ASYMMETRIC INDUSTRY STRUCTURES: MULTIPLE TECHNOLOGIES, FIRM DYNAMICS AND PROFITABILITY.

by

Reinout Michael Koopmans

London School of Economics and Political Science

A Thesis Submitted for the Degree of Doctor of Philosophy

of the

University of London

October 1995

ABSTRACT

Asymmetric Industry Structures: Multiple Technologies, Firm Dynamics and Profitability

The origins of asymmetric firm sizes are analyzed in the first part of this thesis, modelling technology choice in a one shot quantity game with homogeneous goods. For certain sizes of the market more than one technology is chosen in equilibrium. Generally, the larger the market, the higher the fixed cost of the technology that is chosen in equilibrium. The trade off between market size and concentration is non-monotonic, even if for any size of the market only the most fragmented market structure is considered.

In the second part consequences of asymmetric firm sizes are investigated. In Chapter 3 firm dynamics in the chemical sector are examined, distinguishing between the dynamics of scale and scope. The production capacity of firms in homogeneous bulk chemical markets *converges* in size on a market by market basis, resulting in a fragmented industry structure on a disaggregated level. However, the number of products chemical corporations produce within a category of (synthetic organic) chemicals *diverges*, leading to a more concentrated industry structure on higher levels of aggregation. These counteracting forces can potentially explain the persistence of concentration that has been observed in fast growing chemical markets.

In Chapter 4 it is shown that if the observed asymmetry between firms is consistent with a (subgame-perfect) equilibrium of some single or multi-stage game, bounds exist that restrict the degree of asymmetry between the firms' profitability. Their shape is determined by industry factors. In particular, a higher sensitivity of a firm's profitability on its competitor's action rotates the bounds on the profitability-size trade off anti-clockwise. This is tested for homogeneous goods industries using a panel from the FTC Line of Business Data. Allowing for firm specific fixed effects, some strong empirical support is found.

ACKNOWLEDGEMENTS

During the preparation of this thesis I have enormously benefitted from the intellectual guidance of my supervisor Professor John Sutton. His methodological arguments and research agenda have deeply influenced my view on economics, which is reflected throughout this thesis.

Many of the ideas originate or were refined and pushed forward in almost daily discussions with Alison Hole. She read the manuscript more than once and in various stages of its creation. I am deeply indebted to her.

I am very grateful to Professor David Webb for providing me shelter in the Financial Markets Group for four years. The intellectual environment in the Group is widely known to be uniquely stimulating. I am extremely thankful to all my colleagues there, past and present. I benefitted particularly from discussions with and comments by Mike Burkart, Denis Gromb and Godfrey Keller.

The empirical part of this thesis would never have been written without extensive discussions with Ana Lamo and Joep Konings. Parts of Chapter 3 have been written in collaboration with Ana Lamo. I am very grateful to Marvin Lieberman to making the data available the us. Joep Konings, George Symeonidis and Hugh Wills provided useful comments on the empirical part of Chapter 4. Curtis Wagner and George Pascoe at the Federal Trade Commission in Washington have been extremely helpful with the analysis of the Line of Business Data.

Financial assistance from the Foreign Office (FCO), through a Chevening Scholarship, is gratefully acknowledged. So is the financial support from the Financial Markets Group and the Stichting Vreedefonds. The empirical work in Chapter 4 would not have been possible without financial support from the LSE Economics Department and The Royal Economic Society.

Finally I am deeply grateful for support and stimulus that I have received from friends and family, in particular Donata. Their understanding, patience and sympathy have been a great source of strength throughout.

TABLE OF CONTENTS

LIST OF TABLES.

LIST OF FIGURES.

CHAPTER 1

INTRODUCTION.

One of the foremost features of industrial market structures is the considerable heterogeneity that exists among firm characteristics within narrowly defined industries. The literature has adopted a range of approaches to modelling these asymmetries. It ranges from heterogeneity being generated by purely stochastic realizations of firm specific state variables, or 'historical accidents', to deterministic models in which firms operate in a strategic environment with asymmetric equilibrium industry structures.

In the literature most models have both strategic and stochastic elements playing a role, as for example in the theoretical literature on stochastic firm dynamics (Ericson and Pakes (1989), Hopenhayen (1989) and Budd, Harris and Vickers (1993)) or in the literature on (stochastic) patent races (Reinganum (1982), Harris and Vickers (1987) and Katz and Shapiro (1987)). In Lambson (1991) heterogeneity among ex-ante identical firms is explained by the choice of different production technologies, as there is no prior knowledge of future market conditions. In equilibrium firms place different 'bets' on which technology will be the most efficient after the next change in market conditions.

The archetypal example of a purely stochastic model of industry structure is Gibrat's Law, otherwise known as the Law of Proportionate Effect. Gibrat (1931) showed that the distribution of sizes of business firms for narrowly defined industries is typically skewed to the right and close to a lognormal distribution function. He suggests that this might be a consequence of firm growth rates being *independent* of firm sizes. The limiting size distribution of a fixed population of firms is lognormal if the growth rates firms experience over time are independent of their actual size[1,2].

---

[1] Later various alternative stochastic processes have been considered. Ijiri and Simon (1977) show that if growth rates are independent of current size, but there is a fixed probability that an entrant will capture the investment opportunity, the time stationary size distribution is a Yule distribution. Sutton (1995a) shows in this context that if not the growth rate, but absolute growth is independent of current size, an exponential limiting distribution results.

The archetypal example of deterministic models with asymmetric equilibria is a Grab the Dollar game. There are two players and one dollar is on the table. Each of the players can grab the dollar (G) or refrain from grabbing (R). If only one player grabs, he takes it, if both grab, they bang their heads and get a negative pay off, say -1.

Table 1.1 The Grab the Dollar Game.

		Player 2	
		G2	**R2**
Player 1	**G1**	-1, -1	1, 0
	R1	0, 1	0, 0

There are two *asymmetric* pure strategy Nash equilibria ({G1, R2} and {R1, G2}), that are observationally equivalent. There is also a mixed strategy equilibrium in which firms behave strategically, though the outcome is stochastic[3].

Many models that discuss the origins of asymmetry in a strategic environment assume some kind of initial asymmetry among the firms, which drives the asymmetric outcome. This usually refers to differences in technological conditions or strategic asymmetries. The standard textbook treatment of the former is to assume a historically given difference in cost structure, which drives the difference in the

---

[2] Subsequently, Hart and Prais (1956) test for the lognormality of the firm size distribution economy wide, by using those firms quoted on the London Stock Exchange. Simon and Bonini (1958) test if the Yule distribution yields an appropriate description of the firm size distribution. Mansfield (1962) analyzes the size distribution for steel, petroleum and tires and tests for lognormality. Silberman (1967) is the most comprehensive of its kind. He analyzes the seller distribution in 90 four-digit S.I.C. industries and concludes that although the lognormal hypothesis cannot be rejected for specific industries, it is inappropriate to consider the function as a generalized statistical summary measure of industry size distributions in US manufacturing. More recently Cohen and Klepper (1992) analyzed the cross-section distribution of R&D expenses in a number of industries on 2 digit level. They also found highly skewed distributions.

[3] The equilibrium strategies are both firms randomizing with equal probability.

way firms operate, their size and their response to external shocks[4]. It can, however, also be the case that an initial technological advantage puts a firm in a permanently better position through for example scale economies, learning or product proliferation, first mover advantages or more general, strategic asymmetries[5]. However, assuming ex-ante asymmetry, whether it is technological or strategic, does not address the fundamental issue, which is the *source* of the heterogeneity. It only shows which ex-post asymmetries are driven by which ex-ante asymmetries. The asymmetry itself is exogenous and remains as such unexplained[6].

From a methodological point of view assuming ex-ante asymmetry can be interpreted as a problem of missing stages in the extensive form of the game. Any satisfactory explanation of the sources of firm heterogeneity would start with ex-ante symmetric firms, at least in some initial stage of the industry evolution[7]. This might well be long before the firms start their operations. Strategic games with ex-ante identical players and asymmetric equilibria usually have multiple equilibria, reflecting the changing roles players can have in equilibrium, though the outcomes might well be observationally equivalent.

---

[4] The textbook example is a Cournot model with differences in (marginal) costs between firms. See for example Scherer and Ross (1990, Appendix to Chapter 6), and Tirole (1989, Chapter 5). Under perfect competition and technological differences among firms, asymmetric firm sizes might also exist. See Demsetz (1973). In for example Hopenhayen (1989) or Jovanovic and MacDonald (1993), firm asymmetry originates in (unexplained) exogenous cost shocks.

[5] The textbook example of strategic asymmetries is the von Stackelberg model. See d'Aspremont et al. (1983) for a model of cartel stability based on strategic asymmetries (price-leadership). Gilbert and Vives (1986) analyze entry deterrence if incumbents have a first mover advantage. The winner in Harris and Vickers (1985) model of a patent race in the absence of uncertainty and ex-ante identical firms is determined by a first mover advantage. The classic paper on asymmetry through product proliferation is Schmalensee (1978). There is an extensive literature on strategic asymmetries through learning economies. See Cabral and Riordan (1994) and the references therein.

[6] Note that in the Grab the Dollar Game the players are ex-ante identical.

[7] There are a number of game theoretic models of industry structure with no firm level uncertainty in which asymmetric equilibria exist, although the firms are ex-ante identical. For example Shaked and Sutton (1983, 1990), Rosen (1991), Lambson (1991) and d'Aspremont et al. (1983).

This thesis analyzes aspects of asymmetric industry structures both in terms of origin and consequences. The chapters are self-contained and do not necessarily follow on logically from previous ones. Here we present some common themes and show how they can be seen against the background framework that was described above.

One theme that runs through this thesis is the extent to which strategic considerations are important in the explanation of the existence and persistence of firm size inequalities, keeping in mind, though often implicitly, that a purely stochastic model of the Gibrat-type is the principal alternative hypothesis.

A second theme that occurs in Chapters 2 and 3 is the effect of market growth on market structure, in particular with respect to firm asymmetries. Many theoretical models predict a negative relationship between market size and concentration for a given market and technology, suggesting that large firms are at a disadvantage vis à vis small firms or entrants in capturing new market opportunities. Empirically this relationship was found to be not necessarily negative, as for example in Nelson (1963) and Shepherd (1964). Table 1.2 illustrates this point in the context of the chemical sector, which is presented here primarily for later reference. Industries that faced substantial increases in real sales, such as Pharmaceutical Preparations and Surface Active Agents, saw only minimal changes in concentration. Others, such as Industrial Gases and Paints & Allied Products, even faced in increase in concentration although the market size expanded substantially. However, there seems widespread evidence that the relationship is negative if changes in MES are controlled for, as then only the pure entry effect is picked up[8]. Klepper and Graddy (1990) on the other hand, found that in a wide range of product markets the number of producers tends to rise, peak and fall (the "shakeout") to stabilize thereafter at a lower level, though industry sales continue to rise throughout.

The question arises which of these phenomena are driven by changes in the production technology and which are due to strategic advantages of either large or small firms during the growth phase. In Chapter 2 we show that both the persistence of concentration in growing markets and the non-monotonicity of the concentration/ market size relationship can be explained by switches in the technology used by firms in equilibrium as the market grows. We analyze the case in which ex-ante

[8] See for example Evely and Little (1960), Pashigian (1969), Phlips (1970), George and Ward (1975), and Hart and Clarke (1980).

identical firms not only make capacity and entry choices, but also choose their production technology. We find that the equilibrium technology choice depends on the size of the market. If the market is small, firms use the small fixed costs/high marginal costs technology, but as the market becomes larger, firms choose the higher fixed cost/lower marginal cost technology. This switch drives the non-monotonicity in the relationship between market size and concentration and can explain the persistence of concentration in growing markets. For certain market sizes the technologies can also coexist and asymmetric equilibrium structures occur. This reflects the textbook argument of the origins of firm asymmetry, which arises endogenously here.

In Chapter 3 we analyze the dynamics of scale and scope of multi-product chemical corporations, showing that the observed persistence of the concentration can also be explained in terms of two counteracting forces. We first show theoretically that the production capacities of firms on product market level tend to *converge*, leading to a more fragmented industry structure on a market by market level. Secondly, in terms of the number of products firms produce, the chemical corporations tend to *diverge*, with firms that already produce numerous product varieties being more likely to introduce the next product variety. We find empirically that within narrowly defined chemical product markets (6 to 7-digit SIC) firm sizes tend to converge if measured by installed production capacity relative to the market average. In other words, on a considerable disaggregated level differences in installed capacity between chemical businesses tend to become smaller. However, the number of products chemical corporations produce tends to diverge on a (4-digit SIC) sector level. Firms that have already a considerable presence in an industry in the chemical sector are more likely to introduce the next product than firms that have a smaller presence in that particular sector. So firm sizes tend to converge in terms of scale on a market by market basis, but tend to diverge in terms of scope. These counteracting forces are consistent with the persistence of concentration in growing chemical industries that is illustrated in Table 1.2. The explanation is based on firm dynamics that are the result of strategic interaction among firms in the investment process, and does not include any argument related to technological change.

In Chapter 4 we show that the asymmetry in the size distribution of firms has implications for the distribution of profitability (pay offs), if firms act strategically and the (asymmetric) outcome is a Nash equilibrium. More precisely, an asymmetric Nash equilibrium puts an upper and lower bound on the difference in profitability between the firms, where the shape of the bounds depends on industry characteristics.

We test these bounds using the FTC Line of Business data and find that the slope of the profitability-size trade off in a panel of homogeneous goods industries depends on industry characteristics that reflect the sensitivity of the actions of firms to their competitor's pay off. If firm asymmetry is a purely stochastic realization, then there is little reason why these restrictions should be satisfied[9]. Hence, this empirical analysis can be interpreted as a direct test of whether strategic interaction is important in the explanation of firm asymmetries. We conclude that the relationship between profitability and size of business firms can be positive or negative, depending on the characteristics of the industry.

---

[9] Mancke (1974) argued that the empirically observed correlation between size (market share) and profitability can be explained by stochastic factors only and does not have necessarily anything to do with economies of scale or market power. In a response, Caves, Gale and Porter (1977) showed that there is clear evidence of a behavioral component in the relationship between profitability and market share.

TABLE 1.2 Changes in Concentration for Some Growing Chemical Industries* (1963 - 1987).

SIC code	Name	Increase in Real Sales in %**	Absolute Change in Market Share of largest 4 producers, in %
2813	Industrial Gases	93.9	+ 5
2816	Inorganic Pigments	55.2	- 4
2822	Synthetic Rubber	35.4	- 7
2824	Organic Fibres	146.4	- 18
2833	Medicinals and Botanicals	45.2	- 4
2834	Pharmaceutical Preparations	205.1	0
2841	Detergents	71.2	- 7
2843	Surface Active Agents	432.1	- 1
2844	Toilet Preparations	156.5	- 6
2851	Paints and Allied Products	62.9	+ 4
2865	Cyclic Crudes and Intermediates	130.2	- 19
2869	Industrial Organic Chemicals, n.e.c.	172.2	- 20
2893	Printing Ink	181.2	- 3

* Source: Census of Manufactures 1987, US Department of Commerce, Bureau of the Census, Washington D.C.
** Sales are deflated by the Producer Price Index for Chemical and Allied Products (SIC 28). Source: Statistical Abstract of the United States, US Department of Commerce, Bureau of the Census, Washington DC.

CHAPTER 2

MARKET SIZE AND MARKET STRUCTURE WITH MULTIPLE TECHNOLOGIES.

2.1 Introduction.

The relationship between market size and market structure has been one of the key research areas in industrial organisation. A well established result in theoretical models of industrial structure is that in homogeneous goods industries market shares fall monotonically as the market size increases, if firms use a given constant unit cost technology with a fixed cost of entry. The number of incumbents is determined by the size of the fixed cost relative to the size of the market. In the limit, as the size of the market tends to infinity, concentration goes to zero[10]. In models of horizontal product differentiation the monotonicity and the fragmentation result hold in terms of a lower bound to all feasible equilibrium market structures. Multiplicity of equilibria occurs through the existence of multi-plant firms[11]. The lower bound is the most fragmented equilibrium structure, a configuration in which all firms produce a single product. As long as the level of fixed cost is exogenous, the monotonicity and fragmentation result hold in one or the other form. Sutton (1991) calls this the class of 'exogenous sunk cost models'.

They can be distinguished from models in which fixed costs are determined endogenously. Both the monotonicity and the fragmentation result can then fail. Consider a model of vertical product differentiation, such as Shaked and Sutton (1987). The relationship between size and concentration can be positive or negative, depending on the size of the market[12]. Firms deviating and spending more on advertising, increase their production scale in order to recoup the higher sunk costs. A growing market can show increasing (decreasing) concentration, if the optimum

---

[10] The index of market concentration can be defined in various ways for this convergence result to hold. For example as n-Firm Concentration Ratio, or as Herfindahl Index.

[11] See for example Schmalensee (1978), Bonanno (1987) or Shaked and Sutton (1990).

[12] See Sutton (1991, Chapter 2 and 3).

size of the firms increases (decreases) relative to the size of the market. The fragmentation result does not hold if there is always a size of the market above which it is profitable for a firm to deviate, pay a higher sunk cost, increase consumers' willingness to pay, and take a share of the market that is strictly positive. As the market size tends to infinity, fixed costs remain significant relative to the size of the market and market shares are bounded away from zero. Dasgupta and Stiglitz (1980) show that the same mechanism can occur in a homogeneous goods industry if the level of firms' R&D expenditures is a choice variable. Higher R&D expenditures generate a lower unit cost of production and the same kind of mechanism generates the non-convergence result. Sutton (1991) labels this class of models as 'endogenous sunk cost models'[13].

In this chapter we show that the relationship between market size and market structure can be non-monotonic and discontinuous if there is a set of discrete, constant unit costs technologies with fixed costs of entry. Although the relationship is generally negative, the concentration ratio jumps up at certain market sizes, due the firms switching from a high unit cost (low fixed cost) technology to a lower unit cost (higher fixed cost) technology. The optimum size of the firms increases relative to the size of the market and the concentration increases. For larger market sizes only equilibria with the lower unit cost technology exist and the concentration is again falling as the market size increases. The convergence to a fragmented market structure in the limit holds if the number of available technologies is finite.

Although the choice of sunk cost is endogenous, it differs from the endogenous sunk cost models mentioned before in that the choice is discrete rather than continuous. The technologies are standard constant marginal cost technologies with a fixed cost of entry, although unlike the exogenous sunk cost models, firms can choose their production technology. The set up of our model therefore has both features of the exogenous and endogenous sunk cost models, which is also reflected in the results.

In Section 2.2 a Cournot model is presented endogenizing the technology choice. If diffusion of new technologies throughout the industry is indeed as quick as claimed

[13] See for other examples Shaked and Sutton (1982, 1983) and Sutton (1986, 1991, Chapter 3).

by Mansfield (1985)[14], then, at least in the longer run, technology is a choice variable to all firms in the industry. We show that firms which earn the same equilibrium profit will use the same technology. If not, a firm with the low unit cost technology (and a large optimum capacity) would have an incentive to deviate to the higher unit cost technology (and a smaller optimum capacity), reducing the total industry capacity and increasing the market price. The profit of the non-deviating high unit cost firm is higher than before. Since the profit of the deviator is at least as high as the profit of the non-deviating firm, it is worthwhile for the low unit cost firm to deviate. Consequently, in multiple technology equilibria, firms using a higher fixed cost technology will earn a strictly higher net profit than firms using the low fixed cost technology.

In Section 2.3 the market structure is endogenized, using a zero profit condition. Technologies cannot coexist in equilibrium, though single technology equilibria do exist. If the market is relatively small compared to the fixed costs of either technology, an equilibrium exists in which the high unit cost technology is used. If on the other hand the market size is relatively large, an equilibrium exists in which the lower unit cost technology is used. Concentration does not fall monotonically as the market size increases. Moreover, if there are more than two production technologies available, then there can be market sizes for which there is no equilibrium at all. Some technologies are not used in equilibrium for any market size.

The analysis using a zero profit condition is not entirely satisfactory from a game theoretic point of view, though it formalizes the basic intuition in a simple and clear cut way. In Section 2.4 we model entry as a strategic choice variable. Firms can make a positive profit in equilibrium, though entry by an additional firm is not profitable. If the profits of low unit cost firms are sufficiently high, they will not have an incentive to deviate to the high unit cost technology, and more than one technology can be used in equilibrium. Like in a standard Cournot game with entry, multiple equilibria occur if entry and technology choices are made simultaneously. Analyzing the lower bound of the industry concentration, in line with Sutton's (1991) analysis, the non-monotonicity of the fall of industry concentration with increasing

[14] Mansfield (1985) reports views of managers on how quickly new industrial technology leaks out to competitors. Information concerning about the detailed nature and operation of a new product or process leaks out within about one year.

market size turns out to be robust to this alternative way of modelling entry.

2.2 A Model of Technology Choice.

Consider an industry in which firms produce a homogeneous good. Consumer demand is iso-elastic:

$$Q = \frac{S^2}{p} \tag{2.1}$$

where Q is the total quantity produced by the firms in the industry, p is the market price of the good and S^2 is a constant which indicates the size of the market in terms of total sales.

Assume there are k production technologies, where k is finite. All are constant marginal cost (MC) technologies with a given fixed cost (FC). To make the problem non-trivial, assume that a lower MC-technology has a higher FC and vice versa. I.e.

$$\begin{aligned} c_1 > c_2 > \dots > c_{k-1} > c_k \\ \delta_1 < \delta_2 < \dots < \delta_{k-1} < \delta_k \end{aligned} \tag{2.2}$$

where c_i^2 and δ_i^2 are the MC and FC of the i-technology respectively[15].

Assume that the total number of firms in the market $N \geq 2$ is given. The strategy is to simultaneously choose technology and production level. An economic justification for the simultaneous choice of technology and output is that the capacity of a plant is often fixed once it is built. The technology chosen is embodied in the capital. Changing capacity gives scope for changing technology and vice versa. The technical argument is that the analysis is more tractable than a sequential set up.

[15] The size of the market and the costs are defined in quadratic terms to ease notation later on.

The gross profit of a firm that chooses technology i is:

$$\pi_i = (p - c_i^2)q_i \tag{2.3}$$

where q_i is the quantity produced by the firm. The solution concept of the game is a Nash equilibrium in pure strategies.

Call equilibria in which more than one technologies are chosen 'Multiple Technology Equilibria' (MTE), and equilibria in which only one technology is chosen 'Single Technology Equilibria' (STE).

Lemma 2.1. In an MTE the high FC-technology firms will make a higher net profit than low FC-technology firms. ‖

Proof: See Appendix 2A.

Consider an equilibrium in which firms use different technologies. If a firm deviates from a low MC-technology to a higher MC-technology, then the deviator, being the relatively large firm, *could* choose the equilibrium size of high MC firms, which are the relatively small firms. The total production level in the industry would be reduced, given the production of all other firms. The price will be higher than in equilibrium and all firms will make a higher profit. The profit of the deviating firm is higher than the equilibrium profit of the high MC firms, if the deviator chooses the equilibrium size of high MC firms. So, for there not to be an incentive for a low MC firm to deviate, its equilibrium profit must be higher than the equilibrium profit of high MC firms.

Corollary 2.1. Firms that earn the same profit in equilibrium use identical technologies. ‖

Proof. This follows directly from the proof of Lemma 2.1. If the firms did not use the same technology, there would have been an incentive for low MC-technology firms to deviate to a higher MC-technology. QED.

The intuition for this result is, that if firms use a different technology and earn the same profit, there is always an incentive for a low MC-technology firm to deviate to the higher MC-technology (Lemma 2.1). Note that the result applies to losses as well.

2.3 Entry Determined by a Zero Profit Condition.

In this section a zero profit condition is used to endogenize the market structure. The first result that will be derived is that MTE do not exist. Focusing on STE, the relationship between market size and market concentration will be analyzed. It will be shown that the relationship is generally negative, but discontinuous due to switches in the technology that is chosen in equilibrium. The entry choice will be modeled explicitly in the next section, where it will be shown that the qualitative results derived here hold for a wide range of parameter constellations.

The number of firms is treated as a continuous variable and integer problems are ignored for the moment. The zero profit condition dictates that in equilibrium:

$$\pi_i(N_1^*,..,N_k^*) = \delta_i^2 \quad \forall\ i \ \ \textit{for which} \ \ N_i^* > 0$$

where N_i is the number of firms using the i-technology[16].

Corollary 2.1 stated that high FC-technology firms make a higher net profit than the low FC-technology firms in an MTE. This is clearly incompatible with zero profits for all firms. Hence MTE cannot exist.

Corollary 2.2. There cannot be an MTE if entry is determined by a zero profit condition. ‖

Proof. This result follows from the proof of Lemma 2.1, letting $a = 0$.

Corollary 2.2 implies that all firms will use the same technology if a zero profit condition holds for each technology used.

2.3a Two Production Technologies.

Though in general there is a negative relationship between market size and concentration, the relationship can be inverted in this model. The simplest setting to show this is one in which only two production technologies (1 and 2) are available. As before, technology 1 is the higher MC, lower FC-technology.

[16] A * refers to equilibrium values throughout.

Proposition 2.1 Let $S_1 = \dfrac{c_1\delta_2 - c_2\delta_1}{c_1 - c_2}$

(a) The STE $\left\{N_1^*, N_2^*\right\} = \left\{\frac{S}{\delta_1}, 0\right\}$ exists if $S \le S_1$.

(b) The STE $\left\{N_1^*, N_2^*\right\} = \left\{0, \frac{S}{\delta_2}\right\}$ exists if $S \ge S_1$. ||

Proof. See Appendix 2A.

This result states that if the size of the market is small relative to the fixed cost, all firms choose the high MC-technology in equilibrium. If the market size is relative large, all firms will use the low MC-technology. Note that the number of firms in the STE is exactly the same as in a one shot Cournot game with exogenous technology choice, zero profit and iso-elastic demand[17].

The intuition for Proposition 2.1 is that if a high MC-technology firm deviates to the low MC-technology, the industry production level increases, the price falls and all other firms make a lower profit. If the market size is small, the price fall caused by the deviation is relatively large and hence the deviator will make a loss too. If the market size increases, market externalities decrease and the price fall due to the deviation decreases. The deviation becomes profitable if $S > S_1$, and the high MC-technology STE cannot exist.

A low MC-technology STE can only exist if the market size is sufficiently large. A firm deviating from technology 2 to technology 1 will reduce its production and hence raise the market price. This effect will be larger in a small market, and the deviation is relatively more profitable. If $S < S_1$ the deviation by a technology 2 firm will result in a strictly positive pay off for the deviator and a technology 2 STE cannot exist.

Two equilibria exist if the market size is S_1 . One with only technology 1 and the other with only technology 2. The 'marginal' firm in both STE is indifferent between choosing either technology. But as follows from Lemma 2.1, even at market size S_1 an MTE cannot exist. The areas of the two types of STE are not overlapping. This result is due to the zero profit condition and is not very general.

---

[17] See Sutton (1991, Chapter 2).

The market price is lower in larger markets. The price in an i-technology equilibrium can be written as $p_i = (c_i^2 S)/(S - \delta_i)$, which is decreasing in S. The price in the technology 2 equilibrium at S_1 is lower than the price in the technology 1 equilibrium, since $p_2/p_1 = (S_1 - \delta_2)/(S_1 - \delta_1) < 1$. The price falls monotonically as the market size increases, though there is a discontinuity at S_1 .

We now turn to the relationship between market structure and market size. The concentration in the industry will be measured in terms of the 1-Firm Concentration Ratio C, which is the market share of the largest firm in the industry. This coincides with the average market share in a symmetric equilibrium. The results that are derived will also hold qualitatively for alternative measures of concentration, like the Herfindahl Index. Figure 2.1 shows the equilibrium 1-Firm Concentration Ratio as a function of the market size. Firstly, note that the fragmentation result holds and the low MC-technology is used in the limit, as the market size goes to infinity. The model then behaves like a Cournot oligopoly model with only one production technology. In contrast to the one technology case, however, the fall in the concentration is not continuous for smaller firm sizes. At S_1 , C jumps up, due to the switch in the technology chosen by the firms in equilibrium.

The non-monotonicity has the following empirical implication for the comparison of concentration between two markets of different sizes. If the size of the small market is just below S_1 and the size of a large one is just above S_1, the smaller market has a more fragmented market structure than the larger market. In the former one many relatively small firms operate, using the high MC-technology, whereas in the large market a few large firms operate, using the low MC-technology.

The possibility of a small market having a more fragmented market structure than a larger market is not a new result. The contribution of this paper is, however, that one can observe the inverted relationship in homogeneous goods industries. It goes further than Dasgupta and Stiglitz (1980) who analyze endogenous sunk cost in a homogeneous goods industry, but find a (continuous and monotonic) negative relationship between market size and market structure, as their tradeoff between sunk cost and (expected) marginal cost is continuous. Our example shows that heterogeneous demand is not a necessary ingredient to generate the inverted relationship, nor is the effect of higher FC on consumers' 'willingness to pay', as in models of vertical product differentiation.

Figure 2.1

Market Size and Concentration if there are 2 Production Technologies.

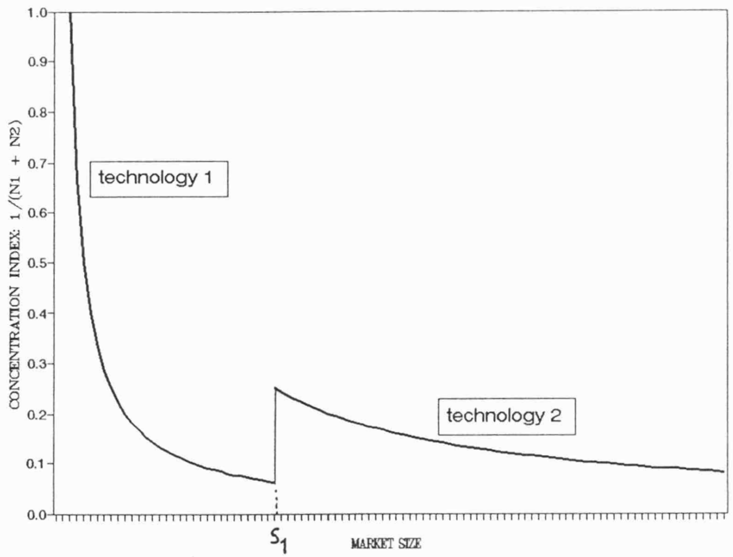


Figure 2.2

Market Size and Concentration if there are 3 Production Technologies.

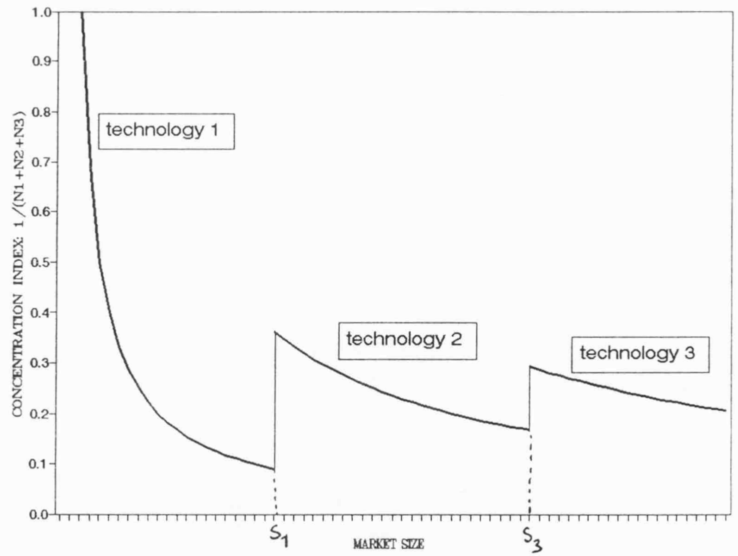


Figure 2.3

Non-Existence of an Equilibrium if there are 3 Production Technologies.

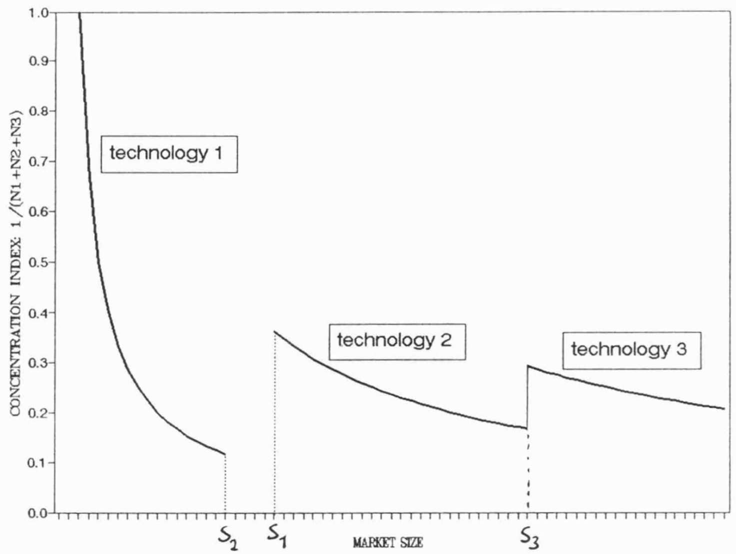
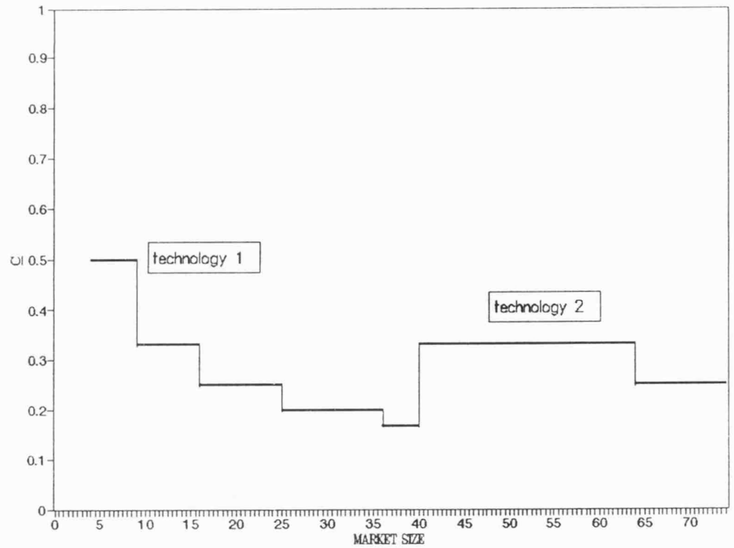


Figure 2.4

Market Size and the Most Fragmented Market Structure $\underline{C}$



2.3b Three Production Technologies.

If there are more than two production technologies, no equilibrium exists for certain market sizes, and some technologies are never chosen in equilibrium whatever the size of the market. To illustrate this in a simple case, assume that there is a third production technology available to produce the homogeneous good. Technology 3 has lower MC and higher FC than technology 2, as in relationship (2.2). As before there exists an equilibrium if no firm has an incentive to deviate from its technology or capacity choice and profits are zero for all firms. In this case, however, each technology can potentially be replaced by two other technologies, or alternatively, each firm can deviate to one of two other technologies. There are no MTE as follows from Corollary 2.1, since with coexistence there is always an incentive for low MC-technology firms to deviate to a technology with higher MC and earn a strictly positive profit.

To establish the existence of a STE, restrictions are derived that rule out incentives to deviate to another technology. The counterpart of Proposition 2.1 for the three technology case is:

Proposition 2.2. Let $S_1 = \dfrac{c_1\delta_2 - c_2\delta_1}{c_1 - c_2}$ $S_2 = \dfrac{c_1\delta_3 - c_3\delta_1}{c_1 - c_3}$ $S_3 = \dfrac{c_2\delta_3 - c_3\delta_2}{c_2 - c_3}$

(a) The STE $\{N_1^*, N_2^*, N_3^*\} = \{\frac{S}{\delta_1}, 0, 0\}$ exists if: $S \leq Min(S_1, S_2)$.

(b) The STE $\{N_1^*, N_2^*, N_3^*\} = \{0, \frac{S}{\delta_2}, 0\}$ exists if: $S_1 \leq S_3$ and $S \in [S_1, S_3]$.

(c) The STE $\{N_1^*, N_2^*, N_3^*\} = \{0, 0, \frac{S}{\delta_3}\}$ exists if: $S \geq Max(S_2, S_3)$. ‖

Proof. See Appendix 2A.

The evolution of the industry structure for a particular parameter constellation is shown in Figure 2.2. It is a special case in the sense that the equilibrium of technology 1 is 'broken' by a deviation of the marginal firm to the 'next' technology (2) if the market size is just above S_1 . Similarly, the equilibrium of technology 2 is 'broken' by a deviation to the 'subsequent' technology (3) for market sizes just above S_3 . In general, however, the equilibrium of technology 1 can also be 'broken' by a deviation to technology 3, in particular if $S_2 \leq S_1$. Then there does not exist an equilibrium if $S \in]S_2, S_1[$, as shown in Figure 2.3. In this area, an equilibrium in technology 1 cannot exist because of a deviation to technology 3, the equilibrium in technology 2 cannot exist because it is profitable for the marginal firm to deviate to technology 1 and an equilibrium with only technology 3 will be broken by a deviation to technology 2. A further result is that if $S_3 < S_1$ there does not exist a technology 2 STE for any market size.

The evolution of the market structure in the three technology case is similar to the two technology case, although there might be market sizes for which there does not exist an equilibrium. If the market is small the high MC-technology is chosen in equilibrium, whereas if the market size is large the lower MC-technologies are chosen. If one ranks markets in order of their size, starting with the smallest, the resulting ranking of the technologies equilibrium would be one from a high MC to a low MC-technology. The fragmentation result still holds, and the fall in concentration for markets of increasing sizes is non-monotonic. Again, it cannot be ruled out that an industry serving a large market is more concentrated than one serving a small market, in particular if a lower MC-technology is used in the large market than in the small market. This result is obvious as long as the technologies are given and different in both markets. It is less obvious if the technology choice is endogenous.

2.4 Entry as a Strategic Choice.

We now introduce entry as a strategic choice, to replace the zero profit condition, The purpose is to show that the non-monotonic relationship between market size and concentration that was derived in Section 2.2 is robust to the difference in modelling entry. The market sizes for which STE with different technologies exist overlap and we might find MTE in these regions.

The set-up of the game is as before, with two production technologies 1 and 2. Their cost structure is as in relationship (2.2). An additional dimension is added to the action space, with a binary choice variable: "enter" or "stay out". Entry decision, technology selection and capacity choice are made simultaneously. The players of the game are a large number of potential entrants. Their pay off is zero if they stay out, and is given by (2.3) if firm j decides to enter the industry and choose technology i.

The solution concept is again a one shot Nash equilibrium in pure strategies. The following necessary conditions ensure that none of the players has an incentive to deviate from its equilibrium strategy.

An equilibrium is an $\{N_1^*, N_2^*\}$, s.t.

(i) no exit: $\pi_i(N_1^*, N_2^*) \geq \delta_i^2 \qquad i = 1,2$

(ii) no entry: $\pi_1^D(N_1^* + 1, N_2^*) \leq \delta_1^2 \qquad \pi_2^D(N_1^*, N_2^* + 1) \leq \delta_2^2$

(iii) no deviation of technology choice:

$$(a) \quad \pi_1^D(N_1^* + 1, N_2^* - 1) - \delta_1^2 \leq \pi_2(N_1^*, N_2^*) - \delta_2^2 \qquad if \ N_2^* \geq 1$$

$$(b) \quad \pi_2^D(N_1^* - 1, N_2^* + 1) - \delta_2^2 \leq \pi_1(N_1^*, N_2^*) - \delta_1^2 \qquad if \ N_1^* \geq 1$$

(iv) oligopoly constraint: there are at least two firms in the market $(N_1^* + N_2^* \geq 2)$

The pay off $\pi_i(N_1^*, N_2^*)$ is the maximal gross profit a firm that uses technology i can earn, if there are N_i firms in the equilibrium using technology i. A deviating firm, whose equilibrium choice is to stay out, but enters instead choosing technology 1 has a maximal gross profit of $\pi_1^D(N_1^* + 1, N_2^*)$. The gross profit of a firm that deviates from technology 2 to technology 1 is $\pi_1^D(N_1^* + 1, N_2^* - 1)$.

Conditions (i) ensure that the marginal entrant makes a positive profit, i.e. has no incentive to stay out. The 'no entry' conditions (ii) ensure that no additional firm enters. The additional entrant would make negative profits, given the choices of all other firms even though it selects its technology and capacity optimally. If the two conditions under (iii) hold, then there is no incentive for any firm that entered to deviate from its technology choice, since if it would do so, it would make a lower overall profit.

The full set of conditions for the existence of MTE and STE is derived in Appendix 2B (Lemmas 2.2 and 2.3). However, the set of conditions cannot be expressed in terms of restrictions on the market sizes alone as before. Only necessary conditions for existence can be presented in those terms.

Proposition 2.3. Let $S_1 = \dfrac{c_1\delta_2 - c_2\delta_1}{c_1 - c_2} \quad S_4^2 = \dfrac{(c_1 + c_2)(\delta_2^2 - \delta_1^2)}{(c_1 - c_2)}$.

(a) A necessary condition for a STE $\{N_1^*, 0\}$ to exist is $S \leq S_1$.

(b) A necessary condition for a STE $\{0, N_2^*\}$ to exist is $S \geq S_4$.

(c) A necessary condition for a MTE $\{N_1^*, N_2^*\}$ to exist is $S \in [S_4, S_1]$, where $S_4 \leq S_1$ is always satisfied. ‖

Proof. See Appendix 2A.

The size of the market for which a high MC-technology equilibrium can exist is bounded from above. Similarly the set of market sizes for which a low MC-technology equilibrium can exist is bounded from below. For relative small markets, if $S \leq S_1$, only high MC-technology STE can exist and for relative large markets, if $S \geq S_4$, only low MC-technology STE can exist. Since $S_4 < S_1$, the areas for both types of STE overlap. This intersection is the set of market sizes for which MTE can exist. Although the intersection always exists, there might not always be MTE, due to the violation of the oligopoly restriction (condition (iv)) or due to the discreteness of the solution (integer values).

If the market size is small, the market structure must be fragmented to 'prevent' a technology 1 firm from deviating to technology 2. If in equilibrium the firms are sufficiently small relative to the market, deviating to technology 2 would imply increasing their size by such an amount that this leads to a sufficiently downward move of the market price that the deviation is not profitable. A larger market size increases the benefits from this deviation, however, as the decreasing effect on the market price becomes weaker. For $S \geq S_1$ the non-deviation restriction of technology choice would dictate such a fragmented industry structure, that the firms using the high MC-technology earn too low a profit to recoup the FC, and the no exit condition (i) is violated. Therefore a technology 1 STE cannot exist. Note that this condition is the same as in Proposition 2.1a.

The set of market sizes for which a technology 2 equilibrium can be supported is bounded from below. The intuition is similar to the one presented earlier. If the market size is small, the increasing effect of a deviation by a technology 2 firm on the market price is large. Hence the deviation is profitable and a technology 2 equilibrium cannot exist.

Since in a MTE there is a 'marginal' firm for every type of technology, the reasoning just presented for both types of STE should hold simultaneously. This is the case for those market sizes for which the non-deviation restrictions of both types of STE hold simultaneously. Therefore MTE can only occur for those market sizes for which both types of STE can exist.

Proposition 2.3 will be illustrated using an numerical example, as no explicit analytical solution for the relationship between market size and market structure can be derived.

Table 2.1: Numerical Values of the Technologies.

technology 1	$c_1^2 = 0.015$	$\delta_1^2 = 1$
technology 2	$c_2^2 = 0.010$	$\delta_2^2 = 4$

Table 2.1 shows the numerical values for the costs of technologies 1 and 2. Table 2.2 shows characteristics of the equilibria that exist for four different sizes of the market. If the market is small, say $S^2 = 5$ or 25, only high MC-technology STE exist. If the market is large, for example if $S^2 = 50$, only low MC-technology STE occur. Has the market size $S^2 = 33$, then three types of equilibria exist:

- two STE in which firms use the high MC-technology (1) ({5,0} and {4,0}). Note that these equilibria are the relatively fragmented ones. The prices in both equilibria are *not* the same.
- one STE in which the firms use the low MC-technology (2) ({0,2}). This is the relatively concentrated equilibrium.
- one MTE ({2,1})[18].

In the MTE the equilibrium net profits of technology 1 firms is 1.06, and of technology 2 firms is 4.25. This is consistent with Lemma 2.1. The net profit of a deviating technology 1 firm is 0.96 and the net profit of a deviating technology 2

[18] The multiplicity of equilibria will not survive in a game of three stages (entry - technology choice - production choice). Given that the pay-off of the outside option is zero, the maximum number of firms that is consistent with any Nash equilibrium for a particular parameter constellation will enter in a Perfect Equilibrium. Or alternatively, the Perfect Equilibrium is the most fragmented Nash equilibrium.

firm is 3.95. Therefore none of the firms has an incentive to deviate.

Table 2.2: Market Structure and Minimum Concentration in Equilibrium.

Market Size S^2	5	25	33	50
Equilibria $\{\hat{N}_1, \hat{N}_2\}$ *Market Price p*	{2,0} *.030*	{3,0} {4,0} *.023 .020*	{5,0} {4,0} *.019 .020* {2,1} {0,2} *.020 .020*	{0,3} *.015*
$\underline{C}$	0.5	0.25	0.20	0.33

Given the multiplicity of equilibria for some market sizes it makes sense to concentrate on the minimum obtainable 1-Firm Concentration Ratio $(\underline{C})$, following the analysis of Sutton (1991). $\underline{C}$ goes up between the market size of 33 and 50, again due to the change in technology that is used by firms in the most fragmented equilibrium. The results from Table 2.2 suggest also that in this case the market price is falling monotonically as the market size is increasing.

In Figure 2.4 the most fragmented market structure $(\underline{C})$ is shown as a function of market size, if technologies 1 and 2 are as given in Table 2.1. The increase in $(\underline{C})$ occurs at $S^2 = 40.5$, due to a switch of technology that is used by the firms in the equilibrium that generates the most fragmented market structure. For $S^2 \leq 40.5$ the low MC-technology 1 is used, for $S^2 \geq 40.5$ the high MC-technology 2. Note that $S_1 = 41.6$ in this example. Although up to that size a technology 1 equilibrium could exist, $\underline{C}$ jumps up at a size just below that due to the integer value of the number of firms. The $\underline{C}$ equilibria are all STE, but this is a special feature of this example.

2.5 Discussion.

2.5a Shakeout and Technology Choice.

This framework can shed some light on the "shakeout" of firms in growing industries, as described by Klepper and Graddy (1990). The market size is increasing over time, and firms play essentially a repeated one shot game, in which they decide to enter for one period. At the end of each period their investment is obsolete. If the discount rate is (close to) zero, the dynamic game is essentially a sequence of independent one shot games. The market structures shown in figures 2.1 to 2.4 can then be interpreted as the dynamic evolution of the industry concentration. If there are more than one production technologies available to the firms, then an occasional "shakeout" can be observed as the industry employs ever larger scale technologies as it grows. So, although the shakeout goes together with a technological change, the new technology could have been around for some time. It is only when the size of the industry increases beyond a certain critical level, that the larger scale technology can be profitably employed.

This explanation is similar in spirit to Jovanovic and MacDonald (1993). Firms operate in a perfectly competitive environment, in which entrants employ a common technology which, after some time, is replaced by a new technology with higher scale economies. The firms that fail to switch to the new, larger scale technology, exit the industry. In their explanation the timing of the shakeout is determined by the (stochastic) arrival of the new technology[19]. In the dynamic interpretation of our model the timing of the shakeout is determined by the growth of the market, given the technology and initial conditions. The large scale technology could have been around for some time, but can only be profitably employed once the market has reached a certain size. This suggests that the "shakeout" can occur if there is no immediate prior technological innovation.

2.5b Plant Size and Market Size.

It has been widely established empirically that the size of plants is positively related to the market size[20]. There are a large number of candidate models that can explain this observed phenomenon[21]. The study by Pryor (1972) is of particular interest in

---

[19] There is an alternative model by Klepper and Simons (1993), based on firms entering by developing some new product variety. They subsequently spend some fixed amount to lower unit production costs. With inertia in sales and some capital market imperfections, the currently larger firms spend more on process innovation as they spread the fixed costs over a larger volume of sales. Some smaller firms exit the industry as a consequence.

[20] The classic study in this area, making an international comparison is Scherer et al. (1975). See Caves (1989) for a summary of this literature.

[21] The optimal size of firms is increasing in the market size in a textbook constant marginal cost Cournot equilibrium. To the extent that this model describes establishment sizes, rather than firm sizes, it explains the observed regularity.

our context. He analyzes a cross-section of 23 western economies and establishes that large plants increase more than proportionately with market size, hence increasing concentration in terms of plants. If the size of firms is determined by the size of plants, then this is exactly the type of result we derived here.

2.5c Exogenous vs Endogenous Sunk Cost Models.

The difference in the limiting behaviour of exogenous and endogenous sunk cost models is the basis of the empirical study by Sutton (1991). The hypothesis is that the endogenous sunk cost model describes industries in which advertising or R&D is important. Levels of advertising or R&D must be chosen before production and marketing. Consequently, in those industries one would expect in the limit the concentration index to be bounded away from zero. However, in industries in which fixed costs are constant, the fragmentation result should hold in the limit, possibly as a lower bound. Sutton (1991) divides a number of food-and-drink related industries into two groups. Both having only a limited amount of R&D expenditure. One group has low levels of advertising and the other has high levels. The hypothesis tested is that for the former group the convergence result holds, whereas for the latter it can fail. The results in this chapter suggest, however, that the division of industries along the lines of Sutton's endogenous and exogenous sunk cost models is only valid if there is only one production technology available in the group with low advertising levels. Otherwise these homogeneous goods industries might have features related to endogenous sunk costs as derived here.

Sutton's test of the limiting behaviour of the industry structure is based on the extrapolation of the observed levels of concentration with finite market sizes. He shows empirically that the concentration index indeed converges to zero in the group of industries with homogeneous goods, whereas the concentration index is bounded away from zero in the case of advertising intensive industries. Our model shows that one cannot rule out observing a positive relationship between concentration and market size in industries with a limited number of production technologies. Hence, extrapolation of the concentration index might not reveal convergence to zero if one has only a relatively small number of observations around the switching point of the technologies. There are two ways out within the context of the model. One is to find observations of the concentration index for much larger markets, which might reveal the limiting behaviour of the industry structure as the market size tends to infinity. The other is to analyze the production technologies used in the market in more detail to see if there is a difference. The above test is only valid if identical technologies are used in markets of different sizes.

2.6 Conclusion.

In this chapter we analyzed the relationship between market size and market structure if there is a finite set of production technologies available to the firms. Unlike the exogenous sunk cost model, the production technology is not given, but is a strategic choice variable of the firm. The model is different from the endogenous sunk cost model in the sense that the technology choice is finite and the trade-off between fixed and unit costs is discrete. It therefore constitutes an intermediate case between the exogenous and endogenous sunk cost model, both in terms of assumptions and results. If the two classes of models are tested against each other, the existence of such intermediate cases have to be taken into account.

Appendix 2A. Proofs of Lemma 1 and the Propositions of Chapter 2.

Lemma 2.1 In an MTE the high FC-technology firms will make a higher net profit than low FC-technology firms. ‖

Proof. The proof consists of showing that if firms using a different technology earn an identical profit there is always an incentive for a low MC-technology firm to deviate to the higher MC-technology. Assume firms use technology m or n earn an identical maximal profit a if $\{N_1,..,N_k\} = \{\hat{N}_1,..,\hat{N}_k\}$. Assume also that $c_m > c_n$, $\delta_n > \delta_m$ and $\hat{N}_m, \hat{N}_n \geq 1$. So the profit of a non-deviating m-technology firm is:

$$\begin{aligned}\pi_m(\hat{N}_1,..,\hat{N}_m,..,\hat{N}_n,..,\hat{N}_k) - \delta_m &= (\hat{p} - c_m^2)\hat{q}_m - \delta_m \\ &= \left[S - c_m^2\frac{\hat{Q}}{S}\right]^2 - \delta_m^2 \\ &= a \qquad (say)\end{aligned} \tag{2A.1}$$

where $\hat{Q}$, $\hat{q}_i = \hat{Q} - c_i^2\hat{Q}^2/S^2$ and $\hat{p}$ are respectively the quantity produced by the industry, the quantity produced by an i-technology firm and the market price that are consistent with $\{\hat{N}_1,..,\hat{N}_k\}$.

The optimum production of an n-technology firm that deviates to the m-technology, given the other firms produce $\hat{Q} - \hat{q}_n$, is: $q_m^D = Q^D - c_m^2 Q^{D2}/S^2$, where $Q^D = \hat{Q} - \hat{q}_n + q_m^D$. The profit of the deviator is then:

$$\begin{aligned}\pi_m^D(\hat{N}_1,..,\hat{N}_m+1,..,\hat{N}_n-1,..,\hat{N}_k) - \delta_m^2 &= \left[S - c_m^2\frac{Q^D}{S}\right]^2 - \delta_m^2 \\ &= \left[S - c_m c_n\frac{\hat{Q}}{S}\right]^2 - \delta_m^2 \\ &> \left[S - c_m^2\frac{\hat{Q}}{S}\right]^2 - \delta_m^2 \\ &> a\end{aligned} \tag{2A.2}$$

The second equality comes from $Q^D = \frac{c_n}{c_m}\hat{Q}$, the first inequality holds because $c_m > c_n$ and the second inequality comes from (2A.1). I.e. the deviant will make a strictly higher profit (if $a \geq 0$) or a smaller loss (if $a < 0$). A necessary condition for this deviation not to be profitable is that the low MC-technology firms earns a profit in equilibrium that is strictly higher than the profit of a higher MC-technology firm. QED.

Proposition 2.1 Let $S_1 = \dfrac{c_1\delta_2 - c_2\delta_1}{c_1 - c_2}$

(a) The STE $\left\{N_1^*, N_2^*\right\} = \left\{\frac{S}{\delta_1}, 0\right\}$ exists if $S \leq S_1$.

(b) The STE $\left\{N_1^*, N_2^*\right\} = \left\{0, \frac{S}{\delta_2}\right\}$ exists if $S \geq S_1$. ‖

Proof.

(a) Two conditions have to be satisfied:

- $\pi_1(N_1^*, 0) = \delta_1^2$ implies the standard result that $N_1^* = \frac{S}{\delta_1}$. (See Sutton (1991), Chapter 2)).

- $\pi_2^D(N_1^* - 1, 1) \leq \delta_2^2$. Using a similar derivation as in (2A.2) this condition implies that

$$\begin{aligned} \pi_2^D(N_1^* - 1, 1) &= \left[S - c_1 c_2 \tfrac{Q^*}{S}\right]^2 \leq \delta_2^2 \\ \Rightarrow \quad Q^* &\geq \frac{S(S - \delta_2)}{c_1 c_2} \end{aligned} \tag{2A.3}$$

if $S - c_1 c_2 \frac{Q^*}{S} > 0$. Note that if $S - c_1 c_2 \frac{Q^*}{S} < 0$ (2A.3) implies that $p^* < c_1 c_2$. Since $p^D = p^* c_2 / c_1$ it follows that $p^D < c_2^2$, in which case it is never profitable to deviate to technology 2. Therefore (2A.3) is a sufficient condition. Standard calculations show that:

$$Q^* = N_1^* q_1^* = \frac{S}{c_1^2}\left[\frac{N_1^* - 1}{N_1^{*2}}\right] = \frac{S^2}{c_1^2}\left[\frac{S - \delta_1}{S}\right] \tag{2A.4}$$

Taking (2A.3) and (2A.4) together gives after rearranging:

$$S \leq \frac{c_1\delta_2 - c_2\delta_1}{c_1 - c_2} = S_1 \tag{2A.5}$$

(b) The same analysis applies here, mutatis mutandis. I.e.:

$$S \geq \frac{c_1\delta_2 - c_2\delta_1}{c_1 - c_2} = S_1 \tag{2A.6}$$

It is now straightforward to show that $\pi_2(N_1^*, 1) < \pi_2^D(N_1^* - 1, 1) \leq \delta_2^2$ holds if (2A.6) is satisfied. The profit of an additional entrant, choosing technology 2 is:

$$\pi_2(N_1^*, 1) = \left[S - c_2 Q^{*\frac{1}{2}}\right]^2 \leq \delta_2^2$$
$$\Rightarrow \quad Q^* \geq \frac{(S - \delta_2)^2}{c_2^2} \tag{2A.7}$$

Restriction (2A.7) is dominated by (2A.3) if (2A.5) holds.
A similar argument holds for $\pi_1(1, N_2^*) < \pi_1^D(1, N_2^* - 1) \leq \delta_1^2$. QED.

Proposition 2.2. Let $S_1 = \dfrac{c_1\delta_2 - c_2\delta_1}{c_1 - c_2}$ $S_2 = \dfrac{c_1\delta_3 - c_3\delta_1}{c_1 - c_3}$ $S_3 = \dfrac{c_2\delta_3 - c_3\delta_2}{c_2 - c_3}$

(a) The STE $\left\{N_1^*, N_2^*, N_3^*\right\} = \left\{\frac{S}{\delta_1}, 0, 0\right\}$ exists if: $S \leq Min(S_1, S_2)$.

(b) The STE $\left\{N_1^*, N_2^*, N_3^*\right\} = \left\{0, \frac{S}{\delta_2}, 0\right\}$ exists if: $S_1 \leq S_3$ and $S \in [S_1, S_3]$.

(c) The STE $\left\{N_1^*, N_2^*, N_3^*\right\} = \left\{0, 0, \frac{S}{\delta_3}\right\}$ exists if: $S \geq Max(S_2, S_3)$. ‖

Proof. The proof of this result is similar to the proof of Proposition 2.1, only that two possible deviations have to be checked. First note that the number of firms in a i-technology equilibrium is equal to S/δ_i , because of the zero profit condition. For the no deviation restrictions of the technology choice, the counterparts of condition (2A.3) in Proposition 2.1 become:

$$\pi_2^D(N_1^* - 1, 1, 0) = \left[S - c_1 c_2 \frac{Q^*}{S}\right]^2 \leq \delta_2^2$$
$$\Rightarrow \quad Q^* \leq \frac{(S - \delta_2)}{c_1 c_2} \tag{2A.8}$$

and

$$\pi_3^D(N_1^* - 1, 0, 1) = \left[S - c_1 c_3 \frac{Q^*}{S}\right]^2 \leq \delta_3^2$$
$$\Rightarrow \quad Q^* \leq \frac{(S - \delta_3)}{c_1 c_3} \tag{2A.9}$$

(2A.8) yields: $S \leq \frac{c_1\delta_2 - c_2\delta_1}{c_1 - c_2} = S_1$, and (2A.9) yields: $S \leq \frac{c_1\delta_3 - c_3\delta_1}{c_1 - c_3} = S_2$.

Checking the non-deviation restrictions for the technology 2 and 3 equilibria, gives the existence conditions under (b) and (c). QED.

Proposition 2.3. Let $S_1 = \dfrac{c_1\delta_2 - c_2\delta_1}{c_1 - c_2}$ $\quad S_4^2 = \dfrac{(c_1 + c_2)(\delta_2^2 - \delta_1^2)}{(c_1 - c_2)}$.

(a) A necessary condition for a STE $\{N_1^*, 0\}$ to exist is $S \leq S_1$.

(b) A necessary condition for a STE $\{0, N_2^*\}$ to exist is $S \geq S_4$.

(c) A necessary condition for a MTE $\{N_1^*, N_2^*\}$ to exist is $S \in [S_4, S_1]$, where $S_4 \leq S_1$ is always satisfied. ‖

Proof.

(a) From Lemma 2.3a, in particular (2B.11) and (2B.12), it follows that a necessary condition for the existence of a STE $\{N_1^*, 0\}$ is that

$$\frac{1 + Z}{c_1^2 + c_1 c_2} \leq \frac{S - \delta_1}{S c_1^2} \tag{2A.10}$$

$$\Rightarrow \quad S^2(c_1 - c_2) + 2c_2\delta_1 S - \frac{(c_1^2\delta_2^2 - \delta_1^2 c_2^2)}{(c_1 - c_2)} \leq 0 \tag{2A.11}$$

Solving for S gives:

$$\Rightarrow \quad S \leq \frac{c_1\delta_2 - c_2\delta_1}{c_1 - c_2} = S_1$$

(b) Lemma 2.3b shows that the STE $\{0, N_2^*\}$ can only exist if Z is defined. So,

$$1 - \frac{(c_1 + c_2)(\delta_2^2 - \delta_1^2)}{(c_1 - c_2) S^2} \geq 0 \tag{2A.12}$$

Solving for S^2:

$$\Rightarrow \quad S^2 \geq \frac{(c_1 + c_2)(\delta_2^2 - \delta_1^2)}{(c_1 - c_2)} = S_4^2$$

(c) From Lemma 2.2 it follows that both the conditions of Propositions (2.3a) and (2.3b) (resp. (2B.10) and (2B.12)) must be satisfied for there to be an MTE. I.e. $S \in [S_4, S_1]$ is a necessary condition. This set is well defined since $S_1 \geq S_4$, as follows from:

$$\frac{(c_1\delta_2 - c_2\delta_1)^2}{(c_1 - c_2)^2} \geq \frac{(c_1 + c_2)(\delta_2^2 - \delta_1^2)}{(c_1 - c_2)} \tag{2A.13}$$

$$\Rightarrow \quad (c_1\delta_1 - c_2\delta_2)^2 \geq 0$$

which is always satisfied. QED.

Appendix 2B Full Set of Existence Conditions for MTE and STE.

In this appendix the full set of existence conditions for MTE and STE are derived, if entry is a strategic choice variable. Lemma 2.2 gives the conditions for the existence of MTE, Lemma 2.3 shows the conditions for existence of both types of STE.

Lemma 2.2. Define: $g(N_1^*, N_2^*) = \frac{N_1^* + N_2^* - 1}{N_1^* c_1^2 + N_2^* c_2^2}$. A MTE exists if $g(N_1^*, N_2^*)$ satisfies:

$$\left[\frac{S-\delta_i}{Sc_i}\right]^2 \leq g(N_1^*, N_2^*) \leq \frac{S-\delta_i}{Sc_i^2} \qquad i = 1,2 \tag{2B.1}$$

and

$$\frac{1+Z}{c_2^2 + c_1 c_2} \geq g(N_1^*, N_2^*) \geq Max\left\{\frac{1}{c_1^2+c_2^2}, \frac{1+Z}{c_1^2+c_1c_2}, \frac{1-Z}{c_2^2+c_1c_2}\right\} \tag{2B.2}$$

where $Z = \sqrt{1 - \frac{(c_1+c_2)(\delta_1^2-\delta_2^2)}{(c_1-c_2)S^2}}$.

If Z is not defined, there is no MTE. ‖

Technically, the 'No exit' constraints (conditions (i)) occur in Lemma 2.2 as the first inequalities in relationship (2B.1). The 'No entry' constraints (conditions (ii)) are the second inequalities of (2B.1). Relationship (2B.2) summarizes the restriction of no deviation in technology choice (conditions (iii)) and the oligopoly condition (iv).

Proof. The proof consists of deriving the restrictions on the parameters of the model as a result of the equilibrium conditions (i) to (iv).

ad (i) 'No Exit'. From profit maximization:

$$\pi_i(N_1^*, N_2^*) = \left[S - c_i^2\frac{Q^*}{S}\right]^2 \geq \delta_i^2$$
$$\Rightarrow \quad Q^* \leq \frac{(S - \delta_i)}{c_i^2}S \tag{2B.3}$$

Since

$$Q^* = N_1^* q_1^* + N_2^* q_2^* = \frac{(N_1^* + N_2^* - 1)}{N_1^* c_1^2 + N_2^* c_2^2} S^2 = g(N_1^*, N_2^*) S^2 \tag{2B.4}$$

(2B.3) implies that if $g(N_1^*, N_2^*) \leq \frac{S - \delta_i}{S c_i^2}$ incumbent firms will make a positive profit.

This is the second inequality in relationship (2B.1).

ad (ii) 'No Entry'. The maximal profit obtainable by an additional entrant who chooses technology 1, given the (equilibrium) choices of all other firms is:

$$\pi_1^D = (N_1^* + 1, N_2^*) = \left[S - c_1 Q^{*\frac{1}{2}}\right]^2 \leq \delta_1^2$$
$$\Rightarrow \quad Q^* \geq \frac{(S - \delta_1)^2}{c_1^2} \tag{2B.5}$$

Using (2B.4) one can show that if $g(N_1^*, N_2^*) \geq (\frac{S - \delta_1}{S c_1})^2$, an additional entrant will make a loss. Along the same lines we can derive the case in which the deviant chooses technology 2. This is the first inequality in relationship (2B.1).

ad (iiia) 'No deviation of technology choice'. A 2-technology firm does not deviate to the 1-technology if:

$$\pi_1^D(N_1^* + 1, N_2^* - 1) - \pi_2(N_1^*, N_2^*) =$$
$$\left[S - c_1 c_2 \frac{Q^*}{S}\right]^2 - \left[S - c_2^2 \frac{Q^*}{S}\right]^2 \leq \delta_1^2 - \delta_2^2 \tag{2B.6}$$

$$\Rightarrow \quad \frac{Q^{2*}}{S^2}\left[c_2^2 + c_1 c_2\right] - 2Q^* - \frac{\delta_2^2 - \delta_1^2}{c_2^2 - c_1 c_2} \leq 0 \tag{2B.7}$$

Solving the quadratic form and using (2B.4):

$$\Rightarrow \quad \frac{1 - Z}{c_2^2 + c_1 c_2} \leq g(N_1^*, N_2^*) \leq \frac{1 + Z}{c_2^2 + c_1 c_2}$$

Where $Z = \sqrt{1 - \frac{(c_1 + c_2)(\delta_2^2 - \delta_1^2)}{(c_1 - c_2)S^2}}$. This is the first part of relationship (2B.2).

ad (iiib). 'No deviation of technology choice'. (A 1-technology firm deviating to the 1-technology). Following the same procedure as in (2B.6) gives:

$$\frac{Q^{*2}}{S^2}\left[c_1^2 + c_1 c_2\right] - 2Q^* + \frac{\delta_2^2 - \delta_1^2}{c_2^2 - c_1 c_2} \geq 0 \qquad \text{(2B.8)}$$

and therefore

$$g(N_1^*, N_2^*) \leq \frac{1 - Z}{c_1^2 + c_1 c_2} \quad \vee \quad g(N_1^*, N_2^*) \geq \frac{1 + Z}{c_1^2 + c_1 c_{20}} \qquad \text{(2B.9)}$$

Note that for the model to be valid $N_1 + N_2 \geq 2$. This implies for a MTE that $g(N_1^*, N_2^*) \geq g(1,1) = (c_1^2 + c_2^2)^{-1}$.

Since $g(N_1^*, N_2^*) \geq (c_1^2 + c_2^2)^{-1} \geq (1 - Z)(c_1^2 + c_1 c_2)^{-1}$, the first inequality of (2B.9) is always dominated by the oligopoly restriction. I.e.

$$g(N_1^*, N_2^*) \geq Max\left\{\frac{1}{c_1^2 + c_2^2}, \frac{1 + Z}{c_1^2 + c_1 c_2}\right\} \qquad \text{(2B.10)}$$

This is the second part of condition (2B.2).

Note that if the root of the quadratic form in (2B.7) is negative and hence Z is not defined, the condition in (2B.6) is always violated. I.e. there is always an incentive for a technology 2 firm to deviate to technology 1. A MTE cannot exist then. QED.

Lemma 2.3

(a) The STE $\{N_1^*, 0\}$ exists if $g(N_1^*, 0)$ satisfies:

$$\left[\frac{S - \delta_i}{Sc_i}\right]^2 \leq g(N_1^*, 0) \leq \frac{S - \delta_1}{Sc_1^2} \qquad i = 1, 2 \tag{2B.11}$$

and

$$g(N_1^*, 0) \geq Max\left\{\frac{1}{2c_1^2}, \frac{1 + Z}{c_1^2 + c_1 c_2}\right\} \tag{2B.12}$$

If Z is not defined, the STE exists if the other conditions of (2B.11) and (2B.12) are satisfied.

(b) The STE $\{0, N_2^*\}$ exists if $g(0, N_2^*)$ satisfies:

$$\left[\frac{S - \delta_i}{Sc_i}\right]^2 \leq g(0, N_2^*) \leq \frac{S - \delta_2}{Sc_2^2} \qquad i = 1, 2 \tag{2B.13}$$

and

$$Max\left\{\frac{1}{2c_2^2}, \frac{1 - Z}{c_2^2 + c_1 c_2}\right\} \leq g(0, N_2^*) \leq \frac{1 + Z}{c_2^2 + c_1 c_2} \tag{2B.14}$$

If Z is not defined, the STE does not exist. ‖

Proof.

(a) The first and second inequalities in relationship (2B.11) are respectively the 'No entry' conditions for either technology (ii) and the 'No exit' (i) conditions. Their derivation is identical to the one in Lemma 2.2. Relationship (2B.12) is derived along the lines of the proof of Lemma 2.2, under (iiib). The first factor is the oligopoly constraint $N_1^* \geq 2 \Rightarrow g(N_1^*,0) \geq g(2,0) = (2c_1^2)^{-1}$. The second factor is the non-deviation constraint of the technology choice and is identical to the derivation under Lemma 2.2, under (iiib). Note that if Z is not defined the inequality in (2B.8) is always satisfied and hence condition (iiib). The technology 1 firms then never face an incentive to deviate to technology 2, and there exists a STE if the conditions (2B.11) and (2B.12) are satisfied.

(b) The logic and algebra of the proof of this result is the same as the one under (a). In particular the derivation in Lemma 2.2, under (i), (ii) and (iiia) apply. Note that in the derivation of the non-deviation constraint in technological choice (iiia) the inequality (2B.7) is never satisfied if the root is negative and Z is not defined. In that case there cannot be a STE in which firms use technology 2. QED

Figures 2.5, 2.6 and 2.7 show the analytical restrictions that are relevant for the existence of MTE and STE. The numerical values of table 2.1 are used. In Figure 2.5 and 2.6 the existence of the technology 1 STE and the technology 2 STE are shown. As the value of $g(N_1^*,0)$ or $g(0,N_2^*)$ for $N_1^*, N_2^* = 2,3,4,..$ coincides with the shaded area, a STE $\{N_1^*,0\}$ and $\{0,N_2^*\}$ exists. The shaded area in Figure 2.7 is where MTE exist[22]. The evolution of the C in Figure 2.4 is consistent with the conditions as shown in Figures 2.5, 2.6 and 2.7.

[22] Note that the oligopoly condition implies that for the MTE $g(N_1^*,N_2^*) \geq g(1,1) = 40$, for the STE with technology 1: $g(N_1^*,0) \geq g(2,0) = 33\frac{1}{3}$ and for the STE with technology 2: $g(0,N_2^*) \geq g(0,2) = 50$.

Figure 2.5

Existence of STE: Technology 1.

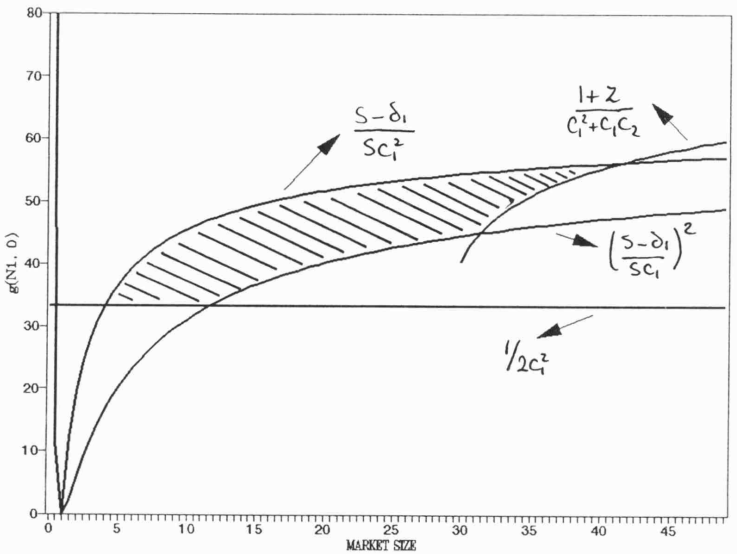


Figure 2.6

Existence of STE: Technology 2

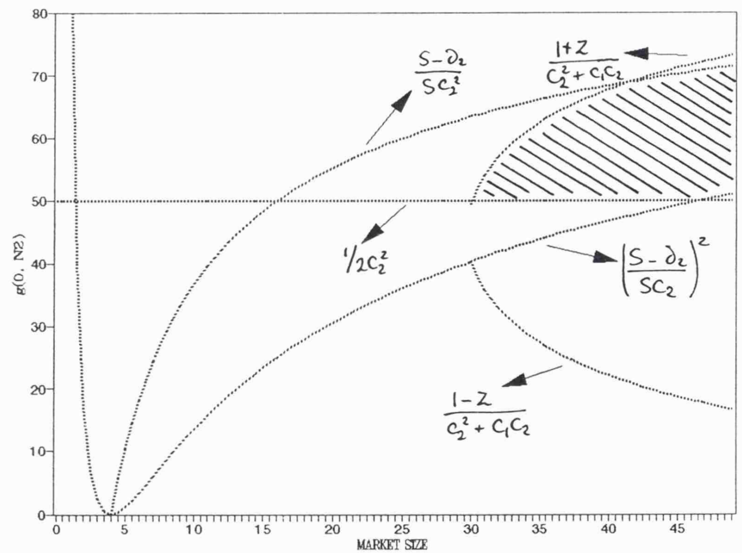
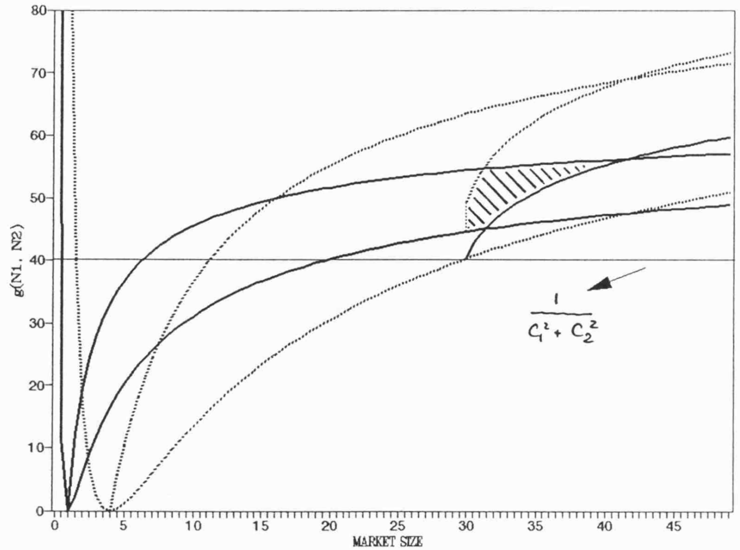


Figure 2.7

Existence of MTE



CHAPTER 3.

CROSS-SECTION FIRM DYNAMICS: THEORY AND EMPIRICAL RESULTS FROM THE CHEMICALS SECTOR.

3.1 Introduction.

Do firm characteristics converge or diverge? Do small firms grow faster than large firms, hence catching up, or do large firms have an inherent advantage in capturing new investment opportunities, dominating the industry in the long run? There is a growing literature addressing these issues[23], which, however, has generated very few robust results that can be tested empirically. Typically any outcome can be observed in equilibrium, depending on - generally unobserveable - rules of the game and specific parameter constellations, like the level of the discount rate[24].

The analysis of the dynamics of the industry structure goes back to the classic papers on stochastic firm growth by Gibrat (1931), Simon and Bonnini (1958) and Hart and Prais (1956). They showed that the typically skewed distribution of firm sizes can

---

[23] Major contributions have been made by Jovanovic (1982) and Cabral and Riordan (1994) on industry dynamics due to 'passive' learning, by Gilbert and Harris (1984) on the evolution of industry structure in a growing (Cournot) market, by Pakes and Ericson (1987) on the strategic investment in dynamic context, known as 'active' learning, and by Hopenhayen (1989) on dynamic competition between firms that face idiosyncratic cost shocks. A parallel literature on R&D and industry structure addresses similar issues although it is more specifically focused on the introduction of new products or technologies. A central question in the latter literature is which firm has the highest incentives to invest in R&D. Is it an entrant or the incumbent monopolist, the efficient or the inefficient firm? Two approaches have been used to address these issues. One is an auction, as in Gilbert and Newbery (1982), Vickers (1987) and Katz and Shapiro (1987), the other a stochastic race, as in Lee and Wilde (1980) and Reinganum (1983).

[24] Although Budd, Harris and Vickers (1992) find that competition tends to evolve in the direction where joint profits are higher, the implication for 'catching up' versus 'increasing dominance' depends on the exact nature of the pay offs. For example in Vickers (1986), the outcome is reversed if price competition is Bertrand rather than Cournot. Reinganum (1982) shows how Gilbert and Newbery's (1982) results that a monopolist will spend more on R&D than an entrant is reversed if the very same question is analyzed in a stochastic race, rather than a bidding game. See Lambson (1991) and Sutton (1995b) for a similar argument.

be generated by assuming independence of proportionate growth rates of current firm size and a fixed probability of entry by new firms. Recently, Sutton (1995a) analyzed the case in which the probability that an incumbent firm captures an investment opportunity of *fixed size* is independent of the current size of the firm. If the probability of capturing the opportunity is non-decreasing in current size, the implied time-stationary size distribution can be interpreted as a "least skewed" distribution. I.e. there exists a lower bound to the associated Lorenz curve, which will be violated if the probability of capturing the project is decreasing in size.

Here this benchmark of strategic independence is taken as a starting point and we ask the question in which direction the outcome changes if strategic interaction is modeled in its simplest form. If firms only differ in their initial size it turns out that it is the effect of winning on the firm's *own* price cost margin (PCM) that determines if small firms tend to catch up with larger ones or large firms dominate the industry in the long run. If the project increases the PCM, for example through lower marginal costs, then the large firm will generally win the project. If it decreases the PCM, then the firm with the small initial capacity is in a better position to win. If, for example, the market price is decreasing in the total market capacity, then installing new capacity decreases the PCM, putting the initially smaller firm in a better position to be the one to do so. Hence the difference between firm sizes are expected to decrease over time. We will call this effect the "demand side effect".

We show that this result holds across a number of alternative static game theoretic specifications. These are a "Grab the Dollar' game, an auction, and a stochastic race. The complexity of the modelling is deliberately kept to a minimum. Since it is well known that in more complex models "anything can happen" it is believed that the results that continue to hold in very simplified specifications are those, that are also most likely to hold empirically in a wide cross-section of industries. Moreover, there are well known examples in the literature of more complex dynamic specifications in which exactly these effects occur. Gilbert and Harris (1984), for example, find in a dynamic Cournot-Nash oligopoly with increasing demand and indivisibilities in installing new capacity, that market forces tend to push the industry towards equal market shares as smaller firms invest to catch up with larger ones. This is consistent with our results, since installing additional capacity implies a decreasing PCM for existing capacity. Farrell and Saloner (1988) and Beggs and Klemperer (1992) find that in the presence of both consumer switching costs and new customers arriving each period, market shares converge to equality, since the larger firm gains relatively more from charging a high price to exploit its current customer base than charging a low price to lock in new customers. The smaller firm can attract the new customers at lower (opportunity) costs, since the loss on its existing customer base from charging a lower price is smaller.

In Sections 3.3 to 3.5 we test the theoretical framework as an explanation for the dynamics of scale on product market level. If the investment project is opening a plant and the market price is decreasing in the total installed capacity, then the initially smaller business is in a better position to capture the opportunity, given the smaller externality on its existing capacity. This empirical prediction goes against much of standard textbook wisdom, in which the chemical sector is the archetypal example of increasing returns. Already large businesses have lower average costs and are therefore in a better position to capture the new investment opportunities, dominating the market in the long run. What we show is that the latter approach is not supported by the data. There is in fact a strong tendency of sizes of businesses to converge.

We use the population data of firms in 24 product markets of the chemicals sector between 1952 to 1983. The conventional way of testing the convergence hypothesis is to estimate the probability of a business being the next one to open a plant or capacity, which, according to the theory, should be decreasing in the business' initial size[25]. Although this test is intuitively appealing, it can be shown that the negative sign of the initial size does not necessarily imply that differences in firm sizes tend to become smaller[26]. We therefore use a methodology initially developed by Quah (1993a,b, 1994) to analyze convergence of per capita income across countries. This approach analyzes the dynamics of the *entire* cross-section distribution, exploiting the time series and cross section information more fully. We use stochastic kernels and transition matrices to characterise the intra-distributional mobility of businesses. We also analyze the long term behaviour of the size distribution of businesses and

[25] A similar test has been done by Gilbert and Lieberman (1987). Evans (1987a, b), Dunne, Roberts and Samuelson (1989) and others have calculated first the average growth rate over the sample period and then tested its dependence on initial size, mainly to show independence of growth rates and size, which is an implication of Gibrat's Law. See Chesher (1979).

[26] This is due to Galton's Fallacy, see section 3.3b.

find that in the period between 1952 and 1983 small firms have been more likely to increase capacity than larger ones, leading to a more fragmented industry structure.

In Section 3.5 we argue that there are theoretical reasons to believe that the converging tendencies are weaker the larger the number of incumbent businesses, as the effect of additional capacity on the market price is smaller, and hence the disadvantage large businesses face in capturing new investment opportunities is smaller. Empirically we find that if there are more businesses in the market, the tendency to converge is weaker. In fragmented product markets we find that businesses close to the mean firm size tend to move away from the mean, although there is only weak evidence of a bimodal long run (ergodic) distribution. There is, however, a strong suggestion that the long run distribution of sizes of businesses has a larger spread the more fragmented the industry structure.

We also test the effect of geographical location on the dynamics of scale. The more geographically spread production facilities are, the less exposed existing capacity is to the price decreasing effect of additional capacity. Essentially, large businesses can protect themselves by opening new capacity elsewhere in the country, given that it is technologically feasible and they are multi-plant firms. The limiting case is a set of local markets, geographically separated due to transport costs. We find empirical support for the hypothesis that the more concentrated the location of production facilities is in the country, the stronger is the tendency of sizes of businesses to converge. We also find that the higher the average number of plants in the industry, the weaker the converging tendencies as the firms have more scope to take advantage of spreading their production geographically.

It turns out that the extent to which small businesses were more likely to install new capacity relative to their larger competitors was much stronger in the 1953-1962 period than in subsequent years. It was the decade with both the highest growth rates and the most concentrated market structures. To disentangle the two effects, subsamples of industries with different growth rates but with similar numbers of businesses were constructed. The effect of growth rates is then analyzed by comparing the dynamics in both subsamples. Market growth does not seem to have an effect on the extent to which business sizes converge in the long run, if conditioned on the number of incumbents.

Since chemical corporations are typically multi-product firms, the firm dynamics are the result of the interaction of two components. That is (1) the dynamics of scale of production capacity on product market level we analyze in Sections 3.2 to 3.5 and (2) the dynamics of scope, or the dynamics of the number of products firms produce. In Section 3.6 we turn to the analysis of the dynamics of scope of chemical corporations. AS the analysis is on higher levels of aggregation, there is generally little substitution between differentiated chemical products on the demand side, at least relative to the homogeneous product markets we analyzed before. Therefore, the converging tendencies due to demand side effects are relatively small. We develop a rudimentary theoretical argument in which the dynamics of scope are driven by technological similarity of the products. If a firm produces products that in terms of production technology are similar to a new variety, then it will be more profitable for that firm to produce the new variety than it will be for a firm that only produces technologically unrelated products. I.e. if a firm has a larger presence in a particular industry in terms of number of products, then it is the more likely candidate to introduce the new variety. This is modeled as a higher probability that the new variety is "technologically close" to one of the firm's existing varieties. This effect, which we will call the "supply side" effect, leads to diverging firm sizes if size is measured in terms of the number of products a firm produces. We find empirical support for this, again using data from the chemical sector.

This analysis suggests that chemical corporations converge in terms of production capacity on a market by market basis, but diverge in terms of the number of products they produce. We conclude that these countervailing forces can serve as an explanation for the persistent nature of concentration levels in the chemical sector, as shown in Table 1.2

The remainder of this chapter is organized as follows. Section 3.2 derives the basic theoretical results. First a 'Grab the Dollar' game is analyzed, augmenting it with a particular pay off structure that depends on the size of firms. Then a similar procedure will be repeated to analyze the cross-sectional allocation in an auction and in a stochastic race. Section 3.3 describes the empirical implications of the theoretical analysis for the chemical sector and points out why the standard cross-section analysis can be misleading. In Section 3.4 we use a novel approach to study firm dynamics, in particular test the hypothesis that firm sizes on a market to market basis converge in the chemicals industry. In Section 3.5 the effects of the number of firms, the geographical location, the average number of plants and market growth on firm dynamics are analyzed empirically. Section 3.6 analyzes the dynamics of scope both theoretically and empirically. In Section 3.7 we discuss the implications and draw conclusions.

3.2 A Theoretical Model of the Dynamics of Scale.

In the literature on stochastic firm growth, industry growth is modeled as a discrete sequence of investment opportunities becoming available to the firms. These projects can be interpreted as involving the opening of a new plant, the introduction of a new product variety, etc. In Sutton (1995a), each opportunity is of the same size in terms of additional production, revenue and profits. The allocation of any project is purely stochastic, with a fixed probability that an entrant will take it up. This is the benchmark case of strategic independence, in which all incumbent firms have an equal probability of taking up the project. Here, the static version of this model is considered, analyzing the arrival of a single project in a simple game theoretic framework. The basic structure of the model is chosen to reflect the idea of this literature on stochastic firm growth as closely as possible. The focus of interest is on the factors that determine if a large or a small firm is more 'likely' to be the next firm to open a plant.

Two firms operate in a market. The large firm (1) has a historically given capacity q_1 , the small one (2) has a historically given capacity q_2 , where $q_1 > q_2 \geq 0$. If $q_2 = 0$, firm 2 is a (potential) entrant. Firm i earns PCM_i per unit of capacity. Assume that a profit opportunity arrives, which can be taken up by either firm. It can be opening another plant, a R&D project or an advertising campaign. We will say a firm 'wins' the project if it is allocated to that firm. The other firm is the 'loser'. The general features of the 'project' are:

Assumption 3.1 It is unique in the sense that it can only be realized by one firm, though both are potential candidates.

I.e. all firms are potential candidates for winning the project, which implies that its arrival is independent of firm characteristics.

Assumption 3.2 The winner receives (pays) a net "fixed profit" (cost) equal to $\pi \in \mathbb{R}$, which is identical for both firms.

Assumption 3.3 Externalities. The project, if won, changes the *PCM* of the winning firm by $\Delta PCM \geq -PCM_i \;\; i = 1,2$ (the externality) and the *PCM* of the loser by $\max\{\gamma\Delta PCM \;,\; -PCM_i\} \;\; i=1,2$ (the market externality). The market externality is assumed to be smaller in absolute value than the externality: $\gamma \in [-1,1]$.

The fixed profit π , which can be interpreted as the profit an entrant would earn, is independent of the firm size, and so are the externalities $(\gamma, \Delta PCM)$. The status quo PCMs might differ among the firms, i.e. $PCM_1 \neq PCM_2$. The validity of the assumption of the market externality being smaller in absolute value than the externality is an empirical issue and depends on the precise characteristics of the project and the industry. This will be discussed below.

Assumption 3.4 The firm that wins implements the project (no "shelving")[27].

The results presented here do not depend on intrinsic differences between the firms, only on differences in initial conditions.

If $\Delta PCM > 0$ and $\gamma > 0$, the project can be interpreted as an unique advertising project which generates positive spillovers for competitors. Alternatively, the project might be a product innovation or an improvement of existing technology, that is licensed to competitors or imitated by them[28]. The positive market externality might be due to diffusion of experience in the industry[29]. The project might be closing a

---

[27] This is primarily relevant for R&D projects, in particular those that can be patented. In the EU national laws typically provide for compulsory licences in two situations. The first is where an invention has not been available within the country to the extent of meeting national demands. The second is where the working of a subsequent invention is prevented by the prior patent.

[28] Foster (1985) describes a strategy "often used in the chemical industry. [BASF] developed a catalyst, which it then improved. The first generation of the catalyst went to its licencees, the second and improved generation into its own plants" (p.119). Katz and Shapiro (1987) quote a number of studies in which competitors imitate innovations at a lower cost than the innovator faced. Although the licencee might have to pay a fee, or imitation involves some R&D expenses, it is generally recognised that this is less than the development cost, and can be normalized to zero within the framework of this model.

[29] Mansfield (1985) finds evidence of a high rate of information diffusion in several industries.

plant in an industry, producing imperfect substitutes ($\gamma \in]0,1[$) or homogeneous goods ($\gamma = 1$).

If $\Delta PCM > 0$ and $\gamma < 0$, the project can again be interpreted as an advertising or (product or process) innovation project, but with negative spillovers. An innovation that is patented might give the holder a competitive advantage and undermine the competitive position of other firms. Even if it would be licensed, the competitors can face a lower PCM due to the royalty fees. An extreme example is where an innovation is 'drastic' in the sense that innovator i can monopolize the market[30]. Then $\gamma \Delta PCM = -PCM_i$.

The constellation $\Delta PCM < 0$ and $\gamma > 0$ occurs for example if one firm adds capacity in a market of products that are imperfect substitutes, or homogeneous goods if $\gamma = 1$. Alternatively, it can be what Katz and Shapiro (1987) call a "major" innovation, an innovation that replaces the existing technology and production capacity based on that technology cannot be operated economically any longer. Then $\Delta PCM = -PCM_i$, where firm i is the innovating firm.

Finally $\Delta PCM < 0$ and $\gamma < 0$. An extreme version of this case is firm i exiting from an industry producing imperfect substitutes or homogeneous goods ($\Delta PCM = -PCM_i$). If the production facilities are dismantled, the capacity reduction will increase the price cost margin of the firms that remain in the industry.

There are at least two examples where $\gamma \in [-1,1]$ is unlikely to hold. One is where goods are vertically differentiated and the introduction of a new variety can (in product space) be located close to the competitor's variety and far from the winner's own existing varieties, hence hurting the profitability of the competitor's varieties more than his own. The other is in the context of horizontal product differentiation. The project is opening a new plant, but there are transport costs. The plant can be located close to a competitor's existing plant, hence immunizing itself from the externality on its existing capacity. In these cases the market externality is higher in absolute value than the externality ($\gamma > 1$).

The focus of this section will be to determine which firm wins the project under which circumstances, i.e. for which parameter constellations

[30] See Arrow (1962).

$(\Delta PCM, \pi, \gamma, q_1, q_2)$ does either firm end up with the project. This will be analyzed in the context of three different games. First a 'Grab the Dollar' game, in which firms choose whether or not to grab the project. The second is an auction in which the firms bid for the project. The third game is a stochastic race, in which firms make strategic investments that increase the probability of winning.

3.2a Grab the Dollar.

Consider a one shot game in which a project is to be allocated (a dollar on the table). The strategy space of the two asymmetric agents is to "grab" the project (G) or to "refrain" (R). The rules of the game are that if only one firm grabs that firm wins the project; if both firms bid, firm 1 will win with probability F . If neither grabs the opportunity it is lost and the PCMs are unchanged. The pay off of firm i is $\pi + (PCM_i + \Delta PCM)q_i$ if it wins, and $(PCM_i + \gamma\Delta PCM)q_i$ if it looses. The pay off is $PCM_i q_i$ if none of the firms bid for the project. Grabbing itself is costless.

(Expected) Pay offs in the Grab the Dollar Game.

		Firm 2	
		Grab $G2$	Refrain $R2$
Firm 1	Grab $G1$	$F\pi + \big(F(1-\gamma) + \gamma\big)\big(PCM_1 + \Delta PCM\big)q_1$ $(1-F)\pi + \big(1 + F(\gamma-1)\big)\big(PCM_2 + \Delta PCM\big)q_2$	$\pi + (PCM_1 + \Delta PCM)q_1$ $\gamma(PCM_2 + \Delta PCM)q_2$
	Refrain $R1$	$\gamma(PCM_1 + \Delta PCM)q_1$ $\pi + (PCM_2 + \Delta PCM)q_2$	$PCM_1 q_1$ $PCM_2 q_2$

A symmetric pure strategy Nash equilibrium in which both firms grab {G1, G2} exists if grabbing has a higher expected pay off than refraining, given that the competitor grabs:

$$\gamma\Delta PCMq_i \le H\big(\pi + \Delta PCMq_i\big) + \big(1 - H\big)\big(\gamma\Delta PCMq_i\big) \quad i = 1, 2$$

$$\Rightarrow \quad \pi \ge (\gamma - 1)\Delta PCMq_i \tag{3.1}$$

where $H = F \quad if\ i = 1$
$H = 1 - F \quad if\ i = 2$

An equilibrium in which neither firm bids {R1, R2} exists if for both firms grabbing yields a negative pay off, given that the competitor refrains:

$$\pi + \Delta PCMq_i \leq 0 \qquad i = 1, 2 \tag{3.2}$$

The conditions for asymmetric equilibria in which firm i bids and j refrains {Gi, Rj} are:

- for firm i it is optimal to bid, given j refrains:

$$\pi + \Delta PCMq_i \geq 0 \Rightarrow \pi \geq -\Delta PCMq_i \tag{3.3a}$$

- for firm j it is optimal to refrain, given that i bids:

$$\gamma\Delta PCMq_j \geq H(\pi + \Delta PCMq_j) + (1 - H)(\gamma\Delta PCMq_j)$$

.

$$\Rightarrow \pi \leq (\gamma - 1)\Delta PCMq_j \tag{3.3b}$$

where $H = F \quad if\, j = 2$.
$H = 1 - F \quad if\, j = 1$

Figures 3.1 and 3.2 show the full set of pure strategy equilibria for $\Delta PCM > 0$ and $\Delta PCM < 0$ respectively. If $\Delta PCM > 0$ the large firm is more willing to incur a high cost ($\pi << 0$) than the small firm, given γ . Hence there exist parameter constellations for which the unique asymmetric equilibrium is {G1, R2}. More importantly, if $\Delta PCM > 0$ there are no parameter constellations for which there is a unique equilibrium in which only the small firm bids. If $\Delta PCM < 0$ then there do not exist parameter constellations for which the unique equilibrium is one in which only the large firm bids. The small firm is less affected by the negative externality and hence is willing to grab for lower levels of the level of π , given γ .

There are three types of equilibria. One is if either firm can win, the second where none wins in equilibrium and the third in which there is a unique equilibrium winner. We consider the last one first. A unique equilibrium winner exists if there is a unique asymmetric (pure strategy) equilibrium.

Proposition 3.1

(i) If $\Delta PCM > 0$ and there exists a unique asymmetric pure strategy equilibrium, then the large firm wins.

(ii) If $\Delta PCM < 0$ and there exists a unique asymmetric pure strategy equilibrium, then the small firm wins.

Proof. There exists an equilibrium in which only the large firm bids if conditions (3.3a) and (3.3b) are satisfied (i = 1, j = 2):

$$-\Delta PCMq_1 \le \pi \le (\gamma - 1)\Delta PCMq_2 \tag{3.4a}$$

And there exists an equilibrium in which only the small firm bids if (i = 2, j = 1)

$$-\Delta PCMq_2 \le \pi \le (\gamma - 1)\Delta PCMq_1 \tag{3.4b}$$

If $\Delta PCM < 0$ then the area defined by (3.4a) is a subset of the area defined by (3.4b) and the only the small firm (2) bids if there is a unique asymmetric equilibrium. If $\Delta PCM > 0$ then the area defined by (3.4b) is a subset of the area defined by (3.4a) and only the large firm (1) bids if there is a unique asymmetric equilibrium. QED.

The figures show that larger differences in initial firm sizes amplify this effect in the sense that the areas for which there exists a unique asymmetric pure strategy equilibrium are larger.

Proposition 3.1 holds if the "no shelving" assumption (3.4) is relaxed. If winning firms have the discretion not to implement the project, the optimal response of a refraining firm in an asymmetric equilibrium becomes:

$$\gamma\Delta PCMq_j \ge H Max\{\pi + \Delta PCMq_j, 0\} + (1 - H)(\gamma\Delta PCMq_j) \tag{3.3c}$$

where $\begin{matrix} H = F & if\, j = 2 \\ H = 1 - F & if\, j = 1 \end{matrix}$.

Firm j only implements the project if the pay off is non-negative. If $\pi + \Delta PCMq_i \ge 0$, then (3.3c) reduces to (3.3b). If $\pi + \Delta PCMq_i \le 0$, (3.3c) becomes

$$\gamma\Delta PCM < 0$$

The consequence is that the areas defined by (3.3a) and (3.3b) are no longer unique asymmetric equilibria, since the refraining firm will be better off grabbing the project and shelving it if won.

Now relax the assumption that the market externality is smaller than the externality, i.e. $\gamma \notin [-1, 1]$. Proposition 3.1 goes through for $\Delta PCM < 0$, but not for $\Delta PCM > 0$. If $\gamma > 1$ and $\Delta PCM > 0$, there exist parameter constellations for which {R1, G2} is an equilibrium, but {G1, R2} is not. This occurs if

$$(\gamma - 1)\Delta PCMq_2 \leq \pi \leq (\gamma - 1)\Delta PCMq_1 \tag{3.5}$$

and the area defined by relation (3.3b) is not any longer a subset of the area defined by relation (3.3a)

We now return to the original game and consider the outcomes in which either firm can win. These can be symmetric equilibria, multiple asymmetric equilibria or mixed strategy equilibria. Which firm is more likely to win the project in the symmetric equilibrium {G1, G2} depends on F. A weak auxiliary assumption would be that the outcome that leads to a lower industry profit is the less likely one. If firm i wins the project, the industry profit is $\Pi_i = \pi + \Delta PCM(q_i + \gamma q_j)$.

Proposition 3.2
Assume that $F \geq \frac{1}{2}$ iff $\Pi_1 \geq \Pi_2$. In symmetric equilibria {G1, G2} the large firm is more likely to win the project if $\Delta PCM > 0$ and the small firm is more likely to win the project if $\Delta PCM < 0$.

Proof. The conditions for {G1, G2} are:

$$\gamma \Delta PCMq_i \leq H(\pi + \Delta PCMq_i) + (1 - H)(\gamma \Delta PCMq_i)$$

$$\Rightarrow \pi \geq (\gamma - 1)\Delta PCMq_i \tag{3.6a}$$

where $H = F \quad if\ i = 1$
$H = 1 - F \quad if\ i = 2$

Since $\Pi_i = \pi + (q_i + \gamma q_j)\Delta PCM$, it follows that

$$\Pi_i - \Pi_j = (1 - \gamma)(q_i - q_j)\,\Delta PCM \tag{3.6b}$$

If $\Delta PCM \geq 0$ then $\Pi_1 \geq \Pi_2$ and hence $F \geq 1/2$.
If $\Delta PCM < 0$ then $\Pi_1 < \Pi_2$ and hence $F < 1/2$.
QED.

The intuition is as before. The outcome that generates the highest industry profits is - here by assumption - the most likely. As long as the market externality is smaller in absolute value than the externality, the equilibrium industry profits are the highest if the large firm wins if $\Delta PCM > 0$ and if the small firm wins if $\Delta PCM < 0$.

There exist parameter constellations for which there are multiple asymmetric pure strategy equilibria. One possible way around this indeterminacy would be to assume that with probability F the equilibrium is played in which firm 1 wins. Assuming that $F \geq \frac{1}{2}$ iff $\Pi_1 \geq \Pi_2$, this would imply that the large firm is more likely to end up with the project if $\Delta PCM > 0$, whereas the small firm is more likely to end up with the project if $\Delta PCM < 0$. Note that in the auction, that will be described in the next section, the indeterminacy is fully resolved in favour of the outcome obtained here using an 'ad hoc' argument.

There is a caveat. Mixed strategy equilibria do not generate any clear cut results concerning which firm is more likely to win the project under which circumstances. The probability of firm i grabbing in a mixed strategy equilibrium if $F = 1/2$ is $p_i = \frac{2(\pi + \Delta PCM q_j)}{\pi + (\gamma + 1)\Delta PCM q_j}$, which is increasing or decreasing in q_j , depending on the sign of $\Delta PCM \gamma \pi$.

Figure 3.1

Unique Equilibria if $\Delta PCM < 0$.

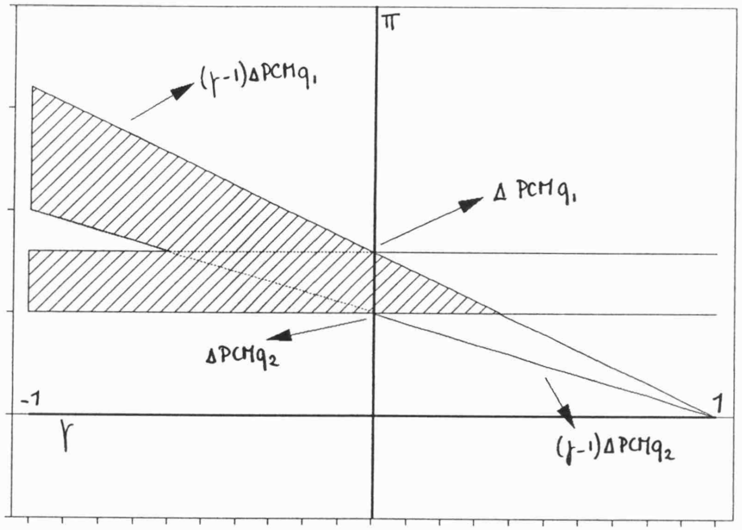
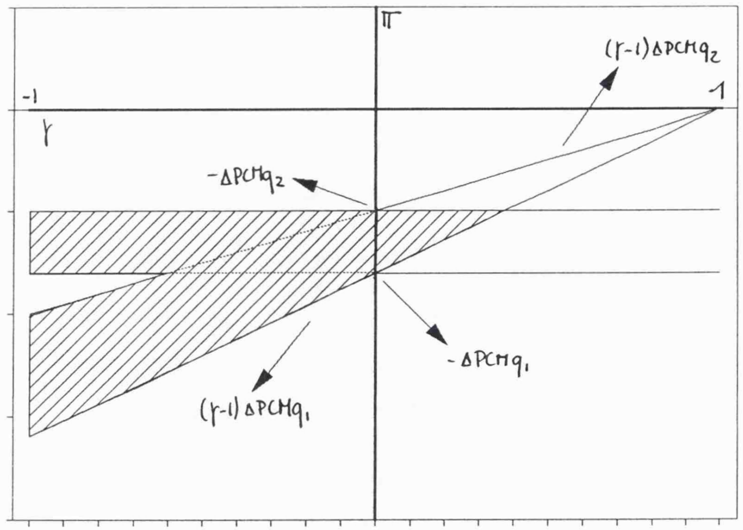


Figure 3.2

Unique Equilibria if $\Delta PCM > 0$.



3.2b Selection in an Auction.

We now turn to a second game, which is essentially a (Dutch) auction with discounting. The dynamic nature of the game allows for a considerably richer way of describing the incentives for the firms to grab the project than before, thereby refining the intuition of the results in the last section. The project is won by the large firm if $\Delta PCM > 0$ and by the small one if $\Delta PCM < 0$. What will be shown is that, with identical discount rates, the large firm is more eager to implement it if $\Delta PCM > 0$, since his opportunity cost of waiting is higher. Secondly, the opportunity cost of losing is higher for the large firm if $\gamma \in [-1, 1]$, hence giving him an incentive to preempt. A similar reasoning holds mutatis mutandis for the small firm if $\Delta PCM < 0$.

The set up of the game follows Katz and Shapiro (1987). However, their analysis is augmented by Assumptions 3.1-3.4.

Technically, the game is a stopping game. As before, there are two firms with historically given capacity $q_1 > q_2 \geq 0$. Decisions are made at discrete dates, at t = 0, δ , 2δ , ... , where $\delta \to 0$. The strategies of the firms are to 'grab', given that no firm has yet grabbed the project, or to wait. Grabbing means winning the project, developing it and realizing the pay off[31]. Waiting means not grabbing at t and deciding again at $t+\delta$. The game ends as soon as one firm grabs the project. If neither of the firms grabs at any finite time, firm i earns PCM_i per unit of existing capacity, generating a continuous stream of $PCM_i q_i$ forever. The incremental profit for the winner consists of a fixed stream of profits (cost), with initial present value equal to $\pi(t) > 0$ (< 0) if the project is grabbed at t. Its current value is continuously differentiable and increasing over time, though at a decreasing rate $\left(\frac{d\pi e^{rt}}{dt} > 0 \ , \ \frac{d^2\pi e^{rt}}{dt^2} < 0\right)$ with finite limit $\lim_{T\to\infty}\pi(T)e^{rT} = \pi^{\infty}$, for example because development costs fall at a decreasing rate over time, or opening a plant becomes more and more profitable due to growing demand. The PCM the winner earns on its existing capacity is changed by ΔPCM . If firm i wins, then, as before, there is an market externality on firm j's profits (the loser). Firm j's PCM changes by $\gamma\Delta PCM$, where $\gamma \in [-1, 1]$.

Both firms have an identical discount rate r. The present value of the pay off of

---

[31] If both firms grab at the same time, either firm will win with equal probability.

winning at time T is:

$$
\begin{aligned}
W_i(T) &= \pi(T) + \int_T^{\infty} e^{-rt}\,\Delta PCMq_i\,dt \\
&= \pi(T) + \frac{e^{-rT}}{r}\Delta PCMq_i \qquad i = 1,2
\end{aligned}
\tag{3.7}
$$

Assume that winning the project initially is not profitable for either firm, i.e. $\pi(0) + \frac{\Delta PCMq_i}{r} < 0 \quad i = 1,2$.

The present value from losing at time T is:

$$
L_j(T) = \frac{e^{-rT}}{r}\gamma\Delta PCMq_j \quad j = 1,2 \tag{3.8}
$$

The equilibrium concept is a subgame perfect equilibrium, confining the analysis to pure strategy equilibria.

Two basic incentives determine the outcome. If firm j never grabs, then firm i's incentive to grab at any date depends upon the pay off from winning only. Following Katz and Shapiro (1987), we call the incremental profit of the winner $W_i(T)$ the "stand alone" incentive. Firm i is willing to grab at T or any time after, if $W_i(T) \geq 0$, since the actual value of winning, $W_i(T)e^{rT}$, is increasing in the grabbing date T. The optimal date to grab if j will never grab, (the "stand alone date") $\hat{T}_i$ is the solution of:

$$
\Delta PCMq_i = e^{r\hat{T}_i}\pi'(\hat{T}_i) \tag{3.9}
$$

The RHS of (3.9) is decreasing in $\hat{T}$ [32]. Therefore, if $\Delta PCM > 0$ firm 1 has an earlier stand alone date than firm 2 and vice versa if $\Delta PCM < 0$.

[32] This is the case because $d^2\pi e^{rt}/dt^2 = d\pi'(t)e^{rt}/dt + rd\pi e^{rt}/dt < 0$, and since $d\pi e^{rt}/dt > 0$, it must be the case that $d\pi'(t)e^{rt}/dt < 0$. The SOC $\pi''(t) + re^{-rt}\Delta PCM < 0$ is satisfied for $t = \hat{T}$, since by (3.9) it is equivalent to $\pi''(\hat{T}) + r\pi'(\hat{T}) < 0$. This is implied by $d\pi'e^{rt}/dt < 0$.

Lemma 3.1
If $\Delta PCMq_i > -r\pi^{\infty}$ for $i = 1,2$, then $\hat{T}_1 < \hat{T}_2$ if $\Delta PCM > 0$ and $\hat{T}_2 < \hat{T}_1$ if $\Delta PCM < 0$.

If $\Delta PCM > 0$ the large firm is more eager to implement the project than the small firm, since the large firm (1) is losing more by a further delay of the implementation of the project, although the firms have identical discount rates. It is this impatience that induces the large firm to implement the project earlier than the small firm would. A similar argument holds for the small firm if $\Delta PCM < 0$.

The other incentive is what Katz and Shapiro (1987) call the "incentive to preempt", which is the difference in profits from existing capacity between winning and losing, $(1 - \gamma)\Delta PCMq_i$. That is, firm i is willing to preempt at T or any date thereafter if for $t \geq T$ $W_i(t) \geq L_i(t)$, $t > T$, even if it is before its stand alone date $(t < \tilde{T}_i)$. The "earliest preemption date" of firm i, $\tilde{T}_i$ is the solution of $W_i(t) = L_i(t)$:

$$-(1 - \gamma)\Delta PCMq_i = r\pi(\tilde{T}_i)e^{r\tilde{T}_i} \tag{3.10}$$

There exists a unique earliest preemption date if $(1 - \gamma)\Delta PCMq_i > -r\pi^{\infty}$, since $W_i(t) - L_i(t)$ is increasing in t. Firm 1 has higher preemption incentives whenever $\Delta PCM > 0$, and because the RHS of (3.10) is increasing in t, firm 1 earliest preemption date is before firm 2's. If $\Delta PCM < 0$, by the same token, firm 2 has the higher preemption incentives and is the first to reach its earliest preemption date.

Lemma 3.2
If $(1 - \gamma)\Delta PCMq_i > -r\pi^{\infty}$, $i = 1,2$ then $\tilde{T}_1 < \tilde{T}_2$ if $\Delta PCM > 0$ and $\tilde{T}_2 < \tilde{T}_1$ if $\Delta PCM < 0$.
This result is driven by the prospect of the firm being worse off if it loses than it had been had it won the project. If $\Delta PCM > 0$ and $\gamma \in [-1,1]$ the large firm has higher preemption incentives than the small firm because the opportunity loss from not wining the project is higher for the large firm. Similarly for the small firm if $\Delta PCM < 0$.

Summarizing, if $\Delta PCM > 0$, then firm 1 has both an earlier stand alone and preemption date. Katz and Shapiro (1987) show that in this case firm 2 cannot win in equilibrium, since firm 1 would always preempt. Similarly, firm 2 will always preempt if $\Delta PCM < 0$ and firm 1 cannot win. Hence Proposition 3.3 is a corollary of their result:

Proposition 3.3

If the selection mechanism is the above stopping game, then:

(i) the large firm is the equilibrium winner if $\Delta PCM > 0$,

(ii) the small firm is the equilibrium winner if $\Delta PCM < 0$.

Proof. The result follows from Lemmas 3.1 and 3.2, and from Katz and Shapiro's (1987) necessary condition for equilibrium no. 5 (p. 407). QED.

If neither firm has a finite stand alone date (i.e. $\Delta PCMq_i < -r\pi^{\infty}, i = 1,2$), there is always an equilibrium without grabbing. If in addition both firms have a finite earliest preemption date (i.e. $(1 - \gamma)\Delta PCMq_i > -r\pi^{\infty} \geq \Delta PCMq_i \, , i = 1,2$), there is a second equilibrium outcome in which the firm with the higher preemption incentives preempts before the earliest preemption date of its competitor (see Katz and Shapiro (1987) - Theorem 1b). These are "self defence" equilibria. Both firms would prefer not to grab, and only do so because the other one does. It can be shown that they can only occur if the market externality is negative, i.e. $\gamma\Delta PCM < 0$ [33].

Before turning to the analysis of the stochastic race, we compare our results to two related models that have been described in the literature. One is Gilbert and Newbery's (1982) model of preemptive patenting, which is essentially a second-price auction with an incumbent monopolist and an entrant bidding for a substitute product. If the monopolist wins, he remains the sole firm in the market. If the entrant wins, the market becomes a duopoly, which reduces the profit of the monopolist. They find that the monopolist will win if entry results in any reduction of total profits below the joint maximizing level. In the unlikely case that the introduction of the substitute by the monopolist increases the profit margin of the existing variety, this is consistent with our results. The more plausible case is, however, that the substitute will decrease the profit margin of the existing variety,

[33] The two conditions are: (i) $\Delta PCMq_i \leq -r\pi^{\infty} \quad i = 1,2$ (neither has a finite $\hat{T}_i$) and (ii) $-r\pi^{\infty} < (1-\gamma)\Delta PCMq_i \quad i = 1,2$ (both have a finite $\tilde{T}_i$). If $\Delta PCM > 0$, (i) and (ii) can only be satisfied simultaneously if $\gamma < 0$, and if $\Delta PCM < 0$ they can only both be satisfied if $\gamma > 0$.

though to a lesser extent than if the entrant would have introduced the variety. In our model this would mean that $\gamma > 1$, which is ruled out by assumption. This example shows, however, that by assuming $\gamma \in [-1, 1]$ economically interesting cases might have been ignored. Furthermore, Gilbert and Newbery (1982) show that "sleeping patents" might occur if the monopolist reduces his overall profit as a consequence of the introduction of the patent. In our model the implementation of the project was assumed.

Katz and Shapiro (1987) consider an R&D project with a time dependent development cost. The loser faces no development costs, but earns a (different) profit flow due to imitation or licensing. The profit flows from winning and losing differ per firm. By choosing the appropriate parameter constellations all possible rankings of stand alone date and earliest preemption date can be generated. Hence either firm winning the project can be an equilibrium outcome. Restricting the pay offs by introducing Assumptions 3.1-3.4 effectively rules out the case in which firm i has greater stand alone incentives, but firm j has greater preemption incentives. Unsurprisingly, the introduction of additional assumptions reduces the number of equilibrium outcomes. However, the interpretation of the model as given here is more general in the sense that the fixed part of the profits can be either a cost or a profit and the externality on the profitability of existing capacity can be either positive or negative. The firms' pay off of winning is nevertheless concave in the grabbing date throughout, as it consists of either a time-increasing fixed profit with a given negative externality or a time-decreasing fixed cost with a given positive externality.

3.2c Selection in a Model of a Stochastic Race.

In this section we analyze a model of a stochastic race, which is similar to Loury (1979), Lee and Wilde (1980) and Reinganum (1983). We will show that the results derived earlier hold qualitatively in this framework. If the pay off structure satisfies Assumptions 3.1-3.4, the large firm has a higher investment rate if $\Delta PCM > 0$ and the smaller firm has a higher investment rate if $\Delta PCM < 0$.

Consider an industry with two firms. As before, firm 1 has a large historically given capacity, firm 2 a small one, so $q_1 > q_2 \geq 0$. They are competing to be the first to win the project. Once a firm wins, the game ends.

The strategy of firm i is to select an investment rate z_i, that determines the probability density that firm i will win at any t. The success date of firm i is a random variable T_i , distributed according to:

$$Pr(T_i \leq t) = G_i(t) = 1 - e^{-h(z_i)t} \tag{3.11}$$

where $z_i \geq 0$ and $h(.)$ is the hazard rate. Assume that the hazard rate is twice continuously differentiable, with $h'(\cdot) > 0, h''(\cdot) < 0, h(0) = 0$ and $\lim_{z \to 0} h'(z) = \infty$, $\lim_{z \to \infty} h'(z) = 0$. The firm commits to a particular level at the start of the race, and pays z_i until one of the firms wins. Until the first success date the flow of profits of firm i is $PCM_i q_i - z_i$. If firm i wins, its flow changes by $\pi + \Delta PCMq_i + z_i$ and firm j's profit flow by $\gamma \Delta PCMq_j + z_j$. Let r again be the common discount rate. The expected profit of firm i as a function of its own and the rival's investment rate is

$$V_i(z_i, z_j) = \int_0^{\infty} e^{-rt} e^{-(h(z_i) + h(z_j))t} \left[h(z_i) \left[\frac{\pi + \Delta PCMq_i}{r} \right] + h(z_j) \left[\frac{\gamma \Delta PCMq_i}{r} \right] - z_i \right] dt$$

$$= \left[h(z_i) \left[\frac{\pi + \Delta PCMq_i}{r} \right] + h(z_j) \left[\frac{\gamma \Delta PCMq_i}{r} \right] - z_i \right] \frac{1}{r + h(z_i) + h(z_j)} \tag{3.12}$$

The probability density of firm i winning at t is $h(z_i)e^{-[h(z_i) + h(z_j)]t}$, generating a pay off of $\frac{\pi + \Delta PCMq_i}{r}$. The probability density of firm j winning is $h(z_j)e^{-[h(z_i) + h(z_j)]t}$, generating a pay off for firm i of $\frac{\gamma \Delta PCMq_i}{r}$. With probability $e^{-[h(z_i) + h(z_j)]t}$ neither firm has won before t and firm i pays z_i .

Proposition 3.4

If the selection mechanism is a stochastic race then in equilibrium

(i) the winning date of the large firm stochastically dominates the winning date of the small firm in the sense of first order stochastic dominance if $\Delta PCM > 0$.

(ii) the winning date of the small firm stochastically dominates the winning date of the large firm in the sense of first order stochastic dominance if $\Delta PCM < 0$.

Proof. See Appendix 3A.

As before, if $\Delta PCM > 0$, both the stand alone and the preemption incentives are higher for the large firm. If the small firm is indifferent between winning and losing, the large firm strictly prefers to win. Hence the large firm has always an incentive to preempt. Moreover, the opportunity cost of waiting is higher for the large firm than for the small one, which makes the larger less patient. The reverse argument holds if $\Delta PCM < 0$.

In the auction model the incentive to preempt dominates the firm's decision as long as the stand alone incentive is non-negative. The main difference here is that preemption is stochastic. Consequently, the impatience of firms to implement the project also becomes relevant for the outcome. But since the larger firm has uniformly higher incentives to invest if $\Delta PCM > 0$ it has a higher probability of winning. By assuming this particular pay off structure both incentives are always aligned.

This is of interest as the Gilbert and Newbery (1982) deterministic auction model and the Reinganum (1983) stochastic race model can give opposite outcomes. In the former an incumbent firm preempts, whereas in the latter the potential entrant has a higher R&D effort. This can only occur if one firm has higher preemption incentives and the other has higher stand alone incentives.

3.3 Application to the Chemical Sector.

In this section we will analyze the theoretical implications of the dynamics of scale in the chemicals sector, which will be tested in the next section. The principal reason for taking this sector is that investment projects can be unequivocally defined as opening or closing a plant or production capacity within a plant. Firms in the chemical sector are typically multi-product firms. The level at which this interpretation of the theoretical framework applies is therefore on business units.

The industries in the data set (see Table 3.1) are typically bulk chemicals, hence homogeneous by nature. Most are intermediate or final petrochemicals. The size of the business unit is measured by its production capacity in a particular market. The advantage that homogeneous products have over differentiated products in testing the theory is that the effect on the price cost margin can be determined under weak conditions. If the market price is decreasing in the total market capacity, then opening capacity will decrease the market price and. Hence, in terms of the terminology of the last section, it must be that $\Delta PCM < 0$ if capacity is increased and $\Delta PCM > 0$ if capacity is reduced, with $\gamma > 0$ throughout. In an industry with differentiated products a typical "project" might be, for example, a combination of increased advertising and increased production. If the former increases the price cost margin and the latter decreases it, the net effect is indeterminate and the empirical implication of the theory unclear.

Within the chemical industry capacity is certainly not the only strategic choice businesses face. R&D programmes are essential in the strategic interaction among businesses (Quintella (1993)). Research in manufacturing technology has resulted in less expensive raw material, such as in the production Acrylonitrile and Vinyl Acetate. In Phenol, a more efficient process based on Cumene Hydroxide has been developed. In some cases the feedstock has changed. For example Phtalic Anhydride used to be produced from Naphtalene, which then changed to ortho-Xylene. Research has focused on increases in size of existing plants, primarily by de-bottlenecking, and by "scaling up" of entire production processes. For example, in the early 1950s the largest Ethylene plants had a capacity of about 100 million pounds per year. In the 1970s the newly constructed Ethylene plants produced well over 1 billion pounds per year (Spitz (1988, Chapter 11)). Similar developments have occurred for Ammonia, Vinyl Chloride, Styrene and Methanol. Hence the "project" of opening a plant or increasing capacity changed over time. This allows for the possibility that businesses, depending on their R&D programme, faced different sets of opportunities in terms of opening additional capacity. In terms of the model the "fixed pay off" π might be business specific, due to businesses using different technologies. However, most production technologies are non-proprietary, particular in petrochemicals. Anecdotal evidence suggests that the diffusion of new production processes is quick which is, according to Spitz (1988), due to engineering contractors learning how to build the scaled-up plants or how to apply new production techniques, which makes the technology available to whoever was willing to pay for it (see p. 424). Any business that opened a new plant, be it an entrant, a small incumbent or a large incumbent, seemed to use the state of the art technology, suggesting that new production technologies were widely available. Mansfield (1985) found that in the petroleum sector 60% of the process technology was available to competitors within 18 months of a business's decision to develop a major new process. This effect was even more pronounced in primary metals, though less in other chemicals. Spitz (1988) described the effect of new technologies on petrochemical industries as follows (p. 393):

If the new route represented a substantial economic improvement, but was not judged to be able to provide a dominant position, the company making the invention usually embarked on a licensing program, settling for the income provided by royalties and catalyst sales, as well as the presumed benefits of becoming a reasonably low cost producer. In other cases the company could not establish a controlling position, because it could not obtain broad patent protection to keep competitors from developing relatively similar process routes. In still other cases, such as Badger-Sherwin Williams' fluid-bed Phtalic Anhydride process, the new technology was not so much better that it forced a wave of shutdowns. Here, the new technology just added one or two new competitors and upgraded the economics of some of the existing producers, who switched to the new process.

Well known exceptions include Du Pont in Titanium Dioxide (Ghemawhat (1984)), where it achieved a dominant position through its proprietary Chloride technology and its position in nylon, which it achieved through selective licensing. Sohio achieved a dominant position in Acrylonitrile, based on a revolutionary propylene technology (Stobough (1987)).

Furthermore, π or ΔPCM might be business specific due to "increasing returns to scale". Although it is well known that there are increasing returns in the chemical sector, they seem to occur primarily on plant level rather than on business unit level. Both Spitz (1988) and Stobough (1987) show significant gains from increasing plant sizes in terms of reducing per unit production costs. The effect of the size of businesses on production costs seems negligible. However, ownership of businesses might matter since firm sizes do seem to play a role in the availability of capital to finance large scale production facilities. But, if there is no significant relationship between the size of an business unit and the firm size, this should not affect the results. Moreover, the typical chemical firm is so large that availability of capital due to lack of credibility does not seem a real issue.

Learning is frequently mentioned as an alternative source of increasing returns. However, in petrochemicals these gains are not business specific according to Spitz (1988, Ch. 10) and Stobough (1987, Ch. 5), since again rapid diffusion of experience throughout the industry undermines any competitive advantage. Although there is evidence of significant "industry wide" learning (Lieberman (1984)), individual companies do not seem to be able to maintain an advantage through more experience.

We therefore claim that it is a reasonable first approximation to assume that in the capacity game all businesses face an equal investment opportunity ("project"), and hence that businesses only differ in their existing capacity.

Another assumption that has to be satisfied for the theoretical results to hold is that the market externality is smaller in absolute value than the externality on the winning business's existing capacity $(-1 < \gamma < 1)$. Spitz (1988, p. 540) describes how prices are often cut as new capacity comes on stream, due to businesses giving discounts in order to fill new capacity. The effect on competitors is likely to be less than the full discount due to transportation costs, which are significant even though most of the US petrochemical capacity is located in the Gulf Coast region (Chapman (1991, Chapter 6)). However, location is still relatively dispersed, due to the dependence of the US petrochemical industries on natural gas liquids as feedstock, rather than oil based raw materials as in Europe. An extensive network of pipelines give businesses considerable freedom in their locational choice without giving up nearness to raw material sources. Oil based feedstock would instead require the location close to refinery complexes and hence a higher degree of geographical concentration.

A priori it is not clear which game is the most appropriate description of the cross-sectional allocation of "projects". However, all three that were described in the last section have a qualitatively identical empirical implication. Hence the *Empirical Hypothesis* can be formulated independently of the precise game.

Empirical Hypothesis:
In growing and declining chemical industries, sizes of businesses tend to converge.

In growing industries there is a sequence of arrivals of new investment projects, each of them being the opening of another plant. This decreases the PCM earned on existing capacity by the winning business, and from Propositions 3.1 - 3.4 it follows that initially small businesses is more likely to win the project. Closing down a plant reduces capacity, increasing the PCM the winning business earns on its remaining capacity. By Propositions 3.1 - 3.4 the business that is larger ex-post is more likely to implement the project[34].

[34] Ghemawat and Nalebuff (1990) find a qualitatively similar result. The mechanism is, however, somewhat different. In a declining industry the optimal size of firms declines over time. Large firms have reached this optimal size and reduce their capacity accordingly. However, their smaller competitors have a suboptimal size and only reduce their capacity once the optimal size is smaller than their actual size.

Table 3.1: Summary Statistics for the Chemical Industries.

Product	First Obs.	Number of Firms Min	Max	Growth Rate	SIC Code	GGINI	Sample*
Acrylic Fibres	1953	3	6	10.3	28242		0, 1
Acrylonitrile	1956	4	6	10.1	28692		0, 1, 6
Aluminum	1956	4	13	4.7	3334	0.61	0, 2, 4
Ammonia	1960	40	64	5.8	28731	0.48	3
Aniline	1961	4	6	11.3	28651	0.52	0, 1, 4
Bisphenol A	1959	3	5	14.2	.		0, 1
Caprolactam	1962	2	4	11.4	28696	0.38	0, 1, 5
Chlorine	1961	30	40	4.6	28121	0.34	3
Cyclohexane	1956	2	14	9.0	28651	0.52	4
Ethyl Alcohol	1958	5	11	4.1	28695	0.49	5
Ethylene	1960	20	26	8.7	29116		3
Ethylene Glycol	1960	9	14	6.0	28696	0.38	0, 2, 5, 6
Formaldehyde	1962	12	18	5.9	28692		0, 2, 6
Isopropyl Alcohol	1964	3	4	3.2	28692		0, 1, 6
Magnesium	1954	1	4	3.1	3339	0.38	5
Maleic Anhydride	1958	3	8	8.6	2865		0, 1
Methanol	1957	8	12	7.7	2869		0, 2
Nylon Fibres	1960	5	23	8.8	28241	0.22	5
Pentaerythritol	1952	4	7	4.6	28696	0.38	0, 1, 5, 6
Phenol	1959	8	13	7.2	28651	0.52	0, 2, 4
Phosphorous Pentasulfide	1965	3	4	5.5	28199	0.36	5
Phthalic Anhydride	1955	6	14	5.6	28651	0.52	0, 2, 4
Polyethylene-LD	1957	8	15	11.1	28213	0.54	0, 2, 4
Polyethylene-HD	1957	2	14	32.7	28213	0.54	0, 2, 4
Sodium Chlorate	1956	3	11	6.9	28197	0.30	0, 1, 5
Sodium Hydrosulfite	1964	3	6	3.5	28197	0.30	0, 1, 5
Sorbitol	1955	2	5	5.9	28696	0.38	0, 1, 5, 6
Styrene	1958	7	13	8.5	28651	0.52	0, 2, 4
Titanium Dioxide	1964	5	6	2.3	28161		0, 1
1,1,1-Trichloroethane	1966	3	4	12.5	.		0, 1, 6
Urea	1960	12	36	9.3	28732	0.28	3
Vinyl Acetate	1960	4	7	11.7	28692		0, 1, 6
Vinyl Chloride	1962	9	14	12.0	28692		0, 2, 6

* Subsample 0 is used in Sections 3.4, 3.5 and Appendix 3B. This is the sample Gilbert and Lieberman (1987) used. Subsamples 1-5 are used in Section 3.5. Subsample 6 is used in column (7), Table 3B.1.

3.3a The Data.

We use capacity data of a sample of 33 industries, most of which produce homogeneous bulk chemicals - see Table 3.1. The majority consists of intermediate or final petrochemicals, although some primary metals and inorganic chemicals are also included. The first year of observation is between 1953 and 1965, while the last is 1983 for all industries. The capacity data are from annual issues of the Directory of Chemical Producers (SRI International), reporting firm and plant capacities of all US producers by product[35].

The data set used in the next section is the subsample used by Gilbert and Lieberman (1987), referred to as sample 0 in Table 3.1. It consists of 24 growing chemical industries. Net output for all products increased from the earliest observation until at least 1975. The sample includes industries with more than three but less than twenty firms. It excludes products for which there is either joint production, or capacity can easily switched from one product to another[36]. Hence it is a sample of industries that serve oligopolistic, growing, homogeneous product markets. If convergence of business sizes is to be observed anywhere, then this sample seems a good candidate.

3.3b The Standard Cross Section Analysis of Convergence.

The empirical hypothesis is tested for growing industries, using two different approaches. The first one, presented in Appendix 3B, is a conventional logit analysis of the probability of business i opening a plant. We find indeed a negative coefficient for the initial condition, though the result is quite unstable. It sounds intuitively appealing to interpret this negative sign as businesses within one industry converging towards a common size. However, a negative sign can be consistent with a growing dispersion in the cross-section distribution of business sizes. This is due

---

[35] See Lieberman (1982), and Gilbert and Lieberman (1987) for a more detailed description of the data set.

[36] See Gilbert and Lieberman (1987). Although output was often consumed elsewhere within the same firm, for all products at least 25% was sold.

to Galton's Fallacy or Regression to the Mean[37]. The proof of the Fallacy for discrete choice models is in Koopmans and Lamo (1995), showing that a negative cross-section coefficient for the initial level (as in Tables 3B.1) is consistent with absence of convergence.

Consequently, it has been argued[38] that the standard deviation of the cross-section distribution should also be considered, suggesting that both a negative coefficient of the initial condition and a decreasing cross-section dispersion over time would be sufficient to show convergence. But these statistics (mean and standard deviation) only capture information about the dynamics of the cross-section distribution if this distribution can be satisfactory described by its first two moments. If not, then mean and standard deviation only imperfectly capture the distribution dynamics, in particular catching-up, changes in the shape of the distribution and intra-distribution mobility.

3.4 The Empirical Analysis of Cross-Section Dynamics.

A more natural way of dealing with convergence is therefore to consider the dynamic behaviour and the cross-section variation of the entire size distribution. We consider an alternative empirical strategy which was suggested by Quah (1993a,b & 1994). It deals with both time-series and cross-section dimensions, based on what in probability theory is called *Random Fields*. These are data structures that have variation of the same order of magnitude in both dimensions. At each point in time there is a cross-section distribution of business sizes, which is simply the realization

[37] See Quah (1993b), Friedman (1992), Huigen et al. (1991), Hall (1987) and Leonard (1986). Strictly speaking this Fallacy refers to growth equations, with the initial size as one of the explanatory variables. Usually the average growth rate is determined and then regressed on the initial condition. See for example Mansfield (1962), Hymer and Pashigan (1962), Singh and Whittington (1975) and Kumar (1985).

[38] In particular by some authors in growth theory, see for example Barro and Sala-i-Martin (1992).

of a random element in the space of distributions. The idea is to describe their evolution over time, which will allow us to analyze intra-distribution mobility, persistence of the business' relative position, and to characterise the long run behaviour. In this framework convergence is understood as the sequence of distributions tending towards a mass-point in the long run.

3.4a The Variable of Analysis.

The theory suggests as the basic variables of analysis the business's size in terms of installed capacity (q), relative to the average business size in the industry. For firm i in industry j this is $CS_t^{ij} = q_t^{ij} \div \Sigma_i q_t^{ij} / N_t^j$ if there are N_t^j firms in the industry at time t. Alternatively business sizes can be measured in terms of the number of plants a business operates (PLS), in which case the variable of analysis is the business's number of plants relative to the industry average.

The central conclusion of the theory for the chemical sector is that small businesses are more likely to install new capacity than their larger competitors. A somewhat stronger implication is that differences between the size of a business and the industry average should shrink over time, and possibly go to zero in the long run.

For CS and PLS convergence is understood as the sequence of their cross-section distribution tending towards a mass point at unity. The normalization is a way to control for overall growth and aggregate fluctuations of the industry, heteroscedasticity, and it allows pooling of industries[39].

There is an important issue of potential entry. We have no indication for how long businesses have been around, waiting in the wings, before they enter. We will make alternative assumptions to test the robustness of the empirical results w.r.t. the various entry assumptions. For example we assume that an entrant has been around

[39] It can be shown that convergence in terms of CS and PLS implies that the absolute difference between business size and industry average goes to zero if the average size is bounded from above.

for seven years before opening its first plant. Given a construction lag of 2 years[40], we include the business in the sample as a potential entrant (with 0 plants) for five years, obviously taking into account the starting year of the sample for that industry[41]. These are CS5 and PLS5. Similarly, though assuming that entrants could have opened a plant for two years only, are CS2 or PLS2.

The empirical analysis has been performed for both CS and PLS under the alternative entry assumption. Figures 3.3a-b, show three dimensional plots of CS2 and PLS2[42]. It is clear that in all cases both time-series and cross-section variations are large, illustrating the importance of studying both dimensions if analyzing dynamics of scale.

3.4b Cross-section Distribution of the Variable.

In the context of random fields the realization of the random element is a cross-section distribution function that can be estimated from the data. Figures 3.4a-d present the cross-section density functions of PLS2 for each period of 3 or 4 years. They have been estimated by non-parametric methods for the available sample of industries[43]. No assumption has been made about the shape nor about the moments of the density function from which the data were drawn. Since our dataset contains population data for each industry, the full cross-section size distribution can be estimated. From these figures the limitations are clear of describing density functions by their first and second moment only, as has been suggested in the literature.

During the 32 years of the sample there is a tendency of PLS to concentrate around

---

[40] See Lieberman (1987).

[41] I.e. if the period between the beginning of the sample for that industry (see Table 3.1) and the opening of the first plant is shorter than five years, accordingly fewer zeros were included.

[42] All the calculations and graphics in this section have been made using Danny Quah's Time-Series Random-Field shell tsrF.

[43] The estimates were obtained using a gaussian kernel with an automatically selected bandwidth (Silverman (1986, 3.4.2)).

the average industry size. However, there are two limitations of the analysis of distribution functions in this context. One is that convergence is generally a limit concept and the cross-section distributions are point in time estimates, available only for 1952-83. Hence we cannot say anything about the long run behaviour of the size distributions. Furthermore, the graphs do not give any information about a business unit's relative position and its movement over time. To deal with these limitations, it is necessary to derive a law of motion for the cross-section distribution in a more formal structure.

3.4c Modelling Dynamics of the Cross-section Distribution.

Let λ_t be the probability measure (one for each year) associated with the cross-section distribution. The simplest probability model that can describe its dynamic behaviour is a first order autoregressive process:

$$\lambda_t = T^*(\lambda_{t-1}, u_t) \tag{3.13}$$

T^* maps the probability measures and a disturbance into another probability measure. Hence T^* encodes information on how the businesses move over time relative to each other. By ignoring the disturbance and iterating, (3.13) can be written as:

$$\lambda_{t+s} = (T^*)^s \lambda_t \tag{3.14}$$

As s goes to infinity, the long run (ergodic) distribution of sizes of business units can be characterised.

The stochastic difference equation in expression (3.13) is unmanageable, but so is (3.14). Given the impossibility of analytic solutions for T^*, we will assume T^* is being generated by the following differential equation:

$$\lambda_{t+1} = \int M(x, A)\, d\lambda_t(x) \tag{3.15}$$

For any probability measure λ on the measurable space ($\mathbb{R}$, R), where $\mathbb{R}$ is the real line and *R* is the Borel sigma algebra, $\forall$ A in R. M is a *Stochastic Kernel*[44], that is, M(x,A) is the probability that the state next period lies in A given that in this period the state is x. T^* is an operator associated with the *Stochastic Kernel* that maps the space of probabilities into itself, and $\lambda_{t+1}(A) = (T^*\lambda_t)A$.

Equation (3.15) measures the probability that the next period's state lies in set A, if the current state is drawn according to the probability measure λ_t. And $(T^*\lambda_t)$ is the probability measure over next period's state, if λ_t is the probability measure over the current period.

The *Stochastic Kernel* allows us to analyze the intra-distribution movements of business units, solving one of the limitations pointed out, but leaving the problem of the analysis of the long run behaviour unresolved, because the *Stochastic Kernel* is infinite dimensional. We can, however, simplify the problem by approximating T^* assuming a countable state space for businesses sizes $S = \{s_1, s_2, .., s_r\}$. In that case T^* is simply a transition probability matrix Q, which makes the difference equation (3.13) tractable[45].

$$\lambda_t = Q(\lambda_{t-1}\ ,\ u_t) \qquad (3.16)$$

Q encodes the relevant information about mobility within the cross-section distribution. But the ergodic distribution of (3.16) can be calculated explicitly. Under some regularity conditions the sequence of powers of matrix Q converges to a matrix which has identical rows describing the ergodic cross-section distribution. This allows us to analyze the long run behaviour of the size distribution.

3.4d Estimation of the *Stochastic Kernel*.

Figures 3.5a-d, 3.6a-c and 3.7a-d are three dimensional plots of some *Stochastic*

---

[44] See Stokey and Lucas (1989).

[45] See Adelman (1958) for an early application of transition matrices in the analysis of firm dynamics.

Kernels for PLS5, PLS2 and CS2, estimated non-parametrically[46]. They describe the transitions from one state to any other in 1 and 5 years respectively. Figures 3.5e-h, 3.6d-f and 3.7e-h present the contours of the kernels in 3.5a-d, 3.6a-c and 3.7a-d respectively.

A slice orthogonal to the plane (t,t+k) and parallel to the t+k axis, represents the probability density that describes the transitions from one point of the time t distribution to another in k periods. The probability mass concentrated along the positive sloped diagonal indicates a high persistence in a business's relative position. A concentration of probability mass along the negatively sloped diagonal implies that businesses overtake each other in size rank. The transition probability describing horizontal lines (parallel to t+k) indicates that there is very low persistence, the probability of being at any point in t+k is independent of the position in t. Finally the mass of probabilities located along a vertical line in size 1 (the industry average) implies convergence in the sense that small businesses grow faster than large ones.

The theoretical results are consistent with the probability mass being both along the negative diagonal, implying a "action - reaction" pattern of opening a plant by alternating businesses and along a vertical line around 1, in which case there is convergence of sizes in a stricter sense.

The graphs show persistence year by year, indicating that the business units remain in their relative position. Particularly for businesses in size class zero, which represents the potential entry state. Not surprisingly, this effect is more pronounced for PLS5 than for PLS2 or CS2.

The results are much more striking for the larger (5 year) horizon, where the probability of transition is no longer clustered along the positive diagonal but along the vertical line in 1, indicating convergence to the industry average. After 5 years,

---

[46] They are obtained using the square of the standard Epanechnikov kernel for estimating the joint density and then re-scaling to obtain the conditional probability (tSrF.).

businesses that were potential entrants initially, will have a positive share of the market. A large probability mass is concentrated under the positive diagonal at zero in period t, indicating that entrants in period t reach the average size (1) in period t+5.

The contours show that in the first decade (between 1955 and 1965) the tendency of the sizes of businesses to converge is the strongest. The estimated kernels are consistently steeper than later estimates, indicating more persistence in subsequent decades. It is worth noting that this corresponds to the decade in which the industries saw their largest market growth and the fewest number of businesses in the industry[47]. This is encouraging for the theory since two essential features are that (i) capacity in the market is increasing and (ii) there is a negative effect of additional capacity on the market price, that decreases as markets become more competitive. This will be analyzed in greater detail in the next section.

Note that these transition kernels are simply point in time estimates, describing what actually happened over the sample period. They are not fitted models. Hence we cannot derive a law of motion, or make any inferences about the long run dynamic behaviour. To deal with these shortcomings, we turn now to the analysis of transition matrices.

3.4e Estimation of the Transition Matrix Q.

The transition matrix Q is analogous to the *Stochastic Kernel,* but in a discrete space. Divide the space of possible values of the sizes of businesses into *r* states. For example, businesses that have a plant share of 0.2 times the industry average to 0.6 times the industry average are in state i = (0.2,0.6). This defines a grid that can be thought as an estimator of the initial unconditional probability distribution λ_{t-k}. Each element of the matrix indicates the probability of transition from one state to

[47] Between 1955 and 1965 the average of annual industry growth rates of market capacity was 14.6%, the average number of firms 6.8. Between 1965 and 1975 the former was 9.3%, the latter rose to 8.01. Between 1975 and 1983 the growth rate dropped to 4.4%, whereas the average number of firms was 7.7.

another: the entry (i,j) is the probability that a business in state i moves to the state j in t periods. Hence every row is a conditional probability vector or the discrete analog of the distribution of the transitions in the figures above, the *Stochastic Kernels*, when cutting the kernel at a point by a plane parallel to t+k axis.

Tables 3.2, 3.3 and 3.4 present some estimators of the transition matrix Q. The grid divides the initial year total observed sample into approximately equal categories, i.e a uniform initial distribution of sizes of business units results by construction. Consequently the length of the defined states varies. Note that they are very narrow around the mean.

The top row of each table gives again the number of states, the second row the upper end of each of them. The first column give the total number of transitions starting from each state over the entire time sample. An estimator of the time invariant transition probability matrix Q is presented in the remaining columns, for a single year transition. Estimates are given for differing measures of size of business units, making alternative assumptions about potential entrants as before, and for differing number of states *r*. Q is calculated as a time average over the sample period. In most cases 1954 is taken as a base year, rather than 1952, in order to increase the number of observations that define the grid (see Table 3.1).

Most of the entries of the matrices are different from zero implying that a transition to almost any state in the distribution can occur within one year. Hence there is substantial mobility. The (negative) diagonal entries are higher in the extreme states than they are in the states closer to the mean. This implies that there is higher mobility around the mean than there is in the tails of the distributions, although this is probably an artefact of the differences in the sizes of states, which are narrower around the mean.

According to the theory we would expect that the transition probabilities from lower to higher states decrease the bigger the business units are already, and vice versa for smaller business units. However, it is rather difficult to extract any general conclusions on this level.

Looking at the long run behaviour reveals more about the dynamics. The last rows of Tables 3.2, 3.3 and 3.4 show estimates of the ergodic distribution associated with the transition matrices. Independently of the initial position of a business unit, the ergodic distribution gives the probability of that business being in a particular state. Recall the states were defined in such a way that the initial distribution is uniform. Though the ergodic distributions are not degenerate at 1, they are unimodal with a peak around this value. Whatever of the position of a business in the initial uniform distribution, the probability of ending close to the average size is higher than the probability of ending up anywhere else.

Table 3.5 shows an estimate of Q for the subsample 1954 to 1964. Comparison the ergodic distribution for this subsample with the one estimated using the full sample confirms the earlier finding that the tendencies to converge are strongest during the years of strong industry growth in industries with a small number of businesses. We also estimated kernels and transition matrices for subsample of industries, leaving out those industries for which there was a proprietary production technology (i.e. Acrylonitrile, Caprolactam and Titanium Dioxide). Qualitatively the conclusions remained unchanged.

In this section we showed that size of production capacity of chemical corporations on product market level tend to converge market by market. This is the first step in the test of the empirical hypothesis. Convergence is a necessary but not sufficient condition for the theoretical framework of Section 3.2 to hold. A number of alternative theories have exactly the same empirical implication. Notably, the existence of an technologically optimal size of the production facility for a business with convex adjustment costs would imply converging dynamics if businesses enter the industry sequentially. Alternatively, the driving force behind convergence could be managerial diseconomies. In the next section we argue that there are two factors that should influence the tendency to converge if the mechanisms presented in Section 3.2 are the relevant ones. We show then that they affect the tendency to converge empirically in the predicted way.

Table 3.2 First Order Transition Matrix, CS2
Time-Stationary, 1954 - 1982

Upper end	**0.236**	**0.440**	**0.657**	**0.903**	**1.284**	**1.877**	**∞**
(r)	(1)	(2)	(3)	(4)	(5)	(6)	(7)
752	*0.72*	*0.09*	*0.05*	*0.05*	*0.04*	*0.03*	*0.01*
754	*0.09*	*0.76*	*0.09*	*0.02*	*0.03*	*0.00*	*0.01*
742	*0.01*	*0.15*	*0.69*	*0.10*	*0.04*	*0.01*	*0.00*
760	*0.01*	*0.01*	*0.19*	*0.67*	*0.09*	*0.03*	*0.00*
777	*0.01*	*0.00*	*0.01*	*0.19*	*0.67*	*0.12*	*0.01*
757	*0.00*	*0.01*	*0.00*	*0.01*	*0.18*	*0.70*	*0.10*
738	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.13*	*0.87*
Ergodic	.058	.131	.151	.166	.172	.167	.155

Table 3.3 First Order Transition Matrix, CS5
Time-Stationary, 1954 - 1982

Upper end	**0.000**	**0.427**	**0.745**	**1.153**	**1.868**	**∞**
(r)	(1)	(2)	(3)	(4)	(5)	(6)
1009	*0.79*	*0.05*	*0.04*	*0.05*	*0.04*	*0.02*
926	*0.07*	*0.85*	*0.07*	*0.01*	*0.00*	*0.00*
937	*0.03*	*0.10*	*0.77*	*0.09*	*0.01*	*0.00*
992	*0.02*	*0.00*	*0.12*	*0.73*	*0.11*	*0.01*
963	*0.00*	*0.00*	*0.00*	*0.15*	*0.76*	*0.08*
931	*0.00*	*0.00*	*0.00*	*0.00*	*0.12*	*0.88*
Ergodic	.100	.162	.177	.194	.197	.170

Table 3.4 First Order Transition Matrix, PLS5
Time-Stationary, 1956 - 1982

Upper end	0.333	0.800	0.917	1.000	1.500	∞
(r)	(1)	(2)	(3)	(4)	(5)	(6)
969	*0.76*	*0.08*	*0.06*	*0.05*	*0.03*	*0.02*
973	*0.06*	*0.77*	*0.11*	*0.03*	*0.02*	*0.00*
989	*0.03*	*0.15*	*0.64*	*0.15*	*0.02*	*0.01*
979	*0.03*	*0.01*	*0.20*	*0.61*	*0.13*	*0.02*
883	*0.02*	*0.01*	*0.02*	*0.19*	*0.69*	*0.07*
928	*0.00*	*0.00*	*0.01*	*0.02*	*0.08*	*0.88*
Ergodic	.111	.180	.196	.189	.160	.164

Table 3.5 First Order Transition Matrix, CS5
Time-Stationary, 1954 - 1964

Upper end	0.000	0.229	0.556	0.862	1.311	2.194	∞
(r)	(1)	(2)	(3)	(4)	(5)	(6)	(7)
288	*0.78*	*0.03*	*0.06*	*0.06*	*0.04*	*0.03*	*0.00*
87	*0.03*	*0.84*	*0.13*	*0.00*	*0.00*	*0.00*	*0.00*
171	*0.02*	*0.04*	*0.82*	*0.10*	*0.02*	*0.00*	*0.00*
180	*0.01*	*0.00*	*0.08*	*0.74*	*0.16*	*0.02*	*0.00*
161	*0.01*	*0.00*	*0.00*	*0.12*	*0.72*	*0.14*	*0.01*
181	*0.01*	*0.00*	*0.00*	*0.01*	*0.15*	*0.73*	*0.10*
182	*0.00*	*0.00*	*0.01*	*0.00*	*0.00*	*0.12*	*0.88*
Ergodic	.033	.029	.108	.163	.230	.227	.210

Figure 3.3a

Time Series and Cross-Section Dimensions

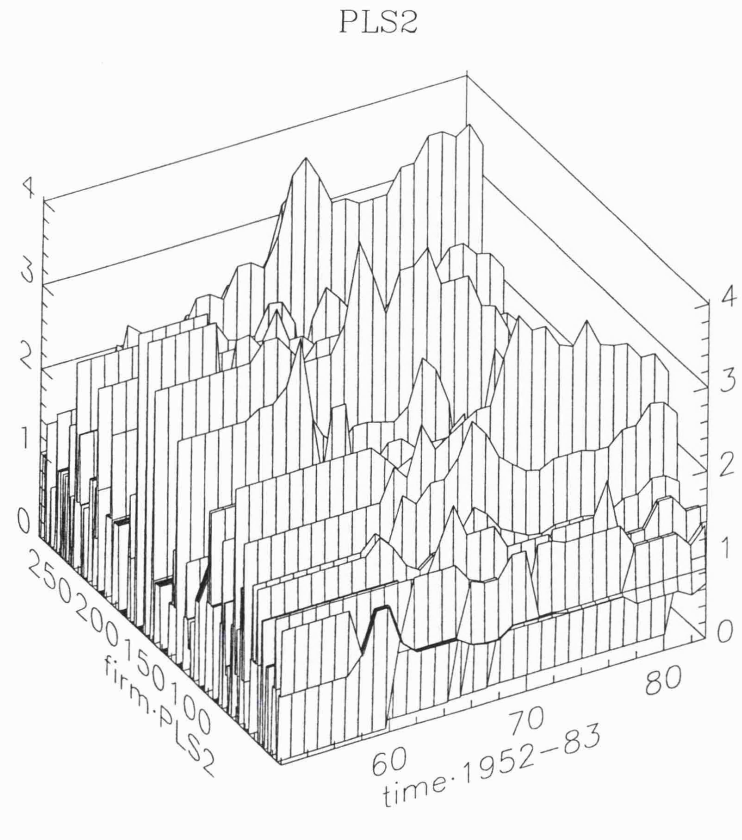


Figure 3.3b

Time Series and Cross-Section Dimensions

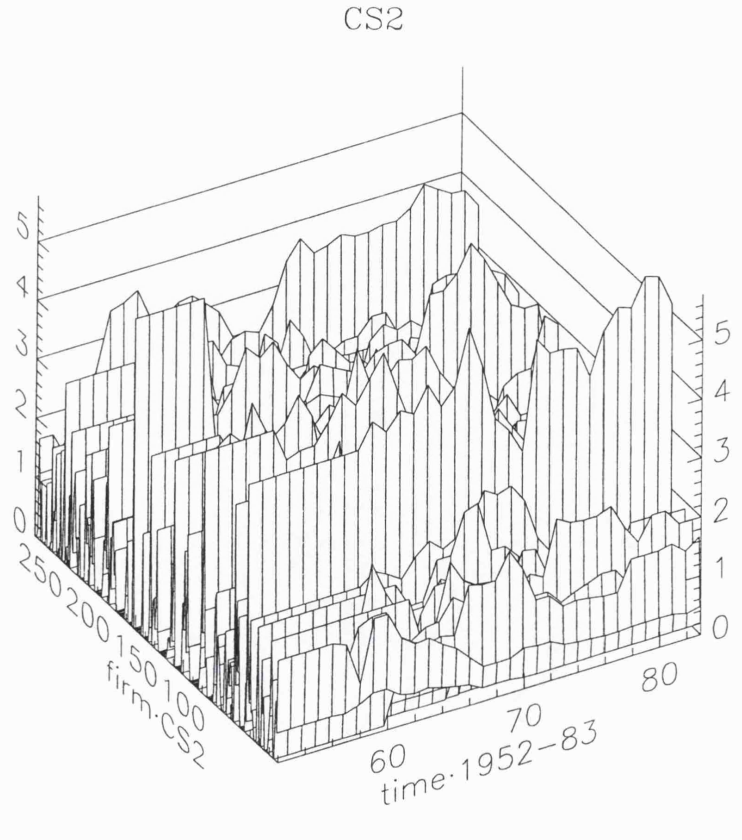


Figure 3.4a

Estimated Density of PLS2 (1954-1957)

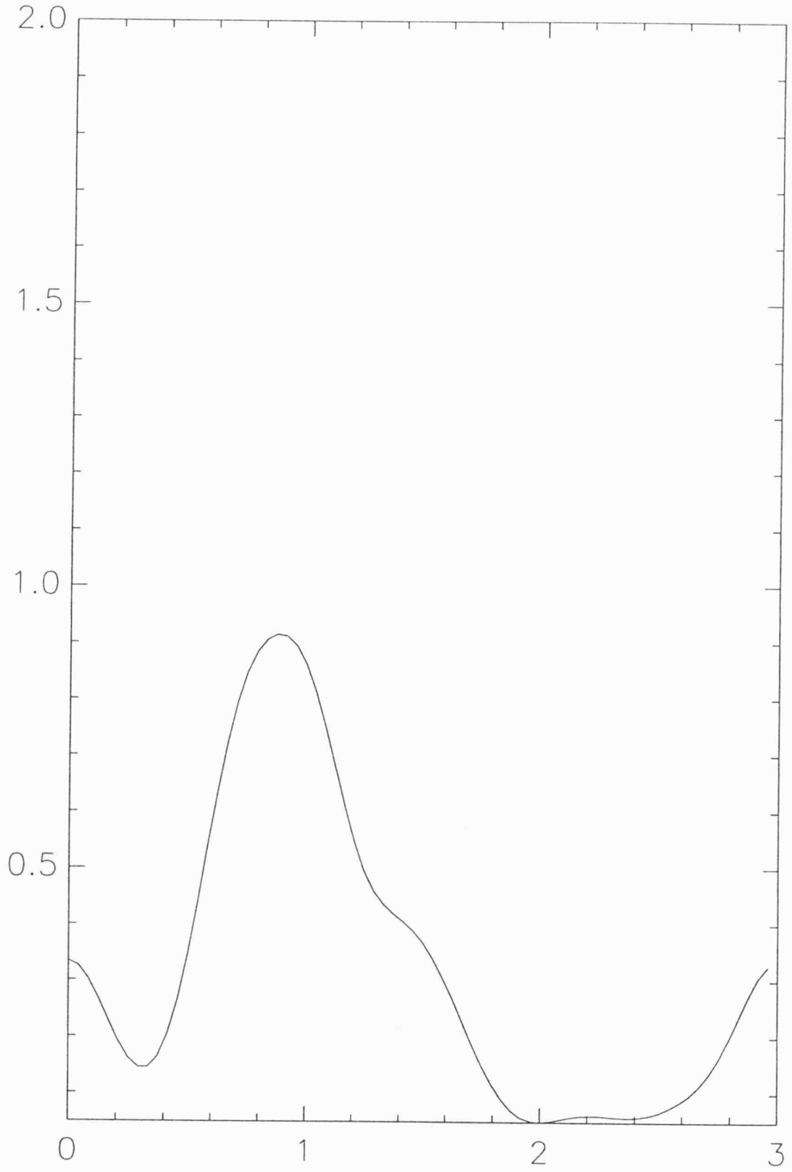


Figure 3.4b

Estimated Density of PLS2 (1961-1964)

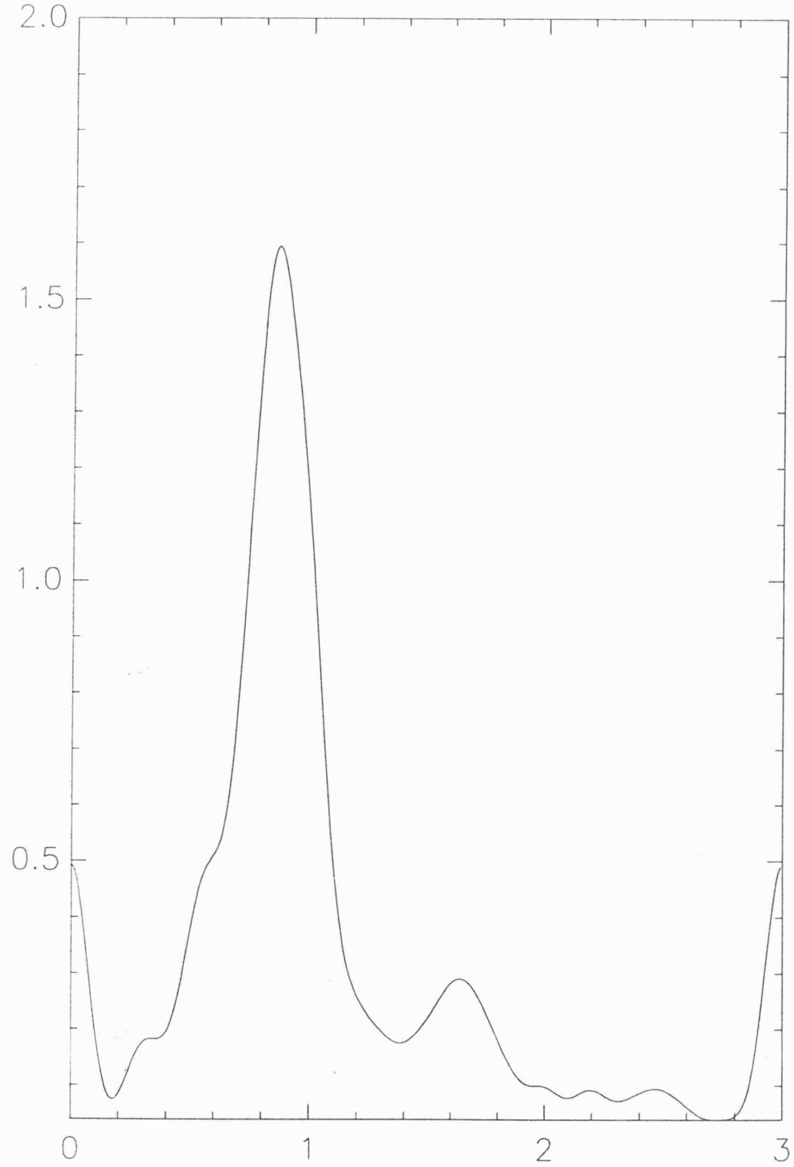


Figure 3.4c

Estimated Density of PLS2 (1969-1972)

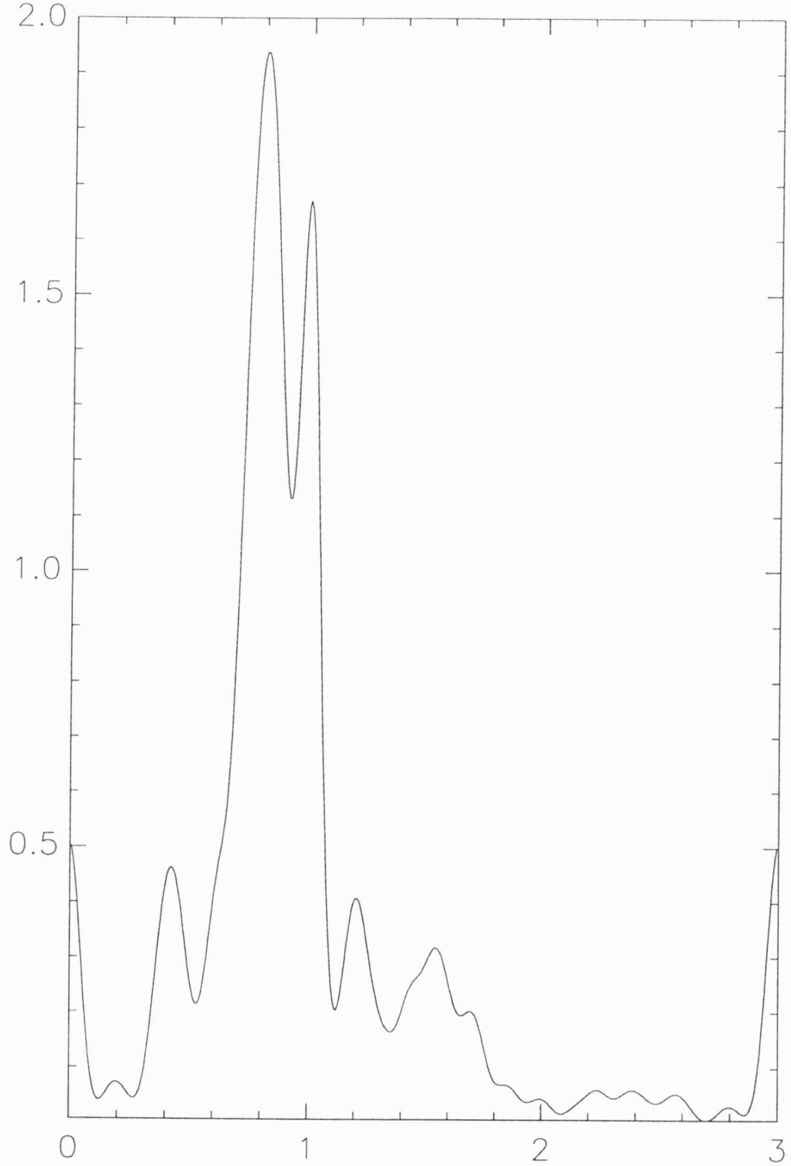


Figure 3.4d

Estimated Density of PLS2 (1977-1980)

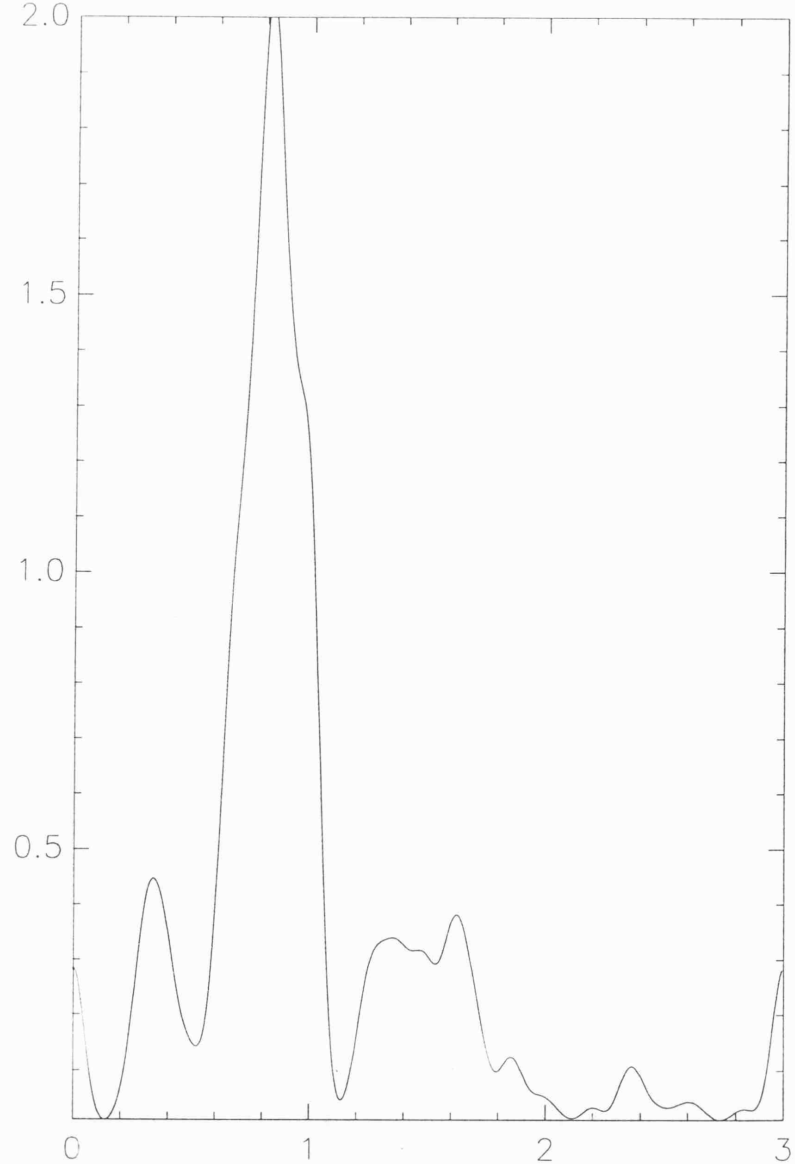


Figure 3.5a

Stochastic Kernel PLS2, 1 Year Transitions, (1953-1982).

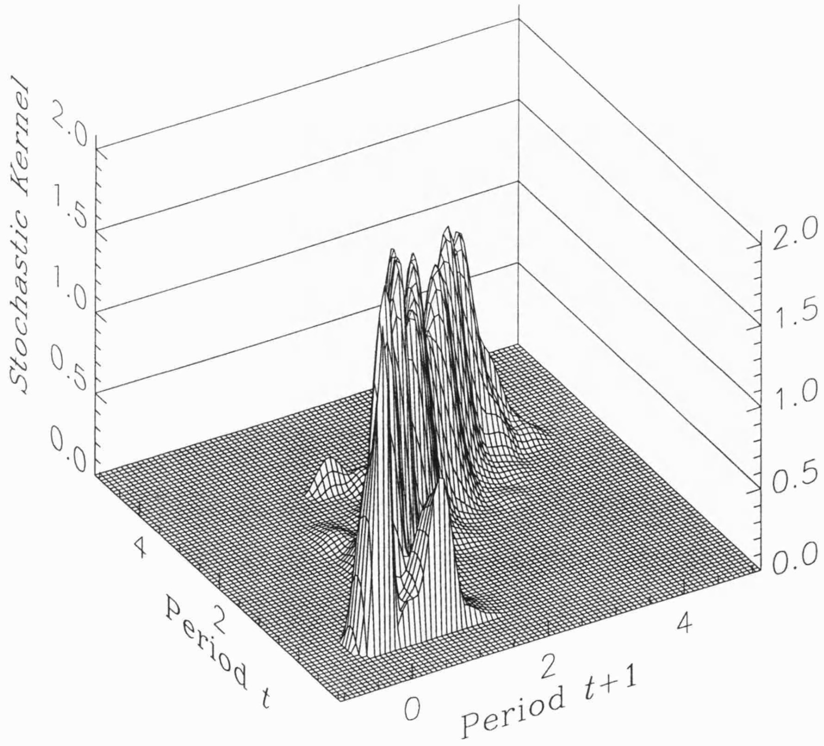


Figure 3.5b

Stochastic Kernel PLS2, 5 Year Transitions, (1953-1958).

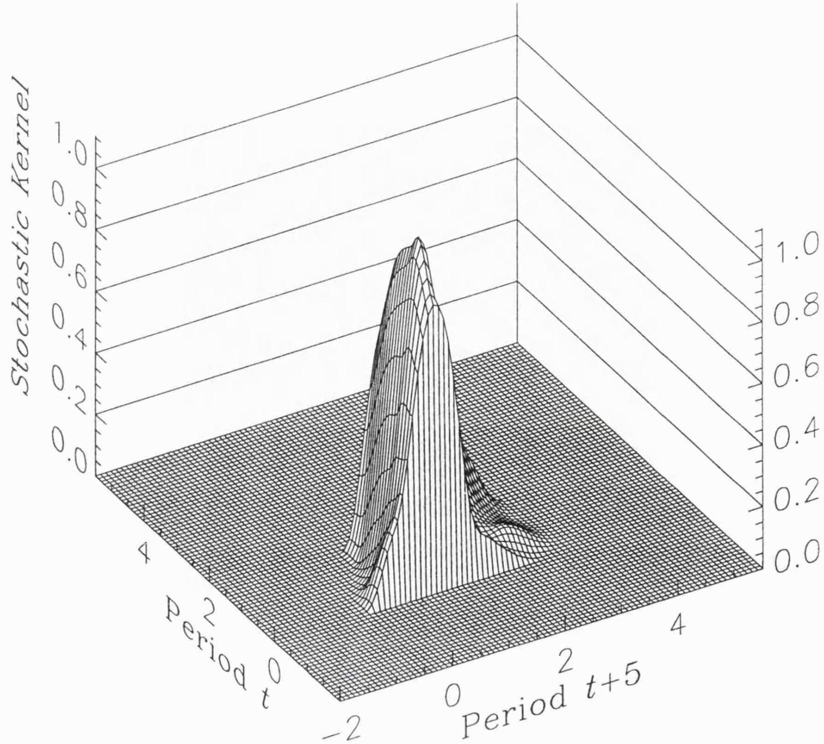


Figure 3.5c

Stochastic Kernel PLS2, 5 Year Transitions, (1963-1968).

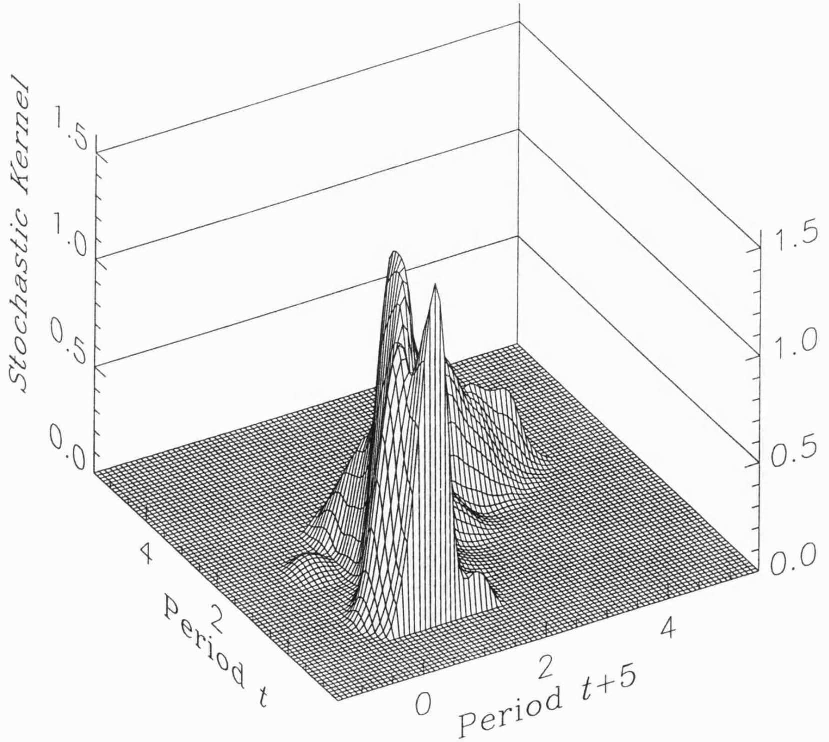


Figure 3.5d

Stochastic Kernel PLS2, 5 Year Transitions, (1973-1978).

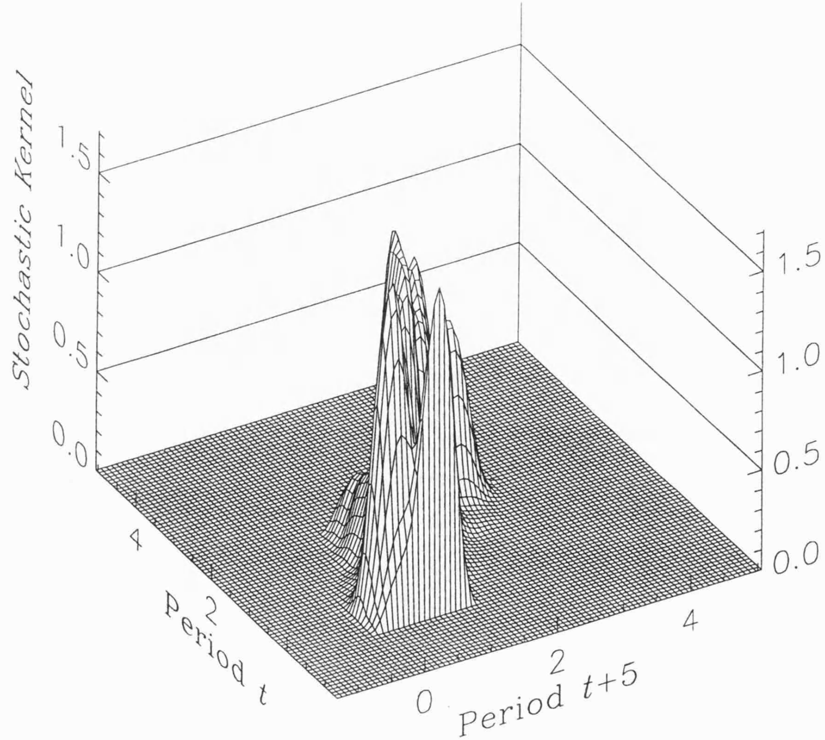


Figure 3.5e

PLS2, 1 Year Transitions, (1953-1983).

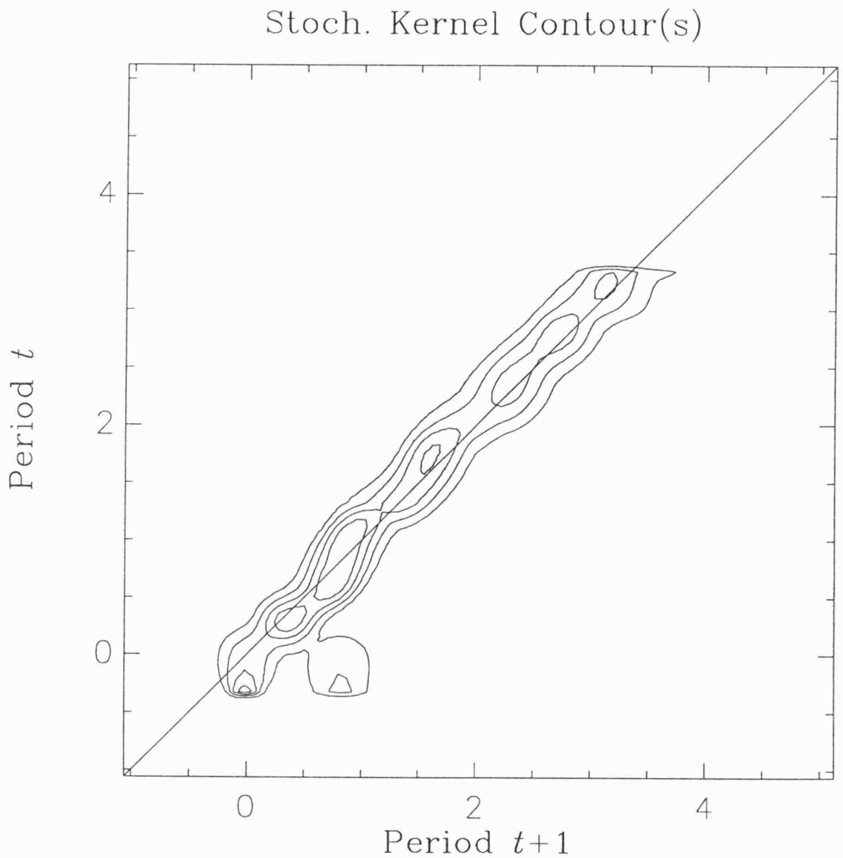


Figure 3.5f

PLS2, 5 Year Transitions, (1953-1958).

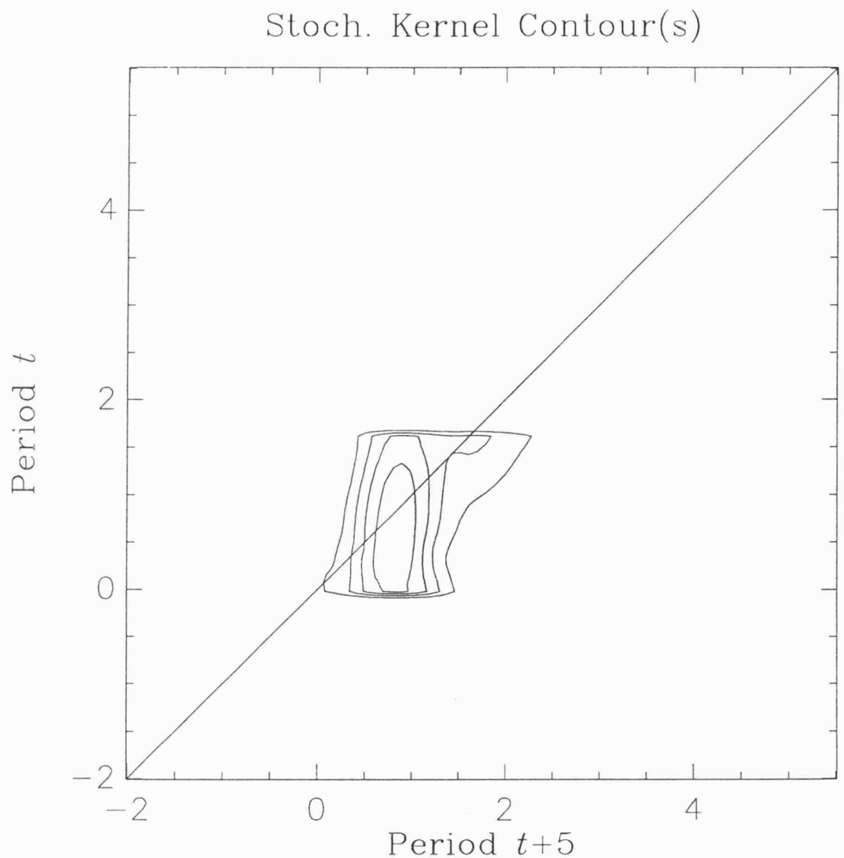


Figure 3.5g

PLS2, 5 Year Transitions, (1963-1968).

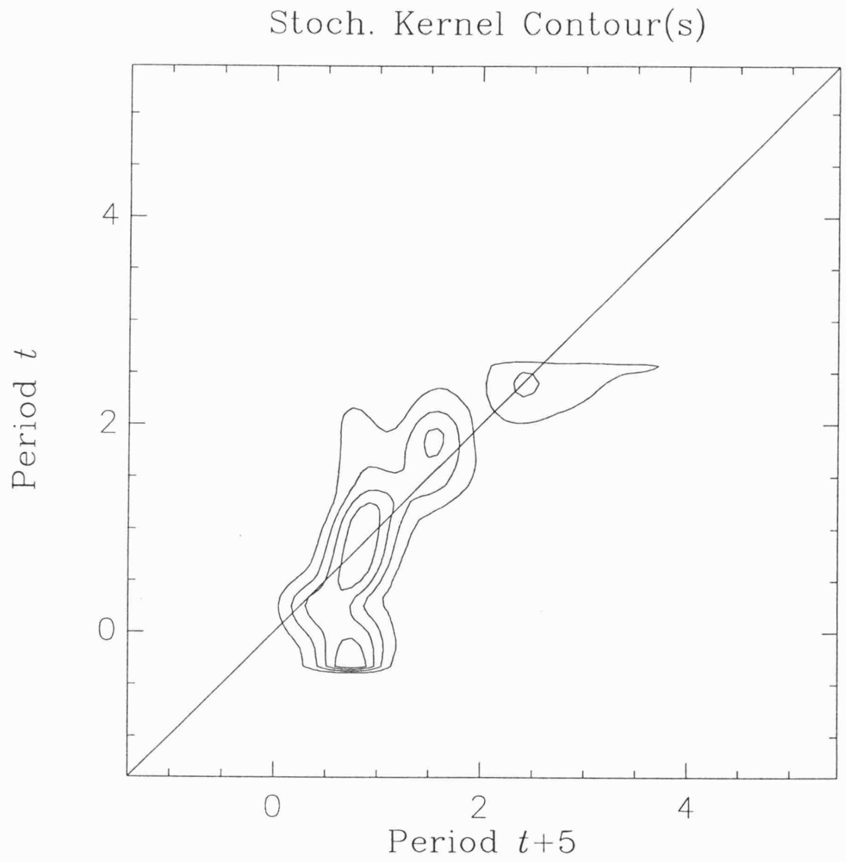


Figure 3.5h

PLS2, 5 Year Transitions, (1973-1978).

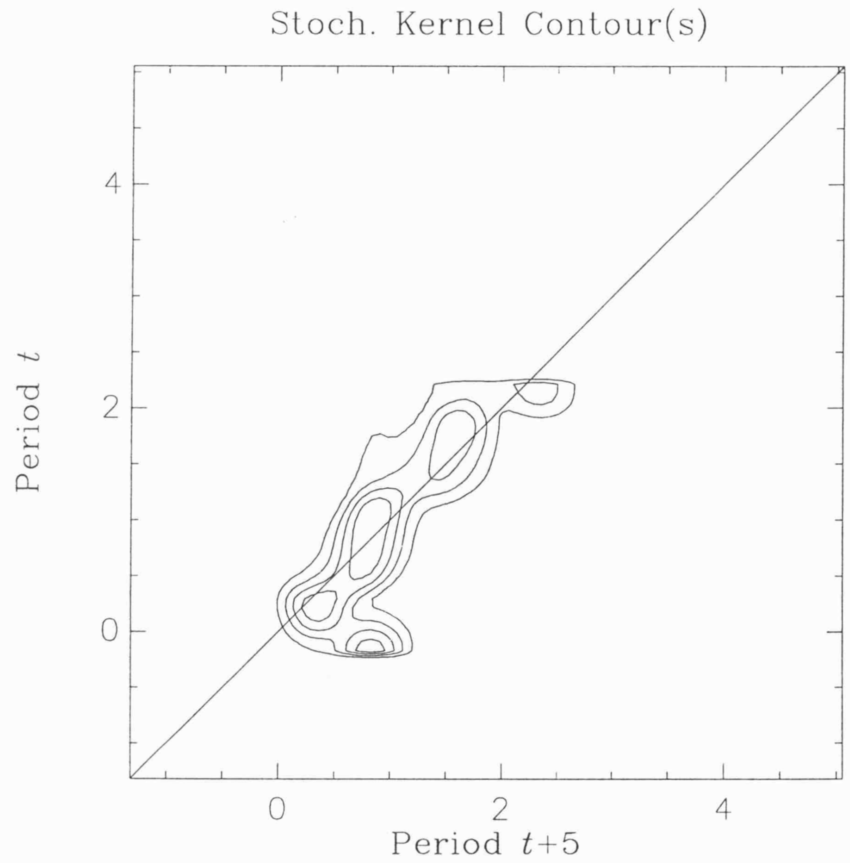


Figure 3.6a

Stochastic Kernel PLS5, 5 Year Transitions, (1953-1958).

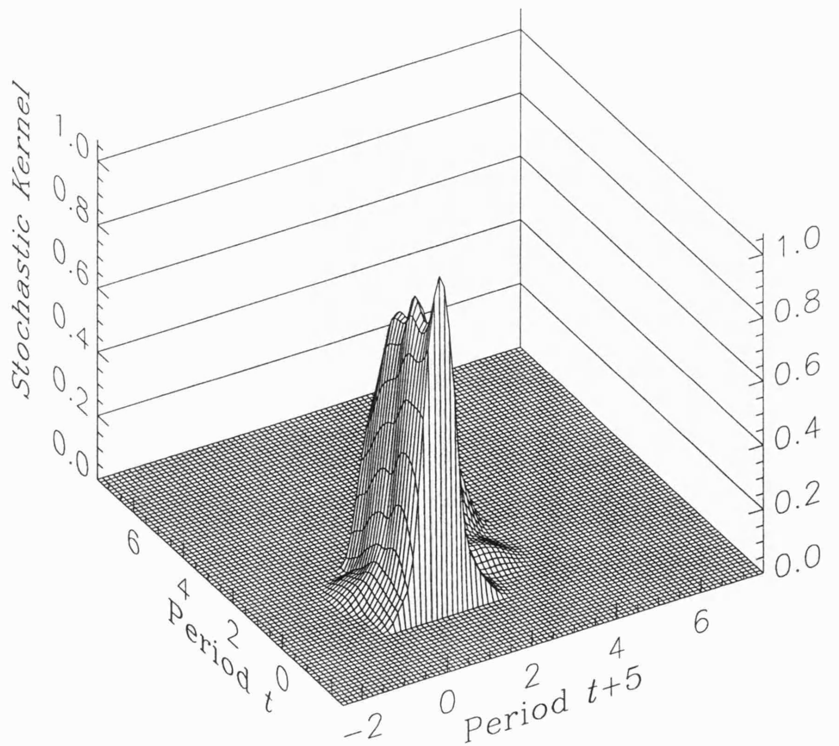


Figure 3.6b

Stochastic Kernel PLS5, 5 Year Transitions, (1963-1968).

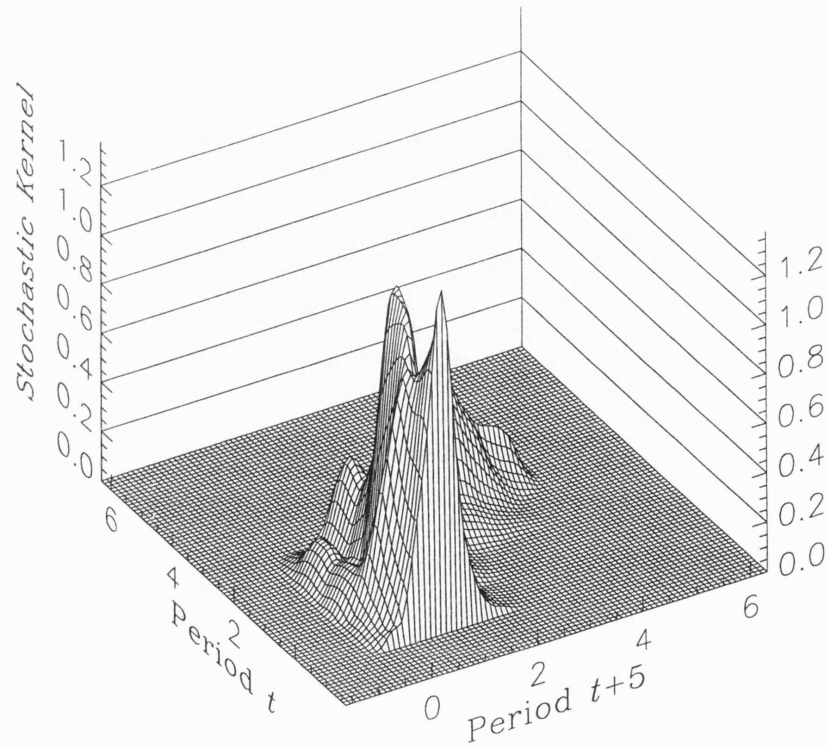


Figure 3.6c

Stochastic Kernel PLS5, 5 Year Transitions, (1973-1978).

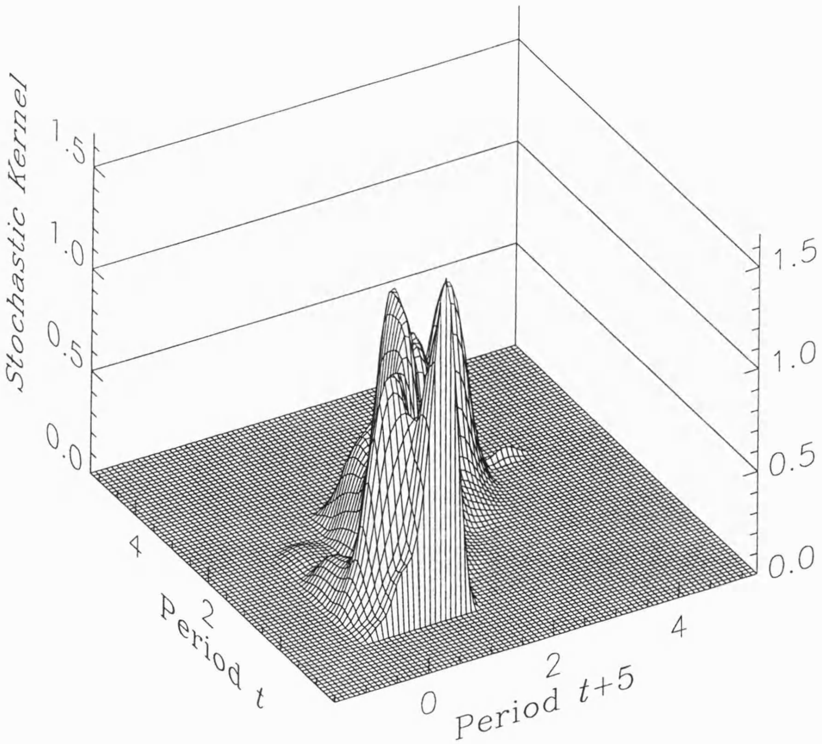


Figure 3.6d

PLS5, 5 Year Transitions, (1953-1958).

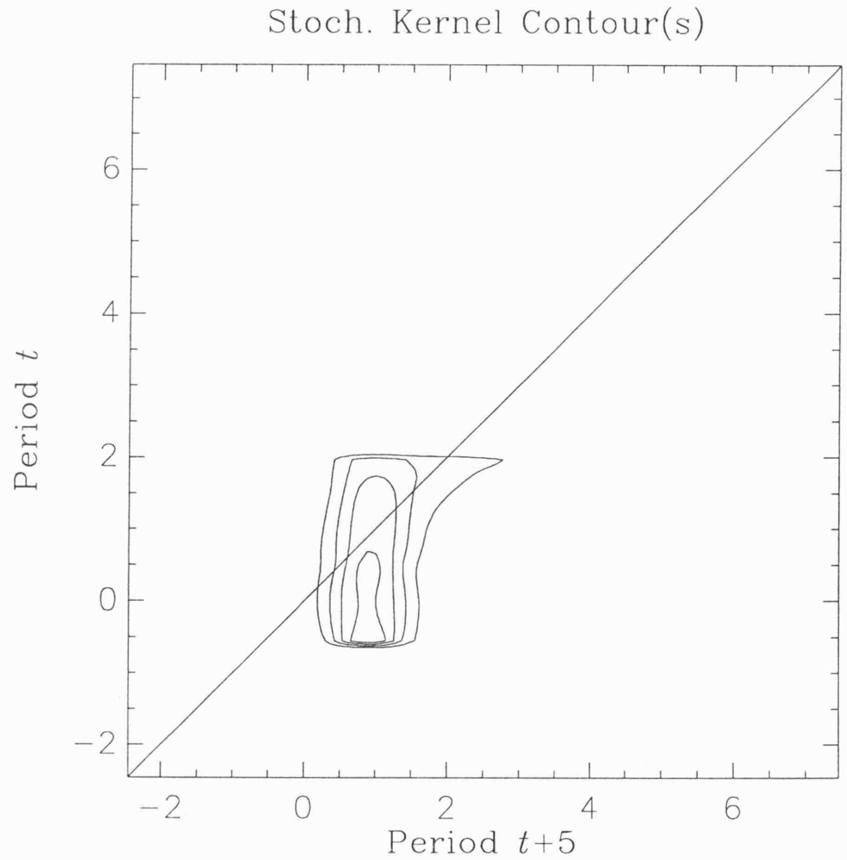


Figure 3.6e

PLS5, 5 Year Transitions, (1963-1968).

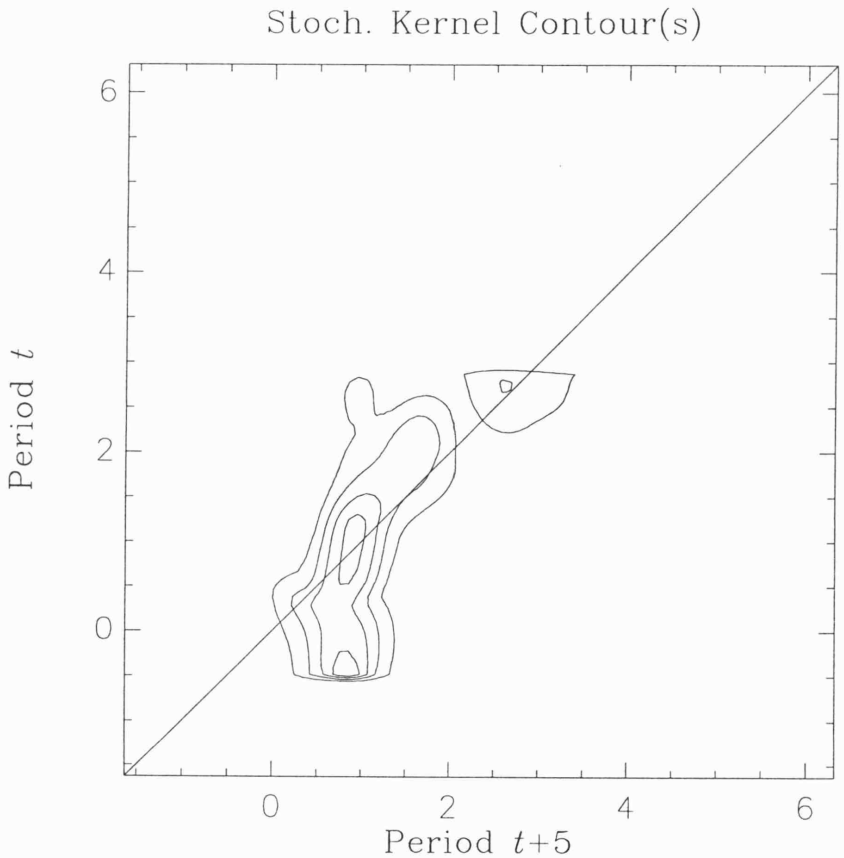


Figure 3.6f

PLS5, 5 Year Transitions, (1973-1978).

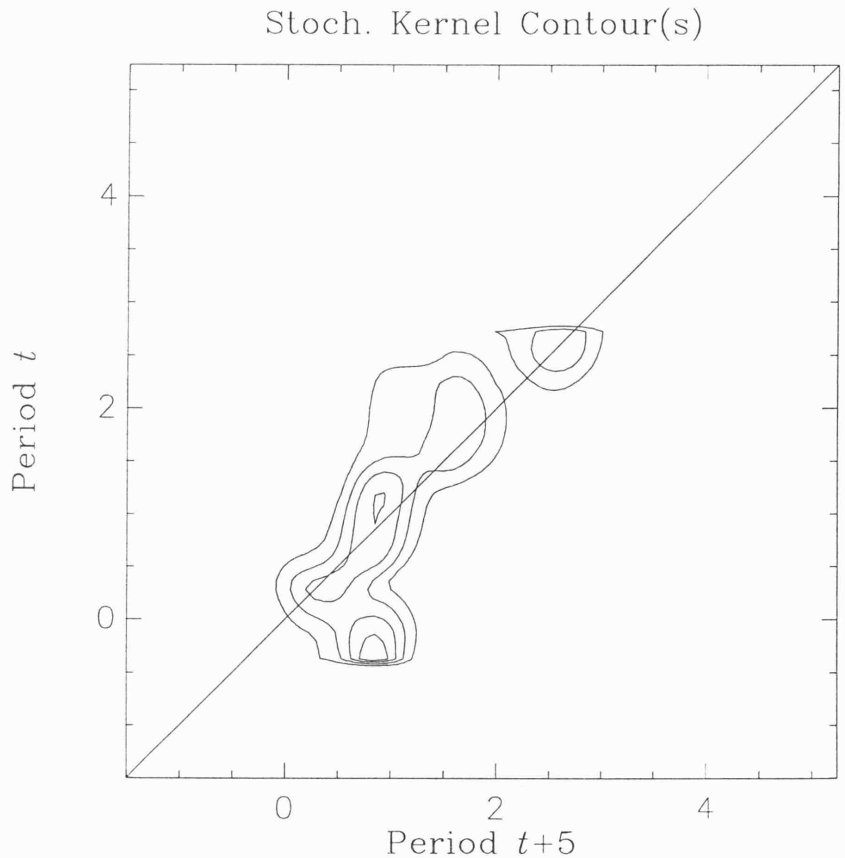


Figure 3.7a

Stochastic Kernel CS2, 1 Year Transitions, (1953-1982).

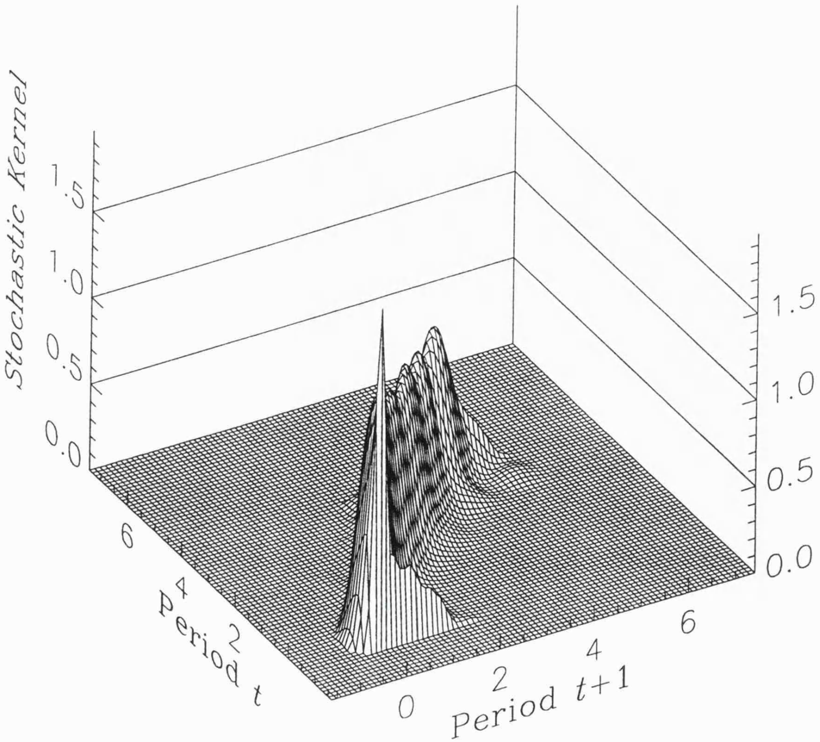


Figure 3.7b

Stochastic Kernel CS2, 5 Year Transitions, (1953-1958).

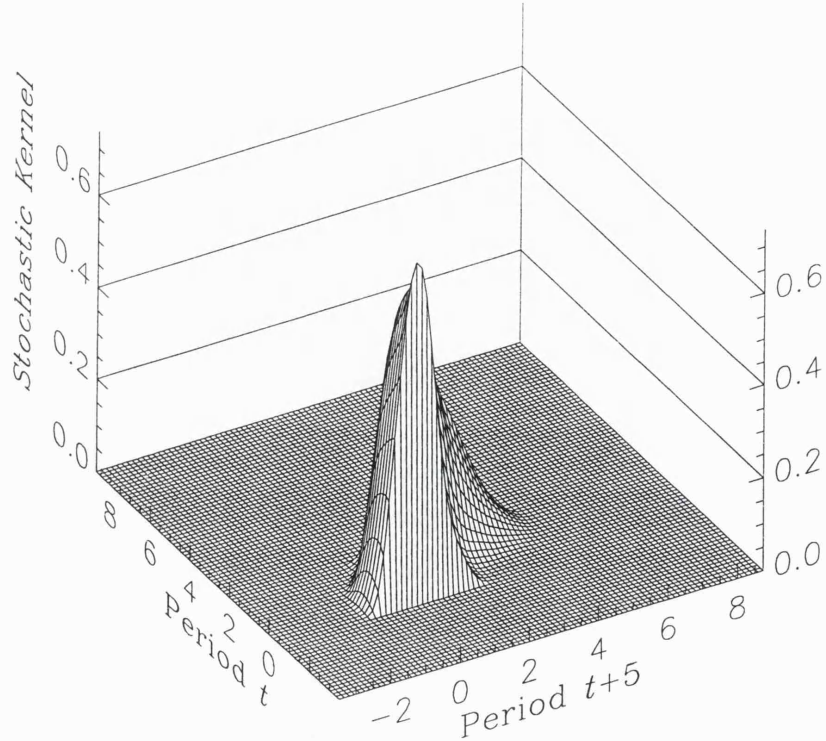


Figure 3.7c

Stochastic Kernel CS2, 5 Year Transitions, (1963-1968).

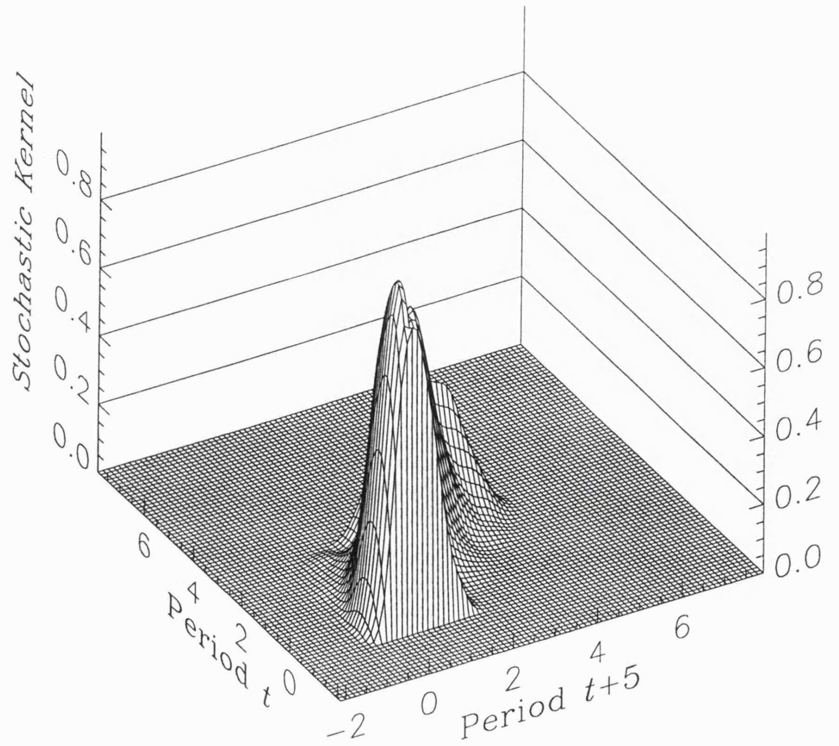


Figure 3.7d

Stochastic Kernel CS2, 5 Year Transitions, (1973-1978).

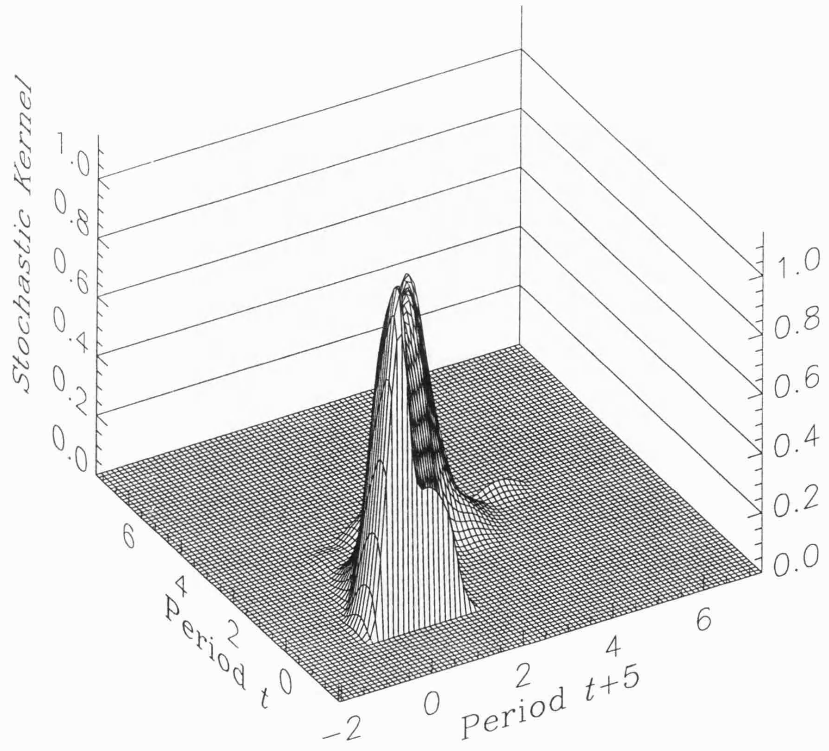


Figure 3.7e

CS2, 1 Year Transitions, (1953-1983).

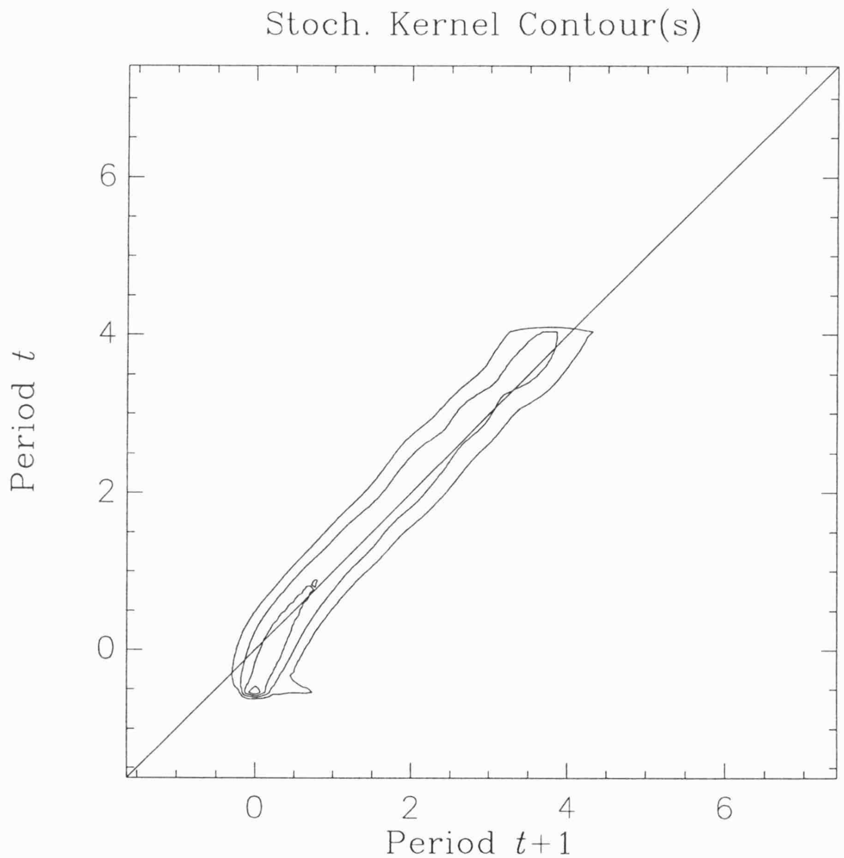


Figure 3.7f

CS2, 5 Year Transitions, (1953-1958).

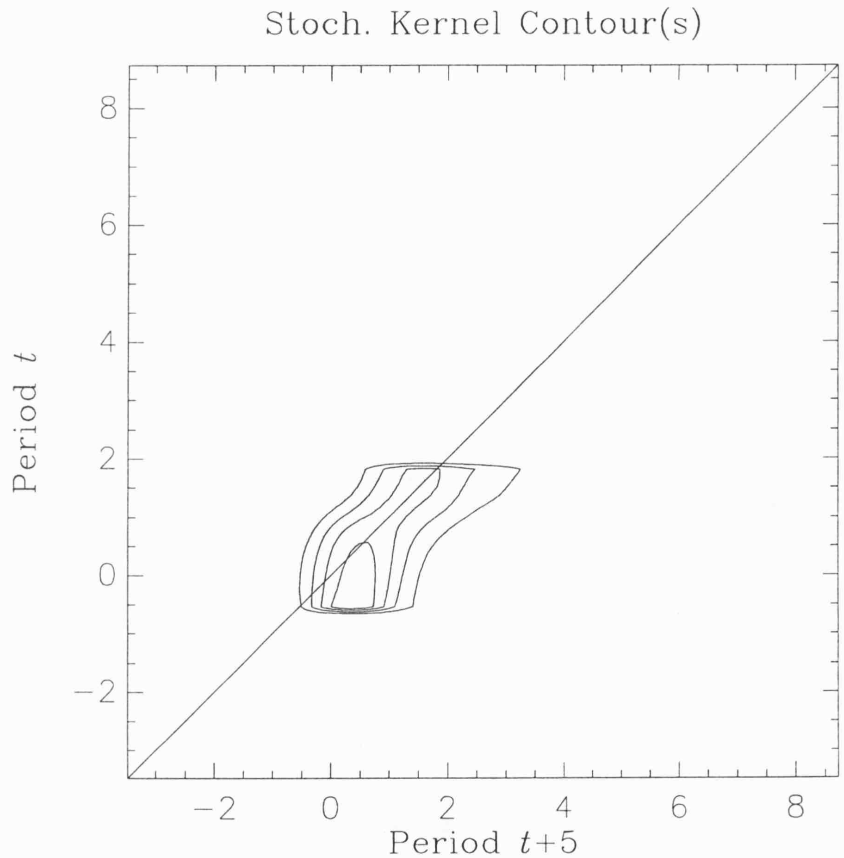


Figure 3.7g

CS2, 5 Year Transitions, (1963-1968).

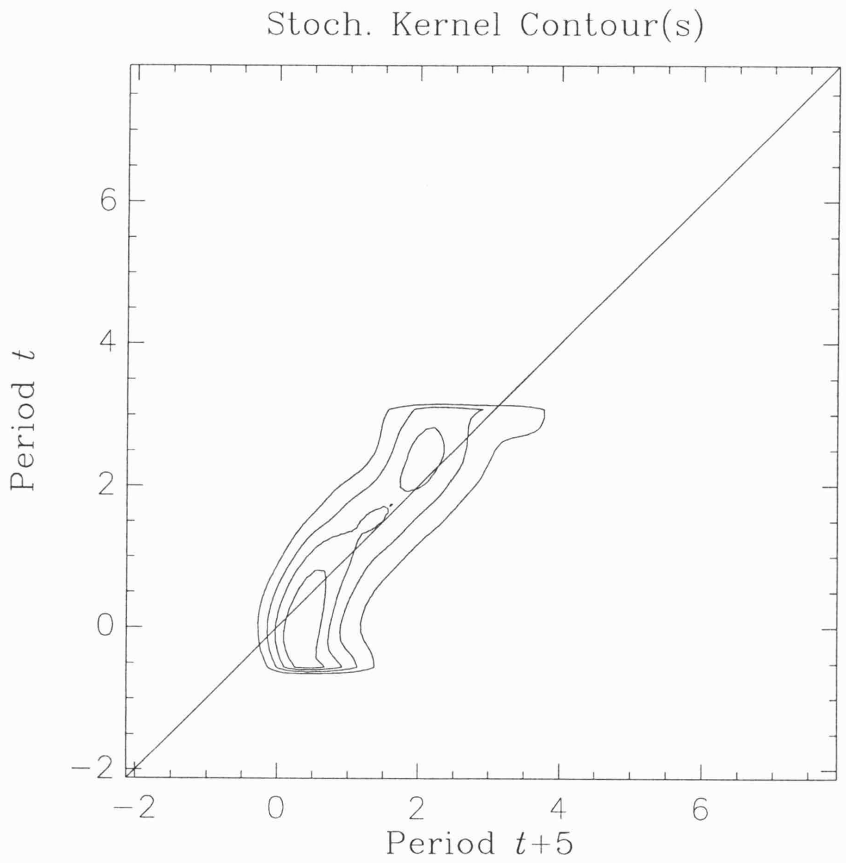
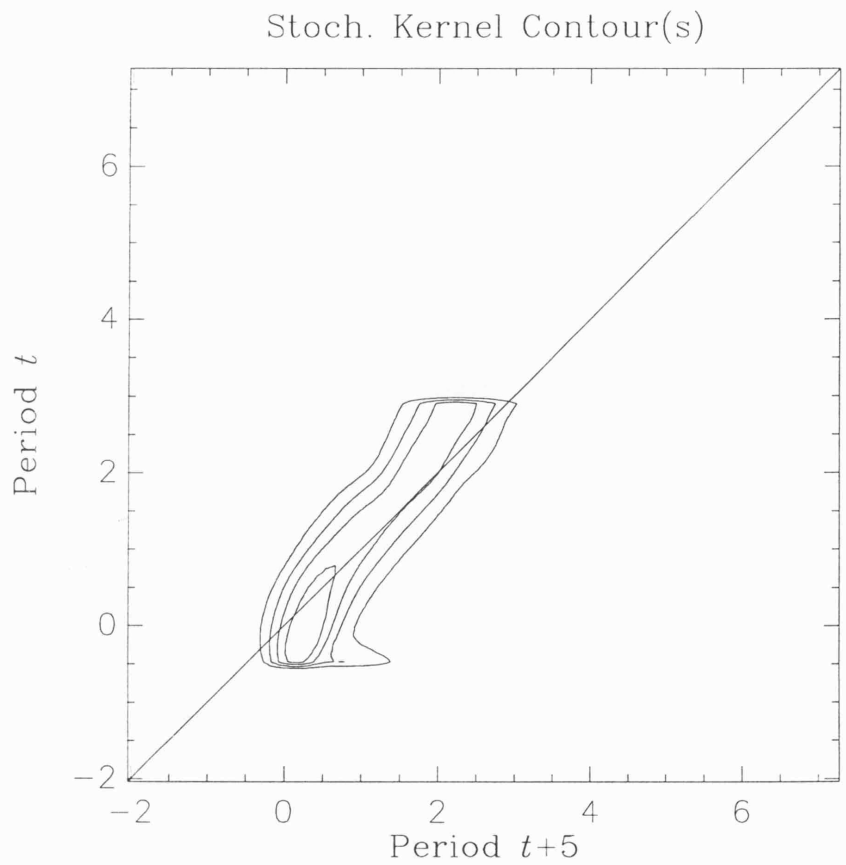


Figure 3.7h

CS2, 5 Year Transitions, (1973-1978).



3.5 The Determinants of the Dynamics of Scale.

3.5a The Theoretical Argument.

Within the theoretical framework of Section 3.2 industry specific characteristics can mitigate the negative effect of adding capacity on the profitability of existing capacity. The stronger these mitigating factors, the smaller the disadvantage of large businesses relative to their smaller competitors and hence the weaker the tendency to converge. Here two such factors will be analyzed, first the number of incumbent businesses in the industry, then the geographical dispersion of production facilities. We also show that market growth as such does not affect the tendency to converge in the long run.

Initially we take a univariate approach to analyze the effects of these factors, employing the non-parametric methods used in the previous section. To round this empirical analysis off, we take a multivariate parametric approach to show that these factors can explain the evolution of firm asymmetries on industry level in a way that is consistent with our theoretical predictions.

(i) The number of incumbent businesses in the industry.

In a large market with many businesses the effect of opening another plant on the price cost margin earned on the existing capacity is likely to be smaller than in a smaller market. If the market is approaching perfect competition, then - by definition - the price effect of an individual business installing new capacity vanishes. If demand is strictly convex to the origin, the opening of a fixed size plant leads to a smaller fall in the market price, the higher the market capacity is for a given level of demand. In that case the number of businesses in the market is no more than a proxy for the market size relative to the MES, or more precisely, the sensitivity of the price w.r.t. additional capacity[48]. The disadvantage large businesses face in capturing investment projects is weaker in bigger than in smaller markets.

[48] As no reliable data on the MES on product market level are available, we did not take the market size relative to the MES as a proxy for this sensitivity.

(ii) The geographical dispersion of production facilities.

The idea here is that for some industries there are technological restrictions on the location of production facilities. For example Aluminium is produced where electricity is cheap. (The Niagara Falls is the classic example.) In other cases high transport costs and geographically concentrated demand favours a high geographical concentration of production facilities. However, businesses in industries that do not face such constraints can always find a location sufficiently far from its own existing capacity that, due to transportation costs, the price cost margin earned on the existing capacity is unaffected. Hence if the production is locationally dispersed, for example as in the market for concrete, then being large may not be a disadvantage in the fight to open the next unit of capacity. If, on the other hand, plants are geographically concentrated, then technological constraints on production are apparently such that it is only feasible at a small number or locations. Under those circumstances a business unit might be forced to open the next plant close to its existing capacity, and the negative effect on the price cost margin earned on existing capacity cannot be avoided. Hence being large puts a business in a disadvantageous position to capture the next investment opportunity.

3.5b The Empirical Methodology and Data.

As in the last section, the dynamics of cross-section size distributions will be examined directly, using the non-parametric estimates of the cross-section distribution, transformation matrices and ergodic distributions. The empirical approach is to split the total sample into various subsamples and analyze the differences in dynamics of scale that occur as a consequence of the conditioning variable. Essentially, we determine if the dynamics of scale differ in the predicted way.

In Section 3.4 various ways to deal with potential entry were analyzed, but the conclusions did not alter qualitatively. In this section we therefore analyze the dynamics of incumbent businesses exclusively. For presentational reasons we took the logarithmic value of the size variable CS.

The sample used in Section 3.4 was supplemented by industries in which a larger number of businesses is in operation, as we are no longer solely interested in oligopolistic markets. Other industries were included on the basis that 5 digit SIC location data were available, in order to increase the number of observations. The location data were taken from the Census of Manufactures 1982[49].

Table 3.6: Summary Statistics of the Subsamples.

Sample Number	Average Number of Businesses	Average Annual Capacity Growth (in %)
1	5.0	8.4
2	11.0	10.5
3	35.1	7.1
4	9.6	11.8
5	6.3	5.9
1a	5.0	12.0
1b	5.2	6.2

3.5c Dynamics of Scale and the Number of Incumbent Businesses.

In Section 3.4 we showed that sizes of businesses tend to converge, whether estimates of matrix Q or its continuous state-space equivalent, stochastic kernels, are used. Here the very same sample is taken, but it is split into three subsamples. Sample 1 consists of those industries that have 10 businesses or less in every year of the sample period. Sample 2 consists of those industries in which more than 10 but less than 20 businesses operate in any year of the sample period. A third sub-

[49] Due to confidentiality rules the Bureau of the Census does not report the production in states if that would reveal the production of individual companies. Therefore, a large number of observations is missing. The difference between the US production of good i and the equivalent sum of the reporting states was distributed proportionally to the manufacturing production between all non-reporting states. See Krugman (1991, p. 57) for a more detailed discussion of the problems associates with the geographical dispersion as reported in the US Census of Manufactures.

sample is constructed containing those industries with 20 or more businesses in any year. Table 3.6 gives some summary statistics of the three subsamples, as well as other subsamples used later.

Tables 3.7 and 3.8 report estimates of the time stationary matrix Q for the pooled Samples 1 and 2, and for Sample 3. The grid has been selected such that an approximately equal number of observations results in each state in the initial year of the sample, as before. Consequently a uniform initial distribution results by construction.

Table 3.7 confirms for incumbent firms our earlier result that sizes of businesses tend to converge relative to the initial distribution for those chemical product markets in which there are less than 20 businesses. The ergodic distribution, based on the endogenously selected grid, is unimodal.

Table 3.8 reports the estimate of Q for the subsample of industries with more than 20 businesses. The ergodic distribution shows that in the long run businesses in this subsample, if anything, are drifting apart relative to the initial (1963) distribution. There is thinning out around the mean and accumulation in the second and sixth state. Over the sample period (here 1963-1982) there has been polarization of the size distribution for industries with a large number of producers. Consequently, relative to the initial distribution the tendency to converge is stronger for the industries with a low number of firms.

If the converging forces are stronger the smaller the number of firms in the market, then the ergodic distribution of the small industry sample should dominate the corresponding distribution of the large industry sample in the sense of second order stochastic dominance. Comparison of the two ergodic distribution of Table 3.7 and 3.8 is difficult as the grid is determined endogenously. In Figure 3.8 the ergodic distributions for the three subsamples are shown for an identical (exogenously fixed) grid, which is equidistance in logarithmic terms. It confirms that the higher the number of businesses the larger the spread of the ergodic distribution. Note that, although the results in Table 3.7 and 3.8 suggest an accumulation in the tails for Sample 3, this effect is not strong enough for a clearly bimodal ergodic distribution to emerge if the grid is equidistant. The size distribution for large industries thinned out in the middle and accumulated mass in the tails, relative to the initial distribution. But there is only a weak suggestion of a bimodal ergodic distribution.

These results are supported by a comparison of the cross-section size distribution for the pooled Samples 1 and 2 on the one hand and Sample 3 on the other, for the years 1963 and 1982. Figure 3.9 shows (smoothed) estimates of the densities of the normalized (log) sizes of businesses, estimated in the same way as the cross-section distributions of Figure 3.4. For the sample of industries with less than twenty businesses there is little change in the cross-section distribution over the two decades. But for industries with a high number of businesses the density tends to spread out over time, though it should be noted that the densities do not capture the intra-distributional dynamics, nor the long run effects. The domain of the density of sample 3 is substantially larger than the domain of the pooled subsample 2 and 3. However, this might just be an artefact of the difference in number of businesses in the industries.

Table 3.7 First Order Transition Matrix - Samples 1 & 2

Time stationary (1963 - 1982)

Upper end	**-1.075**	**-0.579**	**-0.223**	**0.119**	**0.511**	**∞**
(r)	(1)	(2)	(3)	(4)	(5)	(6)
710	*0.86*	*0.06*	*0.02*	*0.04*	*0.02*	*0.01*
705	*0.10*	*0.79*	*0.07*	*0.02*	*0.01*	*0.01*
733	*0.01*	*0.13*	*0.74*	*0.09*	*0.02*	*0.01*
725	*0.00*	*0.01*	*0.16*	*0.72*	*0.09*	*0.01*
710	*0.00*	*0.00*	*0.01*	*0.14*	*0.77*	*0.07*
671	*0.00*	*0.00*	*0.00*	*0.00*	*0.10*	*0.90*
Ergodic	0.134	0.152	0.172	0.181	0.184	0.176

Table 3.8 First Order Transition Matrix - Sample 3

Time stationary (1963 - 1982)

Upper end	**-1.823**	**-1.048**	**-0.534**	**-0.043**	**0.520**	**∞**
(r)	(1)	(2)	(3)	(4)	(5)	(6)
599	*0.92*	*0.02*	*0.04*	*0.01*	*0.02*	*0.01*
568	*0.06*	*0.87*	*0.04*	*0.01*	*0.01*	*0.00*
612	*0.00*	*0.12*	*0.81*	*0.03*	*0.03*	*0.00*
542	*0.00*	*0.01*	*0.11*	*0.78*	*0.07*	*0.02*
545	*0.00*	*0.00*	*0.01*	*0.12*	*0.81*	*0.06*
538	*0.00*	*0.00*	*0.00*	*0.00*	*0.09*	*0.91*
Ergodic	0.152	0.190	0.166	0.140	0.173	0.179

Figure 3.8

Effect of the Number of Firms in the Industry
(Ergodic Distributions)

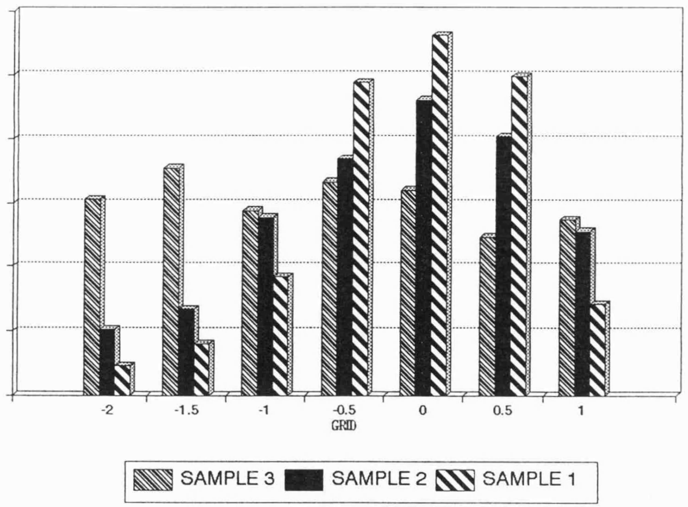


Figure 3.9a

Estimated Density of the Cross-Section Size Distribution Fragmented Industries (Sample 3), 1963.

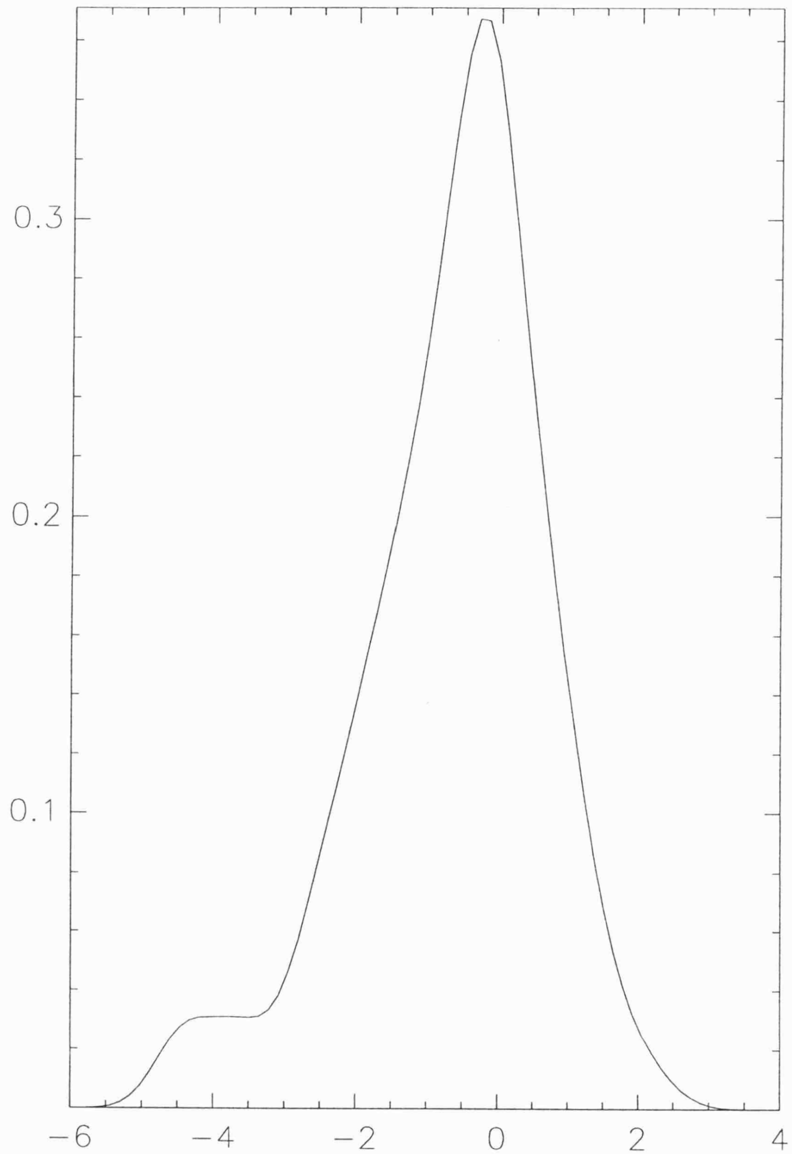


Figure 3.9b

Estimated Density of the Cross-Section Size Distribution Fragmented Industries (Sample 3), 1982

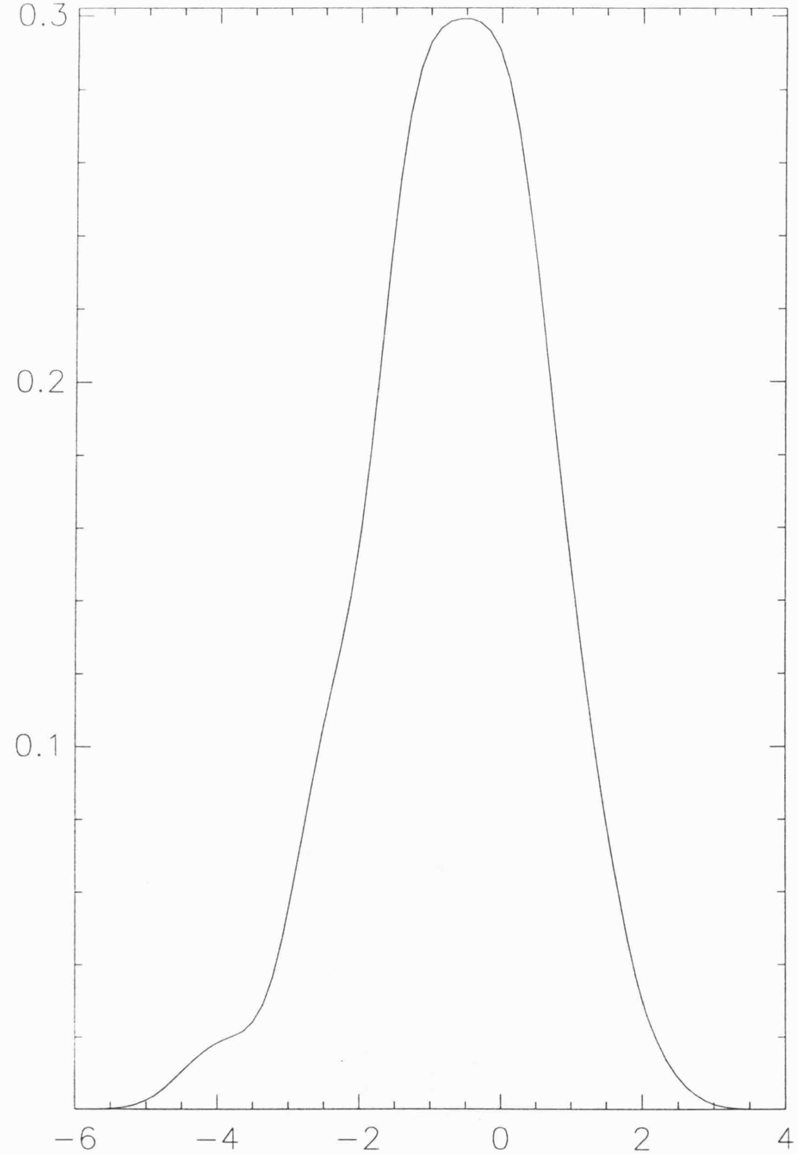


Figure 3.9c

Estimated Density of the Cross-Section Size Distribution
Concentrated Industries (Sample 1 & 2), 1963

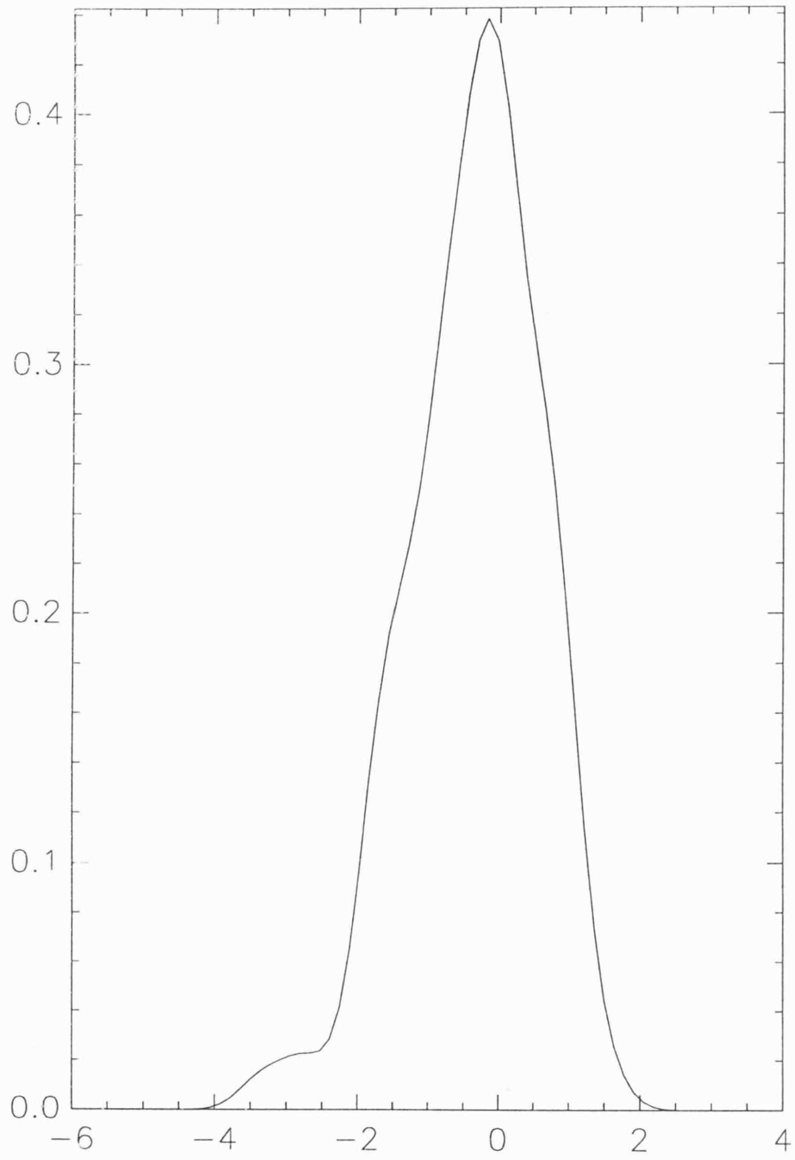


Figure 3.9d

Estimated Density of the Cross-Section Size Distribution
Concentrated Industries (Sample 1 & 2), 1982

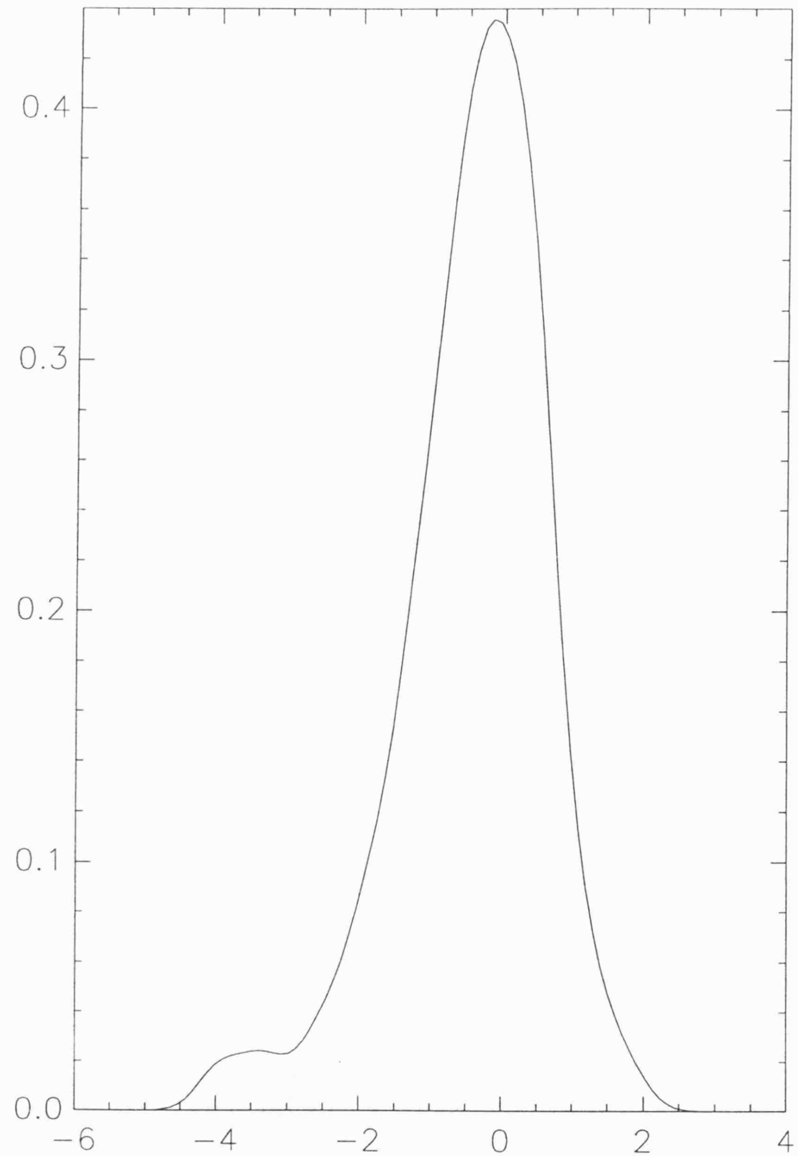


Figure 3.10

**Effect of Geographical Spread of Production Facilities
(Ergodic Distributions)**

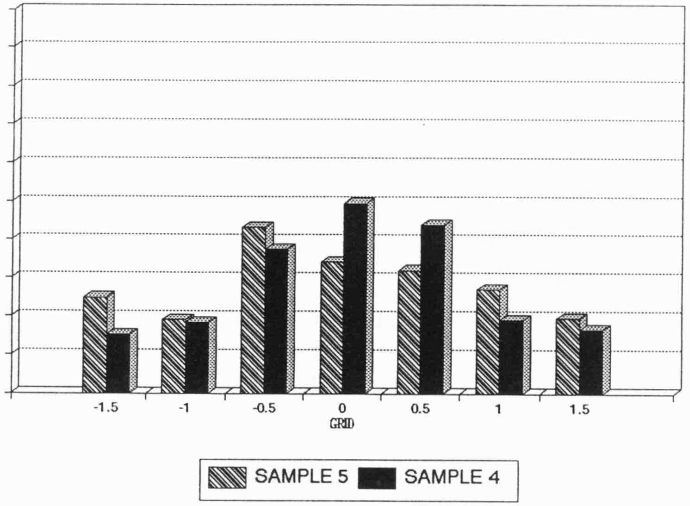
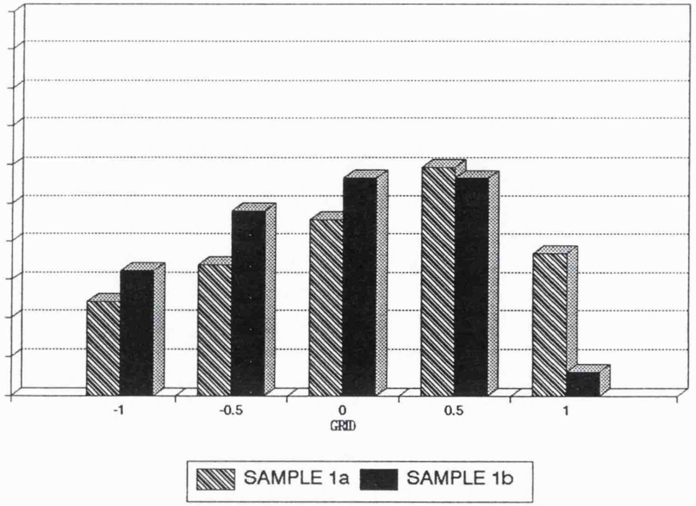


Figure 3.11

**Effect of Market Growth
(Ergodic Distributions)**



3.5d Dynamics of Scale and the Geographical Dispersion of Production.

In this subsection we test the effect of the geographical dispersion of production facilities on the distributional dynamics. Geographical dispersion was measured by a geographical Gini-index GGINI, based on a locational Lorenz curve that is constructed as follows[50]. For each US state both the share in the production of good i and its share in the country's manufacturing production was calculated. All states were ranked by decreasing ratio of the two numbers. The graph of this ranking, with the cumulative proportion of manufacturing production on the horizontal axis and the cumulative production of good i on the vertical yielded a Lorenz curve. The Gini-coefficient is twice the area between the Lorenz curve and the diagonal and hence a measure of the asymmetry of the geographical spread of production facilities, relative to the distribution of US manufacturing across states. A GGINI index equal to zero indicates that the production of good i is proportional to the distribution of manufacturing production in the country. The higher the index, the more concentrated the production of good i relative to manufacturing production.

Only industries with less than 20 businesses in 1982 were analyzed, to guarantee oligopolistic interaction. The index GGINI is calculated for those industries for which 5-digit SIC production location data were available[51], see Table 3.1. The sample of industries was then split into two subsamples, one with locationally concentrated industries, defined as those with Gini-indices (GGINI) more than 0.5 (Sample 4), the other with industries that are locationally dispersed, i.e. GGINI less than 0.5 (Sample 5). Table 3.6 gives summary statistics for the subsamples.

As before, transition matrices (Q) were estimated for both subsamples. Overall the effect was weaker than was found for the number of firms. The most clear cut results were found for the comparison of the ergodic distributions of the two

[50] For a more detailed description see Krugman (1991).

[51] There is an exception for Magnesium and Aluminium, since the 4 digit industry refers to the homogeneous product market.

subsamples, if the grid is fixed equidistantly, to make the comparison feasible. Figure 3.10 gives the ergodic distributions implied by the estimates of Q. The horizontal axis gives the (logarithmic) grid which was identical for both samples. The data confirm the theoretical prediction that the more concentrated the geographical location is, the stronger the tendency to converge, in the sense that more probability mass is concentrated around the mean in the ergodic distribution. The subsample of industries with higher Gini-indices (GGINI) (4) has an ergodic distribution that is more concentrated around the mean relative to the ergodic distribution of the subsample of industries with a low GGINI (5). This result is the more striking since from Table 3.6 follows that Sample 5 has a higher average number of businesses per industry, which, given the results of section 3.5c, suggests a larger spread of the ergodic distribution than that for Sample 4. However, the locational effect seems to dominate the number-of-businesses effect.

3.5e Dynamics of Scale and Market Growth.

In Section 3.4 we reported that the tendency of sizes of businesses to converge is strongest during 1954-1964. There are two candidate explanations for this phenomenon. One is that the growth rate of capacity was higher during that decade than in later years, suggesting that the period was longer in terms of event time than later decades. The higher the annual growth, the more opportunities to open additional capacity arose within a year and therefore the stronger the tendency to converge per unit of time. The other candidate explanation is that, as the number of businesses in the industry was still relatively low, the externality on the market price of the installation of new capacity was relatively strong. In this section these two effects are disentangled and we show that there is no clear effect of market growth on the long run distribution.

In order to condition on the number of businesses, a subsample of industries is taken that show strong variation in growth rates over various time periods, but show little variation in the number of businesses. Sample 1 is split into two subsamples, one between 1953 and 1968 (Sample 1a), the other from 1968 to 1983 (Sample 1b). Table 3.6 gives summary statistics for both subsamples and shows that Sample 1a has an annual growth rate that is roughly twice the annual growth rate of Sample 1b (12.0% and 6.2% resp.). The average number of businesses is, however, almost identical (5.0 and 5.2 resp.). For both subsamples a 5-state transformation matrix Q has been estimated. Figure 3.11 shows the implied ergodic distribution. Both are unimodal, but there is little suggestion of one being more concentrated around the mean than the other. This suggests that the observed tendency to converge is stronger during periods of high growth only because the number of firms was still relatively small and hence the externality strong.

3.5f Pulling Some Threads Together.

Until here the analysis of the origins of the converging tendencies of the sizes of businesses has been restricted to a univariate approach. By splitting the sample one dimensionally into various subsamples we showed that the size of the industry and the geographical spread of production facilities affect industry dynamics in a way that is consistent with the earlier theoretical framework. Now we turn to the multivariate analysis.

The basic proposition that will be tested here is that in industries in which the converging forces are strong, the 1982 (cross-section) size distribution should show less inequality relative to the initial distribution than the corresponding distribution for those industries in which the converging forces are likely to be weak. Inequality of business sizes is measured by the Gini-index[52]. In this section the (cross-section) relationship between the change in the Gini-index over the sample period and the theoretical determinants of the strength of the converging tendencies is tested on product market level.

The earlier analysis is taken as a starting point. That is, the larger the number of

[52] If there are N firms in the industry, with sizes $q_1 \leq q_2 \leq \dots \leq q_N$, and the average size μ, then the Gini-index is $G = (N+1)/(N-1) - [2\Sigma_{k=1}^{N}(N-k+1)q_k]/[N(N-1)\mu]$. The higher the index, the larger the area between the Lorenz curve and the diagonal, reflecting more inequality among sizes of businesses. See Eatwell et al. (1987)

firms in the market, the weaker the converging tendencies are expected to be, and therefore, the more persistent the inequalities of size of business units. Essentially, if in a larger market the price fall associated with the opening of a fixed size plant is smaller, then the larger firm faces less of a disadvantage. The less concentrated the geographical distribution of production facilities are relative to US manufacturing, the weaker the converging tendencies will be, since an already large firm can always find a submarket in which it has no presence as of yet. Therefore, a more persistent inequality among the firms is expected to be observed if the geographical Gini-index (GGINI) is high.

Dunne, Roberts and Samuelson (1989) show that the dynamics in industries with multi-plant businesses are markedly different from those with single plant businesses[53]. To control for this difference, we conditioned on the average number of plants a business in the industry had. The higher this average is, the more scope larger firms have to distribute their production geographically, which relaxes the disadvantages associated with being large. Hence, the higher the average number of plants per firm, the more persistent the inequality among firms is expected to be.

Finally, the number of periods in terms of "event time" is included as a conditioning variable. That is, the number of years between the initial observation, as given in Table 3.1, and 1982 is included as well as the average growth rate of industry capacity.

Our basic specification is:

$$GINI^{i}_{82} = \alpha + \beta GINI^{i}_{t0} + \gamma N^{i} + \delta PLANT^{i} + \zeta GGINI^{i} + \eta GROWTH_{i} + \theta TIME^{i} + \varepsilon^{i} \qquad (3.17)$$

where

[53] They find that large multi-unit plants have both lower failure rates and higher growth rates if successful than large single-unit plants have.

$GINI_{t0}^i$	=	Gini-index of market i at time t0, the initial year of observation. The size of businesses is measured by production capacity.
N^i	=	average number of business in market i over the sample period $\left(N^i = \frac{N_{t0} + N_{82}}{2}\right)$.
$PLANT^i$	=	Average number of plants per business in market i over the sample period.
$GGINI^i$	=	Geographical Gini-index, that is the geographical distribution of production of good i relative to US manufacturing for 1977.
$GROWTH^i$	=	Average growth rate of industry capacity over the sample period $\left(Q_{82}^i / Q_{t0}^i\right)^{1/(82-t0)}$, where Q_t^i is the total industry capacity of product i at time t.
$TIME^i$	=	Sample period (number of years).
ε^i	=	Disturbance term, distributed IID (0, σ_ε^2) across all i

The theoretical predictions are that the coefficients of *N* and *PLANT* (γ and δ respectively) are positive, since the profitability of existing capacity is less sensitive to increasing capacity if the number of firms or the number of plants per firm are high. Hence the converging tendencies are weaker, and the inequalities among firms remain larger. The coefficient of the geographical Gini-index, ζ is expected to have a negative sign, since higher geograpical concentration increases the strength of converging tendencies and therefore leads to less inequality over time. The coefficients of the "event time" variables (GROWTH and TIME) are both expected to be negative, since the longer the sample period in terms of event times, the smaller the Gini-index we expect to observe.

We took all product markets of Table 3.1 for which the geographical Gini-indexes could be determined on 5-digit (SIC) level[54]. These are 20 data points. Table 3.9 give some summary statistics of the variables and raw correlations.

[54] Cyclohexane was left out as the sample starts with two firms of equal size in 1956.

Table 3.9 Descriptive Statistics of the Data.

	MEAN	SD	MIN	MAX
$GINI_{82}$	*.453*	*.151*	*.109*	*.777*
$GINI_{t0}$	*.511*	*.167*	*.288*	*.875*
NUMBER	*10.48*	*10.46*	*2.5*	*43.5*
PLANT	*1.47*	*.519*	*1*	*3.36*
GGINI	*.423*	*.108*	*.220*	*.610*
GROWTH	*1.08*	*.044*	*1.03*	*1.23*
TIME	*23.5*	*3.33*	*17*	*30*

b. Correlations

	$GINI_{82}$	**$GINI_{t0}$**	**NUMBER**	**PLANT**	**GGINI**	**GROWTH**
$GINI_{t0}$	*.457*					
NUMBER	*.458*	*.085*				
PLANT	*.375*	*-.110*	*.170*			
GGINI	*-.333*	*-.110*	*-.021*	*.214*		
GROWTH	*-.266*	*-.197*	*-.053*	*-.294*	*.256*	
TIME	*-.014*	*.312*	*.210*	*-.092*	*.319*	*-.029*

Comparison of the average Gini-index for the initial and final observation reveals a decreasing trend, implying that inequalities among firms tend to become less. That is, the largest x% of the firms tend to represent an ever smaller market share. The geographical Gini-index GGINI indicates that the geographical spread of production in the sample is more concentrated than US manufacturing, indicating that regions tend to specialize in the production of certain chemical products. The annualized growth rates over the sample period varies from 3.3% for Sodium Hydrosulfite to

22.6% for High Density Polyethylene. The sample periods vary between 17 and 30 years.

The raw correlations (Table 3.9b) show a positive association between the Gini-index in 1982 on the one hand, and the number of firms and the average number of plants on the other. The correlation of $GINI_{82}$ with $GGINI$, GROWTH and TIME are all negative, which is consistent with the theoretical predictions.

Table 3.10 Estimated Coefficients (Dependent variable: $GINI_{82}$).

$GINI_{t0}$	N	PLANT	GGINI	GROWTH	TIME	CONST	$\bar{R}^2$	obs
*.431** *(.119)*	*.0046** *(.0012)*	*.136** *(.037)*	*-.522*** *(.207)*	*.234* *(.447)*	*-.0023* *(.0070)*	*.0069*	*.49*	*20*
*.393*** *(.101)*	*.0049** *(.001)*	*.129** *(.032)*	*-.519*** *(.174)*			*.230*	*.55*	*20*

Heteroscedastic (White's) SE in parentheses
* 1% significance level
** 5% significance level

The first row of Table 3.10 shows the results of estimating (3.17) by OLS. Both the coefficients of the average number of firms and the average number of plants are significant on a 1% level, indicating that the more fragmented the industry (or the higher the average number of plants per firm), the more persistent the asymmetries in the sizes of business units. This confirms our earlier result, where we showed using the ergodic distributions that the higher number of businesses in the market the weaker are converging tendencies. The geographical Gini-index GGINI is significant at a 5% level, indicating that the more concentrated production facilities are relative to US manufacturing, the less asymmetries among business sizes tend to persist, i.e. the more are large businesses at a disadvantage in the battle to be the firm to open the next plant. This is also consistent with our earlier results.

Finally the two variables associated with the length of the sample period in terms of "event time" are not significant. The GROWTH variable does not even have the expected sign. This is consistent too with the results we found in the analysis of the ergodic distributions. In the second row of Table 3.10 we estimated (3.17) again, leaving out insignificant variables. This does not have a substantial impact on the remaining estimated coefficients nor their significance. Hence our conclusions remain unchanged.

In this section we tested if two economic factors that in theory determine the tendency of business sizes to converge do indeed affect the degree to which business sizes converge in the US chemical sector. This is relevant as there are a number of competing theories that predict convergence, none of which implies that the number of firms and the geographical spread of production facilities is relevant for the degree of convergence. Empirically this is, however, the case, which strongly suggests that the mechanism we described in Section 3.2 is indeed the relevant one.

3.6 The Dynamics of Scope

In the previous sections we have argued that there are theoretical reasons to believe that sizes of businesses tend to converge, if businesses compete in homogeneous product markets. We showed that there is empirical support for this hypothesis in the chemical sector. However, since chemical corporations are typically multi-product firms, analyzing the dynamics of scale reveals only a partial picture of the overall growth dynamics of chemical corporations. We turn now to the dynamics of scope, i.e. the dynamics of the number of products a firm produces. This section is, however, not more than an illustration to point out that the dynamics of scope are markedly different from the scale dynamics. It is not meant as an in depth analysis, but included here as the dynamics of scope are complementary to the dynamics of scale, and as such crucial to understanding the overall growth dynamics of chemical corporations.

The profitability of existing varieties will be affected if a firm introduces a new variety which is highly substitutable on the demand side. This is the "demand side effect" that was analyzed previously, which puts larger firms at a disadvantage relative to smaller competitors. However, there is also a cost effect. If a firm produces a variety that uses a production technology that is closely related to the technology used by the new variety, then it is probably more profitable for that firm to carry the new product than it is for firms that only produce technologically unrelated products. In other words there is a "supply side effect" that favours already large firms, since there is a higher probability that the new variety is technologically 'close' to one of their existing products.

The chemical sector is an ideal candidate to analyze the dynamics of scope, given the enormous number of chemicals that are produced by a very small number of competitors. In the synthetic organic chemicals, such as cyclic intermediates, pesticides or surface active agents, most products are only produced by one or two firms, although in the industry the number of firms is much higher. Moreover, the substitutability on the demand side between chemicals is of an entirely different order relative to the homogeneous markets we analyzed in previous sections. Hence, if the dynamics of scope are indeed different from the dynamics of scale, then this should be the case in the chemical sector, if anywhere.

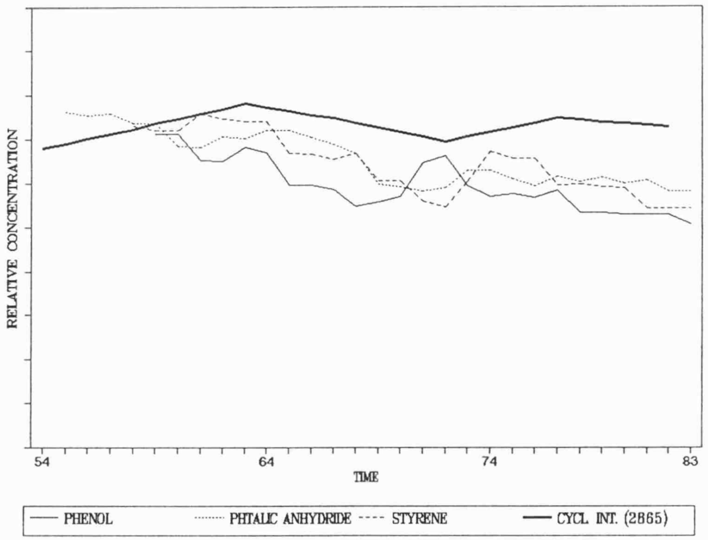
A simple illustration of the differences in dynamics between scale and scope is the comparison of the evolution of concentration on market level versus industry level. Figure 3.12 compares the evolution of the 4-firm concentration ratio C_4 relative to the 8-firm concentration ratio C_8 of Cyclical Intermediates (SIC 2865) with the evolution of C_2 relative to C_4 for Phenol, Phtalic Anhydride and Styrene, which are all cyclical intermediates[55]. This shows that the 2 largest firms lost relatively more market share on product market level, than the 4 largest firms on industry level. On industry level large firms seem to be able to overcome the disadvatages they face on product market level, and maintain their market share.

[55] For the industry level we took the Census of Manufactures data for 1954 - 1982. For the three product market we took annual data from the beginning of the sample until 1983 (see Table 3.1).

Figure 3.12

Evolution of Concentration in Cyclical Intermediates

(1954-1983)



3.6a A Theoretical Framework.

Consider an industry with N firms. Firm i produces n_i differentiated products, where that $n_1 \geq n_2 \geq \ldots \geq n_n$. Assume that a new product can be introduced by one firm only, though all firms are potential candidates. The introduction of the new product does not affect the profitability of existing products through demand side effects, such as business stealing or price effects. However, the profitability of producing the new product to firm depends on the extent to which it produces technologically related products. Rather than modelling "technological closeness" explicitly, we assume that the profitability of the new variety to firm i is the highest of n_i independent draws, π_i ,from a distribution F on $[0, \infty)$. Each draw can be interpreted as implying a certain distance in some technological space between one of the firm's existing products and the new variety. The profit is strictly decreasing in this distance. Consequently the shortest distance implies the maximum profitability.

Which firm wins the product is determined in a second-price sealed-bid auction, in which the firms know the number of products their competitors produce, but not the profitability of the new variety to each of them. In other words, firm i knows n_j but not $\pi_j \quad \forall j \neq i$. The firms submit costless bids $b_i \in [0, \infty)$ simultaneously. The firm with the highest bid wins the right to produce the product and pays the second bid. The other players pay nothing. Hence the pay offs are

$$\begin{array}{ccc} \pi_i - \max_{j \neq i} b_j & \textit{if} & b_i > \max_{j \neq i} b_j \\ 0 & \textit{if} & b_i < \max_{j \neq i} b_j \end{array} \qquad (3.18)$$

If more than one firm bids the highest price, the product is allocated randomly between them. It is well known[56] that in this setting each firm has a strictly dominant strategy, which is to bid its own valuation π_i . In equilibrium the probability that firm i wins, conditional on its valuation, is $\{F(\pi_i)\}^{\Sigma_{j \neq i} n_j}$, which is equal to the probability that firm i's valuation is the maximum. Firm i's unconditional probability of winning is therefore

[56] See for example Wilson (1992, p.236) or Fudenberg and Tirole (1991, p.10).

$$\int_0^{\infty} F(\pi)^{\Sigma_{j\neq i} n_j} n_j F(\pi)^{n_i - 1} f(\pi) d\pi = \frac{n_i}{\Sigma_1^N n_j} \tag{3.19}$$

where $f(.)$ is the probability density function of F. So, the ex-ante probability that firm i wins the project is equal to its market share in terms of number of products. If $n_i = n_j$, the firms have an equal probability of winning the project. If $n_i > n_j$ firm i is more likely to win the project, essentially because it gets more draws and is therefore more likely to have a higher valuation. Consequently, the already larger firm is more likely to be the introducer of the new product and firm sizes diverge relative to the average. What is not captured in this simple framework is the fact that firms tend to specialize in certain areas of the product space. However, a priori it is not clear in which direction this would bias the result.

If there is an exogenous probability p that an entrant will win the product, then it is straightforward to show that the probability that firm i will win the project becomes:

$$\frac{n_i(1 - p)}{\Sigma_1^N n_j}$$

Hence, the probability that a firm will win the product remains proportional to its market share.

An implications is that (expected) growth rates of firms are independent of their size. For firm i the expected growth rate is:

$$\frac{Pr(i\,wins)}{n_i} = \frac{(1 - p)}{(\Sigma_{j=1}^N n_j)} \tag{3.20}$$

The RHS of (3.20) is equal for all firms, and is nothing other than the growth rate of the market the firm operates in.

This framework can be interpreted as a behaviourial basis for Gibrat's Law, which asserts the independence of growth rates of firm size as a basis for the skewed distribution of firm sizes[57]. Gibrat's Law, otherwise known as the Law of Proportionate Effect, is one of the most widely tested propositions in industrial economics[58], though the theoretical foundations for it seem rather weak[59]. Lucas (1967) and Prescott and Visscher (1980) show that Gibrat's Law holds in an environment with constant returns to scale and adjustment costs. However, there is no strategic interaction in either model.

Although this theoretical framework is highly stylized and the auction mechanism not a very plausible way of describing the competitive process between firms of the introduction of new products, it formalizes the really quite obvious point that firm sizes tend to diverge if the large firm has an inherent cost advantage over the small firm in the production of the new variety. Formulating the proposition in somewhat weaker terms, we would expect to find weaker converging tendencies for the dynamics of scope than we found for the dynamics of scale in the product markets, as the demand side effects are relatively weaker and the cost side effects stronger.

The data allow us to define firms and markets at various levels of aggregation, thereby varying the 'average distance' of the products on the supply side. The higher the level of aggregation, the less technologically related the products are on average, as more or less independent subgroups are pooled. This change in average distance between the products does not affect the relative advantage large firms have over small ones. As only the highest draw counts, the advantage of firm i over firm j remains n_i / n_j , as follows from (3.19). The change in average distance affects both the large and the small firm in a similar way, keeping their relative positions

[57] See Scherer and Ross (1990, p. 143) for an outline of the proof of the Law.

[58] See Klomp and Thurik (1995), and Sutton (1995b) for a summary of the literature.

[59] This was already noted by Simon and Bonini (1958). They suggest that the Law of Proportionate Effect holds for firms above the MES, as the cost curve shows virtually constant returns to scale. There is widespread empirical support for the Law holding particularly well for larger firms. Klomp and Thurik (1995) find that in certain industries of the Dutch service sector for which the MES is extremely small the Law holds for the entire size distribution.

unchanged. Hence the diverging effect ought to be the same irrespective of the level of aggregation.

We now turn to the empirical analysis of the dynamics of scope, employing the non-parametric methods that were used in the analysis of the dynamics of scale.

3.6b Variable of Analysis.

The size of a firm is measured in terms of number of products it produces, normalized by the average size of the incumbent firms. That is, if there are N_j firms in market j, the size of firm ij producing n_{ij} products is

$$PRS_{ij} = \frac{n_{ij}}{\frac{\Sigma_{k=1}^{N_j} n_{kj}}{N_j}}$$

To establish what is the total number of firms that can introduce the next product, potential entry is once again an issue. However, taking potential entry into account will only pronounce the diverging effect as it will increase persistence in the lower extreme states. So we restrict our analysis to the dynamics of the incumbent firms only.

3.6c The Data.

The data are from the annual issues of *Synthetic Organic Chemicals, US Production and Sales*, in which the International Trade Commission lists all synthetic organic chemicals that have been produced in a certain year, with their respective producers. The products are grouped according to use and chemical characteristics. Table 3.11 shows summary statistics of the groups of products that are considered here. They were chosen primarily on the basis of (i) a sufficient number of products and firms operating in the industry and (ii) low number of firms per product, in order to minimize the dynamics induced by demand size effects. For Cyclical Intermediates, Flavours and Perfumes and Medical Chemicals only aggregate data were collected, whereas for Pesticides and Surface Active Agents the products were also divided into subcategories (see Table 3.11). Firms are defined in terms of the size of their operations in the 'relevant' market, where 'relevant' refers to the level of aggregation.

3.6d Empirical Results.

Table 3.12a gives the estimated transition matrices for Pesticides and Surface Active Agents. The first line gives the upper end of each state, the second line the number of the state, the first column the number of transitions starting in that state and the remaining is the estimate of the transition matrix, of which the interpretation is as before. Table 3.12b gives the ergodic distributions for all analyzed industries. The (bold) first line of each industry gives the upper end of the states, that were selected such that the initial distributions are uniform. The (italic) second line gives the ergodic distribution. For Surface Active Agents and Pesticides the ergodic distribution is the one implied by the respective transition matrices given in part (a) of the table. For those industries for which we had only 4 year observations, a five state transition matrix was estimated, from which the ergodic distributions follow.

For all industries we find evidence of an ergodic distribution which is bimodal relative to the initial distribution. For Surface Active Agents and Pesticides the two states at both extremes have the highest probability mass, whereas there is thinning out in the middle states. Although the highest probability mass is not in the extreme states throughout, the thinning out in the middle states is substantial, certainly relative to what we found earlier for the product market dynamics. A similar pattern is found for Cyclical Intermediates, Flavours & Perfumes and Medical Chemicals. For Cyclical Intermediates and Medical Chemicals the middle state shows thinning out, for Flavours and Perfumes the trough is in state 4. Although each industry in itself might not be convincing, it is the fact that we find a tendency to thinning out in the middle states throughout that suggests that the dynamics of scope drive firm sizes apart, leading to a few large firms, and possibly some fringe firms. From the current analysis we cannot conclude if the small firms stay in the market or ultimately exit the industry.

Table 3.11 Summary Statistics of Synthetic Organic Chemicals.

	Number of Firms		Number of Products*	Average Number of	Number of Producers
YEARLY OBSERVATIONS:	min	max		Products per Firm*	per Product*
PESTICIDES 1958-1981	69	90	249	4.26	1.33
CYCLIC PESTICIDES	60	81	172	3.42	1.33
Fungicides	24	41	34	1.85	1.47
Herbicides & Plant Hormones	3	41	85	3.03	1.28
Insect Attractants & Repellants	1	6	2	1.00	2.00
Insecticides	23	36	44	2.48	1.30
All Other Cyclic Pesticides	1	12	7	1.29	1.29
ACYCLIC PESTICIDES	30	39	77	2.64	1.34
Fungicides	10	20	17	2.30	1.35
Herbicides & Plant Hormones	10	17	23	1.59	1.17
Insecticides	9	18	21	2.89	1.24
All Other Acyclic Pesticides	10	19	16	1.42	1.69
SURFACE ACTIVE AGENTS 1958-1964	134	184	507	7.80	2.80
CYCLIC	109	151	125	2.53	2.98
ACYCLIC	76	105	382	9.97	2.74
4-YEAR OBSERVATIONS:					
CYCLICAL INTERMEDIATES 1962-1978	142	202	1065	10.46	1.39
FLAVOURS AND PERFUMES 1970-1990	28	58	301	10.75	1.39
MEDICINAL CHEMICALS 1970-1990	101	149	373	3.69	1.27

* last year of sample period.

Table 3.12
Transition Matrices and Ergodic Distributions of Synthetic Organic Chemicals.

(a) 7-State Transition Matrix of Surface Active Agents (1958 - 1964)

Upper End	.124	.248	.434	.678	1.051	1.981	∞
r	(1)	(2)	(3)	(4)	(5)	(6)	(7)
113	*.60*	*.25*	*.06*	*.02*	*.02*	*.03*	*.03*
138	*.42*	*.28*	*.15*	*.06*	*.03*	*.03*	*.03*
120	*.08*	*.32*	*.42*	*.08*	*,08*	*.01*	*.02*
123	*.01*	*.02*	*.13*	*.63*	*.16*	*.03*	*.01*
117	*.02*	*.03*	*.05*	*.14*	*.62*	*.14*	*.01*
114	*.02*	*.04*	*.03*	*.03*	*.10*	*.69*	*.11*
123	*.03*	*.04*	*.02*	*.00*	*.01*	*.11*	*.79*

Pesticides (1958 - 1981)

Upper End	.240	.258	.481	.692	1.055	2.019	∞
r	(1)	(2)	(3)	(4)	(5)	(6)	(7)
351	*.54*	*.23*	*.13*	*.03*	*.03*	*.01*	*.02*
360	*.11*	*.63*	*.09*	*.06*	*.06*	*.01*	*.03*
329	*.23*	*.13*	*.32*	*.22*	*.05*	*.03*	*.01*
341	*.02*	*.08*	*.19*	*.48*	*.18*	*.04*	*.02*
362	*.01*	*.02*	*.09*	*.21*	*.53*	*.12*	*.02*
390	*.02*	*.00*	*.01*	*.02*	*.15*	*.72*	*.09*
380	*.02*	*.01*	*.01*	*.01*	*.03*	*.09*	*.83*

(b) Ergodic Distributions

State	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Surface Active Agent	**.124**	**.248**	**.434**	**.678**	**1.051**	**1.981**	**∞**
	.201	*.144*	*.110*	*.118*	*.133*	*.154*	*.140*
Pesticides	**.240**	**.258**	**.481**	**.692**	**1.055**	**2.019**	**∞**
	.111	*.152*	*.103*	*.138*	*.155*	*.165*	*.174*
Cyclical Intermediates	**.119**	**.276**	**.460**	**1.16**	**∞**		
	.184	*.208*	*.127*	*.221*	*.259*		
Flavours and Perfumes	**.062**	**.153**	**.350**	**1.486**	**∞**		
	.157	*.207*	*.225*	*.176*	*.235*		
Medical Chemicals	**.128**	**.278**	**.628**	**1.751**	**∞**		
	.162	*.203*	*.098*	*.227*	*.311*		

* Four year observations

3.6e The Effect of Disaggregation.

For Pesticides and Surface Active Agents data were available on a more disaggregated level. In this section we will show how demand effects play an increasingly important role as the markets become more disaggregated. In the last section it was shown how in certain industries large firms can always find a geographical submarket in which it has no existing capacity as of yet. By opening a plant there a firm can limit the negative externality on the profitability of its existing capacity, because if transportation costs are substantial, there are only local price effects. A similar pattern can occur in product space. One can think of the industry consisting of a number of submarkets in which products that are close substitutes on the demand side are lumped together, whereas the lumps are relatively differentiated from each other. Within each lump there is a disadvantage to being large: since the products are close substitutes on the demand side, large firms face a disadvantage in the battle to introduce the new variety, relative to smaller ones. However, between the lumps these disadvantages do not exist. A firm with a large number of products can always find a submarket in which it has no presence yet. Introducing a new product does not affect the profitability of existing products through negative demand side effects. The International Trade Commission claims that the products are grouped according to use, rather than chemical characteristics[60].

As we argued before, changes in the technological relatedness ("average distance") does not matter for the supply side effect as long as the profitability of the new variety is determined by the best draw. The advantage large firms have over their small competitors is fully determined by the existing number of products a firm carries, even if we move to different aggregation levels. The demand effect becomes weaker on higher levels of aggregation, since products tend to become less substitutable, whereas supply effects remain constant. The theoretical prediction is therefore that the more aggregated the markets, the stronger the diverging tendencies are relative to the converging tendencies from demand side effects.

[60] See the annual issues of Synthetic Organic Chemicals (Introduction).

We determined PRS on various levels of aggregation for Pesticides and Surface Active Agents. For both industries we pooled the data on each level of aggregation and estimated the transition matrices, fixing the grid equidistantly in logarithmic terms, in order to make the ergodic distributions comparable as we did before. Figure 3.13 shows the ergodic distributions for the three levels of aggregation of Pesticides. On the highest level of aggregation, the industry is assumed to be one market. Measuring PRS on this level, we find that the ergodic distribution is bimodal, with the modes in the extreme states. Disaggregating the Pesticides industry into Cyclic and Acyclic Pesticides and measuring PRS on those levels (2 markets), but pooling the data again for estimation purposes shows that the ergodic distribution is still bimodal, but one mode is not in the extreme state any longer. Analyzing PRS in a similar way for the most disaggregated level, distinguishing 7 markets, reveals a unimodal ergodic distribution. Consistent with the theoretical prediction we find that the converging tendencies are stronger the more disaggregated the markets are. Figure 3.14 shows that a qualitatively similar conclusion continues to hold for Surface Active Agents, although only two levels of aggregation are considered.

Finally we analyzed how these results compare to the Gini-indices of business sizes, to show that they are consistent with our previous results. Table 3.13 shows the Gini-indices for Pesticides and Surface Active Agents on various levels of aggregation. The more disaggregated the markets are, the smaller the Gini-indices are, which holds in all but one case (Acyclic Insecticides is the exception). So, that the differences in firm dynamics on various different levels of aggregation are consistent with the differences in degree of asymmetry among firm sizes on corresponding levels of aggregation. As firm sizes tend to be less diverging at a disaggregated level, the size distribution should also be less skewed, implying a smaller Gini-index.

Figure 3.13

Effects of Aggregation: Pesticides (1958 - 1981).

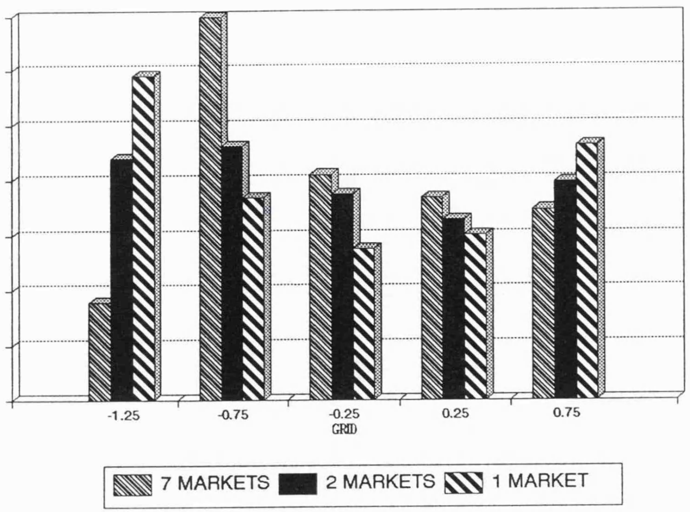


Figure 3.14

Effects of Aggregation: Surface Active Agents (1958 - 1964).

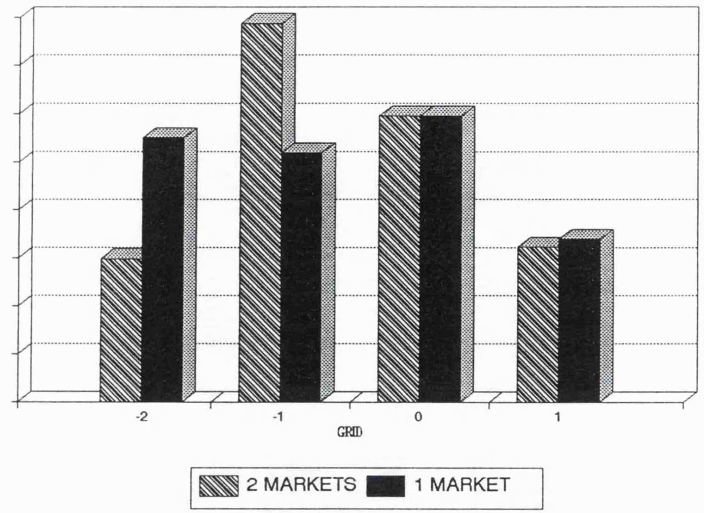


Table 3.13 Gini-indexes of Pesticides and Surface Active Agents.

	Gini-index		
PESTICIDES (1981)	0.5130		
CYCLIC PESTICIDES		0.4830	
Fungicides			0.3077
Herbicides and Plant Hormones			0.4784
Insect Attractants and Repellants			0.0
Insecticides			0.3224
ACYCLIC PESTICIDES		0.3777	
Fungicides			0.3478
Herbicides and Plant Hormones			0.3240
Insecticides			0.4188
SURFACE ACTIVE AGENTS (1964)	0.5699		
CYCLIC		0.4667	
ACYCLIC		0.5604	

3.7 Discussion and Conclusion.

We showed how firm dynamics can be divided into two components, the dynamics of scale and the dynamics of scope. The former is a tendency of firm sizes to converge, which operates if products of competing producers are highly substitutable, as they are at a disaggregated product market level. The dynamics of scope drive firms apart and leads to increasing dominance of large firms. Although the emphasis has been on pointing out the difference in dynamics of scale and scope, we found some support for the claim that as we move to higher levels of aggregation, the disadvantage that large businesses face at the product market level is mitigated on firm level, through opportunities to diversify.

However, the contribution of this chapter is not only to describe the two components of firm dynamics, it is also to show that the mechanism that drives the dynamics of scale is indeed the disadvantage large businesses face relative to smaller competitors through the price decreasing effect of additional capacity. Alternative theories that also imply convergence are for example decreasing returns to scale or managerial diseconomies. We found evidence that the converging tendencies are stronger in those markets where large firms are more likely to suffer the disadvantage of falling prices should they increase capacity. That is, we found that converging tendencies are stronger if the industry is concentrated, if production is geographically concentrated and if firms cannot take advantage of a spreading production geographically as it operates single plant businesses.

We also suggested a mechanism that could explain the diverging tendency in the dynamics of scope, based on production synergies between products with similar characteristics. If a firms has a large presence in a particular area of the product space, then it supposedly has substantial expertise to produce that kind of products, assuming that similar products have related production technologies. This puts the firm with a large presence in a better position to be the introducer of a new variety than a competitor that only produces unrelated products.

If the growth of an industry consists of increasing demand for existing products and a continuous inflow of new products, then the overall effect of growth on firm dynamics and therefore concentration is indeterminate. The analysis of this chapter can thus serve as an explanation for the phenomenon that concentration hardly falls despite dramatic increases in market size, as illustrated in Table 1.2.

Appendix 3A Proof of Proposition 3.4.

Proposition 3.4

If the selection mechanism is a stochastic race then in equilibrium

(i) the success date of the large firm stochastically dominates the winning date of the small firm in the sense of first order stochastic dominance if $\Delta PCM > 0$.

(ii) the success date of the small firm stochastically dominates the winning date of the large firm in the sense of first order stochastic dominance if $\Delta PCM < 0$.

Proof. A full characterization of the equilibrium conditions is given in Reinganum (1983), in particular Proposition 1 on p. 744, which applies mutatis mutandis. The proof here is confined to the issues that differ from that analysis.

The FOCs for the optimal response of firm i if firm j invests z_j are:

$$\frac{h'(z_i)}{r}\left[\pi\left(r + h(z_j)\right) + \Delta PCM q_i\left(r + (1-\gamma)h(z_j)\right) + rz_i\right] = r + h(z_i) + h(z_j) \quad i,j = 1,2 \tag{3A.1}$$

These define implicitly the optimal response functions $\phi_i(z_j, \Delta PCM, \gamma, q_1, q_2)$ $i,j = 1,2$. By the implicit function theorem,

$$\partial\phi/\partial q_i = \frac{-\partial^2 V_i/\partial z_i \partial q_i}{\partial^2 V_i/\partial z_i^2} \tag{3A.2}$$

The denominator is negative by the SOC. The numerator is,

$$-\partial^2 V_i/\partial z_i \partial q_i = \frac{-\Delta PCM}{\left(r + h(z_i) + h(z_j)\right)^2}\left[\frac{h'(z_i)}{r}\left\{r + (1-\gamma)h(z_j)\right\}\right] \tag{3A.3}$$

which is negative if $\Delta PCM > 0$. Hence $\partial\phi_i/\partial q_i > 0$ and t_1 dominates t_2 in the sense of first order stochastic dominance since $h(.)$ is increasing. If $\Delta PCM < 0$ then $\partial\phi_i/\partial q_i < 0$ and t_2 dominates t_1 in the sense of first order stochastic dominance. QED.

Appendix 3B The Logit Model of Plant Openings.

In this appendix we present the results of testing the Empirical Hypothesis using the conventional discrete-choice approach.

The empirical model estimates the probability of business i opening a plant, conditional on at least one business in the industry opens one:

$$Prob\,(x_{ij,t} = 1\,|SUMOP_{j,t} > 0) = F\left(MSH_{ij,t},\ SUMOP_{j,t},\ SUMF_{j,t}\right)$$

Where: (3B.1)

$SUMOP_{j,t}$	= total number of plants opened in industry j between t and t+1.
$x_{ij,t}$	= 1 if business i opens a plant in industry j between t and t+1, 0 otherwise.
F	= logistic distribution function.
$MSH_{ij,t}$	= business i's total capacity in industry j $(q_{ij,t})$ relative to the total industry capacity: $MSH_{ij,t} = q_{ij,t}/\Sigma_i\ q_{ij,t}$.
$SUMF_{j,t}$	= total number of businesses in the industry and potential entrants.

We will test for alternative functional forms of the (log) odds ratio $h(.)$. Our basic specification is linear:

$$h_t = a + b\,MSH_t + c\,SUMOP_t + d\,SUMF_t + u_t \qquad (3B.2)$$

The reason for conditioning on the number of openings (SUMOP) is that the theoretical model does not predict the time-series dynamics of the arrival of investment opportunities. It only makes a statement which business is more likely to win the project given that it is available. The implicit assumption of this procedure is that the arrival process of new investment opportunities (projects) is independent of the business' characteristics. The validity of this will be tested below[61].

[61] See Section 3B.3.

In some of the estimated models MSH will be replaced by an alternative measure of relative business size PLSH, which is the number of plants of business i in industry j relative to the total number of plants currently in operation in j.

A change in the number of businesses competing for the projects changes the probability of business i opening a plant, c.p. the relative sizes. But not only incumbent businesses compete for opening new plants, potential entrants do too. This, as before, begs the question on how to deal with potential entry. In the dataset potential entry is not observable. If businesses actually enter, it is not clear for how long they have been a potential entrant. All that can be said is that it has been for at least two years, since it takes on average two years to build a new plant[62]. We will again make alternative assumptions on how long businesses have been potential entrants before actually entering, to show that the results are robust in this respect. We will also assume that all potential entrants will enter over the course of the sampling period. Exiting businesses remain potential entrants for some years by assumption. Hence SUMF depends on the specific assumption that is made in this respect.

Large businesses will open and close more plants than small businesses due to replacement of old plants by new ones, effectively not changing the industry capacity. To control for this bias, a business opening a plant t will be considered as replacement if the business closed one between t-2 and t+2.

3B.1 Testing Hypothesis 1.

A testable hypothesis H_0 that follows from the Empirical Hypothesis is that the probability of opening the next plant is decreasing in the size of the business, i.e. $b < 0$, although in Section 3.3b we claimed that the interpretation of a negative coefficient of the initial condition as indicating convergence is not necessarily valid.

[62] See Lieberman (1987).

Table 3B.1 show the result of the ML estimators of (3B.2) for alternative specifications. Column (1) reports the estimated coefficients assuming that entrants have been around to grab a project for four years before opening their first plant, taking into account the two years it takes to build one. The coefficient of MSH is negative and significant. Hence we cannot reject H_0 on first sight. The coefficients of both SUMF and SUMOP have their expected signs. Column (2) corresponds to a specification of the model that includes non-linear effects of MSH. The quadratic term is significant. It indicates that for MSH < 0.37 the relationship between the probability of opening and MSH is negative, though it is positive for higher values. It should be noted that 95% of the observations of MSH are below this critical value in the sample.

Some misspecification tests were performed. To test for industry specific effects we estimated the model (3B.2) for each industry. The estimates of b were unstable, often insignificant and not always negative. Allowing b and the coefficient of MSH^2 to be industry specific generated significant estimates only for a few industries. Adding industry dummies to the equation (3B.2) did not show significant industry effects.

Testing for time varying effects, time dummies were added to the original specification (3B.2). They were all insignificant and the estimation results did not change substantially. Splitting the sample into two subsamples, pre and post oil crisis (1973) shows that there is a structural break, but qualitatively the results are unchanged.

The coefficients in column (3) are estimated under the same entry assumption as (1) and (2), but replacing MSH by PLSH. The conclusions are similar to those for (2). Also industry and time effects are as before. In (4) the sample is changed, assuming that all entrants could have opened a plant since the beginning of the sampling period and all exiting businesses stay around as potential entrants until the end of the sampling period. The earlier conclusions remain unchanged.

The results of estimating (3B.2) using the subsample of incumbent businesses only are reported in column (5). In various alternative specifications, both MSH and PLSH become insignificant. Industry by industry, none of the estimates for *b* is significant. However, there might be a problem of sample selection[63] in the sense that we ignore the part of the sample with initial size being zero, since a business does not enter the sample until it has a positive size. The estimation of the model for incumbents should strictly speaking take into account that the initial size being equal to zero is a truncation point in the considered sample[64]. On the other hand, to consider the whole sample as we did in the first place assumes that an entrant's decision of opening a plant can be described by the same model as the incumbent's. This is in line with the theoretical result since both respond to the same motivation, i.e. the fixed profit of the plant and the externality on the existing capacity.

Although these first results show some support for the theoretical results, they seem very sensitive to the exact empirical specification and the sample that is used.

3B.2 Testing Hypothesis 2.

Another testable implication of the theory is that the relative size of the business is a sufficient statistic for determining which business is most likely to win the project. Here we test if any other business specific explanatory variables have significant explanatory power. The variable we will consider is the change in MSH_t over the prior two years: $DSH_t = MSH_t - MSH_{t-2}$. A negative sign of DSH_t indicates that businesses invest to maintain their market share[65]. Gilbert and Lieberman (1987) show that particularly large businesses follow this behaviour, whereas small businesses incremental investment is *positively* related to recent changes in MSH.

[63] For sample selection problems in this context see Hall (1987).

[64] This might be related to Gilbert and Lieberman (1987) finding a positive sign for MSH in their logit model of the probability of incremental capacity expansions. Their sample is also restricted to incumbents.

[65] See Gilbert and Lieberman (1987) for a more detailed economic interpretation.

The model is estimated including this variable and its interaction with MSH to account for the non-linear effect. The sample is the same as in the first three specifications. The results are reported in column (6) of Table 3B.1. The t-statistics of DSH and the interaction term are individually insignificant. To test for the joint significance, we ran a specification excluding DSH and the interaction term and used a Likelihood Ratio (LR) test. The joint significance can be rejected at a 5% significance level[66]. Hence we cannot reject the hypothesis that only the initial condition is a sufficient business specific characteristic for explaining the observed probabilities. Similar conclusions hold if the analysis is done with PLSH.

3B.3 Testing Hypothesis 3.

The theoretical results on the cross-sectional allocation are conditional on the arrival of the project. An implicit assumption underlying the first two tests is that the unconditional dynamics of scale can be split into two parts. The first being the arrival of the projects, the second their cross-sectional allocation. The theoretical results apply to the second part, as they are conditional on the projects being available. This assumes implicitly that the arrival process is independent of individual business characteristics and is entirely determined by industry characteristics, which is the third hypothesis that will be tested here. Rejection would imply that not all businesses are potential candidates of winning the project if it has arrived, since that depends on a business being in a particular state. This again would undermine the game theoretic analysis.

Assume that the probability that a business is in a market in which a project arrives between t and t+1 has a logistic distribution function. The (log) odds ratio $h(.)$ of that distribution function depends on a number of market characteristics, notably past production growth and capacity utilisation. To test for the significance of business specific variables, we estimate a model including both business specific variables and industry wide averages, nesting (3B.2):

[66] Twice the difference in Likelihood Ratio is 0.1, which is distributed as a χ^2 with 2 degrees of freedom.

$$h_t = \alpha + \beta MSH_t + \gamma MSH_t^2 + \delta DSH_t + \zeta CU_t + \eta GROW_t + \theta SUMF_t + u_t$$

where: (3B.3)

- Production growth $GROW_t$ is the annualized four year growth rate of the industry production $(Q_{j,t})$: $GROW_{j,t} = \left(Q_{j,t}/Q_{j,t-4}\right)^{\frac{1}{4}} - 1$
- Capacity utilisation:

$$CU_{j,t} = \frac{Q_{j,t-1}}{\Sigma_i (q_{ij,t} + q_{ij,t-1})} + \frac{Q_{j,t-2}}{\Sigma_i (q_{ij,t-1} + q_{ij,t-2})}$$

The first three explanatory variables (MSH, MSH^2 and DSH) are business characteristics, the last three (CU, GROW and SUMF) are industry averages[67]. Capacity utilisation and production growth have been used by Lieberman (1987) to explain plant openings by entrants and incumbents, and by Gilbert and Lieberman (1987) to explain incremental capacity increases[68]. The definition of these variables follows the latter study. SUMF indicates the level of concentration in the industry, actingessentially as an industry dummy.

The null hypothesis H_0 is that the business specific variables are jointly insignificant. To test for that we estimate a model excluding from (3B.3) all business specific variables.

The estimation results are shown in Table 3B.1, column (7)[69]. Only the results of

---

[67] The annual industry output data are from Synthetic Organic Chemicals, US Production and Sales (ITC).

[68] In Gilbert and Lieberman's logit model of incremental capacity expansions an apparently firm specific variable called BAND is included. This is the percentage by which all producers other than i collectively increase their capacity during the year of observation, testing for firms to "jump on the bandwagon", i.e. invest if competitors do so. They find a significant positive effect for small firms. BAND is, however, highly correlated with the total growth rate of market capacity (SUMOP, ρ = 0.93 for the sample for which (3B.3) was estimated). This implies that in any model in which BAND is included as an independent variable capacity expansion is explained by the industry capacity growth itself.

[69] This model was estimated on a subsample of the industries, referred to as subsample 6 in Table 3.1, due to limited production data availability.

(3B.3) are shown, but the results of the nested model are not very different for the industry averages. A LR-test shows that H_0 cannot be rejected at the 5% confidence interval[70]. Hence the arrival process is independent of business characteristics, and conditioning on the number of plant openings in the industry, as in the earlier tests, is valid. This also validates the theoretical assumption that the cross-sectional allocation can be analyzed given the arrival of the project.

Table 3B.1 Logit Analysis of Plant Openings (Estimated coefficients)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
MSH	-7.14* (1.53)	-15.65* (2.59)		-5.85* (1.53)	3.37* (0.98)	-15.53* (3.02)	1.66 (3.21)
MSH^2		20.92* (5.52)		8.14 (2.81)		22.15 (7.00)	-1.33 (7.14)
DSH						-0.23 (2.93)	-0.54 (2.89)
PLSH			-28.75* (2.59)				
$PLSH^2$			50.50* (5.64)				
SUMOP	0.28* (0.04)	0.25* (0.04)	0.27* (0.05)	0.23* (0.03)	0.16* (0.04)	0.24* (0.04)	
SUMF	-0.18* (0.02)	-0.17* (0.02)	-0.21* (0.03)	-0.11* (0.03)	-0.05 (0.03)	-0.17* (0.03)	-0.39* (0.05)
CU							3.28** (1.38)
GROW							9.88* (3.07)
CONST	0.27	0.27	-1.81	0.13	-2.13	0.45	1.61
No of Obs	1642	1642	1642	2223	807	1522	533
Log Likelihood	-606.8	-584.9	-527.6	-706.4	-291.2	-541.4	-213.9

Heteroscedastic consistent (White) SE in parentheses.
* Significant at 1% (One tailed test).
** Significant at 5% (One tailed test).

70 The statistic is 0.68, which is distributed as a χ^2 with 2 degrees of freedom.

CHAPTER 4

INTRA-INDUSTRY DIFFERENCES IN PROFITABILITY: THEORY AND EVIDENCE[71]

4.1 Introduction.

In the past, both the theoretical and the empirical literature on firm profitability have been dominated by the question whether industry factors (like concentration) or firm specific factors (like firm size) are prime determinants of firm profitability[72]. In this chapter we find that there is a relationship between profitability and size, but the nature of this trade off is determined by industry characteristics

We show in theory that the profitability of large firms is higher *relative* to their smaller competitors in industries in which the firms' pay offs are highly sensitive to competitors' actions. This sensitivity comes for example from firms producing goods that are close substitutes on the demand side, in terms of either horizontal or vertical product differentiation. In other words, we analyze industry factors that determine differences in the slope of the profitability/size trade off across industries. Hence we do not make predictions about the circumstances under which large firms have a higher or lower profitability than their smaller rivals. We show that these can only be made in some extreme cases.

The theoretical approach we take is somewhat non-standard. Rather than starting from the description of the game and characterizing the equilibrium, we *assume* that

---

[71] The conclusions presented herein are those of the author and have not been adopted by the Federal Trade Commission or any entity within the Commission. The FTC's Disclosure Avoidance Officer has certified that the data included in this chapter do not identify individual company line of business data.

[72] See Schmalensee (1989) and Scherer and Ross (1991) for a summary of this literature.

size differences between firms are the outcome of a (subgame-perfect) Nash equilibrium of a game, for which we only specify the order of moves and some weak restrictions on pay offs. Our conclusions are based on the notion that in equilibrium there is no incentive for any of the firms to deviate. The robustness of our result is shown by considering multiple strategic variables and the analysis of a number of different structures of the game.

The advantage of this approach is twofold. Firstly it circumvents the need of an explanation for firm asymmetries, allowing us, however, to make statements about firm profitability *if* those asymmetries are equilibrium outcomes. Secondly, showing the robustness of the results comes at very little costs as we only have to put weak restrictions on profit functions.

We also show that no similarly robust relationship can be derived between profitability and industry concentration. Although Lambson (1987) is a rare attempt to derive such a relationship in a game theoretic setting, we show that it depends crucially on the extensive form of the game.

In the second part of this chapter we test the empirical implications for the profitability-size relationship, in a horizontal product differentiation setting. Using the homogeneous product markets of the FTC Line of Business Data, we test if large firms are more profitable relative to small firms when the industry is essentially one integrated market, or when the industry consists of a set of localised markets. Transport distance is hence interpreted as an empirical proxy for the sensitivity of a firm's profitability to the competitor's action. We find some strong empirical support.

The determinants of firms' profitability have been at the centre of a long lasting debate. The advocates of the Structure-Conduct-Performance paradigm, established by Bain (1951), claim that profitability is determined by industry specific 'barriers to entry', like product differentiation, economies of scale and absolute capital requirements (Baumol (1953)). More recently other barriers have been suggested such as excess capacity (Dixit (1980)), investments in R&D or advertising (Gilbert and Newbery (1982) and Schmalensee (1983)), product proliferation (Schmalensee (1978)), durability of capital (Eaton and Lipsey (1980, 1981)) and contracts with customers (Aghion and Bolton (1987)). With free entry, firms make positive profits due to the integer value of the number of entrants, as was illustrated in Chapter 2. The other point of view is that profitability is determined by firm specific characteristics. Its most prominent advocate, Demsetz (1973), claimed that superior efficiency of large firms is what leads to higher profitability rather than monopolistic power. Although we find that profitability is related to firm size, the nature of this relationship is determined by industry factors, such as the substitutability of products on the demand side.

This theoretical debate has its counterpart in the empirical literature, predominantly in the form of testing firm specific factors (e.g. market share and the firm's advertising and R&D outlays) against industry factors (e.g. concentration, diversification and MES) as determinants of profits. Many empirical studies, like Shepherd (1972), Gale (1972), Gale and Branch (1982), and Smirlock, Gilligan and Marshall (1984) find a significant positive effect of market share on profitability. Most of the studies that use the FTC Line of Business data, such as Ravenscraft (1983), find a positive effect of market share on profitability[73]. Amato and Wilder (1985), Mueller (1986) and Ravenscraft (1983) find a strong effect of the interaction of market share and advertising or R&D outlays. Industry variables, being generally concentration and industry wide advertising-to-sales ratios, generate mixed results. Most of the earlier studies, like Shepherd (1972) and Gale (1972) find a significant positive effect of the concentration ratio, though less significant than the market share variable. For example in studies by Gale and Branch (1982), Odagiri and Yamawaki (1990) and Smirlock et al. (1984), the positive effect of the concentration vanishes if market share is included as an explanatory variable. Studies using the FTC Line of Business data find a weakly negative relationship between concentration

[73] See Ravenscraft et al. (1987), and Scherer and Ross (1987) for an overview of the effect of market share and concentration on profitability in various studies that use this data set.

and profitability. In Ravenscraft (1983) the industry wide advertising-to-sales ratio is significant but the firm level ratio is not, though for R&D it is the other way round.

The organisation of the chapter is as follows. In Section 4.2 we derive our basic theoretical result, that is the relationship between profitability and firm characteristics, such as production level, advertising or R&D outlays. In Section 4.3 we show that an equally robust relationship does not exist for the profitability - concentration relationship. The empirical implications for the profitability-size relationship are tested in Section 4.4. Section 4.5 concludes.

4.2 Profitability and Firm Characteristics.

In this section the relationship between a firm's profitability and its characteristics will be derived. We initially assume that firms have one strategic variable only, say the production level, although the model can be interpreted in terms of other strategic variables such as number of product varieties marketed, the level of advertising and R&D outlays, or the price a firm charges. We show that firms with higher production levels earn higher equilibrium profits whenever there is a negative externality from firm i's action on firm j's pay off. In a very similar way we can put bounds on profitability differences between firm, which are a function of the difference in firm size. We analyze the effects of efficiency differences and strategic asymmetries in order to show how robust the results are. In Appendix 4B we show to what extent our results hold if there are multiple strategic variables.

4.2a Profits and Size.

Consider a game with a set of $n \geq 2$ firms, N. All firms choose a level of a strategic variable - say production - $q_i \in Q_i \subset \mathbb{R}^+$ simultaneously. For simplicity assume that the set of production levels Q_i is identical for all firms, i.e. $Q_i = Q$. Define the profile of production levels as $q = \{q_i\}_{i \in N}$, where $q \in Q^n$. $\pi_i(q_i, q_{-i}) \in \mathbb{R}$, defined over Q^n , is the pay off (net profit) of firm i, if i produces q_i and the profile of all other firms' production levels is $q_{-i} = \{ q_j \}_{j \in N \setminus \{i\}}$.

Assumption 4.1 *Symmetry.*
If $q_i = q_j$ then $\pi_i(q) = \pi_j(q) \quad i,j \in N$.

If two firms choose the same level of production, they will have the same pay offs. This is a weak form of symmetry, which does not imply that the profit functions are identical (no anonymity). Assumption 4.1 is consistent with firms facing a downward sloping average cost curve, but not with cost differences for a given production scale. The latter will be analyzed in the next section.

Assumption 4.2 *Monotonicity.*
π_i is weakly decreasing in q_j, $\forall j \neq i$.

There is a negative market externality. The profit of firm i is decreasing in the production level of all other firms[74]. Assumption 4.2 requires (1) that the externality that firm i exerts on j is of identical sign to the externality of firm j on i and (2) monotonicity. Note that no assumption is made on how profits evolve with changes in the firm's own level of production q_i , given q_{-i} .

In what follows, only pure strategy equilibria will be considered.

Proposition 4.1 *(Profits)*
If Assumptions 4.1 and 4.2 hold and there exists a Nash equilibrium s.t. $q_k^* > q_l^*$, then $\pi_k(q^*) \geq \pi_l(q^*)$. ‖

In an asymmetric equilibrium, firms with higher levels of production earn a higher pay off than firms with smaller production levels. The intuition for this result is that

[74] See Section 4.2 for an example of a set of profit functions that satisfy Assumption 4.1 and 4.2, but do not have an identical functional form (no anonymity).

if a larger firm earns a lower equilibrium pay off than a smaller firm, the former has an incentive to deviate and choose the equilibrium strategy of the latter. The deviation increases the pay off of the (previously) smaller firm, because of the negative externality. Since after deviating both firms have the same production level, they earn the same pay off. The deviating (large) firm will earn a pay off that is at least as high as the *equilibrium* pay off of the smaller firm. Consequently, for there not to be an incentive for the larger firm to deviate, it must earn a pay off that is higher than the equilibrium pay off of the smaller firm.

Proof of Proposition 4.1

The proof is by contradiction. Suppose there is an equilibrium such that $\pi_k(q^*) < \pi_l(q^*)$, $q_k^* > q_l^*$. Consider the deviation by firm k to q_l^* . By Assumption 4.2[75]

$$\pi_l(q_l^*, q_l^*, q_{-l,k}^*) \geq \pi_l(q^*) \tag{4.1}$$

By Assumption 4.1

$$\pi_k(q_l^*, q_{-k}^*) = \pi_l(q_l^*, q_l^*, q_{-lk}^*) \tag{4.2}$$

Hence $\pi_k(q_l^*, q_{-k}^*) \geq \pi_l(q^*) > \pi_k(q^*)$, which is inconsistent with the equilibrium condition. QED.

If the strategic variable is the level of production, there is a positive relationship between the relative size of a firm (its market share) and its absolute profit. Assumption 4.2 occurs naturally through the (negative) market externality of the choice of production level. Alternatively, the strategic variable could be R&D outlays. A positive relationship exists then between profits and R&D outlays if an increase in firm i's R&D decreases the probability of a rival's success. A similar reasoning holds if advertising is the strategic variable, or the number of product varieties marketed by a firm. A 'large' firm is then to be interpreted as the firm with

---

[75] $\pi_l(q_l^*, q_l^*, q_{-l,k}^*)$ is the pay off of firm l, which follows strategy q_l^* , firm k is deviating to q_l^* , and the vector of all other firms' strategies is $q_{-l,k}^*$.

the high level of the strategic variable, be it advertising or R&D, or the number of product varieties. In these cases the negative market externality seems a plausible assumption.

However, the strategic variable can also be the price a firm is charging. Then the market externality is likely to be positive. But, if the strategic variable is assumed to be the inverted price, Assumption 4.2 holds and Proposition 4.1 goes through unchanged. This highlights firstly that in Assumption 4.2 it is the monotonicity rather than the negative market externality that matters. Secondly, it shows that the strategic variable is not necessarily related to any sensible measure of 'size' of the firm. Therefore, the notion that there exists an unconditional relationship between profits and firm size is too general, and depends critically on the strategic variable employed.

Within this theoretical framework no attempt has been made to explain why, starting from an ex-ante symmetric situation, an asymmetric equilibrium might occur. Only consequences have been derived that will be satisfied *if* such equilibria exist[76]. This approach was primarily empirically inspired as most industries show a substantial degree of asymmetry. In the next section we allow for efficiency differences, which can be seen as the driving force behind firm asymmetry.

4.2b Profit Rates and Size.

Proposition 4.1 can be restated in terms of rate of returns, although only at the expense of some generality. Firstly, interpret q_i as an investment, e.g. as the level of production capacity, advertising or R&D outlays. Define the net rate of return of firm i as $r_i(q) = \pi_i(q)/q_i$.

Secondly, assume π_i is strictly positive and continuously differentiable with respect to all elements of q_{-i} . Define the externality of firm j's production on firm i's profit as the partial cross-elasticity: $\varepsilon_{ij} = | \frac{\partial \pi_i}{\partial q_j} . \frac{q_j}{\pi_i} |$. Thirdly, assume this cross-

[76] See section 4.2c for examples of models that fit into this framework.

elasticity is bounded:

Assumption 4.3 *Bounded externality.*

ε_{ij} is bounded from above and below:

$$0 \leq \underline{\varepsilon} \leq \varepsilon_{ij} \leq \bar{\varepsilon}, \forall\, i, j \neq i, q \tag{4.3}$$

This assumption extends Assumption 4.2 in the sense that the externality is not only negative, but the cross-elasticity is also bounded.

We will relax the assumption of ex-ante symmetry (4.1) and allow for efficiency differences to show that our results are robust in this respect. Efficiency differences might occur due to past investments in, for example, R&D, which are not explicitly modelled here. Define the efficiency difference between firm i and j as the difference in profitability if both produce at identical scales: $\delta_{ij}(q) = r_i(q)/r_j(q)$ if $q_i = q_j$. If $\delta_{ij}(q) \geq 1$, then firm i is said to be (weakly) more efficient than j, given q. Assume that

Assumption 4.4

δ_{ij} is bounded from below:

$$\delta_{ij}(q) \geq \underline{\delta}_{ij} \geq 0 \quad \forall\, q \tag{4.4}$$

Assumption 4.4 is a generalization of Assumption 4.1. If δ_{ij} is constant and equal to 1 $\forall\, q$, they are identical.

Proposition 4.2 *(Rate of Return).*

If Assumptions 4.2, 4.3 and 4.4 hold and

- there exists an equilibrium such that $q_k^* > q_l^*$,
- firm k is more efficient than l ($\delta_{kl}(q) \geq 1 \geq \delta_{lk}(q)$, $\forall\, q$)

then there is an upper and a lower bound to the ratio of profit rates r_k^*/r_l^* :

$$\underline{\delta}_{lk}^{-1}\left[\frac{q_k^*}{q_l^*}\right]^{\bar{\varepsilon}-1} \geq \frac{r_k^*}{r_l^*} \geq \underline{\delta}_{kl}\left[\frac{q_k^*}{q_l^*}\right]^{\underline{\varepsilon}-1} \tag{4.5}$$

Proof.

Only the proof of the lower bound is given here. The proof of the upper bound is similar and can be found in Appendix 4A.

Consider firm k deviating from q_k^* to q_l^* . Assumption 4.3 $\varepsilon_{lk} \geq \underline{\varepsilon}$ implies that $-\frac{\partial \pi_l}{\pi_l} \geq \underline{\varepsilon}\frac{\partial q_k}{q_k}$. Taking the integral from q_l^* to q_k^* gives:

$$\ln\pi_l(q_l^*, q_l^*, q_{-kl}^*) - \ln\pi_l(q^*) \geq \underline{\varepsilon}(\ln q_k^* - \ln q_l^*) \tag{4.6}$$

and hence $\pi_l(q_l^*, q_l^*, q_{-k,l}^*) \geq \gamma^{\underline{\varepsilon}}\pi_l(q^*)$, where $\gamma = q_k^* / q_l^*$.
Firm k's equilibrium pay off must satisfy:

$$\pi_k(q^*) \geq \pi_k(q_l^*, q_{-k}^*) = \pi_l(q_l^*, q_l^*, q_{-k,l}^*)\,\delta_{kl}(q_l^*) \geq \pi_l(q_l^*, q_l^*, q_{k,l}^*)\underline{\delta}_{kl} \tag{4.7}$$

and therefore

$$\pi_k(q^*) \geq \gamma^{\underline{\varepsilon}}\,\pi_l(q^*)\,\underline{\delta}_{kl} \tag{4.8}$$

The lower bound follows if the LHS and the RHS of (4.8) are divided by $\pi_l(q^*)\gamma$. QED.

The intuition is similar to that for Proposition 4.1, which continues to hold if $\underline{\delta}_{kl} = 1$ as $\underline{\varepsilon} \geq 0$ and $\gamma > 1$. If the larger firm is always the more efficient firm ($\underline{\delta}_{kl} \geq 1$) and it earns a lower profit than the less efficient firm, then the deviation to the production level of the smaller firm is profitable, not only because of the negative market externality which raises the profits of the small inefficient firm, but also because the deviating firm is more efficient and makes higher profits if they play the same strategy. Efficiency differences shift both bounds upwards.

It is important to note that Proposition 4.2 does not imply any restriction on the sign of the profitability/size relationship. It only says that the bounds rotate anti-clockwise as the (bounds on the) cross-elasticity ($\underline{\varepsilon}$ and $\bar{\varepsilon}$ resp.) increase. However, for certain values of $\underline{\varepsilon}$ and $\bar{\varepsilon}$ we can show that the profitability/size trade off is always positive or negative in the absence of efficiency differences.

Corollary 4.1

If there are no efficiency differences $(\delta_{kl} = 1, \forall q)$ and $\underline{\varepsilon} > 1$, then the large firm always has a higher equilibrium profit rate than the small firm. If, on the other hand, $\bar{\varepsilon} < 1$ then the small firm always has a higher equilibrium profit rate than the large firm. ‖

If in (4.5) $\underline{\varepsilon} > 1$ and $\delta_{kl} = 1, \forall q$, then the lower bound $\gamma^{\underline{\varepsilon}-1} > 1$ and the large firm is always more profitable $(r_k^* > r_l^*)$, implying a positive profitability/size relationship. If $\bar{\varepsilon} < 1$, then the upper bound $\gamma^{\bar{\varepsilon}-1} < 1$ and the small firm is always more profitable $(r_k^* < r_l^*)$, implying a negative profitability/size relationship. The intuition is that if the externality is sufficiently strong $(\varepsilon_{ij} > 1 \quad \forall\, i,j \neq i, q)$, the profits of the small firm are raised by a substantial amount if the large firm deviates. The difference in equilibrium profits must therefore be sufficiently high for an asymmetric equilibrium to exist. If the externality is weak $(\varepsilon_{ij} \in\,]0,1[\quad \forall\, i,j \neq i, q)$, then the small firm's deviation has little effect on the large firm's pay off, and the profits differences must be small for there to be no incentive for the small firm to deviate. Consequently, the small firm has a higher equilibrium profit rate.

4.2c Examples.

In this section we will first work out a simple example of a model of product differentiation, using explicit functional forms. It illustrates what Propositions 4.1 and 4.2 predict and, maybe more importantly, what is not implied by them. We then consider some examples in the literature for which Propositions 4.1 and 4.2 hold.

4.2c.1 Example 4.1.

Consider two firms, k and l, each of which produces one differentiated product. The demand faced by firm i, producing good i is:

$$p_i = q_i^{\alpha_i - 1} q_j^{\gamma - \alpha_i} \quad i \in \{k, l\} \ \wedge \ i \neq j \tag{4.9}$$

where $\alpha_i \in [0, 1)$ and $\alpha_i > \gamma$. The demand is, up to a logarithmic transformation, identical to a standard linear model of product differentiation[77]. The cost of producing q_i is $c(q_i) = q_i^{\delta}$, where $\delta > \alpha_i \ \ i \in \{k, l\}$. Hence the profit of firm i is

$$\pi_i(q_i) = q_i^{\alpha_i} q_j^{\gamma - \alpha_i} - q_i^{\delta} \tag{4.10}$$

In order to make the interpretation of this model similar to Proposition 4.2, we assume that firms choose production levels. They move simultaneously.

The first-order conditions are

$$\alpha_i q_i^{\alpha_i - 1} q_j^{\gamma - \alpha_i} - \delta q_i^{\delta - 1} = 0 \quad i \in \{k, l\} \ \wedge \ i \neq j \tag{4.11}$$

from which follows that the ratio of equilibrium levels of production is[78]:

$$\frac{q_k^*}{q_l^*} = \left(\frac{\alpha_l}{\alpha_k} \right)^{\frac{1}{\alpha_k + \alpha_l - \gamma - \delta}} \tag{4.12}$$

Firm k is larger in equilibrium than firm l if either

$$\alpha_l > \alpha_k \quad \wedge \quad \alpha_k + \alpha_l > \delta + \gamma \tag{4.13a}$$

or

$$\alpha_l < \alpha_k \quad \wedge \quad \alpha_k + \alpha_l < \delta + \gamma \tag{4.13b}$$

is satisfied.

We will show how the various theoretical results of Section 4.2b are satisfied in this

[77] See Tirole (1988, Section 7.1) and Deneckere and Davidson (1985).

[78] The SOC is satisfied if $\alpha_k, \alpha_l < \delta$.

example.

First consider Proposition 4.1. Assumption 4.1 (ex-ante symmetry) is satisfied, since if both firms produce at identical levels $q_k = q_l = \bar{q}$, the pay offs are identical: $\pi_k(\bar{q},\bar{q}) = \pi_l(\bar{q},\bar{q}) = \bar{q}^{\gamma} - \bar{q}^{\delta}$.

Assumption 4.2 (negative externality) is also satisfied as $\gamma < \alpha_k, \alpha_l$.

From (4.11) it follows that the equilibrium profit of firm i is $\pi_i^* = \dfrac{\delta - \alpha_i}{\alpha_i} q_i^{*\delta}$. Hence the equilibrium ratio of profits is

$$\frac{\pi_k^*}{\pi_l^*} = \frac{\delta - \alpha_k}{\delta - \alpha_l}\frac{\alpha_l}{\alpha_k}\left[\frac{q_k^*}{q_l^*}\right]^{\delta} \tag{4.14a}$$

$$= \frac{\delta - \alpha_k}{\delta - \alpha_l}\left[\frac{\alpha_l}{\alpha_k}\right]^{\frac{\delta}{\alpha_k + \alpha_l - \delta - \gamma} + 1} \tag{4.14b}$$

We show in Appendix 4D.1 that if conditions (4.13a) or (4.13b) are satisfied, then $\pi_k^* / \pi_l^* \geq 1$. So the (absolute) equilibrium profits is increasing in equilibrium size, as $q_k^* / q_l^* > 1$. This confirms Proposition 4.1 and is illustrated in Figures 4.1 and 4.2. In each figure equation (4.14a) is simulated for two sets of α and γ, and over a whole range of values for δ. The relationship between (absolute) profits and size is always positive as $\pi_k^* / \pi_l^* > 1$ whenever $q_k^* / q_l^* > 1$. Moreover, π_k^* / π_l^* increases if the size difference increases, though this is outside of the scope of Proposition 4.1. The remainder of Figures 4.1 and 4.2 is explained below.

We now turn to Proposition 4.2. Assumption 4.3 (bounded externality) is satisfied if we make an auxiliary assumption, that is, that the firms need to make a minimum return on costs (or investment) R to be viable. So

$$\frac{p_i q_i - c(q_i)}{c(q_i)} \geq R \qquad i \in \{k, l\} \tag{4.15}$$

Furthermore, we assume that conditions (4.13b) are satisfied, although a similar analysis can be done if (4.13a) holds.

Assumption 4.3 (bounded externality) then becomes[79]:

$$\frac{1 + R}{R}(\alpha_k - \gamma) \geq \varepsilon_{kl}, \varepsilon_{lk} \geq \alpha_l - \gamma \tag{4.16}$$

In order to make the bounds on ε monotonic in γ , we will assume henceforward that $\gamma < 0$.

Assumptions 4.4 is satisfied in the sense that there are no efficiency differences ($\underline{\delta}_{ij} = 1 \ \forall \ ij \in \{kl, lk\}$) .

From (4.11) follows that the equilibrium rate of return is $r_i^* = \frac{\delta - \alpha_i}{\alpha_i} q_i^{*\,\delta - 1}$. And therefore:

$$\frac{r_k^*}{r_l^*} = \frac{\delta - \alpha_k}{\delta - \alpha_l} \frac{\alpha_l}{\alpha_k} \left[\frac{q_k^*}{q_l^*} \right]^{\delta - 1} \tag{4.17a}$$

$$= \frac{\delta - \alpha_k}{\delta - \alpha_l} \left[\frac{\alpha_l}{\alpha_k} \right]^{\frac{\delta - 1}{\alpha_k + \alpha_l - \delta - \gamma} + 1} \tag{4.17b}$$

For this example Proposition 4.2 states that

$$\left[\frac{q_k^*}{q_l^*} \right]^{(\frac{1+R}{R})(\alpha_k - \gamma) - 1} \geq \frac{r_k^*}{r_l^*} \geq \left[\frac{q_k^*}{q_l^*} \right]^{\alpha_l - \gamma - 1} \tag{4.18}$$

In Appendix 4D we show that these bounds are indeed satisfied.

---

[79] See Appendix 4D for details.

What are the empirical implications for the profit rates? Firstly, Proposition 4.2 does <u>not</u> predict that the profitability/size relationship is positive or negative. That either can be the case is illustrated in Figures 4.1 and 4.2 (broken lines). In Figure 4.1 the ratio of equilibrium sizes q_k^* / q_l^* is above 1, whereas the ratio of profit rates r_k^* / r_l^* is above and below one. The larger firm is either more or less profitable, depending on the parameter constellation. Hence there is either a positive or negative relationship between firm size and profitability. For all parameter constellations in Figure 4.2 this relationship is positive. The ratio of equilibrium sizes q_k^* / q_l^* is above 1, and so is the ratio of profit rates r_k^* / r_l^* .

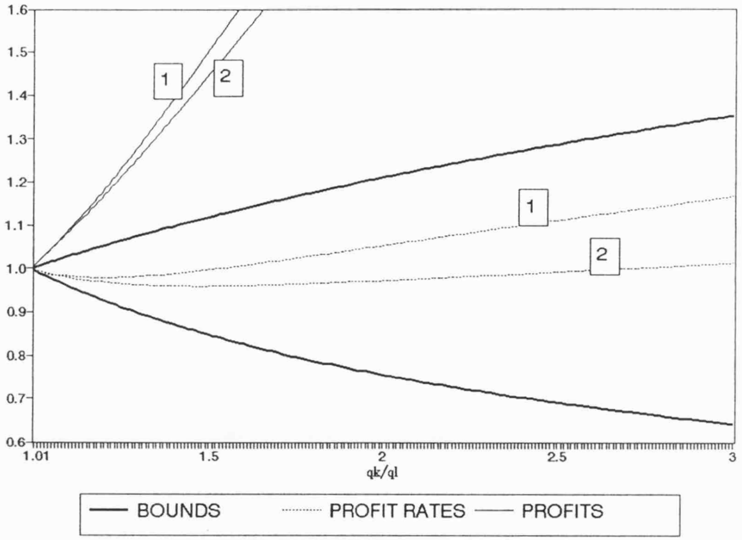
<u>Neither</u> does Proposition 4.2 predict that differences between profitability should become larger or smaller as equilibrium size differences increase. Figures 4.1 and 4.2 illustrate that either can happen, as the broken line can be upward or downward sloping.

The parameters in both Figures 4.1 and 4.2 are chosen such that the bounds on the cross-elasticity of the profit functions (relationship 4.16) are equal for all parameter constellations in each figure. In Figure 4.1 the bounds are lower than in Figure 4.2. The bold lines in Figures 4.1 and 4.2 are the corresponding bounds on the relative profit rates (relationship 4.18). Again, Proposition 4.2 does <u>not</u> restrict these profitability bounds to be upward or downward sloping. Figures 4.1 and 4.2 illustrate that indeed either can happen, depending on the parameter constellation.

So what <u>is</u> the empirical prediction of Proposition 4.2? It is that the bounds on the profit rates rotate anti-clockwise with higher $|\gamma|$, i.e. the higher the (bounds on the) cross-elasticity. This is illustrated in Figure 4.3, where the broken bounds are those corresponding to higher (bounds on the) cross-elasticity (i.e. higher $|\gamma|$). Similarly, $|\gamma|$ is higher in Figure 4.2 than in Figure 4.1, the cross elasticity ε is higher, and hence the bounds on the profit rates in Figure 4.2 are rotated anti-clockwise relative to those in Figure 4.1.

Figure 4.1

Relative Profits and Profitability.
(Example 4.1)



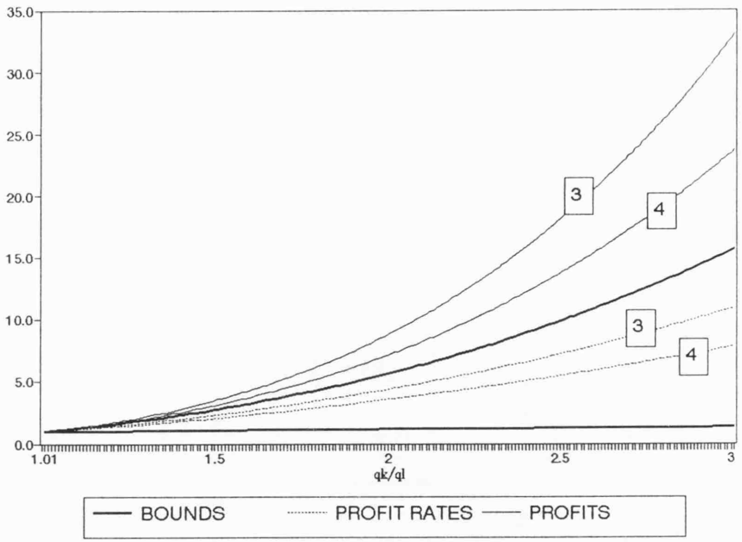
$R = 2$ $\qquad \dfrac{(1 + R)}{R}(\alpha_k - \gamma) = 1.275$ $\qquad \alpha_l - \gamma = 0.60$

1: $\alpha_k = 0.75$ $\qquad \alpha_l = 0.50$ $\qquad \gamma = -0.10$ $\qquad \delta \in [1.72, 42.1]$

2: $\alpha_k = 0.60$ $\qquad \alpha_l = 0.35$ $\qquad \gamma = -0.25$ $\qquad \delta \in [1.69, 55.4]$

Figure 4.2

Relative Profits and Profitability.
(Example 4.1)



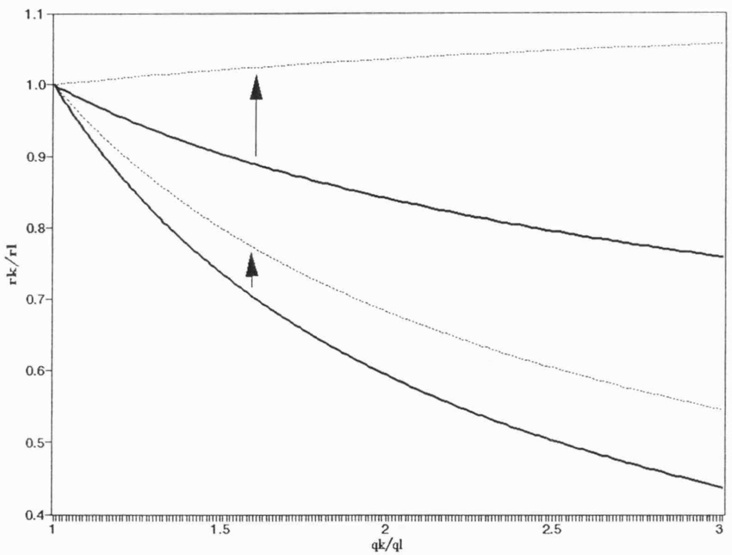
$R = 0.75$ $\qquad \dfrac{(1 + R)}{R}(\alpha_k - \gamma) = 3.50$ $\qquad \alpha_l - \gamma = 1.25$

3: $\alpha_k = 0.75$ $\qquad \alpha_l = 0.50$ $\qquad \gamma = -2$ $\qquad \delta \in [3.62, 44]$

4: $\alpha_k = 0.60$ $\qquad \alpha_l = 0.35$ $\qquad \gamma = -2$ $\qquad \delta \in [3.44, 57.1]$

Figure 4.3

The Effect of an Increase in $|\gamma|$
on the Bounds on Relative Profitability.
(Example 4.1)



Alternatively, assume there is a number of industries with identical R, α_k and α_l . Industries differ in their cross-elasticity γ and cost structure δ . The cross-elasticity can be either high or low (resp. $\bar{\varepsilon}$ or $\underline{\varepsilon}$), through an appropriate choice of γ (resp. $\underline{\gamma}$ and $\bar{\gamma}$). Each industry j has its unique cost structure δ_j . Those industries that have a common ε have identical bounds on the ratio of profitability, but the precise location of the industry equilibrium in the $(r_k / r_l , q_k / q_l)$ space is determined by the industry specific cost structure δ_j . Proposition 4.2 predicts that the bounds of the $\bar{\varepsilon}$ industries are rotated anti-clockwise relative to the bounds of the $\underline{\varepsilon}$ industries. This is what we will test in Section 4.4: the profitability/size trade off in the set of industries with high ε is rotated anti-clockwise relative to the profitability/size trade off of the set of industries with low ε ,

4.2c.2 Examples in the Literature.

An example that fits into the framework of Proposition 4.1 is D'Aspremont et al. (1983). In this model of cartel stability all firms are ex-ante identical. In equilibrium members of the price-leader cartel are the smaller, lower profitability firms, whereas the price-taking, competitive fringe firms are the larger ones, earning higher net pay offs. None of the (small) cartel firms has an incentive to deviate to the competitive fringe, because the market price would fall and the post-deviation profit is lower than the equilibrium profit. Similarly, none of the (large) fringe firms have an incentive to join the cartel and become a 'small' firm. As follows from Proposition 4.1 this can only be the case if the fringe firms earn higher profits than the cartel firms[80].

In Shaked and Sutton's (1990) two stage game of multi-product firms, firms decide on the number of (horizontally differentiated) product varieties they will sell in the first stage of the game, and compete in prices in the second stage. Given the Nash equilibrium of the second stage, the reduced profit function of the first stage is decreasing in the number of varieties marketed by the competitor. Hence, by

[80] In their example this is always satisfied, since $\pi_{fringe} = 2/(4-\alpha^2)^{-2} > [2(4-\alpha^2)]^{-1} = \pi_{cartel}$, where α is the proportion of firms that are members of the cartel.

Proposition 4.1, there is a positive relationship between the scope and profit of a firm[81].

Another example of an asymmetric outcome with ex-ante identical firms is analyzed in 2. If the strategic variable is the level of fixed costs, then the negative externality originates from the fact that higher fixed costs come with lower marginal costs and therefore with higher production levels, which depresses performance of competitors. In these terms, Lemma 2.1 and Proposition 4.1 are directly comparable.

The simplest example for which Proposition 4.3 holds is a Cournot model with constant, but given, asymmetric marginal costs. The low marginal cost firm has a higher market share and earns a higher profit. Another example is Rosen's (1991) model of R&D with asymmetric firm sizes.

However, the results do not hold in some models of vertical product differentiation, of which Shaked and Sutton (1982a, b) are examples. In these models price competition can be relaxed through product differentiation, i.e. through a larger difference in advertising and/or R&D outlays. If the high advertising firm increases its outlays, price competition between the competitors is weakened, and the low advertising firm has more (local) monopoly power and earns a *higher* profit. (See Lemma 4 in Shaked and Sutton (1982a).) If, however, the low advertising firm increases its outlays, the monopoly power of the high advertising firm decreases and it earns a *lower* profit. Assumption 4.2 is violated in the sense that i exerts a different externality on j than vice versa. It is not necessarily the case that the net

---

[81] Shaked and Sutton (1990) give a three good example with simultaneous choice in the first stage. For certain parameters, an asymmetric equilibrium with a two-product firm and a one-product firm exists, though firms are ex-ante identical. Let $\pi(x,y)$ be the profit of a firm that markets x varieties, whereas the competitor markets y. In their example $\pi(2;1) = \frac{4(2-\sigma)(4+3\sigma)^2}{(1+\sigma)(8+4\sigma-\sigma^2)^2}\phi > \pi(1;2) = \frac{4(2-\sigma)(2+\sigma)^3}{(1+\sigma)(8+4\sigma-\sigma^2)^2}\phi$, where $\sigma \in [0,2]$ is the degree of substitution between the products on the demand side and ϕ is the number of consumers in the (two and) three good market.

pay off of the high advertising firm is larger than that of the low advertising firm. Although the revenues of the high advertising firm are higher than those of the low advertising firm, it depends on the shape of the cost curve whether the net profit of the high advertising firm is indeed larger. In Shaked and Sutton (1982a) this is the case for any continuous differentiable cost function that is 'sufficiently convex'. But it is straightforward to show that if the cost function is, for example, a step function the opposite result can hold.

Gilbert and Vives' (1986) model of entry deterrence is an example of another class of models in which Assumption 4.2 might be violated. In their model asymmetric equilibria exist, such that for each firm it is optimal to produce the difference between the - industry wide - entry deterring production level and the total production level of its competitors. If the large firm deviates to being small, additional firms will enter, and the profits of the small firm will decrease. Hence the large firm exerts a positive externality on the small firm, but the externality of the small firm on the large one is negative, like in a standard Cournot model. Assumption 4.2 is violated and Proposition 4.1 does not necessarily hold.

4.2d Strategic Asymmetry.

We now consider strategic asymmetries among firms, in order to show the robustness of the bounds in Proposition 4.2, relation (4.5). Strategic asymmetry is modeled in the context of a sequential move game, in which one of the firms move before the other.

Consider a two stage game with two firms {1, 2}. In stage one, firm 1 decides on its production level $q_1 \in Q$, which cannot be revised in stage two. In the second stage, firm 2 chooses its production level, knowing firm one's choice. Firm 2's strategy is $s(q_1) \in Q$, its best response to q_1 is $s^*(q_1)$. The pay offs are realised after stage two. Let $\pi_i(q)$, defined over Q^2 , be the pay off of firm i if q is the production profile $\{q_1, q_2\}$. The firms can be equally efficient or have different levels of efficiency, denoted by δ_{ij}. The equilibrium concept is subgameperfection[82]

Proposition 4.3 *(Strategic Asymmetries)*
If Assumptions 4.2, 4.3 and 4.4 hold and
- there exists an equilibrium such that $q_k^* > q_l^*$,
- firm k is more efficient than l ($\delta_{kl}(q) \geq 1 \geq \delta_{lk}(q), \ \forall q$)
then, the profit rates satisfy:

$$\zeta_k \underline{\delta}_{lk}^{-1} \left[\frac{q_k^*}{q_l^*} \right]^{\bar{\varepsilon}-1} \geq \frac{r_k^*}{r_l^*} \geq \zeta_l \underline{\delta}_{kl} \left[\frac{q_k^*}{q_l^*} \right]^{\underline{\varepsilon}-1} \tag{4.19}$$

where
$\zeta_k = \left(s^*(q_k^*)/q_k^*\right)^{\bar{\varepsilon}} > 1$ if k=2 and s^* is upward sloping, $\zeta_k = 1$ otherwise.
$\zeta_l = \left(q_l^*/s^*(q_l^*)\right)^{\underline{\varepsilon}} < 1$ if l=2 and s^* is downward sloping, $\zeta_l = 1$ otherwise.

Proof. See Appendix 4A.

The bounds in (4.5) do not necessarily hold if the firms move sequentially. If for example the large firm k moves first, then the small firm will change its action after a deviation by the large firm. If the actions are strategic substitutes[83] in the sense that the small firm increases its production if the large firm deviates to q_l^* , the deviation profit of k will be lower than if l had not revised its production level, due to the negative externality. The fact that the small firm can respond to the deviation of the large firm reduces the lower bound. A similar reasoning holds for the upper bound.

However, under strategic asymmetry it is also the case that with higher (bounds on the) cross-elasticity, the profitability/size trade off is rotated anti-clockwise. Whether the one shot or a sequential move game is the most appropriate description of strategic interaction between firms is ultimately an empirical matter and depends on

[82] See Fudenberg and Tirole (1991).

[83] See Bulow et al. (1985).

the industry under consideration, but this empirical regularity holds independent of the precise structure of the game.

4.3 The Concentration-Profit Relationship.

The positive relationship between concentration and profitability has been a long defended theoretical result. However, in the class of models of industry structure with free entry, Lambson (1987) is a rare example in which this relationship actually occurs, the driving force being the lumpiness of technology. In this section some examples will be analyzed that shows how Lambson's result delicately depends on the strategic form of the game. A negative relationship can be derived if the extensive form is changed from Lambson's two stage set up to a one shot game or to incumbent firms that commit to production levels before potential entrants decide to enter or stay out. The difference in strategic form of market interaction is empirically difficult, if not impossible, to observe. The empirical consequences of Lambson's model are therefore not clear cut.

4.3a Lambson's Argument.

In a symmetric equilibrium, the number of firms N^* (integer) satisfies:

$$\Pi_{N^*} \geq F > \Pi_{N^*+1} \tag{4.20}$$

where F are fixed costs of entry and Π_N is the equilibrium (*gross*) operating profit if N^* identical firms enter. Lambson's (implicit) assumption here is that the game has two stages. Firms decide to enter the industry in the first stage and choose their production level in the second stage.

Define the profit rate as $R = \frac{\Pi_N}{F}$. Then from (4.20) it follows that

$$1 < R \leq \frac{\Pi_{N^*}}{\Pi_{N^*+1}} \equiv R_{N^*}. \tag{4.21}$$

where R_{N^*} puts an upper bound on R for a given market concentration (1/N). If R_{N^*} is decreasing in N^* , the interval $[1, R_{N^*}]$ is smaller for more fragmented industries. Lambson (1987) claims that under fairly weak conditions one can observe a negative relationship between N^* and R_{N^*} . Consequently a positive relationship will occur between profit rate and equilibrium concentration.

Example 4.2: Iso-elastic demand, constant marginal costs and Cournot competition. In this case (4.21) is:

$$1 < R \leq \frac{(N^* + 1)^2}{N^{*2}} = R_{N^*} \tag{4.22}$$

Here R_{N^*} is indeed decreasing in N^* and the positive correlation between rate of return and concentration is likely to occur. One can take this analysis one step further and derive explicitly that firms in more concentrated industries make higher profits. Defining the industry concentration as $C = 1/N$ one can rewrite R_{N^*} as:

$$R_{N^*} = (1 + C^*)^2$$

Hence, the upper bound of profits is higher, the higher the concentration of the industry. Figure 4.4 shows the relationship between profits and market size.

Condition (4.20) is the equilibrium condition of a subgame-perfect equilibrium if firms enter in the first stage. If a firm deviates from its equilibrium strategy of staying out and enters the industry in the first stage, the incumbents will adjust their production choice in the second stage since they are a function of the number of entrants in the first stage. The firms, including the deviant, will be symmetric in terms of production and pay offs. However, the result that R_{N^*} decreases in N^* critically depends on the sequential choice of entry and level of production.

Figure 4.4

Profits in the Sequential Move Game.
(Example 4.2)

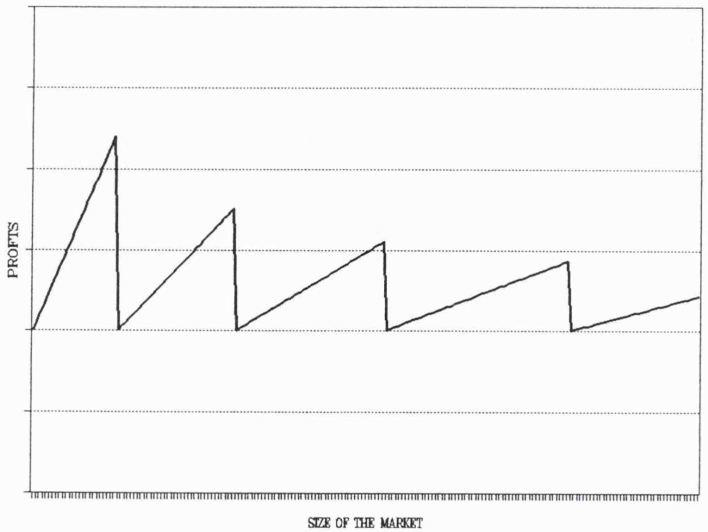
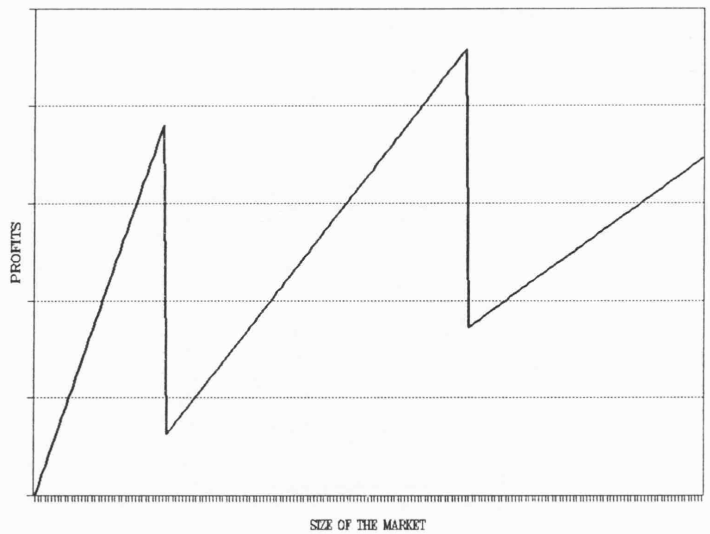


Figure 4.5

Profits in the Simultaneous Move Game.
(Example 4.3)



4.3b Simultaneous Choice of Entry and Production Level.

If the choice of entry and production level is simultaneous or alternatively, if incumbents can commit to output levels in stages of the game prior to entry[84], then the deviant chooses its production optimally *given* the (equilibrium) production choices of all other firms. Incumbent firms will not change their actions in response to the deviation. Though all firms are symmetric in equilibrium, the production level of the deviant will be different in general. Relationship (4.20) is no longer the appropriate equilibrium condition because the deviant's pay off will be different from the incumbents'. For there not to be an incentive for an additional firm to enter, the (net) profit of the deviant should be non-positive. In general,

$$\Pi_{N^*} \geq F > \Pi^D_{N^*+1} \tag{4.23}$$

where $\Pi^D_{N^*+1}$ is the (gross) operating profit which the marginal player that stays out of the industry would earn, if it were to deviate and enter the industry.

Applying a similar derivation as in (4.21) to (4.23) it follows that

$$1 < R \leq \frac{\Pi_{N^*}}{\Pi^D_{N^*+1}} \equiv R^D_{N^*} \tag{4.24}$$

The upper bound to the profit rate $R^D_{N^*}$ might be decreasing in C . There does not seem to be any general economic reason why the profit of the (potential, unobserved) deviant should increase relative to the equilibrium profits as N increases. This will be illustrated below.

There are two types of equilibria in these models of entry deterrence. One is where incumbents behave like Cournot competitors, given the number of entrants N . Their equilibrium strategy is the Cournot best response, given the output of other firms and the number of firms N. The other set of equilibria is where a smaller number of firms enter and 'fill' the market, i.e. produce collectively at an entry

[84] See Gilbert and Vives (1986). In stage one, a given number of incumbents decide on their output level, in stage two a potential entrant can decide to enter and choose its production level.

deterring level. Although they do not follow the Cournot best reply, their strategy is optimal given that additional entry is prevented. In other words, for each firm it is optimal to produce the difference between the industry's minimum entry deterring level and the sum of the production of all other firms. If there is a finite number of potential entrants that enter simultaneously, there exists a size of the market for which it is profitable for the marginal incumbent to deviate to the Cournot behaviour and allow the entrants in.

In the following two examples, these equilibria are illustrated using standard demand functions. In the first, only the Cournot equilibria are derived and it is shown that there exists a negative relationship between profit and concentration where in Lambson's (1987) examples a positive relationship is obtained. In example 4.4 entry deterring equilibria can be calculated explicitly and it is shown that the relationship between maximal obtainable profit for a given number of incumbents is decreasing in the concentration[85].

Example 4.3: Iso-elastic demand, constant (identical) marginal costs and Cournot competition. Then

$$R^D_{N^*} = \frac{C^{*2}}{-2(1 - C^*)^{0.5} + 2 - C^*} = 1 + \sqrt{1 - C^*}$$

which is decreasing in C^* . Hence the positive relationship between the upper bound of profitability and industry concentration is not obtained. This is illustrated in Figure 4.5.

Example 4.4: Linear demand, (identical) constant marginal costs and Cournot competition. If only Cournot equilibria are considered, then

$$R^D_{N^*} = 4$$

Little can be said about the relationship between R and N^* .

[85] See Appendix 4D for the details of these examples.

For entry deterring equilibria, let $R_N^{\max}$ be the maximal (average) equilibrium profit rate that can be achieved for a given number of incumbents (N). This is the profit made in the symmetric equilibrium for which the 'marginal' firm is indifferent between playing the entry deterring strategy or deviating, reducing its production level to the Cournot best response and thereby inducing an entrant to enter the industry. In Appendix 4D it is shown that if there is only one potential entrant, then with linear demand, (identical) constant marginasl costs and Cournot competition

$$R_N^{\max} = 2[3 + 2\sqrt{2} - C]$$

Hence the maximal obtainable profit is decreasing in the market concentration.

The intuition for the difference in outcome between Lambson's (1987) two stage game on the one hand and the one shot game or Gilbert and Vives' (1986) commitment game on the other, is that the collective adjustment as a result of new entry tends to increase with the market size, and hence the price fall which a deviant causes decreases. It therefore becomes profitable for additional entrants to come in at a lower size of the incumbent's profits as the size of the market increases. In the case of a one shot or a commitment game, the incumbents do not adjust their production as a result of additional entry. The profitability of the potential entrant falls relative to the profitability of the incumbents as the market size increases, due to the fact that in a larger market the price will be lower, hence the production level of the deviant must be higher to make entry profitable. This in turn, however, will cause a larger price fall. Incumbents can then, through deterrence, keep successfully additional entrants out at increasing profit levels.

4.4 The Empirical Test.

In this section we test the empirical implications of the model that was derived in Section 4.2, using the homogeneous goods industries of the FTC Line of Business Data. First we derive the empirical model. Then we discuss the data and present the estimation results.

4.4a From Theory to Measurement.

For convenience we restate here our basic result, as formulated in Proposition 4.2[86]

$$\underline{\delta}_{lk}^{-1}\left[\frac{q_k^*}{q_l^*}\right]^{\bar{\varepsilon}-1} \geq \frac{r_k^*}{r_l^*} \geq \underline{\delta}_{kl}\left[\frac{q_k^*}{q_l^*}\right]^{\underline{\varepsilon}-1} \tag{4.5}$$

In the data there are obviously a large number of observations for which the profitability in a certain period is negative. In Appendix 4C we derive the empirical implications of the theoretical model if this is the case. If π is strictly negative then the following version of the model holds:

$$\underline{\delta}_{kl}^{-1}\left[\frac{q_k^*}{q_l^*}\right]^{\bar{\varepsilon}-1} \geq \frac{r_l^*}{r_k^*} \geq \underline{\delta}_{lk}\left[\frac{q_k^*}{q_l^*}\right]^{\underline{\varepsilon}-1} \tag{4C.1}$$

For empirical purposes we transform the model. Taking logarithms, (4.5) and (4C.1) become respectively:

$$\bar{\varepsilon}\ \ln(q_k^*/q_l^*) - \ln\underline{\delta}_{lk} \leq \ln(\pi_k^*/\pi_l^*) \leq \underline{\varepsilon}\ \ln(q_k^*/q_l^*) + \ln\underline{\delta}_{kl}$$

and (4.25a)

$$\bar{\varepsilon}\ \ln(q_k^*/q_l^*) - \ln\underline{\delta}_{kl} \leq -\ln(\pi_k^*/\pi_l^*) \leq \underline{\varepsilon}\ \ln(q_k^*/q_l^*) + \ln\underline{\delta}_{lk}$$

(4.25b)

Although there exist empirical methods to estimate bounds like (4.25)[87], we chose to estimate the associated equations rather than the bounds themselves.

As the profits in (4.25) are measured as cross-sectional differences, we ranked for

[86] The empirical results can be formulated in terms of Proposition 4.3 (relation 4.19), which is probably empirically indistinguishable from (4.5).

[87] See Schmidt (1985), Bauer (1990) and Greene (1993) for an overview of these methods, which are primarily used in the context of productive efficiency. They, however, require quite restrictive assumptions on the distribution of observations inside the bounds, for which there is no theoretical foundation in our model. See Ali and Seitford (1993) for an alternative (deterministic, non-parametric) approach, based on mathematical programming and known as "data envelopment analysis".

each industry and for each year all business units with equal sign of profits in order of decreasing average size over the sample period. For period t the empirical equivalent of business units k and l in (4.25) were respectively the business unit with rank n and n+1 in industry j, given that both had an identical sign of profits. By taking the difference between consecutive ranks, we reduced the loss of degrees of freedom to a minimum. We assumed that the efficiency differences are firm specific, and depend on the sign of profits. For firm i, whose business unit in industry j has rank n, the empirical model is:

$$sign\{\pi_{nj,t}\} \ln(\pi_{nj,t} / \pi_{n+1j,t}) = \varepsilon_{j,t} \ln(q_{nj,t} / q_{n+1j,t}) + \xi^{i}_{+/-} + \nu_{nj,t} \qquad (4.26)$$

where

- $\xi^{i}_{+/-}$ are firm specific effect (efficiency differences).
- $\nu_{nj,t}$ is a disturbance term, which is distributed IID(0, σ^2_ν) over all units nj and t.

The firm dummies $\xi^{i}_{+/-}$ depend on the sign of profits, as the ranking of firms with positive profits might be different than for negative profits[88]. The key coefficient to be estimated in this relationship is the cross-elasticity, or the sensitivity of the action of firm k on the pay off of firm l, $\varepsilon_{j,t}$. The theoretical implication is that in a cross-section of industries we expect the relationship between size and profitability to be more increasing (less decreasing), the higher this sensitivity is. Below we will assume that this sensitivity $\varepsilon_{j,t}$ is a function of a number of observable industry characteristics. To find a satisfactory empirical measure for this sensitivity is a crucial step in the analysis to which we will turn now.

The empirical analysis is confined to homogeneous goods industries, as they are the most likely candidates for which the assumption is satisfied that the only strategic variable is the firm's size of operations in a particular industry, rather than a complex interaction between the levels of production, R&D and advertising. In the absence of vertical differentiation it is the degree of horizontal differentiation that

[88] See Appendix 4C.

determines the sensitivity of the pay off of firm k on the action of firm l. If the goods are intrinsically homogeneous, then this is the geographical location, or more precisely, the extent to which the product can be shipped to alternative locations. Hence transport costs can be interpreted as empirical proxy for the cross-elasticity ε_j . As data on transport costs themselves are not available, we use the average distant shipped (DS) as an inverse measure of transport costs and expect a positive effect of DS on the slope of the profitability/size relationship.

However, the extent to which transport distance measures the cross-elasticity is conditioned by exogenous factors determining location. Some production occurs naturally far away from ultimate markets, and transport distances are large, if it were only to reach customers (mining products is the typical example). To control for this effect we condition on the degree to which the geographical distribution of production differs from consumption, measured by a "geographical Gini-index"[89]. The higher the index, the more the distribution of production differs from the distribution of consumption and the less the proxy DS refers to the cross-sensitivity. We include the geographical Gini-index (GGINI) in the empirical measure of ε_j and expect a negative effect on the profitability/size trade off.

Furthermore, we condition on the extent to which a firm has the possibility of spreading production geographically. If the business units in the industry are typically unit-plant, then there is apparently little scope for geographical diversification, and a high transport distance does not necessarily indicate a high cross-elasticity. We include the average number of establishments per firm in the industry (EF) to control for this effect and expect a positive effect on the profitability/size relationship.

Hence our empirical measure of the cross-elasticity is in industry j is

[89] See Section 3.5c for a more detailed description of the index.

$$\varepsilon_{j,t} = \alpha_1 + \alpha_2 DS_j + \alpha_3 GGINI_j + \alpha_4 EF_j + \hat{\nu}_{j,t} \tag{4.27}$$

where the α_i's will be empirically estimated and

- DS_j Average distance goods are shipped in industry j (Weiss 1972).
- $GGINI_j$ Geographical Gini-index for industry j.
- EF_j Average number of establishments per firm in industry j.
- $\hat{\nu}_{j,t}$ Disturbance term, $IID\,(0\,,\hat{\sigma}_\nu^2)$ over all j and t.

The formulation in (4.25) is such that the bounds hold for a whole set of industries. If the 'true' model differs by industry, then the bounds in (4.25) apply for all industries for which the bounds on the cross-elasticity are identical (see the example in Section 4.2c). Consequently, the bounds are not model dependent, but the profitability/size trade offs themselves are. As we estimate the latter, it can be shown that if we ignore the heterogeneity in α_1 across industries, all coefficients are positively biased, and the standard errors are inflated[90]. Hence there is an indeterminate effect on the significance levels of the estimated coefficients. To account for this, we assume that the same model applies to all industries within a (2-digit) sector, but that a different model applies across the sectors. This implies that α_1 is sector dependent, whereas the other $\alpha's$ are fixed (model independent) parameters.

Following standard practice[91], we will use net operating income (OPI) and market share (MS) as our basic measure of pay offs and size respectively. Summarizing, the empirical model for a firm, whose business unit has rank n in industry j, which is in sector m, is

[90] This is essentially an omitted variable problem, as the heterogeneity requires inclusion of multiplicative dummies. If variables are omitted in an OLS regression, then the estimated coefficients are the 'true' ones plus the OLS estimates obtained when each excluded variable in turn is regressed on the set of included variables. As the latter coefficients are positive, the coefficients are biased upwards. A similar reasoning holds for the standard errors. See Johnston (1984).

[91] See Scherer et al. (1987).

$$sign\{OPI_{nj,t}\} \ln(OPI_{nj,t} / OPI_{n+1j,t}) =$$

$$(\alpha_1^m + \alpha_2 DS_j + \alpha_3 GGINI_j + \alpha_4 EF_j) \ln(MS_{nj,t} / MS_{n+1j,t}) + \xi^i_{+/-} + \tilde{\nu}_{nj,t}$$

(4.28)

where $\tilde{\nu}_{nj,t} = \nu_{nj,t} + \hat{\nu}_{j,t} \ln(MS_{nj,t} / MS_{n+1j,t})$. It can be shown that the error terms $\tilde{\nu}_{nj,t}$ are uncorrelated with the explanatory variables[92], although they are heteroscedastic. This will be taken into account in the empirical estimation.

4.3b The Data.

We use the FTC Line of Business Data as it is one of the very few data sets that gives accounting and financial statistics on business unit level, rather than firm level[93]. The FTC Line of Business Program compiled financial statistics of approximately 450 of the largest US manufacturing enterprises, disaggregated to business unit level, i.e. to a firm's operation in one of 261 manufacturing and 14 non-manufacturing categories[94]. Along with common practice in studies using this data set[95], those industries were dropped that appeared to be primarily residual classifications, as they probably did not even approximately correspond to meaningful markets[96]. Subsequently, we took the industries for which the average advertising and R&D-to-sales ratio over the four years for which observations are

---

[92] This is the case if $E[\hat{\nu}_j \ln^2(MS_{nj,t}/MS_{n+1j,t})] = 0$.

[93] The PIMS dataset is an alternative. See Marshall (1986) for a comparison between the PIMS and the FTC data.

[94] The categories, as defined by the FTC are broadly comparable to 3 to 4-digit SIC level. The FTC collected information from the top 250 *Fortune 500* companies, the top two enterprises in each category, and additional companies, to achieve appropriate coverage of the categories. Hence the sampling is non-random and is not perfectly representative for the US manufacturing sector. See Ravenscraft (1983) for a more detailed description of the full data set.

[95] See Schmalensee (1985) and Kessides (1990).

[96] These were 20.29, 22.12, 23.06, 23.07, 24.05 25.06, 28.17, 29.03, 30.06, 32.18, 33.13, 34.21, 35.37, 36.28, 37.14 and 39.08.

available (1974-1977) was below 1.5%. Table 4E.1 gives the 79 (4-digits) categories. Births and deaths were eliminated to prevent spurious results from specific characteristics of newly-founded or dying businesses, though we will test the extent to which balancing affected our results. The empirical analysis is conducted using the resulting balanced panel of 373 companies for which four years of observations were available. These firms have on average 2.08 lines of businesses in the data set, which gives 774 business units altogether.

Net operating income (OPI) refers to revenues and transfers minus traceable and non-traceable (overhead) costs. The contribution margin (CM) will be used as an alternative measure of profits and is revenues minus the direct costs of operating the revenues. The market share (MS) is the proportion of sales a firm made in the previous year in a particular industry, where sales refers to net operating revenues and transfers. As alternative measures of size we used a firm's beginning of period assets (ASSET), which includes traceable and non-traceable assets, net of depreciation. We also used the beginning of period gross plant, property and equipment (GPPE) as a measure of size, which refers to historic asset values, excluding depreciation. The alternative measures of size have been included as the allocation of non-traceable (overhead) costs has been subject to fierce criticism, as there is an element of subjectivity involved[97].

Distance shipped (DS) is the radius (in 1000 miles) within which 80% of industry shipments occurred. It is based on Weiss (1972)[98] and was previously used by Ravenscraft (1983). The geographical Gini-index (GGINI) is based on geographical production data from the 1972 Census of Manufactures and consumption data from the 1973 Statistical Abstract of the United States. The relative number of establishments per firm (EF) for each industry comes from the 1972 Census of

[97] See Benston (1985) for a detailed comment on the use of business unit level data in empirical studies, in particular in the context of the FTC Line of Business Program. Scherer et al. (1987) responded to this criticism.

[98] The data originate in the 1967 Census of Transportation (Bureau of the Census, USGPO, Washington DC).

Manufactures[99].

Table 4.1 gives some descriptive statistics of the variables. The industry specific parameters only occur as interaction terms in our basic specification (4.28). From Table 4.1 follows that $\ln(MS_{nj,t} / MS_{n+1j,t})$ on the one hand and the interaction term $\ln(MS_{nj,t} / MS_{n+1j,t}) \times GGINI$ on the other are almost perfectly correlated. To avoid multicollinearity the latter is not included in the empirical specifications.

4.3c The Econometric Approach.

Model (4.28) is our basic empirical specification, assuming that $\zeta^i_{+/-}$ are fixed parameters. Below we will assume that the efficiency effects are random, in particular that $\xi^i_{+/-} \sim IID(0, \sigma^2_\xi)$ and $\xi^i_{+/-}$ is independent of the explanatory variables and $\tilde{\nu}_{nj,t}$ [100]. Although the random effects model has stronger assumptions on independence, its advantage is that the loss of degrees of freedom is substantially smaller than for the fixed effects model. Whereas in the fixed effects model a conditional likelihood is estimated, the random effects considers the unconditional likelihood. As a consequence, the random effects model uses the within and between information more fully than the fixed effects estimator. We will formally test which model is a more accurate description of the data.

A well known problem with specifications such as (4.28) is that the explanatory variables are endogenous. Given the complexity of the strategic interaction it seems arbitrary to exclude ex-ante any variable from the structural relationship. Hence there are no theoretically exogenous variables to be used as instruments. Several authors have tried to get around this problem by estimating simultaneous equation models, using lagged endogenous variables as instruments. But as Schmalensee (1989) points out, this is only valid if the relation that is estimated is a long run equilibrium, and deviations from it are random. If not, residuals are serially

---

[99] See Appendix 4E for a detailed description of these variables.

[100] These models are also known as respectively the one-way fixed and random effects error component model, see Baltagi (1995).

correlated and lagged variables are not appropriate instruments. It is, however, unlikely that the US economy was in a long run equilibrium shortly after the first oil shock in 1973, which just preceded the observations in our data set.

Our theoretical result is static and no suitable dynamic model of disequilibrium is available. Moreover, with four time observations there is little hope of achieving a satisfactory estimation of a dynamic model. We estimated a static model and used lagged variables as our measures of size. For ASSET and GPPE these were beginning-of-the-year variables, for the market share variable (MS) this was the volume of sales in t-1. In most of the empirical estimations below we use White's (1980) method to correct for the heteroscedasticity of $\tilde{v}_{nj,t}$.

4.3d The Estimation Results.

The first column of Table 4.2 gives the estimates of (4.26), assuming ε_j is fixed for for each (2-digit) sector. For 10 out of 16 sectors the estimates are positive, though only significantly so in 4 cases. At a 5% significant level (though not at a 1% level), we can reject the restriction that all estimates of ε_j are identical[101]. The positive signs can be interpreted as supporting evidence for Proposition 4.1, which predicts a positive relationship between size and profits.

In the second column we report the estimates of (4.28), allowing the estimates of α_1 to be sector specific. The estimate of α_2 is positive and significant at a 5% level. Hence the larger the geographical market space, the higher the sensitivity of the firms' pay off w.r.t. the actions of competitors, the more the profitability/size trade off is rotated anti-clockwise. The effect of the average number of establishments is negative, although the theoretical prediction was positive, but it is insignificant. We used the averages of DS and EF on sector level to determine the sign of the overall effect of business unit size on profits, i.e. ε_j . Unsurprisingly, they are of identical sign and very similar in magnitude to the estimates of ε_j that

[101] The test statistic is 1.73 $\sim F_{(15,1176)}$, with 5% critical value < 1.70 and 1% critical value > 1.99.

were reported in the first column of Table 4.2. We also estimated (4.28) using the unbalanced panel, but the conclusions were not qualitatively different from the ones we drew on the basis of the balanced panel.

A difference between our specification and most studies that investigate performance using the Line of Business Data is that usually only one year of data is used (see Scherer et al. (1987)). Allowing α_1 to be not only sector dependent but also time dependent does not significantly change the results[102], and neither do additive time dummies[103]. Estimating (4.28) as a cross-section relationship, i.e. year by year, gives qualitatively identical results as before for 1975 and 1977. For 1976 the effect of the distance shipped is positive though insignificant.

If we allow the estimates of α_1 to be different for negative profits, though by the same amount across the sectors, then we find a negative effect that is significant at a 1% level. We also analyzed the effects of replacing the firm dummies by business unit dummies or (time dependent) industry dummies. The results only vary slightly, with the business unit dummies reducing the pure size effect (i.e. the estimates of the sector dependent α_1). With (time dependent) industry dummies the effect of the average number of establishments per firm (EF) is positive though insignificant.

In column three we report the results of estimating (4.28) assuming that the firm specific efficiency effects are random. The reported GLS estimates show that qualitatively the results are unchanged. However, on a 5% significant level we can reject the hypothesis that the coefficients of the fixed effects model are not significantly different from the coefficients of the random effects model (though we cannot reject this on a 1% level)[104]. The most likely explanation for the differences is the possible correlation between the firm specific effect ξ^i and the explanatory

[102] The test statistic is 1.43 $\sim F_{(2,1172)}$, with 5% critical value > 2.99.

[103] The test statistic is 1.94 $\sim F_{(4,1172)}$, with 5% critical value > 3.32.

[104] This is a Hausman specification test, with test statistic 33.41 $\sim \chi^2(18)$, with 1% critical value 34.8 and 5% critical value 28.9.

variables, in particular the size of business units.

We also tested the robustness of our specification w.r.t. alternative measures of profits and size. As there has been fierce criticism on the use of business unit data in this context. Showing that the results do not critically depend on the precise measure of business unit performance and size can counter this criticism to some extent. In the fourth column of Table 4.2 we report the estimates of the variables in (4.28), using the contribution margin (CM) as a measure of performance. As non-traceable (overhead) expenses are not substracted, this measure has the advantage over OPI that it is not subject to arbitrary allocation of overhead costs. On the other side it is subject to which activities take place in the corporate centre and which on business unit level. This might well differ among the firms. The empirical results do not change qualitatively however, only EF is positive, though insignificant, as predicted by the theory.

We also estimated (4.28) using respectively Assets (ASSET) and Gross Plant, Property and Equipment (GPPE) as measures of business unit size. The valuation of assets is subject to arbitrary depreciation methods, but the problem with GPPE is that it is based on historic costs, hence price changes are not taken into account. The results are not reported here, but both measures of size perform 'worse' than MS. The effect of distance shipped (DS) becomes insignificant in both cases, although it remains positive.

Table 4.1 Descriptive Statistics Line of Business Data, 1975 - 1977.

(a) Means by Sector.

Sector	$sign(OPI_{nj,t})\ln\left(\frac{OPI_{nj,t}}{OPI_{n+1j,t}}\right)$	$sign(CM_{nj,t})\ln\left(\frac{CM_{nj,t}}{CM_{n+1j,t}}\right)$	$\ln\left(\frac{MS_{nj,t}}{MS_{n+1j,t}}\right)$	$\ln\left(\frac{ASSET_{nj,t}}{ASSET_{n+1j,t}}\right)$	$\ln\left(\frac{GPPE_{nj,t}}{GPPE_{n+1j,t}}\right)$	DS	EF	GGINI
20	.30	.29	.35	.36	.56	.550	1.48	.72
22	.47	.56	.58	.56	.63	.712	1.43	.89
23	.22	.15	.23	.26	.49	.856	1.11	.78
24	.12	.48	.47	.46	.67	1.40	4.98	.70
25	.39	.76	.39	.38	.44	.735	1.07	.65
26	.22	.26	.29	.29	.44	.637	1.61	.69
27	.41	.55	.31	.38	.67	.304	1.10	.58
28	.22	.20	.13	.17	.28	.828	1.52	.73
29	.06	-.26	.37	.37	.49	1.00	1.98	.74
31	-	-	-	-	-	.650	1.10	.81
32	.34	.59	.52	.51	.65	.282	2.27	.72
33	.26	.22	.36	.37	.55	.672	1.42	.71
34	.28	.21	.34	.34	.64	.473	1.38	.70
35	.69	.39	.79	.92	.87	.812	1.10	.80
36	.32	-.15	.42	.36	.56	.817	1.31	.71
37	.38	-.40	.66	.70	.88	.790	1.15	.72

(b) Correlation Matrix.

	$sign\{OPI_{nj,t}\}\ln\left(\frac{OPI_{nj,t}}{OPI_{n+1j,t}}\right)$	$\ln\left(\frac{MS_{nj,t}}{MS_{n+1j,t}}\right)$	$DS_j\times\ln\left(\frac{MS_{nj,t}}{MS_{n+1j,t}}\right)$	$GGINI_j\times\ln\left(\frac{MS_{nj,t}}{MS_{n+1j,t}}\right)$
$\ln(MS_{nj,t}/MS_{n+1j,t})$	.11			
$DS_j\times\ln(MS_{nj,t}/MS_{n+1j,t})$	.08	.82		
$GGINI_j\times\ln(MS_{nj,t}/MS_{n+1j,t})$	.12	.98	.83	
$EF_j\times\ln(MS_{nj,t}/MS_{n+1j,t})$	.00	.57	.59	.64

-) Cannot be included due to disclosure avoidance.

Table 4.2 Estimated Coefficients of Relationship (4.28).

Column	(1)	(2)	(3)	(4)
Profit Var.	OPI	OPI	OPI	CM
Size Var.	MS	MS	MS	MS
Estimation Method	OLS	OLS	GLS	OLS
Constant α_1 (by sector)				
20	.62	.26	.42	-.33
	(.40)	(.39)	(.29)	(.46)
22	.00	-.41	.09	2.50
	(.37)	(.39)	(.29)	(1.86)
23	-1.74*	-2.23*	-1.33*	-2.07*
	(.62)	(.65)	(.61)	(.63)
24	.39	.04	-.15	.044
	(.21)	(.21)	(.27)	(.34)
25	-1.18	-1.63	-.14	-1.28
	(1.25)	(1.25)	(.64)	(1.02)
26	.69*	.29	.37	-.11
	(.18)	(.21)	(.23)	(.37)
27	-.10	-.50	.03	-.54
	(.28)	(.29)	(.25)	(.46)
28	-.65	-1.16*	-.78	-1.66*
	(.50)	(.52)	(.67)	(.60)
29	.26	-.44	-.27	-1.29*
	(.17)	(.27)	(.35)	(.60)
31	-.17	-.56	-.47	-.43
	(1.38)	(1.38)	(2.30)	(1.40)
32	.81*	.60*	.38	.013
	(.22)	(.23)	(.24)	(.50)
33	.50*	.05	.21	-.72
	(.21)	(.25)	(.21)	(.44)
34	.26	.00	.10	-.12
	(.18)	(.20)	(.18)	(.28)
35	.48*	.00	.30	.34
	(.24)	(.29)	(.29)	(.61)
36	-.05	-.53	-.24	-1.48*
	(.39)	(.42)	(.43)	(.59)
37	.36	-.08	-.01	-.68
	(.20)	(.24)	(.27)	(.49)
DS		.64*	.39*	.16*
α_2		(.18)	(.18)	(.06)
EF		-.03	-.03	.002
α_3		(.02)	(.03)	(.002)
Dummies	firm	firm		firm
	(309)	(309)		(278)
$\bar{R}^2$	.31	.32	.02	.49
OBS	1500	1500	1500	1314

* significant at a 5% level
(All OLS estimated SE are heteroscedastic consistent (White's method).)

4.3 Summary and Conclusion.

Assuming that the observed asymmetric industry structure is the equilibrium configuration of some unknown game structure, we derived an empirically meaningful relationship between firm size and profitability. The more sensitive the profitability of a firm is w.r.t. the actions of its competitor, the higher the profitability of a large firm relative to its smaller rival. Hence in industries in which price competition is relaxed through horizontal or vertical product differentiation, large firms are relatively less profitable than their smaller rivals, compared to industries in which price competition is very tough, given the firm asymmetry. The level of profitability of the large firm can be higher or lower than the profitability of the small firm, depending on the particular characteristics of the industry. In industries in which the sensitivity is very high, large firms are always more profitable than small firms, whereas if the sensitivity is low, then smaller firms are always more profitable.

Testing this proposition using the Line of Business Data shows that firm support can be found in homogeneous goods industries. The sensitivity of firm i's action on firm j's pay off is measured in terms of the average distance over which products are shipped. We show that with higher distance shipped, the profitability/size relationship is rotated anti-clockwise.

Appendix 4A Proofs of the Propositions of Chapter 4.

Proposition 4.2 *(Rate of Return).*

If Assumptions 4.2, 4.3 and 4.4 hold and

- there exists an equilibrium such that $q_k^* > q_l^*$,
- firm k is more efficient than l ($\delta_{kl}(q) \geq 1 \geq \delta_{lk}(q), \ \forall q$)

then there is an upper and a lower bound to the ratio of profit rates r_k^* / r_l^* :

$$\underline{\delta}_{lk}^{-1} \left[\frac{q_k^*}{q_l^*} \right]^{\bar{\varepsilon}-1} \geq \frac{r_k^*}{r_l^*} \geq \underline{\delta}_{kl} \left[\frac{q_k^*}{q_l^*} \right]^{\underline{\varepsilon}-1} \tag{4.5}$$

Proof of the Upper Bound.

Consider firm l deviating from q_l^* to q_k^* . Assumption 4.3 $\varepsilon_{kl} \leq \bar{\varepsilon}$ implies that $\frac{\partial \pi_l}{\pi_l} \geq -\bar{\varepsilon}\frac{\partial q_k}{q_k}$. Taking the integral from q_l^* to q_k^* :

$$\ln\pi_k(q_k^*, q_k^*, q_{-kl}^*) - \ln\pi_k(q^*) \geq -\bar{\varepsilon}(\ln q_k^* - \ln q_l^*) \tag{4.6}$$

and hence $\pi_k(q_k^*, q_k^*, q_{-k,l}^*) \geq \gamma^{-\bar{\varepsilon}}\pi_l(q^*)$,where $\gamma = q_k^*/q_l^*$. For this deviation not to be profitable, firm l's equilibrium pay off must satisfy:

$$\pi_l(q^*) \geq \pi_l(q_k^*, q_{-l}^*) = \pi_k(q_k^*, q_k^*, q_{-k,l}^*)\,\delta_{lk}(q_l^*) \geq \gamma^{-\bar{\varepsilon}}\pi_k(q^*)\underline{\delta}_{lk} \tag{4A.1}$$

Hence

$$\pi_k(q^*) \leq \underline{\delta}_{lk}^{-1}\gamma^{\bar{\varepsilon}}\pi_l(q^*) \tag{4A.2}$$

The upper bound follows if LHS and RHS are divided by $\pi_l(q^*)\gamma$. QED.

Proposition 4.3 *(Strategic Asymmetries)*

If Assumptions 4.2, 4.3 and 4.4 hold and

- there exists an equilibrium such that $q_k^* > q_l^*$,
- firm k is more efficient than l ($\delta_{kl}(q) \geq 1 \geq \delta_{lk}(q), \ \forall q$)

then, the profit rates satisfy:

$$\zeta_k \underline{\delta}_{lk}^{-1} \left[\frac{q_k^*}{q_l^*}\right]^{\bar{\varepsilon}-1} \geq \frac{r_k^*}{r_l^*} \geq \zeta_l \underline{\delta}_{kl} \left[\frac{q_k^*}{q_l^*}\right]^{\underline{\varepsilon}-1} \tag{4.19}$$

where

$\zeta_k = \left(s^*(q_k^*)/q_k^*\right)^{\bar{\varepsilon}} > 1$ if k=2 and s^* is upward sloping, $\zeta_k = 1$ otherwise.

$\zeta_l = \left(q_l^*/s^*(q_l^*)\right)^{\underline{\varepsilon}} < 1$ if l=2 and s^* is downward sloping, $\zeta_l = 1$ otherwise.

Proof.

- Lower Bound.

(i) If l moves in stage 1 (l = 1) and firm k deviates to q_l^* in the second stage, the proof of the bound is like the proof of Proposition 4.2. In this case $\zeta_l = 1$.

(ii) If k moves in stage 1 (k = 1), there are two cases:

a. s^* is (weakly) upward sloping. Then:

$$q_l^* = s^*(q_k^*) \geq s^*(q_l^*) \tag{4A.3}$$

In equilibrium it is not profitable for firm k to deviate to q_l^* if:

$$\begin{aligned} \pi_k(q^*) \geq \pi_k(q_l^*, s^*(q_l^*)) \geq \pi_k(q_l^*, q_l^*) = \\ = \pi_l(q_l^*, q_l^*)\, \delta_{kl}\, q_l^* \geq \gamma^{\underline{\varepsilon}}\, \pi_l(q^*)\, \underline{\delta}_{kl} \end{aligned} \tag{4A.4}$$

Hence

$$\pi_k(q^*) \geq \gamma^{\underline{\varepsilon}}\, \pi_l(q^*)\, \underline{\delta}_{kl} \tag{4A.5}$$

The lower bound follows if both sides of the inequality are divided by $\pi_l(q^*)\gamma$. Note that in this case $\zeta_l = 1$.

b. s^* is downward sloping. Then:

$$q_l^* = s^*(q_k^*) < s^*(q_l^*) \tag{4A.6}$$

In equilibrium it is not profitable for firm k to deviate to q_l^* :

$$\pi_k(q^*) \geq \pi_k(q_l^*, s^*(q_l^*)) \geq \left[\frac{q_l^*}{s^*(q_l^*)}\right]^{\underline{\varepsilon}} \pi_k(q_l^*, q_l^*) =$$

$$= \left[\frac{q_l^*}{s^*(q_l^*)}\right]^{\underline{\varepsilon}} \pi_l(q_l^*, q_l^*)\, \delta_{kl}(q_l^*) \geq \zeta_l \gamma^{\underline{\varepsilon}} \pi_l(q^*) \underline{\delta}_{kl}$$

(4A.7)

Hence

$$\pi_k(q^*) \geq \zeta_l \gamma^{\underline{\varepsilon}} \pi_l(q^*) \underline{\delta}_{kl} \tag{4A.8}$$

The lower bound follows if both sides of the inequality are divided by $\pi_l(q^*)\gamma$. Note that $\zeta_l < 1$, as follows from (4A.6).

- Upper Bound.

(iii) If k moves in stage 1 (k = 1) and firm l deviates to q_k^* in the second stage, the proof of the bound is like the proof of Proposition 4.2. In this case $\zeta_k = 1$.

(iv) If l moves in stage 1 (l = 1), there are two cases:

a. s^* is (weakly) downward sloping. Then:

$$q_k^* = s^*(q_l^*) \geq s^*(q_k^*) \tag{4A.9}$$

In equilibrium it is not profitable for firm l to deviate to q_k^* :

$$\begin{aligned} \pi_l(q^*) &\geq \pi_l(q_k^*, s^*(q_k^*)) \geq \pi_l(q_k^*, q_k^*) = \\ &= \pi_k(q_k^*, q_k^*)\, \delta_{lk}(q_k^*) \geq \gamma^{-\bar{\varepsilon}} \pi_k(q^*)\, \underline{\delta}_{lk} \end{aligned} \tag{4A.10}$$

Hence

$$\pi_k(q^*) \leq \gamma^{\bar{\varepsilon}} \pi_l(q^*)\, \underline{\delta}_{lk} \tag{4A.11}$$

The lower bound follows if both sides of the inequality are divided by $\pi_l(q^*)\gamma$. Note that in this case $\zeta_k = 1$.

b. s^* is upward sloping. Then:

$$q_k^* = s^*(q_l^*) < s^*(q_k^*) \tag{4A.12}$$

In equilibrium it is not profitable for firm l to deviate to q_k^* :

.

$$\pi_l(q^*) \geq \pi_l(q_k^*, s^*(q_k^*)) \geq \left[\frac{q_k^*}{s^*(q_k^*)}\right]^{-\bar{\varepsilon}} \pi_l(q_k^*, q_k^*) = \tag{4A.13}$$
$$= \zeta_k^{-1}\, \pi_k(q_k^*, q_k^*)\, \delta_{lk}(q_k^*) \geq \zeta_k^{-1}\, \gamma^{-\bar{\varepsilon}}\, \pi_k(q^*)\, \underline{\delta}_{lk}$$

Hence

$$\pi_k(q^*) \leq \zeta_k\, \gamma^{\bar{\varepsilon}}\, \pi_l(q^*)\, \underline{\delta}_{lk}^{-1} \tag{4A.14}$$

The lower bound follows if both sides of the inequality are divided by $\pi_l(q^*)\gamma$. Note that in this case $\zeta_l > 1$. QED.

Appendix 4B Multiple Strategic Variables.

In this appendix we analyze the case in which there is more than one strategic variable, for example the level of advertising and production levels. First we assume that all the strategic variables are chosen simultaneously by all firms, then we will analyze the case if they are chosen sequentially.

4B.1 A One Shot Game Formulation.

Assume that there are N firms. Each firm chooses a level of two strategic variables - say production $q_i \in Q \subset \mathbb{R}^+$ and advertising $a_i \in A \subset \mathbb{R}^+$ - simultaneously. Following earlier notation, $q \in Q^n$ and $a \in A^n$ are respectively the profiles of levels of production and advertising. The profit function of firm i is $\pi_i(q, a)$, defined over $Q^n \times A^n$ and strictly positive. The symmetry condition in the case of two strategic variables becomes

Assumption 4B.1: *Symmetry*

If $q_i = q_j$ and $a_i = a_j$ then $\pi_i(q, a) = \pi_j(q, a)$.

As before, Assumption 4B.1 does not imply anonymity, i.e. identical functional forms of the profit functions are not required and it includes cases in which firms face a downward sloping average cost curve. π_i is again continuously differentiable in all elements of q_{-i} . The externality of production of firm i on firm j's pay off satisfies Assumption 4.3, where ε_{ij} is the partial cross-elasticity of the production level of firm j on i's pay off, given a . In general this externality will depend on the profile of advertising levels a . The most likely case is one in which the externality decreases with the level of advertising in the industry. The more advertising, the more the firms create 'local' monopoly power, thereby isolating themselves of the effects of competitor's actions. The bounds in Assumption 4.3 hold globally, unconditional on the level of advertising. The boundary values $\underline{\varepsilon}$ and $\overline{\varepsilon}$ can most easily be interpreted as 'limiting' externalities, in the sense that the level of advertising in the industry goes to zero or becomes very high.

Advertising of firm i has a negative externality on firm j's pay off, given q . Along the lines of Assumption 4.2 assume that

Assumption 4B.2: *Monotonicity.*

π_i is weakly decreasing in all elements of a_{-i} , given q .

Proposition 4B.1 *(Multiple Strategic Variables)*

If Assumptions 4B.1, 4B.2 and 4.3 (bounded externality) hold and there exists an equilibrium such that $q_k^* > q_l^*$ and $a_k^* > a_l^*$, then the ratio of profit rates satisfies

$$\frac{r_k^*}{r_l^*} \geq \left[\frac{q_k^*}{q_l^*} \right]^{\underline{\varepsilon} - 1} \tag{4B.1}$$

Proof.

Consider firm k deviating from q_k^* to q_l^* . Since $\varepsilon_{lk} \geq \underline{\varepsilon}$:

$$\pi_l (q_l^*, q_l^*, q_{-k,l}^*, a^*) \geq \gamma^{\underline{\varepsilon}} \pi_l(q^*, a^*) \tag{4B.2}$$

In equilibrium it is not profitable for firm k to deviate to $\{ q_l^*, a_l^* \}$. Therefore:

$$\begin{aligned} \pi_k(q^*, a^*) &\geq \pi_k(q_l^*, a_l^*; q_{-k}^*, a_{-k}^*) = \pi_l (q_l^*, a_l^*; q_l^*, q_{-k,l}^*, a_l^*, a_{-k,l}^*) \geq \\ &\geq \pi_l(q_l^*, a_l^*; q_l^*, q_{-k,l}^*, a_k^*, a_{-k,l}^*) \geq \gamma^{\underline{\varepsilon}} \pi_l(q^*, a^*) \end{aligned} \tag{4B.3}$$

where the second inequality holds by Assumption 4B.2 and $a_k^* > a_l^*$. Hence

$$\pi_k(q^*, a^*) \geq \gamma^{\underline{\varepsilon}} \pi_l(q^*, a^*) \tag{4B.4}$$

The lower bound follows if both sides of the inequality are divided by $\pi_l (q^*) \gamma$.
QED.

Hence the same lower bound apply for the case of multiple strategic variables as does for the single strategic variable case (if $\delta_{kl} = \delta_{lk} = 1 \quad \forall q$), although in the former the bound might not be as tight as in the latter due to the negative externality of advertising. The weaker the negative externality of advertising the tighter the bound will be.

The bound in (4B.1) does not hold if the ranking of the levels of the strategic variables across firms in equilibrium is not perfectly correlated[105]. If the firm with a high equilibrium level of production has the lower level of advertising, then it is not even necessarily the case that this firm has the higher equilibrium pay off than its competitor.

4B.2 A Two-Stage Game Formulation.

Consider the two stage version of this advertising game. In the first stage all firms choose the level of advertising $a_i \in A$ simultaneously, in the second stage they choose their level of output $q_i \in Q$ as a function of the advertising profile a . The (net) pay offs $\pi_i(q, a)$ are strictly positive and are realised at the end of the second stage. Assume that advertising shifts the profit function upwards (c.p.):

Assumption 4B.3

$\pi_i(q, a) \geq \pi_j(q, a)$ if $a_i > a_j$ and $q_i = q_j$.

I.e. if two firms have an equal production level, but different levels of advertising, the firm with the higher levels of advertising has higher profits. This assumption replaces the symmetry assumptions (4B.1).

Proposition 4B.2

If Assumptions 4B.3 and 4.3 (bounded externality) hold and there exists a subgame-perfect equilibrium such that $q_k^* > q_l^*$ and $a_k^* > a_l^*$, then the ratio of profit rates satisfies the bound in relationship (4B.1).

[105] I.e. if it is not the case that $q_k^* \geq q_l^* \wedge a_k^* \geq a_l^* \; \forall \, k, l$.

Proof.

The proof of this proposition follows from the proof of Proposition 4B.1. Relationship (4B.2) continues to hold if k deviates to q_l^* in the second (price-competition) stage. For that deviation not to be profitable in equilibrium it must be the case that

$$\pi_k(q^*, a^*) \geq \pi_k(q_l^*, q_{-k}^*, a^*) \geq \pi_l(q_l^*, q_l^*, q_{-k,l}^*, a^*) \geq \gamma^{\varepsilon}\pi_l(q^*, a^*) \tag{4B.5}$$

as follows from 4B.3 and $a_k^* > a_l^*$. Relationship (4B.4) continues to hold. QED.

What cannot be shown in this set up is that the ratio of profits necessarily satisfies an upper bound.

Appendix 4C Negative Net Operating Income.

In this appendix we derive the theoretical results (in particular relationship 4.5) for the case in which the pay offs are negative.

The model of Section 4.2b applies. However, we assume $\pi_i(q_i, q_{-i})$ is strictly positive and define efficiency differences slightly differently: if $\pi_i(q), \pi_j(q) < 0$ and $\delta_{ij}(q) \geq 1$, then firm j is said to be more efficient than firm i. The following then holds:

Proposition 4C.1 *(Negative Rates of Return).*
If Assumptions 4.2, 4.3 and 4.4 hold, and
- there exists an equilibrium such that $q_k^* > q_l^*$
- firm k is more efficient than firm l ($\delta_{kl}(q) \geq 1 \geq \delta_{lk}(q) \quad \forall\, q$),
then there is an upper and a lower bound to the ratio of profit rates r_k^* / r_l^* :

$$\underline{\delta}_{lk}^{-1} \left(\frac{q_k^*}{q_l^*} \right)^{-\underline{\varepsilon}-1} \geq \frac{r_k^*}{r_l^*} \geq \underline{\delta}_{kl} \left(\frac{q_k^*}{q_l^*} \right)^{-\bar{\varepsilon}-1} \tag{4C.1}$$

Proof.
Upper Bound.
Consider firm k deviating from q_k^* to q_l^* . As $\varepsilon_{lk} \geq \underline{\varepsilon}$, it follows that $-\frac{\partial \pi}{(-\pi_l)} \geq \underline{\varepsilon} \frac{\partial q}{q_k}$. Taking integrals from q_l^* to q_k^* gives:

$$\ln(-\pi_l(q_l^*, q_l^*, q_{-kl})) - \ln(\pi_l(q^*)) \leq -\underline{\varepsilon}(\ln q_k^* - \ln q_l^*)$$

Hence $\pi_l(q_l^*, q_l^*, q_{-k,l}^*) \geq \gamma^{-\underline{\varepsilon}} \pi_l(q^*)$. For the deviation not to be profitable, firm k's equilibrium pay off must satisfy:

$$\pi_k(q^*) \geq \pi_k(q_l^*, q_{-k}^*) = \pi_l(q_l^*, q_l^*, q_{-k,l}^*)\, \delta_{kl}(q_l^*) \geq \gamma^{-\underline{\varepsilon}} \pi_l(q^*) \underline{\delta}_{lk}^{-1} \tag{4C.2}$$

as $\underline{\delta}_{lk}^{-1} \geq \delta_{kl} \geq \underline{\delta}_{kl}$. Therefore

$$\pi_k(q^*) \geq \underline{\delta}_{lk}^{-1}\gamma^{-\underline{\varepsilon}}\pi_l(q^*) \tag{4C.3}$$

The upper bound follows if the LHS and the RHS of 4C.3 are divided by $\pi_l(q^*)\gamma$.

Lower Bound.

Consider firm 1 deviating from q_l^* to q_k^*. As $\varepsilon_{kl} \leq \bar{\varepsilon}$, it follows that $-\frac{\partial \pi}{(-\pi_k)} \leq \bar{\varepsilon}\frac{\partial q}{q_l}$. Taking integrals from q_l^* to q_k^* gives:

$$\ln(-\pi_k^*) - \ln(-\pi_k(q_k^*, q_k^*, q_{-kl})) \leq -\bar{\varepsilon}(\ln q_k^* - \ln q_l^*)$$

and therefore $\pi_k(q_k^*, q_k^*, q_{k,l}^*) \geq \gamma^{\bar{\varepsilon}}\pi_k(q^*)$. For this deviation not to be profitable, firm l's equilibrium pay off must satisfy:

$$\pi_l(q^*) \geq \pi_l(q_k^*, q_{-l}^*) = \pi_k(q_k^*, q_k^*, q_{-k,l}^*)\,\delta_{lk}(q_l^*) \geq \gamma^{\bar{\varepsilon}}\pi_k(q^*)\underline{\delta}_{kl}^{-1} \tag{4C.4}$$

as $\underline{\delta}_{kl}^{-1} \geq \delta_{lk} \geq \underline{\delta}_{lk}$. Hence

$$\pi_k(q^*) \leq \underline{\delta}_{kl}\gamma^{-\bar{\varepsilon}}\pi_l(q^*) \tag{4C.5}$$

The upper bound follows trivially if the LHS and the RHS of (4C.5) are divided by $\pi_l(q^*)\gamma^{-1}$. QED.

Appendix 4D Examples.

Example 4.1.

(i) $\pi_k^* / \pi_l^* \geq 1$ *if either conditions (4.13a) or (4.13b) are satisfied.*

- If

$$\alpha_l > \alpha_k \quad \wedge \quad \alpha_k + \alpha_l > \delta + \gamma \tag{4.13a}$$

hold, then $\delta - \alpha_k > \delta - \alpha_l$ and $\alpha_k + \alpha_l - \delta - \gamma > 0$. Hence from (4.14b) follows immediately that $\pi_k^* / \pi_l^* \geq 1$.

- If

$$\alpha_l < \alpha_k \quad \wedge \quad \alpha_k + \alpha_l < \delta + \gamma \tag{4.13b}$$

define $f(\alpha) = (\delta - \alpha)\alpha^{\eta}$, where $\eta = -1 - \delta / (\alpha_k + \alpha_l - \gamma - \delta)$. $f(\alpha)$ is increasing if $\alpha \leq \eta\delta / (\eta + 1) = \alpha_k + \alpha_l - \gamma$, from which follows that $f(\alpha_k) > f(\alpha_l)$ and hence $\pi_k^* / \pi_l^* = f(\alpha_k) / f(\alpha_l) \geq 1$.

(ii) The bounds on the cross-elasticity:

$$\frac{1 + R}{R}(\alpha_k - \gamma) \geq \varepsilon_{kl}, \varepsilon_{lk} \geq \alpha_l - \gamma \tag{4.16}$$

- Lower bound.

$$\begin{aligned} \varepsilon_{ij} &= \frac{(\alpha_i - \gamma)\, q_i^{\alpha_i} q_j^{\gamma - \alpha_i}}{q_i^{\alpha_i} q_j^{\gamma - \alpha_i} - q_i^{\delta}} \\ &= \alpha_i - \gamma + \frac{\alpha_i - \gamma}{q_i^{\alpha_i} q_j^{\gamma - \alpha_i} - 1} \end{aligned} \tag{4D.1}$$

The fact that $\pi_i > 0$ implies that $q_i^{\alpha_i - \delta} q_j^{\gamma - \alpha_i} > 1$. Hence $\varepsilon_{ij} \geq \alpha_i - \gamma$.

- Upper Bound.

The auxiliary assumption $\dfrac{p_i q_i - c(q_i)}{c(q_i)} \geq R$ implies that

$$q_i^{\alpha_i - \delta} q_j^{\alpha_j - \gamma} \geq 1 + R \tag{4D.2}$$

Relation (4D.1) can be rewritten as

$$q_i^{\alpha_i} q_j^{\gamma - \alpha_i} = 1 + \frac{\alpha_i - \gamma}{\varepsilon_{ij} - \alpha_i + \gamma} \tag{4D.3}$$

Substituting (4D.3) into (4D.2) and rewriting gives the upper bound

$$\frac{1 + R}{R}(\alpha_k - \gamma) \geq \varepsilon_{kl}, \varepsilon_{lk}$$

(iii) The bounds in relationship (4.18) are satisfied (Proposition 4.2):

$$\left[\frac{q_k^*}{q_l^*}\right]^{(\frac{1+R}{R})(\alpha_k - \gamma) - 1} \geq \frac{r_k^*}{r_l^*} \geq \left[\frac{q_k^*}{q_l^*}\right]^{\alpha_l - \gamma - 1} \tag{4.18}$$

- Lower bound. Substituting the equilibrium size ratio $\frac{q_k^*}{q_l^*}$ (i.e. relationship (4.12)) into the lower bound of (4.18), gives after rearranging

$$\frac{\delta - \alpha_k}{\delta - \alpha_l}\left[\frac{\alpha_l}{\alpha_k}\right]^{\frac{\alpha_k}{\alpha_k + \alpha_l - \gamma - \delta}} \geq 1 \tag{4D.4}$$

Define $g(\alpha) = (\delta - \alpha)\alpha^{\theta}$, where $\theta = \alpha_k / (\gamma + \delta - \alpha_k - \alpha_l)$. Relationship (4D.4) can be rewritten as $g(\alpha_k) / g(\alpha_l) \geq 1$. $g(\alpha)$ is increasing if $\alpha \leq \theta\delta / (\theta + 1) \leq \alpha_k$. As $\alpha_k \geq \alpha_l$ by (4.13b), it must also be the case that $g(\alpha_k) > g(\alpha_l)$. Therefore the inequality in (4D.4) is satisfied.

- Upper bound. Substituting the equilibrium size ratio $\frac{q_k^*}{q_l^*}$ (i.e. relationship (4.12)) into the upper bound of (4.18), gives after rearranging

$$1 \geq \frac{\delta - \alpha_k}{\delta - \alpha_l}\left[\frac{\alpha_l}{\alpha_k}\right]^{-\kappa} \tag{4D.5}$$

where $\kappa = \dfrac{\alpha_l - (1/R)(\alpha_k - \gamma)}{\gamma + \delta - \alpha_k - \alpha_l}$.

Define $h(\alpha) = (\delta - \alpha)\alpha^{\kappa}$. As $h(\alpha)$ is decreasing if $\alpha \in (\frac{\kappa\delta}{\kappa + 1}, \delta)$, it must be the case for (4D.5) to hold that $\alpha_l \geq \dfrac{\kappa\delta}{\kappa + 1}$, since then $h(\alpha_l) > h(\alpha_k)$. This is the case only if

$$\frac{\delta - \alpha_l}{\alpha_l} \geq R \tag{4D.6}$$

The auxiliary assumption $\dfrac{p_i q_i - c(q_i)}{c(q_i)} \geq R$ is always satisfied in equilibrium if

$$\frac{\delta - \alpha_k}{\alpha_k} \geq R \tag{4D.7}$$

If (4D.7) is satisfied, (4D.6) is also satisfied as $\alpha_k > \alpha_l$ and so the inequality in (4D.5) holds.

Example 4.3

Demand: $P = S/Q$, where P is the price level, Q is the quantity and S is a measure of market size. Profit maximization gives an equilibrium profit of $\Pi(N) = S/N^{*2}$. The profit of an additional entrant, which chooses its production level optimally $(q^D = \dfrac{S\sqrt{Q^*}}{c} - Q^*)$, given the production level of all other firms, can be rewritten as:

$$\Pi^D = \left(S^{\frac{1}{2}} - c^{\frac{1}{2}}Q^{*\frac{1}{2}}\right)^2 = \frac{S}{N^*}\left(N^{*\frac{1}{2}} - (N^* - 1)^{\frac{1}{2}}\right)^2 \tag{4D.8}$$

where c is the marginal cost, and $Q^* = \dfrac{(N^* - 1)S}{cN^*}$. (See 2, Appendix 2A for details.) Hence

$$\frac{\Pi(N)}{\Pi^D(N)} = R^D_{N^*} = \frac{1}{N^*[2N^* - 1 - 2(N^{*2} - N^*)^{1/2}]} = \frac{C^{*2}}{-2(1 - C^*)^{0.5} + 2 - C^*} \tag{4D.9}$$

Set $x = \sqrt{1 - C^*}$, then $R^D_{N^*} = (1 + x)^2 = (1 + \sqrt{1 - C^*})$, which is decreasing in C^* .

Example 4.4

If demand is linear $P = a - bQ$, profit maximization yields an equilibrium profit of

$$\Pi(N^*) = \frac{(a - c)^2}{b(N^* + 1)^2} \tag{4D.10}$$

The profit of the deviant is $\Pi^D = (a - c - b(Q^* + q^D))q^D$.
From the FOC it follows that $q^D = \frac{a - c - bQ^*}{2b} = \frac{a - c}{2b(N^* + 1)}$, hence

$$\Pi^D(N^*) = \frac{(a - c)^2}{4b(N^* + 1)^2} \tag{4D.11}$$

From (4D.10) and (4D.11) it follows that $R^D_{N^*} = 4$. Hence the upper bound to profitability is *independent* of concentration.

The Entry Deterring Equilibria.

The first step is to derive the maximal average profit that can be achieved by incumbents for any given market structure. The second step is to establish how this maximal profit changes with the number of incumbents (= concentration). The analysis is confined to symmetric equilibria, for reasons that become clear later.

From equation (4D.11), it follows that the Cournot quantity deters entry if $a \leq c + 2(N + 1)\sqrt{bF}$. For larger a's, the N incumbent firms can play a deterring strategy by having a production level that is higher than the (one shot) Cournot outcome. The market price will be lower and the pay off of the entrant is non-positive. Hence he stays out. It can be shown that the profit of the incumbents is decreasing in the (entry deterring) industry production level. Consequently, the equilibrium industry production level under deterrence is the minimum level for which entrants do not have incentives to enter, i.e. the industry production level $\hat{Q}^*$ for which the profit of an hypothetical entrant is zero:

$$\Pi^D(\hat{Q}^*) = \frac{1}{4b}(a - c - b\hat{Q}^*)^2 = F \tag{4D.12}$$

Therefore $\hat{Q}^* = (a - c - 2\sqrt{bF}) / b$ and the corresponding profit of incumbents is $\hat{\Pi}^* = [2\sqrt{bF}(a - c - 2\sqrt{bF})] / bN$. The profit function conditional on playing the entry deterring strategy, is increasing in the measure of market size a. So is the entry preventing output $\hat{Q}^*$. Let the $a(N)$ for which the maximal average profit and the largest entry preventing output are achieved be $a^{\max}(N)$. This corresponds to that symmetric equilibrium for which the marginal player is indifferent between playing the equilibrium deterring strategy on the one hand, or deviating, allowing entry and playing the optimal Cournot response on the other hand[106].

From profit maximization it follows that the optimal capacity of a deviating incumbent is $\hat{q}^D = [(a - c - b(N - 1)\hat{Q}^*/N)^2] / 2b$. The profit of this deviator is $\hat{\Pi}^D(N) = [(a - c - b(N - 1)\hat{Q}^*/N)^2] / 8b$. $a^{\max}(N)$ is the solution of $\hat{\Pi}^*(N) = \hat{\Pi}^D(N)$. Solving the quadratic form gives $a^{\max}(N) = c + 2[(3 + 2\sqrt{2})N + 1]\sqrt{bF}$. Consequently,

$$R_N^{\max} = \frac{2[(3 + 2\sqrt{2})N - 1]}{N} = 2[3 + 2\sqrt{2} - C]$$

[106] Gilbert and Vives (1986, p. 76) observe that the largest entry preventing output $\hat{Q}^*$ is only an equilibrium if incumbents have equal shares. Therefore only symmetric equilibria are analyzed.

Appendix 4E The Line of Business Data.

This appendix contains a description of the subsample of homogeneous goods industries, and the variables that are used in the empirical analysis of this 4. For a more detailed description of the Line of Business data see The Statistical Report of the Annual Line of Business Report (FTC Bureau of Economics).

LB Variables.

OPI	Operating Income: CM - Other Non-traceable Expenses (Media Advertising, Administration, etc.).
CM	Total Net Operating Revenues - Costs of Operating Revenues.
SALES	Total Net Operating Revenues and Transfers: Revenues from Outsiders + Transfers from other LB's, Foreign Section and Domestic Regulated Section.
GPPE	Gross Plant, Property and Equipment.
ASSETS	Total traceable and non-traceable assets. That is Gross Plant, Property and Equipment, minus accumulated depreciation (depletion and amortization) plus Inventories and All Other Assets. The allocation of non-traceable assets is at the discretion of the reporting firm.

Industry Variables.

DS	Distance Shipped: Radius within which 80% of industry sales occur, in thousands of miles. From Weiss (1972), based on the 1967 Census of Transportation.
EF	Average Number of Establishments per Firm. From: 1972 Census of Manufactures.
GGINI	Geographical Gini-index: geographical distribution of production in the industry, relative to the geographical distribution of aggregate consumption, based on resp. the 1972 Census of Manufactures and the 1973 statistical Abstract of the US. See Krugman (1991) and 3 for a more detailed explanation.

Table 4E.1 Summary Statistics of the Subsample of Line of Business Industries (1974-1977).

Code	DS	EF	GGINI	# BUs	Description
2001	1.006	1.08	.63	20	Meat Packing, and Prepared Meat Products
2002	-.709	1.26	.73	12	Poultry Dressing, Poultry and Egg Processing
2003	SMSA	1.24	.65	11	Fluid Milk
2011	.229	1.34	.74	22	Prepared Feeds for Animals/Fowls
2013	.736	1.58	.81	8	Wet Corn Milling
2016	.866	1.33	.93	9	Cane Sugar
2017	.866	3.81	.90	6	Beet Sugar
2021	.523	1.45	.69	24	Fats and Oils
2023	.420	1.33	.87	2	Malt
2201	.729	1.62	.90	24	Weaving Mills - Cotton, Synthetics and Silk
2203	-.689	1.16	.86	6	Narrow Fabric Mills
2205	.816	1.04	.86	10	Knit Outerwear Mills
2207	.698	1.13	.87	24	Knitting Mills, nec
2210	.552	1.46	.90	20	Yarn and Thread Mills
2211	.840	2.00	.99	2	Tire Cord and Fabric
2301	.946	1.19	.77	13	Men's and Boy's Suits and Coats
2303	.816	1.07	.79	26	Women's and Misses' Outerwear
2305	-.816	1.12	.80	6	Children's Outerwear
2401	SMSA	1.01	.86	9	Logging Camps and Logging Contractors
2402	1.379	11.8	.77	29	Sawmills and Planing Mills
2403	1.858	1.07	.63	31	Millwork, Plywood and Structural Members
2404	-.327	1.66	.71	15	Wood Buildings and Mobile Homes
2503	.698	1.11	.68	8	Office Furniture
2504	.904	1.04	.68	9	Public Building and Related Furniture
2505	.779	1.04	.61	18	Partitions and Fixtures
2601	1.002	1.33	.81	4	Pulp Mills
2602	.797	1.76	.76	27	Paper Mills, except Building Paper
2603	.754	2.04	.68	20	Paperboard Mills
2605	.319	1.39	.67	9	Envelopes
2606	.801	1.25	.65	20	Bags, excl. Textile Bags
2608	.586	1.06	.72	10	Stationary, Tablets and Related Products
2610	.379	1.57	.63	29	Paperboard Containers and Boxes
2701	SMSA	1.09	.64	5	Newspapers
2705	SMSA	1.02	.56	35	Commercial Printing
2706	.900	1.25	.59	12	Manifold Business Forms
2709	SMSA	1.01	.63	4	Printing Trade Services
2814	.828	1.52	.73	33	Fertilizers
2901	1.325	2.10	.79	31	Petroleum Refining
2902	.296	1.72	.64	18	Paving and Roofing Materials
3101	.650	1.10	.81	6	Leather Taning and Finishing
3201	.617	.898	.92	4	Flat Glass
3202	.336	4.33	.76	7	Glass Containers
3204	.492	1.11	.69	11	Products of Purchased Glass
3205	.191	2.64	.70	11	Cement, Hydraulic

FTC	DS	EF	GGINI	# BUs	Description
3210	.144	1.11	.63	17	Concrete Products, including Brock and Brick
3211	SMSA	1.24	.64	13	Ready-Mixed Concrete
3212	.264	1.49	.75	6	Lime
3213	.312	2.59	.77	5	Gypsum Products
3214	.319	1.04	.75	4	Cut Stone and Stone Products
3301	.401	1.51	.68	56	Blast Furnace and Basic Steel Products
3302	.470	1.15	.68	32	Iron and Steel Foundries
3303	1.555	2.82	.92	8	Primary Copper
3304	1.073	1.33	.82	5	Primary Lead
3305	.979	1.50	.86	5	Primary Zinc
3306	.997	2.58	.86	8	Primary Aluminium
3307	1.328	1.18	.77	14	Primary Nonferrous Metals, nec
3308	.835	1.17	.69	13	Secondary Nonferrous Metals
3309	.862	1.43	.68	17	Aluminium Sheet, Plate and Foil etc.
3310	.697	1.33	.73	29	Nonferrous Rolling and Drawing, nec
3311	.927	1.66	.74	26	Nonferrous Wire Drawing and Insulating
3312	.415	1.04	.67	16	Nonferrous Foundries
3401	.362	2.98	.66	13	Metal Cans
3402	.237	1.59	.76	7	Metal Barrels, Drums and Pails
3409	.464	1.07	.58	13	Fabricated Structural Metal
3410	.803	1.12	.59	13	Metal Doors, Sash, Frames, Molding and Trim
3412	.504	1.06	.59	36	Misc. Metal Work
3413	.467	1.06	.75	25	Screw Machine Products, Bolts, etc.
3414	.339	1.11	.75	19	Metal Forgings, ferrous and Nonferrous
3415	.650	1.17	.90	19	Automotive Stampings
3416	.650	1.26	.81	6	Crowns and Closures
3418	SMSA	1.04	.66	9	Metal Coating, Engraving and Allied Services
3513	.842	1.11	.78	17	Machine Tool Accessories and Measuring Devices
3517	-.718	1.08	.88	6	Textile Machinery
3519	.771	1.07	.80	7	Paper Industry Machinery
3603	.817	1.31	.71	27	Motors and Generators
3703	.977	1.13	.65	4	Buses
3704	.900	1.37	.73	6	Combat Vehicles, Tanks
3706	.708	1.19	.74	10	Truck Trailers
3709	.755	1.04	.72	26	Ship and Boat Building and Repairing

*) Cannot be included due to disclosure avoidance.

-) Not Available.

SMSA Standard Metropolitan Statistical Area

References

Adelman, I. G. "A Stochastic Analysis of the Size Distribution of Firms" *Journal of the American Statistical Association*, vol. 53, pp. 893 - 904.

Aghion, P. and P. Bolton (1987) "Contracts as a Barrier to Entry" *American Economic Review,* vol. 70, pp. 388-401.

Ali, A.M. and L.M. Seitford (1993) "The Mathematical Programming Approach to Efficiency Analysis", 3 in H.O. Fried, C.A.K Lovell and S.S. Schmidt (eds.) "The measurement of Productive Efficiency", Oxford University Press, Oxford, England.

Amato, L. and R.P. Wilder (1985) "The Effects of Firm Size on Profit Rates in US manufacturing" *Southern Economic Journal,* vol. 52, pp. 181 - 90.

Annual Line of Business Report 1973-1976, Statistical Report of the Bureau of Economics to the Federal Trade Commission, Washington DC.

Arrow, K. (1962) "Economic Welfare and the Allocation of Resources for Inventions". In R. Nelson "The Rate and Direction of Inventive Activity", Princeton University Press, Princeton, NJ.

D'Aspremont, C., A. Jacquemin, J.J. Gabszewicz and J.A. Weymark (1983) "On Stability of Collusive Price Leadership" *Canadian Journal of Economics,* vol. 16, pp. 17 - 25.

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* vol. 34, pp. 585 - 612.

Bain, J.S. (1951) "Relation of Profit Rate to Industry Concentration: American Manufacturing, 1936 - 1940" *Quarterly Journal of Economics,* vol. 40, pp 293 - 324.

Baltagi, B.H. (1995) "Econometric Analysis of Panel Data" John Wiley & Sohns, Chichester, England.

Barro, R.J. and X. Sala-i-Martin (1992) "Convergence" *Journal of Political Economy*, vol. 100, pp. 223-251.

Bauer, P. (1990) "A Survey of Recent Econometric Developments in Frontier Estimation" *Journal of Econometrics,* vol. 46, pp. 21-39.

Baumol, W.J. (1959) "Business Behaviour, Value and Growth", MacMillan, New York.

Beggs, A. and P. Klemperer (1992) "Multi-Period Competition with Switching Costs", *Econometrica,* vol. 60, pp. 651-666.

Benston, G.J. (1985) The Validity of Profits-Structure Studies with Particular Reference to the FTC's Line of Business Data" *American Economic Review*, vol. 75, pp. 37 - 67.

Bonanno, G. (1987) "Location Choice, Product Proliferation and Entry Deterrence", *Review of Economic Studies,* vol. 54, pp. 37 - 45.

Budd, C., C. Harris and J. Vickers (1992) "A Model of the Evolution of Duopoly: Does the Asymmetry Between Firms Tend to Increase or Decrease?", *Review of Economic Studies,* vol. 60, pp. 543-573.

Bulow, J., J. Geanakoplos and P. Klemperer (1985) "Multimarket Oligopoly: Strategic Substitutes and Complements" *Journal of Political Economy,* vol. 93, pp. 488-511.

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

Caves, R.E. (1989) "International Differences in Industrial Organisation", Ch. 23 in R. Schmalensee and R.D. Willig (eds) Handbook of Industrial Organisation, North Holland, Amsterdam.

- , B. T. Gale and M. E. Porter (1977) "Interfirm Profitability Differences: Comment" *Quarterly Journal of Economics,* vol. 92, pp. 667-675.

Census of Manufactures, Bureau of the Census, US Department of Commerce, Washinton DC.

Census of Transportation, Bureau of the Census, US Department of Commerce, Washington DC.

Chapman K. (1991) "The International Petrochemical Industry", Blackwell, Oxford.

Chesher, A. (1979) "Testing the Law of Proportionate Effect" The Journal of Industrial Economics, vol. XXVII, pp. 403 - 411.

Cohen, W. M. and S. Klepper (1992) "The Anatomy of Industry R&D Intensity Distributions" *The American Economic Review* , vol. 82, pp. 773-799.

- and R. C. Levin (1989) "Empirical Studies of Innovation and Market Structure" , 18 in: R. Schmalensee and R. D. Willig "Handbook of Industrial Organization", vol. 2, North Holland, Amsterdam.

Comanor, W. S. and T. A. Wilson (1974) "Advertising and Market Power", Harvard University Press, Cambridge, Mass.

- and - (1979) "The Effect of Advertising on Competition: A Survey" *Journal of Economic Literature,* vol. XVII, pp. 453-476.

Dasgupta, P. and J. Stiglitz (1980) "Industrial Structure and the Nature of Innovative Activity", *Economic Journal,* vol. 90, pp. 266 - 93.

Davis S. J. and J. Haltiwanger (1992) "Gross Job Creation, Gross Job Destruction and Employment Reallocation" *Quarterly Journal of Economics*, vol. CVII, pp. 819-863.

Demsetz, H. (1973) "Industry Structure, Market Rivalry, and Public Policy" *Journal of Law and Economics,* vol. 16, pp. 1 - 10.

Deneckere, R. and C. Davidson "Incentives to Form Coalitions with Bertrand Competition" *The RAND Journal of Economics*, vol. 16, pp. 383 - 402.

Desdoigts, A. (1994) "Changes in the World Income Distribution : A Non-Parametric Approach to Challenge the Neoclassical Convergence Argument" PhD thesis, European University Institute, Florence.

Dixit, A. (1980) "The Role of Investment in Entry Deterrence" *Economic Journal,* vol. 90, pp. 95 - 106.

Dunne, T., M. Roberts and L. Samuelson (1988), "Patterns of Firm Entry and Exit in U.S. Manufacturing Plants" *Rand Journal of Economics,* vol. 19, pp. 495-515.

-, - and - (1989) "The Growth and Failure of U.S. Manufacturing Plants" *Quarterly Journal of Economics*, vol. CIV, pp. 671-698.

Eaton, B.C. and R.G. Lipsey (1980) "Exit Barriers as Entry Barriers: The Durability of Capital as a Barrier to Entry" *Bell Journal of Economics*, vol. 11, pp. 721-729.

- and - (1981) "Capital, Commitment and Entry Equilibrium" *Bell Journal of Economics* vol. 12, pp. 539-604.

Eatwell, J., M. Milgate and P. Newman (1987)"The New Palgrave: a Dictionary of Economics", The MacMillan Press Ltd, London.

Ericson, R. and A. Pakes (1989) "An Alternative Theory of Firm & Industry Dynamics", unpublished manuscript, Columbia University.

Evans, D. (1987a) "Tests of Alternative Theories of Firm Growth" *Journal of Political Economy,* vol. 95, pp. 657-674.

- (1987b) "The Relationship Between Firm Growth, Size and Age: Estimates for 100 Manufacturing Industries" *Journal of Industrial Economics,* vol. 35, pp. 567-581.

Evely, R. and I. Little (1960) "Concentration in British Industry" Cambridge University Press, Cambridge.

Farrell, J. and C. Shapiro (1988) "Dynamic Competition with Switching Costs" *Rand Journal of Economics,* vol. 19, pp. 123-137.

Foster, R. (1985) "Innovation: The Attacker's Advantage", Summit Books.

Fudenberg , D and J. Tirole (1991) "Game Theory" MIT Press, Cambridge, Mass.

Gale, B.T. (1972) "Market Share and Rate of Return" *The Review of Economics and Statistics,* vol. 54, pp. 412 - 23.

Gale, B.T. and B.S. Branch (1982) "Concentration Versus Market Share: Which Determines Performance and Why Does it Matter?" *The Antitrust Bulletin,* vol. 27, pp. 83 - 105.

George, K.D. and T.S. Ward (1975) "The Structure of Industry in the EEC" Cambridge University Press, Cambridge, England.

Ghemawat, P. (1984) "Capacity Expansion in the Titanium Dioxide Industry" *The Journal of Industrial Economics*, vol. 33, pp. 145-163.

- and Nalebuff, B. (1990) "The Devolution of Declining Industries" *Quarterly Journal of Economics,* vol. 105, pp. 167-186.

Gibrat, R. (1931) "Les Inégalités Économiques Applications", Paris: Libraire du Recueil Sirey.

Gilbert, R. J. and R.G. Harris (1984) "Competition with Lumpy Investment" *The Rand Journal of Economics,* vol. 15, pp. 197 - 212.

- and M. Lieberman (1987) "Investment and Coordination in Oligopolistic Industries" *Rand Journal of Economics*, vol. 18, pp. 17-33.

- and D.M.G. Newbery (1982) "Preemptive Patenting and the Persistence of Monopoly" *American Economic Review*, vol. 72, pp. 514 - 26

- and X. Vives (1986) "Entry Deterrence and the Free Rider Problem", *Review of Economic Studies*, vol. 53, pp. 71 - 83.

Greene, W.H. (1993) "The Econometric Approach to Efficiency Differences" 2 in H.O. Fried, C.A.K Lovell and S.S. Schmidt "The measurement of Productive Efficiency", Oxford University Press, Oxford, England.

Hall, B. (1987) "The Relationship between Firm Size and Firm Growth in the US Manufacturing Sector" *The Journal of Industrial Organisation*, vol. 35, pp. 583-606.

Harris, C. and J. Vickers (1985) " Perfect Equilibrium in a Model of Race" *Review of Economic Studies,* vol. 52, pp. 192-209.

- and - (1987) "Racing with Uncertainty" *Review of Economic Studies,* vol. 54, pp. 1-21.

Hart, P.E. and R. Clarke (1980) "Concentration in British Industry 1935 - 1975", Cambridge University Press, Cambridge, England.

- and S. Prais (1956) "The Analysis of Business Concentration: A Statistical Approach" *Journal of the Royal Statistical Society*, Serie A, vol. 119 pp. 150 - 81.

Huigen, R.D., A.J.M. Kleijweg and G. van Leeuwen (1993) "The Relationship Between Firm Size and Firm Growth in Dutch Manufacturing Estimated on Panel Data" Research Paper 9105, Research Institute for Small and Medium-sized Business in the Netherlands, Zoetermeer October.

Hopenhayen H.A. (1989) "A Dynamic Stochastic Model of Entry and Exit to an Industry", mimeo, Graduate School of Business - Stanford University, Research Paper 1057.

Ijiri, Y. and H.A. Simon (1975) Skew Distributions and the Sizes of Business Firms, North-Holland, Amsterdam.

Johnston, J. (1984) "Econometric Methods" McGraw-Hill, Singapore.

Jovanovic, B. (1982) "Selection and Evolution of Industry" *Econometrica,* vol. 50, pp. 649 - 70.

- and G. M. MacDonald (1993) "The Life-Cycle of a Competitive Industry", NBER Working Paper No. 4441, Cambridge MA.

Katz, M.L. and C. Shapiro (1987) "R&D Rivalry with Licensing or Imitation" *American Economic Review,* vol. 77, pp. 402 - 20.

Kessides, I. N. (1990) "Internal versus External Market Conditions and Firm Profitability: an Exploratory Model" *The Economic Journal,* vol. 100, pp. 773 - 792.

Klepper, S. and E. Graddy (1990) "The Evolution of New Industries and the Determinants of Market Structure" *Rand Journal of Economics,* vol. 21, pp. 27 - 44.

Klepper, S. and K. Simons (1993) "Technological Change and Industry Shakeouts", Manuscript, Carnegie Mellon University, Pittsburgh.

Klomp, L. and R. Thurik (1995) "The Growth of Firms in Dutch Services" CASBEC, Erasmus University Rotterdam, unpublished.

Koopmans, R. and A. Lamo (1995) "Cross-Section Firm Dynamics: Theory and Empirical Results from the Chemical Sector" CEP Working Paper No 229, London School of Economics

Krugman, P. (1991) "Geography and Trade", MIT Press, Cambridge Mass.

Kumar, M.S. (1985) "Growth, Acquisition Activity and Firm Size: Evidence from the United Kingdom", *The Journal of Industrial Economics*, vol. XXXIII, pp. 327-338.

Lambson, V. E. (1987) "Is the Concentration-Profit Correlation Partly an Artifact of Lumpy Technology? *American Economic Review,* vol. 77, pp. 731 - 733.

- (1991) "Industry Evolution with Sunk Costs and Uncertain Market Conditions" *International Journal of Industrial Organization,* vol. 9, pp. 171-196.

Lamo, A. R. (1994) (tba) PhD thesis, LSE, London, in progress.

Lee, T. and L.L. Wilde (1980) "Market Structure and Innovation: A Reformulation" *The Quarterly Journal of Economics,* vol. 94, pp. 429 - 36.

Leonard, S.J. (1986) "On the Size Distribution of Employment and Establishments" NBER Working Paper No 1951, NBER, Cambridge MA.

Lieberman, M. (1982) "The Learning Curve, Pricing and Market Structure in the Chemical Processing Industries", PhD Dissertation, Harvard University, Cambridge, Mass.

- (1984) "The Learning Curve and Pricing in the Chemical Processing Industries" *The Rand Journal of Economics*, vol. 15, pp. 213 - 28.

- (1987) "Post-entry Investment and Market Structure in the Chemical Processing Industries" *Rand Journal of Economics* vol. 18, pp. 533- 549.

Loury, G.C. (1979) "Market Structure and Innovation" *The Quarterly Journal of Economics,* vol. 93, pp. 395 - 410.

Lucas Jr, R.J. (1967) "Adjustment Costs and the Theory of Supply" *Journal of Political Economy*, vol. 75, pp. 321-334.

Mancke, R. B. (1974) "Interfirm Profitability Differences", *Quarterly Journal of Economics,* vol. LXXXVIII, pp. 181-193.

Mansfield, E. (1962) "Entry, Gibrat's Law, Innovation and the Growth of Firms" *American Economic Review*, vol. 52, pp. 1023-1051.

- (1985) "How Rapidly Does New Industrial Technology Leak Out?" *Journal of Industrial Economics,* vol. 34, pp. 217-223.

Marshall, C.T. (1986) PIMS and the FTC Line of Business Data: A Comparison" unpublished doctoral dissertation, Harvard University Graduate School of Business Administration, Cambridge, Mass.

Mazòn, C. "Is Profitability Related to Market Share? An Intra-Industry Study in Spanish Manufacturing" Working Paper no 9327, Banco de España, Madrid.

Mueller, D.C. (1986) "Profits in the Long Run", Cambridge University Press, Cambridge, UK.

Nelson, R. L. (1963) "Concentration in the Manufacturing Industries of the United States" Yale University Press, New Haven.

Odagiri, H. and H. Yamawaki (1990) "The Persistence of Profits in Japan" in: D.C. Mueller "The Dynamics of Company Profits", Cambridge University Press, Cambridge, UK.

Pakes, A. and R.E. Ericson (1988) "Empirical Implications of Alternative Models of Firm Dynamics", SSRI WP # 8803, University of Wisconsin - Madison.

Pashigan, P. (1969) "The Effect of Market Size on Concentration" *International Economic Review*, vol. 10, pp. 291-314.

Phlips, L. (1971) "Effects of Industrial Concentration", North Holland, Amsterdam.

Prais, S.J. (1975) "The Evolution of Giant Firms in Britain", Cambridge University Press, Cambridge, England.

Prescott, E.C. and M. Visscher (1980) "Organization Capital" *Journal of Political Economy*, vol. 88, pp. 446-461.

Pryor, F.L. (1972) "The Size of Production Establishments in Manufacturing" *Economics Journal*, vol. 82, pp. 547 - 566

Quah, D. (1993a) "Empirical Cross-Sectional Dynamics in Economic Growth" *European Economic Review*, vol. 37, pp. 426-443.

- (1993b) "Galton's Fallacy and Tests of the Convergence Hypothesis" *The Scandinavian Journal of Economics*, vol. 95, pp. 427-443.

- (1994) "Convergence Empirics Across Economies with (Some) Capital Mobility", mimeo, London School of Economics.

Quintella R.H. (1993) "The Strategic Management of Technology in the Chemical and Petrochemical Industries" Pinter Publishers, London.

Ravenscraft, D.J. (1983) "Structure-Profit Relationships at the Line of Business and Industry Level" *The Review of Economics and Statistics,* vol. 65, pp. 22 - 31.

Reinganum, J.F. (1982) "Uncertain Innovation and the Persistence of Monopoly" *American Economic Review,* vol. 73, pp. 741 - 48.

- (1989) "The Timing of Innovation: Research Development and Diffusion" in: R. Schmalensee and R.D. Willig, *Handbook of Industrial Organization*, vol. I, Amsterdam, pp. 849 - 908.

Rosen, R.J. (1991) "Research and Development with Asymmetric Firm Sizes", *The Rand Journal of Economics*, vol 22, pp. 411 - 429.

Scherer, F.M., A. Beckenstein, E. Kaufer, R.D. Murphy, (1975) "The Economics of Multi-Plant Operation", Harvard University Press, Cabridge, Mass.

Scherer, F.M., W.F. Long, S.Martin, D.C. Mueller, G. Pascoe, D.J. Ravenscraft, J.T. Scott and L.W. Weiss (1987) "The Validity of Studies with Line of Business Data: Comment" *American Economic Review,* vol. 77, pp. 205 - 223.

Scherer, F. M., and D. Ross (1987) "Industrial Market Structure and Economic Performance" 3rd edition, Houghton Mifflin Company, Boston.

Schmalensee, R. (1978) "Entry Deterrence in the Ready-to-Eat Breakfast Cereal Industry" *Bell Journal of Economics,* vol. 9, pp. 237 - 60.

- (1983) "Advertising and Entry Deterrence: an Explanatory Model" *Journal of Political Economy,* vol. 91, pp. 636 - 53.

- (1985) "Do Markets Differ Much?" *American Economic Review,* vol. 75, pp. 341 - 51.

- (1989) "Inter-Industry Studies of Structure and Performance" in R. Schmalensee and R.D. Willig (eds) "Handbook of Industrial Organization, vol. II, 16, North Holland, Amsterdam.

Schmidt, P. (1985) "Frontier Production Functions", *Econometric Reviews,* vol. 4, pp. 289-328.

Scott, J.T. and G. Pascoe (1986) "Beyond Firm and Industry Effects on Profitability in Imperfect Markets" *Review of Economics and Statistics*, vol. 68, pp. 284 - 292.

Shaked, A. and J. Sutton (1982a) "Natural Oligopolies and the Gains from Trade", Unpublished manuscript, London School of Economics.

Shaked, A and J. Sutton (1982b) "Relaxing Price Competition through Product Differentiation" *Review of Economic Studies*, vol. 49, pp. 3 - 14.

- and - (1983) "Natural Oligopolies" *Econometrica*, vol. 51, pp. 1469 - 83.

- and - (1987) "Product Differentiation and Industrial Structure" *Journal of Industrial Economics*, vol. 36, pp. 131 - 46.

- and - (1990) "Multi-product Firms and Market Structure" *Rand Journal of*

Economics, vol. 21, pp. 45 - 62.

Shepherd, W. G. (1964) "Trends of Concentration in American Manufacturing Industries 1947-1958" *The Review of Economics and Statistics,* vol. XLVI, pp. 200-212.

- (1972) "The Elements of Market Structure" *The Review of Economics and Statistics,* vol. LIV, pp. 25 - 37.

Silberman, I. H. (1967), "On Lognormality as a Summary Measure of Concentration" *American Economic Review,* vol. 57, pp. 801-831.

Silverman, B.W. (1986) "Density Estimation for Statistics and Data Analysis" Chapman and Hall, New York, NY 10001.

Simon, H. and C. Bonini (1958) "The Size Distribution of Business Firms" *American Economic Review,* vol. 48, pp. 607 - 17.

Singh, A. and G. Whittington (1962) "Firm Size and Rate of Growth" *Journal of Political Economy,* vol. 70, pp. 556-569.

Smirlock, M., T. Galligan and W. Marshall (1984) "Tobin's q and the Structure - Performance Relationship" *American Economic Review,* vol. 74, pp. 1051 - 60.

Spitz P. H. (1988) "Petrochemicals, The Rise of an Industry" John Wiley & Sons, New York.

Statistical Abstract of the United States, Bureau of the Census, US Department of Commerce, Washington DC.

Stobaugh R. "Innovation and Competition" Harvard Business School Press, Boston MA.

Stokey, N. L. and R.E. Lucas, Jr. (1989) "Recursive Methods in Economic Dynamics", Harvard University Press, Cambr. Mass.

Sutton, J. (1986) "Vertical Product Differentiation: Some Basic Themes" *American Economic Review (Papers and Proceedings)*, vol. 76, pp. 393 - 98

- (1991) "Sunk Costs and Market Structure", MIT Press, Cambridge, Mass.

- (1995a) "The Size Distribution of Businesses, Part I: A 'Bounds' Approach" STICERD Discussion Paper No. EI/9, Economics of Industry Group, STICERD, London School of Economics.

- (1995b) "Gibrat's Legacy", unpublished manuscript.

Synthetic Organic Chemicals, US Production and Sales, International Trade

Commission, Washington DC. (Annual Issues).

Tirole, J. (1988) "The Theory of Industrial Organization", The MIT Press, Cambridge, Mass.

Vickers, J. (1986) "The Evolution of Market Structure when there is a Sequence of Innovations" *The Journal of Industrial Economics,* vol. 35, pp. 1 - 12.

Weiss, L. W. (1972) "The Geographic Size of Markets in Manufacturing" *The Review of Economics and Statistics*, vol. 54, pp. 245-266.

- and G.A. Pascoe (1986) "Adjusted Concentration Ratios in Manufacturing, 1972 and 1977. Statistical Report of the Bureau of Economics to the Federal Trade Commission, Washington DC.

White, H. (1980) "A Heteroscedastic-Consistent Covariance Matrix Estimator and a Direct Test of Heteroscedasticity" *Econometrica,* vol 48, pp 817 - 830.

Wilson, R. (1992) "Strategic Analysis of Auctions", Chapter 8 in: R.J. Aumann and S. Hart (eds.), Handbook of Game Theory, vol I, Elsevier Science Publishers, Amsterdam.