic specification of the form,

$$SVAR_i = \alpha + \beta_0 MS_i + \beta_1 (K/S)_i + \beta_2 ISV_j + \varepsilon_i \qquad (14)$$

where $SVAR_i$ is sales variance of firm i[12], MS_i is Market Share of firm i, $(K/S)_i$ is the capital sales ratio of firm i and ISV_j is the variability of total sales in industry j. Where the expectation is that $\beta_0<0$, $\beta_1<0$ and $\beta_2>0$ because the prediction of the flexibility hypothesis is that, after controlling for capital intensity, there will be a negative relationship between sales variance and market share within industries, while an industry with greater sales variance will tend to lead to firm with greater sales variance. The need to control for capital intensity as one element of the cost structure is identified by Mills and Schumann. As flexibility is a feature identified by Stigler which can be achieved by substituting variable for fixed costs then one would expect sales variability to be inversely related to the capital output ratio, as a proxy of the importance of fixed costs.

Subsequent to this initial basic regression, in order to counter the possibility that market power might be causing the inverse relationship, Mills and Schumann truncate the sample progressively according to the level of concentration in the industry, so that industry level differences in market power can be catered for. The data is formed of a sample of 856 US companies using data from 1970-1980. The results of the analysis are that the predictions are met in terms of each variable having the predicted sign and being significant. In terms of the control for the degree of market power the coefficients remain of the predicted sign and are significant irrespective of the level of concentration with the exception of the effect of capital intensity which is insignificant in

[12] All three papers cited above contain discussion of the appropriate empirical measure of sales (or employment) variability in this context, using some form of log variance adjusted for trend. As this is not the focus here we will not discuss this issue further.

determining sales variability for firms in low concentration industries, a point not commented upon by Mills and Schumann. On the strength of these latter results they conclude that it is flexibility and not market power which drives these results.

The second empirical study by CDS replicates the same estimating equation (14) with the exception that the data used is derived from size class data for the United States from 1977-1988. No attempt is made to control for the level of concentration in the industry. The results obtained confirm the previous findings in terms of being significant and of the correct sign. The third study by Zimmermann adopts a more wide ranging approach and therefore is more interesting. Unable to obtain data on German companies on output the focus is on the variability of employment and capacity utilisation. Three different measures of size, including both absolute and relative measures, are compared for their properties. In addition, variables controlling for inventories, order backlogs and export propensity as well as dummy variables for consumer and investment goods are included as independent variables. There is also an attempt to control for concentration by truncating the sample into low concentration industries and comparing results. There is no attempt however to control for capital intensity or sales variability at the industry level. In an interesting development, Zimmermann also derives ways of approximating within the theoretical framework, elements of the cost function and can then estimate their relation to firm size.

Zimmermann's results relating to employment flexibility confirm the usual inverse relationship between firm size and employment variability and this occurs irrespective of whether the sample is truncated according to the level of concentration or not. This offers support for the Mills and Schumann interpretation. However, the estimates of the parameters of the cost function contradict this interpretation. In principle there should be a trade-off between minimum average cost and flexibility. Those firms with high minimum average cost will be more flexible because they are at a static cost disadvantage which they have to overcome by their flexibility in the dynamic context. As Mills and Schumann more formally state "If competitive equilibrium supports technologically heterogeneous firms, that is, firms with different cost parameters, the more flexible firms have greater minimum average costs." (1985, p760). However Zimmermann finds a weak positive relationship between the level of minimum average costs and firm size. In all cases the estimated coefficient is positive but in only three out of six cases is the point estimate significant. These cases are where the size variable is relative rather absolute. In none of the cases is the expected negative relationship found. This result clearly contradicts the Mills and Schumann hypothesis. Zimmermann therefore suggests that the relationship is more complex and that further work is required.

To briefly summarise the discussion thus far, the inverse relationship between sales (or employment) variability and relative firm size appears to be relatively robust across data sets and to the inclusion (or omission) of a number of control variables. This result would therefore appear to be well on the way to becoming an empirical regularity of the form espoused by Schmalensee. This it to be expected given we have already established on a theoretical level that it should be the case that the relationship should hold in a wide range of circumstances. Attempts at slightly more complex corroboration of the flexibility hypothesis have yielded more contradictions and therefore indicate that there are issues still to be resolved.

V.ii. Criticisms of previous studies

Having up to this point uncritically summarised the salient points of the empirical studies we must now turn to consider some of the potential problems which arise. The transfer from a theoretical model to an empirically estimable framework is nearly always complex and almost never exact, therefore the points that are raised relate largely to the specification of the relevant variables and their economic interpretation rather than to questions of the precise functional form, debate about which can often be fairly arcane. There is one exception to this rule which is the examination of the contradiction outlined above, discovered by Zimmermann, which is discussed below. The problem which arises here would appear to be an attempt at too close a relationship between the precise theoretical framework and the empirical relationship which means that the framework is insufficiently versatile to cope with alternative theoretical propositions. This is dealt with in the following paragraph.

The contradiction is that the flexibility hypothesis predicts an inverse relationship between firm size and minimum average costs, but Zimmermann in contrast finds a positive relationship. We show now that it can be reasonably expected that in the case of a misspecified framework, i.e. trying to fit a model developed using perfect competition to data generated within an imperfectly competitive environment, that such an apparently contradictory result can be generated.

Reproducing some of the key equation from the exposition in section III, the point of minimum average costs is given by,

$$\frac{c(x_i^*)}{x_i^*}=\beta_i+\sqrt{\frac{2\alpha_i}{\gamma_i}} \quad (6)$$

where x^*_i is the output at the point of minimum average cost.

Combining this with the expected profit function (8) and imposing a zero profit condition

$$E\pi_i(p)=\frac{\gamma_i(\bar{p}-\beta_i)^2}{2}-\alpha_i+\frac{\gamma_i\sigma_p^2}{2} \quad (8)$$

yields the key empirical relation

$$(\bar{p}-\beta_i)^2+\sigma_p^2=2\alpha_i/\gamma_i \quad (15)$$

which can be substituted back into (6) to give a value of minimum average cost in terms of β_i and the first two moments of price. This is the technique used by Zimmermann to obtain estimates of minimum average cost.

If we instead consider the imperfect competition framework the results concerning the profit function are altered such that adopting the terminology of section IV, i.e. $\gamma^*=\rho\gamma/(\rho+\gamma)$.

$$E\pi_i(p)=\frac{\gamma_i^*(2\gamma-\gamma_i^*)}{\gamma_i}\frac{(\bar{p}-\beta_i)^2}{2}-\alpha_i+\frac{\gamma_i^*(2\gamma-\gamma_i^*)}{\gamma_i}\frac{\sigma_p^2}{2} \quad (8a)$$

There is no need to set this equal to zero because we are operating within an imperfectly competitive environment therefore instead expected profit may be characterised by a level of profit described by some function $g(z,s_i)$ which depends upon a set of potentially relevant parameters z and some measure of firm size s_i. Imposing this and rearranging yields a relation analogous to (15) of the form,

$$(\bar{p}-\beta_i)^2+\sigma_p^2=\frac{2\alpha_i\gamma_i}{\gamma_i^*(2\gamma_i-\gamma_i^*)}+\frac{2g(z,s_i)\gamma_i}{\gamma_i^*(2\gamma_i-\gamma_i^*)} \quad (15a)$$

Equation (15a) demonstrates two things. Firstly, that any attempt to measure the left hand side of (15a) in order to approximate minimum average costs will, if there is imperfect competition and positive profits, include some element of profits $g(z,s_i)$ within the measurement. If profit is increasing in firm size, the measurement error will be greater for large firms. Secondly, we can show that even this assumption of positive profits is not necessary to generate the distortion. If expected profits are zero, so that the second term on the right hand side of (15a) drops out, the first element of the right hand side will still be incorrectly measured. It is straightforward to show that $\frac{2\alpha_i\gamma_i}{\gamma_i^*(2\gamma_i-\gamma_i^*)}>2\alpha_i/\gamma_i$ so long as firms do not behave competitively and it is also trivial to show that this inequality becomes greater as a firm's market power increases. Therefore, a distortion in the measure of average variable cost will arise as a result of the existence of market power. For both of these reasons we can expect that the relationship between the measured minimum average cost and firm size will be biased upwards due to the measurement error arising from the use of an incorrect framework.

Paradoxically, this argument actually strengthens the case for a form of flexibility hypothesis because it does not rule out the possibility that, despite Zimmermann's findings there may indeed be a negative relationship between minimum average costs and firm size. The implications and interpretation would have to be slightly different, however, because imperfect competition is implied.

Turning to the analysis of sales variability, the first, perhaps obvious, but nonetheless important point is that the theory relates to production flexibility the data analysed is sales rather than output data[13]. The implicit assumption is that the good is non storable and can be sold instantaneously. However, these are not good working assumptions for most industrial firms. While it is perfectly possible to abstract from inventories and production lags in theory, the minimum necessary adjustment from an empirical point of view would appear to be that the variability of output is considered rather than sales. An attempt to incorporate inventories, and

[13] In the absence of output data employment fluctuations have been analysed which represent a more plausible alternative although without implausibly detailed data at this level of generality the intensity of labour use, necessary to calculate the labour input, is not measurable.

also potentially production lags into the framework, both theoretically and empirically, would enable a more broadly interesting and engaging line of approach to be adopted.

The second point concerns the incorporation of the capital sales ratio which on a theoretical level is crucial and therefore could only be excluded from empirical work with some strong justification. The theory is designed to explain the difference in the output variability of firms *within* the same industry but the measure of capital sales ratio used is the absolute level. Although it is clearly an empirical issue, given the wide differences in technology across industries, the absolute level of the capital sales ratio is likely to vary to a greater extent across industries than within industries. Therefore, the appropriate capital sales ratio is that of the firm relative to the industry average, and not the absolute level. This point would appear to be confirmed by, and provides a potential explanation for, the reducing significance of the capital sales ratio in Mills and Schumann's study as firms in more concentrated industries are progressive excluded from the sample. If the capital sales ratio is indeed picking up inter industry effects, it is performing a different economic function to that which it is intended for and a re-specification would be appropriate. A further point to be made is that the inclusion of the capital sales ratio is intended as a indicator of the importance of fixed costs relative to variable costs. A problem with this proxy is that only the fixed costs of plant are included; ideally (quasi) fixed labour, advertising and research and development costs etc., which will all enter into the cost function as fixed costs.

This problem has implications for their further experiments on the effects of concentration on the relationship. By truncating the sample according to the mean level of concentration and performing separate regressions of (14) for each group, MS suggest that this can eliminate market power explanations for the sales variability phenomenon. It can equally be argued that this is not an appropriate rejection anyway because again, the market power must differ between firms in the same industry, rather than across the board as is implied by MS. It is clear from their results that changing the degree of concentration has negligible implications for the coefficient on market share, but has important implications for the coefficient β_1 which declines in value and significance as the mean level of concentration falls. This would appear to indicate further that the effect of the capital sales ratio is an inter industry effect rather than an intra industry effect.

The fourth point and perhaps the most important point to be made is that it is possible that the empirical approach of finding the determinants of sales or output variability is misconceived. The important aspect is the nature of the cost structures faced (or adopted) by small and large firms. From an empirical point of view we need to investigate these directly, although this is likely to be a much more difficult task. From the perspective adopted here, we have seen that output variability only tells us about β in (1) but not the other parameters of this cost function. We turn to this cost function analysis in the following subsection.

V.iii. Estimating short run cost functions[14]

Before going on to outline our own approach, we look at evidence from previous studies of cost structures, but before we look at the studies of cost structure *per se* we need to look at the related intermediate work on inventory behaviour. This is useful because the initial path of research on inventory behaviour mirrors, to some extent, the initial research path followed in the flexibility debate. In the empirical inventory literature, as in the flexibility literature, the focus has been dominated by the analysis of the relative variability of sales, production or inventories rather than on the empirical analysis of the cost function which drives the theoretical results. Much empirical work on inventories has found that production tends to be more variable than sales, (as noted above the distinction between production and sales has been entirely omitted from the flexibility debate) which is empirically the reverse of the production smoothing argument which is implied by upward sloping marginal cost curves. There has been a vast literature which has sought to explain this anomaly but very little of this literature has taken the apparently radical step of

[14] Zimmermann (1995) was published after the bulk of this chapter had been written. The broad direction in which Zimmermann's methodology is pointing, i.e. attempting to identify specific parameters of the cost function, is similar to the direction that this chapter pursues. Given the criticisms expressed in the previous section and the analysis that follows, it is clear that the specific ways in which that paper, and this work, follow such a direction are very different.

refuting the idea of upward sloping marginal cost curves. A number of empirical articles[15], for example, seek to impose upon the data the constraint that the slope of the marginal cost curve must be positive or disregard data which give the "perverse" result of a negative slope. Ramey (1991) and Guariglia and Schiantarelli (1994) both find that in estimating cost functions on US and UK data, the implied slope of the marginal cost curve is negative. Specifically they are interested in estimating cost functions of the form given in (14) with the addition of terms for output and inventory adjustment costs. In general their estimates of γ_i are negative, a result which is ruled out by Mills and Schumann by assumption. If the marginal cost curve is negatively sloped, this would imply two things: firstly a theory of production bunching rather than production, smoothing that would explain why production tends to be more variable than sales; secondly the existence of imperfect competition as perfect competition is inconsistent with increasing returns[16].

Empirical evidence on long run costs indicates the presence of substantial economies of scale[17] however there is much less evidence[18] on the nature of short run costs which are clearly the focus for the current discussion. This focus on the long run appears slightly odd in this context, although the whole force of the structure conduct performance paradigm is oriented to looking at the long run. In the presence of uncertainty or instability, in this case demand fluctuations, it is even more important to view the long run as a series of short runs, because if instability is ever present, and that seems likely, one will only ever observe the short run solutions. What evidence on short run costs that exists tends to be in the form of survey evidence. This evidence indicates that in the short run costs are constant until the capacity level of output, is reached after which costs rise quite rapidly. A problem with these analyses are that they are fairly old, mostly dating from the immediate post war period when, according to the jargon, both the UK and US were still

[15] A number of these are cited in Ramey (1991).

[16] It is this aspect and the complications that necessarily entail from imperfect competition that one suspects is the cause of the unwillingness of (mainly US) empirical observers and theoretical analysts to contemplate decreasing marginal costs.

[17] A brief summary of the methods and results of these studies is contained in Hay and Morris (1991, p47-54).

[18] For example Scherer and Ross (1990) cover the determinants of economies of scale in the long run in considerable detail in their overview (p97-141) but restrict discussion of the determinants of short run costs to a single footnote.

employing "fordist" techniques of mass production. The argument of many of the advocates of the flexibility hypothesis is that there has been a radical change in the nature of production which is dated from the early 1970's onwards. Only evidence from the later period would be deemed useful from this point of view. The presence of capacity constraints of the form outlined above does not readily fit into the theoretical framework of MS as it implies a discontinuity or kink in the marginal cost function. This evidence would however be in accordance with Stigler's view of a technology which is divisible but not particularly adaptable. In this sense, the empirical evidence which directly examines cost structures is not particularly favourable to the MS flexibility position.

Studies of cost structures are fraught with difficulties which is probably why they have been avoided for the most part. It has been argued, in fact, that the study of cost structures in cross section is not feasible. Friedman (1962, p142) argues that "it [is] impossible to define the average cost of a particular firm for different hypothetical outputs independently of conditions of demand." Certainly any attempt to derive cost parameters on the basis used by Zimmermann will encounter this difficulty, making extensive use as it does of the variance of prices and outputs. However, Friedman's point largely applies to the case of attempting to identify long run average costs. It will be argued here however that estimating short run cost functions does offer the best way forward for the analysis of the flexibility approach.

There are two reasons for arguing for this approach; firstly directly estimating a short run cost function allows for adaptability of functional form and comparisons of alternative functional forms within recognised statistical procedures; secondly, *contra* Friedman, it can be argued that by concentrating on the short run cost function analysis, it is possible to move away from the problems of imperfect/perfect competition. With regard to the adaptability of the functional form it can and has been argued that many features of small firms may lead to a greater degree of flexibility[19]. We do not dispute any of these contentions, but merely point out that they are only important in assessing flexibility, in terms of the definition adopted here, in so far as they

[19] Such as a greater use of more flexible technologies.

influence the parameters of the short run cost function. The primary task is to attempt to identify the parameters of this short run cost function and then it may be possible to relate these estimates back to the factors that 'cause' flexibility. The only point of contention in the first stage is therefore the statistical problem of the appropriate parameterisation of the cost function.

Turning to the second point *contra* Friedman, estimating a cost function will usually only require, in terms of assumptions on behaviour, that firms are cost minimisers. Even this may be relaxed slightly. In the case of firms where fixed costs are inflated through managerial discretion and perks, or where excessive expenditures on advertising or other means of acquiring market power are incurred, it may still be possible to identify important parameters of the cost function. It would not, in this case and will not in general in the method advocated below, be possible to identify the level of minimum average costs. It is by moving away from the need to examine average costs that we circumvent the problems identified by Friedman.

To illustrate the general approach let us examine the attempts at estimating the "linear quadratic" cost function introduced by Holt *et al* (1960). This has been the most commonly used functional form adopted by those interested in studying inventory behaviour. It implies a current period total cost function of the form.

$$C(Y_t)= \alpha_0 + \alpha_1 Y_t + \frac{1}{2}\alpha_2(Y_t)^2 + \frac{1}{2}\alpha_3(\Delta Y_t)^2 + \frac{1}{2}\alpha_4(I_t - \alpha_5 S_{t+1})^2 \quad (16)$$

Where Y_t is current period production, $\Delta Y_t = Y_t - Y_{t-1}$, I_t is the inventory level in period t and S_{t+1} is the level of sales in period t+1. In principle, this cost function is relatively flexible because it is straightforward to add additional cubic terms. The final term in (16) is interpreted as an inventory holding cost. This arises because of the potential desire of firms to hold inventories against the possibility of stockout or backlog rather than simply a production bunching/smoothing exercise and there is, therefore, a desired inventory sales ratio to be maintained. From estimating the parameters of such a cost function we can establish the slope of the marginal cost function (ignoring the inventory terms) as $\alpha_2+(1+\delta)\alpha_3$, where δ is the discount factor. This would enable a form of test of the flexibility hypothesis which focuses on the slope of the marginal cost curve.

This is, therefore, a form which is more in tune with the hypothesis developed by Stigler. Estimating such relations for firms of differing sizes and different market shares may and should, if the flexibility hypothesis is correct, yield differing estimates of the slopes of the marginal cost curves. Clearly, there are problems associated with the estimation of such relations which are not trivial. In practice α_1 from (16) is expanded to be a function of factor prices so that,

$$C(Y_t)= \alpha_0 +(\alpha_{11} + \alpha_{12}pm_t + \alpha_{13}w_t)Y_t + \frac{1}{2}\alpha_2(Y_t)^2 + \frac{1}{2}\alpha_3(\Delta Y_t)^2 + \frac{1}{2}\alpha_4(I_t - \alpha_5 S_{t+1})^2 \quad (17)$$

where pm_t is the price of intermediate inputs and w_t is the wage rate. Noting that $Y_t \equiv S_t + I_t - I_{t-1}$ (18) and defining the firms objective function as the profit function,

$$V = E_0 \sum_{t=0}^{\infty} \delta^t (p_t S_t - C(Y_t)) \quad (19)$$

we can substitute (17) and (18) into (19) and maximise with respect to the current level of inventories and sales. At this point we should note that it is conventional in the previous literature to assume that sales are given, which means that only one of the first order conditions is operative, because the only decision to be made by the firm concerns the level of inventories. If we allow a downward sloping demand curve, however, the two first order conditions associated with maximisation of (19) are given by (20) and (21).

$$\begin{aligned} &E_t(-\alpha_{11}(1-\delta) - \alpha_{12}(pm_t - \delta pm_{t+1}) - \alpha_{13}(w_t - \delta w_{t+1}) - \alpha_2(Y_t - \delta Y_{t+1}) \\ &-\alpha_3(\Delta Y_t - 2\delta\Delta Y_{t+1} + \delta^2 \Delta Y_{t+2}) - \alpha_4(I_t - \alpha_5 S_{t+1})) = 0 \end{aligned} \quad (20)$$

$$E_t(p_t + p'(S_t)S_t - \alpha_{11} - \alpha_{12}pm_t - \alpha_{13}w_t - \alpha_2 Y_t - \alpha_3 \Delta Y_t + \delta^{-1}\alpha_4\alpha_5(I_t - \alpha_5 S_{t+1})) = 0 \quad (21)$$

As stated above, (21) is usually disregarded and a form of (20) is estimated, where it is necessary for purposes of identification that some restriction is placed upon the parameters in (20). Two prominent restrictions of this form are that α_4=1 the restriction used by Ramey (1991) and α_3=1 used by Blanchard (1983). In principle, the two equations should be estimated simultaneously with parameter restrictions across the regressions. This has not been attempted to our knowledge, presumably because of the statistical complexity involved. In order to achieve results within this framework, the (implicit) assumption has been made that by estimating (20) as a single equation does not lead to significant biases. From an empirical perspective, and combined with the theoretical arguments given above, it would appear that this approach offers the best way forward for a better understanding of the differences between the cost structures of small and large firms.

Before we consider previous estimates of (20) we require a slight digression. It is useful to observe that equations in the form of a restricted version of (21) have been at the centre of the recent upsurge in single industry studies of market power[20]. In the usual framework, (see Bresnahan (1982)) there are no inventories hence production equals sales. As a result, there is an identification problem; it is not possible to empirically separate out the slope of the demand curve $p'(Y_t)$ from the slope of the marginal cost curve α_2. Bresnahan (1982) has suggested one solution to this problem which is contingent upon the adoption of a specific functional form for the demand curve. Equation (21) above offers an alternative solution. Whereas previously the slope of the demand curve depended upon production by introducing inventories the demand curve depends on the level of sales and the cost curve depends upon the level of production. Therefore assuming both sales and production are available the two aspects can be empirically separated.

V.iv. Previous estimates for the US Automobile industry

It would be helpful from the perspective of equation (20) to examine previous estimates which can then be used to shed light on this issue. As far as we are aware no study has particularly examined the issue of firm size in this context, hence there are no general results to be reported. There are interesting papers, however, the results from which can be used to shed light on the issues discussed here. In particular, the information in Blanchard (1983) and the re-estimates using the same data set by Ramey (1991) can be used to good effect. These papers present estimates of a form similar to (20) for the mass production US car makers. As one of the world's most important industries this is a useful case to examine.

Estimates are presented for ten models, reproduced here in Appendix 2, mostly produced by a division of a larger company, as opposed to estimating at the firm level. Ramey reports that in estimating a form of (20) in each case the value of α_2 is negative, as is the overall value of

[20] For a fuller description of this resurgence see Bresnahan (1989)

marginal cost. This alone clearly contradicts the Mills and Schumann theoretical framework. In the context of diminishing marginal costs, the variability of output will increase as the slope of the marginal cost function becomes more negative as firms incentive to bunch production increases. Overall the estimates reported for the various models do not vary to an enormous degree, although some significant differences do exist[21]. For Blanchard, most of the estimates of α_2 are poorly determined, only one is positive and significant at the 5% level, but that is the smallest division, Cadillac. The other estimates vary but are not significantly different from zero, implying constant marginal costs for all of the other divisions. Ramey adopts a different restriction and obtains the negative results reported above. The lowest (most negative) values of the slope of the marginal cost curve are obtained for Oldsmobile and Chevrolet while the highest (least negative) values are for Cadillac, Buick and American Motors. Chevrolet is the largest producer while Oldsmobile ranks 5th out of ten in terms of production. Cadillac, Buick and American Motors rank 10th, 6th and 9th respectively. These results are difficult to interpret without in addition to these estimates comparing the separate estimates for α_2 and α_3. The estimates of α_3, the output adjustment costs, are lowest and are insignificantly different from zero for the three of the smallest divisions, Buick, Cadillac and American Motors and are highest for the largest division Chevrolet. However the estimates of α_2 are also most strongly negative for Chevrolet although there is relatively little relationship between size and the other estimates. For interpretation of these estimates, the lower output adjustment costs for the smaller divisions indicates some support for the flexibility hypothesis. The indications are also that there are significant short run economies of scale for most divisions.

From the data given in Blanchard we can also obtain the coefficient of variation of production for the specific models. Examining this data appears to confirm most clearly what has been suggested with regards the relationship between the variability of production and size. Ford (the second largest producer) has the lowest coefficient of variation of output, Chevrolet and Chrysler-

[21] As a consequence of the need to impose a restriction upon equation (11) the parameter estimates are contingent upon the imposed restriction. For Blanchard (1983) that restriction is that the output adjustment cost is the same for each model of car. For Ramey (1991) the restriction is that the inventory holding cost is the same for each model.

Plymouth[22] (first and third in size) rank 7th and 8th in terms of the coefficient of variation of output. The smallest models Cadillac, American Motors and Mercury Lincoln rank 1st, 2nd and 4th respectively in terms of the coefficient of variation of output. That small firms (or in this case small divisions of firms) have greater proportionate variability of output is therefore clear. While the evidence of the cost function estimates gives some indication that this would be the case, the presence of substantial short run economies of scale also point to a more complex relationship than that advocated by Mills and Schumann.

VI. Flexibility, capital intensity and variations in price cost margins

The original interest in the flexibility hypothesis from the point of view of this chapter was in terms of the point of comparison with the work of Schmalensee. In order to consider this we need to observe the variance of profitability and price cost margins. If we examine the ratio of price minus average variable cost to price, this can be defined for each of the situations described above.

In the Mills and Schumann approach, the price cost margin is determined by substituting (2) into the expression for average variable cost,

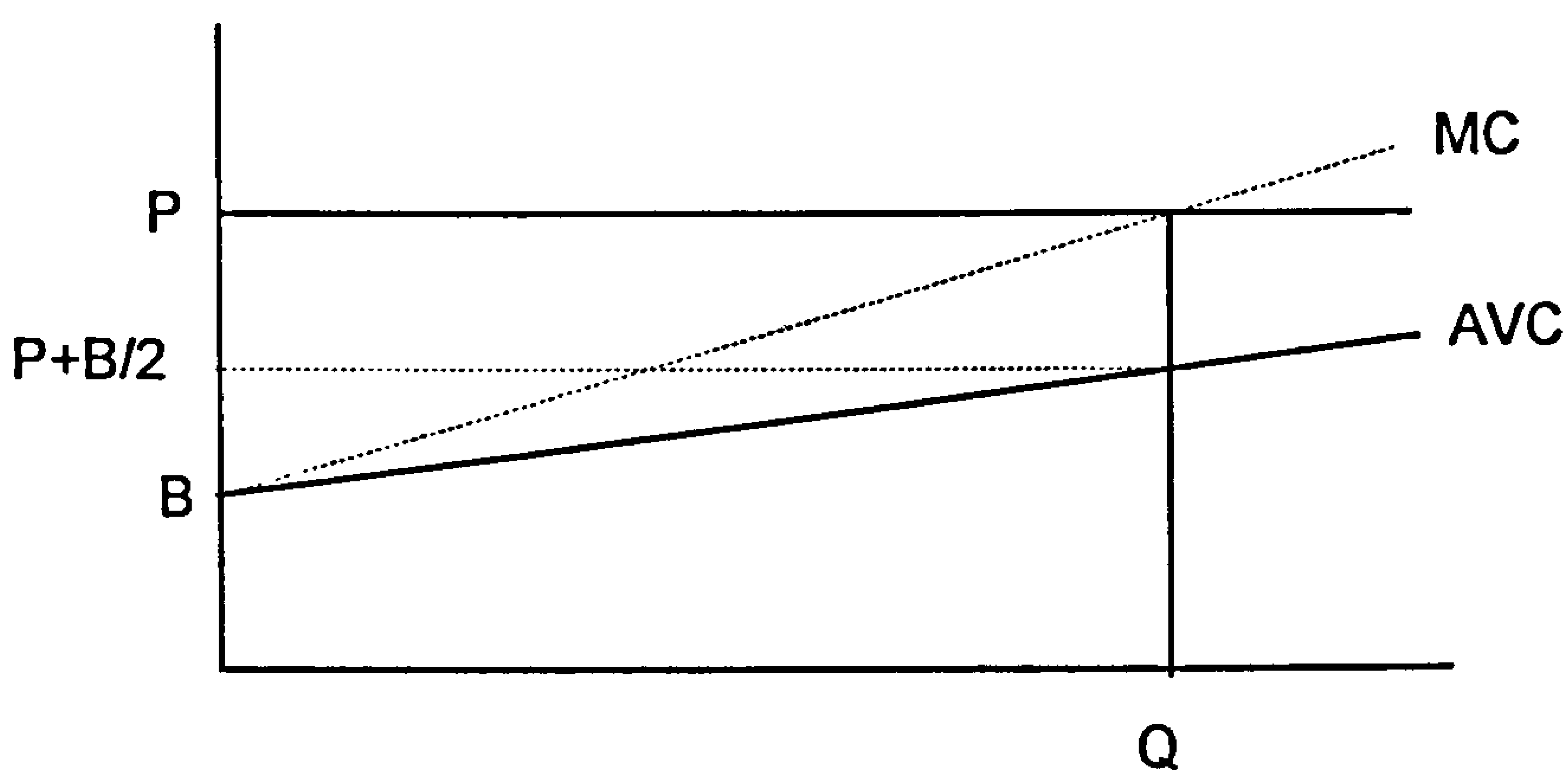
$$\frac{p-\beta_i-\frac{x_i}{2\gamma_i}}{p}=\frac{p-\beta_i}{2p} \quad (22)$$

The only explicit consideration of this relation is found in Chappell, Mayer and Shughart (1993). They suggest that because $\delta PCM/\delta p=\beta_i/2p^2$, then as demand, and hence price, increases so those firms with the higher value of β_i will increase their price cost margins to a greater extent than those firms with a lower β_i. Since those firms are also the flexible firms, in the Mills and Schumann sense of having a higher elasticity of supply $\eta_i(p)$, it can be shown that the more flexible and therefore for empirical purposes smaller firms, will increase their price cost margins to a greater extent than less flexible firms. The intuition behind this derives from the fact that the

22 Both of the Chrysler divisions had lower coefficients of variation than would be expected given their market shares. This may be a result of the institutional characteristics of the company. It is certainly the case that faced with the Japanese penetration of the US market in the 1980's, Chrysler appeared to be the least able of the big three to adapt.

price cost margin in this context is merely a measure of the difference between marginal and average variable cost. This can be illustrated diagramatically in Figure 1. As the slope of the average variable cost curve is half of the slope of the marginal cost curve, for any realisation of price the distance between price and average variable cost will be half the distance between price and the intercept of the marginal and average variable cost curves. Therefore the price cost margin depends only on β not on the slope of the marginal cost curve.

Figure 1



In the imperfect competition model the price cost margin has been defined earlier in (13) and can be rearranged as,

$$\frac{(p-\beta_i)(1-\dot{\gamma}_i/2\gamma_i)}{p}=\frac{(p-\beta_i)}{2p}\frac{(\rho_i+2\gamma_i)}{(\rho_i+\gamma_i)} \quad (23)$$

Clearly as long as this expression is positive, then δ^2PCM/$\delta p \delta\beta$ is still positive, hence the Mills and Schumann flexibility result will still hold. One should note however that the reverse is true if one adopts the Stiglerian definition of flexibility as typified by the slope of the marginal cost curve. This follows immediately from noting that δ^2PCM/$\delta p\delta\gamma<0$. In this case, for a given β, the flatter the marginal cost curve, i.e. the more flexible in the Stiglerian sense, the less price cost margins will respond to price fluctuations. It is this difference in result which causes the apparent contradiction discussed in Scherer and Ross (1990, p442).

Scherer and Ross discuss the impact of capital intensity on cyclical movements in price cost margins and suggest that firms with high levels of capital intensity will tend to have greater fluctuations in their price cost margins. In their analysis (ibid., p286-289) however, the crucial discriminating feature between capital intensive and other firms, is in fact the slope of the marginal cost curve rather than the level of fixed costs incurred. For Scherer and Ross, capital intensive firms possess relatively steep marginal cost curves, whilst for Stigler flexible firms have a high proportion of variable costs and correspondingly, the marginal cost curve is relatively flat. It is this factor which drives their results concerning the effect of capital intensity rather than the capital intensity itself.

If we recall Schmalensee's finding that the profitability advantage of large to small firms is counter cyclical, this is consistent with the explanation of flexibility given by Mills and Schumann but not that by Stigler. To explain this difference we should note once again that the Mills and Schumann results are actually driven by static inefficiency rather than dynamic efficiency. Schmalensee's finding is therefore consistent with the fact that small firms are statically inefficient and is not wholly consistent with the notion that small firms are dynamically efficient. The dynamic efficiency and capital intensity explanations are very similar as should be expected from Stigler's discussion of flexibility. However, they clearly have different predictions to the Mills and Schumann hypothesis. The alternative explanation considered by Schmalensee is that variations in market power will tend to lead to differences in the cyclical movements. A consideration of these issues involves an additional set of theories which are relatively complex in themselves. Further discussion of this aspect is not pursued here but is contained in Chapter 5 which deals separately with the issues of fluctuations in the degree of market power across the business cycle.

VII. Conclusions

A large amount of material has been covered in this chapter so a number of conclusions are warranted.

One clear conclusion which must be drawn from this analysis is that it is not possible to infer from the measured output variability anything about the curvature of the marginal cost curves of firms. There are two main reasons for this; firstly, the coefficient of variation of output can be shown to depend not on the slope of the marginal cost curve but its intercept; secondly, it is not possible to empirically separate demand effects from supply effects because of the equivalence of prediction. The first reason is important for the interpretation of the relations. It indicates that empirical measures of output variation are positively related to static inefficiency but that this does not necessarily imply a positive relationship with dynamic efficiency. Analytically the second aspect is less important once one takes the first into account. The failure to distinguish between the slope of the marginal cost curve and the slope of the perceived marginal revenue schedule is less of a difficulty when one considers that, in the absence of perfect competition, both are subject to the strategic decision making of the firm. The degree of flexibility not only depends upon the cost function but also upon the competitive stance adopted.

A second conclusion, which derives from the consideration of price cost margins, is that, in the same way that there is a general empirical regularity that price cost margins are positively related to market share, the theoretical framework considered here points to the existence of an important empirical regularity[23] of the inverse relationship between market share and measured output variability which industrial economists should be aware of.

The final conclusion is that rather than focusing on the variability of output, future empirical research should instead focus on attempts to estimate the relevant parameters of the cost function. Initial attempts to do this within a relatively tractable framework have been documented in this chapter. The evidence from the US motor industry suggests some limited support for the notion that output adjustment costs are lower for smaller firms, but also indicates substantial economies of scale for most firms. The clear suggestion for future work is that more attempts at identifying the parameters of the cost function at an industry level are required.

[23] Of the form which Schmalensee (1989a) suggests are so important for industrial organisation.

Appendix 1: Flexibility and output adjustment costs

To maintain simplicity of analysis we restrict the demand function to be completely inelastic and take a value which realises an (endogenous) price p in odd periods and a price p* in even periods. We restrict the analysis to perfect competition, the extension to imperfect competition is straightforward but tedious and unilluminating. We assume a discount factor of 1, again for simplicity. The cost function for the firm (suppressing the i subscript) in odd periods is therefore defined by,

$$c(x^*,x)= \alpha+\beta x+\frac{x^2}{2\gamma}+\delta(x-x^*)^2 \quad \text{(A1)}$$

where x is output in odd periods and x* is output in even periods and α, β, γ and δ are all positive constants[24]. The intertemporal profit function is defined by

$$\pi(x^*,x)= p^*x^*-\alpha-\beta x^*-\frac{x^{*2}}{2\gamma}-\delta(x^*-x)^2+px-\alpha-\beta x-\frac{x^2}{2\gamma}-\delta(x-x^*)^2 \quad \text{(A2)}$$

Maximising with respect to x and x* and rearranging the first order conditions we can solve for x* and x
such that,

$$x^*=\frac{4\gamma\delta.\gamma(p-\beta)+(4\gamma\delta+1)\gamma(p^*-\beta)}{8\gamma\delta+1} \quad \text{(A3)}$$

and

$$x=\frac{4\gamma\delta.\gamma(p^*-\beta)+(4\gamma\delta+1)\gamma(p-\beta)}{8\gamma\delta+1} \quad \text{(A4)}$$

Clearly if $\delta=0$ i.e. there are no output adjustment costs these will collapse to equation (2) in the main text. We can observe the variability of output noting that.

$$x-x^*=\frac{\gamma}{8\gamma\delta+1}(p-p^*) \quad \text{(A5)}$$

Therefore the variability of output will rise as the marginal cost curve becomes flatter and the cost of output adjustment falls. The coefficient of variation of output evaluated at the mean price can be found, defining $\bar{x}=(x^*+x)/2$ and $\bar{p}=(p^*+p)/2$ as

$$cv_{\bar{x}}=\frac{(x-\bar{x})}{\bar{x}}=\frac{(p-\bar{p})}{\bar{p}}.\frac{\bar{p}}{(\bar{p}-\beta)}=cv_{\bar{p}}.\eta(\bar{p}) \quad \text{(A6)}$$

which is as in (4) in the main text. In principle introducing adjustment costs in this way changes little, merely confirming a three way observational equivalence between the conjectural variation term, the slope of the marginal cost curve and the level of output adjustment costs. In practice the framework in which we have observed adjustment costs is unrealistically restrictive and we can expect demand to behave in a less stable more complex fashion which will lead current period optimal output to depend upon previous period output and the expectations upon demand realisations.

[24] Strictly γ may be negative so long as output adjustment costs, in the form of δ, are sufficiently high.

Appendix 2: The US Automobile Industry 1966-1979

The data contained in this appendix are reproduced or generated using the information provided in Blanchard (1983) and Ramey (1991)

Model	Market Share % (Domestic Sales)	C.V. production	Blanchard estimates of α_2/α_3
Chevrolet	25.4	.298	0.29
Pontiac	8.9	.355	-0.04
Oldsmobile	8.6	.389	1.21
Buick	7.6	.343	0.88
Cadillac	3.1	.400	1.85*
Ford	21.8	.257	0.11
Mercury Lincoln	5.9	.365	0.24
American Motors	3.1	.392	1.66
Dodge	5.9	.268	3.29
Chrysler Plymouth	9.7	.275	2.69

Notes: Market Shares are calculated from information contained in Blanchard (1983) Table 1. CV production is the coefficient of variation of production, calculated from Tables 1 and 2 of Blanchard (1983). The estimates of α_2/α_3 are from Blanchard (1983) Table 7B. * indicates that the estimate is significantly different from zero at the 5% level but no standard errors are reproduced.

Model	Ramey estimates of α_2	Ramey estimates of α_3	Ramey estimates of slope of MC
Chevrolet	-1.45*	0.40*	-0.651*
Pontiac	-0.88*	0.18*	-0.516*
Oldsmobile	-1.27*	0.24	-0.798*
Buick	-0.53	0.07	-0.394*
Cadillac	-0.41*	0.06	-0.300*
Ford	-0.73*	0.14*	-0.446*
Mercury Lincoln	-0.90*	0.17*	-0.551*
American Motors	-0.61*	0.09	-0.424*
Dodge	-0.94*	0.19*	-0.569*
Chrysler Plymouth	-0.80*	0.14*	-0.523*

Notes: These estimates are taken from Ramey (1991) Table 3. The slope of the Marginal cost curve is calculated as $\alpha_2+(1+\delta)\alpha_3$. The parameter estimates are reported, * indicates that the estimate is significantly different from zero at the 5% level but no standard errors are reproduced. Full estimates are presented in Ramey (1991).

Chapter 3. Vertical Integration, Input Bargaining and the Empirical Specification of Price Cost Margins

I. Introduction

In recent years it has become common when specifying empirical price cost margins to use net output as opposed to the more usual gross output in the denominator. Examples of this include Hart and Morgan (1977), Conyon and Machin (1991a, 1991b), Haskel and Martin (1992, 1994) and Conyon (1995) for studies of the UK, and Baldwin *et al.* (1984) and Dickson (1991) for studies using Canadian data[1]. As Conyon and Machin (1991b) and Conyon (1995) point out, with gross output as the denominator the point estimate on concentration from a regression of margins on concentration is lower than with net output as the denominator and the estimate is frequently insignificant. This tendency is discussed in Conyon and Machin (1991b) and the argument put forward for the redefinition of the margin is the need to control for the degree of vertical integration. In addition to these empirical studies, there have been criticisms made from a theoretical point of view concerning the need to incorporate the degree of vertical integration explicitly into a consideration of industry relations. This chapter seeks to analyse the logic of these arguments and the implications for the specification of the empirical relationship. In particular, the chapter argues that only through a consideration of the relationships between end-producer firms and firms providing inputs and services, can an adequate resolution of the issue of working with net and gross output be found. The chapter proceeds as follows: in the next section there is a brief and critical discussion of the rationales used for the use of the alternative (net output denominator) margin; secondly a consideration of the theoretical relations under alternative assumptions about the way in which bought in inputs are treated by firms; thirdly, these theoretical relations are translated into empirical specifications which attempt to take into account the elements considered by the theory. Finally, some evidence is presented using a panel of industry level UK manufacturing data for 1980-1991 in order to empirically compare the predictions of the different approaches.

[1] That the finding occurs in at least two countries indicates that it is not solely a problem with a single data source, such as the UK census of production, although we are not aware of studies which have identified the issue in the US.

However before we consider the treatment of material inputs as a potential influence on the price cost margin we need to consider other sources of potential bias arising from the use of price cost margins. In particular Fisher (1987) has criticised the use of the ratio of profits to sales as a measure of monopoly power because of the accounting conventions used. Fisher's conclusion is that the use of the profits to sales ratio is untenable once differing conventions for assessing profits are taken into account. Kay and Mayer (1986) and Geroski (1988) address the same issue but are more favourable about the possibilities for using accounting data albeit within circumscribed limits.

Fisher's main point is that firms differ in their assessment of the cost of capital particularly in the absence of a well functioning second hand market and that their assessment of the cost of capital will generally differ from that conventionally used by economists. Therefore for any particular investment project the profits to sales ratio will not generally be equal to the Lerner index and will deviate by the extent to which accounting depreciation of capital differs from economic depreciation. The point made by Kay and Mayer (1986) is that while in any one year profits may not reflect the true economic return of a project, because the cost of capital accounted for may differ from the true cost of capital in that year, it is still the case that in the long run a project with a higher than average rate of return will yield higher reported profits, therefore in evaluating the long run rate of return the profits to sales ratio may still have a role to play.

Nevertheless we must still take into account the possibility of biases that arise from differing accounting conventions between firms. One suggestion made by Fisher (1987) and adopted by Machin (1991) is to transform the profit to sales ratio using $-\log(1-\frac{\pi}{S})$ as the dependent variable because the extent of the bias is proportional to unity minus the measured profit to sales ratio. Machin (1991) reports that adopting the transformed dependent variable makes no substantive difference to his results. However, even adopting the transformation there is still the potential for bias if the adoption of accounting conventions is correlated with the right hand side variables. As one of the most important possible influences is the growth rate of the firm it is important to include this in regressions. Similarly advertising expenditures and other investments that may have effects over several years will lead to differences in the timing of reported profits hence it is important to control for these expenditures. While Schmalensee (1989a; p965) reports that firms of different sizes appear to adopt different accounting practices and therefore the extent of bias may be correlated with market share there is no similar finding for differences between industries of differing levels of concentration.

The results reported in the chapter include measures of both advertising expenditure and the growth rate of firm size in the estimation. A potential problem that remains with this study is our failure to control for capital intensity due to the absence of data. In addition we must be aware, in interpreting the short run results of the later chapters, that one of the determinants of the time path of small firms profits is their accounting practices. However, as suggested earlier, the problem is substantially lessened when considering the long run performance of firms which is less susceptible to these differences in accounting practices.

II. The relation between theoretical framework and empirical specification

Steedman (1992) has provided a general critique of the absence of vertical linkages from industrial economic theorising. Steedman adopts a Sraffian framework for considering some of the problems perceived with the Kaleckian approach to price determination. While the target is Kaleckians, the same point applies to any partial equilibrium approach to price determination. The first relevant point made[2] is that, when one considers price determination using a matrix of inputs and outputs, prices are determined not only by the mark up in the industry concerned, but also by the mark-ups in many (possibly all) of the other industries. In one sense, this is obvious and doesn't contradict much empirical work which is interested in determining prices relative to costs and not the price level itself. The second point Steedman makes denies the usefulness of the notion of a vertically integrated industry. "The 'vertically integrated' industry is not an actual

[2] There are a number of interesting points made in the article which are not directly relevant to the current issue.

industry but, rather, a theorist's hypothetical construction, consisting of various bits of actual industries. The 'vertically integrated' bicycle industry, for example, contains (most of) the actual bicycle industry and bits of the steel industry, the rubber industry, the chemical industry, the electricity industry, etc." (ibid. p137) Steedman therefore suggests that dealing with vertically integrated industries is unhelpful and should be abandoned. This appears to be an extreme stance and not a particularly interesting one. All industries are linked by some vertical structure and we have to incorporate this into our analysis somehow. Making certain formalisations concerning the relationships between the vertically related industries enables us to focus more precisely on the bicycle industry, for example, in isolation from the influences brought upon it by the steel, rubber and chemical industries. What alternative have we if we want to reduce very complex phenomena to a tractable level? It is perhaps worthy of note that Steedman's algebraic example (ibid. p 138) that purports to cast doubt on the validity of vertically integrated concepts is not as damning as appears to be suggested. In the example the price level is determined by the equation, $p=(u+pA)(I+m)$

where p is a vector of prices of both inputs and outputs (within the input-output framework adopted all commodities potentially can be both inputs and outputs) u is a vector of direct unit non material costs, A is matrix of input coefficients, I is an identity matrix and m is a vector of mark-ups. Vertically integrated direct unit costs are therefore defined by $u^*=u(I-A)^{-1}$ and $p=u^*(I+\mu)$, where the μ is the vector of vertically integrated mark-ups. Following from this, Steedman shows that $\mu>m(I-A)^{-1}$ so that "one does not obtain the vertically integrated mark-ups, μ, simply by vertically integrating the mark-ups, m" (ibid. p138). However, if we define the mark-up not as the ratio of price to direct unit cost, as Steedman does, but as the price cost margin so that $m'=(p-(u+pA))/p$ and $\mu'=(p-u^*)/p$ then $\mu'=m'(I-A)^{-1}$ and one does indeed obtain the vertically integrated price cost margins, μ', simply by vertically integrating the price cost margins, m'.

However, having rebutted Steedman's criticism on one level, one should not necessarily draw the conclusion that using these vertically integrated price cost margins is a valid procedure. The foregoing discussion may appear slightly arcane, but it is useful because it is illustrative of two potential problems with obtaining vertically integrated margins that we have to consider in this chapter. Firstly, the relation is derived from a system of fixed input output coefficients, there is no possibility for substitution between inputs. This derives from the assumption that the A matrix is fixed and invariant to the relative prices of commodities. In reality, we have to address the likelihood of input substitution as a result of changes in the price of raw materials. Secondly, the vector of prices is determined as a system, therefore this form of determination of margins is inconsistent with a partial equilibrium determination of margins in which firms act as price takers for material input prices. To be consistent with the framework, firms must take into account the effect their own price decisions will have upon the pricing decisions of material input producers. Both of these important points will be explored below in more detail within a more conventional theoretical setting of oligopoly price determination.

One aspect of this framework which is not considered by Steedman is the potential dynamic factors which can provide the link between a framework in which input prices are assumed to adjust to firm decisions, and a framework in which firms act as price takers for inputs. For example, the effects of a change in demand at the industry level may influence output prices but in the common case of some lag in production the input prices will have been determined in the previous period(s) and therefore will not adjust[3]. This point, which is akin to the one of the issues raised by Coutts, Godley and Nordhaus (1978), will not be explored further here, where it would introduce additional complexity, but will be considered with the other dynamic aspects of the analysis in Chapter 5.

II.i. Establishing a theoretical framework and the rationales for an alternative margin

Under the assumption of constant marginal cost the Cowling and Waterson (1976) framework yields a relationship between price cost margin and market share at a firm level of the form,

$$\frac{p - avc_i}{p} = \frac{\Pi + F}{R} = \frac{s_i(1+\lambda_i)}{\eta} \qquad (1)$$

[3] A paper considering the effects of production lags on price inertia at the macroeconomic level for this precise reason is Lindbeck and Snower (1994). One should note however while this paper examines price sluggishness it does not examine any variation over time in the Lerner index only in the price level.

where s_i is the market share of firm i i.e. y_i/Y, λ_i is the conjectural parameter $\delta(Y-y_i)/\delta y_i$, η is the absolute value of the industry price elasticity of demand and avc_i is the average variable cost of firm i. The definition of the numerator of the left hand side is profits plus fixed costs which, when specifying price cost margins, is usually defined as gross output minus the cost of material inputs minus labour costs[4]. The denominator is sales revenue or gross output. Using net as opposed to gross output for the denominator is equivalent to multiplying the left hand side by the ratio of gross to net output. A crude but standard measure of vertical integration is the ratio of net output to gross output (see Scherer (1980) p79) so the adjustment for vertical integration is simply multiplying the conventional price cost margin by the reciprocal of a measure of vertical integration.

This approach is relatively unsatisfying as the precise reasoning and implicit assumptions about the role of material inputs in the determination of price cost margins which lie behind the adjustment are not clear. The reasoning is partially attributed to Cowling (1982) by Conyon and Machin (1991b), however Cowling's argument appears to be slightly different. In particular when Cowling talks of vertical integration in this context (1982; p161-2) it is only a limited form of vertical integration i.e. the elimination of intra industry transactions. As is shown by Cowling, the existence of intra industry transactions will lead to problems of aggregation at the industry level because while these are netted out of the numerator, because sales enter positively and material inputs negatively, so that intra industry transactions cancel each other out, the denominator will increase with a rise in intra industry transactions as it merely measures the aggregated sales. To free the measure from this bias we would need to control for the degree of intra industry transactions. In this case, the correct measure of the denominator will be gross output minus inputs purchased from within the industry. Unfortunately, this data is not generally available, or at least not for most years. To relate this to the original debate, using net output as the denominator implies, according to Cowling's framework, the very strong assumption (which clearly does not accord with reality) that all inputs are purchased from within the industry. Since

[4] As we are deducting variable costs, labour costs are ordinarily defined as production worker wages and not salaries which are regarded as overheads.

this cannot be adequately sustained, the use of the alternatively defined margin must be rationalised in a different way, or be used as means of comparison alongside the conventional margin, which is how it is used by Cowling.

The second approach which is adopted by Haskel and Martin (1992,1994) is to start in their modelling framework with a value added production function. Clearly, Haskel and Martin do not believe that material inputs are unimportant for the analysis but the implicit assumption is that material inputs can be netted from the production function at no significant cost. However, both theoretically and empirically, it is not clear that working with a value added production function is possible without cost. On an empirical level, significant differences do exist between different specifications of the margin, as is evidenced in Conyon and Machin (1991b), and assuming a value added production function does not help us understand why these differences exist. On a theoretical level, as Rotemberg and Woodford (1993) show, a well defined value added function is only possible under the assumption of price taking behaviour in both the product and factor markets, an assumption which is not particularly relevant to the circumstance of our investigation. In the case of Cournot competition, the implied value added production function will contain within it an element of the degree of market power so it is not neutral to the focus of our investigation[5]. In particular, by working below with a very general functional form, which includes a class of separable functions, we can reject Haskel and Martin's (1994, p29) assertion that in assuming a gross output function which is separable, movement to a value added production function is costless.

A third approach to the problem has been put forward in a recent consideration by Dickson (1994). Addressing the specific issue of the difficulty of obtaining a significant correlation between margins and concentration, Dickson considers the role of the elasticity of demand as an omitted variable in price cost margin concentration regressions. Dickson compares the case[6] of

[5] In essence this also follows from the argument made above in relation to Steedman (1992) concerning the consistency of price taking within input markets with the assumption of a neutral vertically integrated measure of the price cost margin.

[6] The following exposition is taken from Dickson (1994, p330-331) but some of the symbols have been changed to conform with the rest of the chapter. In addition to the three problems mentioned below there

firm's marking up value added prices p_V over per unit value added costs ac_V with the conventional case of firms marking up finished goods prices on per unit costs which include material inputs. In the former case the equilibrium solution is characterised by (2) and also $\eta=\eta_V(V/S)$ where η is the conventionally defined elasticity of demand.

$$\frac{p_v - ac_v}{p_v} = \frac{\pi + F}{V} = \frac{H}{\eta_v} \qquad (2)$$

In this case, using the standard approach, the omitted elasticity variable from the regression, η, will actually be a a function of the value added elasticity η_V and V/S. If V/S is correlated with the measure of concentration, this will induce bias in the estimates from the original regression. Dickson's concludes that, because information on elasticities is not available, the value added measure should be used to control for "stage of production effects" and he therefore focuses on the "effective rate of market power" and not the "nominal rate of market power". This conclusion seems misplaced for three reasons: firstly, the notion of controlling for "stage of production effects" would appear *ad hoc* and not directly related to the forgoing discussion, which is related to elasticities; secondly, the existence of bias is conditional on the assumption that firms mark up solely on so called value added costs rather than costs including material inputs[7], this would appear to be a matter for empirical resolution rather than for assumption; thirdly, if one wants to control for the hypothesised omission of V/S, the empirically correct response would appear to be to include it as a right hand side regressor along with other "control" variables such as advertising intensity. This response has the additional advantage that the issue of whether firms mark up on conventional costs or just value added costs is not prejudged. A further issue regarding the econometric estimation of such relationships is the question of time variation. If the stage of production effects described by Dickson are constant over time then estimation of a panel including fixed effects will eliminate the omitted variable bias because all industry specific effects which do not vary over time are extracted. Dickson does not discuss this aspect at all, but there are two views. Cowling and Waterson (1976) proceed on the assumption that the elasticity of

---

is an additional, potentially more damaging query with the method Dickson adopts. The supposed difference between margins arises when firms maximise value added profits i.e. p_Vq_i-c_{Vi} rather than conventionally defined profits pq_i-c_i. However, under most definitions p_V=p-$p_m(m/q)$ in which case the two measures of profit are identical. As Dickson does not define his terms precisely, it is not clear whether the proposed change is qualitative or merely one of definition.

[7] It is also conditional upon the correlation between concentration and V/S.

demand is constant over time and therefore estimation in the form of first differences or, in their case, year on year ratios, eliminates the effect of demand elasticities. On the other hand, it has been argued (see, Harrod 1936) that this assumption of the constancy of demand elasticities may be incorrect and therefore it may be possible for differences in demand movements to effect price cost margins. We attempt to control for this element in the estimations reported below.

II.ii. Developing an alternative theoretical framework

The previous section has critically considered the alternative rationales for the use of the alternative margin and found each wanting. This chapter will seek to yield a superior resolution of this issue by considering other theoretical relations. A principal argument of this chapter is that it is only by looking at the treatment of material inputs in the pricing decision and, in particular, imperfect competition in the factor market, that we find a satisfactory reason for adjusting the margin, but that it appears preferable in this situation to control for this on the right hand side of an empirical equation as opposed to the left hand side. We assume a production function where there are two inputs to the production process materials and labour. This is merely for simplicity of exposition and labour may be considered to be a general term for internal inputs while materials may be considered to include payments for subcontracting etc. and therefore covers all bought in elements. The production function for the firm is of a general form $y = f(l,m)$ where y is output, m is material input, l is labour input. This function is assumed to be continuous and differentiable but no further restrictions are placed upon it. The industry faces an inverse demand function $p(Y)$ where $Y=\Sigma y$ over the firms in the industry. The profit function for each firm is $\pi = py - p_m m - wl$, where p_m is the price of material inputs and w is the wage. It is assumed initially that p_m and w are regarded as fixed by the firm. The firm chooses m and l to maximise profits. Using Euler's theorem and rearranging the first order conditions yields,

$$\frac{py - p_m m - wl}{py} = (1-\gamma) + \gamma\frac{s_i(1+\lambda_i)}{\eta} \quad (2)$$

where γ is the degree of homogeneity of the production function. Therefore, if the production function is linearly homogeneous, (2) reduces to (1) and the conventional definition of the margin with gross output as the denominator is the implied correct specification. There appears no need to correct for vertical integration[8] and this applies whether the function is separable or not.

One possible treatment of materials costs is that they are not marked up on at all, but are treated as a fixed cost in the short run at least. In this case introducing a short run production function which depends only upon the labour input y=g(l) and maximising with respect to labour input yields.

$$\frac{py-wl}{py}=(1-\gamma)+\gamma\frac{s_i(1+\lambda_i)}{\eta} \qquad (2a)$$

where γ is the degree of homogeneity of the production function g(l). This does imply a respecification of the price cost margin but alters the numerator rather than the denominator.

A related alternative to this is if it is assumed that material inputs are used in fixed proportions to the other inputs. This would appear to be a reasonable approximation in many cases although it does abstract from the firm's decision on the degree of vertical integration. We assume material input used in a constant proportion of s_m=m/y and a Leontief production function of the form[9]

$$y=\min\left(\frac{g(l)}{1-s_m},\frac{m}{s_m}\right).$$

Assuming efficient use of inputs, the profit function becomes $\pi=py-wl-p_m s_m y$. Maximising with respect to labour input and rearrangement of the first order conditions in this case yields,

$$\frac{py_i-wl_i-p_m m_i}{py_i}=(1-\gamma)\frac{py_i-p_m m_i}{py_i}+\gamma\frac{s_i(1+\lambda_i)}{\eta} \qquad (3)$$

the right hand side of which collapses to the right hand side of (1) if the function g(l) is linearly homogeneous. Clearly, this does not justify the use of a value added deflated margin in empirical relations although it does indicate that there may be a role for the Scherer measure of vertical integration as an explanatory variable if the function g(l) is not linearly homogeneous. As a point of note, it does not appear to be a readily justifiable assumption that g(l) is linearly homogeneous as it would imply a fairly contradictory "fixed proportions" production function which, contingent

---

[8] Vertical integration is here taken to be vertical integration over all inputs, to avoid confusion with the Cowling argument which is orthogonal to the following discussion which should more correctly be referred to as correcting for intra industry transactions.

[9] This particular functional form is taken from the discussion in Rotemberg and Woodford (1993)

upon a particular value of bought in inputs, exhibited constant returns to the labour input. The expectation of a pure fixed proportions technology would be that g(l) would in fact be homogeneous of degree zero.

We now investigate the consequences of imperfect competition in the factor market, in particular the market for bought in goods and services as opposed to internal activities. Initially, we can model some degree of monopsonistic power in the market for bought in inputs. We use the original general form of the production function, but the price of inputs p_m depends upon the quantity demanded by the firm, so the profit function is of the form $\pi = py - wl - p_m(m)m$. Again maximising with respect to m and l and using Euler's theorem yields,

$$\frac{py - wl - p_m m}{py} = (1-\gamma) + \gamma \frac{s_i(1+\lambda_i)}{\eta} + \varepsilon \frac{p_m m}{py} \qquad (4)$$

where $\varepsilon = \frac{\delta p_m}{\delta m}\frac{m}{p_m}$ is the reciprocal of the elasticity of derived demand for bought in inputs. This elasticity will clearly depend upon the degree of monopsony power being exerted by the downstream firm. This suggests that a correct specification for this framework again requires that the denominator is gross output but that the ratio of material inputs to gross output should enter as a right hand side variable from which we may be able to determine the extent of monopsony power.

The next step we wish to take is to introduce the bargaining of material input prices and, in particular, the efficient bargaining of material input prices. While it is not uncommon to consider bargaining over wages, it may appear strange initially to consider bargaining over input prices. However, this is merely the framework that is encapsulated in the notion of bilateral monopoly where a monopsonist and monopolist face each other in the market for material inputs. The standard result of this framework is that while the optimal level of material input m^* may be agreed upon, the price at which that volume of inputs is transacted is indeterminate[10]. By introducing a Nash bargaining framework, we may solve explicitly for the input price in terms of the relative bargaining strengths of the upstream and downstream firm respectively. Specifically,

[10] See Waterson (1984) Ch 6 for a discussion of this.

if we take the profits of the upstream material input producer to equal $\pi_u=(p_m-c_m)m$, where, for simplicity, we assume a constant marginal cost of input production and the profits of the downstream firm are as defined above. The Nash bargain is of the form $N=\pi_u^{\sigma}\pi_d^{1-\sigma}$ [11], where σ is the bargaining power of the input producer so that, if σ=0 the input producer has no bargaining power at all while, if σ=1 the end producer has no bargaining power. Maximising the Nash bargain with respect to m and π_d with respect to l, then substituting, yields the relation,

$$\frac{py-p_m m-wl}{py}=(1-\sigma)(1-\gamma)+(1-\sigma)\gamma\frac{s_i(1+\lambda_i)}{\eta} \qquad (5)$$

It is readily apparent from (5) that if σ=0, i.e. the input supplier has no bargaining power, then (5) collapses to (2). If σ=1, i.e. the end producer has no bargaining power, the input supplier is able to extract the full surplus, hence the price cost margin is equal to zero.

Having derived this relation in (5) we can now make explicit the nature of the assumption implied when "controlling for the degree of vertical integration". If it is assumed that $\sigma = p_m m/py$, i.e. the ratio of material input costs to revenue, then (1-σ) is equal to the ratio of value added to revenue. By substituting this into (5), multiplying through by revenue and dividing by value added, we then derive a relation analogous to (2) but with value added as the denominator.

$$\frac{py-p_m m-wl}{py-p_m m}=(1-\gamma)+\gamma\frac{s_i(1+\lambda_i)}{\eta} \qquad (6)$$

This gives an indication of the nature of the assumption being made when using net output as the denominator, but it should be noted that this is a special case of only <u>one</u> of the frameworks we have looked at.

Clearly, the next question which must be considered is, how reasonable is the assumption that $\sigma = p_m m/py$? Under this assumption, if there are no material inputs so the downstream firm is completely vertically integrated, σ must be zero and also value added must be equal to revenue so the adjustment would appear appropriate in this case. The problem is that $p_m m/py$ is partly

[11] The implicit assumption is that the "outside option" yields zero profits. While this appears broadly sensible it is clear that if the bargaining power of the upstream firm becomes large an outside option for the downstream firm of producing the good itself, i.e. vertical integration, may yield positive profits and become desirable.

determined by σ, itself because even if m/y is a fixed constant, p_m/p will depend upon relative bargaining strength. While this is not a unique problem (everything within this framework is determined simultaneously apart from, γ and η, which are regarded as exogenous), it likely to be of interest to investigate the causes of changes in σ. Relevant influences will be the degree of both buyer and seller concentration in the respective factor markets, but will also depend upon the particular technological features of production, for example, relating to asset specificity in the transaction cost framework of analysis. It is because of these influences that the option of making the rather crude assumption that $\sigma = p_m m/py$ is not an attractive one.

II.iii. Consequences for the empirical specification

In the previous section, we established that under a number of circumstances there is a role for the ratio of bought in input costs to sales as an explanatory variable in a price cost margin regression. In this section, we explore the interpretation and specification of empirical relations. Clearly, the omission of this ratio from the right hand side only has implications for the interpretation of the coefficient on market share if it is correlated with the market share term. In this case, the coefficient is likely to be biased according to the standard framework of omitted variable bias. As we have discussed above it is likely that the ratio is likely to be correlated with the degree of concentration in the industry in the case of monopsony and bargaining and, if these models are correct, some bias will indeed be induced.

Using relation (5) demonstrates another reason why the link between concentration and price cost margin may be biased downwards when failing to take into account such bargaining even in the absence of the omitted variable effect. A linear estimating equation of price cost margin on market share will be of the form. $PCM_i = \delta_0 + \delta_1 Mshare_i + u_i$ where $\delta_0 = (1-\sigma)(1-\gamma)$ and $\delta_1 = [(1-\sigma)\gamma(1+\lambda_i)]/\eta$. Therefore as σ rises the coefficient on market share falls so it is possible that a positive and significant link may fail to be established.

If we examine a basic regression of price cost margin at the industry level on concentration and the ratio of purchased inputs to sales, the coefficients can be directly interpreted in relation to four of the models. We have an estimating equation[12] of the form,

$$PCM_i = \delta_0 + \delta_1 Conc_i + \delta_2 (\frac{M}{R})_i + v_i \quad (7)$$

where v_i is an error term, $M=p_m m$ and $R=py$. Figure 1. indicates the expected values of the different coefficients in the different models.

Figure 1 Predicted coefficients for the different frameworks

	δ_0	δ_1	δ_2
Standard Model	$1-\gamma$	$\gamma(1+\lambda)/\eta$	0
Quasi Fixed Material Inputs	$1-\gamma$	$\gamma(1+\lambda)/\eta$	-1
Leontief Production Function	$1-\gamma$	$\gamma(1+\lambda)/\eta$	$\gamma-1$
Monopsony	$1-\gamma$	$\gamma(1+\lambda)/\eta$	ε

Therefore in all of these frameworks we expect similar coefficients on the constant and market share terms. The difference of interpretation occurs with the second term which is predicted to be zero in the first case, to be exactly equal to -1 in the second case and may be positive or negative in the third case. It is expected to be negative in the monopsony case given the conventional assumption that the elasticity of derived demand is negative.

In terms of the specification for the bargaining model, we are really interested in finding a proxy for bargaining strength which is, of course, unobserved. By allowing a more structured functional form for the bargaining parameter, or by trying to identify the effect of concentration on monopsony, we can further develop the empirical specification. Therefore if we model σ in a way

[12] Clearly additional 'control' variables may be warranted but are omitted from the exposition for simplicity.

that $\sigma = \theta_0 + \theta_1(M/R)$ or if we model the derived demand parameter explicitly so that $\varepsilon = \phi_0 + \phi_1 CONC$ we obtain the interactive specification.

$$PCM_i = \delta_0 + \delta_1 Conc_i + \delta_2 (\frac{M}{R})_i + \delta_3 ((\frac{M}{R})_i * Conc_i) + v_i \qquad (8)$$

The additional element which it is important to explore, given the focus of many recent studies, and which has not been addressed until now[13], is the extent to which the alternative definitions of the margin have different cyclical properties. Clearly, any difference in the cyclical fluctuations in the margins depends upon the cyclical movements in the ratio of material inputs to final sales. In considering, *a priori*, any movements in this ratio, it is perhaps useful to separate out movements in the unit input requirements m/y, and movements in relative prices p_m/p. There would appear to be little reason to expect unit input requirements to vary much according to business fluctuations except as a second order effect of substitution between labour and bought in inputs as their relative prices change. Given the focus here on fixed coefficient technologies and, additionally, given that substitution effects are more likely to occur in response to long term rather short term relative price movements, it would appear to be a reasonable working assumption that changes in unit input requirements will largely be unaffected by business fluctuations. The sole point which could be made is that if in times of boom firms find themselves capacity constrained, they may choose to subcontract out some, or a larger proportion, of their work. In this case one would expect m/y to rise in booms and fall in slumps as firms adjust to these capacity constraints.

In analysing the relative movements of p_m/p it is necessary to separate out further the semi manufactured inputs and raw material inputs. As many of the manufactured goods sold will be both outputs and inputs, there would appear to be no immediate consequences for the ratio p_m/p. In the case of raw materials there is an established literature on the determinants of prices in the business cycle. Kalecki (1939) argued that the prices of raw materials were essentially demand determined and therefore fluctuated more widely than finished goods prices. If this is the case

[13] Haskel and Martin (1994, p29) indicate that their result on the effect of capacity constraints on price cost margin is unaffected by differences in the definition of margin. The other studies do not examine this issue at all.

p_m/p will fall in recessions and rise in booms. Combining these two elements, in both cases the expectation, although small in the case of m/y, is that M/R will move pro-cyclically and that, therefore, V/R will move counter-cyclically. What are the implications of counter-cyclical movements in V/R for the movements in price cost margins? If we assume that price cost margins, using the conventional gross output denominator definition, display no cyclical movement[14] then in recessions net output denominator price cost margins will fall as V rises relative to R and this margin will rise in booms. The implication is, therefore, that price cost margins defined using the net output denominator will tend, if our hypotheses are correct, to be procyclical relative to the conventional margin. A further point is that if the hypothesis with regard to subcontracting as a response to capacity constraints is correct, and if such subcontracting occurs at least partially within the industry, then the price cost margin measure adjusted for intra industry transactions will similarly respond differently to the cycle. The bias will be in the same direction tending to make the margin adjusted for intra industry transactions procyclical relative to the conventionally defined margin.

In addition to the empirical tests of the effect of some form of imperfect competition in the product market, we also wish to investigate what may be termed the "Cowling effect" of intra industry transactions. In order to do this we need to make use of the UK input output tables which indicate the extent of intra industry transactions. As has been noted above, the problem with this empirical exercise is that the input output tables of the UK are only published infrequently, namely they have been published for 1979, 1984, 1989, 1990 and 1991[15]. Nevertheless with the information available for these latter years, it is possible to use a first difference framework which removes fixed effects in order to identify if the use of a margin adjusted for intra industry transactions produces empirical results which are quantitatively different from those obtained using the standard margin. It is not possible, however, without using some form of interpolation concerning the rate of intra industry transactions, to create a full

[14] Clearly this is not necessarily the case and is assumed solely because we are interested in the relative cyclicality of the different margins.
[15] The plan is that they will be published annually from 1989 but at the time of writing these are the only years available which make use of the SIC 1980 classification.

panel from 1980 onwards which would facilitate direct comparison. Interpolation does not appear a particularly attractive approach because, following the above discussion, it is difficult to assert that if significant changes in the extent of intra transactions are occurring, that they are a smooth secular phenomenon, as opposed to a phenomenon which is sensitive to particular cyclical or other temporal factors. The use of data from the earlier years would also necessitate a smaller sample because changes in the industry classification make the matching process more difficult.

III. Results of the empirical analysis

The first empirical exercise is therefore to create the three alternative margins which are necessary to test the Cowling intra industry transactions hypothesis. The first margin is defined as net output minus operative wages over net output: the second margin is defined as net output minus operative wages over gross output. Both of these measures can be readily obtained from the Census of Production for the relevant years at the three digit level of aggregation. In using the input output tables to generate the measure of intra industry transactions, we need to perform a matching of the CSO industry categories used in this table with the three digit SIC classification. This matching proved possible at the three digit level of aggregation for 48 industries. A further 10 observations were added which were amalgamations of more than one three digit industry but which correspond to a single CSO input-output table industry. In this case, the measure of the third price cost margin was (Σnet output-Σoperative wages)/(Σgross output * IT) where IT is the ratio of purchases of inputs from within the CSO industry to total sales in that CSO industry. Concentration for these ten observations was calculated as a weighted average of the concentration ratio in the component three digit industries. The growth rate of sales was calculated as the first difference of logs of the total sales of the industries taken as a group.

Table 1 presents the summary data of the tightly defined sample of 48 industries and the whole sample of 91 industries (excluding miscellaneous categories and categories for which data was incomplete). This illustrates that, although in the tightly defined sample almost half of the SIC three digit industries have been discarded, the samples have very similar means and standard deviations so there does not appear any systematic bias in the selection of industries by this method. As a point of note, the mean value of IT across the whole sample was 6% but this varied between industries. It was effectively negligible in some industries but in other industries intra industry purchases amounted to over 25% of all sales.

In estimating the relationship between margins and concentration, further independent variables have been included as control variables. These are a measure of advertising intensity, designed to proxy the extent of product differentiation and potential entry barriers, and the growth rate of industry sales. Price cost margins may vary with demand conditions[16] and, as noted above, this effect may differ between measures of price cost margins. In addition to these measures of import share and distributors trading margins are included. This data is taken from the CSO input output tables. Import share proxies the degree of international competition while distributor's trading margins control for the degree of downstream market power. This latter point addresses the reason advocated for using the alternative margin by Dickson (1994). He suggests that the omission of the industry elasticity of demand will potentially bias results if it is correlated with concentration. As has been recognised by Waterson (1980), the derived elasticity of demand will depend on the extent of buyer (downstream) market power. Empirically, this downstream market power is proxied by Waterson either as the weighted average of buyer concentration or as the weighted average of downstream price cost margins. Using distributor's trading margins, therefore, seeks to control in this way for such downstream market power effects. In relation to the argument by Dickson, by incorporating this we are therefore trying to separate out the intra industry transactions effects from those potentially caused by omission of factors relating to the elasticity of demand. If the inclusion of distributor's trading margins matters for the establishment of a relationship between concentration and margins, then a part of Dickson's argument would be confirmed.

Table 2 contains the results of OLS levels regressions for the different specification of margins for 1989-1991. The use of these years is somewhat problematic. It is well recognised that the

[16] See Domowitz, Hubbard and Petersen (1986) and Haskel and Martin (1992) in addition to later chapters of this thesis for evidence of this.

relationship between concentration and price cost margins is weaker during periods of recession than periods of boom. Manufacturing output was stagnant through 1989 and fell in both 1990 and 1991 without picking up until the start of 1993. Therefore we may not expect to find a strong relationship between concentration and price cost margins using data from these years. The other point of note is that it is customary for concentration and advertising intensity and frequently other right hand side variables to be treated as endogenous within this kind of framework and therefore an instrumental variables approach to be used. There remains a problem with this, one that has been argued strongly in this context by Schmalensee (1989a), namely that valid contemporaneous instruments do not exist and therefore lagged values of variables must be used as instruments. If we were to do that in this case, we would lose at least one third of our sample, and depending on the result of instrument validity tests, possibly two thirds of the sample[17]. OLS is therefore used to maximise the available sample but with the acknowledgement of the potential problems therein.

The results from Table 2 in levels offer little support for the argument that controlling for intra industry transactions has an important impact upon the relationship. It is only with the net output margin that a positive and significant relationship is established. With both other margins, the point estimate is negative and the t statistic insignificant. Both advertising and distributor's margins have effects of similar magnitude and significance irrespective of the definition of the margin although the sign of the coefficient on distributor's margin is the opposite of that expected from Waterson (1980). Waterson found that greater downstream market power led to a more inelastic derived demand for the upstream firm which in turn led to higher price cost margins. In this case, downstream margins have a negative impact on upstream margins. This is more indicative of a framework where downstream distributors can squeeze upstream margins within some kind of bargaining context. Omission of the distributor's trading margin variable did not

[17] It could easily be anticipated that x_{it-1} is not a valid instrument hence x_{it-2} must be used instead. Indeed attempts were made to treat concentration as endogenous and used lagged values on which there is much information as instruments. However these regressions did not perform well, producing negative R squared statistics and thus appearing to reject these lagged values as valid instruments. Another possible way of getting around this problem would be to use lagged values of concentration in the OLS regressions. Estimations of this form did not produce substantially different results from those presented however and are therefore not reproduced here.

have any effect on the coefficients on the other variables and this would therefore count against Dickson's account of the difference between margins. Imports do not have a significant effect irrespective of margin used. This accords with previous evidence from the UK (e.g. Conyon and Machin (1991a)) which fails to find a strong effect for imports. The final point of note is that the coefficient on industry demand is only significant for the net output denominator margin, this would appear to confirm our expectation concerning the relative cyclical movements in input prices. There are no important differences between the smaller and larger samples, suggesting one does not lose much economic information by conflating SIC groups in the fashion described above, but a wider coverage of manufacturing industry is not possible.

Table 3 repeats the regressions of Table 2 but in first differences so as to remove any unobservable industry specific effects which could be correlated with the right hand side variables. As stated in section II.i. the removal of fixed effects will control for the omission of any non time varying control variables whether observable or not. The results are not easy to interpret because in all cases the effect of concentration on margins is insignificant. As suggested earlier this could be due to fact that the time period for which the data was available is a period of recession. While the point can be made that the point estimates and level of determination are similar for the net output margin and the margin adjusted for intra industry transactions and both have a stronger level of determination than the conventional margin, the overall level of determination is too poor to make any strong conclusions. The only consistent results are for the growth rate of industry sales where in this case all margins are positively related although the magnitude of the effect is greater for the net output denominator margin. We have only examined data for three years and this is, therefore, not a particularly comprehensive test of the hypothesis[18]. However, the evidence from Table 2 indicates that using the net output denominator to control for intra industry transactions is inappropriate and that failing to correct for intra industry transactions will not seriously distort empirical results. The results from Table 3 qualify this to some extent but are too poorly determined for any substantive conclusions.

[18] In particular looking at only two years does not enable one to look at the secular trend in intra industry purchases.

The second part of the empirical investigation involves a re-specification of the price cost margin concentration relationship to account for monopsony or bargaining in the market for bought in inputs. Table 4 gives the means and data descriptions for the variables used in this empirical section. Table 5 reports results for two way fixed effects estimates[19] using a panel of census industries from 1984 -1991 and including yearly dummy variables to control for aggregate cyclical or secular demand effects. Regression 1 uses the alternative net output denominator margin as the dependent variable. Regression 2 uses the more conventional gross output denominator margin. Regressions 3 and 4 are analogous to the specifications given in equations 7 and 8 in section II using the gross output denominator margin. The first difference instrumental variables estimator from Arellano and Bond (1988) is used as all of the right hand side variables are considered to be endogenous. The growth rate of industry sales and a proxy for advertising intensity are included as control variables on the right hand side to account for any industry specific demand effects or the effects of product differentiation or advertising related entry barriers. Both of these variables are treated as endogenous as is customary in industrial organisation studies of this kind (see Schmalensee, 1989a).

Regressions 1 and 2 merely illustrate the conventional wisdom that the a net output margin will obtain a larger and better determined coefficient on concentration than the gross output denominator margin. Neither of the control variables are significant in these or any of the other regressions. In the case of demand growth this is likely to be a result of the inclusion of the yearly dummies which will absorb economy wide demand fluctuations, these dummies are always jointly significant. In the case of the advertising proxy, this would appear to indicate that the effect of advertising is largely a fixed effect with relatively little variation over time. There are no problems of second order serial correlation with the estimated regressions. Regression 3 yields an improvement in the level of determination of the coefficient on concentration although a decline in its absolute value relative to the standard case. The coefficient on M/R in this regression is

[19] Estimates in levels were attempted as well but these indicated strong serial correlation which could indicate some form of mispecification such as omission of unobservables. These results are therefore not reported.

negative and significant. Referring back to Figure 1 this fits the monopsony story and the fixed coefficient approach if there are widespread diseconomies of scale in the use of labour. In fact we cannot reject the hypothesis that the coefficient on M/R is equal to -1. A value of -1 for this coefficient is predicted by the quasi fixed material input framework. Such a value is also consistent with an interpretation that the function g(l) in the fixed proportions production function is homogeneous of degree zero. As discussed earlier, this would imply a "pure" fixed proportions technology. However that the function g(l) in the fixed proportions framework is homogeneous of degree zero also, noting from Figure 1, implies that the coefficient on concentration should be equal to zero. This is rejected by the data. Only the monopsony and quasi fixed material inputs approaches are consistent with the findings. This would imply that there may be a problem of specification in conventional estimations which arises from their failure to take into account either the existence of monopsony power or the precise treatment of material inputs. We would argue that including the ratio of bought in inputs to sales as an explanatory variable is helpful in resolving this problem.

In principle, separate effects could occur if we differentiate between bought in goods and bought in services[20]. As, in the data, we can break down bought in inputs into these two separate components, a further investigation which included these in the estimation either separately or jointly was made. Both in terms of volume and in terms of statistical estimation the role of bought in goods dominated that of bought in services. In terms of the theory, this would imply that firms have greater monopsony power over suppliers of goods than over the subcontractors they use. The separate effects are not well determined perhaps due to collinearity and therefore the results are not reported.

The attempt to further discriminate between the bargaining approach and the other theories using the interactive specification is less straightforward. Regression 4 in Table 5 indicates an increase in the explanatory power of the right hand side variables, as evidenced from comparing the Wald

[20] These are industrial services such as subcontracting rather than non industrial services which could be regarded as overheads.

Statistics from (3) and (4), but no separate effect of either concentration or the interaction term can be discerned. This could be ascribed to the existence of some degree of collinearity as frequently occurs when using an interactive specification. A Wald test of the joint significance of concentration and the interaction term gives a value of 5.26 and the test is distributed as χ^2 with 2 degrees of freedom. This value is significant at the ten per cent level but not at the five per cent level. Thus, there is limited, but not particularly compelling, evidence of collinearity and we have to conclude that the interactive specification is not to be preferred. From a theoretical point of view, this would appear to count against a bargaining framework explanation of the relationships. However, as the evidence was not conclusive, a further attempt to test this was made and is discussed below.

The alternative approach to this test involves a logarithmic transformation of the estimating relationship, so that a log linear specification is adopted in preference to a linear model. As the formal theoretical framework is essentially non linear or at least is only approximately linear the estimation of log linear relationships between price cost margins and concentration is not uncommon. There are a number of disadvantages to the use of log linear specifications however. It is unclear what the original functional form would be that would indicate a purely log linear relationship between the variables. In particular, the inclusion of a demand growth or advertising sales ratio control variable leads to problems of interpretation. If either of these variables takes a value of zero, i.e. demand is constant, or there is no advertising in the industry then the implication of the log linear specification is that the price cost margin must also take a value of zero. In terms of the theory driving the relationship, however, stagnant sales should not necessarily imply zero price cost margin.

Having noted the problems which arise with this form of specification there is also a simple way of testing the implicit assumption about the appropriate specification of dependent variable in a log linear framework. To see this, note that the alternative price cost margin is defined as $\frac{\pi+F}{V}=\frac{\pi+F}{S}\frac{S}{V}$, taking logs of which becomes $\ln(\frac{\pi+F}{S})-\ln(\frac{V}{S})$. Within a log linear specification the appropriateness of the implicit empirical assumption simply becomes a matter of estimating a log linear model using the conventional gross output denominator price cost margin as the dependent variable. By including the log of the ratio of net output to gross output as an independent variable, a t test to see if the coefficient on this variable is significantly different from unity is sufficient. This approach was tried but with the reservations expressed above concerning the use of log linear specifications borne in mind.

The results from these estimations are displayed in Table 6. In none of the estimations is there evidence of significant second order serial correlation and the Sargan test of instrument validity does not raise any problems. Therefore, despite the theoretical misgivings about a log linear specification, there do not appear to be any statistical problems of mispecification. The first column contains a simple log linear specification using the net output denominator price cost margin, the second column presents the same specification but using the gross output denominator margin. These results merely confirm the general finding that a positive and significant relationship with concentration only exists when using the net output denominator margin. Column three replicates column two, with the addition of the log of the ratio of net output to gross output. In this case, the point estimate on the added variable is 1.3968 with an estimated standard error of .1358. Therefore, we can reject the hypothesis that the coefficient is equal to unity at the one per cent level of significance. The implication is that we can also reject the hypothesis that the use of the net output denominator margin is statistically correct for this data set with a high degree of confidence. As stated earlier this does not necessarily mean that it is either theoretically correct or empirically appropriate for all data sets, but for the current data set it does not seem appropriate. Column four of Table 6 replaces the log of the ratio of net output to gross output with the log of the ratio of bought in inputs to gross output. This is largely included for purposes of completeness and comparison but it further confirms that the inclusion of this ratio tends to increase the significance of concentration in these estimations.

IV. Conclusion

In terms of practical work this chapter examines formal rationales for the approach adopted by Conyon and Machin (1991b) and Haskel and Martin (1992) in specifying price cost margins. The chapter clearly sets out the assumptions which are implicit in the adopted frameworks over the treatment of bought in inputs and finds little theoretical justification for the use of the alternative margin. The chapter finds that only under very specific theoretical assumptions is a net output denominator margin appropriate, i.e. an efficient bargaining framework between input suppliers and input purchasers where the bargaining strengths depend upon the importance of bought in inputs as a proportion of final sales. The chapter shows that, in general, the appropriate theoretical correction for the degree of vertical integration depends upon the specific framework hypothesised for the relationship between end user and input provider. However, while there seems to be little justification for the alteration of the left hand side variable there are a range of circumstances when a version of the ratio of bought in inputs to total sales should be included as a right hand side variable. Additionally the chapter has brought out a different aspect of the adjustment to control for vertical integration, namely the adjustment of price cost margin measures for intra industry transactions.

Empirical support for the theoretical conclusion is not particularly easy to come by. The proposition is that the ratio of bought in inputs to sales should be included as a right hand side "control" variable in price cost margin regressions as this simultaneously accounts for a range of theoretical models. By the same token this leads to problems with regards interpretation of a significant effect of the control variable as no specific theoretical model is preferred. However, the empirical evidence we have presented indicates there is a role for such a control variable within the data set we have compiled. Interestingly, while the theoretical framework which implies a redefined price cost margin has been rejected on the grounds that it is based upon narrow theoretical assumptions, it cannot be rejected empirically within the current data set with a high degree of confidence. In terms of prescription for future empirical work, the implication is that as a matter of course control on the right hand side of an estimating equation is to be empirically and theoretically preferred if the treatment of bought in inputs is unknown. However, if knowledge about the nature of bargaining is available or, if the relevant assumptions are deemed acceptable, the use of an redefined price cost margin may be justified.

The evidence presented here suggests that, from a theoretical standpoint, the assumptions of the framework necessary to justify the alternative margin are sufficiently restrictive to be excluded from the ordinary run of things. A statistical test of the assumption was proposed however and for the current data set the assumption was rejected on this test. An alternative specification was suggested which nests a series of theories. The results presented here are consistent with two possible theoretical explanations: firstly, that firms possess monopsony power over suppliers of inputs of goods and services; secondly, that the production function used by firms is of a fixed proportions form which is not admitted by the conventional theory. It is not clear whether any further attempt to distinguish between these two positions would be possible. This chapter has therefore suggested a rejection of one approach to specifying price cost margins regressions and the adoption of an alternative less restrictive specification which is preferred both on statistical grounds for the current sample and in terms of theoretical suitability.

In addition to these conclusions, the empirical results have cast doubt upon the need to control for intra industry transactions when specifying the price cost margin. These results are however presented with the caveat that the justification for adjusting for intra industry transactions (Cowling, 1982) refers to secular trends while the data available for testing are from a short sample period.

Table 1 Data Description, Intra Industry Transaction Data 1989-1991

1a.

Variable	Obs	Mean	Std. Dev.	Min	Max
Conc	273	.429473	.2304068	.0854078	.9922145
PCM1	273	.7226212	.0949878	.4940711	.9647428
PCM2	273	.3211874	.0892142	.1248008	.5948678
IT	273	.0603163	.06106	0	.26047

1b.

Variable	Obs	Mean	Std. Dev.	Min	Max
Conc	144	.4223217	.2261698	.0854078	.9922145
PCM1	144	.739362	.095897	.5462555	.9218515
PCM2	144	.3324778	.0958325	.1248008	.5948678
IT	144	.0566883	.0560254	0	.26047

1c.

	PCM1	PCM2	PCM3	Conc
PCM1	1.0000			
PCM2	0.3640	1.0000		
PCM3	0.3963	0.9798	1.0000	
Conc	0.3456	-0.1333	-0.1047	1.0000

Notes:

1. Conc is the five firm concentration ratio by sales; PCM1 is the net output denominator price cost margin, (net output-operative wages)/net output; PCM2 is the gross output denominator price cost margin, (net output-operative wages)/gross output; PCM3 is the price cost margin adjusted for intra industry transactions, (net output-operative wages)/(gross output*(1-IT));IT is the ratio of intra industry transactions in industry final demand.
2. 1a lists the variable descriptions for a full sample of 91 industries. 1b lists the variable descriptions for the 48 industries capable of a close matching in the UK CSO input-output tables. 1c Gives the raw correlations of the three price cost margins in the narrow sample of 48 industries.
3. Data is taken from Census' of Production, Business Monitor PA1002 for 1989, 1990 and 1991 and Economic Trends 1992, 1993 and 1994.

Table 2 Price cost margins, concentration and intra industry transactions (levels 1989-1991)

	(1)	(2)	(3)	(4)	(5)	(6)
Constant	.554 (23.9)	.255 (9.17)	.258 (9.21)	.567 (24.9)	.253 (9.92)	.265 (10.5)
Concentration	.162 (6.25)	-.045 (1.46)	-.034 (0.99)	.145 (5.79)	-.046 (1.61)	-.033 (1.06)
Advertising	.200 (9.27)	.216 (7.32)	.231 (7.40)	.179 (8.18)	.199 (8.39)	.206 (8.16)
Dist margins	-.231 (4.04)	-.254 (3.88)	-.317 (4.53)	-.133 (1.89)	-.181 (3.61)	-.241 (4.32)
Import share	.093 (1.62)	-.024 (0.40)	.046 (0.73)	.020 (0.39)	-.046 (1.01)	-.003 (0.07)
Growth	.240 (3.28)	.054 (0.68)	.060 (0.69)	.263 (4.02)	.105 (1.52)	.109 (1.48)
No of Obs	144	144	144	174	174	174
R^2	.4862	.3522	.3456	.4149	.3775	.3539
Root MSE	.0699	.0785	.0834	.0736	.0747	.0794

Notes:
1. The dependent variables are PCM1 in (1) and (4), PCM2 in (2) and (5) and PCM3 in (3) and (6). Dependent variables are defined as follows:
PCM1=[(Net Output-Operative Wages)/Net Output]
PCM2=[(Net Output-Operative Wages)/Gross Output]
PCM3=[(Net Output-Operative Wages)/(Gross Output*(1-INTGRO))]
where INTGRO it the ratio of Intra Industry Transactions to Gross Output.
2. Absolute values of Heteroskedasticity robust t statistics in parentheses. All regression are OLS in levels for 1989-1991.
3. No of Obs is the number of industry sic groups used in estimation. There are 48 SIC industries which can be closely matched to the input output tables. Another 10 observations can be constructed as amalgams of SIC groups. Details of these SIC groups are given in the data appendix.
4. Concentration is the five firm concentration ratio, Advertising is defined as the ratio of "cost of other non industrial services" to industry gross output, Dist. Margins is the ratio of distributors trading margins to industry sales, Import share is the share of imports in sales. Grow is the growth rate of industry sales. Sources for data are given in the data appendix.

Table 3 Price cost margins, concentration and intra industry transactions (first differences 1989-1991)

	(1)	(2)	(3)	(4)	(5)	(6)
Constant	-.001	.001	-.0001	-.003	-.0003	-.002
	(0.42)	(0.52)	(0.06)	(1.56)	(0.18)	(0.91)
ΔConc	.078	.019	.061	.059	.020	.058
	(1.03)	(0.45)	(1.12)	(0.99)	(0.53)	(1.19)
ΔAdvertising	-.008	.047	.056	.002	.050	.058
	(0.31)	(1.72)	(1.97)	(0.10)	(2.03)	(2.22)
ΔDist margins	-.005	-.032	-.054	-.073	-.090	-.114
	(0.08)	(0.54)	(0.90)	(1.30)	(1.85)	(2.34)
ΔImport share	-.086	-.053	-.191	-.074	-.074	-.164
	(0.48)	(0.34)	(1.13)	(0.47)	(0.57)	(1.17)
ΔGrowth	.113	.067	.059	.093	.051	.044
	(2.83)	(2.70)	(1.93)	(2.70)	(2.02)	(1.65)
No of Obs	96	96	96	116	116	116
R^2	.3761	.2481	.2028	.2904	.1764	.1501
Root MSE	.0168	.0175	.0207	.0195	.0149	.0228

Notes: 1. All dependent and independent variables are as defined for Table 2, save that they are in first differences.
2. Absolute values of Heteroskedasticity robust t statistics are in parentheses. All regression are by OLS.
3. The number of units of observation are as for Table 2, but one year's set of observations have been lost through first differencing.

Table 4: Data Description: Price Cost Margins and Bought in inputs (1984-1991)

Variable	Obs	Mean	Std. Dev.	Min	Max
PCM1	760	.7201993	.0963362	.3925997	.9647428
PCM2	760	.3230236	.0954088	.1187585	.6525966
Conc	760	.4347352	.2255123	.07	.9922145
Grow	760	.0508818	.0993468	-.3961051	.5215284
Ads	760	.0653215	.0295866	.0174203	.1786424
P/R	760	.5514347	.1118349	.2281243	.8431002

	PCM1	PCM2	P/R	Conc	Ads	Grow
PCM1	1.0000					
PCM2	0.4487	1.0000				
P/R	0.0031	-0.8850	1.0000			
Conc	0.2288	-0.1207	0.2441	1.0000		
Ads	0.4318	0.6038	-0.4519	0.0072	1.0000	
Grow	0.1021	0.0797	-0.0406	-0.0998	-0.0327	1.0000

Notes:
1. Conc is the five firm concentration ratio by sales; PCM1 is the net output denominator price cost margin, (net output-operative wages)/net output; PCM2 is the gross output denominator price cost margin, (net output-operative wages)/gross output; Grow is the first difference of the log of total industry sales; Ads is a proxy for the advertising sales ratio. P/R is the ratio of bought in inputs to gross output.
2. The table lists the variable means for a sample of 95 industries from 1984-1991 inclusive.
3. Data is taken from Census' of Production, Business Monitor PA1002 for various years.

Table 5: Price Cost Margins and Bought In Inputs

Instrumental Variables First Difference Equations 1984-1991

Variable	(1) ΔPCM1	(2) ΔPCM2	(3) ΔPCM2	(4) ΔPCM2
Constant	0.003	-0.007	0.005	0.005
	(0.825)	(1.324)	(2.243)	(2.519)
ΔCONC	0.4184	0.2311	0.1584	0.1476
	(2.497)	(1.539)	(2.243)	(0.764)
ΔGROW	0.0025	0.0052.	0.0011	0.0009
	(0.171)	(0.416)	(0.158)	(0.143)
ΔADS	0.3773	0.2984	-0.0286	-0.0333
	(0.799)	(0.757)	(0.160)	(0.186)
ΔP/R			-0.9856	-1.0043
			(8.112)	(6.042)
Δ(P/R)*CONC				0.0002
				(0.060)
Obs.	760	760	760	760
Wald 1.	7.208(3)	3.179(3)	110.1(4)	129.5(5)
Wald 2.	43.20(8)	32.24(8)	94.34(8)	89.69(8)
Sargan	4.98(2)	5.00(2)	5.60(2)	5.66(2)
SC1	-2.951	-3.220	-2.448	-2.451
SC2	-0.650	-0.482	-0.536	-0.545

Notes:
1. Dependent variable is PCM1 in regression 1, PCM2 in the other regressions. Both as defined in note 4 Table 1. All regressions include a full set of time dummies. P/R is the ratio of bought in inputs of goods and services to gross output.
2. All regressions are by instrumental variables using the Arellano and Bond DPD program. All right hand side variables are treated as endogenous. Instruments are the full set of time dummies, the first lags of each of the independent variables, the second lag of the change in concentration and second lag of the level of concentration.
3. Absolute values of robust one step t statistics are in parentheses.
4. Wald1 is a test of the joint significance of the right hand side variables. Wald2 is a test of the joint significance of the time dummies. Both are distributed as Chi-squared with degrees of freedom in parentheses.
5. Sargan is the test of over identifying restrictions, degrees of freedom in parentheses.
6. SC1 and SC2 are robust tests of first and second order serial correlation respectively. They are distributed as N(0,1). With first differencing an MA(1) error process is generated therefore it is expected and admissible within the model for significant and negative first order serial correlation to exist.

Table 6: Price Cost Margins and Bought In Inputs: logarithmic specifications

Instrumental Variables First Difference Equations 1984-1991

Variable	(1) ΔLPCM1	(2) ΔLPCM2	(3) ΔLPCM2	(4) ΔLPCM2
Constant	0.004 (0.670)	-0.033 (2.835)	0.019 (2.241)	0.008 (0.620)
ΔLCONC	0.2664 (2.111)	0.2868 (1.207)	0.2582 (2.368)	0.3249 (2.152)
ΔGROW	-0.0019 (0.073)	0.0029 (0.047)	-0.0025 (0.116)	0.0100 (0.264)
ΔLADS	0.0818 (1.366)	0.1813 (1.371)	0.0413 (0.894)	0.1105 (1.284)
ΔLVI			1.3968 (10.28)	
ΔL(P/R)				-1.5845 (4.679)
Obs.	760	760	760	760
Wald 1.	6.043(3)	3.234(3)	146.6(4)	31.69(4)
Wald 2.	35.36(8)	29.80(8)	51.46(8)	22.17(8)
Sargan	2.745(2)	2.540(2)	1.760(2)	0.063(2)
SC1	-2.967	-3.454	-2.811	-3.038
SC2	-1.243	-0.535	-1.447	-0.003

Notes:
1. Dependent variable is ΔLPCM1, first difference of log of PCM1, in regression 1, ΔLPCM2, first difference of log of PCM2 in the other regressions. LVI is the log of the ratio of net output to gross output. LADS is the log of the proxy for advertising sales ratio. L(P/R) is the log of the ratio of bought in inputs of goods and services to gross output. All regressions include a full set of time dummies.
2. All regressions are by instrumental variables using the Arellano and Bond DPD program. All right hand side variables are treated as endogenous. Instruments are the full set of time dummies, the first lags of each of the independent variables, the second lag of the change in concentration and second lag of the level of concentration.
3. Details of all diagnostic tests are as in the notes to Table 5.

Data Appendix:

<u>Data Sources</u>

The sources for the data are various issues of the annual report of the census of production Business Monitor publication PA1002, and various issues of Economic Trends which now publish the core elements of the UK Input Output Tables.

In constructing the variables gross output was used rather than total sales to eliminate noise created by inventory movements. Net output is calculated by deducting purchases of materials and fuel and the cost of industrial services received from gross output and making an adjustment for the change in the stocks of materials and fuel. In making use of data on the two elements it is therefore possible to differentiate between bought in inputs in the form of goods and inputs in the form of services. Advertising is proxied by the ratio of the "cost of other non industrial services" to gross output. This is an imprecise definition because in addition to advertising services, payments for licensing, use of trademarks and patents, which roughly accord with the economic intuition of advertising as product differentiation, the cost of other non industrial services also includes payments to lawyers, accountants etc. This measure has been used previously however (e.g. Dowrick (1990)) and can be justified on the grounds that the use of professional services such as lawyers and accountants is unlikely to exhibit substantial variance across industries in comparison to advertising services. Total sales was used to generate the sales growth data, defined as the first difference of log of total sales, as a demand side proxy.

The data from the Input output tables was constructed as follows. The volume of intra industry transactions is the recorded value on the diagonal of Table 3: Demand for Products (the "combined use matrix"). Data on Imports, Distributors trading margins and the total supply of products are taken from Table 2: Supply of Products.

I would like to thank Martin Conyon for providing me with part of this data.

Data Sample

The SIC(80) groups used in the initial estimation of the intra industry transaction hypothesis presented regressions (1)-(3) of Tables 2 and 3 are listed below. For each of these groups a close matching between the industry classification used in the input output tables and the SIC group (three digit) classification could be found. In addition to those SIC groups for which a close matching could not be found, groups with miscellaneous categories or missing data were also eliminated from the full sample of manufacturing SIC groups.
224 241 242 247 248 251 255 256 257 258 259 260 314 316 320 321 322 324 325 326 329 330 341 342 343 344 346 361 364 411 412 413 419 421 422 428 429 431 436 437 438 451 455 467 471 472 475 483

In addition to these groups a further ten industry categories could be formed as combinations of two or more SIC groups. These combinations of SIC groups are listed below. In these cases the variables are constructed as aggregations of the groups (in the case of sales hence sales growth) or weighted averages (in the case of concentration and price cost margins).

1 221 222 223
2 243 244 245 246
3 323 327
4 351 352 353
5 362 363
6 371 372 373 374
7 424 426 427
8 432 433 434
9 441 442
10 461 462 463 464 466

Chapter 4: Strategic Asymmetries and Intra Industry Performance

I. Introduction

We have argued in the previous chapters that strategic asymmetries are important in understanding differences between large and small firms within industries whether those asymmetries are focused on the competitive stance or on decisions concerning the cost structure employed. Indeed, in chapter 2 we have argued that in some cases the two can be conflated in empirical terms because of the observational equivalence that can occur. There have been a number of previous attempts to investigate these asymmetries both theoretically and empirically. In this chapter we will review these attempts and argue that one of the key elements that has been lacking from previous empirical studies is an attempt to exploit both time series and cross section elements of a data set. While there have been a number of empirical studies using such panel data to investigate firm level performance for the U.K.(e.g. Machin and Van Reenen (1993) and Geroski, Machin and Van Reenen (1993)) these have not addressed directly the issue of asymmetry between firms or groups of firms. In addition, these studies suffer[1] from the reliance upon accounting information for what are, in large part, diversified companies for their data, hence they have difficulties mapping industry level information into their data set. We therefore argue that there is both a need for, and scope for, further testing models of such asymmetries using a data set that can address the important statistical criticisms that can be made of the previous tests. In this chapter, we also present the results of such empirical tests and compare these results with the previously obtained results. In section II, we consider the theories that have been put forward including the dominant firm competitive fringe model and the mobility barrier hypothesis put forward by Caves and Porter as ways of explaining intra industry performance. Section III critically reviews the previous empirical studies of intra industry performance. Section IV discusses the statistical problems that need to be addressed in an empirical study of this kind.

[1] We do not wish to imply from this that such studies make use of data invalid to their purpose merely that the data used would have limitations for studies with our focus.

Section V sets out the alternative empirical specifications of the intra industry performance relations and section VI presents and analyses the results of the empirical analyses.

II. Theories of strategic asymmetry and intra industry performance

In this section we will consider four approaches to strategic asymmetry from the literature. The four approaches are the standard Clarke Davies / Cowling Waterson framework that seeks to explain firm performance while allowing for differences in both costs and in competitive stance; the dominant firm competitive fringe model; the mobility barrier hypothesis of Caves and Porter (1977), and the "animal motif" framework associated with Fudenberg and Tirole (1984). While this cannot pretend to be, and space constraints preclude, a comprehensive survey, these four approaches are arguably the most widely used models when seeking to explain intra industry performance. The four approaches overlap to some extent, as the dominant firm approach can be considered to be a special case of both the Clarke Davies and Caves and Porter approaches.

II.i. The Clarke Davies model of intra industry performance

The Clarke and Davies (1982) elaboration of the Cowling Waterson model is essentially a model of short run profit maximisation in the context of a homogeneous product and a fixed number of firms. Within these two constraints, the model is flexible in that it permits variation across firms both in cost structure and in competitive stance and both of these factors can account for variation in performance. The model has been considered extensively within industrial economics textbooks[2] and therefore we will not discuss the model in detail.

The degree of strategic asymmetry involved in this model is limited. The models that follow address some specific asymmetry such as first mover advantage(s) or alternatively the advantage(s) associated with incumbency. In the Clarke Davies model, the options for strategic asymmetry lie in adoption of differing competitive stances and, if the cost structure is regarded as endogenous to some degree, different cost structures. To some extent, this is a direct consequence

[2] Waterson (1984) Chapter 2 presents a relatively comprehensive exposition and discussion of the model.

of the short term profit maximisation basis of the model which excludes analysis of the long run factors that influence the institutional context of price setting. This omission implies a difficulty of interpretation for it is customary in the empirical analyses (see section III) to include numerous 'long run' 'control' variables in the analysis that have little justification on a formal theoretical level but which are relevant from the point of view of the empirical determination of performance in the long run.

II.ii. The dominant firm competitive fringe framework

The dominant firm - competitive fringe model was developed initially by Forchheimer (1908) and was then subsequently developed by Saving (1970) and Encaoua and Jacquemin (1980) to incorporate the notion of a cartel of dominant firms and the possibility of entry. The idea is straightforward enough, a group of small firms act as price takers in a market offering a supply curve of output which is upward sloping. This supply curve leaves the dominant firm(s) with a residual demand curve and the dominant firm(s) is (are) assumed to act as a monopolist (cartel) with regard to the residual demand curve. The profitability implications are equally clear, so long as the competitive fringe has a supply curve of finite elasticity the dominant firm(s) will make supranormal profits. As the supply curve of the fringe becomes more elastic so more output is taken up by the fringe, the smaller will be the residual demand curve and the greater will be the fringe market share. The determinants of the shape of the supply curve of the can be split into two factors the shape of the marginal cost curve and the propensity towards short run entry by fringe firms. If the small competitive firms are flexible, in the sense described in Chapter 2, the slope of their marginal cost curves will be relatively small and this will pass over, through the horizontal aggregation, in forming an elastic fringe supply curve. Additionally, if, as price rises many small but inefficient firms can enter into the industry, the supply curve will be more elastic. The conventional assumption[3] of the competitive fringe model is that the supply curve is fairly

[3] This is not necessarily an assumption for mathematical convenience. Cowling's arguments below imply that the elasticity of a firm's supply/marginal cost curve will be a key determinant of whether it forms part of the dominant group or the competitive fringe. By implication firms with elastic supply responses will tend to be co-opted into the dominant firm group.

inelastic however, otherwise the dominant firm(s) would be left with a very small portion of demand.

From the point of view of changes over time we can analyse demand changes for a linear demand curve in two ways; an outward shift in the demand curve and a steepening of the demand curve. In the former case, the relative size of the small firm sector will tend to decrease unless the supply curve is very elastic, in which case the small firm sector will be fairly large to start with. The intuition behind this is that with an inelastic supply curve as the optimal price rises with the increase in demand so the output of the fringe is relatively unresponsive, so any increase in total demand will be mainly met by the larger firms. If the supply response of small firms is elastic then the increase in demand will largely be met by the small firms, but with an elastic supply response by small firms the size of the residual demand curve will be small and, therefore, the scope for dominant firm output reduced. In the case of a steepening of the demand curve the reverse will tend to be the case. The likelihood is that the share of output produced by small firms will rise, the exception being if the small firms have a significant absolute cost advantage, an outcome which doesn't tend to fit well within the confines of the model. The intuition for this is that with a steepening of the demand curve the possibility exists for increasing monopoly power, this will tend to raise the incentive for the dominant firm to restrict output, hence raise price. This raising of price will bring forth additional supply by the smaller firms.

The previous discussion has proceeded on the basis that the characteristics of the supply curve are exogenously given. While this is necessary for a simple exposition of the model, it is insufficient for a thorough understanding of the workings of industries. Cowling (1982, p9) has rejected the notion of a competitive fringe as a useful way of analysing industries. Two prominent points are made, the first is that small firms in an industry are unlikely to be producing similar products to large firms and indeed may be producing products for supply to the larger firms, a factor examined in Chapter 3 of this thesis. The second point is that if the fringe becomes too strong a constraint on their operation the oligopoly group will take steps to eliminate or assimilate the fringe firms[4]. Elimination is likely to take the form of some kind of predation, assimilation is likely to take the form of take-over. The implication for modelling of these relations is that the supply curve of the fringe is likely to be fairly inelastic because if it were not the oligopoly group would take the necessary steps to ensure that it was. The clear exception is the case of low entry barriers and a weak oligopoly group, in which case the supply curve will tend to be elastic and the oligopoly group will be relatively powerless to do anything about it. Encaoua and Jacquemin (1980) develop their model incorporating this recognition of the possibility that the oligopolistic group can act to limit the supply curve, more explicitly in their framework the level of expenditures which are intended to curb entry is a decision variable along with price. The result is a form of dynamic limit pricing model with non price competition added into the framework.

On the assumption that most strong oligopoly groups will have within their power the ability to erect some form of entry barrier, it becomes clear from this discussion that the reasoning of the competitive fringe dominant firm model can be turned on its head. It is not the size of the competitive fringe which places an exogenous constraint upon the actions of the dominant firms, instead it is the strength of the dominant firms which determines the size and relative importance of the small firm sector. Such an interpretation is almost true by definition. The competitive fringe, as price takers, simply react to changes in prices in their respective markets adapting their output, the dominant group is made up of strategic actors who will clearly see it as in their own advantage to manipulate the industry in a way which minimises the effectiveness of the constraint. The competitive fringe model should not be considered an unusually restrictive model when one recognises that Saving's framework is merely a limit case of strategic asymmetry between firms, where in this limit case one firm (or one group of firms) does not, in fact, act strategically at all.

[4] Aaronovitch (1955) makes a persuasive argument that trade associations play (or at least played at the time of writing) a key role in the 'policing' of smaller firms acting in order to enforce the interests of 'the big combines' on the industry as a whole. Accounts of the behaviour of the Building Societies Association in running its interest rate "cartel" in the interests of the larger societies prior to its disbandment in 1984 would appear to confirm that this is not merely a historical curiosity.

The Saving version of the dominant firm model could be seen to suffer from the same restriction in terms of scope of strategic asymmetry as the Clarke Davies model. In the particular formulation of Encaoua and Jacquemin (1980), more factors are permitted that additionally influence the outcome than the direct stance of the firms with regard to pricing. We have argued, following Cowling, that this attitude can also be applied to the more limited forms of the dominant firm model such that many kinds of strategic decision making will influence the relative positions of the dominant firm(s) and the fringe. In this sense, in its conception of strategic asymmetry the dominant firm model may be considered a halfway house between the Clarke Davies type model and the Caves and Porter analysis.

II.iii. Mobility barriers and entry barriers

By viewing the allocation of firms to a dominant group or competitive fringe as the result of an endogenous strategic decision making framework, on the part of the firms as well as influenced by the exogenous parameters faced by those firms, the framework of the competitive fringe model can be seen as part of the more general theory of "mobility barriers" contained in the Caves and Porter (1977, 1978) notion of strategic groups within industries. In the terms we have discussed above, the fringe 'strategic' group is limited by the dominant groups ability to construct appropriate mobility barriers.

The Caves and Porter (1977) analysis is interesting, not least because it predates and anticipates many of the future developments that were to be formalised in the game theoretic developments of the subsequent years[5]. The key element of the Caves and Porter theory is the assertion that there may be subgroups of firms within industries with "differing structural characteristics". Moreover, these subgroups are not simply a result of misdefinitions of industries, as the firms may be producing an identical product yet be classified into subgroups based upon the means of provision or production of this product. Once the existence of such subgroups is accepted, the conventional analysis of entry barriers may be applied to movement between subgroups and,

[5] For example, the intuition of the findings of Gilbert and Vives (1986) discussed below were anticipated almost exactly (Caves and Porter (1977) p247-8).

indeed, the strategic decisions conventionally associated with entry deterrence could also become pivotal in defining the subgroups.

Analytically, once the existence of subgroups is acknowledged, the framework can be applied to these subgroups in the same way that the standard theories of entry are applied to the industry. However, it should be noted that some of the problems that can potentially occur with entry deterrence such as difficulties of co-ordination of oligopolists are eased when applied to industry subgroups because there is likely to be greater mutual interest between the firms that make up a separate subgroup than between firms in different subgroups. The primary implication of the theory, from the point of view of performance, is that "easy entry into an industry's fringe group is therefore no guarantee against monopoly profits"(ibid. p257).

II.iv. Animal motifs, strategic substitutes and strategic complements

Fudenberg and Tirole (1984) present a taxonomy of business strategies that formalises much of the expected relationship between leader and follower firms. Unlike most of the game theoretic explosion, this exposition posits asymmetry between firms and accounts for the strategies adopted by the leader (the followers as a result of their second mover status are reduced to "reactive" decision makers) in terms of the effect on the follower. The discussion takes the form of a two stage game in which in the first stage the leader firm chooses the level of a strategic variable, which may in principle be any strategy but is commonly a capacity or advertising intensity decision, to adopt. In the second stage, the two firms compete with respect to a different strategic variable, commonly price or quantity, assuming that firm two chooses to compete in the second stage at all[6]. There are two key determinants of the first stage strategies adopted by the leader firms. Firstly, does the level of the strategic variable increase or decrease the follower's profits in the second period? In the terms set out by Fudenberg and Tirole, is the strategy tough or soft? Secondly, are the strategies employed in the second stage competition strategic substitutes or

[6] If entry does not occur the leader firm acts as a monopolist albeit constrained by the strategy adopted in the first stage.

strategic complements?[7] The answers to these two questions, combined with the follower firms entry decision determine the adopted strategy. The four animal motifs used by Fudenberg and Tirole in principle comprehensively characterise the possibilities. The insight added by this taxonomy is that the strategies adopted by the leader firm are contingent upon the nature of the relations. Therefore, in answering the question implicitly posed by Encaoua and Jacquemin, does advertising expenditure limit the scale of the follower firms? the answer must be ambiguous. In certain industries advertising by dominant firms is likely to inhibit the scale of entry by the follower firms. However, if the advertising raises prices for all firms this is likely to increase the number and scale of the follower firms.

A further issue that arises for all of the models of strategic asymmetry is the ability of oligopolists to act in concert as a single monopolist. Conventionally such co-ordination is influenced by the number of firms in the group and the degree of heterogeneity within the group (see Scherer and Ross (1990) Chp 8). A question addressed by Gilbert and Vives (1986) is whether in a noncooperative environment oligopolists will be able to sustain such strategies. Their model is specific to their context (a top dog strategy is required by the model) but the intuition spills over into other models. In principle, if one firm's action deters or limits entry or mobility, the other firms benefit and can therefore free ride on the strategic action of their rival. This could potentially lead to a failure of co-ordination in the conventional public good sense. Gilbert and Vives show that this will not tend to occur if the spoils from deterrence are distributed according to the contribution to the deterrence strategy. The further point that oligopolists may not replicate the full monopoly solution is of only marginal relevance. The important point is that it is fully conceivable that non co-operative oligopolists can deter or limit entry and the precise (and usually empirically unobservable) conditions under which it will occur are of lesser importance. From an empirical point of view the implication is that for some strategic actions, those where the rewards are distributed according to the participation in the strategic action rather than being equally

[7] Strategic substitutes and complements as Bulow, Geanakopolous and Klemperer (1985) define and elaborate upon.

distributed according to the consequence of the action(s), the degree of co-ordination amongst the oligopoly group may be irrelevant to the success of the actions.

Can we empirically operationalise this taxonomy? There have been a number of theoretical examples given of the different strategies but we are not aware of attempts to investigate this empirically. *Ex ante* it is difficult to see that it is possible to identify specific industries on the basis of the taxonomy because of the concomitant difficulties of giving *ex ante* responses to the two relevant questions. *Ex post* the problem of tautology arises, we may categorise industries to different motifs on the basis of observed performance and strategies but do not have any clear prediction. Indeed, observation of performance is not trivial, the observed capacities do not help us differentiate, for example, between lean and hungry and top dog cases because we cannot observe the counterfactual level of capacity associated with a decision taken in the absence of (potential) rivals. On a less ambitious level we can perhaps suggest that there exist some industries in which advertising leads to dominance while others where advertising through spillovers tends to lead to equal levels of performance, if not equal sizes. The effect of advertising is therefore potentially variable across different industries. This would therefore reformulate Bain's (1956) idea that combination of strategies where MES interacts with advertising and concentration to give differing levels of performance.

III. Previous empirical tests of the theories of intra industry performance

As discussed above the extent to which the fourth approach to intra industry performance can be incorporated into empirical testing is difficult to assess and there is only one formal empirical study[8] that has attempted to examine this issue. Cross industry tests of the dominant firm - competitive fringe are similarly rare although there have been a number of studies of single industries[9]. Perhaps as the dominant firm framework is a special case of the Clarke Davies framework, it has been regarded as incorporated within the cross industry tests of this framework.

---

[8] Slade (1995a) analyses the saltine cracker market, see Slade (1995b) for a general discussion of strategic models.

[9] Roberts (1984) examines the coffee market, Suslow (1986) the aluminium industry and Yamawaki (1985) the iron and steel industry.

Given these gaps in the literature, we focus, for this discussion of previous studies, on the two established forms of test; firstly, those based around the Clarke Davies framework; secondly, the tests examining the Caves and Porter mobility barriers hypothesis.

Before considering the specific papers, it is worth examining the general overview presented in Schmalensee (1989a) in relation to intra industry hypotheses. This overview can be encapsulated in the four stylised facts in relation to intra industry studies.

4.10. The profitability of industry leaders in U.S. manufacturing may be positively related to concentration; the profitability of firms with small market shares is not. (ibid. p983)

4.11. In samples of U.S. firms or business units that include many industries, market share is strongly correlated with profitability; the coefficient of concentration is generally negative or insignificant in regressions including market share. (ibid. p984)

4.12. Within particular manufacturing industries, profitability is not generally strongly related to market share. (ibid. p984)

4.13. The estimated effect of market share on profitability in U.S. manufacturing industries is positively related to the industry advertising sales ratio. (ibid. p985)

It is worth emphasising at this stage that the main theoretical focus for the previous 'facts' was the debate concerning the efficiency/market power explanations for profitability. We are not in this paper concerned with such discussions. The debate has, as such academic disputes are wont to, disappeared without any effective resolution. The reasons for this would appear to be relatively clear. Firstly, there does not appear to be any consensus on an appropriate empirical test for distinguishing between the hypotheses, and, indeed, it is severely questionable as to whether such a test could exist. Secondly, despite the fact that evidence has been presented that has apparently contradicted each of the hypotheses, this has merely led to either a restatement of the hypothesis in a way that permits acceptance of the evidence or rejection of the validity of the evidence. In short, neither side were learning (or willing to learn) a great deal from the empirical tests and in such an environment, academic interest is destined to wane. While those who maintain an interest in such disputes may find the empirical results presented in this chapter to be of interest they are not intended for such a purpose. Instead, the aim is the more limited one of reassessing empirical regularities or stylised facts within a more sophisticated statistical framework.

III.i. Estimating forms of conjectural elasticity

One of the main empirical foci for analysing intra industry profitability has been the estimation of conjectural elasticities or other versions of conjectural variation terms. We briefly review three alternative forms of estimation for these studies and consider their application to our context. All centre around the Clarke and Davies (1982) reworking of the Cowling - Waterson (1976) model. At the firm level Clarke and Davies reformulate the Cowling and Waterson relation,

$$\frac{p-c_i'}{p}=\frac{s_i(1+\lambda_i)}{\eta} \quad (1)$$

where $\lambda_i=\frac{\partial Q_{-i}}{\partial q_i}$ is the standard conjectural variation term as

$$\frac{p-c_i'}{p}=\frac{s_i+\alpha_i(1-s_i)}{\eta} \quad (2)$$

where $\alpha_i=\frac{\partial Q_{-i}}{\partial q_i}\frac{q_i}{Q_{-i}}$ is a conjectural elasticity. Implicitly the relation between λ and α is given by $\lambda_i=\frac{\alpha_i(1-s_i)}{s_i}$. One of the difficulties in making such a model empirically workable is that α is not identifiable without placing strong restrictions on the parameters of the estimated model[10]. Using the Clarke Davies reformulation Kwoka and Ravenscraft (1986) model α as a function of various structural variables permitting it to depend, in turn, on the level of concentration and the market shares of the large firms within the relevant industry. They add various additional variables but these are not entered into the specification as explaining α merely as "control" variables that are presumably external to the theoretical specification adopted. Machin and Van Reenen (1983) approaching the same theoretical model, respecify the conjectural term as $\lambda_i=\frac{\alpha_{1i}(1-s_i)}{s_i}+\frac{\alpha_{2i}}{s_i}$ although the theoretical justification for such a formulation is unclear and the only change to the estimable model is an implied firm specific intercept term the inclusion of which is justified on statistical grounds alone. In the Machin and Van Reenen study all additional

---

[10] For example, despite placing a structure on α Kwoka and Ravenscraft still estimate α/η rather than α itself.

relevant variables, including a lagged dependent variable, are added interactively because α_1 is assessed as depending upon them. Kwoka and Ravenscraft consider the possibility that the estimated coefficients will vary across firms[11] and replicate Porter's finding that advertising matters for large firms but not for the follower firms. They also report estimates of interaction terms that indicate behaviour more aggressively competitive than Cournot behaviour ($\alpha<0$) on the part of all firms and especially those follower firms in industries where minimum efficient scale is important[12]. Geroski (1990b) argues that this observation of rivalrous behaviour may be a quirk of the Line of Business data used which is drawn from a single year, 1975, which is at an unhelpfully turbulent point in the business cycle. Both studies partition the data according to (different) industry characteristics and estimate separate results for these.

III.ii. Empirical tests of the Caves and Porter hypothesis

The strategic management literature is replete with studies of strategic groups models. See Reger and Huff (1993) for a survey of single industry studies. Schmalensee (1985) has examined the side issue of how important are industry factors relative to firm specific factors in determining profitability. Again within the strategic management literature this has been developed and discussed by, *inter alia*, Rumelt (1991), Cool and Dierickx (1993) and Powell (1996). By comparison the economics literature has been relatively silent on these issues. There are nevertheless difficulties in attempting to empirically analyse these issues. In principle, in order to avoid arbitrarily partitioning the data, detailed information on the firms sufficient to divide them into their respective subgroups would be necessary. However, even if this data were available, which it is not in this cross industry context, it would be difficult to avoid tautology, as by creating subgroups in the image of the theory, or otherwise, one could not 'prove' the validity of, or 'falsify', the theory. Caves and Porter (1977, p254) argue that allocation to subgroups will be

[11] Machin and Van Reenen, by estimating a fixed effects model, permit intercepts to vary across firms but do not compare otherwise.

[12] The chain of reasoning explaining this result is not completely clear. The implication appears to be that small firms suffer from cost disadvantages in industries where minimum efficient scale is high and this therefore encourages aggressive behaviour on the part of the firms that operate at suboptimal levels. A point of note, which makes the use of minimum efficient scale problematic, is that the correlation between minimum efficient scale and the cost disadvantage of sub optimal output is not necessarily exact as illustrated by Scherer *et al.* (1975). This reasoning also fails to explain why small firms in industries with low entry barriers should act in a less competitive manner than those protected by entry barriers.

according to the different endowments of the firms and, more importantly, that these endowments are distributed at random. Below we argue that this is unlikely to be the case, instead positing a much greater role for (some) firms in choosing their strategic group. The consequence of this allocation problem is that some arbitrary partitioning of the data does occur, invariably according to the size of the firms.

From the perspective of this thesis the comparison of firms of different sizes is our major focus and we are therefore not too troubled by such an imposition. It should be noted that for a more detailed attempt at testing the specific hypothesis it would appear to be preferable to also partition the industries according to other factors such as advertising intensity vertical integration etc.. However it can be argued within this context that size is a privileged measure for differentiating between firms. This derives from a belief that large firms have an advantage over small firms arising from their ability to reap economies of scale not just in production but more particularly, within the current context, with respect to advertising, research and development or the terms of access to capital. Therefore, in the absence of legal restriction or resource monopoly we would not expect a large firm[13] to choose to locate in a subgroup that did not offer the best return. To paraphrase Steindl (1945), large firms can do anything that small firms can do but the reverse is not true[14]. This inability on the part of small firms to reap the benefits that accompany increased size therefore represent the fundamental strategic asymmetry which can be privileged above others. It can be further argued that this size related asymmetry is likely to more pronounced for mobility barriers than for entry barriers precisely because a number of potential entrants will be large firms seeking to diversify into a new market and for these firms this asymmetry does not exist. To counter this argument, we have to accept the empirical observation that in a not insignificant number of industries small firms do out-perform the larger firms. In section II of chapter 1 we have outlined some of the potential reasons given, especially in Waterson (1989), for this finding. One of these reasons is that there are niche markets that may

[13] Or a firm with considerable access to resources.

[14] Steindl was building upon the argument known as the "principle of increasing risk" advocated by Kalecki (1939; Chapter 4). The basis for such capital market disadvantages faced by small firms has been criticised by Stigler (1968 Chp 10) but has been extended more recently by Auerbach (1988).

only be accessible to small firms or which require specific or local knowledge to exploit. In this context, we have to accept that size is not uniquely privileged in providing asymmetry. However, we would still maintain that, from an empirical point of view, it is the most important readily observable indicator of asymmetric opportunity.

The main empirical test of the Caves and Porter hypothesis[15] is contained in Porter (1979) which investigates 38 US industries for 1963-1965 and examines the correlation between the performance of leader and follower groups, where, in this case the leader group is taken to be the largest firms with a combined 30% share of the market, and the followers constitute the rest of the firms in the industry. The results of the multiple regression results reported by Porter indicate important differences in the relative determinants of the performance of the leaders and followers. In each case, the measure of performance is the rate of return on equity as the industry data source used is the accounting data size classes provided as part of the US IRS data set[16]. The main explanatory variables used by Porter as part of his hypothesis were the four firm and eight firm concentration ratios and the relative average market shares of the leader and follower firms as measures of market structure. In addition, measures of advertising intensity, minimum efficient scale and minimum capital requirements and combinations of these were included as measures of barriers to either entry or mobility. A measure of industry sales growth and dummy variables for a regional market or a convenience goods industry were also included.

Estimation was performed separately for three different level of observation, the industry level, leader level and follower level. Concentration emerged as a significant and negative determinant for the industry level equations and for the follower equations. The effect of concentration on the leader performance is weakly positive but insignificant. The growth rate of industry demand has a

[15] Caves and Porter (1977) p252 list a number of pieces of supporting evidence in footnotes. The only other published paper we are aware of that addresses this hypothesis directly is Bradburd and Ross (1989). This paper uses the FTC 1975 line of business data to examine the relative distribution of large and small firm sales in five digit relative to four digit industries and the relative performance of the large and the small firms.

[16] This is the same data set used by Schmalensee (1989b) although Porter merely uses three years of data which are averaged, whereas Schmalensee uses a series of years from 1953-1983 to construct his data.

positive and significant effect upon the performance of the follower firms but has no significant effect upon the performance of the leader firms. The overall effect at the industry level is positive but poorly determined. The most notable contrast exists for the advertising sales ratio and the measures of scale economies/capital requirements. Advertising intensity measured at the industry level had a strong positive effect on the performance of the leader firms, but had a negative albeit poorly determined effect on the follower firms. Increasing minimum efficient scale had a positive effect on the performance of the leader firms and a negative but insignificant effect on the performance of the follower firms. Capital requirements, measured as the capital required to build a plant of minimum efficient scale, was a significant and positive determinant of follower performance. Neither of the dummy variables were significant at any reasonable level of significance.

These results are very suggestive of the Caves and Porter hypothesis. It suggests that high advertising is something which will tend to benefit the larger firms in an industry. Such arguments are consistent to a degree[17] with the dual structure arguments made by Sutton (1991). The evidence with regard to economies of scale is less clear cut. Minimum efficient scale appear to increase leader performance but not follower performance this is consistent with a simple story of cost disadvantage from being unable to achieve the required scale economies. By contrast, the presence of large capital requirements tends to increase the performance of the follower firms but has no effect upon the leader. Such a finding is consistent with an interpretation of capital requirements as a barrier to entry which therefore protects the follower firms in the industry as imperfections in the capital market restrict the entry of small firms. However, this capital requirement is not of benefit to the relatively larger firms, which is consistent with a view that such imperfections in the capital market are unlikely to affect potential large scale entrants.

IV. Empirical testing of intra industry performance using panel data

[17] Sutton's arguments largely depend for closure upon the assumption that profits are zero in long run equilibrium for all firms, which is inconsistent with observed higher profit performance.

Using the information presented within the UK Census of Production, we can create a data set which contains both a time series and a cross sectional dimension on industry size classes with which a superior econometric analysis of these issues can be examined. There are two major econometric problems which arise with the tests we have discussed in the previous section. Both arise from the cross section nature of the data set available. In this section, we discuss these difficulties and consider the additional econometric consequences of using panel data to attempt to understand these relations.

The first criticism that can be made of cross sectional work is that observations of variables in any particular year may not be equilibrium observations, yet the hypotheses being tested imply that the data are generated by a long run equilibrium relationship. The conventional solution to this problem following on from the early studies of Bain (1951,1956) was to average data over a period of years (5 years in Bain, 1951) in order to establish an average relationship on the assumption that disequilibrium observations will tend to even themselves out over that period. More specifically, the key requirement from a statistical point of view is that "deviations from long run equilibrium must be uncorrelated with the independent variables employed" (Schmalensee, 1989a; p953).

This approach has received specific theoretical justification from Pesaran and Smith (1995) who consider the problem of estimating the long run relationship when there is parameter heterogeneity across units of observations and Mairesse (1990) who considers the possibility of errors in variables. *A priori*, we cannot rule out either of these possibilities in our data set. Pesaran and Smith suggest that in data sets where both N and T are of a reasonable size[18] there are four types of estimation that can be used to estimate the long run relationship. The first is to pool the data into a panel for estimation purposes, the second is to aggregate the data and estimate an aggregate long run time series relationship, the third method involves estimating time series relationship for each of the n observations and averaging the coefficients, the final method

[18] Specifically they impose the inexact requirement that T is large enough to run "sensible" regressions for each unit (firm /industry).

is to average over time and estimate a quasi cross section relationship. The second and third methods are not open to us as, with a maximum of 13 annual observations, 'sensible' time series estimation cannot be considered. The question reduces, in this instance, as to whether it is better to estimate a pooled regression making use of both the time series and cross section information or to estimate a cross section regression. There are clear difficulties in the adoption of the latter approach when interested in aggregate effects that vary over time because these are lost from the analysis. However, Pesaran and Smith indicate that the cross section approach is under certain assumptions[19] robust to dynamic mispecification, whereas the pooled estimates are subject to potentially serious biases if there is either heterogeneity of the coefficients on lags of the dependent variable across groups, or the exogenous variables are autoregressive processes and have heterogeneous coefficients. Since neither can be ruled out *a priori*, such possibilities must be taken into account.

Mairesse (1990) argues that, if variables are measured with error, cross section estimation on data averaged over time will minimise the effect of random errors producing consistent estimates, whereas estimates based upon first differences will amplify such errors leading estimates to be poorly determined and potentially biased towards zero. Mairesse also gives backing to a short run/long run ("transitory effects"/"permanent effects") interpretation of the difference between the within group and between group estimates respectively.

The main alternative approach to estimating long run relationships that has been developed is Pesaran and Smith's first method, i.e. making use instead of both the cross section and time series properties of data to explicitly model adjustment to disequilibrium as an autoregressive distributed lag process from which long run relationships can be inferred. In the US, the data from the size class analyses is only available infrequently[20], therefore it is not possible to exploit

---

[19] The key assumptions are that the coefficients and the regressors are independent and the regressors are strictly exogenous.

[20] This largely arises from the fact that the US census of manufactures is only published every five years. Therefore while the IRS size class analyses of accounting data are published annually the industry level data necessary to test some of the relevant hypotheses only exhibit time series variation every five years.

such time series aspects of the relationship. While for Porter (1979) some of the measures of the variables used in the empirical analysis are averaged over two or three years in order to minimise the distortion of disequilibrium observations, for other variables only a single year's observation is available for each industry. By building up a panel of observations on intra industry performance for the UK, we are able to exploit the time series aspects of the relationship to directly estimate long run relationships.

The second aspect of the criticism of the cross sectional nature of the studies relates to mispecification of the empirical relationship. The statistical indicator of this mispecification is the correlation between the independent variables and the error term. This mispecification can be decomposed into two elements. The first is explored by Schmalensee (1989a; p954-5) in suggesting that many of the variables must be regarded as endogenous in the usual heuristic sense[21] of being variables directly or indirectly determined by the strategic choices of the firms in the industry hence determined within the theoretical system. In addition, by reflecting on the short run nature of the decision making process we must also include as endogenous those variables which do not satisfy the statistical criteria of weak exogeneity (see Hendry (1995; p162-4)), for example, as a result of the dependence of the decision making process, upon the expected rather than the realised value of one of the independent variables. Within the context of a cross sectional estimation framework, Schmalensee denies the possibility of the existence of relevant contemporaneous exogenous variables that would enable unbiased estimation of these kind of relationships to be performed, as he puts it "in general there are no theoretically exogenous variables that can be used as instruments to identify and estimate any structural equation" (Schmalensee (1989a; p954)). However in permitting the exploitation of time series aspects of data sets it is suggested that appropriate instruments can be found. Even here Schmalensee qualifies his argument suggesting that "panel data....can yield consistent structural estimates *if an explicit model of disequilibrium behaviour is employed.*" (ibid; p956, emphasis added). Without further elaboration, which Schmalensee does not give, this statement is slightly ambiguous. It is

[21] Using the Koopmans (1950) definition of exogeneity "A variable is exogenous if it is determined outside the system under analysis" cited in Hendry (1995; p157)

unclear whether the criticism refers to the practice of estimating static relationships with panel data thereby merely exploiting the increased degrees of freedom and making use of the availability of lagged values for the instrument matrix, in which case the model does not contain any disequilibrium framework. If this is the limit of the criticism we can accept it, such models address the second problem of finding appropriate instruments but do not address the initial problem of correlation between the extent of disequilibrium and the independent variables. However, if Schmalensee is going further and suggesting that a *specific* form of equilibrium adjustment needs to be explicitly specified prior to estimation we would have to disagree. It seems unlikely that we are going to be able to specify *a priori* that firms adopt any specific process of adjustment or equilibrium correction with any satisfactory degree of confidence. In fact, such an assumption is unnecessary because with the exception of formulations that predict highly non-linear relationships between variables, many such formulations are readily reducible to restricted forms of autoregressive distributed lag models[22]. Imposing an *a priori* restriction is unlikely to be warranted in most cases.

A further and closely related aspect of the use of cross section data that needs to be considered is the inevitable omission of industry or firm specific factors which cannot be directly observed. Such factors when applied to the intra industry case presented here would include, by way of example, the degree of tightness of co-ordination between the dominant group within the industry which can only be imperfectly measured through observable structural variables and which are, at very least, likely to be partially determined by intrinsically unobservable elements such as industry norms and customs and the degree of trust and communication built up over time between members of the dominant group. If such unobservables are correlated with the included independent variables their omission will lead to bias in the point estimates on these right hand side variables. Therefore if, as seems plausible from both a heuristic and a statistical standpoint in this particular context, such unobservables are not randomly distributed across industries but are correlated with the observable structural variables of interest, a fixed effects approach to

22 See Hendry (1995) Chapter 7 for an account of some of the commonly used restricted forms and their relation to the autoregressive distributed lag model.

estimation exploiting the time series characteristics of a panel is necessary to eliminate the bias that results.

V. Alternative specifications of intra industry performance relations

In this section we will develop the empirical specifications that we will use to analyse the various approaches taken to study intra industry performance.

If we initially turn to the dominant firm competitive fringe model, this suggests that there are separate determinants of the performance of the fringe and the dominant groups and explicit relations describing these can be found. Following Saving (1970) we can determine the equilibrium relation for the Lerner index of the dominant firms as

$$\frac{p-c}{p}=\frac{C_k}{\eta_m+(1-C_k)\varepsilon_R} \quad (3)$$

where C_k is the k firm concentration ratio and the dominant group is made up of k firms. η_m is the (absolute value of the) price elasticity of the industry demand curve and ε_R is the elasticity of supply of the fringe firms. As has been suggested earlier, this elasticity may depend upon the competitive stance adopted by the fringe firms. The less competitive the stance and the steeper the joint marginal cost curve of the fringe, the lower will be this elasticity. In principle, the Lerner index of the fringe firms should be equal to zero if they are fully competitive. In practice, because measurement of the Lerner index usually relies upon the difference between price and average variable cost rather than marginal cost, the measured Lerner index will not be zero. In the case of the cost functions considered in Chapter 2[23] the fringe price cost margin is equal to $\frac{1}{2\varepsilon_R}$. The Lerner index of the fringe firms therefore depends upon the nature of the marginal cost function as well as the competitive stance. Any factor that increases the elasticity of supply reduces the price cost margin.

Relation (3) is entirely static in conception but Encaoua and Jacquemin (1980) have extended the framework to a dynamic analysis where expenditures on the part of the dominant groups limit the

[23] See equations 1 and 13 from chapter 2 of this thesis.

scale of entry. Their prediction from this point of view is unusual as they predict that while a greater responsiveness of the fringe to such expenditures (limiting their scale of entry to a greater extent) increases the degree of monopoly power for the dominant firms, the advertising sales ratio itself will be inversely related to the degree of market power. One of the reasons for this prediction, which runs counter to most predictions in relation to advertising intensity[24], is their assumption that such expenditures only serve to reduce the flow of entry and do not affect the price that can be charged. Non price competition will tend to increase the price being charged which will actually increase the supply of the small firms unless the mobility or entry barrier form of advertising dominates. In the Fudenberg and Tirole (1984) taxonomy, if there is no mobility barrier such advertising constitutes a fat cat strategy within this context, if it creates a mobility barrier advertising is likely to constitute a top dog strategy. It is therefore not possible to determine *a priori* the effect of advertising either on the fringe or the dominant group. The Encaoua and Jacquemin characterisation of the long run equilibrium is almost identical to (1) the difference being that ε_R must be interpreted as a discounted intertemporal elasticity of supply.

From an empirical point of view, both Saving and Encaoua and Jacquemin predict that the Lerner index of the dominant group will be positively (if non-linearly) related to the concentration ratio and negatively related to the industry price elasticity of demand and the elasticity of supply of the fringe. The effect of advertising intensity, whilst strictly predicted to have a negative effect within their framework, can work both ways in a less restrictive framework. More generally, because the elasticities are not observed and are unobservable we are forced into observing them through their potential determinants. Therefore, any factors that can plausibly effect either of these elasticities are candidates for inclusion in an empirical specification that attempts to discover the determinants of performance.

Porter's study examines relations of the form

$\pi_{ij} = f(Conc, MES, A/S, Cap, Grow, etc.)$

[24] The prediction is consistent with the Austrian school view of advertising as, for example, expounded in Kirzner (1973; Chapter 4) although the intuition is very different.

where π_{ij} are the measured profits of size group i in industry j. Conc is a measure of industry concentration, MES is a measure of Minimum efficient scale, A/S is the industry level advertising sales ratio, Cap is a measure of absolute capital requirements, Grow is the growth rate of industry sales. No specific functional form is suggested and no explicit theoretical framework from which such a functional form could potentially be derived is posited. All of the variables can be approximated to some degree within our data set.

With respect to MES, the Porter measure is derived by attempting to observe the median firm size using the size class analyses. While this measure has been relatively widely used within published studies, it is nevertheless arbitrary in design and has the relatively unsatisfactory implication that half of firms producing within every industry operate at inefficient output levels. Our measure of minimum efficient scale is simply the mean size in terms of sales of the firms in the smallest size class. As the UK census of production contains fewer size classes than the US data identifying the median firm is a more complicated task for the UK. Since the size distribution is skewed towards the smaller firms, it is likely that in a majority of cases the median firm would be in the smallest size class band, it would appear more consistent to apply this mean size in all cases. The mean size of the firms in the smallest size class represents the extent of size related barriers to entry into the industry faced by small firms. Within the Caves and Porter theoretical framework entry barriers to the industry per se may be relatively low although entry into the dominant group are likely to be more significant. This is in keeping with the general impression given by the Scherer's studies (esp. Scherer *et al.* 1975) that the size of minimum efficient scale of production relative to market size generated from engineering estimates is relatively small and the cost disadvantage from operating at levels substantially below minimum efficient scale is similarly low for many industries. Therefore, measures of minimum efficient scale are much more likely to represent obstacles to small entrants than large scale entry.

The Porter measure of capital requirements is simply obtained by multiplying the estimate of MES by the industry capital sales ratio. Capital stock figures are not available for the industries covered by the UK Census of Production and therefore we have to improvise an alternative. For this we use the average value of net capital expenditure for the smallest firms. For such a measure to approximate the capital requirements of entry requires the additional strong assumption that depreciation rates are constant across industries. We are relatively unhappy about having to make this rather unrealistic assumption but in the absence of better data there appears to be little alternative[25].

The measures of advertising intensity and industry growth are at the industry level. In Porter's analysis, an attempt is made to discriminate between the advertising expenditures of the small and large firms, but this proves unsuccessful. This is probably due to the existence of economies of scale in advertising which imply that while the size of dominant firms may increase, the advertising expenditures required in order to maintain that dominance do not increase in proportion. An analogous criticism is made in Schmalensee (1992) of the argument put forward by Sutton (1991) that as an industry (in which advertising expenditures are endogenous) increases in size so advertising expenditures must similarly increase in order to prevent entry. In practice, the important distinction both in Sutton's and Porter's work appears to be between industries in which advertising is intensive and industries in which it is limited. The industry level measures of advertising intensity and industry growth used here have been described in the data appendix to chapter 3 of this thesis.

We can therefore provide by means of comparison with the Porter results four different forms of estimation. Initially we estimate a cross section OLS regression using the data on the relevant variables averaged for the years 1984-1992 for purposes of comparison as this most nearly approximates the method Porter uses for generating results. Secondly, we can include a dynamic adjustment parameter within the cross section framework following Pesaran and Smith (1995; p88) that potentially controls for disequilibrium. Thirdly, we can further obtain long run estimates of the effects of the relevant variables by estimating a transformed distributed lag

[25] We can comfort ourselves slightly by noting that Porter's proxy for capital requirements requires that we make the assumption that the median firm has a capital sales ratio equal to the industry capital sales ratio. Given the wide empirically observed differences in capital sales ratios between large and small firms (see for example Caves and Pugel 1980) this assumption seems equally lacking in realism. Unfortunately two bad proxies do not make a good one.

model. Finally, we can make use of the Arrellano and Bond (1988, 1991) estimator for dynamic fixed effects models to control for the potential omission of unobservables.

Previous studies, both theoretical and empirical, while giving directions for analysis are not conclusive in providing empirical specifications for the analysis of intra industry performance. Neither in the case of choosing the appropriate variables for inclusion nor in the conversion of theoretical equilibrium relations into empirical estimating equations do we have any clear prior knowledge that we can be confident about. If we are therefore to analyse these issues empirically we are forced unwillingly, but perhaps not surprisingly, into a form of ad hoc exposition[26].

VI. Analysis of Intra industry performance in UK Manufacturing

Elementary descriptive statistics are presented in Tables 1-3. As a guide to interpretation we have used the following terminology for partitioning the industry: the "large" firms are the 5 largest firms in an industry taken as a group; the "fringe" firms are all the firms in the industry save this largest 5; the "small" firms are the firms with 1-99 employees, the smallest size category available within the UK census of production. In a small number of industries, the fringe consists entirely of small firms but in the majority of cases the fringe includes, in addition to the small firms, firms with more than 99 employees. Following from this discussion of the discontinuity in the data between 1983 and 1984 contained in Chapter 1, all estimation and summary statistics are for the period 1984-1992.

Without resorting to estimation we can initially segregate industries in terms of the relative performance of the large and small firms and this proves instructive in confirming basic intuitions concerning the potential existence of strategic groups. In the upper quartile of industries[27] when assessed in terms of the relative performance of large and small firms are 8 of the industries in the "food, drink and tobacco" sector and 4 industries centred around pharmaceuticals, man made

[26] The merits and demerits of this ad hocery are an ongoing question for industrial economics and are discussed further in Chapter 6.
[27] The upper quartile is marked by relative performance of large firms that is at least 17% greater than for small firms.

fibres and other specialised chemicals. In the latter case these industries are prominent in the UK for intensive research and development activity, in the former case Sutton (1991) has argued for a strong role for advertising in determining the structure of the industry. Indeed, both of these sectors are also heavily represented in the upper quartile of our measure of advertising intensity[28]. Therefore from this rough assessment non price competition elements would appear to be important in determining relative performance. By way of contrast, industries in the lowest quartile[29] are dominated by the metalworking, mechanical and electrical engineering and, with the exception of motor vehicles, transport sectors.

Table 1 indicates that there is a closer correlation between the price cost margins of the different industry subgroups than between the two definitions of price cost margin[30]. It is also clear, and unsurprising given the overlapping definition, that the fringe and small firm price cost margins are more highly correlated (greater than 0.95) than either is correlated with the large firms price cost margins. Between price cost margins of the same definition all of the correlation coefficients are greater than 0.75. From Table 2 we can see that with the exception of the concentration ratio and the measures of minimum efficient scale, the correlations between the independent variables are relatively low which at least points to reduced worry concerning collinearity between variable that could lead to potentially misleading estimation results. Table 3 presents the means and standard deviations of the relevant variables for the period 1984-1992.

VI.i. Analysis of the cross section (between industry) variation

The first estimation is simply an attempt to replicate the cross section results of Porter (1979) with our data set. As there are differences in variable definition between that study and this as

[28] This is also somewhat reassuring from the point of view that our measure of advertising is a fairly rough proxy. A fuller description of the industry distribution according to our measure of advertising intensity is contained in the data appendix to this Chapter.

[29] The lower quartile is marked by relative performance of small firms that is at least 6% greater than for large firms.

[30] The two definitions are as used in Chapter 3. The standard margins is defined as (net output - operative wages)/ gross output. The alternative margin has the same numerator but net output in the denominator. We have argued in Chapter 3 that it is more appropriate in empirical analysis to use the conventional margin and correct for vertical integration in other ways. Results for the net output denominator price cost margin are reported for purposes of comparison.

well as differences in the industry make up of the samples we would not expect too close a correlation between that study and the analysis presented here, but it is instructive to observe and following the Pesaran and Smith (1995) argument that this presents a potentially relatively robust means of obtaining estimates of the long run parameters. The results for this is presented in Table 4. Separate estimates are presented for the large, fringe and small firm sub-samples. In each case three different specifications are reported. Initially, the conventional measure of the price cost margin using gross output as the denominator is used as the measure of performance but following the argument in Chapter 3 correction is made for the ratio of material inputs to sales. Secondly, estimation in the absence of such a correction was employed. Thirdly, the adjusted margin using net output as the denominator was used. In addition to the reported estimates, an additional adjustment variable to potentially correct for dynamic specification was included. This variable was never significant at the 5% level and as Pesaran and Smith (1995) argue that results omitting this adjustment variable can be robust we do not report these results in detail.

In terms of an overview of the results presented in Table 4, the results for the fringe and the small firms are sufficiently similar that they can be considered together. Given the overlap in observation and the previously mentioned high correlation this should not be surprising and we will not discuss the two separately. Additionally, it is clear that the results differ substantially depending on whether or not differences in vertical integration are controlled for and we therefore discuss these results separately. The only clearly consistent result is that advertising intensity has a positive well determined effect on price cost margins for all industry subgroups and regardless of the measure/correction used. The effect is still clearly positive for the fringe or small firms although the estimated coefficients are marginally lower for these subgroups. The estimated effect of advertising is much greater without a correction for vertical integration.

Looking initially at the results with a correction, the effect of the other variables is less uniform. Large firms' price cost margins are positively and significantly related to concentration, an expected result given that concentration is the combined market share of these firms. For the small/fringe firms, concentration is not significantly different from zero in any situation, in addition, the market share of small firms has no effect on their own or large firm price cost margins, the former at least being contrary to expectation. The MES and Capital requirements control variables are again not consistent with expectations. MES has a significant negative effect on large firm price cost margins and a negative yet poorly determined effect on the fringe/small firms. This result is replicated using the alternative measure of MES although these estimations are not reported. By contrast capital requirements has no effect on the margins of any of the subgroups. The only other result is that for all subgroups the mean industry growth rate has a positive effect on price cost margins but this effect is of a larger magnitude and is only significant for the large firms.

Turning to those results that do not control for vertical integration. With the exception of advertising intensity the only well determined coefficients are capital requirements which enter negatively, and market share of the small firms which also has a negative effect, but only on the small/fringe firms. In both cases the negative effect is contrary to expectation. Concentration has a poorly determined negative coefficient as should not be surprising given the focus of Chapter 3.

In general, it would appear there are two points that can be drawn from these results. Firstly, small firms do appear to benefit from being in an advertising intensive industry contrary to Porter's results but do not benefit from being in a concentrated industry, a result consistent with Porter's results. Secondly, the similarities between the results would appear to be greater than the differences. We can investigate this more clearly using direct tests by pooling the data or using measure of relative performance. This is investigated in the results in Table 5.

Before we move on to discuss these comparative results it is necessary to consider a relevant point concerning the usefulness of analysing small firm data. It could be argued that the information on small firms may be little more than white noise because of the difficulties associated with collection of data on the smallest firms. Examining the F tests in Table 4 refutes this argument because in each case the coefficients are jointly significant at the 1% level. Therefore we can conclude that the information on small firms is not simply white noise.

Establishing this is important for the results which follow, because if the information on small firms was simply white noise with a relatively high variance it would be impossible to establish whether the determinant of small firm performance are different from large firm performance[31].

This second question is addressed in the results presented in Table 5. This presents the results of estimates of the following relations,

$$PCM_{ic} = X'\beta + DX'\Gamma + u_{ic} \quad (4)$$

PMC_{ic} is the price cost margin of size class c in industry i, X is a matrix of independent variables including a column of ones, β and Γ are vectors of coefficients and D is a dummy variable transformation with zeros for the large firm size class and ones for the small/fringe size class. The test for differences between the determinant of large and small firm performance is therefore a joint test of the parameters in Γ. Alternatively following the empirical approach adopted by Kwoka and Ravenscraft and Machin and Van Reenen we can combine all size classes into a common sample and make use of interactive terms with the mean market share of the size classes.

$$PCM_{ic} = MS_{ic}\beta_0 + X'\beta + MS_{ic}X'\Gamma + u_{ic} \quad (5)$$

This allows us to interpret the estimates as unrestricted forms of the specifications, discussed in section III.i. above, used to estimate conjectural variation elasticities. The third approach involves straightforward estimation of the linear model implied in Table 4 but modifies the dependent variable creating a relative performance measure (PCM_{il}/PCM_{is}) thereby approaching the question from a different but related direction.

Table 5 reports the results of these estimations. The results for fringe firms as before do not add anything to the results for small firms and therefore only the latter are reported. In the attempt to separate out significant differences in the determinants of performance from (4), in none of the cases is the joint test of the Γ parameters significant and indeed none are individually significant. Therefore, we must conclude that using this framework we cannot detect significant differences. This would argue against the validity of an intra industry split and in favour of an industry level

[31] It would in this case of course be impossible to establish anything with precision about the performance of small firms.

approach. Similarly a joint test of $\beta_0+\Gamma$ is never significant indicating that the effect of market share as a determinant of performance within this sample is negligible. Again the industry effects are more important determinants of performance.

VI.ii. Analysis of the pooled sample

Turning to the pooled estimates we have to make use of a dynamic specification for the analysis. As suggested earlier an autoregressive distributed lag model imposes the least restrictions upon the specification and therefore it was adopted as a way of modelling the data. The implications of a dynamic specification in a fixed effects panel data model requires estimation using first differenced instrumental variables. At this we must raise the problem of the interpretation of such a relationship. Instead of viewing this as a way of estimating a relationship with fixed effects removed, we could alternatively view it as an estimation procedure that identifies a short run relationship between variables, i.e. simply a differenced data framework, see Hendry (1995). This makes a problematic interpretation of the results. We cannot directly infer that this is a long run relationship at all in the absence of fixed effects. Baltagi and Griffin (1984, p643) conclude that "the within estimator offers a good estimator of the short run effects but can severely underestimate the long run response." We must therefore bear this in mind when interpreting the first difference estimates.

Table 7 reports the results of the first difference estimation of the relationship. Table 8 gives separate estimates for producer and consumer goods industries. Table 9 gives the results of the pooled estimates for the relative performance of large and small firms. Relative to the cross section estimates we drop the capital requirements measure from the specification. While the mean over time of investment per firm may be a reasonable proxy for capital requirements at the cross sectional level, the time varying behaviour of investment per firm cannot be expected to perform the same function.

Initial estimates used an Autoregressive Distributed Lag ADL(2,2) representation of the data, however, joint tests of the significance of the variables dated t-2 were in all cases rejected. However, Baltagi and Griffin (1984, 642-3) note the weak power of mispecification tests aimed at determining the appropriate length of lag structure in pooled models and therefore some caution has to be exercised. Adopting an ADL(1,1) framework and using the net output denominator margin, joint tests of the variables dated t-1 were only uniformly significant at the

10% level but were significant at the 1% level for the gross output margin with the inputs to sales ratio on the right hand side. Dropping the variables dated t-1 led to a failure of the Sargan statistics for each of the models, indicating model mispecification, and in the case of the net output denominator margin led to a failure of the tests for second order serial correlation, reinforcing the conclusion of mispecification at that level. We focus, therefore, on the ADL(1,1) estimates as the appropriate dynamic specification. Separate estimates were made for the net output margin and the gross output margin with input/sales ratio on the right hand side. Columns (1), (2) and (3) are estimates for the large firms, (4), (5) and (6) for the small firms. Estimates for the fringe firms were made but were sufficiently similar to the small firms estimates that they are not reported separately. In each case the estimates are free from second order serial correlation. Yearly dummy variables are included and are in each case jointly significant.

The first aspect to be considered is the degree of persistence of the performance measures, indicating the pace at which any adjustments to disequilibrium are made. In general, estimates of the coefficient on the lagged dependent variable are relatively low compared with previous estimates using firm and industry data in the UK. Studies such as Machin and Van Reenen (1993), Haskel and Martin (1994) and Geroski, Machin and Van Reenen (1993) all report coefficients in a first differenced framework between 0.4 and 0.5[32], although there is evidence that they display heterogeneity across firms and industry sectors. The estimates in columns (1) and (4) of Table 7, where the inputs to sales ratio and lagged dependent variable are instrumented but the other independent variables are treated as exogenous, our estimates are in a similar range for the large firms (0.468) and slightly lower for the small firms (0.388). If small firms are subject to a greater degree of competition than large firms we may expect their performance to be less persistent and therefore the coefficient to be lower. However, when the other right hand side variables are instrumented, significant estimates of the coefficient on the lagged dependent variable ranged between 0.2 and 0.3 but in a number of cases were lower than this and imprecisely determined. When separating out the sample into consumer and producer goods industries, only in the case of large firms in producer goods industries is there a significant

[32] Haskel and Martin (1992) report estimates more comparable with those reported here.

persistence effect. It may be the case that incorporating lagged effects of the right hand side variables, effects that are largely excluded from the previous UK studies, captures some of the persistence effects that might otherwise show up in the lagged dependent variables. In these latter estimates, we cannot identify with any degree of confidence differences between the degree of persistence of performance of large and small firms, confirming the general finding of the persistence of profits literature (see Mueller (ed.), 1990).

With respect to the right hand side variables, the most striking difference between the pooled and the cross section estimates is the lack of effect of advertising in the pooled estimates[33]. The coefficients on advertising are frequently negative, especially for the small firms, and most commonly insignificant. There are two aspects that have to be considered in interpreting this result. Firstly, advertising may well be conceived of as a fixed effect leading to product differentiation and indicating differences between industries. However, in the short run, bouts of advertising competition may reduce price cost margins for the competing firms. The short run effect may therefore dominate in the pooled estimates. The second possibility that has to be considered is measurement error. We have acknowledged that our measure of advertising intensity is a proxy that contains other information. Griliches and Hausman (1986) argue that the within estimates, in the context of measurement error, can accentuate the noise component of the measurement error and that this will both magnify the variability of estimates and bias the coefficient towards zero. This can account for the fact that the estimates are relatively poorly determined but unless the measurement error is of a sufficiently systematic nature it cannot account for the negative coefficient estimates. In addition, if there is measurement error in the advertising intensity variable instruments dated t-2 are no longer valid and instruments dated t-3 or further back must be used. However, the use of such instruments, in estimations not reported, had a negligible impact on the previously estimated coefficients on the advertising intensity variable. Therefore, we would prefer to interpret the advertising intensity relationship as a correctly (albeit imprecisely) determined short run effect.

---

[33] This finding is reported but passes without comment in Domowitz, Hubbard and Petersen (1987).

The estimates of the effects of the growth rate of industry sales are in most cases positive but are not all well defined and coefficient estimates are much smaller than the cross section estimates. The magnitude of the coefficients in the aggregate estimations are similar for small and large firms. When we separate into producer and consumer goods samples in Table 8, it is only for producer goods that there is a significant positive effect. The estimates for consumer goods are small and poorly determined. There is a tradition for assuming that the prices of producer goods may be more responsive to demand conditions because of steeper marginal cost schedules (see Kalecki (1939)) and our finding is consistent with that. From Table 9 we can see that for producer goods, the growth rate of industry sales tends to reduce the performance of large firms relative to small firms again contradicting the cross section results where the reverse is true.

With respect to the structural variables, contemporaneous concentration has a positive effect on the price cost margin of small firms and a negative effect on the price cost margin of small firms, This effect is more pronounced for producer goods industries. The lagged concentration term has the opposite effect lowering large firm price cost margins and raising small firm margins. The net effect is uniformly positive for large firms, and is normally, though not exclusively, negative for small firms. The positive impact effect on large firm price cost margins is to be expected given that rising concentration implies rising market shares for these large firms. The lagged effect implies that not all of the initial impact may be long lived however. Increases in market share may bring temporary benefits but these effects are overstated if the changes in market share do not significantly influence the other elements of structure or conduct of the other firms in the industry. The implied long run effects are of similar magnitude to the cross section effect on large firms. The rationale for the effect of concentration on small firms is less clear. The cross section estimates are positive but are not well determined. A negative coefficient may result because the increased market share of the large firms conditional on the market share of the mean market share of the small firms implies that some of the small firms may have left the industry to make way for the large firms. Therefore, increases in concentration may be coincident with pressure on the ability of small firms to survive. Again not all of the effect may be long lived if the normal pattern of conduct/performance of the industry is resumed fairly quickly.

With respect of the average market share of the small firms, the net effect is uniformly positive but with some evidence that the impact effect of contemporaneous changes in market share is offset by subsequent readjustments. This again contradicts the negative coefficients in the cross section estimation. In the estimates for all manufacturing it is noticeable that both the large and small firms benefit from a situation of larger small firms in the industry and the magnitude of the effect is greater for the large firms. However, it is noticeable that the size of the effect of market share on small firms is substantially greater in consumer goods industries than in producer goods industries. Indeed, the net effect of changes in market share in producer goods is negligibly small for small firms. This contradiction is made even more stark in Table 9 where it can be seen that increases in the market share of the small firms decrease the performance gap between large and small firms for consumer goods and increase the gap for producer goods. Minimum efficient scale has a positive effect on the performance of large firms in producer goods industries as may be expected if size and production costs matter to a greater extent in these industries, but the effect is negative otherwise.

In general there is sufficient differences of magnitude and sign of the coefficients between cross section and first difference estimates to indicate that they are not necessarily measuring the same phenomena, i.e. the long run effects of the variables. Following from the discussion in section IV above, we accept the cross section estimates as the 'true' long run and interpret the first difference estimates as a description of the short run relationships between the variables. In most cases, this is not problematic for our interpretation of intra industry performance. We can view that, in the long run, factors such as concentration and advertising intensity help to determine the framework within which prices are set, feeding through into performance. However, we can also learn something about the way in which short run changes in concentration and advertising intensity affect the short run outcome from the pre-existing framework and may have longer lived effects as the framework adjusts over time.

VII Conclusion

In conclusion we can usefully separate out the cross sectional results from the pooled estimate results. These estimates of the long run performance of the large and small firm sectors are interpreted as the degree of immunity of these sectors from competitive pressures especially in the form of entry pressures either into the industry or into the relevant strategic group. The results confirm that the main observable determinant of this long run performance is advertising intensity and that this is a quantitatively more important factor for the larger firms than the smaller firms. The share of the large firms in the industry is the other key determinant of the relative long run performance of the large and small firms. The other potential determinants do not appear in this data set to be important measures of long run performance.

Examining the pooled first difference estimates we can see that the disappearance of the effect of advertising intensity effect in these estimates is evidence, following the argument of Mundlak (1978), of the correlation of the fixed effects with advertising intensity. The results of the pooled estimates do not point to any consistent determinants of performance and indeed point to important differences in performance between sub sections of manufacturing industry. The short run performance of consumer goods industries does not appear to be readily explainable in terms of the structural variables used. There is greater success in this endeavour in relation to producer goods industries indicating that the share of the small and large firm groups and the growth rate of the industry are all important in determining short run intra industry performance. These results indicate that for producer goods industries, at least, it is true to say that short run performance of the groups is affected by the relative strengths of the small and large firm groups.

We can use these results to readdress the stylised facts suggested by Schmalensee (1989a) and considered in section III above. We can provide evidence in favour of the long run validity of 4.10 in our sample. Large firms benefit unambiguously from high concentration but small firms receive weak positive benefit when assessed in cross section and the impact is broadly negative in first differences. We can also provide some support for the assertion in 4.11 that in cross industry sample, market share has a positive impact on performance in the short run, but in the long run, small firms do not, in our sample, benefit from higher market shares, the effect in cross section being small but negative. 4.12 states that within industries profitability is not strongly related to market share. Our test on relative performance indicate that this is not true with respect to producer goods industries where increasing market shares for the small firms reduces the performance gap and increasing market shares for large firms increases the gap. For consumer goods industries we must accept that no such relationship exists.

Table 1: Correlations between Price Cost Margins (1984-1992)

	lpcm	fpcm	spcm	lpcml	fpcml	spcml
lpcm	1.0000					
fpcm	0.7387	1.0000				
pcm	0.6891	0.9091	1.0000			
lpcml	0.5125	0.2390	0.2010	1.0000		
fpcml	0.3681	0.4056	0.3304	0.7323	1.0000	
spcml	0.3124	0.2931	0.3276	0.7104	0.9027	1.0000

Table 2: Correlations between Independent Variables (1984-1992)

	conc	mes	ms	adint	grow	cap
conc	1.0000					
mes	0.6769	1.0000				
ms	0.1923	0.4463	1.0000			
adint	0.1866	0.1429	-0.0752	1.0000		
grow	-0.0875	-0.0715	-0.1047	-0.0622	1.0000	
cap	0.1712	0.0941	0.0247	0.1261	-0.0186	1.0000

Table 3: Means and Standard Deviations (1984-1992)

Variable	Mean (Standard Deviation)
spcm	0.306(0.084)
spcml	0.706(0.085)
fpcm	0.310(0.088)
fpcml	0.714(0.091)
lpcm	0.325(0.109)
lpcml	0.718(0.113)
conc	0.446(0.226)
mes	0.728(0.678)
ms	0.011(0.027)
adint	0.149(0.063)
gr	0.047(0.099)
cap	0.336(1.162)

All variables are as defined in the Data Appendix to this Chapter.

Table 4: Cross Section Estimates of Intra Industry Performance (1984-1992)

	MES	CAP	CONC	MY	MS	GROW	ADINT	CONS	F TEST
Large Firms									
PCM	-0.025[b]	0.003	0.092[a]	-0.825[a]	-0.003	0.375[b]	0.428[a]	0.673[a]	15.80
	(.012)	(.009)	(.029)	(.054)	(.140)	(.175)	(.072)	(.032)	(6,79)
PCM	-0.012	-0.016[b]	-0.019		-0.048	0.558	2.052[a]	0.190[a]	10.06
	(.016)	(.008)	(.046)		(.318)	(.351)	(.332)	(.031)	(6.80)
PCM1	-0.044[c]	-0.001	0.180[a]		-0.333	0.743[c]	0.923[a]	0.501[a]	20.93
	(.024)	(.018)	(.057)		(.288)	(.387)	(.156)	(.031)	(6,80)
Fringe Firms									
PCM	-0.006	0.001	0.037	-0.751[a]	-0.045	0.218	0.311[a]	0.666[a]	11.05
	(.011)	(.008)	(.025)	(.035)	(.108)	(.136)	(.064)	(.022)	(6,79)
PCM	0.004	-0.024[a]	-0.031		-0.377[b]	0.454	1.591[a]	0.209[a]	18.01
	(.015)	(.008)	(.046)		(.188)	(.274)	(.290)	(.023)	(6,80)
PCM1	-0.009	0.0001	0.076		-0.350	0.427	0.734[a]	0.561[a]	15.91
	(.023)	(.020)	(.054)		(.229)	(.296)	(.134)	(.027)	(6,80
Small Firms									
PCM	-0.006	-0.004	0.039	-0.699[a]	-0.067	0.175	0.276[a]	0.641[a]	8.66
	(.010)	(.008)	(.027)	(.031)	(.105)	(.149)	(.064)	(.019)	(6,79)
PCM	0.013	-0.023[b]	-0.036		-0.502[a]	0.402	1.331[a]	0.222[a]	16.46
	(.015)	(.009)	(.051)		(.169)	(.298)	(.316)	(.024)	(6,80)
PCM1	-0.016	-0.012	0.099[c]		-0.264	0.303	0.693[a]	0.563[a]	12.18
	(.022)	(.021)	(.055)		(.217)	(.306)	(.130)	(.025)	(6,80)

Notes:

1. Variable descriptions are contained in the Data Appendix. MY is the ratio of material inputs to gross output for each industry subgroup.
2. Huber heteroskedasticity consistent standard errors are in parentheses.
3. F Test is a test of the joint significance of the independent variables, excluding MY. Degrees of freedom are in parentheses. All of the F Tests are significant at the 1% level.

Table 5 Cross Section Estimates: Interactive specifications (1984-1992)

Independent. Variables	(1) PCM	(2) PCM1	Independent. Variables	(3) PCM	(4) PCM1
D	0.027	0.062	MES	-0.008	-0.012
	(.019)	(.040)		(.010)	(.020)
MES	-0.024^{b}	-0.044^{c}	CAP	0.001	-0.007
	(.011)	(.024)		(.011)	(.024)
CAP	0.001	-0.001	MY	-0.774^{a}	
	(.008)	(.018)		(.034)	
CONC	0.084^{a}	0.181^{a}	MSHARE	-0.049	-0.426
	(.027)	(.057)		(.168)	(.355)
MY	-0.771^{a}		GROW	0.160	0.361
	(.034)			(.152)	(.308)
MS	-0.032	-0.333	CONC	0.039	0.076
	(.129)	(.288)		(.027)	(.052)
GROW	0.410^{b}	0.743^{c}	ADINT	0.359^{a}	0.736^{a}
	(.174)	(.387)		(.065)	(.136)
ADINT	0.406^{a}	0.923^{a}	ADINT*MSH	-0.032	1.179
	(.072)	(.156)		(.830)	(1.83)
MES*D	0.015	0.029	CONC*MSH	0.310	1.044^{c}
	(.015)	(.032)		(.287)	(.601)
CONC*D	-0.038	-0.082	MES*MSH	-0.126	-0.377^{b}
	(.039)	(.079)		(.091)	(.189)
GROW*D	-0.285	-0.440	GROW*MSH	1.987	3.551
	(.235)	(.493)		(2.04)	(4.77)
MS*D	0.043	0.070	CAP*MSH	-0.016	-0.024
	(.165)	(.361)		(.075)	(.169)
ADINT*D	-0.088	-0.230	CONS	0.670^{a}	0.557^{a}
	(.096)	(.203)		(.021)	(.026)
CAP*D	-0.003	-0.010	$\bar{R}^2$	0.8373	0.3442
	(.012)	(.028)			
CONS	0.647^{a}	0.501^{a}	F-Test	0.93 (6,161)	1.42 (6,162)
	(.024)	(.031)			
$\bar{R}^2$	0.8378	0.3426			
F-Test	1.70 (6,159)	1.54 (6,160)			

Notes:

1. Variable descriptions are contained in the Data Appendix.
2. Huber heteroskedasticity consistent standard errors are in parentheses. a b c indicate a coefficient significant at the 1, 5, and 10% levels respectively.
3. F Test is a test of the joint significance of the interaction terms for (1) and (2) and the joint significance of MSHARE and the interaction terms in (3) and (4). Degrees of freedom are in parentheses. None of the F Tests are significant at the 10% level.

Table 6: Cross Section Estimates of Relative Performance (1984-1992)

Independent Variables	(1) RPCM	(2) RPCM	(3) RPCM1	(4) RPCM1
MES	-0.058[c]	-0.045[c]	-0.040[c]	-0.048[b]
	(.033)	(.026)	(.024)	(.022)
CAP	-0.008	-0.009	0.018	-0.002
	(.016)	(.015)	(.011)	(.010)
CONC	0.218[c]	0.177[b]	0.120[c]	0.152[a]
	(.111)	(.080)	(.072)	(.055)
RMY	-1.187[a]	-1.259[a]		
	(.123)	(.118)		
MS	0.179	0.175	-0.220	-0.112
	(.422)	(.380)	(.351)	(.353)
GROW	0.736	0.152	0.604[b]	0.389[c]
	(.465)	(.376)	(.301)	(.229)
ADINT	0.623[b]	0.747[a]	0.272	0.217
	(.293)	(.280)	(.181)	(.179)
CONSTANT	2.046[a]	2.117[a]	0.919[a]	0.923[a]
	(.138)	(.119)	(.032)	(.031)
$\bar{R}^2$	0.6833	0.6792	0.1160	0.0764
F-Test	5.43[a]	3.21[a]	12.18[a]	7.75[a]
	(6,79)	(6,79)	(6,80)	(6,80)

Notes:

1. Variable descriptions are contained in the Data Appendix. RPCM is the price cost margin of large firms/ price cost margins of small(fringe) firms. RPCM1 is the same ratio but for a net output denominator price cost margin. (1) and (3) are the ratio with small firms (2) and (4) with fringe firms. RMY is the ratio of material inputs to gross output for the large firms/ the ratio of material inputs to gross output for the small(fringe) firms.
2. Huber heteroskedasticity consistent standard errors are in parentheses. a b c indicate a coefficient significant at the 1, 5, and 10% levels respectively.
3. F Test is a test of the joint significance of the independent variables, excluding RMY. Degrees of freedom are in parentheses. All of the F Tests are significant at the 1% level.

Table 7: First Difference Estimates (1985-1992)

Independent Variables	(1) LPCM(t)	(2) LPCM(t)	(3) LPCM(t)	(4) SPCM(t)	(5) SPCM(t)	(6) SPCM(t)
MES(t)	0.011 (2.22)	-0.004 (0.37)	-0.013 (0.43)	-0.002 (0.66)	-0.021 (2.16)	-0.050 (1.86)
MES(t-1)	-0.009 (1.85)	-0.010 (1.01)	-0.013 (0.46)	-0.008 (1.51)	0.012 (1.37)	0.037 (1.28)
GROW(t)	0.042 (3.24)	0.038 (1.38)	0.058 (0.87)	0.040 (4.27)	0.028 (1.43)	0.030 (0.52)
GROW(t-1)	0.021 (1.54)	0.012 (0.88)	-0.0003 (0.010)	0.011 (1.45)	0.005 (0.59)	0.021 (0.99)
ADINT(t)	0.046 (1.22)	-0.151 (1.43)	-0.419 (1.94)	-0.026 (0.84)	-0.029 (0.42)	-0.397 (1.35)
ADINT(t-1)	0.045 (1.39)	0.072 (1.07)	-0.092 (0.53)	0.016 (0.73)	-0.030 (0.61)	0.0006 (0.004)
MS(t)	0.352 (4.52)	0.069 (0.96)	0.219 (1.69)	0.131 (1.57)	0.142 (2.37)	0.208 (1.58)
MS(t-1)	-0.148 (1.49)	0.150 (1.58)	0.712 (2.12)	0.008 (0.06)	-0.063 (0.65)	0.272 (1.00)
CONC(t)	0.098 (4.50)	0.085 (1.67)	0.219 (1.57)	-0.057 (3.00)	-0.065 (1.32)	-0.107 (0.74)
CONC(t-1)	-0.063 (2.45)	-0.024 (0.48)	0.021 (0.15)	0.032 (1.61)	0.057 (1.40)	0.216 (2.00)
MY(t)	-1.020 (8.11)	-1.001 (16.5)		-0.835 (9.55)	-0.883 (16.3)	
MY(t-1)	0.494 (3.09)	0.297 (1.96)		0.351 (3.25)	0.249 (2.24)	
PCM(t-1)	0.468 (2.94)	0.243 (1.66)	0.032 (0.22)	0.388 (3.96)	0.261 (2.36)	0.192 (1.94)
Serial Correlation	*-0.182*	*-0.176*	*-1.675*	*0.381*	*0.253*	*0.109*
Sargan	*19.7(13)*	*50.6(43)*	*76.1(45)*	*7.9(13)*	*45.8(43)*	*45.6(45)*
Time Dums	*22.1(8)*	*16.7(8)*	*26.6(8)*	*41.9(8)*	*38.2(8)*	*43.6(8)*
Wald	*73.2(10)*	*35.2(10)*	*45.0(10)*	*35.0(10)*	*12.2(12)*	*13.2(10)*

Notes:

1. Variable definitions are contained in the Data Appendix. Dependent variable in (1),(2),(4) and (5) is standard price cost margin. (3) and (6) the net output denominator margin.
2. Estimation is by the first difference instrumental variables GMM method outlined in Arrellano and Bond (1988). In (1) and (4) only MY(t), MY(t-1) and PCM(t-1) are instrumented. In (2),(3),(5) and (6) all right hand side variables are instrumented. Instruments are all independent variables and the lagged dependent variable dated t-2 and back.
3. Asymptotically consistent and robust t ratios are in parentheses.
4. *Serial Correlation* is a robust test of second order serial correlation distributed as N(0,1). *Sargan* is the test of overidentifying restrictions degrees of freedom in parentheses. *Time Dums* is a Wald test of the joint significance of the yearly dummy variables. *Wald* is a joint test of the independent variables excluding MY.

Table 8: First Difference Estimates: Producer/Consumer Goods (1985-1992)

Independent Variables	Consumer Goods (1) LPCM(t)	(2) SPCM(t)	Producer Goods (3) LPCM(t)	(4) SPCM(t)
MES(t)	-0.014	-0.022	0.029	-0.010
	(0.91)	(1.81)	(2.47)	(1.02)
MES(t-1)	-0.013	0.005	-0.012	-0.003
	(1.26)	(0.72)	(1.29)	(0.31)
GROW(t)	-0.016	0.010	0.068	0.045
	(0.76)	(0.61)	(3.98)	(2.31)
GROW(t-1)	0.004	0.011	0.019	0.017
	(0.29)	(0.645)	(1.10)	(1.99)
ADINT(t)	-0.091	-0.021	-0.041	-0.124
	(1.20)	(0.52)	(0.41)	(1.52)
ADINT(t-1)	-0.041	-0.020	0.045	-0.034
	(0.56)	(0.45)	(0.89)	(0.58)
MS(t)	0.282	1.110	0.112	0.135
	(0.432)	(1.73)	(2.08)	(3.55)
MS(t-1)	0.099	1.015	0.055	-0.123
	(0.094)	(1.56)	(0.92)	(1.28)
CONC(t)	0.078	-0.015	0.097	-0.105
	(0.94)	(0.31)	(3.03)	(3.09)
CONC(t-1)	0.020	0.048	-0.065	0.030
	(0.38)	(0.95)	(1.47)	(0.99)
MY(t)	-0.896	-0.915	-0.909	-0.969
	(17.9)	(15.2)	(21.6)	(14.9)
MY(t-1)	0.099	0.167	0.244	0.106
	(0.53)	(1.26)	(2.46)	(0.58)
PCM(t-1)	0.050	0.150	0.217	0.102
	(0.264)	(1.07)	(2.107)	(0.60)
Serial Correlation	*0.390*	*1.494*	*-1.378*	*-0.940*
Sargan	*60.7(43)*	*58.9(43)*	*67.4(43)*	*54.4(43)*
Time Dums	*16.1(8)*	*36.5(8)*	*20.9(8)*	*23.9(8)*
Wald	*7.9(10)*	*20.8(10)*	*81.9(10)*	*33.8(10)*
Observations	*328*	*328*	*368*	*368*
Industries	*41*	*41*	*46*	*46*

Notes: 1. Dependent variable is the standard gross output price cost margin.
2. Details of the estimation procedure and diagnostic tests are as in notes 2, 3 and 4 to Table 7.

Table 9 First Difference Estimates: Relative Performance (1985-1992)

Independent Variables	All (1) RPCM(t)	Consumer (2) RPCM(t)	Producer (3) RPCM(t)
MES(t)	-0.007 (0.05)	0.071 (0.60)	0.018 (0.14)
MES(t-1)	-0.256 (2.23)	-0.223 (2.03)	-0.092 (1.11)
GROW(t)	-0.352 (1.38)	-0.168 (0.64)	-0.187 (1.19)
GROW(t-1)	0.027 (0.30)	0.004 (0.030)	-0.125 (1.37)
ADINT(t)	-0.437 (0.49)	-0.281 (0.35)	-1.059 (1.26)
ADINT(t-1)	-0.574 (0.77)	0.178 (0.35)	0.458 (0.68)
MS(t)	-0.441 (0.41)	-2.729 (0.34)	0.129 (0.25)
MS(t-1)	2.525 (1.81)	-10.10 (1.07)	1.801 (3.28)
CONC(t)	0.795 (1.08)	0.446 (0.72)	0.688 (1.74)
CONC(t-1)	-0.791 (1.38)	0.397 (0.71)	-0.670 (1.59)
RMY(t)	-2.288 (4.93)	-2.238 (6.08)	-1.977 (6.80)
RMY(t-1)	0.375 (2.17)	-0.029 (0.22)	0.127 (0.63)
RPCM(t-1)	0.106 (1.43)	0.015 (0.18)	-0.024 (0.28)
Serial Correlation	*0.366*	*0.828*	*-0.314*
Sargan	*36.8(43)*	*73.5(43)*	*56.0(43)*
Time Dums	*5.9(8)*	*7.2(8)*	*7.0(8)*
Wald	*27.1(10)*	*22.7(10)*	*61.2(12)*
Observations	*696*	*328*	*368*
Industries	*87*	*41*	*46*

Notes: 1. Dependent variable is the ratio of the standard price cost margin of the large firms to the price cost margins of the small firms.
2. Details of the estimation procedure and diagnostic tests are as in notes 2, 3 and 4 to Table 7.

Data Appendix

Data is taken from the UK census of production for the relevant years. Variables are as defined in the data appendix to Chapter 3 with the exception of those listed below.

The data on intra industry groups is taken from the size class analyses (Table 13) of the UK Census of Production. The "large" firm group is the largest five firms. The "small" firm group is those firms with between 1-99 employees. The "fringe" firm group is obtained by subtracting the large firms from the industry totals.

MS: The mean market share by sales of the firms with 1-99 employees
MES: The measure of minimum efficient scale is the average sales per establishment of establishments operated by the largest five firms.
AMES: The alternative measure of minimum efficient scale is the average sales per establishment of establishments operated by the fringe firms.
CAP: Capital requirements is measured as the average net capital expenditure per establishment of the establishments operated by the fringe firms
ADINT: The advertising intensity variable using is the ratio of "cost of other non industrial services" to net/gross output at the industry level. Clearly the interpretation of the results on this variable are important for the chapter as a whole. Therefore need to assure ourselves that this is indeed a good proxy for advertising intensity. Through analysis of the distribution of industries according to this variable we can attempt to see whether it is indeed approximating advertising intensity. The industries in the upper quartile in terms of the mean of Adint measured in terms of gross output include "Pharmaceuticals", "Soap and toilet preparations", "Ice cream, cocoa, chocolate and sugar confectionery", "Spirit distilling and compounding", "Wines, cider and perry" and "Toys and sports goods". All of these would intuitively be on most lists of relatively high advertising industries. However deflating by gross output is potentially problematic for industries with a low ratio of net output to gross output. Using the measure of Adint deflated by net output adds to the previous list "Domestic-type electrical appliances", "Motor vehicles and their engines", "Soft drinks" and "Tobacco industry". Again these are consumer goods industries commonly associated with advertising expenditures. We would therefore argue that this proxy for advertising intensity is actually picking up the desired effects and can indeed be interpreted as such.

Methodological Appendix: Dynamic Panel Data Estimation

This appendix briefly describes the appropriate estimation procedure for a dynamic panel data models with fixed effects. Fuller descriptions are given in Hsiao (1986; p71-6) and Baltagi (1995; p125-132)

The model is a straightforward linear model with a single lag of the dependent variable on the right hand side. Adding exogenous independent variables is statistically relatively unproblematic and therefore ignored. There are observations for T time periods on N industries.

$$y_{it} = \beta + \gamma y_{i,t-1} + \eta_{it} \qquad i=1,......N \quad t=1,.....T$$

β is a constant and γ is the parameter of interest. We can decompose the error term into three elements,
$\eta_{it} = \alpha_i + v_t + \varepsilon_{it}$ where α and ν are the industry and time specific effects respectively that are constant across time and industries. ε represent unobserved effects over both time and industry.

The conventional way of estimating such relationships was to use least squares but incorporate dummy variables for each industry and time period or equivalently estimate least squares in mean deviation form. Early estimates of this form such as Balestra and Nerlove (1966) produced relative low estimates of the γ parameter. Nickell (1981) identified that estimating such relations would result in a biased estimate of γ if T was small even if N→∞. This bias results from correlation between the mean deviation of the lagged dependent variable and the error term, a correlation that only tends to zero as T→∞. The solution initially proposed to this problem by Anderson and Hsiao (1982) was to estimate the model in first differences of the form described below, suppressing the time dummies from the exposition.

$$(y_{it} - y_{i,t-1}) = \gamma(y_{i,t-1} - y_{i,t-2}) + (\varepsilon_{i,t} - \varepsilon_{i,t-1})$$

Because of the correlation between $(y_{i,t-1} - y_{i,t-2})$ and $(\varepsilon_{i,t} - \varepsilon_{i,t-1})$ instruments must be found and Anderson and Hsiao suggested both $y_{i,t-2}$ and $(y_{i,t-2} - y_{i,t-3})$ were valid instruments hence either or both could be used to estimate γ accurately.

Arrellano and Bond (1988, 1991) subsequently questioned the efficiency of the Anderson Hsiao estimator and developed a Generalised Method of Moments (GMM) estimator that made use of a larger instrument set by making use of further lags of the lagged dependent variable as they became available, i.e. towards the end of the data set more lags exist and using them as instruments enhances the efficiency of the estimator. It is this last estimator that is used in this and the subsequent chapter when estimating dynamic models in the pooled sample.

This framework may also be used for the estimation of general instrumental variable models in the absence of dynamic specification to provide efficient estimates in the presence of correlation between contemporaneous right hand side variables with the error term. (See Blundell *et al* (1992) for a use in this context) As long as ε_{it} is serially uncorrelated lagged values of the endogenous right hand side variable dated t-2 and further back are valid instruments. First differencing generates a MA(1) error process therefore tests of higher order serial correlation are necessary in order to establish the validity of the estimation procedure. In each case robust tests of second order serial correlation are reported.

Making use of the additional lags as instruments means that the equation will be overidentified as there will be too many instruments relative to the number of included variables, or in the method of moments context, there too many moment equations to solve in order to identify the parameters. In order to identify the parameters some of the moment restrictions generating these excess equations have to be relaxed. Sargan (1958) has presented a test of the validity of the overidentifying restrictions and this test is reported together with the parameter estimates.

Chapter 5: The Dynamic and Cyclical Behaviour of Intra Industry Performance

I. Introduction

This chapter seeks to examine how the performance of the size groups of firms within industries may vary across the business cycle. While the cyclical performance of firms and industries has recently been a prominent topic for discussion within industrial economics and macroeconomics, or at least some variants thereof, especially new Keynesian macroeconomics, there is a dearth of studies that look at intra industry performance in this context. The only study we are aware of is the study of US firms in Schmalensee (1989b). This study will be discussed in this chapter and we will produce estimates for the UK that may be compared with Schmalensee's results. As noted in the previous chapter, one of the reasons for this absence of previous work is the limited number of data sets that contain the requisite information. In this chapter we combine the panel of intra industry groups in UK manufacturing used for the previous chapter with data relating to cyclical movements at the industry level that has been taken from the C.B.I. Industrial Trends survey. This survey presents data on capacity levels and adequacy of inventories that enables us to study the issues in greater detail than merely relying on conventional measures of the cycle such as the growth rate of industry sales.

The structure of the chapter is as follows: firstly we will discuss the theoretical models that generate expectations concerning differences or similarities in the cyclical performance of large and small firms; secondly we selectively[1] summarise some of the previous empirical studies looking at cyclical performance; thirdly we present the empirical frameworks which form the basis for our analyses of the data; fourthly we present the results from the various empirical analyses; finally we offer some conclusions based upon these results.

[1] Relatively comprehensive coverage of the studies up to the end of the 1980's is contained in Schmalensee (1989a) and Carlton (1989) therefore we concentrate upon the more recent studies in our survey.

II. Theories of cyclical performance

In considering the determinants of cyclical performance we can usefully separate the short run from the long run effects. Our focus is mainly on the short run determinants but we have to take into account, in addition, the way that long run effects such as the effects of entry can vary in their influence over the business cycle.

II.i. Demand fluctuations and short run profit maximisation

One of the simplest ways to analyse cyclical movements and their effect on performance is to represent demand fluctuations by a change in the elasticity of demand. This argument follows Harrod (1936) in assuming that, in addition to a secular fall over time in the absolute value of the elasticity of demand, demand will become more elastic in the slump and more inelastic in booms. The reasoning behind this assumption is that demand becomes more (in)elastic as consumers become more (less) price conscious due to lower (higher) or more (less) uncertain real incomes[2]. A more elastic demand schedule can be conceived of in two ways: firstly as a flatter demand schedule as substitutes become more attractive, we can call this demand shift type I; alternatively, a shift inwards of the demand curve, due to a change in real incomes, but retaining a constant slope, which will also lead to a higher elasticity, this can be called demand shift type II.

We start with the conventional formula for empirical price cost margins based upon short run profit maximisation. Initially, therefore, we are focusing on essentially comparative static analysis of the determination of price cost margins. Under the conventional assumption that marginal cost is constant the empirical price cost margin is equal to (1).

$$\frac{p-avc_i}{p}=\frac{s_i(1+\lambda_i)}{\eta} \quad (1)$$

[2] Kalecki (1939 p35) cites Harrod as suggesting "In the slump consumers 'resent and resist the curtailment of their wonted pleasures....Their efforts to find cheapness become strenuous and eager. Nor are commercial firms exempt from this influence upon their purchase policy; they, too, have received a nasty jolt and must strain every nerve to reduce costs.' "

An alternative formulation occurs if the cost function is of the quadratic form used by Mills and Schumann (see chapter 2),

$$c(x_i) = \alpha_i + \beta_i x_i + \frac{x_i^2}{2\gamma_i} \quad (2)$$

leading to a price cost margin of the form,

$$\frac{p - avc_i}{p} = \frac{s_i(1 + \lambda_i + \kappa_i)}{\eta} \quad (3)$$

In (3) $\kappa_i = -X'(p)/2\gamma_i$ and $X'(p)$ is the slope of the demand curve. Therefore κ_i is constant under the assumption of a demand curve with constant slope, demand type II, and is variable otherwise. Note that constant marginal cost in this framework implies that marginal cost is equal to average variable cost and $\gamma_i \to \infty$ so $\kappa_i = 0$. Note also that κ_i can be negative if marginal cost is falling, i.e. $\gamma < 0$. Taking the total derivative of (3) with respect to η yields (4) if the demand shift is of type II and (5) in the case of a type I demand shift.

$$\frac{dPCM_i}{d\eta} = -\frac{1}{\eta}\frac{s_i(1 + \kappa_i + \lambda_i)}{\eta} + \frac{\partial s_i}{\partial \eta}\frac{(1 + \kappa_i + \lambda_i)}{\eta} + \frac{\partial \lambda_i}{\partial \eta}\frac{s_i}{\eta} \quad (4)$$

$$\frac{dPCM_i}{d\eta} = -\frac{1}{\eta}\frac{s_i(1 + \kappa_i + \lambda_i)}{\eta} + \frac{\partial s_i}{\partial \eta}\frac{(1 + \kappa_i + \lambda_i)}{\eta} + \frac{\partial \kappa_i}{\partial \eta}\frac{s_i}{\eta} + \frac{\partial \lambda_i}{\partial \eta}\frac{s_i}{\eta} \quad (5)$$

The relations are fairly complex and therefore we need to be careful in interpreting them. The first term in both (4) and (5) is the direct effect and implies that as demand becomes more elastic the fall in price cost margins will be larger for firms with higher price cost margins, which *ceteris paribus* will be the larger firms. In the second term, the indirect effect of a change in demand causing a change in market share, the sign of $\partial s_i/\partial \eta$ will clearly vary from firm to firm. Following from the discussion of output variation in Chapter 2 we would expect that as large firms are less prone to fluctuations their market shares will rise in recessions and fall in booms. Therefore $\partial s_i/\partial \eta$ will be positive for large firms and will be negative for small firms. Conditional on the assumption that the final two terms in (5) are zero[3], this indicates that the sign of the change in margins for small firms is (theoretically) unambiguous and negative. The sign for large

[3] This implicitly assumes that either the demand shock is of type I or marginal costs are constant and also assumes that demand changes do not affect competitive stance.

firms is however ambiguous as the indirect effect of a change in market share could, in principle, outweigh the direct effect so that $\partial s_i/\partial\eta - s_i/\eta$ becomes positive. This will be the case if the elasticity of market share with respect to the price elasticity of demand is greater than unity. Therefore, the conclusion must be that, if there is no impact of demand movements on market shares, i.e. $\partial s_i/\partial\eta=0$, the price cost margins of large firms will be pro-cyclical and will be more pro-cyclical than those of small firms. If the prediction and empirical observation of chapter 2 is correct, so that $\partial s_i/\partial\eta>0$ for large firms, this effect is reduced to the extent that it is possible for large firms to exhibit counter cyclical price cost margins and therefore for the ratio of large firm price cost margins to small firm price cost margins to be counter cyclical.

To complete the picture we also need in addition to introduce the possibility that either the degree of collusion will vary with the elasticity of demand[4] or that the degree of distortion contained in measured price cost margins will vary with demand. Citing Kalecki (1939, p35) again "in the slump, cartels are created to save profits", "entrepreneurs avoid price cuts because it may induce their competitors to do likewise" and "cartels are not afraid that outsiders will appear". These arguments have been subsequently reiterated by Cowling (1983). Cowling suggests that although the initial response to an increase in the degree of excess capacity resulting from the onset of recession may be the cutting of margins in order to restore planned levels of inventory, a subject considered in more detail in subsection II.iv. below, the most prevalent medium and long term response to slump is to reinforce the degree of monopoly maintaining oligopolistic prices. However, Cowling considers an alternative to this norm, cut throat competition in order to eliminate rivals, a suggestion also considered by Kalecki. For this to occur Cowling (p 345) argues that there has to be "some fundamental asymmetry in the oligopoly group which was either not recognised or not acted on [previously] ...". While, as Cowling states, this is "unlikely to constitute the general case", it could form an important subset of cases.

---

4 Clearly we do not have to pose a direct relationship. The degree of collusion is intuitively more likely to respond to the factors associated with an up or downturn in economic activity, that cause the elasticity of demand to change, rather than respond directly to the change in elasticity of demand.

For simplicity of analysis we now assume that $\partial s_i/\partial\eta=0$ to eliminate the second term from (5). We can therefore rewrite (5) as (6).

$$\frac{dPCM_i}{d\eta}=-\frac{s_i}{\eta}\left(\frac{1}{\eta}+\frac{\kappa_i}{\eta}-\frac{\partial\kappa_i}{\partial\eta}+\frac{\lambda_i}{\eta}-\frac{\partial\lambda_i}{\partial\eta}\right)=-\frac{s_i}{\eta}\left(\frac{1+\kappa_i(1-\sigma_\kappa)+\lambda_i(1-\sigma_\lambda)}{\eta}\right) \quad (6)$$

From (6) we can see that ordinarily the price cost margin will be more pro cyclical the larger is the price cost margin. The extent of the counteracting effects depends upon σ_K and σ_λ, the elasticities of κ and λ with respect to the price elasticity of demand. Type I demand changes require that σ_K be positive but by definition this elasticity cannot vary between firms[5]. We can see from (6) that in order for the change in the degree of collusion to create countercyclical price cost margins the following relation must be satisfied[6], $\sigma_\lambda>>1$. This implies that the increase in the conjecture term must be more than proportionate to the increase in the elasticity of demand. Taken separately it would appear that to generate countercyclical price cost margins it is necessary for either the conjectural term or market share to respond elastically to changes in the price elasticity of demand. Alternatively the effects could act in combination, reinforcing each other, although logically, increases in the degree of collusion on the part of an individual firm would tend to reduce rather than increase market share.

As a point of note we can also investigate from these relations the theoretical bias that results from using average variable cost rather than marginal cost. The difference between the two measures is expressed in (7)

$$\frac{mc_i-avc_i}{p}=\frac{s_i\kappa_i}{\eta} \quad (7)$$

The total derivative of which can be expressed as (8).

$$\frac{d\left(mc_i-avc_i\right)/p}{d\eta}=-\frac{s_i}{\eta}\frac{\kappa_i}{\eta}+\frac{s_i}{\eta}\frac{\partial\kappa_i}{\partial\eta}+\frac{\kappa_i}{\eta}\frac{\partial s_i}{\partial\eta} \quad (8)$$

[5] This is because the *proportionate* change in κ only depends upon the effect on the slope of the demand curve which is common to all firms. This interpretation of (6) does point to the theoretical possibility that counter cyclical measured price cost margins could result from a high value of σ_K i.e. a large change in distortion between high and low demand states.

[6] In this situation, where $\partial s_i/\partial\eta=0$, the condition, $\sigma_\lambda>>1$, is necessary but not sufficient.

It can be readily established that (8) is negative (i.e. the difference is procyclical) unless either market share or κ are positively elastic in respect of the price elasticity of demand. As neither of these cases can be assumed with any degree of generality it can be inferred that, allowing for the stated exceptions, the tendency is that using average variable cost in the measured margin will lead to a procyclical bias. This is intuitively the case if one considers studies such as Bils (1987) where marginal cost rises very steeply in times of booms and therefore price - marginal cost values are low at this point in the cycle compared to the relation between price and average variable cost. Such is also the case when considering the case of formal capacity constraints, where at the capacity point marginal cost is infinite, price is determined solely by demand, and price may therefore increase without limit relative to average variable cost in periods of high demand. Of greater import for our focus is whether the bias is greater or lesser depending upon firm size. From (8) we can state that if $\partial s_i/\partial\eta=0$ the procyclical bias will be greater for firms with larger market shares. However, in this case, large firms are also predicted to be have more procyclical price cost margins in the standard framework. Also if $\partial s_i/\partial\eta\neq0$ but $\partial\kappa_i/\partial\eta=0$ the direction of bias follows the large firms. If large firms' price cost margins are procyclical (countercyclical) the bias is also procyclical (countercyclical). Therefore we can see that the bias will tend to amplify or overstate the 'true' relative time paths rather than confuse them.

II.ii. Capacity constraints and short run profit maximisation

Before incorporating long run factors into this discussion we consider the role of capacity constraints in pricing decisions. The argument that firms are subject to a capacity constraint is relatively rarely advanced in empirical industrial organisation studies. This seems rather surprising given the early focus of models of imperfect competition on the relationship between excess capacity and imperfect competition (see Steindl 1952). In recent years the main study, comparable in form to the current study, that has investigated the issue of capacity constraints has been Haskel and Martin (1994). The way in which their model is theorised is fairly restrictive. They borrow from Bresnahan's (1982) consistent conjectures model and approximate capacity constraints with the slope of the marginal cost curve. In the same vein as the Kreps and

Scheinkman (1983) model, as firms approach capacity constraints, firms output levels more closely approximate Cournot rather than Bertrand levels. In this way it can be seen that capacity constraints influence the consistent conjecture adopted by the firm and therefore feed through to competitive stance. We would argue however that the equation of a change in the implicit conjecture with a change in competitive stance is a misleading confusion of conduct for performance. The output restriction, in the case of a binding capacity constraint, is involuntary, at least in the short run. The firm wishes to increase its output level but is unable to do so, and therefore the implied lower level of output is of a different order to output restrictions associated with monopolistic conduct[7].

The role of capacity utilisation or excess capacity in facilitating the oligopolistic co-ordination which in turn determines the price cost margin has been a subject of contention. Spence (1977) argues that the existence of excess capacity gives a greater incentive to deviate from any collusive agreement and will therefore reduce the mark-up. It has also been suggested however that excess capacity is a double edged weapon increasing both the incentive to deviate and the power to retaliate against deviators. Cowling (1983) argues that although the initial response to the creation of excess capacity may be the weakening of collusion when firms recognise their mutual adversity the degree of collusion hence the mark-up will increase.

An alternative way to approach the question of capacity constraints is simply to examine the constrained short run profit maximising decision of the firm[8].

The firm chooses output level, q, to maximise,

$\pi = pq - c(q) - F$ s.t. $q \leq k$ where k is the level of capacity owned by the firm. The implicit assumption is in the short run the marginal cost of producing an output level greater than k is

---

[7] There is an asymmetry involved here. The recognition that a rival firm is likely to be capacity constrained may well lead the firm to adopt a different competitive stance, a possibility incorporated into the puppy dog strategy of Fudenberg and Tirole (1984).

[8] Bresnahan and Suslow (1989) make use of this kind of framework in examining the behaviour of a single industry.

infinite, or at least prohibitive. Maximising this profit function yields two potential situations. If the constraint does not bind we have the conventional relationship,

$\frac{p-mc_i}{p}=\frac{s_i(1+\lambda_i)}{\eta}$ and q<k . The alternative situation is where

$\frac{p-mc_i}{p}=\frac{s_i(1+\lambda_i)}{\eta}+\psi$ if q=k. Where ψ is the Legrange multiplier associated with the capacity constraint. In the case of a vertical marginal cost curve at q=k and the perceived marginal revenue curve intersecting with this vertical portion of the marginal cost curve, the ratio between price and marginal cost will be equal to zero as quantity is fixed and price is determined by the intersection of the effective demand curve with the capacity level of output. In this case, the ratio of price to average variable cost will rise, in theory without limit, according to the increase in the level of demand and will not normally, in this case, be influenced by the conventional structural determinants of the price cost margin.

II.iii. Entry, disequilibrium and cyclical performance

The next aspect we have to consider is the long run effects which must be considered in terms of the response to entry and other adaptations of industry structure/ conduct which may occur over the business cycle. Most fundamental is the question of the extent to which entry responds to the business cycle and/or long run profitability, effects which may temper the profitability effects of the business cycle. In addition we have to consider the effects of entry on the composition of the industry. If boom conditions attract marginal firms into the industry in greater numbers the effect will be to reduce average profitability of the small firms, on the apparently reasonable assumption that marginal entrants are small[9]. We can relate this to theories going back to Steindl (1952, esp. Chapter 5) that regard a theory of cost differentials between firms as central to explanations of industry performance. Steindl directly posits the concept of a marginal firm in an analysis that distinguishes between industries where entry is difficult and industries where small producers are plentiful. This point also relates back to the point made by Kalecki (1939) that cartels are more prevalent in slump conditions because the threat of entrants moving in to benefit from the higher

[9] As support for this general position we can point to Geroski (1995, p424) who characterises the entry process by new ("de novo") firms as "more common but less successful than entry by diversification". New small scale entry by firms accounts for a large proportion of both entries and subsequent exits.

profits is diminished. The implication of both of the above arguments is that the supply curve of fringe firms will shift between periods of recession and boom. The slope of the schedule will be steeper in times of recession as the flow of entrants is stemmed and flatter in booms as the number of potential entrants increases.

An additional factor, that could lead to a potential flattening of the supply schedule of fringe firms, is the availability of finance. Either through improving the credit position of existing firms seeking to expand or the provision of credit to those seeking to enter, a relaxation of monetary policy on the part of the government or a less restrictive bank lending policy will tend to flatten the fringe supply curve. Gertler and Gilchrist (1991) develop a model which encompasses the asymmetry between firms access to finance in the context of demand fluctuations[10]. In a model where sales are randomly determined they establish that as the potential costs of bankruptcy weigh more heavily for smaller firms an increase in demand more than proportionately reduces the risk of bankruptcy for these smaller firms. Also an expansion of the money supply by reducing the adverse affects of the credit market on smaller firms enables a higher proportionate level of output. Therefore changes in both demand and the money supply, which in the model operate independently, yield a more elastic sales response from smaller than larger firms. They note that the model can be readily extended to incorporate differences in the level of internal funding. This approach is very similar in approach to the Mills and Schumann (1985) arguments which again focus on the relative elasticities of supply but in this case the difference in the elasticities is due to differences in the choice of technology.

It would appear plausible that this finance constraint argument could also be applied to entry on the part of large firms in the form of diversification. The availability of funds for large scale entry or take-over is likely to be more limited in periods of recession and firms may wish to

[10] Gertler and Gilchrist (1991) incorporate imputed bankruptcy costs into the marginal cost schedules of the firm and thereby alter the slope of the marginal cost curve for a small firm. As the supply schedule here is a horizontal aggregation of many firms we do not need to adopt this approach but can merely identify possibility of finance operating as a constraint on output as a way of obtaining a steeper supply schedule in a recession period.

'consolidate' or 'focus on core activities' at these times. The effect is therefore to reduce the external competitive pressure in periods of recession and increase such competitive pressure in time of boom. We can relate these arguments back in turn to the competitive fringe model considered in the previous chapter.

In examining the dominant firm model it is possible under the assumption that demand and marginal cost curves are linear and of the regular slope[11] to look for changes in price cost margins in response to demand changes. Hence, while the Lerner index will always be zero for the fringe firms, the price cost margin will be positive[12]. A upwards shift in the intercept of the demand curve will lead to an increase in price, hence, will lead to an increase in supply by the fringe. As the price cost margin for the fringe is a measure of the divergence between marginal and average variable cost, and this, under our assumptions, will increase with output, so the price cost margin will increase with a demand increase of this form. The same prediction applies to the price cost margins of the dominant firms as well as long as their marginal cost curve is not rising too steeply. These results are to be expected given the general preponderance in favour of procyclical price cost margins when abstracting from the changes in the degree of monopoly power.

Placed in a dynamic context it can be seen that there are potential advantages for the dominant group from having a fringe of small firms. It can be readily shown within the context of the model that the optimal change in output in response to a horizontal shift in the demand curve is lower both in absolute and proportionate terms for a dominant firm facing a fringe of small firms than for a monopolist with a similar industry demand curve. The same is true of the optimal dominant firm price which will fluctuate less than the monopoly price. As one of the acknowledged ways in

[11] That is the industry demand curve is linear and downward sloping, the supply curve of the fringe is linear and upward sloping and the marginal cost curve of the dominant group is linear and not downward sloping. All the comparative static results stated here are based upon this relatively general framework. The specifics of the algebra needed to develop the results is exceptionally tedious yet straightforward and therefore not reported.

[12] If the fringe cost/supply function is akin to the Mills and Schumann function the price cost margin of the fringe will be as described in equation (7) of Chapter 3.

which monopoly power is used is to reduce risk it follows that dominant firms may actually encourage a fringe of small firms which act to smooth out demand fluctuations This is likely to be more especially desired when there are substantial costs of holding excess capacity or operating at levels of overcapacity. Again the desirability of the fringe is subject to the proviso that the loss in profits is not sufficiently great to outweigh the benefits of demand smoothing.

II.iv. Inventories, price adjustment and the business cycle

The role of inventories across the business cycle is a subject of importance to macroeconomists but has been relatively neglected by industrial economists especially in applied work[13]. In the context of this chapter the importance of inventories lies in the trade-off of the relative costs of price and output adjustment in the context of variable or cyclical demand. The costs of price adjustment lie in two main areas, firstly and arguably of lesser importance is the "menu cost" argument that there are costs of changing prices because of the costs of informing customers of the new prices, secondly there is the cost in terms of market perturbance of the change in price. This market perturbance takes different forms; there is a potential loss of consumer loyalty if consumers dislike uncertainty concerning prices; in an oligopolistic context, there is also the cost of disruption to a tacit, or less commonly overt, collusive agreement, which may break down or weaken. On the other hand there are similar costs associated with output adjustments. There are similar direct costs of adjusting the level of production in terms of extra hiring, overtime, sackings or lay offs, and in terms of the capital/capacity costs of variable production. Indirect costs include stockouts that can lead to a loss of consumer loyalty and the presence of excess capacity can put pressure upon tacit collusive agreements.

The presence of such obstacles to both output and price flexibility implies a general role for inventories[14] or order book adjustment in terms of the smoothing out of output, sales or price

[13] An exception to this is the stream of papers mainly from US economists attempting to assess the validity or otherwise of the production smoothing hypothesis. Reference to some of this literature has been made in Chapter 2.

[14] The most common explanation for holding inventory is the production smoothing argument. This focuses on the presence of increasing marginal costs implying that a temporary positive shock to sales

fluctuations. Against these motives have to be weighed up the costs of holding inventory. On a direct level; for production to stock industries, these will constitute the storage and opportunity costs of leaving produced goods unsold; for production to order industries the direct costs are the more conventional costs of capacity adjustment. However it is equally clear that either the presence of inventory or order book adjustment still possess the potential for market perturbance effects. As Rotemberg and Saloner (1989) point out inventory can be used as a strategic weapon that can help or hinder maintenance of tacit collusion, excessive or insufficient inventory levels can have a similar impact as excess capacity or capacity constraints on the pricing of oligopolists.

As the firm is involved in a three way decision on prices, output and inventories, each of which have differing costs and impacts, it becomes difficult to obtain hard predictions. Rotemberg and Saloner argue that oligopolists will tend to invest in inventory in booms as a way of deterring the temptation to cheat which is greatest in periods of relatively high demand. They therefore suggest a procyclical time path for the inventory sales ratio in oligopolistic industries. It should be noted that if stockout costs are high, in terms of the loss of consumer loyalty, the level of inventories should be correlated with the level of sales because as sales rise the likelihood of stockout for a given level of inventory also rises.

The most common prediction is that in the context of inventory or order book adjustment price fluctuations will tend to be lower than otherwise[15]. This is an important part of the basis for the theory of normal cost pricing[16]. The idea of normal cost pricing is that "output price is set by taking a constant percentage over average normal historical current cost" (Godley and Nordhaus, 1972, p854). A proviso to this is that this percentage may not remain constant over time. In fact

---

will increase the cost of production in the absence of inventory. Clearly this omits many of the alternative motives for price and output stability discussed above.

[15] See Amihud and Mendelson (1983) for a theoretical examination of the price smoothing role of inventories and Amihud and Mendelson (1989) for empirical evidence concerning inventories and market power.

[16] Fuller discussion of the issue of pricing behaviour can be found in Hay and Morris (1991, chapter 7), Sawyer (1983), Domberger (1983) and Coutts, Godley and Nordhaus (1978). An example of a recent empirical study examining the question is Geroski (1992).

over the period they examine, the 1960's, Coutts, Godley and Nordhaus (1978) observe a sustained decline in the mark-up whereas Geroski (1992) finds a sustained rise in the mark-up over the 1980's. However the strong implication of the theory is that prices are invariant to short term movements in demand or cost, the underlying rationale for which is that the various costs of short term price adjustments exceed the benefits that arise from stability.

In applying these arguments to the firm size question it is not possible to investigate pricing at the level we require because observations on price are only available at an industry level hence no distinction between small and large firms can be made at an empirical level. However we can attempt to derive some suggestions for the relative behaviour of large and small firms. If small firms are more flexible than large firms in terms of having lower output adjustment costs, although this assumption is an open question as we have discussed in chapter 2, there will be a lesser need for use of either inventory or order book adjustment on the part of those firms. Similarly small firms are less likely to be engaged in collusive agreements and therefore the costs of perturbing such agreements are likely to be low. Unless advertising makes consumers less concerned by the such perturbations, consumer loyalty effects would appear unlikely to differ substantially between large and small firms[17].

Cowling (1983) argues that the growth of multi-plant firms will also reduce the impact of demand fluctuations on price in an analogous way to inventory/excess capacity as the effect on costs of adjusting production for these firms will be less than for single plant firms. Multi plant firms are able to change output by changes in the number of plants whereas the implied marginal cost curve for a single plant firm will be steeper. One would therefore expect margins to be less affected by the cycle for those large firms with more plants. For this reason one may expect the effect of the cycle to be less for larger multi-plant firms than for their smaller counterparts.

[17] In theory some firms may choose to compete on reliability of delivery relative to some other quality standard but there seems to be little justification for a systematic size related difference of this kind.

The main exception to this conclusion of inventories reducing price variability is the tendency to hold "sales" when inventory levels become excessive. Thus if demand falls unexpectedly firms may lower prices for a limited period to restore normal levels of inventory. If such "sales" are preannounced for a fixed period they need not create a market perturbance if both consumers and producers anticipate a relatively rapid resumption of the previous market situation. Clearly however such "sales" can lead to market disruption if there are asymmetries in the distribution of the demand shock or adjustment costs between firms. In this context we may expect an impact reduction in prices as excessive inventories accrue but that this effect is not necessarily long lasting.

II.v. The persistence of profits

Linking in with the notions of long run effects interacting with dynamic effects is the persistence of profits literature. This literature was developed in a number of studies[18], however we focus on Geroski (1990a) as a representative exposition of the theoretical framework developed in this literature. Geroski's first point is that when considering the antitrust implications of structure performance relations we should be interested in the long run effects i.e. persistent profits not short term fluctuations around the long run. The point of note is that while the long run is the main point of interest from the conventional structure performance perspective, and was the main focus of the previous chapter, the additional focus of this thesis is on the short term responses to business fluctuations. As we have tried to show, in subsection II.iii. above, the two interact closely, for example the ability to obtain profits above the norm in a boom (or slump) period is still likely to depend upon the susceptibility to entry or intra industry mobility. If anything when considering the short run responses it is the speed of response that is most important.

The second point is empirical, when we observe profits we usually observe current period profits when the theoretical point of interest (subject to the proviso given above) is long run equilibrium

[18] Mueller (1986) was the principle monograph study but a number of studies including Levy (1987) and Geroski and Masson (1987) and the other studies reproduced in Mueller (ed.) (1990) developed the perspective.

profits. If there are systematic measurement errors arising from this difference, and in particular if these systematic measurement errors are correlated with the other explanatory variables we will observe biased estimates of both the point estimates and covariance matrix.

A time series model of profits is then developed by Geroski that incorporates these elements and indicates an appropriate specification. Let[19] π_t be firm i's profits in time t and πp_t be the competitive rate of return in the long run, $\rho_t=\pi_t-\pi p_t$ is the excess profits in time t. Deviations from the competitive rate of return are governed by luck and entry factors where the former are orthogonal to the latter and entry factors can be loosely interpreted as all factors affecting the market behaviour of market participants so these are factors that are not random and are endogenous to the determination of profits. These can be formalised in equation (9)

$$\Delta\rho_t=\theta_0+\gamma_0 E_t+\gamma_1\rho_{t-1}+\mu_t \qquad (9)$$

where $\Delta\rho_t=\rho_t-\rho_{t-1}$, E_t is the rate of entry in period t, and μ_t are the luck factors in period t and is i.i.d. with mean 0 and finite and constant variance. Both γ_0 and γ_1 are presumed to be constant and negative. The model is equivalent to a distributed lag framework where the coefficient $\gamma_1=(\gamma^*-1)$ where γ^* is the coefficient on the lagged term in a levels equation of excess profits. If E_t is not observed, a dependence of entry on the previous levels of excess profits, relative to the norm of excess profits, can be specified such as of the form

$$E_t= \phi(\rho_{t-1}-\rho^*) + \varepsilon_t \qquad (10)$$

where ρ^* is the norm of excess profits. Thus the coefficient on the lagged term in a levels equation of excess profits, with entry omitted, would become, through substitution, $\gamma^* +\phi\gamma_0$. The second part of this is simply the bias induced in the coefficient from the omission of entry as a right hand side variable.

Econometrically the main studies estimated these lagged profits models separately for many firms and then looked for systematic differences between the estimated coefficients by regressing them on many structural factors that may influence either the degree of persistence or the level of

[19] The presentation closely follows Geroski (1990a) but the notation has been changed slightly.

excess profits. The question may be posed as to whether this is the most efficient method of identifying this information. By estimating the initial coefficients while intentionally omitting relevant variables that are correlated with the lagged dependent variable both the estimated coefficients and standard errors will be biased. The bias in the coefficients may not be a problem[20] as the magnitude of the bias is the subject of the second stage regressions. However it is not clear that the estimates will be efficient relative to a single stage procedure that takes into account such heterogeneity in the coefficients. Such a single stage procedure would involve pooling the data and permitting the estimated coefficient on the lagged dependent variable to systematically vary across units with the movement of the relevant variables[21].

In terms of assessing the usefulness of the persistence of profits literature, we can see that it has made the significant step of shifting the focus of empirical analyses in industrial economics to the presence of long run supranormal profits and the process of dissipation of temporary supranormal profits. From an econometric point of view the literature has led to a general proliferation of the usage of dynamic specifications in industrial economic analyses and a clear framework for the interpretation of such specifications.

II.vi. An overview of the different theories

A lot of material has been covered in the previous discussions and it would appear necessary to draw some overall implications from the survey of theories we have undertaken. On a simplistic level it is possible to say that models of short run profit maximisation, with a number of qualifications and exceptions that have been discussed, predict procyclical price cost margins for all firms. There is also an implication that large firms may experience more procyclical price cost margins than small firms. Incorporating elements of intertemporal profit maximisation into the

[20] An example of an occasion when such first stage coefficient estimate biases may prove problematic is when observations are deleted from the sample for the second stage estimation on the basis of the possibly incorrect inference that the coefficient falls outside an acceptable range, e.g. Cubbin and Geroski (1990) eliminate estimates that are greater than 0.95 on the basis of the instability of estimates so close to unity.

[21] Implicitly a limited form of this approach is adopted in the interactions reported by Machin and Van Reenen (1993, Tables I and III).

framework can reverse or mitigate this finding; the dynamics of collusion may lead to countercyclical margins which would be expected to have a greater impact upon large firms; mechanisms exist for smoothing the effects of the cycle such as inventories, order backlogs and excess capacity, the relative effects and usage of these mechanisms by large and small firms would appear to be an open question. We have also suggested that through the process of entry the business cycle may have longer lasting effects upon the performance of the industry and the firms in industries. This requires an explicitly dynamic framework for dealing with these issues.

III. Previous empirical studies of cyclical performance

As stated earlier the only previous study of intra industry performance over the business cycle that we are aware of has been Schmalensee (1989b). Before we discuss this paper we will examine the other recent studies of cyclical performance at the industry or firm level as these studies logically (but not necessarily in time) precede Schmalensee's developments.

The first major panel data studies to examine cyclical performance at the industry level were Domowitz, Hubbard and Petersen (DHP)(1986, 1987, 1988). These studies all made use of a panel of US manufacturing industries from 1958-1981. The studies examined the instability of the concentration margins relationship over time and in particular noted the weakening of the relationship over time such that the strength of the relationship was much weaker in the late 1970's and early 1980's than in the 1960's. This finding was subsequently challenged by Salinger (1990) who argued that with a wider industry sample and adjustments to the measure of concentration to take into account both import competition and regional markets within the U.S. the relationship remained robust through the 1970's and to the mid 1980's. It is also interesting to note that Wachtel and Adelsheim (1977) provide clear evidence that, for the period 1946-1960, highly concentrated industries exhibit a countercyclical pattern of price cost margins while medium and low concentrated industries exhibited a procyclical pattern. The limited evidence from this study for the period from 1961 onwards is of procyclical price cost margins for all industries. This suggests that there may have been a structural break in the relationship at a much earlier stage than the 1970's, and that this break that is omitted from the DHP studies because it occurred at the start of the time period of their study.

The exception to the finding of universal procyclicality, in the post 1960 period, is Bils (1987). Bils' study is, however, based upon estimates of marginal cost[22] that rise steeply as output increases. The implication of the estimates provided is that firms reach forms of capacity constraints at high levels of output. We have noted, in section II.ii. above, that, as capacity constraints are reached, price and marginal cost converge, in the limit, and price increases in relation to average variable cost. Bils' estimates can be made perfectly consistent with the estimates from DHP by noting that these latter studies make use of a measure of performance that proxies the relationship between price and average variable cost, not the relationship between price and marginal cost.

One of the main findings of the 1986 study was that price cost margins are more procyclical in concentrated industries. However, as Schmalensee (1989b, p353) notes, there may be two secular trends at work which are creating this correlation. DHP use both the growth rate of industry sales and the aggregate rate of unemployment as measures of the business cycle. However unemployment is strongly trending over the period as, in the data set that DHP use, is the relationship between concentration and margins. A negative and significant coefficient on the interaction between unemployment and concentration therefore may be the spurious correlation of two trends. The two may or may not be related but if they are both trending we cannot tell if the correlation is economically meaningful or not.

Studies for Japan, Odagiri and Yamashita (1987), covering 1958-1982, and Germany, Neumann, Böbel and Haid (1983), covering 1965-1977, also confirm the finding of a greater strength on the relationship between concentration and price cost margins in times of expansion relative to

[22] An important reason for this observed rapid rise in marginal cost is the importance of increased overtime payments in the estimates of marginal cost for levels of output above normal.

recession and find no evidence that highly concentrated industries are less procyclical than industries with lower levels of concentration.

UK studies of cyclical performance at the industry level include Haskel and Martin (1992, 1994) examining UK data for 1983-1986. The 1992 paper concentrates on the role of trade unions and unemployment in the determination of industry level price cost margins. It is found that sectoral level unemployment has a positive direct effect on price cost margins but that the interaction effect with concentration is negative. The coefficients and variable means imply that for industries with below average levels of concentration unemployment has a positive (i.e. counter cyclical) effect while for industries with higher than average levels of concentration the effect is negative (i.e. pro-cyclical). Haskel and Martin qualify these results because of the short time period and differences in the data between their results and the DHP results An important point to note is that Haskel and Martin use the net output denominator price cost margin whereas DHP use the more conventional margin. It is useful to note from their results that the finding is absent in the levels estimates, the coefficients being poorly determined and, in the case of unemployment of the opposite sign, and is only present in the first difference estimates.

Haskel and Martin (1994) empirically examines the effect of capacity constraints on the price cost margins of industries. The theoretical framework adopted has been briefly discussed in section II.ii. above. Using data for 1983-1989 and again adopting the net output denominator margin they estimate number of interactive specifications, the effect of capacity constraints in a linear model proving insignificant. In the most readily interpretable format of interaction between concentration and capacity constraints they once again find a negative (yet insignificant) effect for capacity constraints but that a positive interaction effect. The estimated positive interaction effect indicates that capacity constraints have a positive effect on margins for mean concentration levels and that this increases with concentration[23].

[23] This is based on our calculations from a rescaling of their reported results. Taking the coefficients as reported and basing calculations on the reported means of the variables indicates a negative effect for all concentration levels, a finding that is inconsistent with the rest of their statistical results. It was therefore assumed that an unreported rescaling of both coefficients and means had been made. No calculations of

Machin and Van Reenen (1993) examine firm level cyclical effects using accounting data for the UK. Using an accounting approximation to the price cost margin they find procyclical margins in the sense that they are negatively related to the aggregate unemployment rate. However they do not investigate industry or firm level cyclical effects nor do they investigate differences in cyclical effects between firms of different sizes.

While the focus of Schmalensee's paper is explicitly wedded to the debates concerning market power/efficiency explanations of inter and intra industry performance, a debate we have deliberately side stepped within this thesis as it is not seen as presenting a useful way forward for increasing understanding, the results of the analysis can be interpreted more generally and applied outside of that framework. Schmalensee's principle objective is to attempt to identify the reason why industry level price cost margins are more procyclical in concentrated industries. In particular the objective is to find if the prominent influence is inter industry or intra industry differences in cyclical performance. To this end Schmalensee estimates for each year of his data set, 1953-1983, elasticities of profitability with respect of firm size[24], seeking to abstract the pure size effects from industry effects. These estimated elasticities are then examined for cyclical variation using various aggregate measures of the cycle including capacity utilisation. There were two clear results from this analysis. Firstly the inter industry component of the elasticity was positive and made a significant contribution to the total effect using conventional measures approximating the price cost margin[25]. However the inter industry component exhibited no significant variation over time. All of the time series variation in the overall elasticity was generated by time series variation in the pure size, or intra industry, effect. Secondly this time varying intra industry component, hence the overall elasticity, was both declining secularly over

this form are reported by Haskel and Martin and similarly the significance of such interactions is not assessed.

[24] Detailed discussion of the estimation technique used is reserved for section V.i. below.

[25] Schmalensee tests 12 different measures of profitability, 6 using assets as the size measure in the denominator and 6 using sales as the denominator. In the former case the inter industry effect was negligible in size and/or insignificant. The latter are the more conventional approximations to the price cost margin and are discussed above.

time and behaved in a countercyclical manner. This indicates that, for Schmalensee's sample, there is little role for industry effects in determining the procyclical profitability in concentrated industries. Instead we must conclude that large firms are less procyclical than small firms. Schmalensee squares this with the DHP results by arguing that the missing element must be the industry level effect that differs between concentrated and unconcentrated industries. That is, industry factors that affect the profitability of *both* large and small firms, keeping the inter industry relationship between the two constant, must be at work.

IV. Framework for analysis

The starting point for the empirical analysis of the cyclical and dynamic aspects of intra industry performance is that most of the previous analyses are partial in focus. Many studies focus upon one element of the business cycle (e.g. inventories or capacity constraints) in isolation. We have seen from section II of this chapter that there are many different factors that potentially come in to play in analysing cyclical performance. Ideally we wish to simultaneously incorporate as many of these factors as possible. The reasons for the partial nature of some of the previous analyses is the theoretical, and usually, by implication empirical, complexity associated with the multiple decisions implicit in a fully specified intertemporal profit maximisation framework. Whilst parsimony may be a suitable resolution of this problem from a theoretical perspective, enabling clarity of analysis, it is less attractive for empirical work where the potential costs of omitting relevant variables are usually greater than the costs of including irrelevant variables. In general therefore we have sought to make the specification of the empirical analysis as general as possible. This contains costs in itself on two grounds; firstly it can tend to widen the distance between empirical specification and any relevant theoretical specification; secondly the interrelationships between variables may be difficult to empirically separate due to possible collinearities. For these reasons the analysis may appear "data-driven". At all stages however we have attempted to be mindful of the predictions arising from the discussion in section II.

The empirical analysis takes four forms. Firstly in section V.i. we present and discuss the summary statistics of the data we have collected. The second and third aspects of the analysis present results from our data set using frameworks developed in previous relevant analyses. In V.ii. we present estimates of the Schmalensee (1989b) elasticity of profitability with respect to firms size. We are using for all of the analysis in this chapter a data set that is similar in construction to the one used by Schmalensee. By presenting our own estimates we can provide a comparison and in addition add to the variety of descriptive statistics. In section V.iii. we present analysis within the persistence of profits framework of the dynamic adjustment processes of large and small firms. The focus of the persistence of profits literature has been on estimating the speed of adjustment to the normal profit level whilst also trying to identify the factors that influence these pressures. The main focus in this regard has been on the role of entry (or exit) as a dissipation device for supranormal (or sub-normal) profits. However a series of other variables have been used in trying to examine the speed of adjustment[26]. In our analysis we focus perhaps to a greater extent on the disequilibrium adjustment aspect of the analysis examining the effect of the length of the production run and multiplant firms on the adjustment process.

The fourth element to the empirical analysis is the estimation of the cyclical determinants of large and small firm performance. It has been argued in section II that there is a variety of data available for which there are sound theoretical reasons to expect to influence performance levels. Most analyses include only one measure of the cycle, yet not all measures of the cycle measure precisely the same phenomena. We feel that this is an interesting empirical area to develop and therefore we attempt to distinguish between the different potential determinants. Foremost amongst these are the twin aspects of the use of stocks and capacity utilisation but in addition we examine the effect of the growth rate of industry sales. Because we are using a panel approach we can also control for aggregate cyclical factors through the use of yearly dummy variables.

[26] By way of example Cubbin and Geroski's (1990) study of the determinants of the speed of adjustment for 239 UK companies includes advertising intensity, concentration, market share , growth and capital intensity among other explanatory variables

V. Results of the analyses

V.i. Summary analysis of data

The data set used is the same panel of UK manufacturing size classes used in the previous chapter. In addition this data has been merged with data taken from the CBI Industrial Trends Survey on the proportion of firms operating below capacity and the net proportion of firms with excessive levels of inventory. A description of this CBI data and the methods used to compile it are contained within the data appendix to this chapter.

Table 1 reports the coefficients of variation for the relevant cyclical variables, separating out the high, medium and low concentration industries. Both the inventory sales ratio and the price cost margin are, in general, more variable for the large firms than for the small firms. It has been suggested that, in theory, the inventory sales ratio may vary more as large, or oligopolistic firms have a wider set of motives for varying inventories and our results would appear to confirm this. However as most of these motives relate to a desire to smooth prices or costs it is perhaps surprising that the price cost margins of the large firms are more variable. It may be the case that the entry and exit of marginal firms serves to sufficiently dampen the variability of small firm price cost margins. However it is also the case that for both large and small firms price cost margins are more variable in highly concentrated industries. This combined with the greater variability of large firm price cost margins points to the existence of market power as the factor increasing the variability of price cost margins. It may therefore be the case that large firms have higher variability of price cost margins in spite of their greater reliance upon inventory.

It should also be noted that price cost margins are lower for both small and large firms in highly concentrated industries as a result of the lower level of vertical integration in those industries. Using the alternative net output based price cost margin, that partially corrects for such differences in vertical integration, the statistics for which are not reported, the more concentrated industries have higher price cost margins but the same pattern of higher coefficients of variation in the industries with higher levels of concentration is repeated. There is no substantial evidence of a greater variability of the stocks to output ratio in more highly concentrated industries. This finding runs counter to the Rotemberg and Saloner and other market power arguments that the inventory sales ratio should be more variable for both large firms and for firms in more highly concentrated industries. The only slight confirmation for this is that the stocks to output ratio is higher for larger firms and the variability is higher for larger firms in medium and low concentration industries. The latter argument relies upon an intra industry view of market power rather than an industry wide view.

V.ii. Estimates of the profitability (dis)advantage of large firms

The next step of the application to the UK data is to estimate versions of Schmalensee's profitability advantage of large firms for the period 1980-1992. Schmalensee's methodology for calculating the elasticities of profitability is described in Schmalensee (1989b; p342-345) and this method is followed almost precisely here, with the exception that the size classes used are just the largest and smallest firms[27] and the size class bands are in terms of enterprise employment rather than assets. The method involves regressing gross profits, π_i, which can be measured in three different ways, on some measure of size and the number of firms per size class.

The three measures of gross profits used are, (i) the standard measure taken from the census, in this case net output minus operative wage costs. (ii) a 'synthetic' measure of gross profits which is the level of gross profits that each size class would have if the size class price cost margin was the same as the industry price cost margin. This measure of gross profits removes within industry variation and therefore focuses upon inter industry differences in profitability, (iii) an 'adjusted' measure of gross profits which removes the inter industry component focusing only on the intra industry component. More precisely,

(i) π_{ic} is the gross profits of size class c in industry i. The 'synthetic' measure is then calculated by,

[27] There are relatively fewer size class bands in the UK census data due to the greater level of disaggregation used, many industries only have two or three size bands. For this reason it was decided to focus on the two size bands for which there is consistent data for all industries i.e. 1-99 employees and the largest five enterprises in the industry.

(ii) $\pi^{s}_{ic} = \frac{\pi_i}{s_i} s_{ic}$ where π_i and s_i are gross profits and the measure of size at the industry level and s_{ic} is the size measure for class c in industry i. The 'adjusted' measure is then calculated by,

(iii) $\pi^{a}_{ic} = \pi_{ic} - (\frac{\pi_i}{s_i} - \frac{\pi_m}{s_m}).s_{ic}$ where π_m and s_m are gross profits and a measure of size at the level of manufacturing as a whole. The two measures of size used were net output and gross output as these are the most commonly used in forming price cost margins. The regression then takes the form,

$\pi_{ic} = \alpha n_{ic} + \beta s_{ic} + \varepsilon_{ic}$ where n_{ic} is the number of firms in size class c of industry i.

A regression of this form[28] is estimated for each year, for each of the measures of gross profits and for the alternative size measures. Each regression can be used to form an estimated elasticity of profitability with respect to firm size which is derived (Schmalensee p342) as

$$\phi = -\frac{\alpha}{\alpha + \beta \bar{s}}$$

where $\bar{s}$ is the mean size of firm and relevant standard errors can also be calculated for these estimates[29]. This leads to six different estimates for ϕ.

	Original	Synthetic	Adjusted
Net Output	ϕ_1	ϕ_3	ϕ_5
Gross Output	ϕ_2	ϕ_4	ϕ_6

The result of estimating three different elasticities for each size measure is that it becomes possible to decompose the elasticity of profitability with respect to size into those effects which can be attributed to inter industry characteristics and those which result from the profitability (dis)advantage of large to small firms across the economy as a whole. The original elasticity

28 In addition there is a constant term included. All estimation is performed by OLS. Inference is based upon the heteroskedasticity robust covariance matrix.

29 The standard errors are calculated using the usual Taylor Series approximation for testing non-linear restrictions (see Greene (1993), p218-220) where

$$\text{var}(\phi) = \frac{\beta \bar{s}}{(\alpha + \beta \bar{s})^2} \text{var}(\alpha) + \frac{\alpha \bar{s}}{(\alpha + \beta \bar{s})^2} \text{var}(\beta) + \frac{\alpha \beta \bar{s}^2}{(\alpha + \beta \bar{s})^4} \text{cov}(\alpha\beta)$$

should be a summation of the synthetic and adjusted estimates. Schmalensee regressed the estimated elasticities on various measures of the business cycle and discovered that they were countercyclical. As we have only thirteen annual estimates such small sample regression seems inappropriate in this case. Instead the estimates are presented in Tables 2 and 3 and are displayed graphically in Figures 1-4.

Figure 1 presents the measures using net output as the size measure, Figure 2 presents the gross output estimates. From Figures 1 and 2 it is clear that for both measures of size the original elasticity of profitability with respect to firm size (ϕ_1 and ϕ_2) are negative at the start of the period. Using gross output as the measure of size gives an original measure of the profitability advantage which is negative throughout the period. When we decompose this into an intra industry effect and an inter industry effect we can see that the original effect is dominated by the inter industry effect which is strongly negative throughout and becomes more strongly so towards the end of the period, i.e. in the 1990's recession. This indicates that there is a strong industry effect that profitability is higher where small firms are more important. This is only weakly countered by the positive effect of size *per se*. This contrasts with the net output based estimates in which both the pure size effect and the industry effects are both negative at the start of the period, although they are insignificantly different from zero, indicating a weak overall profitability advantage for small firms. This is reversed through the period but, throughout, the industry effect is insignificant and it is the pure size effect which has the dominant effect on the overall estimate. By point of note, Schmalensee's estimates of the inter industry effect were also of negligible size relative to the pure size effect. The presence of a significant inter industry effect with the gross output estimates can probably be explained in terms of the important inter industry differences in material input usage we have discussed in chapter 3. We can also note that Schmalensee restricted himself to 2 digit industry codes hence may have lost some of the inter industry heterogeneity present in the 3 digit industries used here.

Figure 3 compares the intra industry effects for the two measures and Figure 4 makes the same comparison for the inter industry effects. The time path of the intra industry estimates is remarkably similar both exhibiting a fairly consistent upward trend which appears to slow or be reversed in the two recession periods at the beginning and end of the period. The effect of the boom in the latter part of the 1980's would appear to have been the accumulation of a pure size profitability effect. However, as the time period is relatively short we must exercise caution before adopting a complete cyclical interpretation of the results. Manufacturing output in the UK was stagnant in 1989, fell in 1990 and 1991 and did not resume an upward trend until the start of 1993, after the end of the period of analysis. The peak years for UK manufacturing, based on the CBI evidence in chapter 1, were 1987 and 1988. ϕ_6 , the net output measure of the pure size effect, peaks in 1989 and remains higher through the subsequent recession than it was in 1987 or before. The measure increase once again in 1992, before manufacturing output as a whole recovered. The gross output measure continues to rise until 1990 and exhibits no significant subsequent decline. It, therefore, appears likely that there is secular effect, brought on by industrial restructuring in the early and mid 1980's, which are dominating the cyclical effect. This argument would appear particularly compelling in relation to the gross output estimates which appear devoid of cyclical movement.

The inter industry effects are more markedly different. There is no discernible time trend from the estimates but in both cases there is a fall (rise) in the profitability (dis)advantage of large firms in the 1990's. This would indicate that industries where large firms are important have performed badly in the recent recession, but overall the industry effect does not vary very much. It is interesting to note from Figures 3 and 4 that the difference between the two measures of profitability, which has been explored in more detail in chapter 3, is almost exclusively an industry effect.

V.iii. Persistence of profits estimates

The next element in the analysis of dynamic performance is to investigate the degree of persistence, the determinants of persistence and how these may differ between large and small firms. If the small firms do form a competitive fringe we should expect a low level of persistence as any profits will be short term as they will be competed away by entry into the fringe.

The standard autoregression in profits of the form,

$$PCM_t = \alpha + \lambda PCM_{t-1} + \varepsilon_t \quad (11)$$

was estimated for both the large and small firm size classes for each industry[30]. Estimation was by OLS. It may be argued that, with 13 annual observations per size class, such small sample regressions are inappropriate. We would note that, in most of the previously published studies, the time periods have been less than 20 years in duration, therefore our estimates offer a reasonable comparison. These generated estimates were then regressed on the various structural variables. As has been suggested above the statistical properties of such estimation procedures are open to question. Therefore, in addition, we estimate interactive specifications within a pooled model in order to directly estimate, in a single equation, the determinants of persistence. Initially we estimate a model of the form,

$$PCM_{it} = \alpha + \beta_0 X_{it} + \beta_1 (X_{it} * PCM_{it-1}) + \lambda^* PCM_{it-1} + \varepsilon_{it} \quad (12)$$

X_{it} is an independent variable such as concentration[31]. The determinants of persistence can be assessed using this approach though investigation of the β_1 coefficient and the predicted value of the persistence effect can be obtained from the equation, $\lambda = \lambda^* + \beta_1 X_{it}$. These estimates are obtained in both levels and first differences. Estimating in levels imposes a constant value of α across industries. First differencing permits α to vary across industries. However, by first differencing, following from the discussion in chapter 4, we also shift the focus onto the short run rather than long run effects of the independent variables. We have seen that concentration and advertising may differ in their short and long run effect and we wish to be sure that we capture

[30] It has been more conventional for estimates of persistence to be based upon firm level data. However, this has not been the case exclusively, see Kessides (1990) for an industry level example.

[31] The estimates use the full set of independent variables. We present the case of a single independent variable for simplicity.

accurately the effects of persistence. Therefore, we consider, in addition, a specification of the form,

$$PCM_{it} = \alpha_i + \beta_0 X_i + \beta_1 (X_i * PCM_{it-1}) + \lambda^* PCM_{it-1} + \varepsilon_{it} \quad (13)$$

where $X_i = \frac{1}{T}\sum_{t=1}^{T} X_{it}$. First differencing to eliminate the industry specific fixed effects yields an estimating equation of the form,

$$\Delta PCM_{it} = \beta_1 (X_i * \Delta PCM_{it-1}) + \lambda^* \Delta PCM_{it-1} + \Delta\varepsilon_{it} \quad (14)$$

where $\Delta PCM_{it} = PCM_{it} - PCM_{it-1}$. Given this specification, despite the first differencing and the use of a pooled sample, the estimated degree of persistence depends upon the mean level of the independent variables.

Considering the initial estimates from (11), contrary to most of the previous studies in the persistence literature, the estimates of λ were substantially lower on average for the small firm size class, .224, than for the large firm size class, .381, despite the fact that three values of λ greater than 0.95 were excluded from the sample of large firms because of the implied instability of the predicted path of profits. The mean predicted long run levels of profitability, $\alpha/(1-\lambda)$, were almost identical for the two groups, .324 for small firms, .327 for large firms. This contradiction between the two results may indicate that while individual small firms may maintain a high level of profitability in spite of potential or actual entry the small firms as a group are responsive to such profitability changes and entry and exit do occur as a result altering the performance of the small firms. Nevertheless it is clear that adjustment is still not instantaneous even for the small firm sector. The magnitude of the mean coefficient for the large firms is on the low side for individual firms estimates. This may be the result of inter or intra large firm group mobility which may lower the persistence of performance for the group below that of the firm.

The estimated coefficients from (11) were regressed, using OLS, on various independent variables to attempt to analyse the determinants of the degree of persistence. The independent variables used were the conventional structural parameters considered in this framework, advertising, concentration and small firm market share. In addition, following the arguments from section II we include a proxy measure of the length of production run to see if firms are slower to adjust if their production run is longer. We also include the number of plants per large enterprise as a measure of one aspect of the ability of large firms to smooth production. The results of these estimates are presented in Table 4. The results of the pooled estimates of (12) and (14) are reported in Table 5.

Table 4 reports F-Tests indicating that a significant relationship can be determined for the standard estimates. However, there are problems with the diagnostic tests in the pooled methods of estimating the determinants of persistence[32]. The levels estimates of (12) indicate significant first order serial correlation, which may be attributed to dynamic mispecification and/or the omission of the fixed effects. Estimating (12) and (14) in first differences eliminates serial correlation, but the Sargan statistic fails for the small firm sample in estimating (14) and fails for the large firm sample in estimating (12). It should also be noted that for the small firms from the test of the joint significance of the interaction terms in the first difference estimates of (12) and (14) we cannot reject the null of no relationship. We therefore have to exercise some caution when interpreting these estimates.

The most striking result concerning the estimated persistence parameters is the effect of the mean market share of the smaller firms. The market share of the smaller firms increases the speed of adjustment of the performance of the large firms group implying that the large firms experience a greater degree of competitive pressure if on average the small firms are larger relative to the size of the market. This effect is present in both the conventional and the pooled estimates. The exception is the first difference estimates of (12). We have noted above our concerns that these estimates may be picking up the short run effects and this result appears to confirm this concern. Small firms with higher market shares increase competitive pressures but increases in the mean market shares of the small firms do not. The presence of a small firms group with relatively high market shares would appear to lead to a greater threat of mobility into the oligopolistic group,

[32] It should be noted that we have ignored the problem of serial correlation, in common with the general persistence literature, in the estimates of (11).

using the Caves and Porter framework, or alternatively a greater degree of head on competition. Either effect could lead to a reduced ability of the large firms to hold on to any temporary degree of monopoly they can create. For the small firms the effect of their own market share is not well determined and the estimates are contradictory. The pooled estimates indicate more rapid adjustment whereas the conventional estimates indicate slower adjustment. The signs of the coefficients on concentration are consistent with such an interpretation implying a greater ability to hold on to monopoly power for large firms and a more marginalised rapid adjustment rate for the small firms group. However, these effects are not significant in any of the estimates.

Perhaps surprisingly advertising intensity does not appear to have a strong effect on the ability to sustain an increased price cost margin for either group of firms. The sign is consistently negative for the large firms indicating a reduced ability to hold onto excess profits, but this effect is only significant in one of the cases. Longer production runs appear to increase the speed of adjustment for the large firms and reduce the speed of adjustment weakly for small firms. Whilst the latter may seem intuitive the former does not. The measure of the length of production run is, by construction, highly correlated with the inventory - gross output ratio and it may be that the increased speed of adjustment is rather due to a greater level of inventory usage. The multiplant coefficient is of the 'right' sign implying smoothing of adjustment but is insignificant for the large firms in all estimates and for the small firms in the pooled estimates. The effects of material inputs is not generally well determined but the sign of coefficients imply that the large firm group can hold on to its higher price cost margins better when vertically integrated.

The determinants of the predicted long run price cost margin are analysed in Table 4, advertising intensity is the only significant determinant and this applies to both small and large firms. It is not possible given the specifications reported in Table 5 to observe the implied long run value of the price cost margin. It is clear from the predicted coefficients generated by (11) that small firms are quicker to adjust to normal levels of profitability. However, the ability to determine the factors that effect the speed of adjustment for small firms is lower. The estimates point to the market shares of the small firms and the length of the production run as important determinants of the speed of adjustment for large firms and to a lesser degree advertising and concentration.

V.iv. The determinants of cyclical performance

The next part of the analysis is an attempt to discover the cyclical determinants of the performance of small firms and large firms. The focus is upon attempting to discover whether some cyclical factors affect small firms/large firms to a greater or lesser degree. Consistent data is available for the period 1984-1992. Data between 1980 and 1984 is used for instrument purposes but not in the direct estimation. In addition to the cyclical components we include concentration, small firm market share and their lagged values, and the lagged dependent variable in the estimation process. These are the relevant included variables from the analysis in chapter 4. Further lags of these variables did not prove collectively significant hence we restrict the lag length to 1 in each of these cases. The cyclical variables considered are the proportion of firms reporting below capacity operation (PBC) and the net proportion reporting excessive levels of finished good inventory(EINV), both of these variables are taken from the CBI Industrial trends survey. In addition the growth rate of industry sales (GROW) and the ratio of stocks and work in progress to gross output (at the size class level) (INVSAL) are included from the UK Census of Production data.

Determining the appropriate lag length for the cyclical variables is made more difficult by the need to maximise the number of observations available for estimation. The lag length of three was chosen as a compromise between this interest of maximising the number of observations and the desire to have as complete a description of the dynamic process as possible. In most of the cases considered for the large firms, a joint test of the variables dated t-3 proves significant at the 5% level. For the small firms, however, such a test is only significant in the sample of production to stock industries. This is consistent with a more rapid response to cyclical factors on the part of small firms. Formally, the presence of significant tests of the variables dated t-3 might indicate a potential problem of underspecification and the initial adoption of an insufficiently general approach[33]. We should note, however, that all of the previous panel data studies on the issue, discussed in section III above, have only made use of contemporaneous independent variables and restricted any dynamic specification to the inclusion of a lagged dependent variable[34]. The results presented here therefore represent a significant advance on the kind of dynamic specification previously adopted.

A further point of note regarding the specification is that we use the conventional (gross output) definition of the price cost margin and correct for vertical integration by included the ratio of material inputs as a right hand side variable. This correction is important. As has been shown in section V.ii. above, at an aggregate level, the conventional price cost margins exhibits less cyclical variation than the net output price cost margin. Estimation using the conventional margin with no correction yielded little cyclical movement and indeed frequently led to insignificant aggregate time effects measured by the yearly dummy variables. Estimation using the net output margin produces a more pronounced cyclical effect akin to the estimates presented below. Our results therefore differ from those of Haskel and Martin (1992) who report that estimation without correction did not substantially effect their results.

As we have no problems, given the focus of this chapter, with producing short run estimates of the parameters only the first difference estimates that implicitly exclude fixed effects from the data are considered. Estimation in levels always proved susceptible to serial correlation and was therefore rejected. Estimation was by the GMM procedure outlined in Arrellano and Bond (1988). Following the concerns expressed in Schmalensee (1989a) regarding the exogeneity of any contemporaneous variables and, as all of the cyclical variables are potentially subject to

[33] The merits and demerits of parsimony in the selection of dynamic specification are considered in more detail in Hendry (1995) esp. Chap 15 and Keuzenkamp and McAleer (1995).
[34] An exception to this is Berg (1986) which examined the effect of excess capacity on collusion in Norwegian manufacturing industry making use of up to 6 lags of excess capacity, but not within a panel, instead estimating for different sectors. Berg's findings demonstrated important differences between the contemporaneous impact effect of excess capacity and the subsequent lagged effects demonstrating the benefits in this context of a more general specification.

strategic manipulation of the firms concerned, we regard all of the contemporaneous cyclical variables as endogenous and they are therefore instrumented.

Coefficient estimates and t ratios for the estimates of the cyclical determinants of price cost margins for both small and large firm groups for all manufacturing industries are given in Table 6. The results of a subdivision of the sample into consumer goods and producer goods industries is reported in Table 7 and a subdivision into production to stock and production to order industries is reported in Table 8. Diagnostic tests for all estimates are presented in Table 9.

The diagnostic tests indicate no problem of serial correlation in any of the results. In only one case does a test of the joint significance of PBC, EINV and GROW fail to reject the null of no relationship. This case, the large firm group in consumer goods industries, is apparently problematic because, in addition, it is one of two cases where the time dummies are not jointly significant and the Sargan test for overidentifying restrictions fails. This is, therefore, an indication of some kind of dynamic mispecification for this case. In the other cases, only for the large firms in the production to order subdivision are there indications of problems from the Sargan statistic. There is limited evidence of collinearity between the cyclical variables. In two cases the joint test is significant while individual tests on each of these variables are all insignificant. However the correlation matrix reveals that the highest correlation between the independent variables is 0.5, the correlation between the contemporaneous values of EINV and PBC.

Overall there is little evidence that the performance of small firms is less responsive to cyclical factors than large firms. The overall cyclical assessment needs to combine not only the estimated capacity, inventory and demand parameters but also the aggregate, economy wide effects captured by the annual dummy variables. In all cases the time dummies are jointly significant at the 5% level for the small firms estimates. For the large firms there are no significant aggregate effects for either the production to stock or the consumer goods subdivisions. In the case of consumer goods we have already identified potential mispecification as an issue, this mispecification may be either a cause or an effect of this absence of aggregate effects.

The time dummies are important because they indicate a different time path over the cycle for the large and small firms. In the aggregate the large firms exhibit a relatively traditional procyclical path. The dummy variables indicate a peak in 1987/88, a trough in 1990 and a substantial recovery in 1992. This is also the time path in the producer goods and production to order subdivisions. No substantial movement is indicated for the other subdivisions. For the small firms, considering all manufacturing industries, there is a slight peak in 1987/88 followed by much larger declines in 1990 and 1992. These troughs in 1990 and 1992 are common to all subdivisions except production to order industries. The timing of the peak varies across subdivisions with producer goods experiencing a peak in 1989 and consumer goods in 1988. The smaller peak is consistent with the theories discussed in section II above that predicted that the effect of demand changes would be to increase large firm margins to a greater extent. The relative behaviour in 1992 is somewhat enigmatic, not only because, as the last year of the data, it anticipates questions concerning the performance in the subsequent years. 1992 was a year in which manufacturing output was stagnant after the declines of 1990 and 1991. It is therefore surprising that such a strong resurgence in margins of large firms should occur in the absence of capacity pressures or substantial demand growth. The fact that the rise was concentrated in producer goods industries may indicate that it is a result of increases in the degree of collusion, in the form of the conventional raw material 'cartels' response to prolonged recession, which were reported at this time[35]. If this was the reason for the rise the small firms apparently did not benefit, indeed their price cost margins fell indicating greater competitive pressures. The reasons for the apparent absence of any discernible cyclical movement in consumer goods industries is unclear.

[35] The Economist 22/7/1994 reports tensions amongst several informal 'cartels' that had been set up over the previous two years.

The cyclical determinants of performance for small firms appear relatively constant across the sub samples of manufacturing industry. In each of the cases reported, PBC has a net positive effect on small firms performance, EINV has a negative effect on small firms performance and GROW has a positive effect on small firms performance. EINV is predicted to have a negative effect on performance if it induces "sales" of surplus stock or price cutting behaviour to restore a balance between production and sales. Similarly the growth rate of industry sales, as a general indicator or the state of demand, should lead to increased performance through the effects on the elasticity of demand discussed in section II.ii. above. The effect of excess capacity is perhaps more surprising[36]. Capacity feeds through into performance in many ways so there is no unambiguous expectation of sign on these coefficients. It would be more strongly expected that large firms would benefit from excess capacity as a tool for enforcing collusion, but this expectation is not universally met. The magnitude of these effects varies from sample to sample but for small firms the sign does not. PBC and GROW play a quantitatively larger role in consumer goods industries and EINV a larger role in producer goods industries. GROW also plays a more important role in production to stock industries rather than production to order industries for small firms.

The pattern is less clear for the large firm sample. Overall there is a weak positive effect of PBC but this masks a negative (procyclical) effect for producer goods industries and a positive (countercyclical) effect for consumer goods industries. EINV similarly reduces price cost margins of large firms in producer goods and production to stock industries but has negligible net effects otherwise. GROW has a strong positive effect on the price cost margins of large firms in producer goods and production to stock industries and a weaker effect elsewhere.

Subsequent to these estimates, we present in Table 10 estimates of the cyclical determinants of relative performance, measured by the ratio of large firm price cost margin to small firm price

[36] Cowling (1983) has pointed to a potential problem with the measure of excess capacity. The measure only identifies the proportion of firms reporting excess capacity not the degree of excess capacity. This distinction is potentially important because a recorded increase may simply indicate that excess capacity is becoming more widespread *within* the industry rather than increasing in depth.

cost margin. These estimates attempt to identify the significance of differences in the determinants of the performance of the two groups of firms. This analysis includes that same subdivision of manufacturing industry. Diagnostic tests are presented in Table 11. The diagnostic tests indicate failure of the Sargan test for the consumer and producer goods subdivisions. The consumer goods example is to be expected given the problem with the large firms estimates in this subdivision. Overall there is only a significant effect of the cyclical parameter in the producer goods and production to stock subdivisions. The analysis of the time dummies for this set of estimates confirms the above discussion of the separate estimates. There is a peak in the relative performance advantage in 1987/1988, a subsequent trough in 1990/1991 followed by a rapid recovery in 1992. Again this effect is concentrated in the producer goods and production to order industries. The time dummies are jointly insignificant in the other subdivisions.

EINV has a uniformly positive impact effect which indicates that small firms fare worse, the effects on both large and small individually is negative, under conditions of oversupply and this effect appears to persist for at least a year. This may indicate that the cost of holding surplus stocks are greater for small firms hence they must sell at a loss while they can. Similarly large firms fare better when capacity is limited and worse when excess capacity reappears, although there appears to be a response lag of at least a year to this effect. This effect is particularly marked for the producer goods and production to order sectors. The effect of demand growth appears to be relatively evenly distributed across large and small firms except in the producer goods industries where large firms benefit, in the long run at least, from higher demand growth to a greater extent.

VI. Conclusion

In conclusion we can identify five main results in this chapter. Firstly, whilst differences between the cyclical performance of large and small firms exist they are not dramatic. It cannot be stated unambiguously from our results that large or small firms are more or less responsive to the business cycle. The differences that exist are, in particular, unstable across subsectors of industry and are not uniformly related to any particular measure of the business cycle either in aggregate or in terms of industry specific measures. This result which slightly overturns our prior beliefs will be discussed to a greater extent in the next chapter.

Secondly, the analysis of the time dummies, in the estimates of the cyclical determinants of performance, and the estimates of the Schmalensee elasticities both point to difficulties in separating out cyclical from secular factors. The time path of the performance advantage of large firms has some cyclical elements but cannot be explained purely as a procyclical phenomenon. The Schmalensee estimates point to a secular rise in the performance advantage of large firms the causes of which may lie in corporate and industrial restructuring over the period rather than cyclical factors. In the relatively short time periods that we have to analyse here, in common with most of the previous studies of the trade cycle, the business of separating trend from cycle is difficult. The absence of a significant downturn in the profitability of large firms in the 1990's recession is a particular feature of these aggregate estimates.

Thirdly, the analysis of the time effects on relative performance have indicated a more complex cyclical path than could be accounted for in examining either the aggregate or the industry level growth of manufacturing output or demand. We have identified important differences in the response of the different subsectors both in their response to aggregate shocks embodied in the time dummies and in response to local demand, capacity and inventory levels. We have offered some potential explanations for these differences between the subsectors, but there is a need for further work to check both the robustness of the results contained here and to examine more closely the reasons why different types of industries react differently to the cycle.

Fourthly, the analysis of the dynamic adjustment processes provides evidence that the performance of the small firms sector does adjust more rapidly to both cyclical factors and general disequilibrium disturbances. If we combine these estimates with the previous estimates which indicate that, at the firm level, there is little evidence of more rapid adjustment by smaller firms we can attribute this result to an effect affecting the small firm sample, i.e. the effects of greater entry and greater intra group mobility leading to smoother adjustment than for the more stable large firm group, rather than a result purely driven by differential sizes.

Fifthly, we have been able to provide evidence that the presence of relatively larger small firms leads to a more rapid adjustment on the part of the large firms. In this case we cannot say definitively whether this is because the presence of larger small firms leads to greater changes in composition of the large firms group or because there is a greater competitive threat or a combination of the two. Intuitively the presence of a number of firms with market shares significant enough to approach the largest five is likely to provide both of these potential threats to the existing large firms.

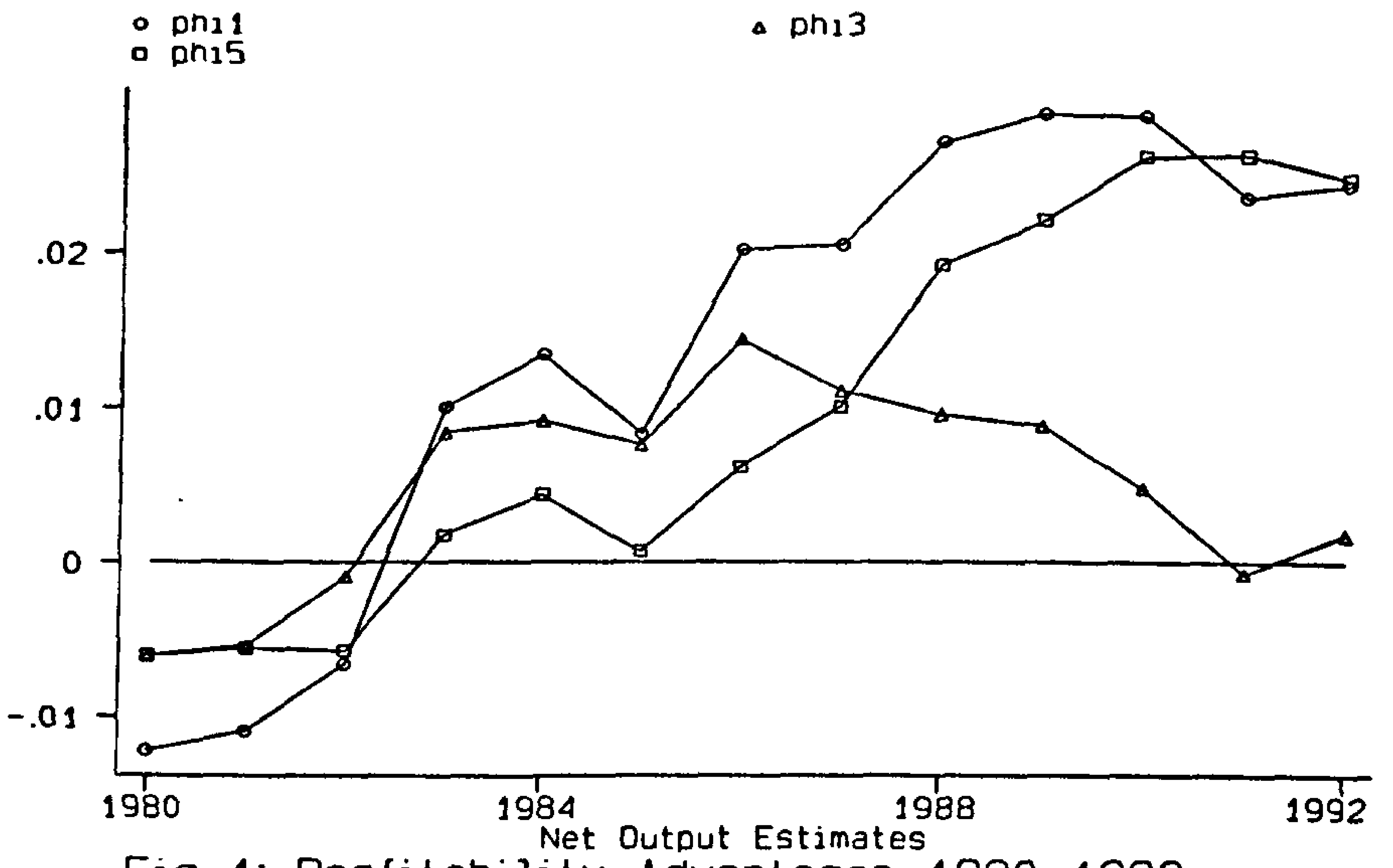


Fig 1: Profitability Advantages 1980-1992

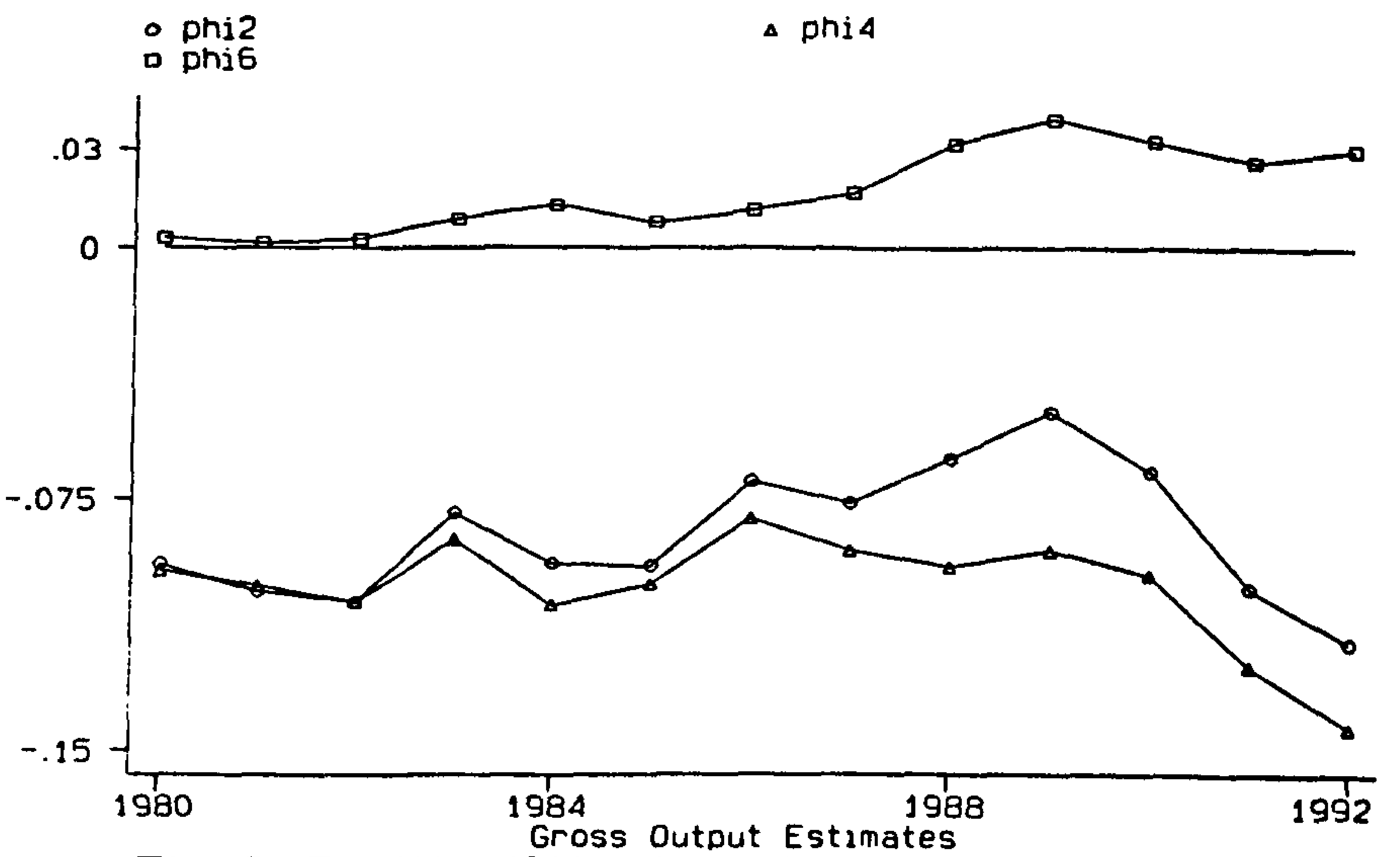


Fig 2: Profitability Advantages 1980-1992

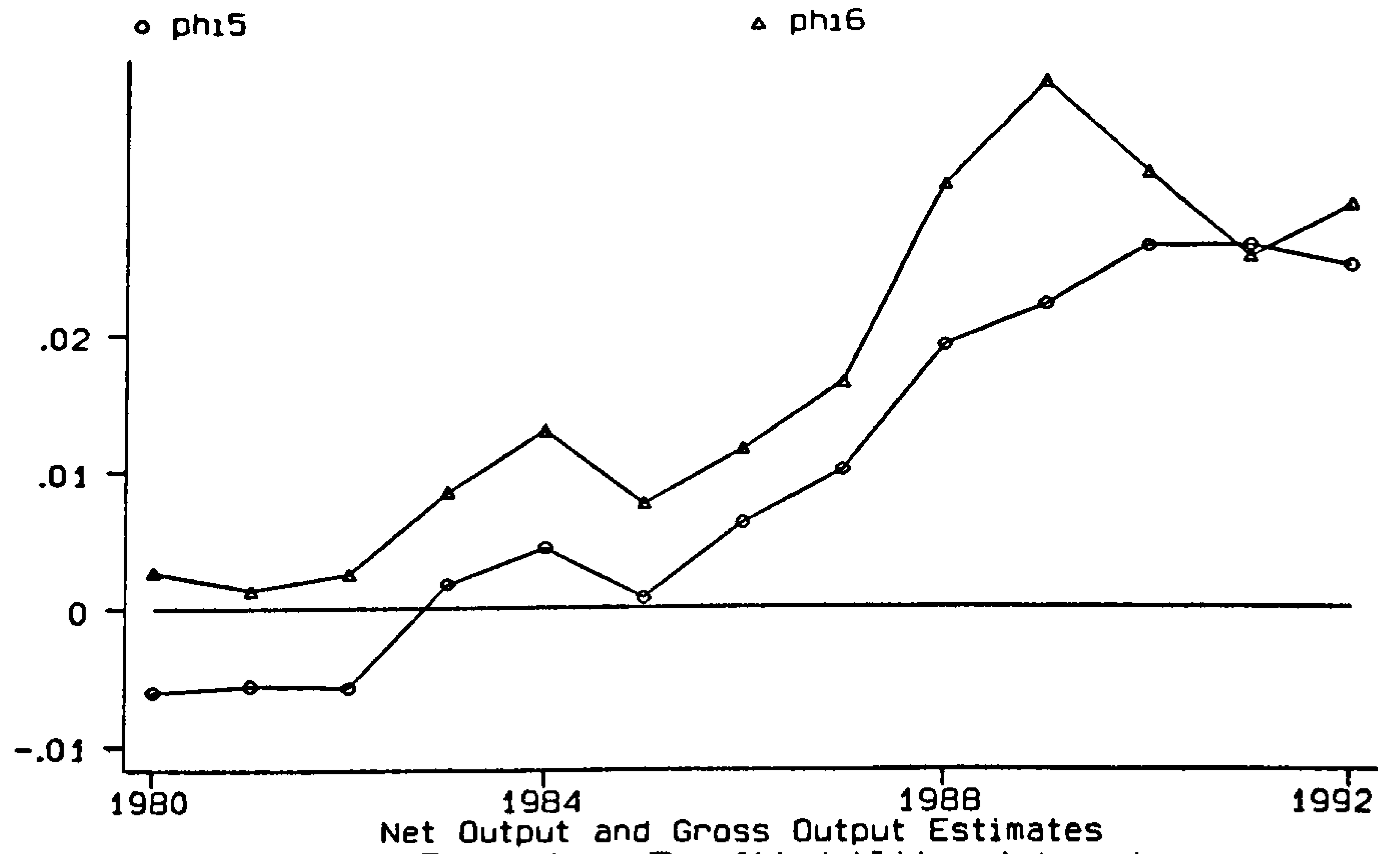


Fig 3: Intra Industry Profitability Advantage

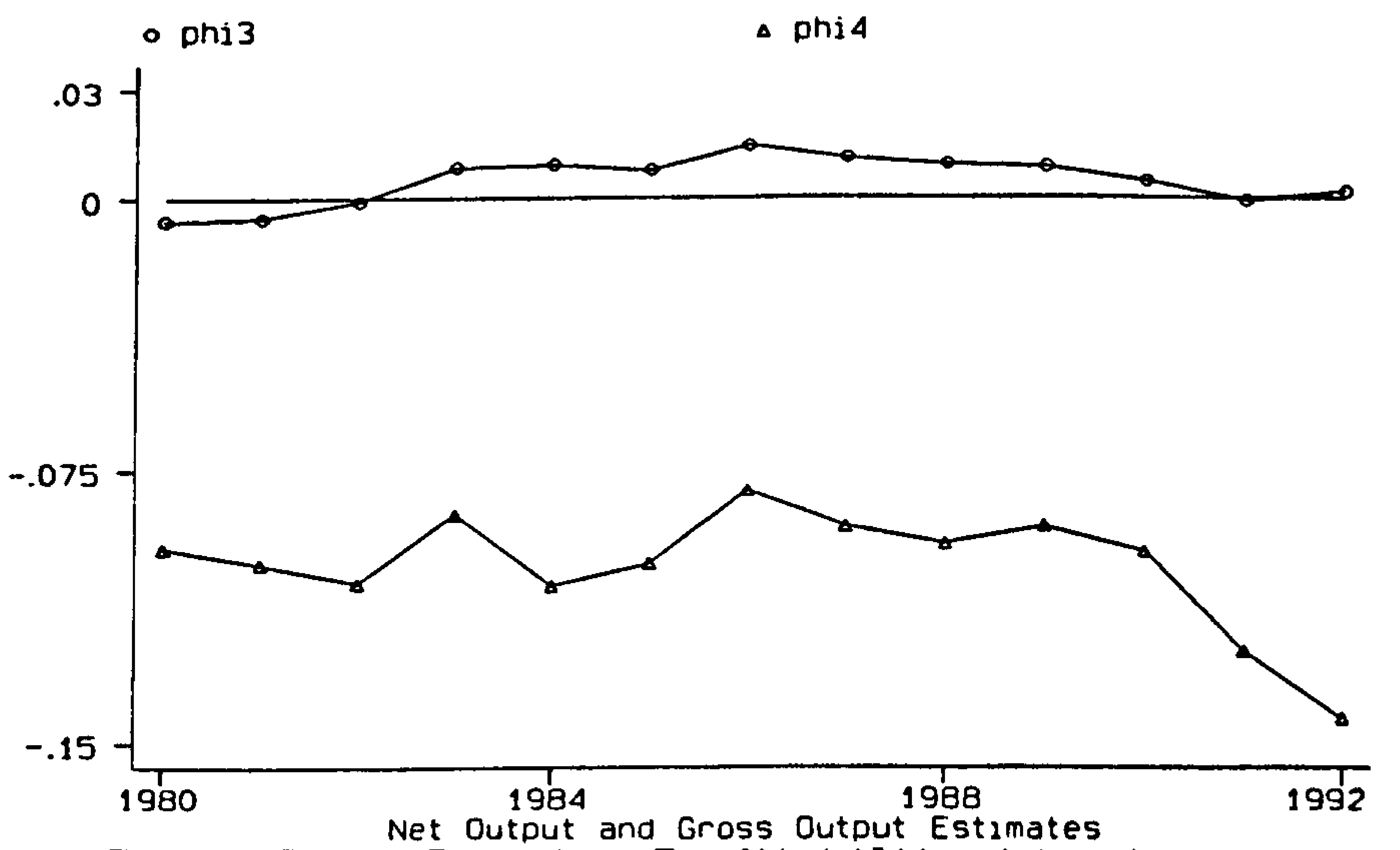


Fig 4: Inter Industry Profitability Advantage

Table 1: Summary Statistics ranked by Industry Concentration

		Concentration		
	All	High	Medium	Low
Coefficient of Variation				
INVSAL (large)	.187	.197	.156	.209
	(.094)	(.083)	(.061)	(.123)
PCM (large)	.148	.171	.141	.131
	(.077)	(.092)	(.073)	(.058)
INVSAL (small)	.157	.203	.122	.146
	(.079)	(.101)	(.043)	(.058)
PCM (small)	.113	.176	.089	.073
	(.090)	(.124)	(.042)	(.037)
Excess Inv.	1.086	1.332	.971	.957
	(.624)	(.836)	(.447)	(.453)
Below Capac.	.331	.355	.327	.310
	(.075)	(.093)	(.074)	(.047)
Grow	16.04	17.5	19.6	11.1
	(37.5)	(30.4)	(54.6)	(19.4)
Mean				
INVSAL (large)	.214	.249	.187	.205
	(.186)	(.292)	(.094)	(.101)
PCM (large)	.317	.293	.321	.337
	(.097)	(.090)	(.101)	(.099)
INVSAL (small)	.169	.175	.163	.169
	(.075)	(.105)	(.052)	(.060)
PCM (small)	.302	.286	.304	.316
	(.077)	(.087)	(.075)	(.068)
Excess Inv.	.152	.149	.166	.142
	(.080)	(.086)	(.083)	(.069)
Below Capac.	.583	.574	.586	.588
	(.083)	(.093)	(.088)	(.068)
Grow	.043	.036	.046	.047
	(.028)	(.030)	(.030)	(.024)
No of Industries	87	29	29	29

Notes: INVSAL is the ratio of stocks and work in progress to gross output. PCM is the conventionally defined price cost margin. Excess Inv. is the net proportion of firms reporting excessive rather than insufficient levels of inventory. Below Capac. is the proportion of firms reporting below capacity operation. Grow is the growth rate of industry sales. Standard Deviations are reported in parentheses. High, Medium and Low Concentration samples are the top, middle and bottom thirds of the sample ranked by concentration.

Table 2: Estimates of Profitability Elasticities

Net Output based estimates

Year	ϕ_1	ϕ_3	ϕ_5
1980	-0.012	-0.006	-0.006**
1981	-0.011	-0.005	-0.006*
1982	-0.007	-0.001	-0.006*
1983	0.010	0.008	0.002
1984	0.013	0.009	0.004**
1985	0.008	0.008	0.001
1986	0.020*	0.014	0.006*
1987	0.020*	0.011	0.010*
1988	0.027**	0.010	0.019**
1989	0.029**	0.009	0.022**
1990	0.029**	0.005	0.026*
1991	0.024	-0.001	0.026**
1992	0.024	0.002	0.024*

Notes:

1. ϕ_1, ϕ_3, and ϕ_5 are the original, synthetic and adjusted estimates respectively, as defined in the text.
2.* indicates test distributed as N(0,1) that estimate is different from zero, significant at the 5% level

** indicates same test, significant at the 1% level

Table 3: Estimates of Profitability Elasticities

Gross Output based estimates

Year	ϕ_2	ϕ_4	ϕ_6
1980	-0.095*	-0.096*	0.004
1981	-0.103*	-0.101*	0.001
1982	-0.106*	-0.106*	0.003
1983	-0.079*	-0.087*	0.008
1984	-0.094*	-0.107*	0.013**
1985	-0.095**	-0.101**	0.008
1986	-0.070*	-0.081*	0.012*
1987	-0.076*	-0.090*	0.016*
1988	-0.063*	-0.095*	0.031**
1989	-0.050*	-0.090*	0.038**
1990	-0.067*	-0.098*	0.032*
1991	-0.102**	-0.125*	0.025*
1992	-0.118**	-0.143**	0.029*

Notes:

1. ϕ_2, ϕ_4, and ϕ_6 are the original, synthetic and adjusted estimates respectively, as defined in the text.
2.* indicates test distributed as N(0,1) that estimate is different from zero, significant at the 5% level

** indicates same test, significant at the 1% level

Table 4: Persistence of Profits Estimates (UK manufacturing 1980-1992)

	λ_s	λ_l	$(\alpha/1-\lambda)_s$	$(\alpha/1-\lambda)_l$
CONC	-0.114 (0.73)	0.220 (1.38)	0.046 (1.05)	0.044 (1.68)
MS	1.575 (1.69)	-3.416 (4.83)	-0.174 (0.59)	-0.277 (1.34)
ADINT	0.642 (1.01)	-0.078 (0.12)	0.377 (2.62)	0.542 (4.87)
PROD	0.029 (1.22)	-0.019 (1.96)	0.0003 (0.04)	-0.002 (1.70)
MP	0.016 (2.70)	0.004 (0.54)	0.0007 (0.76)	0.002 (1.18)
MY	0.514 (1.41)	-0.533 (1.80)	-0.353 (1.20)	-0.853 (11.87)
CONST	-0.292 (1.32)	0.675 (3.80)	0.447 (2.95)	0.702 (15.05)
F-Test	2.89 (6,80)	6.34 (6,77)	4.31 (6,80)	28.8 (6,77)
Observations	87	84	87	84

Notes:

1. Dependent variables are based upon the estimated coefficients of a regression of the form $PCM_t=\alpha+\lambda PCM_{t-1}+\varepsilon_t$ for each industry. s and l refer to the small and large firms size classes respectively. Three observations were excluded from the large firm sample because of the implied instability of the process from the estimated coefficients.
2. CONC is the five firm concentration ratio, MS is the average market share of the small firms, ADINT is a measure of advertising intensity, PROD is the length of the production run, MP is number of plants per large enterprise, MY is the ratio of material inputs to gross output. All independent variables are industry or size class means for the period 1980-1992.
3. All estimation is by OLS. Heteroskedasticity robust t ratios are in parentheses. F-Test is a F-test of the joint significance of the independent variables, degrees of freedom in parentheses. All tests reject the null of joint insignificance at the 5% level.

Table 5: Dynamic Persistence Estimates (UK manufacturing 1985-1992)

	SPCM			LPCM		
	(i)	(ii)	(iii)	(iv)	(v)	(vi)
CONC	0.036 (2.08)	-0.121 (1.49)		0.028 (1.22)	0.128 (1.68)	
CONC*PCM	-0.074 (1.31)	0.157 (0.63)	-0.587 (0.93)	0.011 (0.16)	-0.325 (1.86)	0.338 (0.49)
MS	0.223 (1.84)	0.108 (0.62)		0.246 (2.44)	-0.025 (0.19)	
MS*PCM	-1.506 (2.09)	-0.014 (0.02)	-5.077 (1.47)	-1.936 (3.59)	0.301 (0.45)	-7.288 (5.19)
ADINT	0.023 (0.35)	-0.250 (1.19)		0.184 (2.23)	0.004 (0.02)	
ADINT*PCM	0.379 (1.79)	0.570 (0.89)	-0.378 (0.17)	-0.010 (0.04)	-0.724 (0.86)	-6.358 (2.18)
PROD	0.007 (1.07)	0.008 (1.41)		0.001 (0.35)	-0.004 (1.53)	
PROD*PCM	-0.015 (0.81)	-0.016 (1.13)	0.033 (1.16)	-0.006 (1.45)	0.007 (1.56)	-0.057 (2.99)
MP	0.000 (0.04)	0.004 (1.67)		-0.000 (0.05)	-0.005 (1.23)	
MP*PCM	-0.0003 (0.13)	-0.011 (1.64)	0.001 (0.03)	0.0001 (0.07)	0.009 (1.47)	0.003 (0.29)
MY	-0.380 (8.58)	-0.646 (4.46)		-0.322 (9.82)	-0.941 (5.46)	
MY*PCM	-0.224 (1.90)	-0.865 (1.94)	-0.404 (0.21)	-0.603 (7.16)	-0.066 (0.14)	-0.892 (1.13)
PCM_{t-1}	0.556 (4.13)	0.485 (1.77)	0.880 (0.70)	0.772 (12.9)	0.143 (0.46)	2.039 (3.47)
Serial Corr.	6.894	-0.656	-0.669	7.874	-1.439	-0.207
Sargan	N/A	52.5	76.2	N/A	83.7	55.7
Time	2.99	56.4	26.0	12.8	27.3	25.6
Wald	19.6	6.96	8.39	94.3	13.0	44.1

Notes:

1. Dependent variables are the price cost margins of the small and large firm groups in (i)-(iii) and (iv)-(vi) respectively.
2. Estimation is by OLS in levels in (i) and (iv). Estimation in (ii), (iii), (v) and (vi) is by first differenced instrumental variables using the Arellano and Bond GMM procedure. Instruments are lags of the dependent and independent variables dated t-2 and back. Robust one step t ratios are in parentheses.
3. CONC is the five firm concentration ratio, MS is the average market share of the small firms, ADINT is a measure of advertising intensity, PROD is the length of the production run, MP is number of plants per large enterprise, MY is the ratio of material inputs to gross output. The interaction terms in (i),(ii),(iv) and (v) are the contemporaneous independent variable multiplied by the lagged dependent variable. e.g. CONC*PCM=$CONC_t$*PCM_{t-1}. In (iii) and (vi) independent variables are the industry or size class means for the period 1980-1992 interacted with the lagged dependent variable.
4. Serial Corr. is a robust test of first order correlation in (i) and (iv), and a robust test of second order correlation in (ii), (iii), (v) and (vi). The test is distributed as N(0,1). Sargan is the test of overidentifying restrictions distributed as χ^2 with 43 degrees of freedom. Time is a test of the joint significance of the yearly dummy variables distributed as χ^2 with degrees of freedom in parentheses. Wald is a joint test of the significance of the interaction terms distributed as χ^2 with 6 degrees of freedom.

Table 6: Business Cycle Estimates - All Manufacturing

	$LPCM_t$	$SPCM_t$
PBC_t	0.200 (1.26)	0.129 (1.03)
PBC_{t-1}	-0.042 (0.35)	0.168 (1.48)
PBC_{t-2}	0.027 (0.24)	0.132 (1.33)
PBC_{t-3}	0.004 (0.02)	0.310 (1.80)
$EINV_t$	0.086 (0.58)	-0.172 (1.51)
$EINV_{t-1}$	0.135 (0.87)	-0.060 (0.36)
$EINV_{t-2}$	0.142 (0.95)	-0.064 (0.40)
$EINV_{t-3}$	-0.263 (1.99)	-0.186 (1.76)
$GROW_t$	0.063 (2.86)	0.050 (2.17)
$GROW_{t-1}$	0.057 (1.84)	0.043 (2.11)
$GROW_{t-2}$	0.052 (2.14)	0.038 (2.02)
$GROW_{t-3}$	0.060 (3.04)	0.029 (1.88)
$INVSAL_t$	-0.003 (0.13)	0.016 (0.32)
$INVSAL_{t-1}$	-0.054 (1.55)	-0.009 (0.17)
$INVSAL_{t-2}$	0.063 (1.99)	0.011 (0.25)
$INVSAL_{t-3}$	-0.055 (1.50)	0.047 (0.66)
MY_t	-0.963 (21.4)	-0.890 (17.8)
MY_{t-1}	0.104 (0.91)	-0.022 (0.19)
$CONC_t$	0.039 (0.71)	-0.135 (2.99)
$CONC_{t-1}$	-0.097 (2.10)	0.033 (0.70)
MS_t	0.333 (3.95)	0.261 (2.62)
MS_{t-1}	0.135 (1.33)	-0.027 (0.35)
PCM_{t-1}	0.136 (1.07)	-0.006 (0.05)

Notes: Estimation is by first difference IV. LPCM is the price cost margins of the large firms. SPCM is the price cost margin of the small firms. Robust one-step t ratios are in parentheses. Coefficients on PBC and EINV have been multiplied by 10. Instrument information, data definitions and diagnostic tests are reported in Table 9 and its notes.

Table 7: Business Cycle Estimates: Consumer/Producer Goods Industries

	Consumer		Producer	
	$LPCM_t$	$SPCM_t$	$LPCM_t$	$SPCM_t$
PBC_t	0.030 (0.15)	0.200 (1.37)	-0.054 (0.34)	0.068 (0.49)
PBC_{t-1}	0.204 (1.65)	0.287 (2.19)	-0.002 (0.02)	0.181 (1.91)
PBC_{t-2}	0.181 (1.18)	0.213 (1.94)	0.026 (0.19)	0.197 (1.98)
PBC_{t-3}	0.281 (1.54)	0.421 (1.88)	-0.357 (2.13)	0.099 (0.64)
$EINV_t$	-0.019 (0.10)	-0.065 (0.38)	-0.049 (0.43)	-0.195 (1.53)
$EINV_{t-1}$	0.158 (0.811)	-0.042 (0.28)	-0.000 (0.01)	-0.232 (1.80)
$EINV_{t-2}$	0.065 (0.401)	-0.165 (1.28)	0.019 (0.18)	-0.183 (1.61)
$EINV_{t-3}$	-0.120 (0.585)	-0.226 (1.70)	-0.281 (2.21)	-0.157 (1.92)
$GROW_t$	0.475 (0.02)	0.044 (2.06)	0.059 (3.95)	0.040 (2.50)
$GROW_{t-1}$	0.027 (1.10)	0.046 (2.56)	0.041 (2.05)	0.017 (0.95)
$GROW_{t-2}$	0.023 (1.14)	0.042 (2.63)	0.045 (2.56)	0.005 (0.28)
$GROW_{t-3}$	0.014 (0.80)	0.030 (1.85)	0.029 (1.54)	0.006 (0.39)
$INVSAL_t$	0.032 (0.87)	-0.021 (0.51)	-0.026 (0.70)	-0.033 (0.54)
$INVSAL_{t-1}$	0.065 (1.19)	-0.017 (0.43)	-0.022 (0.68)	0.042 (0.82)
$INVSAL_{t-2}$	-0.085 (1.87)	-0.034 (1.18)	0.110 (5.29)	0.003 (0.04)
$INVSAL_{t-3}$	0.044 (0.72)	0.011 (0.21)	-0.012 (0.31)	0.069 (0.95)
MY_t	-0.858 (13.0)	-0.977 (18.8)	-0.950 (17.6)	-0.888 (10.7)
MY_{t-1}	0.107 (0.77)	-0.212 (1.55)	0.150 (1.17)	0.059 (0.53)
$CONC_t$	0.096 (1.09)	-0.051 (0.98)	0.056 (1.03)	-0.130 (3.68)
$CONC_{t-1}$	-0.002 (0.03)	0.022 (0.32)	-0.065 (1.43)	0.040 (0.76)
MS_t	-0.632 (0.72)	1.375 (2.19)	0.348 (4.98)	0.124 (1.81)
MS_{t-1}	-0.363 (0.34)	1.514 (2.85)	0.042 (0.69)	-0.076 (1.18)
PCM_{t-1}	0.064 (0.48)	-0.209 (1.40)	0.125 (0.96)	0.043 (0.39)

Notes: See notes to Tables 6 and 9

Table 8: Business Cycle Estimates: Production to Stock/Order Industries

	Production to Stock		Production to Order	
	$LPCM_t$	$SPCM_t$	$LPCM_t$	$SPCM_t$
PBC_t	0.120 (1.02)	-0.038 (0.29)	0.005 (0.03)	0.100 (0.59)
PBC_{t-1}	0.246 (1.36)	0.149 (2.07)	0.106 (0.96)	0.188 (1.40)
PBC_{t-2}	0.184 (1.19)	0.077 (0.49)	0.075 (0.49)	0.200 (1.87)
PBC_{t-3}	0.087 (0.48)	0.251 (2.18)	-0.036 (0.21)	0.014 (0.09)
$EINV_t$	-0.435 (1.96)	-0.195 (1.29)	0.182 (1.03)	-0.109 (0.84)
$EINV_{t-1}$	-0.351 (1.38)	0.098 (0.55)	-0.018 (0.15)	-0.071 (0.55)
$EINV_{t-2}$	-0.290 (1.29)	0.046 (0.26)	0.017 (0.14)	-0.057 (0.48)
$EINV_{t-3}$	-0.206 (1.30)	-0.067 (0.39)	-0.232 (1.89)	-0.122 (1.48)
$GROW_t$	0.047 (2.34)	0.035 (1.58)	0.055 (3.00)	0.077 (2.76)
$GROW_{t-1}$	0.058 (2.48)	0.037 (1.96)	0.029 (1.29)	0.038 (1.62)
$GROW_{t-2}$	0.045 (2.14)	0.053 (2.47)	0.029 (1.34)	0.026 (1.19)
$GROW_{t-3}$	0.023 (1.10)	0.051 (3.02)	0.017 (1.09)	-0.021 (0.96)
$INVSAL_t$	0.007 (0.21)	0.052 (1.95)	-0.009 (0.31)	-0.089 (1.15)
$INVSAL_{t-1}$	0.036 (1.59)	-0.020 (0.93)	-0.042 (1.49)	0.114 (1.41)
$INVSAL_{t-2}$	-0.062 (1.69)	-0.034 (0.94)	0.086 (3.99)	-0.131 (1.69)
$INVSAL_{t-3}$	0.046 (2.11)	-0.074 (1.60)	-0.035 (0.96)	0.044 (0.55)
MY_t	-0.852 (17.2)	-0.862 (17.6)	-0.983 (23.2)	-0.966 (24.9)
MY_{t-1}	0.051 (0.33)	0.118 (1.17)	0.143 (1.28)	0.014 (0.10)
$CONC_t$	0.114 (2.85)	-0.088 (2.00)	0.089 (1.19)	-0.122 (2.83)
$CONC_{t-1}$	-0.012 (0.27)	0.051 (1.18)	-0.090 (1.75)	0.023 (0.53)
MS_t	0.244 (3.18)	0.225 (2.13)	-0.009 (0.012)	-0.112 (0.27)
MS_{t-1}	0.056 (0.83)	0.033 (0.36)	0.030 (0.05)	0.010 (0.02)
PCM_{t-1}	0.024 (0.14)	0.142 (1.47)	0.139 (1.27)	-0.024 (0.16)

Notes: See notes to Tables 6 and 9

Table 9: Diagnostic Tests for Tables 6, 7 and 8

	All Manufacturing			
	LPCM	SPCM		
Sargan	64.2(57)	55.3(57)		
Time Dums	25.1(8)	34.3(8)		
Wald	37.1(12)	35.4(12)		
Serial Correlation	-0.319	-0.065		
PBC	2.92	5.04		
EINV	6.62	5.01		
GROW	12.2	8.04		
INVSAL	5.75	1.12		
T-3	27.5	9.06		
Observations	696	696		
	Consumer		Producer	
	LPCM	SPCM	LPCM	SPCM
Sargan	95.4(57)	75.0(57)	67.6(57)	61.8(57)
Time Dums	9.54(8)	34.6(8)	20.9(8)	21.7(8)
Wald	9.15(12)	26.2(12)	89.9(12)	34.0(12)
Serial Correlation	0.437	-0.118	-0.919	-1.352
PBC	4.56	9.45	7.79	5.30
EINV	1.84	4.77	11.2	4.60
GROW	2.44	8.08	20.8	9.52
INVSAL	4.55	4.76	34.8	3.06
T-3	3.78	7.93	34.4	3.76
Observations	328	328	368	368
	Production to Stock		Production to Order	
	LPCM	SPCM	LPCM	SPCM
Sargan	69.1(57)	64.1(57)	80.9(57)	69.5(57)
Time Dums	7.24(8)	18.9(8)	19.4(8)	24.3(8)
Wald	56.3(12)	43.1(12)	52.0(12)	72.4(12)
Serial Correlation	0.823	-0.351	-0.955	0.447
PBC	3.23	7.89	1.69	3.90
EINV	4.96	12.9	5.98	2.44
GROW	16.7	11.2	15.3	23.0
INVSAL	15.4	7.76	18.9	4.27
T-3	9.73	19.1	7.48	3.87
Observations	240	240	456	456

Notes:

1. Estimation is by first differenced instrumental variables using the GMM procedure of Arrellano and Bond(1988). Instruments are lags of the independent and dependent variables dated t-2 and back.

2. PBC is the percentage of firms reporting below capacity operation. EINV is the net proportion of firms reporting excessive finished goods inventory levels. INVSAL is the ratio of stocks and work in progress to gross output. GROW is the growth rate of industry sales. MY is the ratio of material inputs to gross output. CONC is the five firm concentration ratio. MS is the mean market share of the small firms.

3. Sargan is a test of overidentifying restrictions distributed as χ^2 with degrees of freedom in parentheses. Time Dums is a joint test of the significance of the yearly dummy variables distributed as χ^2 with degrees of freedom in parentheses. Serial Correlation is a test of second order serial correlation distributed as N(0,1). Wald is a joint test of the significance of PBC, EINV and GROW distributed as χ^2 with degrees of freedom in parentheses. PBC, EINV, INVSAL and GROW are joint tests of the 4 coefficients on each of these variables plus their lags. T-3 is a test of the joint significance of the 4 coefficients dated t-3. These tests are distributed as χ^2 with 4 degrees of freedom. Critical values for these latter tests are 9.49 and 13.3 at the 5% and 1% levels respectively

Table 10: Business Cycle Estimates: Relative Performance

	All $RPCM_t$	Consumer $RPCM_t$	Producer $RPCM_t$	P/Stock $RPCM_t$	P/Order $RPCM_t$
PBC_t	0.192 (1.14)	-0.008 (0.04)	-0.042 (0.34)	0.087 (0.79)	-0.256 (1.92)
PBC_{t-1}	-0.212 (1.34)	-0.109 (0.82)	-0.235 (1.98)	-0.296 (1.79)	-0.302 (2.06)
PBC_{t-2}	-0.064 (0.41)	0.088 (0.56)	-0.189 (2.42)	0.019 (0.13)	-0.290 (2.29)
PBC_{t-3}	0.200 (0.96)	0.144 (0.79)	-0.019 (0.11)	-0.065 (0.36)	-0.049 (0.33)
$EINV_t$	0.286 (2.00)	0.216 (1.25)	0.125 (1.53)	0.283 (2.07)	0.402 (2.76)
$EINV_{t-1}$	0.221 (1.45)	0.211 (1.13)	0.036 (0.36)	0.175 (0.77)	0.107 (0.94)
$EINV_{t-2}$	0.156 (1.09)	0.112 (0.67)	-0.022 (0.23)	0.048 (0.27)	0.119 (1.28)
$EINV_{t-3}$	-0.165 (1.50)	0.193 (0.88)	-0.188 (1.94)	-0.090 (0.44)	-0.034 (0.27)
$GROW_t$	-0.066 (0.32)	-0.072 (0.42)	-0.113 (0.86)	-0.017 (0.09)	-0.263 (1.26)
$GROW_{t-1}$	0.109 (0.44)	-0.152 (0.54)	0.026 (0.23)	0.113 (0.59)	-0.421 (1.53)
$GROW_{t-2}$	0.060 (0.28)	-0.200 (0.90)	0.113 (0.97)	-0.137 (0.64)	-0.306 (1.26)
$GROW_{t-3}$	0.164 (0.665)	-0.216 (1.16)	0.355 (2.93)	-0.168 (0.72)	0.179 (1.24)
$RINVSAL_t$	-0.092 (3.16)	0.037 (0.27)	-0.042 (1.43)	0.217 (1.52)	-0.037 (1.41)
$RINVSAL_{t-1}$	0.052 (0.83)	0.012 (0.12)	0.016 (0.52)	0.040 (0.36)	0.023 (0.48)
$RINVSAL_{t-2}$	-0.039 (1.99)	0.024 (0.23)	-0.023 (1.46)	0.182 (1.86)	-0.025 (1.20)
$RINVSAL_{t-3}$	0.115 (1.43)	-0.027 (0.27)	0.037 (0.74)	0.120 (1.35)	0.038 (0.52)
RMY_t	-2.261 (7.16)	-2.380 (8.73)	-1.770 (9.34)	-2.525 (8.07)	-1.729 (7.32)
RMY_{t-1}	-0.033 (0.15)	-0.159 (0.66)	-0.022 (0.12)	-0.163 (0.61)	0.122 (0.42)
$CONC_t$	0.974 (1.50)	0.720 (0.67)	0.529 (1.31)	0.570 (1.09)	0.691 (1.34)
$CONC_{t-1}$	-1.266 (2.01)	-0.078 (0.11)	-0.321 (0.82)	-0.797 (1.84)	0.263 (0.44)
MS_t	0.075 (0.05)	-6.727 (0.71)	1.513 (3.48)	-0.253 (0.26)	6.032 (1.25)
MS_{t-1}	2.790 (2.21)	-11.90 (1.21)	1.694 (3.66)	1.447 (2.24)	-4.980 (1.03)
$RPCM_{t-1}$	0.015 (0.20)	-0.023 (0.32)	-0.064 (0.63)	-0.021 (0.36)	-0.005 (0.04)

Notes: RPCM is the ratio of the large firms' price cost margin to the small firms' price cost margin. Robust one-step t ratios are in parentheses. Instrument information, data definitions and diagnostic tests are reported in Table 11 and its notes.

Table 11: Diagnostic Tests for Table 10

	All RPCM	Consumer RPCM	Producer RPCM	P/Stock RPCM	P/Order RPCM
Sargan	69.3(57)	99.1(57)	91.1(57)	74.0(57)	73.9(57)
Time Dums	15.7(8)	4.31(8)	23.1(8)	10.1(8)	18.6(8)
Wald	18.9(12)	5.50(12)	36.0(12)	15.6(12)	28.2(12)
Serial Correlation	0.490	0.222	-0.235	-0.508	0.650
PBC	4.40	3.28	6.22	5.24	10.6
EINV	8.97	1.82	7.92	5.87	8.51
GROW	1.76	1.61	10.1	4.12	5.57
INVSAL	13.6	0.87	6.25	17.9	3.64
T-3	6.17	3.09	17.9	2.19	2.74
Observations	696	328	368	240	456

Notes:

1. Estimation is by first differenced instrumental variables using the GMM procedure of Arrellano and Bond(1988). Instruments are lags of the independent and dependent variables dated t-2 and back.
2. PBC is the percentage of firms reporting below capacity operation. EINV is the net proportion of firms reporting excessive finished goods inventory levels. RINVSAL is the large firms' ratio of stocks and work in progress to gross output divided by the small firms' ratio of stocks and work in progress to gross output. GROW is the growth rate of industry sales. RMY is the large firms' ratio of material inputs to gross output divided by the small firms' ratio of material inputs to gross output. CONC is the five firm concentration ratio. MS is the mean market share of the small firms.
3. Sargan is a test of overidentifying restrictions distributed as χ^2 with degrees of freedom in parentheses. Time Dums is a joint test of the significance of the yearly dummy variables distributed as χ^2 with degrees of freedom in parentheses. Serial Correlation is a test of second order serial correlation distributed as N(0,1). Wald is a joint test of the significance of PBC, EINV and GROW distributed as χ^2 with degrees of freedom in parentheses. PBC, EINV, INVSAL and GROW are joint tests of the 4 coefficients on each of these variables plus their lags. T-3 is a test of the joint significance of the 4 coefficients dated t-3. These tests are distributed as χ^2 with 4 degrees of freedom. Critical values for these latter tests are 9.49 and 13.3 at the 5% and 1% levels respectively
4. Unlike the estimates presented in Tables 6, 7 and 8 the coefficients on PBC and EINV in Table 10 have not been multiplied by 10.

Data Appendix

PRODL: The series for the length of production run were constructed in accordance with the method set out in Coutts, Godley and Nordhaus (1978, p37-41). Formally
PRODL=(2*INV)/[GROUT*[1+(2*MAT)/(3*GROUT)]]   where INV is the level of stocks, GROUT is the level of gross output, MAT is the level of material inputs.
MP: No of establishments/businesses operated by the largest five enterprises per enterprise

Other data series taken from the UK Census of Production are as defined in the Data Appendix to Chapter 4.

CBI Data Series

The creation of the excess capacity series data was taken from the CBI quarterly industrial trends survey, Question 4.
"Is your present level of output below capacity?(i.e. are you operating below a satisfactory level of operation)"
To which respondents fill in one of three boxes "Yes", "No" or "N/A" . The proportions answering "Yes" were collected for each industry category.

The creation of the excessive inventory data is based upon Question 5c from the survey.
"Excluding seasonal variations do you consider that in volume terms your present stocks of finished goods are " "More than adequate" "Adequate" "Less than adequate" "N/A"
The measure is the difference between the proportion of firms reporting "More than adequate" and the proportion reporting "Less than adequate". A positive figure indicates on balance excessive stocks of finished goods, a negative figure indicates on balance insufficient stocks.

In each case the annual figure is calculated as the mean of the 4 quarterly figures.

Matching of CBI industry categories to SIC and MLH categories.

The industry categories used and their correlation with the 1980 SIC categories are given in the Table below. In the majority of cases there is a direct correlation between CBI categories and the SIC industries, although most of the CBI categories cover more than one three digit SIC code. In some cases a three digit SIC code is covered by more than one CBI category. In this case a weighted average measure has been constructed where the weights are value added of the four digit SIC code. The additional problem in constructing the series was that some of the CBI categories changed at the end of 1983, having previously been based upon the 1968 MLH industry classification. Because of the broad nature of the categories this made less of a difference than may be imagined and most of the categories are essentially unchanged. However a number of industries were affected and the details of these are given below.

223 2234=Fabricated Metal Goods 2235=Ferrous Metals
251 2514,2515=Resins and Plastics 2511,2512,2513,2516=Industrial and Agricultural Chemicals
255 2551=Consumer Chemicals 2552=Industrial and Agricultural Chemicals
256 2562,2564,2565,2567,2568=Ind and Ag Chemicals 2563=Consumer Chem 2569=Rubber
311 3111=Ferrous Metals 3112=Non Ferrous Metals
455 4555=Furniture 4556,4557=Other textiles
467 4671=Furniture 4672=Timber

CBI category	SIC codes
Ferrous Metals	221,222,223
Non Ferrous Metals	224
Building Materials	241-246
Glass and Ceramics	247,248
Industrial Chemicals	2511-2516,2562-2569
Agricultural Chemicals	2513,2568
Pharmaceuticals	255,257,258,259
Man Made Fibres	260
Foundries	311,312
Metal Goods n.e.s.	313,314,3162-3169
Hand Tools	3161
Constructional Steelwork	3204
Heavy Industrial Plant	3205
Agricultural Machinery	321
Metal Work	3221
Engineers Small Tools	3222
Industrial Machinery	323,324,327,3285-6
Contractor's Plant	325
Industrial Engines	3281,3283,3287-8
Heating	3284
Other Mechanical Eng.	326,3289,329
Office Machinery	330
Electrical Industrial Goods	341-3,3442,347-8
Electronic Industrial Goods	3441,3443,3444,3453
Electrical Consumer Goods	346
Electronic Consumer Goods	3452,3454
Motor Vehicles	351,352,353
Shipbuilding	361
Aerospace	362,363,364,365
Instrument Engineering	371,372,373,374
Food	411-423
Drink and Tobacco	424-429
Woollen Textiles	431
Spinning and Weaving	432,433,434
Hosiery	436
Textile Consumer Goods	438,4555,4557
Other Textiles	435,437,439,4556
Footwear	451
Leather	441,442
Clothing and Fur	453,456
Timber products	461-466
Furniture	467
Pulp and Paper	471
Paper and Board Products	472
Printing and Publishing	475
Rubber Products	481,482
Plastics Products	483
Other	491-495

SIC	Pre 84	Post 84
221,222	Ferrous Metals	Ferrous Metals
223	pt Fabr Metals	Ferrous Metals
	pt Ferrous Metals	
224	Non Ferrous Metals	Non Ferrous Metals
241-246	Building Materials	Building Materials
247,248	Glass and Ceramics	Glass and Ceramics
251	Pt Resins and Plastics	Pt Ind Chemicals
	Pt Ind and Ag Chem	Pt Ag Chemicals
255	Pt Consumer Chemicals	Pharmaceuticals
	Pt Ind and Ag Chemicals	
256	Pt Ind and Ag Chemicals	Pt Industrial Chemicals
	Pt Rubber	Pt Ag Chemicals
	Pt Consumer Chemicals	
257-9	Consumer Chemicals	Pharmaceuticals
260	Man Made Fibres	Man Made Fibres
311	pt Ferrous metals	Foundries
	pt Non Ferrous metals	
321	Ag Machinery	Ag Machinery
323	Textile Machinery	Ind Machinery
325	Contractor's Plant	Contractor's Plant
326	Other Mech Eng	Other Mech Eng
327	Ind & Off Machinery	Ind Machinery
329	Other Mech Eng	Other Mech Eng
330	Electronic	Office Machinery
341,342	Power and Industr Elect	Electrical Industrial
346	Electrical Consumer	Electrical Consumer
347	Power and Industr Elect	Electrical Industrial
351,352,353	Motor Vehicles	Motor Vehicles
361	Shipbuilding	Shipbuilding
362,363,364	Other Vehicles	Aerospace and Oth Vehicles
371,372,373,374	Instrument Engineering	Instrument Engineering
411-416,419,421,422	Food	Food
424,426-429	Drink and Tobacco	Drink and Tobacco
431	Woollen Textiles	Woollen Textiles
432-434	Spinning and Weaving	Spinning and Weaving
435	Other Textiles	Other Textiles
436	Hosiery	Hosiery
437	Other Textiles	Other Textiles
438	Other Textiles	Textile Consumer
441,442	Leather	Leather
451	Footwear	Footwear
453,456	Clothing and Fur	Clothing and Fur
455	Pt Furniture	Pt Textile Consumer
	Pt Other Textiles	Pt Other Textiles
461-464	Timber	Timber
466	Misc Manufactures	Timber
467	Pt Furniture	Furniture
	Pt Timber	

471	Paper	Pulp and Paper
472	Paper	Paper and Board
475	Printing	Printing
481,482	Rubber	Rubber
483	Resins and Plastics	Plastics
491	Fabricated Metal Gds	Other
492-494	Miscellaneous	Other

Chapter 6: Concluding Remarks

This thesis started by noting that there was a tendency for economists and politicians alike to imply that small firms were more dynamically efficient than large firms. The initial response to this claim was that there was relatively little hard evidence to support the claim and that therefore evidence would need to be provided in order to adjudicate on the claim. In this final chapter we will summarise the results that have been generated within the thesis and then reassess this basic question. We will then outline proposals for future research that can further help us to shed light on this question. In addition we will address a subsidiary question that has been exposed in the process of completing the research of the appropriate methodological stance to adopt in conducting research of this form.

The summary statistics presented in Chapter 1 pointed to the fact that the small firm sector has grown in relative importance in UK manufacturing industry over the past 25 years while there has been a relative decline in the importance of very large firms over the same period. At the same time however the performance, measured in terms of price cost margins, of the remaining large companies has increased steadily over the same period and at a more rapid rate than for the smaller firms although the small firms did also increase their performance steadily over the period. In terms of most of the cyclical variables both large and small firms appeared to follow each other closely in a procyclical fashion and the important differences were in terms of the secular decline in the reliance on the use of excess capacity and stocks on the part of the larger firms, which had been much greater in the 1970's. From this data there is therefore no clear evidence one way or the other on the relative dynamic efficiency of large or small firms.

In chapter 2 we assessed one of the most influential literatures that have been associated with the economic concept of small firms using dynamic efficiency as a means of competitive survival when confronted with larger firms with lower per unit costs. We demonstrated that there is an observational equivalence between this flexibility theory and the more conventional theory that firms performance varies according to unit costs. Indeed we showed that the crucial element driving the link between sales variability and market share is in fact differences in unit costs and not differences in the slope of the marginal cost curve. This finding clearly does not demonstrate that small firms are less flexible than large firms. Instead the finding merely casts doubt on the interpretation of some of the major economic tests of the flexibility theory suggesting an alternative interpretation that does not imply superior dynamic efficiency on the part of small firms.

Chapter 3 did not address the issue of the dynamic efficiency of small firms but instead suggested an empirical and theoretical respecification taking into account imperfect competition in the market for bought in inputs that is adopted in the analysis of chapters 4 and 5. Chapter 4 examined the short and long run determinants of the relative performance of large and small firms. We found that in the long run large firms do experience higher levels of performance and that this higher performance is particularly marked in industries where there is high advertising intensity and high level of concentration. This indicate that competitive pressures do play a part in lowering long run performance for small firms and that small firms are less able to overcome these competitive pressures using strategies such as advertising. In our sample, however, the short run performance of both large and small firms was not found to be strongly related to any of the conventional structural characteristics of industries.

In chapter 5 we explicitly focused on the dynamic and cyclical determinants of performance of the small and large firm sectors. If there were clear differences in the relative dynamic efficiency of large and small firms it is here that we would expect to find the most contrasting results between the different intra industry sectors. However as was noted in the conclusion to chapter 5 there is no such clear cut evidence. This is somewhat surprising given the strength of conviction with which many argue that small firms possess greater dynamic efficiency and certainly overturns prior beliefs. Instead we find evidence that if anything it was the larger firms which benefited from the boom in the late 1980's and that they appear to have retained this advantage through the recession that has followed at the start of the 1990's. The limited evidence from considering subsectors of industry indicate that it is for the large firms that demand growth and the inventory sales ratio are most important in certain sectors for the determination of performance. In the case of the former this greater response is as predicted by conventional short run profit maximising models of behaviour.

In summary we can therefore answer our original question by affirming that there is still relatively little hard evidence that small firms are more dynamically efficient than large firms. Indeed perhaps because we have cast an element of doubt on the validity of some of the previous evidence we may feel that we have less evidence of this proposition than before. Nevertheless we cannot claim to have categorically decided the issue and there is a good deal of scope for further tests that may be able to shed more light on the issue. We now turn to consider what kind of new directions might be usefully pursued.

We have indicated in chapter 2 that it is still an open empirical question as to whether small firms are indeed more flexible in terms of their cost structures than large firms. However, a potentially viable way forward in the attempt to find a more complete answer to the question has been presented in the form of an alternative empirical framework. Studies of individual industries that can obtain estimates of the short run cost functions of the firms or that examine the relative sales and production variability of individual product lines are the only way forward in answering this question that do not suffer from the observational equivalence problem that affects the previous studies. These therefore represent one aspect of future research. An additional issue that needs to be considered in this context is the role of inventories in facilitating production flexibility or production smoothing. While this topic has received much attention in general in the literature the number of studies that address the issue of firm size is limited.

In terms of the empirical analysis of the determinants of performance the discussion must clearly move onto new territory. In particular, the analysis of the business cycle has to move beyond the question of whether or not profitability is procyclical or countercyclical. The focus needs to be turned on to the specific determinants of price setting behaviour, both in terms of the different roles or inventories, capacity, order books, demand growth, input prices, etc. in the short run, and in terms of the historically contingent aspects of the analysis that relate to the longer term planning horizon such as modes of organisation of production, buyer-supplier relations, adoption of technology and, of course, the long term strategic interaction between rivals. However in addition to the role of inventories as determining performance and helping to determine the relative flexibility of large and small firms the study inventory usage in itself can help to enhance our understanding of the business cycle in relation to large and small firms.

Overall the scope for further investigation therefore lies in completing studies of individual industries, investigating the relative cyclical performance of the firms in those industries, and in switching the focus onto the relative ability of firms to absorb and adapt to demand fluctuations through their inventory usage.

Having addressed the primary question of this thesis we wish to now address a secondary focus; the methodological aspects of research and the validity of economic and econometric findings that had been generated both by previous research and the research contained within this thesis. The remaining part of this conclusion is devoted to remarks on this process of assessing the validity of economic arguments.

The methodological concern has coincided with, although it has not been driven by, a debate, if not a crisis, within the mainstream economics profession concerning the relationship between theory and empirical testing (see, for a recent example, the special issue of Journal of Econometrics 1995 Vol 67 No 1). A number of solutions have been presented as ways forward to closing the current gap that exists between theory and empirical analysis. We would argue that technocratic solutions to the problem are unlikely to be of much use. One such attempt to restore the status of econometric analysis, extreme bounds analysis developed by Leamer (1978) and applied to the concentration profits relationship by Cooley (1982), has not proved widespread as a tool[1]. The problem with all technocratic solutions lies in the realisation that, *whatever* techniques are used, findings are still susceptible to the prior beliefs and prejudices of the researcher[2]. The other extreme of reactions to the impasse is to resort, as McCloskey (1986) does, to rhetoric as the arbiter of science. Stettler (1995), with conscious irony, has described McCloskey's concept as a "market for ideas". McCloskey's implied call for a free market in ideas where the good ideas will prosper and the weak will fail is similarly problematic to more conventional notions of free markets. As in most markets the exploitation of first mover advantages, consumer loyalty and high switching costs are likely to lead to the market dominance of established views. From this perspective McCloskey's suggestion of persuasion or rhetoric as the standard of ideas seems nothing more than an argument for the perpetuation of the status quo, i.e. the neo-classical orthodoxy. As descriptive reality it may illustrate the problems associated with closing the gap between theory and empirics in what, to labour the analogy, may be considered a segmented market. However, it does not offer either a way forward or a prescription of method[3].

A third approach to the gulf between theory and empirical reality, one that is ultimately pessimistic, suggests that the theoretical and empirical modes are essentially couched in different modes of thinking. Empiricists or econometricians are likely to be structuralists by nature. Theorists tend to be functionalists. This would appear to be especially true of game theorists but is also true of most theoretical economists. By positing a rational individual a functionalist perspective is forced on to all economic phenomena because all economic phenomena can be (or ought to be according to this logic) potentially explained as the outcome of the interaction of rational economic agents[4]. However, maintaining a distance from the impositions of theory

[1] Technocratic solutions are clearly not immune to criticism from within their own framework, for such criticism of Leamer, and implicitly Cooley, see McAleer *et al* (1985) and Spanos (1986).
[2] Gould (1981) reports examples of *unconscious* data manipulation by scientists that biases results in favour of priors.
[3] We would argue that the rigid application of methodology is to be feared as much as the absence or ignorance of method. Hendry (1995 p5) suggests in his prescription for good econometric methodology that one should 'think brilliantly', 'be creative' or if that and luck fail, stick to theory.
[4] For a clear and interesting example of how functionalism can prove less than illuminating by imposing an unsuitable function on (in this example non economic) phenomena and how a structural approach can prove more amenable, see Gould (1995).

enables analysis of the data generating processes in a way which may be unexplained by theory and could at least in principle be unexplainable by credible theory. This points to the other side of functionalism, there is a desire for completeness, which is only achieved by ascribing a function or motive to each action, yet in doing so *all* action becomes sensible.

Geroski (1990a) cites Gould (1987) advocating a "consilience of induction" which is based upon the idea of many different tests which point to a particular conclusion. However, this would appear to be valid only as an *ex post* justification of methodology, rather than a starting point, as the validity of such an approach could only reject claims of verificationism if the tests are independent. But how can one set out *ex ante* to consider tests which are defined to be independent because such tests are, of necessity, constrained by the view of the researcher. Eclecticism does not constitute a way through to truth when consensus is determined within existing power relations.

The conclusion from this discussion is that all we may hope for is to work towards and try to maintain an effective balance between theory and empirical analysis. We also need to be aware that this balance is ever shifting. Progressive developments in the two elements are neither uniform in magnitude nor linear in direction. Changes in the questions of interest generated from the politico-economic situation and developments in the tool kits of both theorists and applied economists occur at different rates and have different rates of diffusion. We would like to imagine that there are forces, such as the theorists need for empirical support of a hypothesis, and the empirical economists need to place the search for a description of the 'data generating process' within the context of some hypothesis concerning the nature of that process, tending towards balance in this relationship. However the increasing segmentation of economics into sub disciplines and the strength of resistance to any change in the basic tenets of the neo-classical core of the subject may indicate that these balancing forces are perhaps weaker now than they were twenty years ago.

Bibliography

Aaronovitch, S., 1955, Monopoly: A study of British monopoly capitalism, Lawrence and Wishart.

Akerlof, G. and J. Yellen, 1985, A near-rational model of the business cycle, with wage and price inertia, Quarterly Journal of Economics, 100, 823-838

Amihud, Y. and H. Mendelson, 1983, Price smoothing and inventory, Review of Economic Studies, 50, 87-98

Amihud, Y. and H. Mendelson, 1989, Inventory behaviour and market power: an empirical investigation, International Journal of Industrial Organization, 7, 269-280

Anderson, T., and C. Hsiao, 1982, Formulation and estimation of dynamic models using panel data, Journal of Econometrics, 18,47-82

Arrellano, M. and S. Bond, 1988, Dynamic panel data estimation using DPD: A guide for users, IFS working paper 88/15

Arrellano, M. and S. Bond, 1991, Some tests of specification for panel data: Monte Carlo evidence and application to employment equations, Review of Economic Studies, 58, 277-297

Auerbach, P., 1988, Competition, Blackwell

Axelrod, R., 1984, Evolution of Cooperation, Penguin

Bain, J.S., 1951, Relation of profit rate to industry concentration: American Manufacturing 1936-1940 Quarterly Journal of Economics, 65, 293-324.

Bain, J.S., 1956, Barriers to New Competition Harvard University Press

Baldwin, J., P. Gorecki and J McVey, 1984, Imports, secondary output, price cost margins and measures of concentration: Evidence for Canada, Discussion Paper 263, Economic Council of Canada.

Balestra, P. and M. Nerlove, 1966, Pooling cross-section and time-series data in the estimation of a dynamic model: the demand for natural gas, Econometrica, 34, 585-612

Baltagi, B., 1995, Econometric Analysis of Panel Data, John Wiley

Baltagi, B., and J. Griffin, 1984, Short and long run effects in pooled models, International Economic Review, 25, 631-645

Benoit, J-P. and V. Krishna, 1987, Dynamic duopoly: prices and quantities, Review of Economic Studies, 54, 23-35

Berg, A., 1986, Excess capacity and the degree of collusion: some estimates for the Nowegian manufacturing sector, International Journal of Industrial Organization, 4, 99-108

Best, M., 1990, The New Competition: institutions of industrial restructuring Polity Press

Bils, M., 1987, The cyclical behaviour of marginal cost and price, American Economic Review, 77, 838-855

Blanchard, O., 1983, The production and inventory behaviour of the American automobile industry, Journal of Political Economy, 91, 365-401

Blundell, R., S. Bond, M. Devereux and F. Schiantarelli, 1992, Investment and Tobin's Q: evidence from company panel data, Journal of Econometrics, 51, 233-257

Bolton, 1971, Small Firms: report of the committee of inquiry on small firms, Cm4811, HMSO

Bradburd R.M.and D.R.Ross, 1989, Can small firms find and defend strategic niches? A test of the Porter hypothesis, Review of Economics and Statistics, 71, 258-262.

Bresnahan, T., 1981, Duopoly models with consistent conjectures, American Economic Review, 71, 934-945.

Bresnahan, T., 1982, The oligopoly solution concept is identified, Economics Letters, 10, 87-92.

Bresnahan, T., 1989, Empirical studies of industries with market power, Chapter 17 in Schmalensee, R and R Willig Handbook of Industrial Organization Vol II, North Holland

Bresnahan, T. and V. Suslow, 1989, Short run supply with capacity constraints, Journal of Law and Economics, 32, S11-42

Bulow, J., J. Geanakopolous and P. Klemperer, 1985, Multimarket oligopoly: strategic substitutes and complements, Journal of Political Economy, 93, 488-511

Carlsson, B., 1989, Flexibility and the theory of the firm, International Journal of Industrial Organisation, 7, 179-203

Carlton, D., 1989, The theory and facts of how markets clear: is industrial organization valuable for understanding macroeconomics?, Chapter 15 in R. Schmalensee and R.Willig (eds), Handbook of Industrial Organization Vol I, North Holland

Caves, R., and M. Porter, 1977, From entry barriers to mobility barriers, Quarterly Journal Of Economics, 91, 241-261.

Caves, R., and M. Porter, 1978, Market structure, oligopoly and stability of market shares, Journal of Industrial Economics, 26, 289-313.

Caves, R and T. Pugel, 1980, Intraindustry differences in conduct and performance: viable strategies in US manufacturing industries, New York University Monograph Series in Finance and Economics 1980-2.

Chappell, W., B.Das, and W Shughart, 1993, Demand Fluctuations and Firm Heterogeneity, Journal of Industrial Economics, 41, 51-60.

Chappell, W., W. Mayer and W. Shughart, 1993, Firm heterogeneity and production flexibility: evidence from price cost margins of large and small firms, Bulletin of Economics Research, 45, 229-244.

Clarke, R., and S. Davies, 1982, Market Structure and price cost margins, Economica, 49, 277-287

Clarke, R., S. Davies and M. Waterson, 1984, The profitability concentration relation: market power or efficiency?, Journal of Industrial Economics, 32, 435-450

Conyon, M., 1995, Industry Profit Margins and Concentration: Evidence from UK Manufacturing, International Review of Applied Economics, 9, 275-290.

Conyon, M. and S. Machin, 1991a, The determination of profit margins in UK manufacturing, Journal of Industrial Economics, 39, 369-383

Conyon, M. and S. Machin, 1991b, Market Structure and the empirical specification of profit margins, Economics Letters, 35, 227-231

Cool, K and I. Dierickx, 1993, Rivalry, strategic groups and firm profitability, Strategic Management Journal, 14, 47-59.

Cooley, T., 1982, Specification analysis with discriminating *priors: an application* to the concentration profits debate, Econometric Reviews, 1, 97-128

Coutts, K., W. Godley, and W. Nordhaus, 1978, Industrial Pricing in the UK Cambridge University Press

Cowling, K., 1982, Monopoly Capitalism Macmillan

Cowling, K., 1983, Excess capacity and the degree of collusion: oligopoly behaviour in the slump, Manchester School, 51, 341-359

Cowling, K. and M. Waterson, 1976, Price cost margins and market structure, Economica, 43, 267-274

Crampton, P. and T. Palfrey, 1990, Cartel enforcement with uncertainty about costs, International Economic Review, 31, 17-47

Cubbin, J., and P. Geroski., 1990, The persistence of profits in the United Kingdom, Chapter 9 in D. Mueller (ed) The Dynamics of Company Profits, Cambridge University Press

Davidson, C. and R. Deneckere, 1990, Excess capacity and collusion, International Economic Review, 31, 521-542

Del Monte, A and F. Esposito, 1992, Flexibility and industrial organisation theory, in A Del Monte Recent Developments in the Theory of Industrial Organisation, Macmillan.

Demange, G., and J-P. Ponssard, 1985, Asymmetries in cost structures and incentives towards price competition, International Journal of Industrial Organization, 3, 85-100

Demsetz, H., 1973, Industry structure, market rivalry and public policy, Journal of Law and Economics, 16, 1-9

Dickson, V. 1991, The relationship between concentration and prices and concentration and costs, Applied Economics, 23, 101-106

Dickson, V., 1994, The empirical specification of price cost margins, Economics Letters, 44, 329-332

Domberger, S., 1983, Industrial Structure, Pricing and Inflation, Martin Robertson

Domowitz, I., Hubbard R.G. and Petersen, B., 1986, Business cycles and the relationship between concentration and price cost margins, Rand Journal of Economics, 17, 1-17

Domowitz, I., Hubbard R.G. and Petersen, B., 1987, The intertemporal stability of the concentration margins relationship, Journal of Industrial Economics, 35, 13-34

Domowitz, I., Hubbard R.G. and Petersen, B., 1988, *Market structure* and cyclical fluctuations in US manufacturing, Review of Economics and Statistics, 70, 55-66

Dowrick, S., 1990, Wage pressure, bargaining and price cost margins in UK manufacturing, Journal of Industrial Economics, 38, 239-267

Dunne, P and A. Hughes, 1992, The changing structure of competitive industry in the 1980's, Chapter 4 in Driver, C and P. Dunne (eds), Structural Change in the UK Economy Cambridge University Press

Encaoua, D., and A. Jacquemin, 1980, Degree of monopoly, indices of concentration, and the threat of entry, International Economic Review, 21, 87-105.

Fisher, F., 1987, On the misuse of the profits-sales ratio to infer monopoly power, Rand Journal of Economics, 18, 384-396.

Forchheimer, K., 1908, Theoretiches zum umvollständingen monopole, Schmollers Jahrbuch, 1-12

Friedman, M., 1962, Price Theory: a provisional text, Frank Cass

Fudenberg D.and J. Tirole, 1984, The fat cat effect, the puppy dog ploy and the lean and hungry look, American Economic Review, 74, 361-368.

Geroski, P., 1988, In pursuit of monopoly power: recent quantitative work in industrial economics, Journal of Applied Econometrics, 3, 107-123.

Geroski, P., 1990a, Modelling persistent profitability, Chapter 2 in D. Mueller (ed) The Dynamics of Company Profits, Cambridge University Press

Geroski, P., 1990b, Market Dynamics and Entry, Oxford University Press

Geroski, P., 1992, Price dynamics in UK manufacturing: a microeconomic view, Economica, 59, 403-419

Geroski, P., 1994, Market Structure, Corporate Performance, and Innovative Activity, Oxford University Press

Geroski, P., 1995, What do we know about entry?, International Journal of Industrial Organization, 13, 421-440

Geroski, P., S. Machin and J. Van Reenen, 1993, The profitability of innovating firms, Rand Journal Of Economics, 24, 198-211.

Geroski, P., and R. Masson, 1987, Dynamic market models in industrial organization, International Journal of Industrial Organization, 5, 1-14

Gertler, M., and S. Gilchrist, 1991, Monetary policy, business fluctuations and the behaviour of small manufacturing firms, mimeo

Gibbons, R., 1992, A Primer in Game Theory, Harvester Wheatsheaf

Gilbert, R., and X. Vives, 1986, Entry deterrence and the free rider problem, Review of Economic Studies, 53, 71-83

Godley, W., and W. Nordhaus, 1972, Industrial Pricing in the UK, Economic Journal, 82, 853-882

Gould, S.J., 1981, The Mismeasure of Man, Penguin

Gould, S.J., 1987, Wonderful Life, Penguin

Gould, S.J., 1995, Male nipples and clitoral ripples, in Adam's Navel, Penguin

Green E., and R.Porter, 1984, Noncooperative collusion under imperfect price information, Econometrica, 52, 87-100

Green, F., 1992, Recent trends in British trade union density: how much a compositional effect?, British Journal of Industrial Relations, 30, 445-458.

Greene, W., 1993, Econometric Analysis, 2nd edition, Prentice Hall.

Griliches, Z., and J. Hausman, 1986, Errors in variables in panel data, Journal of Econometrics, 31, 93-118

Guariglia, A. and F. Schiantarelli, 1994, Production smoothing, firms' heterogeneity and financial constraints: Evidence from a panel of UK firms, Boston University mimeo

Haltiwanger J. and J. Harrington, 1991, The impact of cyclical demand movements on collusive behaviour, Rand Journal of Economics, 22, 89-105

Harrington, J., 1989, Collusion among asymmetric firms, International Journal of Industrial Organization, 7, 289-307

Harrod, R.F., 1965[1936], The Trade Cycle, A.M. Kelley

Hart, P. and E. Morgan, 1977, Market structure and economic performance in the United Kingdom, Journal of Industrial Economics, 25, 177-193

Haskel, J. and C. Martin, 1992, Margins, concentration, unions and the business cycle, International Journal of Industrial Organization, 10, 611-632

Haskel, J. and C. Martin, 1994, Capacity and Competition, Journal of Industrial Economics, 42, 23-44

Hay, D. and D. Morris, 1991, Industrial Economics and Organization: Theory and Evidence, 2nd Edition, Oxford University Press.

Hendry, D., 1995, Dynamic Econometrics Oxford University Press

Henley, A., 1994, Industrial deconcentration in the UK, Manchester School, 62, 40-59

Hirst, P., and J. Zeitlin, 1988, Reversing Industrial Decline, Berg

Holt, C., Modigliani, F., Muth, J. and H. Simon, 1960, Planning Production, Inventories, and Work Force, Prentice Hall.

Hsiao, C., 1986, The Analysis of Panel Data, Cambridge University Press

Kalai, E. and W. Stanford, 1985, Conjectural variations strategies in accelerated cournot games, International Journal of Industrial Organization, 3, 133-152

Kandori, M., 1991, Correlated demand shocks and price wars during booms, Review of Economic Studies, 58, 171-180

Kalecki, M., 1939, Essays on the Theory of Economic Fluctuations Allen Unwin

Kay, J. and C. Mayer, 198, On the application of accounting rates of return, Economic Journal, 96, 199-207.

Kessides, I., 1990, The persistence of profits in U.S. manufacturing industries, Chapter 4 in D. Mueller (ed) The Dynamics of Company Profits, Cambridge University Press

Keuzenkamp, H., and M. McAleer, 1995, Simplicity, scientific inference and econometric modelling, Economic Journal, 105, 1-21

Kihlstrom R. and X. Vives, 1992, Collusion by asymetrically informed firms, Journal of Economics and Management Strategy, 1, 371-396

Kirzner, I., 1973, Competition and Entrepreneurship, University of Chicago Press

Klein B.H., 1984, Prices, wages and business cycles: A dynamic theory, Pergamon.

Koopmans, T.C., 1950, When is an equation system complete for statistical purposes? Chapter 17 in T.C. Koopmans (ed) Statistical Inference in Dynamic Economic Models John Wiley

Kreps, D., and J. Scheinkman, 1983, Quantity pre-commitment and *Bertrand competition* yield Cournot Outcomes, Rand Journal of Ecnomics, 14, 326-337.

Kreps, D., P.Milgrom, J. Roberts and R. Wilson, 1982, Rational cooperation in the finitely repeated prisoner's dilemma, Journal of Economic Theory, 27, 245-252

Kwoka, J.E. and D. Ravenscraft, 1986, Cooperation v rivalry: Price cost margins by line of business, Economica, 53, 351-363

Leamer, E., 1978, Specification Searches, Wiley

Levy, D., 1987, The speed of the invisible hand, International Journal of Industrial Organization, 5, 79-92

Lindbeck, A. and D. Snower, 1994, Price Inertia and Production Lags, Seminar Paper 494, Institute for International Economic Studies, University of Stockholm

Luxemburg, R., 1970[1900], Reform or Revolution, in Rosa Luxemburg Speaks, Pathfinder

McAleer, M., A. Pagan and P. Volker, 1985, What will take the con out of econometrics?, American Economic Review, 75, 293-307

McCloskey, D., 1986, The Rhetoric of Economics, Wheatsheaf

Machin, S., 1991, Unions and the capture of economic rents: An investigation usng British firm level data, International Journal of Industrial Organization, 9, 261-274

Machin, S., and J. Van Reenen, 1993, Profit margins and the buisness cycle: evidence from UK manufacturing firms, Journal of Industrial Economics, 41, 29-50.

Mairesse, J., 1990, Time series and cross sectional estimates using panel data: why are they different and why should they be equal?, in Hartog, J., G. Ridder. and J Theeuwes, Panel Data and Labor Market Studies, North Holland

Marschak, T., and R. Nelson, 1962, Flexibility, uncertainty and economic theory, Metroeconomica, 14, 104-107.

Marx, K., 1981[1894], Capital Volume 3, Penguin

Milgrom, P. and J. Roberts, 1990, The economics of modern manufacturing: technology, strategy and organisation, American Economic Review, 41, 511-528.

Mills, D., 1984, Demand fluctuations and endogenous firm flexibility, Journal of Industrial Economics, 33, 55-71.

Mills, D., 1986, Flexibility and firm diversity with demand fluctuations, International Journal of Industrial Organisation, 4, 302-315.

Mills, D., and L. Schumann, 1985, Industry structure with fluctuating demand, American Economic Review, 75, 758-767.

Millward, N. and M. Stevens, 1986, British Workplace Industrial Relations 1980-1984 Gower

Mueller, D., 1986, Profits in the Long Run, Cambridge University Press

Mueller, D., 1990, The Dynamics of Company Profits, Cambridge University Press

Mundlak, Y., 1978, On the pooling of time series and cross section data, Econometrica, 46, 69-85.

Neumann, M., I. Böbel, and A. Haid. 1983, Business cycle and industrial market power: an empirical investigation for West German industries, 1965-1977, Journal of Industrial Economics, 32, 187-196

Nickell, S., 1981, Biases in dynamic models with fixed effects, Econometrica, 49, 1417-1426

Odagiri, H. and T. Yamashita, 1987, Price mark-ups, market structure, and business fluctuations in Japanese manufacturing industries, Journal of Industrial Economics, 35, 317-331

Oi, W., 1982, Heterogeneous firms and the organization of production, Economic Inquiry, 21, 147-171

Osborne, M., and C. Pitchik, 1987, Cartels, profits and excess capacity, International Economic Review, 28, 413-428

Pesaran M.H. and R. Smith, 1995, Estimating long run relationships from dynamic heterogeneous panels, Journal of Econometrics, 68, 79-113.

Piore M., and C. Sabel, 1984, The Second Industrial Divide, Basic Books

Porter, M., 1979, The structure within industries and companies performance, Review of Economics and Statistics, 61, 214-227

Powell, T., 1996, How much does industry matter? An alternative empirical test, Strategic Management Journal, 17, 323-334.

Ramey, V., 1991, Nonconvex costs and the behaviour of inventories, Journal of Political Economy, 99,306-334

Rees, R., 1993, Collusive equilibrium in the great salt duopoly, Economic Journal, 103, 833-848

Reger R. and A Huff, 1993, Strategic groups: a cognitive perspective, Strategic Management Journal, 14, 103-124.

Roberts, K., 1985, Cartel behaviour and adverse selection, Journal of Industrial Economics, 33, 401-413.

Roberts, M., 1984, Testing oligopolistic behaviour, International Journal of Industrial Organization, 2, 367-383.

Rotemberg, J. and G. Saloner, 1986, A supergame theoretic model of price wars during booms, American Economic Review, 76, 390-407

Rotemberg, J. and G. Saloner, 1989, The cyclical behaviour of strategic inventories, Quarterly Journal of Economics, 104, 73-98

Rotemberg, J. and M. Woodford, 1993, Dynamic general equilibrium models with imperfectly competitive product markets, mimeo

Rowthorn, R., and T. Wells, 1987, Deindustrialization and Foreign Trade, Cambridge University Press

Rumelt, R., 1991, How much does industry matter?, Strategic Management Journal, 12, 167-185

Salinger, M., 1990, The concentration-margins relationship revisited, Brookings Papers in Economic Activity: Microeconomics 287-335

Sargan, J., 1958, The estimation of economic relationships using instrumental variables, Econometrica, 26, 393-415

Saving T., 1970, Concentration ratios and the degree of monopoly, International Economic Review, 11, 139-146

Sawyer, M., 1983, Business Pricing and Inflation, Macmillan

Scherer, F.M., A. Beckenstein, E. Kaufer and R. Murphy, 1975, Economics of Multi-Plant Operation: An international comparisons study, Harvard University Press

Scherer F.M., 1980, Industrial Market Structure and Economic Performance 2nd Edition Rand McNally

Scherer, F.M. and D. Ross, 1990, Industrial Market Structure and Economic Performance, 3rd Edition, Houghton Mifflin.

Schmalensee R., 1978, Entry deterence in the ready-to-eat breakfast cereal industry, Bell Journal of Economics, 9, 305-327

Schmalensee, R., 1985, Do markets differ much?, American Economic Review, 75, 341-351

Schmalensee R., 1987, Competitive advantage and collusive optima, International Journal of Industrial Organization, 5, 351-367

Schmalensee, R., 1989a, Inter industry studies of structure and performance, Chapter 16 in Schmalensee, R and R Willig Handbook of Industrial Organization Vol II, North Holland

Schmalensee, R., 1989b, Intra Industry Profitability Differences in US Manufacturing, Journal of Industrial Economics, 37, 337-357.

Schmalensee, R., 1992, Sunk Cost and Market Structure: a review article, Journal of Industrial Economics, 40, 125-137.

Shaked, A and J. Sutton, 1987, Product differentiation and industrial structure, Journal of Industrial Economics, 36, 131-146

Singh, S., M. Utton and M. Waterson, 1995, Strategic behaviour of incumbent firms in the UK, mimeo

Slade, M., 1986, Conjectures, firm characteristics and market structure, International Journal of Industrial Organization, 4, 347-369

Slade, M., 1989, Price wars in price setting supergames, Economica, 56, 295-310

Slade, M., 1992, Vancouver's gasoline-price wars:an empirical exercise in uncovering supergame strategies, Review of Economic Studies, 59, 257-276

Slade, M., 1995a, Product rivalry with multiple strategic weapons: an analysis of price and advertising competition, Journal of Economics and Management Strategy, 4, 445-76

Slade, M., 1995b, Empirical games: the oligopoly case, Canadian Journal of Economics, 28, 368-402

Spanos, A., 1986, Statistical Foundations of Econometric Modelling, Cambridge University Press

Spence, M., 1977, Entry, capacity, investment and oligopolistic pricing, Bell Journal of Economics, 8, 534-544

Staiger R. and F. Wolak, 1992, Collusive pricing with capacity constraints in the presence of demand uncertainty, Rand Journal of Economics, 23, 203-220

Steedman, I., 1992, Questions for Kaleckians, Review of Political Economy, 4, 125-151

Steindl, J., 1945, Small and Big Business, Blackwell

Steindl, J., 1952, Maturity and Stagnation in American Capitalism Blackwell

Stettler, M., 1995, The rhetoric of McCloskey's rhetoric of economics Cambridge Journal of Economics, 19, 391-403

Stigler, G., 1939, Production and distribution in the short run, Journal of Political Economy, 47, 305-327

Stigler, G., 1964, A Theory of oligopoly, Journal of Political Economy, 72, 44-61

Stigler, G., 1968, The Organization of Industry, Irwin Dorsey

Suslow, V., 1986, Estimating monopoly behaviour with competitive recycling, Rand Journal Of Economics, 17, 389-403

Sutton, J., 1990, Explaining everything, explaining nothing: game theoretic models in industrial economics, European Economic Review, 34, 505-512

Sutton, J., 1991, Sunk Costs and Market Structure, MIT Press

Wachtel, H., and P. Adelsheim, 1977, How recession feeds inflation: price markups in a concentrated economy, Challenge, 20, 6-13

Waterson, M., 1980, Price Cost Margins and Successive Market Power, Quarterly Journal of Economics, 94, 135-150

Waterson, M., 1984, Economic Theory of the Industry, Cambridge University Press

Waterson, M., 1989, Differential efficiency, profitability and competition policy, University of Reading mimeo

Williamson, O., 1970, Corporate Control and Business Behaviour: an inquiry into the effects of organizational forms and enterprise behaviour, Prentice Hall

Yamawaki, H., 1985, Dominant firm pricing and fringe expansion: The case of U.S. iron and steel industry, 1907-1930, Review of Economics and Statistics, 67, 429-437.

You, J.I., 1995, Small firms in economic theory, Cambridge Journal of Economics, 19, 441-462

Zimmermann, K.F., 1995, Flexibility in the face of demand fluctuations: Employment, capacity utilisation and industry structure, International Journal of Industrial Organization, 13, 179-193