Market Dynamics, Evolution and Smallness

(Marktdynamiek, evolutie en kleinbedrijf)

PROEFSCHRIFT

ter verkrijging van de graad van doctor aan de Erasmus Universiteit Rotterdam op gezag van de Rector Magnificus Prof. dr P.W.C. Akkermans M.A. en volgens besluit van het college voor promoties

De openbare verdediging zal plaatsvinden op vrijdag 21 november 1997 om 13,30 uur

door

Martin Anthony Carree

geboren te Schiedam
Promotiecommissie

Promotor: Prof. dr A.R. Thurik

Overige leden: Prof. dr J. van Sinderen
Prof. dr L.E. Sleuwaegen
Prof. dr A. van Witteloostuijn
PREFACE

During my academic training in econometrics at the Erasmus University Rotterdam I was a student-assistant to two professors of the Econometric Institute. Johan Koerts was the first. We developed a course on multivariate analysis. Roy Thurik was the second. We did research on small business presence and entry and exit of firms. Next to that I assisted Ben Bode in his final stage of his PhD-thesis. All three inspired me in various ways and contributed to this thesis by paving the way to it.

This PhD-research was financially supported by NWO (ESR) under grant number 510-27-022. Roy Thurik took care of most of the rest. I benefitted from his suggestions and support throughout the last five years. He also stimulated me to go to international conferences and to Carnegie Mellon University. Two institutes have helped me significantly. These are the Tinbergen Institute and the EIM Small Business Research and Consultancy. The Tinbergen Institute provided me with a pleasant working environment and supported some of my trips financially. The EIM Small Business Research and Consultancy provided me with data for my research.

Next to Roy Thurik, two persons have contributed strongly to a chapter of my thesis. They are Luuk Klomp (Chapter 7) and Steven Klepper (Chapter 9). I have enjoyed working together with them. Useful comments were given by my small committee members Jarig van Sinderen, Leo Sleeuwaegen and Arjen van Witteloostuijn. I am also grateful to David Audretsch, René Belderbos, Peter Brouwer, Pieter Fris, René den Hertog, Clive Jie-a-Joen, Aad Kleijweg, Marcel Lever, Jeroen Potjes and Yvonne Prince for their contributions, large and small, to the continuous process of my research into market dynamics, evolution and smallness. I also thank Carien, Ine and Sylvia (TI), and Wilma (EOV) for important background support. Several colleagues have been pleasant company. I especially remember Albert de Vaal, Alex Koning, André Lucas, Bert Menkveld, Cees Dubbelman, Eric de Laat, Eric Vermeulen, Erno Kuiper, Ferrie Pot, Jan van Dalen, Jan de Kok, Joachim Stibora, Jurgen Sweegers, Huang Xin, Krijn Schep, Liang Peng, Piet-Hein Admiraal, Ralph Wildeman, Richard Paap and Stefan Straetmans.

Finally, I want to thank my parents and Jasperina. September 1997 is a month I will not forget. It was a period of finishing off my thesis, but also of receiving our first child, Annemarie.

Martin Carree
Rotterdam, September 1997
PREFACE

The fields of information processing and management have evolved significantly in recent years, with a proliferation of techniques and methodologies. This transformation has been driven by advances in technology and the increasing importance of data in decision-making processes. The goal of this introduction is to present a comprehensive overview of the field, highlighting key concepts and providing a foundation for further exploration.

This text is designed for students, researchers, and professionals who wish to gain a deeper understanding of information processing and management. It covers a range of topics, from foundational theories to practical applications, and includes case studies and examples to illustrate the concepts discussed.

We hope that this introduction will serve as a valuable resource for anyone seeking to understand the dynamic landscape of information processing and management. Whether you are a student just starting your academic journey or a seasoned professional looking to expand your knowledge, we believe this book will provide the insights and tools you need to succeed.

Yours sincerely,

[Author's Name]
## CONTENTS

1 Introduction
   1.1 Rethinking scale  
   1.2 Developments in Industrial Economics since the 1970s  
   1.3 Causes of the increase in small firm presence  
   1.4 Consequences of the increase in small firm presence  
   1.5 Outline of the book  

2 A Survey of Causes of Increased Small Business Presence  
   2.1 Introduction  
   2.2 The limits of size  
   2.3 Development of the share of small firms in the Netherlands and the U.S.  
   2.4 Seven reasons why the share of small firms has increased  
   2.5 Concluding remarks  

3 Profitability and Number of Firms: Their Dynamic Interaction  
   3.1 Introduction  
   3.2 The dynamic interaction between profitability and number of firms  
   3.3 Data  
   3.4 Empirical results  
   3.5 Conclusions  

4 Entry and Exit in Retailing: Displacement and Replacement  
   4.1 Introduction  
   4.2 Patterns of entry, exit and their determinants  
   4.3 The determinants of entry and exit in retailing  
   4.4 Effects and estimation of displacement and replacement  
   4.5 Data and empirical results  
   4.6 Average replacement and displacement periods and persistence of profits  
   4.7 Conclusions  

5 The Carrying Capacity and Entry and Exit Flows in Retailing  
   5.1 Introduction  
   5.2 The carrying capacity  
   5.3 Entry, exit and disequilibrium  
   5.4 Data and characteristics of the retail sector  
   5.5 Two implications of the oligopoly model  
   5.6 Empirical results for entry and exit flows in retailing  
   5.7 Summary and discussion  
   5.A Appendix
6 Small Firms and Economic Growth

6.1 Introduction
6.2 Competitive selection with differing rates of entry and adjustment
6.3 Large firm presence in European manufacturing
6.4 Empirical results
6.5 Changes in market structure and productivity in Dutch manufacturing
6.6 Smallness and economic growth

7 Small Business, Job Creation and Testing for Convergence

7.1 Introduction
7.2 Regression fallacy
7.3 Testing the convergence hypothesis
7.3.1 Convergence of productivity
7.3.2 Two new test statistics
7.3.3 Simulation experiment
7.3.4 Empirical results for 22 OECD countries
7.3.5 Conclusion
7.4 Convergence of firm sizes

8 The Life Cycle of a Competitive Industry

8.1 Introduction
8.2 The evolution of prices, output and number of firms
8.3 Empirical results
8.4 Simulated industry evolution
8.5 Conclusion

9 A First-Mover Advantage Model

9.1 Introduction
9.2 A first-mover advantage model
9.3 Some preliminary results

10 Epilogue

SUMMARY IN DUTCH

REFERENCES
CHAPTER 1

INTRODUCTION

The focus of this thesis is on market dynamics and market evolution. A common denominator throughout this thesis is the emphasis on the role of small firms. Understanding market processes that affect the structure and performance of industries is the primary goal of this book. To this end new models are introduced and confronted with empirical data. This empiricist view is a key common factor of the chapters of this book.

Market processes have been at the heart of economic science from the very outset. Economists devote much energy to investigations into the development of markets over time. These studies range from research on diffusion of new production technologies to research on the persistence of profits and from research on merger waves to research on growth and hazard rates of small market participants. In the current post-socialist era there is broad consensus among economists that competitive markets are the best guarantee for economic progress but there is still much discussion about how these markets actually work. Nowadays, governments seek to promote economic growth by providing the best possible conditions for the functioning of markets. The main instruments are the competition and anti-trust policy, privatization and deregulation, support of new ventures and industrial policy. The choice of instruments will depend on the specific industry setting and requires a careful examination of the effects on market processes in order to prevent undesirable consequences.

There are four elements in this thesis, that are essential to progress on the positive research question of how markets work and on the normative research question of how markets should work. First, market dynamics and market evolution are the subject of analysis. We are not so much interested in static market equilibria but more in the way markets evolve over time and the speed of this process. The competitive selection process of market participants is reflected most clearly in the processes of entry and exit. Three studies are almost exclusively devoted to these changes in the industry composition (Chapters 3 through 5). Second, special attention is paid to the role of small firms. The interest in small businesses has increased considerably in the last decade and many economists consider them as a major factor in current industries as innovators and specialists. The majority of entries and exits are small firms and some of the new businesses will grow into the large firms of tomorrow or even lay the foundation of completely new industries. Third, several new dynamic models are introduced in the thesis (Chapters 3, 4, 5, 8 and 9). The models provide a systematic
summary of the main interactions between variables of interest in markets in motion. Fourth, each of the models is confronted with extensive data sets from various sources. A rigorous empirical examination of the new models is pursued throughout the studies and some novel results are discussed in detail.

In this introductory chapter a short review is given in Section 1.1 of the changing attitude of economists with regard to small firms. This is followed in Section 1.2 by an overview of some recent developments in the field of Industrial Economics. In Section 1.3 six important causes of the increase in small firm presence are presented. The consequences of this increase are the topic of Section 1.4. The contents of the rest of this thesis are discussed in Section 1.5. It is followed by a brief overview of the collection of publications which form the basis for the chapters of this thesis.

1.1 Rethinking scale

Small firms seemed doomed to disappear and were largely ignored in economic science during most of this century. Giant corporations were seen as the most powerful engines of progress (Schumpeter (1950), Galbraith (1956)) and the constant decrease in the share of small firms in most Western economies seemed perfect evidence. The economic threat from centralized planned economies reporting strong productivity growth and technological development also led to a widespread faith in 'scale economies'. It was the era of the hierarchical industrial firm growing progressively larger through exploiting economics of scale and scope (Chandler (1990) and Teece (1993)). Only big firms would have the capacity to be efficient and, hence, small firms could be dismissed as unimportant. The conglomerate merger wave of the late 1960s seemed to have set the case.

Policy makers had no reason to question their belief in social engineering during the post war period of strong economic growth. The entrepreneur, presented as the prime cause of economic development by Schumpeter (1934), seemed to be dead and gone. The neo-classical price theory, assuming away the function of the entrepreneur, dominated economic science (Baumol (1968), Leibenstein (1968)). The economist's idea of competition became that of an achieved state of efficient allocation instead of an ongoing dynamic process. Economic policy was aimed at promoting stable markets necessary to accommodate mass-production technologies used by large corporations. The presumption that large firms were superior to small ones in terms of economic performance pushed policy makers in the direction of encouraging mergers, subsidizing large-scale enterprises and protecting state enterprises both in East and West (Acs (1996a)).

Times have changed. The centralized planned economy has proved to be a complete failure, and the question comes around why economists could ever have
claimed that such systems would work (Buchanan (1991)). It has now become an elementary proposition among economists that economies with private ownership of the means of production work better than economies where individual ownership is absent. The formerly centralized planned economies have had a hard time providing the empirical evidence for this. A characteristic of these economies was the almost complete absence of small firms, and this extreme monopolization was one of the major factors leading to the collapse of state socialism (Ellman (1993), Acs (1996a)). The development of small enterprises is considered a vital part of the current transition process in former Eastern Europe.¹

Times have also changed in Western economies. There is ample empirical evidence that the share of small business in manufacturing in Western economies has risen since the 1970s (Acs and Audretsch (1990a, 1993b)). Large firms have been downsizing and restructuring in order to concentrate on ‘core business’ again. In the meantime the entrepreneur has risen from the dead. High-technology innovative small firms have come at the forefront of technological development in many industries. Piore and Sabel (1984) claim that an ‘industrial divide’ has taken place: a moment in history when the path of technological development itself is at issue. The model of mass-production is being replaced by a model of flexible production, promoting the viability of small firms. After the failure of macroeconomic policy during the economic crises of the 1970s and early 1980s policy makers abandoned the ‘demand’ approach of stabilizing markets and started to shift to a more ‘supply’ approach of stimulating entrepreneurship.

Many of the policy changes in market economies in the last two decades have a common characteristic of shifting economic activities from the public sector to the private sector. Thousands of state firms in Eastern and Western Europe and in many developing countries have been privatised (Boycko et al. (1996)) and a strong deregulation movement has swept the world. Countries promote domestic competition in order to be able to compete successfully in the world market, directly in line with recommendations by Porter (1990). The very notion of competition has also changed back again. Static welfare theory has lost ground to evolutionary theories of innovation and competition in line with writings by Schumpeter and Hayek (Ikeda (1990)). At the same time the growing importance of small businesses in the market economy has put them on top of the research agenda in Industrial Economics (Acs (1996b), Acs and Audretsch (1989a), You (1995)). Jensen (1993) compares the changes which have been taking place since the early 1970s with earlier industrial revolutions: “The major

¹See for example Russia’s Shatalin Plan, which “is built on the assumption that society needs small enterprises to orient production to the needs of every person, to fight the dictatorship of monopolies in consumer and production markets, and to create a favourable environment for quick introduction of new scientific and technological ideas” (Nolan (1995), p. 82).
restructuring of the American business community that began in the 1970s and is continuing in the 1990s is being brought about by a variety of factors, including changes in physical and management technology, global competition, regulation, taxes, and the conversion of formerly closed, centrally planned socialist and communist economies to capitalism, along with open participation in international trade. These changes are significant in scope and effect; indeed, they are bringing about the *Third Industrial Revolution*" (p. 835, italics added).

Small and new firms have been claimed by 'small business economists' to play an important role in the economy serving as agents of change through their entrepreneurial activities, being the source of important innovative activity, stimulating industry evolution and creating an important share of the newly generated jobs (Acs (1996a)). The now much criticized study of Birch (1981), claiming that four out of five new jobs had been created by small firms, is often mentioned as the starting point of the increased interest in small firms.\(^2\) The real figure probably lies somewhat lower, but the interest did not fade away. Technological innovation by small firms is nowadays seen as one of the major reasons of America's regaining leadership in the world economy (Acs and Audretsch (1990a), U.S. Small Business Administration (1996)).

The changes in firm-size distributions and concentration ratios have usually been quite modest in most industries over the last decades. This could give the false impression of a low degree of turbulence. Audretsch (1995) shows that the degree of turbulence has been rising strongly since the 1960s in the U.S. One of the indications of the increased volatility is the time it took to replace one-third of the Fortune 500. In the 1950s and 1960s this took two decades, in the 1970s one decade and in the 1980s one-third of the Fortune 500 were replaced in only five years. Summary statistics of firm-size distributions fail to uncover such turbulence in market shares (Davies and Geroski (1997)). The newly emerging 'entrepreneurial' economy is remarkably turbulent when compared to the 'managed' economy dominated by large conglomerates (Audretsch and Thurik (1997)).

### 1.2 Developments in Industrial Economics since the 1970s

The scientific field of Industrial Economics flourishes like never before if one would consider the number of new journals as a reliable measure. A common theme in those new journals and in received journals as well is the increased attention for industry dynamics and industry evolution. The titles of some journals leave little doubt about that, e.g. *Industrial and Corporate Change, Journal of Business Venturing, Journal of* ____________

\(^2\) Probably the first economist in the Netherlands to recognize the revival and potential of small firms was Van der Zwan (1981).
**Evolutionary Economics and Structural Change and Economic Dynamics.** The idea that the economy is in a state of incessant restructuring and that industry structure, technology and the supporting institutions co-evolve gains more and more support (Nelson (1995, 1996), Caballero and Hammour (1996)). Industrial economists have tended to focus more and more on "the very competitive-entrepreneurial process which is at the heart of the market process itself" (Kirzner (1973), p. 20). Three very influential studies on the evolution of industry appeared in the year 1982. These were Nelson and Winter's book *An Evolutionary Theory of Economic Change*, Jovanovic's article 'Selection and the Evolution of Industry' and Gort and Klepper's article 'Time Paths in the Diffusion of Product Innovations'. A large collection of articles on learning, diffusion, selection and innovation was published in the subsequent 15 years (Dosi et al. (1995)).

Traditionally, industrial economists have focused mainly on market concentration (Scherer and Ross (1990) and Shepherd (1990)). Monopoly power was the main concern and researchers aimed at determining the extent of concentration, the degree to which it was detrimental to economic performance, and ways in which public policy could be designed to cope with problems of monopoly power. Their analyses were usually characterized by the use of a static framework and the neglect of smaller firms. After the 1970s the ‘New Industrial Organization’ became increasingly popular (Jacquemin (1987) and Tirole (1988)). The strategic interactions between market participants are the main focus of this research paradigm. This required some innovative methodological aspects, like noncooperative game theory. The field of Industrial Economics became more theoretical and less interested in empirical and case studies. However, research in the ‘New Industrial Organization’ paradigm still concentrated on large firms and still used a rather static framework. Jacquemin (1987) claims that the paradigm entails a dynamic approach but studies are usually satisfied with considering sequential (read two-period) decision making.

A second important branch in Industrial Economics has been the research into the dynamics and evolution of industry as mentioned before. Studies in this area concentrate on the processes of firms entering, growing, surviving or exiting, displacing incumbents, innovating and changing the market place. The diversity of approaches and their selection are thought to be essential to economic growth. Recent contributions to the field include Audretsch (1995) and Baldwin (1995). The field has clear roots in Schumpeter's seminal writings on economic development. However, it was not until the early 1980s that the study of industry dynamics and evolution blossomed. Early contributions like Mansfield (1962) suffered from limited availability of data and the paucity of adequate statistical models. The increasing quantity and quality of data resources and new econometric techniques like hazard rate analysis, error-correction models and panel data techniques, have increasingly enabled researchers to perform studies of market processes.
One may discriminate between industry dynamics and industry evolution on basis of the time span. Industry dynamics is concerned usually with relatively short time periods. It studies for example entry and exit rates, the growth of firms and the persistence of profits often using cross section or panel data. Industry evolution is more concerned about relatively long time periods. It studies the industry from its starting point to its current state and tries for example to explain the diffusion of innovations and the time paths in number and sizes of firms. Small firms play a vital role both in industry evolution and in industry dynamics. They are the seedbed of new industries (Beesley and Hamilton (1984)) and they constitute the major part of entering and exiting firms. They are an important source of innovation as they represent a wide variety of productive approaches to innovation collectively pursued in the industry (Cohen and Klepper (1992a)).

1.3 Causes of the increase in small firm presence

The limits of organization size have not been totally forgotten during the period of the expansion of large firms. Early contributions by Robinson (1934) and Coase (1937) and later contributions by Leibenstein (1966), Williamson (1967) and Arrow (1974) discuss the possible inefficiencies of large organizations. The revival of small and medium-sized enterprises has made economists even more critical about the merits of largeness. The first economists who explicitly concentrated on the reasons of the growing role of small firms since the 1970s were Brock and Evans (1989). In a review article of the state of small business economics they mention six reasons of the upswing of small business. These are:

(1) The 1970s and 1980s appear to be a period of creative destruction as described by Joseph Schumpeter. New industries like the computer and biotechnology industry have developed largely as a result of small innovative entrants. Prusa and Schmitz (1991) show for example that new firms have a comparative advantage in creating new software categories.

(2) Recent technological advances, most notably the enormous decrease in computer cost, have probably reduced the minimum scale of entry. Carlsson (1989a) argues in his study of metalworking industries that the emergence of new computer-based technology has improved the productivity of small and medium scale production relative to standardized mass-production.

(3) The integration of the world economy due to lower trade barriers has led to stronger foreign competition. Rising import competition has affected especially those industries
which could be characterized as tight oligopolies, like the U.S. automobile industry (Shepherd (1982)). It has also led to greater variability of sales which may have benefitted small firms using flexible production technologies (Fiegenbaum and Karnani (1991), Mills and Schumann (1985)).

(4) Smaller businesses are usually more labour-intensive than their larger counterparts. These firms may therefore have benefitted more from the supply of labour due to increased labour-force participation of women and the entry of the baby boomers into the labour market. The high unemployment rates in the 1970s and early 1980s not only caused real wage levels to decrease but may also have increased entry and decreased exit (Storey (1991)).

(5) Next to supply factors there has also been an increase in the demand for variety. This is partly caused by an increasing welfare (Jackson (1984)), the increase in international competition (Acs et al. (1990)) and cross cultural influences. As a consequence many speciality companies supplying a market niche have entered the market place. Bradbur and Ross (1989) find evidence that small firms may prosper in such niches when defended by mobility barriers.

(6) Deregulation has strongly affected the share of small firms in some industries, like the telecommunication sector. Phillips (1985) has shown for U.S. manufacturing how small firms have benefitted from the relaxation of entry regulations. Deregulation often goes hand in hand with privatization. Almost all of the former public companies downsize considerably when they are confronted with the market regime.

In Chapter 2 we will go into some more detail on these and an additional reason for the increase in small firm presence. But now we turn to the consequences of smallness.

1.4 Consequences of the increase in small firm presence

Some small business economists, like Acs et al. (1997), argue that small firms are the engines of global economic growth. This opinion is not shared by most other economists who would claim a much more prominent role for large enterprises. We discuss four approaches to the research question of the consequences of the presence of small firms:

(1) The first approach stresses the point that markets with relatively high small firm presence are expected also to display a relatively high degree of market competition because of the less concentrated market structure. Industrial economists have put a lot
of effort in the investigation of the relation between the structure of industries and their performance. Initially, their main concern was the loss of allocative efficiency due to a high degree of market concentration. This misallocation burden may be substantial for some industries but is generally thought to amount in total to only one or two per cent of GNP (Shepherd (1990), p. 132). A second concern became the loss of internal efficiency due to lack of competitive pressure. The concept of 'X-inefficiency' was introduced by Leibenstein (1966). A collection of empirical studies have shown since that the amount of X-inefficiency is significant and increasing with market concentration (Frantz (1988), ch. 8). A formal exposition of how internal inefficiencies in hierarchies develop can be found in Leibenstein (1987). Both of these two concerns are static in nature. Recently, Nickell (1996) has tried to measure the dynamic effect of lack of competition. He finds that competition, measured either by increased number of competitors or by lower level of rents, is associated with higher rates of total factor productivity growth. That is, market power appears to lead not only to a less efficient allocation of resources and more slack but also to less growth.

(2) Evolutionary economists take a somewhat different view on structure and performance. They consider both as endogenous and are more interested in long term effects of market settings. They usually stress the importance of competitive selection (Jovanovic (1982)). Entry and exit which have barely any effect on market structure in the short term may radically alter the market settings in a somewhat longer term. Eliasson (1995) shows that lack of entry may affect economic performance in a non-negligible extent only after a period of about 25 years. Many evolutionary models suggest that the importance of new and small firms may severely be underestimated when only looking at the short term. Firms which were started but one or two decades ago may in some cases grow into enterprises responsible for important technological progress. A disadvantage of many evolutionary models is their sensitivity to behavioural assumptions about the market participants.

(3) The entrepreneur was defined away by mainstream neoclassical economics but is regaining its position in economic science again. The positive effect of entrepreneurial activity on economic growth has become an elementary proposition among economists.

---

³See Leibenstein and Maital (1992) and Button and Weyman-Jones (1992) for an introduction into the measurement of X-inefficiency. In order to give a balanced account of the progress of microeconomic theory, it should be noted that neoclassical micro economists consider the concept of X-inefficiency as invalid because it cannot be reconciled with profit maximization or cost minimization (Stigler (1976)).

⁴The basic reference of the incorporation of the principles of biological evolution and natural selection into economic systems is Alchian (1950).
It is often argued that entrepreneurship is found particularly inside small firms and not so much inside large firms. Of course, there are exceptions like Intel and 3M (Burgelman (1994)), but the suboptimal incentive structures in many large corporations, due to very limited property rights for employees and bureaucratic inertia, may cause large enterprises not to be as ‘entrepreneurial’ as their smaller counterparts (Acs et al. (1997), Ginzberg and Vojta (1985) and Leibenstein (1987)).

(4) The fourth approach is mainly concerned with the role of small firms in creating new products and production processes. Small firms may have important strengths and weaknesses in a dynamic context (Nooteboom (1994)). In the early years of his career Schumpeter tended to stress the strengths of smallness while in later years he emphasized the strengths of large industrial enterprises. The changing role of firm size and competition as determinants of innovative activity over the industry life cycle or in the economy have again received considerable attention in the last decade. Arguments range from small firms being an important source of innovation especially in highly innovative industries (Acs and Audretsch (1987) and Prusa and Schmitz (1991)) and increased product market competition having a positive effect on the adoption of new technologies (Goel and Rich (1997)) to small firms playing only a marginal role in innovative activity when considering the entire industry life cycle (Cohen and Klepper (1996a)) and large firms being increasingly able to have important complementarities at the level of manufacturing, marketing and engineering (Milgrom and Roberts (1990)). Cooperation in the small firm sector may also provide important opportunities to overcome some economies of scale present in the development of new technologies (Oughton and Whittam (1997)).

1.5 Outline of the book

The industry is the unit of analysis in this thesis. The subject of analysis is market dynamics and market evolution with special attention for the role of small firms. I have intended to make each chapter independently readable without referring extensively to other chapters. This may sometimes lead to a mild form of redundancy across the separate chapters.

An overview of the development of the share of small firms in manufacturing and service industries in the Netherlands and the U.S. can be found in Chapter 2. Reasons for the increase of this share in most Western economies during the last two decades are reviewed afterwards. This discussion is an extension of Section 1.3. Chapter 3 concentrates on the determinants of net entry and the interaction between profitability and net entry. A new simultaneous equation model is developed and confronted with an extensive data set of Dutch retail industries. Chapter 4 continues with an overview
of the empirical literature of (dis)incentives to enter and exit and considers the possibility of entry causing exit (displacement) and exit causing entry (replacement). The displacement and replacement effects are studies in a new two-equation model which is applied on data of Dutch retail industries. The chapter introduces a new measure of market dynamics, viz. the replacement and displacement periods. Attention is paid to the recent decision of the Dutch Ministry of Economic Affairs to abolish some specific legal requirements to start a retail venture. In Chapter 5 a formal model of the local oligopoly structure in retailing is proposed and its parameters are estimated using the same data source. An important element of Chapters 3 and 5 is the error correction process. Following the Kirznerian notion of entrepreneurship we assume that (potential) entrepreneurs correct for market (structure) errors.

Chapter 6 contains a study on the effect of the structure of manufacturing industries on the growth of output for a sample of 13 European countries. It is followed by a short analysis of the effect of changes in market structure on productivity in Dutch manufacturing. The chapter also introduces a simple simulation model of competitive selection. Chapter 6 builds on Section 1.4 by testing the consequences of market structure on growth of productivity. Chapter 7 discusses how one can judge whether small firms grow faster than their larger counterparts or vice versa. It is a contribution to the currently heated debate about the role of small firms in the job generating process. Two new test statistics are derived which are used to examine whether complete economies converge and to examine convergence of firm size or growth rates.

In Chapter 8 we go from industry dynamics to industry evolution. An in-depth study of the U.S. tire industry during most of this century is made. Special attention is given to demand factors and the continuous learning-by-doing process in the tire industry as explanations of the shakeout in the number of producers which occurred in the 1920s. It comments on an article by Jovanovic and MacDonald (1994a) who propose a theory of industry evolution. Chapter 9 proposes a conceptually simple theory of industry evolution which can predict different aspects of industrial evolution. It is capable of predicting entry and exit rates of firms by cohort and age, of the firm-size distribution and of the R&D intensities per firm. A preliminary empirical investigation of the model is performed using entry and exit data of the U.S. tire industry. In the Epilogue some elements of a general model of the development of the presence of small business in modern economies are suggested. It combines indications that concentration within markets grows over time while the number of markets increases as well. These elements influence the development of the overall share of small firms. Periods in which this share increases may be followed by periods in which it decreases and vice versa.

Figure 1.1 summarizes the causal links which are the subject of research in the coming chapters. Central in the book and the figure is Industry Structure and
Composition. The composition and structure of industries is determined by the Entry and Exit of firms and by Changes in Market Shares of incumbents. The effects of Industry Structure and Composition on Market Growth and on Changes in Market Shares are investigated in Chapters 6 and 7, respectively. The effects of Technological Progress on Entry, Exit and Changes in Market Shares are covered in Chapters 2, 8 and 9. The interrelationship between Entry and Exit and the effect of Entry/Exit Barriers on these variables are studied in Chapter 4. The other causal links which involve effects of Market Growth and Industry Structure and Composition on (expected) Profitability and of Profitability on Entry and Exit are mainly covered in Chapters 3 and 5. Some elements of these effects are studied in Chapters 4 and 8, as well. Market growth is in Figure 1.1 supposed to affect entry and exit through its effect on (expected) profitability. The model discussed in Chapter 9 entails a dependency of Technological Progress on Industry Structure and Composition.

![Diagram of causal links]

Figure 1.1: An overview of the causal links covered in the book

Most of the chapters in this thesis are based upon previously written articles and papers. Below an overview is given of the chapters and the corresponding articles and papers. Some other publications which are related to the subject of this thesis include Carree, Potjes and Thurik (1993), Potjes, Carree and Thurik (1994) and Waarts, Carree and Wierenga (1991).

Chapter 2 is based upon three articles. Data on the share of small firms in the Dutch economy are derived from “Recent Developments in the Dutch Firm-Size Distribution”, published in Small Business Economics. The case study of the Dutch printing industry is based upon the article “Groeiende Rol van MKB in Grafische Bedrijfstak”, published in Graficus Magazine, a Dutch trade journal. The discussion on the reasons of increased small business presence is largely derived from "Het Aandeel van het
Midden- en Kleinbedrijf, published in Bedrijfskunde, a Dutch journal in the field of Business Economics. Chapter 3 is largely based upon the article "Profitability and Number of Firms: Their Dynamic Interaction in Dutch Retailing", published in the book Market Evolution: Competition and Cooperation. Additionally, some parts of the article "The Dynamics of Entry, Exit and Profitability: An Error Correction Approach for the Retail Industry", published in Small Business Economics, are used. Chapter 4 is based upon two articles: "Entry and Exit in Retailing: Incentives, Barriers, Displacement and Replacement", published in Review of Industrial Organization, and "De Gevolgen van de Nieuwe Vestigingswet", published in ESB, a Dutch journal on policy issues. Chapter 5 is an extension of the paper "The Carrying Capacity and Entry and Exit Flows", which was published in the Tinbergen Institute Research Bulletin. The chapter is under revision at an international journal. The empirical analysis on small firm presence in European industries is derived from "Small Firms and Economic Growth in Europe", to be published in Atlantic Economic Journal. An extended version of this article is published as an OCFEB Research Memorandum. Chapter 7 is largely a combination of two articles: "Small Business and Job Creation: A Comment", published in Small Business Economics and "Testing the Convergence Hypothesis: A Comment", to be published in Review of Economics and Statistics. The research paper on the life cycle of the U.S. tire industry, presented in Chapter 8, is published as a TRACE Discussion Paper. Finally, Chapter 9 is based on the paper "Estimating a Structural Model of Industry Shakeouts", of which a revised version is to be published in Journal of Evolutionary Economics.
CHAPTER 2

A SURVEY OF CAUSES OF INCREASED SMALL BUSINESS PRESENCE

The increasing importance of small firms in modern day economies is central in the current survey. A short review is given of the theoretical literature of the determinants of firm size and firm-size distribution. This is followed by some empirical material on 'smallness' in the U.S. and the Netherlands. Evidence on the industry level is provided for the printing industry in three European countries. The data show that the long-term trend towards larger firm sizes has stopped, at least in manufacturing. The second part of the survey goes into detail about the seven reasons why small firms have regained market shares.
2.1 Introduction

During the 1980s an unexpected reappraisal of the role and importance of small firms emerged. Some economists started to appreciate the important role of small firms in technological change, as agents of change and in the job generating process (Acs and Audretsch (1993a)). Two important reasons for this appreciation were the economic stagnation and high unemployment of the 1970s and early 1980s and the increasing awareness of diseconomies of scale. Regarding the latter, Jensen (1993, p.842) argues that “It is far less valuable for people to be in the same geographical location to work together effectively, and this is encouraging smaller, more efficient, entrepreneurial organizing units that cooperate through technology”. One of the first to question the widespread faith in large scale economies is Gold (1981) who finds little support for scale potentials from either the theoretical or empirical literature.

The scientific field of Industrial Economics has been preoccupied with the behaviour and role of large enterprises at least since World War II. This can be concluded from the vast amount of empirical studies of the effects of market concentration (Cubbin (1988), Curry and George (1983), Schmalensee (1989) and Waterson (1984)). Only a few empirical studies have concentrated on the determinants of the share of small and medium-sized firms in industries. Examples are Acs and Audretsch (1989d), Baldwin and Gorecki (1985), Carree, Potjes (1993), Schwalbach (1989) and White (1982).

The empirical fact that firms come in different sizes, small and large, was also largely neglected in the theoretical literature. The principal determinant of firm size was assumed to be the long-run average cost curve. The efficient firm size could simply be found at that point where the cost was minimized. The traditional analysis predicts that different firm sizes in a homogeneous product market may only persist in case the long-run average cost curve is flat over some range (You (1995)). Two important papers extended the concept of equilibrium firm size to that of equilibrium firm-size distribution. Lucas (1978) argues that the observed size distribution of business firms results from the distribution of persons by managerial talent. More talented managers run the larger firms while persons with managerial talent below a cutoff level become workers. Kihlstrom and Laffont (1979) focus on risk aversion as the determinant which explains who becomes an entrepreneur and who works as a labourer. Following Knight (1921)

---

5 The new attention for small businesses emerges from the foundation of Small Business Economics in 1989 and the Global Conferences on Small Business Economics, of which the third edition was held in Rotterdam in August 1994. See Carree and Lever (1994) for a summary of this conference.

6 Recently, Jovanovic (1994) has shown that in a model with both managerial ability and wage-working ability, some (extreme) cases may arise in which the worst managers in a group of persons end up doing the managing. He gives the example of professional sports.
they argue that the higher the risk aversion the less likely an individual will start an enterprise. The least risk averse entrepreneurs are predicted to operate the largest firms. Both Lucas and Kihlstrom and Laffont have entrepreneurial ability (to manage production or to take risks) as the specific scarce input that limits the size of a firm.

The two papers were swiftly followed by studies extending equilibrium firm-size distribution models to explain other empirical regularities. Calvo and Wellisz (1980) extend Lucas' model by introducing a learning process through which managers acquire the necessary knowledge and by considering the age of entrepreneurs. Rosen (1982) models the distributions of firm size, span of control, and rewards as the joint outcome of market assignments of personnel to hierarchical positions. Oi (1983) introduces monitoring costs to explain a wide range of phenomena, like the capital intensity, injury frequency rates and the incidence of shift work.

There have been at least three additional theories alongside that of assuming a distribution of entrepreneurial abilities over individuals to explain an equilibrium firm-size distribution. In the first theory the firm-size distribution is derived from the stochastic growth process of firms. Formal analyses and empirical tests of this theory can be found in Ijiri and Simon (1977). If one assumes that the growth rate of size is normally distributed and independent of size and over periods, then one may derive that the limiting firm-size distribution is log-normal. The second theory explains the coexistence of small and large firms from adjustment costs when changing the size of the company (Lucas (1967)). A related theory in organisation science is the structural inertia theory developed by Hannan and Freeman (1984). They argue that organisations ought to become to some extent inert in order to be successful. The third theory was developed in Mills (1984) and Mills and Schumann (1985). Large and small firms coexist in this theory because large firms display more efficient production but lower flexibility, while it is the other way around for their smaller counterparts. Dosi (1988) supports this theory in the field of R&D investments and innovative output: "each technological paradigm is characterized by different trade-offs between flexibility and economics of scale" (p. 1153). For a more extensive treatment of these theories, see Brock and Evans (1986,1989) and You (1995).

Studies on the equilibrium firm-size distribution are strongly related to new models of optimal firm size and organization of production. Changes in the factors which determine the size of the firm will of course influence the firm-size distribution as well. The next section discusses some of these models. They give some idea of the factors...

---

7 See Hannan et al. (1990) for this and alternative assumptions. They show that size-localized competition may lead to a distribution with peaks at isolated subpopulations.

8 See Baum (1996) for a survey of the empirical evidence on this theory.
that limit the size of large, hierarchical, firms. Some of the causes of increased small business presence are directly related to the theoretical limitations to the number of employees in large firms. Before we discuss the seven causes of the shift in the firm-size distribution, we first present some data on the development of the share of small firms in the Netherlands and the U.S. In Section 2.3 some data are presented for the Fortune 500 firms, self-employment rates, the Dutch economy (1978-89), the American economy (1993-94), and the printing industry. These data show that the share of small firms in manufacturing industries has risen considerably. It is followed by Section 2.4 in which we discuss seven reasons why these shares have been rising since the mid 1970s. This section is an extension of Section 1.3. Section 2.5 concludes.

2.2 The limits of size

Oliver Williamson derives much of his fame from the field of transaction cost economics, but his 1967 analysis of control loss as a limitation to firm size also stands out as a seminal contribution. A decade or more before other researchers did so, Williamson showed how the cumulative effects of control loss between successive hierarchical levels could be responsible for the optimal firm size to be finite. He concludes that "an irreducible minimum degree of control loss results from the simple serial reproduction distortion that occurs in communicating across successive hierarchies. If, in addition, goals differ between hierarchical levels, the loss in control can be more extensive." (p. 135). That is, there are two major limitations to firm size. The first is the problem of **coordination**. The second is the problem of **incentives**. Subordinates may shirk on the job or pursue their own interests when effort is not observed by their superiors. Large hierarchical companies, governments and centrally-planned economies all suffer from these two basic problems. A short discussion of some interesting papers which have been published on the subject of limits of firm can be found below. Two important books about the disadvantages of large companies are written by Ginzberg and Vojta (1985) and Leibenstein (1987). They provide a clear description of how large size may lead to less alertness, less motivation and difficult communication.

Three apparently independent studies on the limit imposed on firm size by the heights of the hierarchical structure appeared in the late 1970s. Beckmann (1977) proposes a model in which 'supervision' is produced in each layer of the hierarchy except the bottom level in which 'supervision' and labour input are combined into final output. Calvo and Wellisz (1978) show that Williamson's model is valid only in case of imperfect supervision and furthermore show a rationale to pay higher wages to

---

9See Hess (1983) for an extensive survey of the way in which problems of coordination and incentives affect the organization of production.
supervisors than to supervisees even if persons have identical abilities. Keren and Levhari (1979) concentrate on the costs caused by delays in decision making in hierarchies and formulate the object of the hierarchy as the minimization of the time taken to prepare a set of instructions for the workers. A general result in their analysis is that spans of control in business organizations should increase when going down the levels of hierarchy. They extend this analysis in two later papers (Keren and Levhari (1983, 1989)).

Qian (1994) presents a model which he claims to be a combination and generalization to the models of Williamson, Calvo and Wellisz and Keren and Levhari. He shows that in an optimal hierarchy wages and effort levels decrease as one moves down the hierarchy. However, even his comprehensive model does not include some new considerations. Camacho (1991), for example, introduces adaptation costs which increase with the size of firm. He derives that the greater the variability of the environment the smaller the optimal size. Aron (1988) includes the possibility of diversification into multiple product lines. She derives that the optimal firm size and the degree of diversification increase in the managerial ability and are therefore positively correlated cross sectionally. A new and interesting approach was taken by Gifford (1992a,1992b,1993). In her model entrepreneurs have to choose between evaluating and restoring current projects and evaluating and adopting new projects. The optimal allocation of attention implies a bound on the size of the firm. She is also able to derive a relation between optimal innovation and firm size. Wiggins (1995) focuses on the key advantage provided by ownership to explain why entrepreneurial activities are carried out in small enterprises rather than in large firms where the entrepreneur is an employee. He predicts that worker-owned, entrepreneurial enterprises are more likely to emerge when the likelihood of success of the enterprise is small such as in the case of pathbreaking innovations. A recent approach to the limits of size can be found in Otani (1996). He builds upon Hayek’s normative view of the economic system to suggest why firms cannot expand infinitely. The main reason for this is the limitation of the ability of entrepreneurs to collect knowledge.

2.3 Development of the share of small firms in the Netherlands and the U.S.

The continuous increase of average firm and plant size in many manufacturing industries reversed in most Western economies during the 1970s. In this section some empirical evidence is presented for the increasing importance of small firms in nowadays Western economies. The evidence will be presented in four separate sections. The first concentrates on the Fortune 500, the 500 largest industrial companies in the U.S. The second discusses the development in self-employment rates in the U.S. and other Western countries. This is followed by a short description of some
developments in the share of small firms in the Dutch and U.S. economy. In the last section a case study of the European printing industry is presented. This section is based on Carree and Prince (1993). There are other indications that the share of small firms has increased. Gartner and Shane (1995) claim, for example, that the number of organizations per 1000 people has more than doubled in the U.S. during the 1980s.

(i) The Fortune 500

A popular measure to show the increased importance of small and medium-sized firms is the share of the Fortune 500 companies in employment and output. The share of the 500 largest U.S. industrial companies in total U.S. employment dropped from 21 per cent in 1969 to 11 per cent in 1991 (Jovanovic (1993), p.203), but this is partly caused by the decrease in the share of manufacturing employment in total U.S. employment. A similar change can be seen in the share of the Fortune 500 in U.S. GDP. This share was 37% in 1954 and rose to 58% in 1979 but then started to drop quickly to only 42% in 1989 (Carree, Den Hertog and Thurik (1993)). Carlsson (1989a, p. 28) provides data which are not influenced by the decreasing share of manufacturing in the total economy. He shows that the share of the Fortune 500 companies in total manufacturing employment dropped from 78.7% in 1975 to 72.5% in 1985. In the same period the share of these firms in total shipments dropped from 83.2% to 77.2%.

(ii) Self-Employment

The fraction of the labour force that is self-employed has decreased in most Western economies until the mid 1970s. Since then the self-employment rate has started to rise again in several of these economies. Blau (1987) observes that the proportions of both male and female self-employed in the nonagricultural labour force declined during most of this century but bottomed out in the early 1970s and started to rise until at least 1982. Evans and Leighton (1989b) show that the rate of unincorporated self-employment among white men fell from an average of 8.2% in 1967 to 7.6% in 1975 and afterwards rose to 8.6% in 1983. However, they also note that this figure dropped again to 7.7% in 1985. A comprehensive study on the development of self-employment rates is Acs et al. (1991). They document cross-section and time-series data on industrialized and lesser-developed countries. They show that of 23 OECD countries, 15 had an increase in the self-employment rate during the 1970s and 1980s. The self-employment rate started to rise again in for example Germany, Italy, New Zealand, Spain, Sweden and the U.S. It did not so however in for example France, Japan and the Netherlands. The weighted average of the self-employment rate in OECD-countries is reported to rise slightly from 8.4% in 1978 to 8.9% in 1987. It is important to note that Acs et al. think
that self-employment will continue its downward trend despite the recent reversal of this trend in many industrialized countries. They argue that the factors which have stimulated self-employment were only temporary.

The determinants of the individual choice between self-employment and employment have been investigated intensively in the last decade. Aside from the studies mentioned above, recent empirical work includes Evans and Jovanovic (1989), Evans and Leighton (1989a), De Wit and Van Winden (1989,1990) and Holtz-Eakin et al. (1994a,1994b). An important conclusion from these studies is that liquidity constraints play a role when considering to set up a firm. Subsidies for individuals starting a new firm are therefore likely to increase self-employment. It implies that a policy of stimulating entrepreneurship should include special attention for venture capital. See De Wit (1991,1993) for an extensive survey of models of self-employment and De Wit and Van Winden (1991) for a comprehensive model to determine the number of self-employed individuals in an economy.

(iii) Small firms in the American and Dutch Economy

Loveman and Sengenberger (1991) and Acs and Audretsch (1993a) provide some data on the development of small business presence in a collection of Western countries. They show that the growing importance of large business has come to a halt somewhere in the 1970s. In this section we concentrate on the development of the firm-size distributions in the Dutch economy and on recent developments in the U.S. economy. Table 2.1 gives an overview of 31 Dutch second-digit industries. The first 16 industries (ISIC 31-39) are manufacturing industries and the last 15 industries are non-manufacturing industries (ISIC 50-951). The 31 industries cover 91% of total employment (3.7 million f.t.e.’s) in the private sector in 1987. Firms without employees are excluded from the analysis. The table shows that the average firm size has gone down for most manufacturing industries during the 1978-89 period. Very substantial decreases can be found in textiles (ISIC 321), wearing apparel (ISIC 322), electrical engineering (ISIC 383) and instrument engineering (ISIC 385). The overall picture in non-manufacturing industries is the reverse: most of these industries show an increase in the average firm size over the 1978-89 period. This is most striking in banking and finance (ISIC 81) and business services (ISIC 832). In the last column of Table 2.1 a different measure of the development of the firm-size distribution is presented. It is the development of the ratio between the number of firms which employ 1 to 10 employees and the number of firms with 100 or more employees. The two measures usually suggest the same change in the firm-size distribution. Two exceptions are insurance (ISIC 82) and renting of machines and other movables (ISIC 833).
<table>
<thead>
<tr>
<th>ISIC</th>
<th>SBI</th>
<th>Industry</th>
<th>EMP</th>
<th>MEAN</th>
<th>RATE</th>
</tr>
</thead>
<tbody>
<tr>
<td>31</td>
<td>20,21</td>
<td>Food, beverages and tobacco</td>
<td>161</td>
<td>+10</td>
<td>-6</td>
</tr>
<tr>
<td>321</td>
<td>22</td>
<td>Textiles</td>
<td>38</td>
<td>-36</td>
<td>+105</td>
</tr>
<tr>
<td>322</td>
<td>23</td>
<td>Wearing apparel except footwear</td>
<td>21</td>
<td>-43</td>
<td>+131</td>
</tr>
<tr>
<td>323/4</td>
<td>24</td>
<td>Leather, footwear and leather products</td>
<td>9</td>
<td>-18</td>
<td>+71</td>
</tr>
<tr>
<td>33</td>
<td>25</td>
<td>Wood products, including furniture</td>
<td>44</td>
<td>-16</td>
<td>+96</td>
</tr>
<tr>
<td>341</td>
<td>26</td>
<td>Paper and paper products</td>
<td>25</td>
<td>-4</td>
<td>+10</td>
</tr>
<tr>
<td>342</td>
<td>27</td>
<td>Printing and publishing</td>
<td>65</td>
<td>-18</td>
<td>+53</td>
</tr>
<tr>
<td>351/2</td>
<td>29,30</td>
<td>Chemicals, artificial &amp; synthetic filaments and staple fibres</td>
<td>85</td>
<td>+7</td>
<td>-24</td>
</tr>
<tr>
<td>355/6</td>
<td>31</td>
<td>Rubber and plastic products</td>
<td>25</td>
<td>-22</td>
<td>+49</td>
</tr>
<tr>
<td>36</td>
<td>32</td>
<td>Building materials, earthenware, glass and glass products</td>
<td>37</td>
<td>-6</td>
<td>+12</td>
</tr>
<tr>
<td>381</td>
<td>34</td>
<td>Metal products except machinery and transport equipment</td>
<td>93</td>
<td>+3</td>
<td>-2</td>
</tr>
<tr>
<td>382</td>
<td>35</td>
<td>Mechanical engineering</td>
<td>85</td>
<td>-27</td>
<td>+61</td>
</tr>
<tr>
<td>383</td>
<td>36</td>
<td>Electrical engineering</td>
<td>109</td>
<td>-53</td>
<td>+120</td>
</tr>
<tr>
<td>384</td>
<td>37</td>
<td>Transport equipment</td>
<td>71</td>
<td>-7</td>
<td>+3</td>
</tr>
<tr>
<td>385</td>
<td>38</td>
<td>Instrument engineering</td>
<td>12</td>
<td>-34</td>
<td>+122</td>
</tr>
<tr>
<td>39</td>
<td>39</td>
<td>Other manufacturing industries</td>
<td>7</td>
<td>+6</td>
<td>-7</td>
</tr>
<tr>
<td>50</td>
<td>51</td>
<td>Construction</td>
<td>300</td>
<td>-4</td>
<td>-1</td>
</tr>
<tr>
<td>50</td>
<td>52</td>
<td>Installation on construction projects</td>
<td>82</td>
<td>0</td>
<td>+9</td>
</tr>
<tr>
<td>61</td>
<td>61-64</td>
<td>Wholesale trade, commodity brokers and agents</td>
<td>284</td>
<td>-11</td>
<td>+25</td>
</tr>
<tr>
<td>62</td>
<td>65,66</td>
<td>Retail trade</td>
<td>282</td>
<td>+13</td>
<td>6</td>
</tr>
<tr>
<td>63</td>
<td>67</td>
<td>Hotels, restaurants and cafes</td>
<td>61</td>
<td>+18</td>
<td>-16</td>
</tr>
<tr>
<td>711</td>
<td>71,72</td>
<td>Railways and road transport</td>
<td>113</td>
<td>+6</td>
<td>-29</td>
</tr>
<tr>
<td>7121/3</td>
<td>73</td>
<td>Sea transport</td>
<td>34</td>
<td>-16</td>
<td>+2</td>
</tr>
<tr>
<td>719</td>
<td>76</td>
<td>Supporting services to transport</td>
<td>24</td>
<td>-14</td>
<td>+67</td>
</tr>
<tr>
<td>81</td>
<td>81</td>
<td>Banking and finance</td>
<td>104</td>
<td>+36</td>
<td>-27</td>
</tr>
<tr>
<td>82</td>
<td>82</td>
<td>Insurance</td>
<td>49</td>
<td>+7</td>
<td>+27</td>
</tr>
<tr>
<td>831</td>
<td>83</td>
<td>Real estate</td>
<td>25</td>
<td>+15</td>
<td>-6</td>
</tr>
<tr>
<td>832</td>
<td>84</td>
<td>Business services</td>
<td>194</td>
<td>+29</td>
<td>-17</td>
</tr>
<tr>
<td>833</td>
<td>85</td>
<td>Renting of machines and other movables</td>
<td>4</td>
<td>+10</td>
<td>+40</td>
</tr>
<tr>
<td>9</td>
<td>9</td>
<td>Other services</td>
<td>514</td>
<td>+14</td>
<td>-11</td>
</tr>
<tr>
<td>951</td>
<td>68</td>
<td>Repair of consumer goods</td>
<td>66</td>
<td>-18</td>
<td>+148</td>
</tr>
</tbody>
</table>

Note: ISIC stands for the ‘International Standard Industrial Classification’ of 1974 (revision 2). SBI stands for the ‘Standaaard Bedrijfslndeling’ of the Netherlands Bureau of Statistics (CBS). EMP is the total number of employees in 1000s in 1978. MEAN is the percentage change of the average employment over the 1978-1989 period and RATE is the percentage change of the ratio of number of firms in the size class 1-10 employees to number of firms in the size class 100 or more employees. Source: Carree and Thurik (1991). Netherlands Bureau of Statistics (CBS) and EIM Small Business Research and Consultancy.
Much of the empirical evidence suggesting that the share of small firms increases, especially in manufacturing, dates from the 1970s and 1980s. However, recent evidence on the U.S. economy does not suggest that large firms have regained their position of the 1950s and 1960s. In Table 2.2 an overview is given of the growth of employment in the eight sectors of the private sector in the U.S. During the year 1994 employment growth was more than average in industries which were dominated by firms with fewer than 500 employees in 7 out of 8 sectors. The only exception was the construction sector, but in that sector industries dominated by firms with less than 500 employees barely perform worse in terms of employment growth. Industries dominated by large businesses (500+ employees) show an employment growth which is strongly lower than average except in the mining and finance sectors where it is about equal to the average.

Table 2.2: Percentage change in employment by size category in the U.S. in 1994

<table>
<thead>
<tr>
<th>Sector</th>
<th>Total</th>
<th>SB dominated ind.</th>
<th>Indeterminate ind.</th>
<th>LB dominated ind.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mining</td>
<td>-3.73</td>
<td>0.65</td>
<td>-5.26</td>
<td>-3.32</td>
</tr>
<tr>
<td>Manufacturing</td>
<td>1.98</td>
<td>3.40</td>
<td>2.96</td>
<td>0.31</td>
</tr>
<tr>
<td>Construction</td>
<td>7.48</td>
<td>7.12</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>Wholesale Trade</td>
<td>3.18</td>
<td>3.45</td>
<td>1.56</td>
<td>0.60</td>
</tr>
<tr>
<td>Retail Trade</td>
<td>3.50</td>
<td>4.64</td>
<td>6.08</td>
<td>2.08</td>
</tr>
<tr>
<td>Transportation</td>
<td>3.99</td>
<td>5.12</td>
<td>5.08</td>
<td>-0.58</td>
</tr>
<tr>
<td>Finance</td>
<td>0.74</td>
<td>2.86</td>
<td>-9.89</td>
<td>0.16</td>
</tr>
<tr>
<td>Services</td>
<td>4.56</td>
<td>5.02</td>
<td>7.58</td>
<td>2.65</td>
</tr>
</tbody>
</table>

Source: U.S. Small Business Administration (1996), Table 1.13. Small Business dominated industries are industries in which firms with fewer than 500 employees account for 60 per cent or more of total employment. Large Business dominated industries are industries in which firms with more than 500 employees account for 60 per cent or more of total employment. Other industries are classified as 'Indeterminate industries'. NA indicates lack of industry representation within a size category. Data exclude self-employed workers.

(iv) The case of the European printing industry

The empirical evidence presented so far covers the entire economy. In this section we concentrate on a specific industry, viz. the printing industry. Table 2.1 shows that the Dutch printing and publishing industry has an increase of the share of small and medium-sized businesses which is somewhat above the average of manufacturing industries. The average SME share over the 1971-87 period in this industry was 48.9% but it increased over this period with 4.6 percentage points. A more in-depth study has shown that this increase is even stronger in the third-digit-level printing industry.
(publishers are left out). The following table shows how this share has developed in Denmark, the U.K. (publishers included) and the Netherlands.

Table 2.3: The printing industry in Denmark, the U.K. and the Netherlands

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of firms</td>
<td>1907</td>
<td>2052</td>
<td>11155</td>
<td>19232</td>
<td>1893</td>
<td>2592</td>
</tr>
<tr>
<td>Employment</td>
<td>29142</td>
<td>32990</td>
<td>323300</td>
<td>313500</td>
<td>54642</td>
<td>44786</td>
</tr>
<tr>
<td>SME share</td>
<td>58.9</td>
<td>63.5</td>
<td>37.6</td>
<td>47.0</td>
<td>44.8</td>
<td>56.9</td>
</tr>
</tbody>
</table>

Source: Carree and Prince (1993). The SME share is measured as the percentage of employment in firms with less than 100 employees in total employment. The data are for the printing industry (SBI 27.1) for Denmark and the Netherlands while the data are for the printing and publishing industries (SBI 27.1 and 27.3) for the UK.

The table indicates that the SME share has grown independent of whether total employment has risen or has decreased. The yearly accounts of the KVGO (Koninklijk Verbond van Grafische Ondernemingen) allow for a detailed picture of the development of the size distribution in the Dutch printing industry.

Table 2.4: Size classes in the Dutch printing industry measured in number of employees

<table>
<thead>
<tr>
<th>Year</th>
<th>1-4</th>
<th>5-9</th>
<th>10-19</th>
<th>20-49</th>
<th>50-99</th>
<th>100-199</th>
<th>200+</th>
<th>firms</th>
<th>employment</th>
</tr>
</thead>
<tbody>
<tr>
<td>1977</td>
<td>5.2</td>
<td>8.0</td>
<td>12.2</td>
<td>17.2</td>
<td>14.9</td>
<td>14.0</td>
<td>28.5</td>
<td>2269</td>
<td>43148</td>
</tr>
<tr>
<td>1982</td>
<td>6.0</td>
<td>8.5</td>
<td>12.4</td>
<td>18.3</td>
<td>15.4</td>
<td>12.1</td>
<td>27.3</td>
<td>2523</td>
<td>43836</td>
</tr>
<tr>
<td>1987</td>
<td>5.6</td>
<td>8.7</td>
<td>13.0</td>
<td>21.7</td>
<td>14.0</td>
<td>14.1</td>
<td>22.9</td>
<td>2441</td>
<td>43105</td>
</tr>
<tr>
<td>1988</td>
<td>5.5</td>
<td>8.9</td>
<td>12.7</td>
<td>22.7</td>
<td>13.3</td>
<td>13.3</td>
<td>23.7</td>
<td>2491</td>
<td>44573</td>
</tr>
<tr>
<td>1989</td>
<td>4.7</td>
<td>8.1</td>
<td>12.0</td>
<td>22.3</td>
<td>14.7</td>
<td>14.2</td>
<td>24.0</td>
<td>2770</td>
<td>53686</td>
</tr>
<tr>
<td>1990</td>
<td>4.7</td>
<td>7.9</td>
<td>12.3</td>
<td>23.1</td>
<td>15.4</td>
<td>14.9</td>
<td>21.7</td>
<td>2853</td>
<td>55653</td>
</tr>
<tr>
<td>1991</td>
<td>4.8</td>
<td>8.2</td>
<td>12.3</td>
<td>23.3</td>
<td>15.9</td>
<td>14.1</td>
<td>21.4</td>
<td>2942</td>
<td>56986</td>
</tr>
<tr>
<td>1992</td>
<td>4.8</td>
<td>8.8</td>
<td>12.0</td>
<td>24.4</td>
<td>16.2</td>
<td>13.2</td>
<td>20.6</td>
<td>3009</td>
<td>56841</td>
</tr>
</tbody>
</table>

*From 1989 on the coverage of firms is improved because FGE members are also included. Source: yearly accounts of the KVGO.

Table 2.4 shows that the share of medium-sized firms between 20 and 50 employees has grown considerably. This is probably the consequence of an increase in the demand for specialized products with a short delivery time and the increasing technical possibilities to meet this demand. These medium-sized firms are probably small enough to react quickly and large enough to finance the necessary new machinery. An additional reason for the decrease in the average firm size in the printing industry is the introduction of DTP (Desk Top Publishing). Many of the pre-press activities are
performed outside the printing-office nowadays. A recent development is the entry of small 'copy shops' which benefit from the low investment costs of computer and printing facilities.

2.4 Seven reasons why the share of small firms has increased

We now turn to the reasons why the firm-size distribution has shifted in the last two or three decades. The shift from mass production to specialized and differentiated production processes which has been taking place since the mid 1970s plays a central role in the increased changes of survival and growth for small and medium-sized enterprises. Several large companies have suffered from this shift while a lot of small firms have profited. See Case (1992) and Jensen (1993) for several interesting examples from the American industrial landscape. In this section we discuss seven reasons why specialized small firms have prospered in a period in which the majority of large firms were downsizing to survive. The first six reasons come from Brock and Evans (1989) and were already mentioned in Section 1.3. The last reason can be found in Carlsson (1989a). See also Acs and Audretsch (1990a), Aiginger and Tichy (1991), Evans (1991) and Loveman and Sengenberger (1991). The relative importance of these reasons will differ between industries because the role and characteristics of small and medium-sized firms differ as well (Nooteboom (1986b)).

(i) creative destruction

Piore and Sable (1984) use the term Industrial Divide, while Jensen (1993) prefers the term Third Industrial Revolution to describe the substantial restructuring of the business community during the last 25 years. Many economists agree that we experience a period of Creative Destruction, as described in Schumpeter (1950). The most obvious evidence is the emergence of new industries like the computer and biotechnology industries. Small firms play an important role in these new industries. Acs and Audretsch (1987) provide empirical evidence that small firms have the relative innovative evidence in industries which are highly innovative. Audretsch (1991) presents evidence that new firms have higher changes of survival in industries with high small-firm innovative activity when compared to industries in which this activity is low. Innovative strategies can often compensate for the scale disadvantages of small firms (Porter (1979)). Evidence for the comparative advantage of small firms in inventing radically new products or processes is given in Prusa and Schmitz (1991) and Rothwell (1983,1984). This is in line with theories expressed by Schumpeter (1939) and Williamson (1975) who emphasize that entrepreneurs play a significant part in the establishment of new industries, while large firms have advantages at later stages
because of the high costs of innovations or because of the need of financial or managerial resources. Cohen and Klepper (1996a) presents some empirical evidence for this change in the role of firm size over the industry life cycle. They emphasize the importance of the R&D cost spreading advantage of large size.

A period of creative destruction is one in which not price competition counts but instead "competition from the new commodity, the new technology, the new source of supply, the new type of organization" (Schumpeter (1950), p. 84). Small firms may survive or prosper in such periods by means of strategic instruments like flexibility and innovative activity (Acs and Audretsch (1989d), Caves and Pugel (1980) and Fiegenbaum and Karnani (1991)). An important incentive for these small firms is that the entrepreneurs own the company (Wiggins (1995)). Kraft (1989), for example, shows for a sample of West German firms operating in the metal industry that firms led by hired managers tend to innovate less intensively than owner-managed enterprises. In case innovative activity is the prime strategic instrument specialized and highly skilled employees of large companies are often the ones who start new ventures (spin-offs). They can then exploit the comparative advantages of small firms, like alertness, strong motivation and the capability of supplying specialized products (Nootoebum (1987)).

(ii) technological advances

The introduction of 'Numerical Controls' (NC) during the 1970s meant a reversal of a technological trend. The new technologies were not exclusively aimed at making mass production processes even more efficient, but also provided important opportunities for product differentiation, flexibility and innovation. The large decreases in computer costs have on average reduced the importance of scale economies in many sectors and small technology-based firms started to challenge large companies which still had every confidence in mass production techniques (Carlsson (1989a) and Thurik (1993)). A particularly interesting study on the benefits from high technology manufacturing programs is Meredith (1987). He argues that small firms are just as well, or better, equipped to implement technological advances, like CAD/CAM, and predicts the factory of the future to be a small factory. Small business owners often can more clearly see the benefits and problems of their automation decisions as they are very selective in where they invest their capital and because they are close to the operational level. They can use the new technologies to further extend the competitive advantages of providing service, fast innovation, customization and fast reaction in the marketplace and within their own facilities. Jovanovic (1993) notes another effect of modern technology: "recent advances in information technology have made market-based coordination cheaper relative to internal coordination and have partially caused the recent decline in firm size and diversification" (p. 221, italics added).
Acs et al. (1990) note that the large companies were the first to introduce NC as they had better financial resources and research facilities. However, the recent increase of external research facilities like universities has given small firms the possibility to engage in external R&D (Dosi (1988), Nelson (1990), Den Hertog and Thurik (1993) and Audretsch et al. (1996)). Engaging in external R&D has its own disadvantages, like threatening the exclusivity of the resulting new knowledge (Arrow (1962)) or small-numbers-bargaining hazards stemming from specialized R&D capabilities (Pisano (1990)). Pisano provides empirical evidence on biotechnology R&D projects that transaction cost problems may indeed motivate firms to internalize R&D.

These considerations show the importance of a ‘strategic management of technology’ for small firms (Acs and Audretsch (1990a)). This is especially the case when innovating firms cooperate. Inter-firm collaboration for innovation and new forms of coping with research cooperation have increased strongly over the years. Smith et al. (1991) see such joint ventures as one of the manifestations of wider changes in industrial organisation which have resulted from the structural crises of the 1970s and 1980s. Networks of small firms and specialized suppliers in ‘industrial districts’ have drawn the attention of many economists (Piore and Sable (1984), Porter (1990), Jacobs et al. (1990), Krugman (1991) and Loveman and Sengenberger (1991)). The delicate balance between competition and cooperation among the small firms is the key operating principle of industrial districts, according to You (1995), who also compares them with the Japanese subcontracting system.

The extent to which small and medium-sized firms benefit from technological changes depends on the type of technology, the organizational form of the firm and the speed of diffusion (Acs and Audretsch (1988), Jovanovic and Lach (1989), Poutsma et al. (1987) and Jovanovic and MacDonald (1994b, especially p. 35)). An interesting division of the role of small and large firms in the continuous advance of technology is suggested by Rothwell (1983,1984). His theory is one of ‘dynamic complementarity’ (see also Nootenboom (1987,1994)). Large and small firms are thought to complement and succeed each other in the innovation and diffusion process. Rothwell mentions the U.S. semiconductor industry as a classical example of dynamic complementarity: “Existing large firms provided much of the basic technology, venture capital and technically skilled personnel which were essential to new technology-based firm startups; the new technology-based firms provided the risk-taking entrepreneurial drive and rapid market exploitation. It was a synergistic relationship.” (p. 9-10).

(iii) integration of world markets

Globalisation is a challenge for many managers of large cooperations. Lower trade barriers, the economic boom in South-East Asia, the increase of OPEC countries’
income due to improved terms of trade (although partly offset by some local conflicts) and the collapse of communism have enlarged the 'world marketplace' for Western countries. The integration of the world economy may provide opportunities for exploiting scale. However, it is also a threat to large firms. Import competition has swept away entire industries, like the U.S. production of televisions, and has stripped other industries of their largest firms, like the Dutch ship-building and textiles industries. The U.S. automobile industry, decades long a tight oligopoly with very limited innovation, has had a narrow escape.\textsuperscript{10} Shepherd (1982) shows that foreign competitors were especially successful in traditional mass production industries, which often had an oligopolistic market structure. Large cooperations affected by import competition often lobby for protection, but this makes the final verdict probably only worse. See for example Lenway et al. (1996) on the damages of trade protection for the U.S. steel industry. They claim that it frustrates the Schumpeterian development of an industry.

The increased international competition and the volatility in exchange rates after the abolishment of fixed rates in the early 1970s have led to greater variability of sales. Selective investment in flexible technologies is a superior strategy to cope with this variability when compared to massive investment in greenfield plants. By adopting production technologies that permit them to quickly adapt to fluctuating demand, small firms may enter and compete successfully with larger enterprises which enjoy the advantage of static production efficiency (Mills and Schumann (1985) and Das et al. (1993)). Not only may small firms be more flexible than large ones, as assumed by Camacho (1991), but the group of small firms is flexible in itself (Nelson and Winter (1982)). The composition of the group may change quickly through entry and exit.

(iv) changes in the supply of labour

The supply and composition of labour has changed strongly since during the 1970s and 1980s. The labour participation of women and the level of education have increased drastically. During the early 1970s the entry of 'baby boomers' (born between 1945 and 1950) increased the supply of labour in many countries which were involved in World War II. The increase in the supply of labour has been a major factor in limiting the rise of real wage levels, which has been beneficial for the usually more labour-intensive smaller enterprises. Many economists have observed that wages paid by small companies are on average lower than those paid for similar work by large companies (Bosch (1990), Brown and Medoff (1989), Loveman and Sengenberger (1991).

\textsuperscript{10}Engelbrecht (1997) finds that the increased openness to international trade forced U.S. manufacturing firm to cut nonproduction employment. The organizational 'fat' build up in previous years adversely affected U.S. export competitiveness to an important extent during the 1980s.
Oosterbeek and Van Praag (1995) and Schmidt and Zimmerman (1991)). This may provide an explanation why 'sub-optimal' firms manage to survive even when there are economies of scale in the industry (Audretsch et al. (1995)). The increase in the average level of education over the last decades has enlarges the group of potential entrepreneurs which may introduce new products and new organizational forms.

There are two additional issues concerning the influence of the composition of the labour force on small business presence. First, many Western countries have become much more multicultural since the 1960s. This has led to many new firms providing specialized products to specific cultural groups. In the past some cultural groups, most notably the Jews, did not have many alternatives outside starting their own business. Shapero and Sokol (1982) argue that this 'refugee hypothesis' is still valid in modern day societies. In his discussion of the supply of entrepreneurship Alexander (1967) also argues that the potential role of marginal social groups should not be underestimated. He provides examples of the Chinese in South-East Asia and the Indians in Africa. Second, some economists argue that an increase in unemployment also increases the number of potential entrepreneurs and hence leads to more small businesses. Unemployment is seen as a 'push factor' to start an enterprise (see Storey (1991) for a survey). Evans and Leighton (1989a) find for example that "People who switch from wage work to self-employment tend to be people ... who experienced relatively frequent or long spells of unemployment as wage workers" (p. 532). This is consistent with the study of Highfield and Smiley (1987) on the influence of microeconomic and macroeconomic factors on the entry process and with Acs et al. (1991) who find for OECD countries that a 10 increase in the unemployment rate is associated with a 1.5 per cent increase in the self-employment rate. The strong increase in unemployment from the 1960s to the early 1980s may have enlarged the pool of potential entrepreneurs, although it strongly depends on the abilities of the jobless.

(v) increasing variety in consumer demand

The increasing incomes and wealth have enabled individuals to get higher in Maslow’s Hierarchy of Needs (Kotler (1986), p.179). Physiological and safety needs are satisfied for most people in Western countries. This allows individuals to consider the ‘higher’ social, esteem and self-actualization needs. As a result the demand for variety increases. Jackson (1984) shows in a utility maximization context how an increase in wealth leads to the consumption of a larger variety of products. Cross cultural influences have also enlarged the demand for variety. People become more and more aware of alternatives to their own culture in areas of food, sports, magazines, literature, music, furniture, etc. Acs et al. (1990) suggest that international competition has increased exposure to foreign products.
One way to model the increase in demand for variety is the use of a multidimensional brand map. An increase in the number of dimensions corresponds to increased demand for variety and will lead to an increase in the number of brands. A promising instrument for understanding the competitive positions of the various brands in a market is the ‘defender model’ (Hauser and Shugan (1983)). Estimation methods using supermarket scanning data to derive the multidimensional brand maps in the ‘defender model’ can be found in Shugan (1987) and Waarts et al. (1991). The estimation method presented in the last paper allows for the simultaneous estimation of the parameters of a preference distribution together with the brand coordinates in a multidimensional space.

Small firms are often the most obvious suppliers of new and specialized products. They do not profit from large scale economies but such cost disadvantages become less relevant as the price elasticities of specialized products are usually much lower. Extensive specialization also leads to mobility barriers which enable small firms to prosper for long periods of time (Bradburd and Ross (1989)). The decrease in diversification as reported by Jovanovic (1993) suggests that large firms are not capable of entering into such market niches.

The importance of the growth in variety for long-term economic growth is brought forward by e.g. Saviotti (1996,1997). He claims that not only the overall variety increases in the course of the economic development but that this increase is as well a necessary requirement for this development (Saviotti (1996), p. 96). Fragments of a theory of consumers' decisions which develop over time and with real income are given in Pasinetti (1993, p. 107-109). The main challenge of such a theory is to explain the association between changes in absolute level of total consumption and changes in its composition.

(vi) deregulation and privatization

Phillips (1985) finds that small firms have dominated in both the creation of new businesses and new jobs in deregulated industry sectors in the U.S. in the early 1980s. Deregulation in transportation and financial services have helped small business to gain an increasing share of total sales and total employment in these industries. This confirms some preliminary empirical evidence as provided by Shepherd (1982). Industry deregulation may have large positive benefits like new business formation, lower prices, new products and new jobs. Winston (1993) provides an impressive survey of the effects of deregulation for many U.S. industries and Koedijk and Kremers (1996) provide empirical evidence that deregulation in Europe has indeed promoted economic growth. Olley and Pakes (1996) show that "aggregate productivity increases sharply after each of the two periods in which the [U.S. telecommunications equipment] industry underwent
changes that decreased regulation" (p. 1265). See also OECD (1995a, p. 39-49) for an overview of the ongoing process of deregulation in many OECD-countries. In Australia sectors like electricity, telecommunications, port authorities, shipping, aviation and broadcasting were deregulated within a time span of only a couple of years. The same tendency to deregulate many industries can be found in countries like Finland, Italy and Sweden. Small firms are likely to benefit from this deregulation wave. A particularly interesting example of economic reform, including privatization and deregulation, is New Zealand. Evans et al. (1996) provide a detailed account of the economic reforms in that country from 1984 to 1995.

In the last decade thousands of state firms all over the world have gone private (Boycko et al. (1996)). A recent example is Telekom in Germany. One reason for privatization is the current opinion that running firms is not a task for the government and should be left to the private sector. Another important factor behind privatization is the poor performance of public companies. They are inefficient because they address the objectives of politicians, of which employment is a key one, instead of seeking maximum efficiency. Money losing coal mines in Western Europe, for example, were closed long after it became clear that they had no viable future. Politicians are in such cases eager to avoid massive job losses in the short run. The aim of privatization is to improve efficiency through restructuring firms and reallocating control rights over employment from politicians to managers. It will lead to a decrease of (excess) employment to be able to compete successfully in an unprotected environment. It can also take the form of breaking up the public enterprise into separate companies. In many cases privatization goes hand in hand with the rapid introduction of a competitive market to prevent monopoly situations. Privatization is therefore likely to affect the firm-size distribution in some specific industries. See Boycko et al. (1996) for an interesting model of privatization that explains the inefficiency of public companies and the increases in efficiency after they have gone private.

(vii) back to core business and outsourcing

Merger waves have been a major factor in the shaping of industrial structure of the U.S. and European countries. The first two merger waves were around 1900 and 1925 and the third merger wave took place in the late 1960s (De Jong (1988a, 1988b)). The third wave is often described as a conglomerate merger wave because many large firms expanded into unrelated markets through takeovers. The credo seem to be ‘the bigger the better’ and some conglomerates acquired dozens of firms in only a couple of years. However, the days of reckoning were swift to come. Much of the third merger wave was reversed in the 1970s and early 1980s (De Jong (1988c)). The conglomerate movement was not able to create economic value and the new management fashion became to
concentrate on your 'core competences' (Carlsson (1989a)). Firms that started to concentrate on one thing have indeed outperformed the (stock) market (The Economist, 21 Dec. 1996). Jovanovic (1993, p. 203) reports that the trend to more diversification reversed around 1977 and that the 1980s were characterized by corporate spin-offs and divestment. Aiginger and Tichy (1991) are probably correct when blaming much of the current 'back-to-basics' and downsizing (or rightsizing) fashion on the opportunistic conglomerate wave of the late 1960s.

There are basically three ways in which large companies may start to benefit from 'smallness' (Loveman and Sengenberger (1991)). First, large firms may decentralize production but still remain owner of the new plants. This implies that the firm size is not affected but that only the plant sizes decrease. Second, large firms may start to use a franchise-system. The ownership of the plants or shops is given up but there remains nevertheless a strong interconnection between the franchiser and franchisees. Third, large firms may sell parts of the company for example through a management buy-out. These new companies will usually remain supplier or client of the firm but this interconnection may become less strong over the years. An example of such an interconnection is the Japanese sub-contracting system in which a group of small firms supply to the large company (Caves (1989), Goto (1982) and Patrick and Rohlen (1987)). A related development is 'outsourcing'. Nowadays, many companies have non-core activities like cleaning, security and catering run by outside contractors. This is also the reason behind the strong increase in small consultancy companies providing specialist services to large companies.

2.5 Concluding remarks

Simon (1991) comes to the following conclusion in his review on organizations and markets: "organization size and degree of integration, and the boundaries between organizations and markets, are determined by rather subtle forces. The wide range of organizational arrangements observable in the world suggests that the equilibrium between these two alternatives may often be almost neutral, with the level highly contingent on a system's history" (p. 41-42). This ought to make us cautious when we imagine that we know the precise reasons why the size distribution shifts in one direction or the other. In this chapter we have discussed some theories about (optimal) size and size distributions. We concentrated on disadvantages of size and neglected potential remedies for these disadvantages (Radner (1992), Burgelman (1994)) and possible advantages of size which are stressed by transaction cost economists (Williamson (1975), Teece (1980), Hendrikse and Schreuder (1987)) or, for example, in the field of R&D (Cohen and Klepper (1996a) and Klepper (1996)) and with regard to important complementarities at the level of manufacturing, marketing and engineering
(Milgrom and Roberts (1990)). We do not suggest that these remedies and advantages are not relevant, but we concentrate on possible explanations for the increased share of small firms in several industries.

The theoretical considerations were followed by a section in which we presented some empirical evidence on the increasing importance of small firms, especially in manufacturing industries. The main part of this chapter was devoted to seven reasons why small business presence may have increased. It is obvious that the significance of these causes differs widely across industries. An important question is how these seven reasons may enable us to predict the future of small business. Hollander et al. (1967) were not so pessimistic about this future in the late 1960s: "At the very advanced stage of economic activity and technology in the United States and the rapid growth in prospect for both, there is no reason to doubt that there is a continuing sound base for viable small business." (p. 199). The prospects of small firms have even improved considerably since then. However, the significance of some causes like 'changes in the supply of labour' and 'deregulation and privatization' will probably decrease over the next decades. The 'baby boomers' will retire in the first 15 years of the next century and the number of firms suitable for privatization will become smaller. Regulations will converge to the EC standard in most European countries and will not adjust quickly after that. The 'back to core business and outsourcing' wave may dry up within the next decade, but opinions of business economists are divided (The Economist (Dec. 21 1996)). These considerations suggest that the speed of the increase of the share of small business will fall. In the Epilogue we discuss the main elements of a general theory of the development of small business presence in market economies. One of the consequences of this theory is that periods of increase of the share of small firms will be followed by periods of decrease in this share and vice versa.
longitudinal and experimental research regarding the effects of training and education on innovation and productivity. We have (Kletsow et al., 1995; Hattula et al., 1996) respectively the need to innovate, the need to train, and the need for education. However, the relationship between these needs and the actual implementation of training and education programs is still not fully understood. The need for training and education is influenced by various factors, such as the availability of resources, the level of competition, and the organizational culture. The implementation of training and education programs requires a systematic approach that takes into account the specific needs and circumstances of each organization. The benefits of training and education programs include increased productivity, improved employee satisfaction, and the development of new skills and knowledge. However, the effectiveness of these programs depends on various factors, such as the quality of the training materials, the level of commitment from the employees, and the support from the management.
CHAPTER 3

PROFITABILITY AND NUMBER OF FIRMS: THEIR DYNAMIC INTERACTION

High profitability is assumed to attract new firms, whereas a rise in the number of firms is thought to create a pressure on profitability. The purpose of this study is to investigate this dynamic interaction between profitability and number of firms in the Dutch retail sector. A two equation error correction model is developed and tested using a panel data set of 36 Dutch shoetypes covering the 1978-1988 period. The rationale of the model lies in Kirznerian ideas on the role of the entrepreneur as discoverer and exploiter of profit opportunities. Entrepreneurial activity appears to be high in Dutch retailing. Excess profits are predicted to vanish within a 5-year period and demand growth is followed by an increase in the number of stores within a one-year period.
3.1 Introduction

Dynamic market modelling has received broad attention in Industrial Economics (Nelson and Winter (1982), Eliasson (1984,1991), Geroski and Masson (1987), Carlsson (1989b), Schmalensee (1989), Klepper and Graddy (1990) and Mueller (1990)). This study is concerned with a central theme in this literature: the development of market structure and performance over time. In fact, we investigate the growth (or decline) of both profit levels and number of firms over time in the retail sector. We seek to provide answers on questions like: What is the effect of an increase in the number of firms on profitability? What is the effect of high or increasing profitability on the growth of the number of firms? What is the effect of demand growth on profitability and on the entry and exit flows in the industry? Does unemployment lead to an increase in the number of firms as unemployed are forced into entrepreneurship?

This study stands out from other studies in that it deals with the retail sector. There are several justifications for this choice. First, retailing has a strong contribution to the economy: for instance, in the Netherlands it accounts for about 23% of the total number of economically active enterprises and for about 13% of total labour force in the private sector in 1988 (Bode (1990)). There has however been scarce attention in Industrial Economics for non-manufacturing industries, let alone for the retail sector. Second, retailing is a comparatively simple entrepreneurial activity to which there are only limited entry and exit barriers. This implies that adjustment processes generated by entrepreneurs discovering and exploiting profit opportunities can be expected to have a direct effect in the sense that, for instance, the setting up or closing down of stores occurs within only a limited time period. Interesting evidence on the speed of the competitive and entrepreneurial process in the Dutch economy may be derived from such a low barrier industry. Third, retailers can be expected to react in a more predictable and more direct way on market incentives and disincentives than entrepreneurs in manufacturing industries. This is a consequence of a more clearly structured market environment in retailing, where elements of innovation, complex and time-consuming decision-making procedures and international competition generally play a more limited role than in manufacturing.

In an earlier study (Carree and Thurik (1994)), we used a similar dynamic market model to empirically investigate distinct competitive forces like actual competition among incumbents, new firms competition and potential competition, and various inflationary processes like cost, demand and wage inflation, in the retailing sector. This chapter provides additional information on the speed and empirical relevance of the ad-

11Clearly, the accessibility of some extensive panel data sets which have been merged for the purpose of the present analysis plays an important role in this choice.
justment processes in retailing. Our model is discussed in Section 3.2 and Section 3.3 is used for a description of the data set of 36 shoptypes for the period 1978-1988. Section 3.4 presents the empirical results and Section 3.5 is used for conclusions.

3.2 The dynamic interaction between profitability and number of firms

Entrepreneurial activity is a vital but elusive concept in theories of economic development and market dynamics (Hébert and Link (1989)). More than a century ago, Walras suggested that the definition of the entrepreneur is the key to our specific approaches to economic science (Walker (1986)). In recent years the importance of the entrepreneur as innovator, risk bearer and competitor is becoming more clear (Lumpkin and Dess (1996)). The importance of entrepreneurship has repeatedly been brought forward by prominent economists like Knight (1921), Schumpeter (1934), Kirzner (1973,1997) and Baumol (1968,1990). The entrepreneur is the key figure both for Schumpeterians and for Neo-Austrians. There is however a fundamental difference between the Schumpeterian notion of entrepreneurship and the Neo-Austrian (read Kirznerian) ideas on entrepreneurship. Schumpeter takes the entrepreneur to be the innovator who aims to achieve a profitable monopoly position, viz. the creator of disequilibria. For Kirzner, however, the role of the entrepreneur is to discover and exploit profit opportunities leading to the kind of adjustment processes necessary to move markets toward the equilibrium state (Kirzner (1973,1979)). In his view competition and entrepreneurship coincide because the competitive market process is basically entrepreneurial. Essential to this concept of competition in market-process theory is "the free entry of rivals, each in a incessant race to better the others" (Ikeda (1990), p.79). In this study we concentrate on entrepreneurial activity as a competitive force to adjust profits towards their long-run equilibrium levels. Our model of the dynamic interaction between profitability and number of firms in retailing is based on the Kirznerian notion of entrepreneurship. This implies that we have to specify profit opportunities (market incentives) and the ways of their exploitation. For Kirzner the efficiency of the market system depends crucially upon the degree of success with which entrepreneurs discover and exploit these profit opportunities.

Two fundamental indicators for profit opportunities in an industry are the prevailing level of profitability and the market growth. Profitable and growing markets generally provide more opportunities for entrepreneurs to act successfully than contracting markets with many firms experiencing losses. The exploitation of these profit opportunities can be achieved either by incumbents or by entrants into the market. Incumbents may try to enhance their competitive position by decreasing prices or increasing advertising efforts or level of service. Entrants may find market niches or may profit from the lack of ability of incumbents to adjust to developments in market demand.
Many of these entrants may be unemployed workers who use entrepreneurship as an escape out of unemployment (Storey (1991)). High unemployment may stimulate potential independents to enter and may discourage incumbents to exit.

The model which is presented in equations (3.1) and (3.2) is of an error correction mechanism type (Salmon (1982) and Gilbert (1986)). Shoptypes which provide entrepreneurs high profitability are assumed to be "in error", i.e. in disequilibrium. As the entrepreneurs in this study are mainly small independents for whom profits generally equal their income (before taxes), we define excess profit, i.e. the market "error", as the difference between this profit and gross modal wage.

\[(3.1) \quad \Delta \Pi_{it} = \Delta M_{it} - \alpha_0 + \alpha_1 (\Pi_{i,t-1} - M_{i,t-1}) + \alpha_2 (\Delta \Pi_{i,t-1} - \Delta M_{i,t-1}) + \alpha_3 \Delta NOF_{it} + \alpha_4 \Delta NOF_{i,t-1} + \alpha_5 \Delta CS_{it} + \alpha_6 \Delta CS_{it-1} + \alpha_7 (\Delta K_{it} - \Delta Q_{it}) + \epsilon_{1it}\]

\[(3.2) \quad \Delta NOF_{it} = \beta_0 + \beta_1 (\Pi_{i,t-1} - M_{i,t-1}) + \beta_2 (\Delta \Pi_{i,t-1} - \Delta M_{i,t-1}) + \beta_3 \Delta CS_{i,t-1} + \beta_4 \Delta CS_{i,t-2} + \beta_5 UN_{t-1} + \beta_6 \Delta UN_{t-1} + \beta_7 \Delta NOF_{i,t-1} + \epsilon_{2it}\]

In these equations \(\epsilon_{1it}\) and \(\epsilon_{2it}\) are possibly correlated random errors and the variables are logarithms of:

- \(\Pi_{it}\) : average profit in shoptype \(i\) in year \(t\) in 1980 prices
- \(M_{it}\) : gross modal wage in year \(t\) in 1980 prices
- \(NOF_{it}\) : number of firms in shoptype \(i\) in year \(t\)
- \(CS_{it}\) : consumer spending for products sold in shoptype \(i\) in year \(t\) in 1980 prices
- \(K_{it}\) : average value of total costs in shoptype \(i\) in year \(t\) in 1980 prices
- \(Q_{it}\) : average value of sales in shoptype \(i\) in year \(t\) in 1980 prices
- \(UN_{it}\) : number of unemployed in 1,000 persons in year \(t\)

The market "error" in our model, which presents an opportunity for entrepreneurial action, is the difference between profit and modal income. More precisely, in case \(\alpha_1\) is negative and \(\beta_1\) positive, we find all \(\Pi_{it}\) exceeding \(M_{it} - \alpha_0/\alpha_1\) in long-term equilibrium.

---

\(^{12}\)In Carree and Thurik (1994), the dependent variable of the profitability equation is \(\Delta \Pi_{it} - \Delta Q_{it}\). Influences of distinct competitive forces on retail pricing practice and the role of retailing in the inflationary process were the major topics in that study.
to have a downward effect on profitability, while all \( \Pi_i \) exceeding \( M_1 + \beta_0/\beta_1 - \beta_0 UN_1/\beta_1 \) in long-term equilibrium lead to a larger number of firms, which again squeezes profitability in case \( \alpha_3 \) is negative. We expect \( \alpha_0/\alpha_1 \) to be negative because independents not only supply labour but also are to be compensated for their managerial responsibilities and risk borne on their capital investment. The benchmark level of profit for net entry, \( M_1 + \beta_0/\beta_1 - \beta_0 UN_1/\beta_1 \), depends on the number of unemployed. High unemployment may stimulate potential independents to enter and demotivates incumbents to exit. A disequilibrium on the labour market may therefore generate entrepreneurial activity which exerts a downward pressure on disequilibria in the product markets, i.e. on excess profits.

The hypotheses concerning the signs of the parameters incorporated in this model are discussed in eight groups (\( H_1 \) through \( H_8 \)):

\[
H_1: \quad \alpha_1 < 0, \quad \alpha_2 > 0
\]

An excess profit situation is one of instability, which attracts entrepreneurial activity not only from within the shoptype but also from diversifying firms in other shoptypes or from vertically integrating wholesalers or manufacturers. This activity brings profits down: we expect \( \alpha_1 \) to be negative.\(^{13}\) We have no strong a priori reasons to expect the lagged endogenous variable to have any effect in equation (3.1). Nevertheless, we incorporate this variable to allow for sustained growth or decline of profits vis-à-vis modal wage. Such a structural shift may be a consequence of slowly evolving revaluation of the entrepreneurial activity of shopkeepers, for example by growing or declining managerial responsibilities. This could be a reason for a positive \( \alpha_2 \).

\[
H_2: \quad \alpha_3 < 0, \quad \alpha_4 > 0
\]

As the number of firms grows competitive forces are expected to increase. This implies that \( \alpha_3 \) is expected to be negative. Earlier experience with increases in the number of firms would however make a shoptype less sensitive to such new firms competition. A positive \( \alpha_4 \) would be consistent with this view. An extreme situation is the one in which \( \alpha_3 = -\alpha_4 \): changes in the growth of number of firms (\( \Delta\Delta NOF_{it} \)) have an effect on the development of profits vis-à-vis modal wage, not so much the growth of the number of

\(^{13}\)Empirical studies in the literature on the persistence of profits (Mueller (1990)) usually find this effect of excess profitability on the development of profitability to be negative. There are however different opinions about the extent of this effect. Geroski and Masson (1987) find a small effect and regard the competitive process to be extremely slow, while Levy (1987) reports a strong effect with adjustment to long-run equilibrium levels in only five years.
firms ($\Delta NOF_n$). However, we hypothesize that $\alpha_3 < -\alpha_4$ because we expect less recent changes in the industry structure to have more a limited effect than recent ones.

$H_3$: $\alpha_3 > 0 \quad \alpha_4 > 0$

Growing consumer spending on retail services is assumed to lift the pressure on increases of retailers' profits. A positive $\alpha_5$ is expected. We also incorporate $\Delta\Delta CS_n$ as indicator for unexpected demand shocks: temporary deviations from demand growth over time for specific retail services may facilitate profit increases even more.\(^{14}\) This indicates a positive value for $\alpha_6$.

$H_4$: $\alpha_7 < 0$

Rises in the cost-output ratio are likely to endanger a favourable development of profit as cost shifts are likely to be only partially passed on to customers (in the same period).\(^{15}\) We expect the parameter $\alpha_7$ to be negative. In the long run cost changes are entirely passed on in case of a negative $\alpha_1$ due to the elimination of disequilibrium by the error correction mechanism.

$H_5$: $\beta_1 > 0 \quad \beta_2 > 0$

High profits will encourage entry into a shop type and discourage exit.\(^{16}\) The parameter $\beta_1$ is expected to be positive. Growing profits vis-à-vis modal wage ($\Delta\Pi_{i,t-1} - \Delta M_{i,t-1}$) may also provide an incentive to enter a shop type. This implies a positive $\beta_2$.

$H_6$: $\beta_3 > 0 \quad \beta_4 > 0$

---

\(^{14}\) Liebowitz (1982) regards sales growth minus past sales growth as a more accurate predictor of changes in profit rates than sales growth.

\(^{15}\) Nootenboom et al. (1988) using 16 Dutch shop types for the period 1976-1983, find that deviations from normal costs are passed on to customers for about fifty per cent only.

\(^{16}\) De Wit and Van Winden (1990) use the difference between logarithms of profit in case of self-employment and wage in case of wage-employment to determine whether an individual chooses to become self-employed. They find a greater probability of becoming self-employed when this difference grows. The effect of excess profit on net entry at shop type level can be seen as an aggregation of individuals making decisions on entry and exit on the basis of their perceived discrepancies between attainable profit as shopkeeper and wage as employee. Geroski (1995) claims that a slow reaction of entry to high profits is a stylized result in the empirical literature on entry. He argues that this may be a consequence of entry showing much less 'between' industry variation than profitability does.
Growing consumer demand provides opportunities for new entrants. Viable sub-market niches may emerge or incumbents may simply lack the ability to expand quickly enough. As speed of perception and exploitation of these opportunities for entry is unknown both $\Delta CS_{u-1}$ and $\Delta CS_{u-2}$ are incorporated. We expect both a positive $\beta_3$ and a positive $\beta_4$.

$$H_7: \quad \beta_3 > 0 \quad \beta_4 > 0$$

High and growing unemployment discourages shopkeepers to exit as both variables are indicators for poor job opportunities. It also implies that there are a large number of potential entrepreneurs: unemployed may try to become self-employed in retailing to improve upon their financial situation. Evans and Leighton (1989a) present evidence that unemployed workers are more likely to enter self-employment than employees. Retailing may be an especially attractive sector for new business start-ups as financial and knowledge barriers are low. Both $\beta_5$ and $\beta_6$ are expected to be positive. The change in unemployment may very well be a more adequate measure of the tension on the labour market than unemployment, because of the sustained high unemployment level in the Netherlands. Those who are already unemployed for several years may be expected not to have the same amount of entrepreneurial ability and perseverance to become self-employed as workers who have become unemployed only recently. See Storey (1991) for a survey on the effect of unemployment on new business start-ups.

$$H_8: \quad \beta_7 > 0$$

Gort and Konakayama (1982) argue that perceptions of profit opportunities are positively related to the successful experience of others in the market. This implies that entrepreneurial activity may tend to be autocorrelated: successful entry may demonstrate the attractiveness of a shoptype to entrepreneurs. A positive $\beta_7$ is expected. One might argue that some new entrants already disappear after a one-year period due to inadequate preparations and too optimistic expectations. This could disturb interpretations on the parameter $\beta_7$ as representing the demonstration effect because a lower value for this parameter would be the result of such a correction.

### 3.3 Data

In this study a data set of 36 Dutch shoptypes for the period 1978-1988 is used. The source of the data is an ongoing panel of independent, mainly small Dutch retailers called *Bedrijfsignaleringssysteem* (interfirm comparison system) which is operated by the Research Institute for Small and Medium-Sized Business (EIM) in Zoetermeer. In
Carree and Thurik (1994) we provide an overview of the shoetypes and the observation periods for which data for these shootypes are available.\textsuperscript{17} We have a total of 341 data points while on average a data point is computed using observations from about seventy individual retail stores. The consumer spending on product packages and modal wage data are from the Central Bureau of Statistics (CBS) in Voorburg, unemployment data were retrieved from the UN Monthly Bulletin of Statistics and data on number of firms were gathered by the Central Registration Office (CRK) in The Hague. Summary statistics for the variables used in the model are shown in Table 3.1.

Table 3.1: Summary statistics of the variables used

<table>
<thead>
<tr>
<th>variable</th>
<th>MIN</th>
<th>MAX</th>
<th>MEAN</th>
<th>STD</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\Delta \Pi_{t,t-1} - \Delta M_{I,t}$</td>
<td>-1.13</td>
<td>0.99</td>
<td>0.037</td>
<td>0.262</td>
</tr>
<tr>
<td>$\Pi_{t,t-1} - M_{I,t-1}$</td>
<td>-1.29</td>
<td>1.09</td>
<td>0.117</td>
<td>0.415</td>
</tr>
<tr>
<td>$\Delta \Pi_{t,t-1} - \Delta M_{I,t-1}$</td>
<td>-1.13</td>
<td>0.99</td>
<td>0.039</td>
<td>0.276</td>
</tr>
<tr>
<td>$\Delta NOF_{I,t}$</td>
<td>-0.11</td>
<td>0.10</td>
<td>-0.001</td>
<td>0.026</td>
</tr>
<tr>
<td>$\Delta NOF_{I,t-1}$</td>
<td>-0.11</td>
<td>0.08</td>
<td>-0.004</td>
<td>0.025</td>
</tr>
<tr>
<td>$\Delta CS_{I,t}$</td>
<td>-0.22</td>
<td>0.21</td>
<td>0.002</td>
<td>0.058</td>
</tr>
<tr>
<td>$\Delta CS_{I,t-1}$</td>
<td>-0.29</td>
<td>0.19</td>
<td>0.006</td>
<td>0.062</td>
</tr>
<tr>
<td>$\Delta K_{I,t} - \Delta Q_{I,t}$</td>
<td>-0.19</td>
<td>0.17</td>
<td>-0.005</td>
<td>0.054</td>
</tr>
<tr>
<td>$\Delta CS_{I,t-1}$</td>
<td>-0.22</td>
<td>0.24</td>
<td>-0.004</td>
<td>0.064</td>
</tr>
<tr>
<td>$\Delta CS_{I,t-2}$</td>
<td>-0.22</td>
<td>0.24</td>
<td>-0.007</td>
<td>0.062</td>
</tr>
<tr>
<td>$\Delta UN_{I,t-1}$</td>
<td>-0.08</td>
<td>0.44</td>
<td>0.127</td>
<td>0.202</td>
</tr>
</tbody>
</table>

Note: MIN, MAX, MEAN and STD stand for minimum, maximum, mean and standard deviation.

3.4 Empirical results

Equations (3.1) and (3.2) are estimated using Three Stages Least Squares (SAS-module SYSLIN). We provide estimation results both for 'no fixed effects' ($\alpha_{0}$ and $\beta_{0}$ are equal for all shoetypes) and 'fixed effects' ($\alpha_{0}$ and $\beta_{0}$ are allowed to differ across shoetypes). Using fixed effects we allow the long-term equilibrium relation between

\textsuperscript{17}Data are available for 14 shoetypes during the 1978-88 period, for 19 shootypes during the 1982-88 period, for two shoetypes (electrical appliances, more than 25% repairs and electrical appliances, mixed assortment) during the 1982-87 period and for one shootype (supermarkets without butcher's shop) during the 1978-83 period.
profits and modal wage to differ across shoptypes. It is likely that profits in shoptypes characterized by high capital requirements, high risks and high managerial responsibilities will tend to have higher long-term equilibrium values than in shoptypes in which these are low (Carree and Thurik (1994)). Estimation results are presented in Table 3.2.

We find confirmation for profit adjustment to excess profits ($\alpha_1 < 0$). The adjustment rate (the extent of the effect of $\Pi_{t-1} - M_l_{t-1}$ on $\Delta \Pi_t - \Delta ML_t$) is rather slow (about 10% in a one-year period) in case shoptype-specific effects are excluded. This would imply an almost complete profit adjustment in about 30 years. However, the adjustment rate increases substantially to about 50% in a one-year period at their inclusion. This implies an almost complete profit adjustment in about 5 years consistent with the empirical evidence found by Levy (1987) for U.S. manufacturing industries. It is most interesting to find the speed of adjustment to be roughly the same for different sectors (retailing versus manufacturing) in different countries (the Netherlands versus the United States). The parameter estimate of $\alpha_2$ is not significant (at a 5%-significance level). There is no indication of change in profits vis-à-vis modal wage to have any effect on this change in the next period.

Considering the dynamic interaction between profitability and number of firms results are that $\hat{\alpha}_3 < 0$, $\hat{\alpha}_4 > 0$ and $\hat{\alpha}_3 < -\hat{\alpha}_4$, as expected. However, none of these results are statistically significant. Entry and exit may have only a limited direct effect on industry profitability because they are most widespread in the group of relatively small firms who make up the competitive fringe (MacDonald (1986)). Increases of retailers' profits appear to be facilitated by demand growth ($\alpha_5 > 0$) but not by changes in this variable, $\Delta \Delta CS_{it}$. A favourable development of their profits is however endangered by rises in the cost-output ratio ($\alpha_7 < 0$). Presumably, this results from firms' market positions preventing cost shifts to be entirely passed on to customers within a one-year period or from limited awareness of cost changes during such a period.

Excess profits and growing consumer demand appear to be key incentives to new firm formation in retailing because both the parameter estimates of $\beta_1$ and $\beta_3$ are positive and significant. This is consistent with empirical evidence in most studies on entry in manufacturing industries (Acs and Audretsch (1989b)). The effect of excess profit ($\Pi_{t-1} - ML_{t-1}$) rises strongly at the inclusion of fixed effects. This is not surprising because entrepreneurs will confront expected profits with capital requirements, risks and managerial responsibilities when considering entry (and exit). Entrepreneurial reaction (in setting up or closing down stores) to changes in consumer demand seems to be quick: $\Delta CS_{it-2}$ has no significant effect on $\Delta NOF_{it}$. The parameter estimate of $\beta_2$ is also not statistically significant indicating that entrepreneurs are not too impressed by short-term profit changes.
<table>
<thead>
<tr>
<th>param</th>
<th>variable</th>
<th>no FE</th>
<th>FE</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\alpha_0$</td>
<td>constant</td>
<td>0.053</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(3.2)</td>
<td></td>
</tr>
<tr>
<td>$\alpha_1$</td>
<td>$\Pi_{i,t-1} - M_{i,t-1}$</td>
<td>-0.107</td>
<td>-0.492</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(2.4)</td>
<td>(4.3)</td>
</tr>
<tr>
<td>$\alpha_2$</td>
<td>$\Delta \Pi_{i,t-1} - \Delta M_{i,t-1}$</td>
<td>-0.020</td>
<td>0.081</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.3)</td>
<td>(1.2)</td>
</tr>
<tr>
<td>$\alpha_3$</td>
<td>$\Delta NOF_{it}$</td>
<td>-8.208</td>
<td>-4.055</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(1.7)</td>
<td>(0.8)</td>
</tr>
<tr>
<td>$\alpha_4$</td>
<td>$\Delta NOF_{it-1}$</td>
<td>5.527</td>
<td>2.150</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(1.6)</td>
<td>(1.0)</td>
</tr>
<tr>
<td>$\alpha_5$</td>
<td>$\Delta CS_{it}$</td>
<td>1.049</td>
<td>1.468</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(1.9)</td>
<td>(3.1)</td>
</tr>
<tr>
<td>$\alpha_6$</td>
<td>$\Delta \Delta CS_{it}$</td>
<td>-0.661</td>
<td>-0.671</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(1.1)</td>
<td>(1.2)</td>
</tr>
<tr>
<td>$\alpha_7$</td>
<td>$\Delta K_{it} - \Delta Q_{it}$</td>
<td>-1.993</td>
<td>-1.530</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(7.7)</td>
<td>(5.9)</td>
</tr>
<tr>
<td>$\beta_0$</td>
<td>constant</td>
<td>-0.015</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(1.1)</td>
<td></td>
</tr>
<tr>
<td>$\beta_1$</td>
<td>$\Pi_{i,t-1} - M_{i,t-1}$</td>
<td>0.0053</td>
<td>0.0198</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(1.9)</td>
<td>(3.5)</td>
</tr>
<tr>
<td>$\beta_2$</td>
<td>$\Delta \Pi_{i,t-1} - \Delta M_{i,t-1}$</td>
<td>0.0023</td>
<td>-0.0076</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.5)</td>
<td>(1.6)</td>
</tr>
<tr>
<td>$\beta_3$</td>
<td>$\Delta CS_{it-1}$</td>
<td>0.129</td>
<td>0.120</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(6.0)</td>
<td>(5.4)</td>
</tr>
<tr>
<td>$\beta_4$</td>
<td>$\Delta CS_{it-2}$</td>
<td>0.002</td>
<td>0.008</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.1)</td>
<td>(0.4)</td>
</tr>
<tr>
<td>$\beta_5$</td>
<td>$UN_{i,t-1}$</td>
<td>0.0022</td>
<td>-0.0017</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(1.0)</td>
<td>(0.7)</td>
</tr>
<tr>
<td>$\beta_6$</td>
<td>$\Delta UN_{i,t-1}$</td>
<td>0.0174</td>
<td>0.0178</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(2.6)</td>
<td>(2.6)</td>
</tr>
<tr>
<td>$\beta_7$</td>
<td>$\Delta NOF_{i,t-1}$</td>
<td>0.674</td>
<td>0.387</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(14.7)</td>
<td>(6.3)</td>
</tr>
</tbody>
</table>

System-R² | 0.439 | 0.545 |

Note: Estimations results are for no fixed effects (no FE) and for fixed effects (FE). Numbers in parentheses are t-statistics.
The level of unemployment \(\left(UN_{t-1}\right)\) appears to have no effect on changes in number of stores. However, an increase in the number of unemployed does seem to be followed by an increase in the number of retail firms (the parameter estimate of \(\beta_6\) is positive and significant). One is tempted to conclude that people who have become unemployed only recently have a higher propensity of using self-employment (in the retail sector) as an alternative to being unemployed than people who have had no working experience for a longer period. The parameter estimate of \(\beta_5\) is positive and strongly significant. The demonstration effect appears to have empirical relevance in the retail sector.\(^{18}\)

3.5 Conclusions

Measuring entrepreneurial activity is a daring venture as many different theories on the nature of entrepreneurship exist (Hébert and Link (1989)). In this study we concentrate on the Kirznerian notion of entrepreneurial activity as a competitive force to adjust profits to their long-term equilibrium levels. An error correction model of the dynamic interaction between profitability and number of firms in retailing is developed and tested on a panel data set of 36 Dutch shoptypes for the period 1978-1988.

One would expect entrepreneurial activity to be high in the retail sector because financial entry and exit barriers are low and because starting and running a store is a comparatively simple entrepreneurial activity. Some evidence for this is found in the effects of excess profits and demand growth. The fact that excess profits vanish within a 5-year period demonstrates that excess profits arouse fierce competitive action. Excess profits also stimulate entry of new stores (entrepreneurs). Demand growth is followed by an increase in the number of stores already within a one-year period. (Recently) unemployed workers appear to have opportunities of becoming self-employed in the retail sector as increases in unemployment have a positive effect on the change in the number of stores.

Profitability and number of firms are two important market phenomena which deserve detailed attention. There are however more elements of market structure and performance, like growth and survival rates of firms and adoption of organizational or technical innovations. The current chapter can only be considered as a first step in understanding the causes and effects of market processes. In the next chapter we will extend our analysis to gross entry and exit rates. Special attention is given to entry and exit barriers and displacement and replacement effects in the entry and exit processes.

\(^{18}\)Robertson and Symons (1992) argue that biases arise when parameters vary across panels, shoptypes, while this is not allowed for in estimation. In the case of a few time periods and many panels they find that dynamic properties tend to be over-estimated, i.e. the coefficient of a lagged dependent variable is overstated. The decline of our parameter estimate of \(\beta_5\) from 0.674 to 0.387 when fixed effects are included is consistent with the Monte Carlo results presented in Figure 4 of their paper.
The level of myeloma immunoglobulin (IgM) in the serum of some patients with multiple myeloma is elevated, and this elevation may correlate with the patient's clinical course. Treatment with melphalan and prednisone has been shown to reduce the level of IgM and improve the patient's condition. However, the mechanism of action of this treatment is not fully understood.

In recent studies, it has been suggested that an increase in the number of plasma cells in the bone marrow may contribute to the elevation of IgM levels. This hypothesis is supported by the observation that patients with a high level of IgM often have a more aggressive form of the disease. Further research is needed to confirm these findings and to develop more effective treatment strategies.
CHAPTER 4

ENTRY AND EXIT IN RETAILING: DISPLACEMENT AND REPLACEMENT

In this chapter the determinants of entry and exit and the interrelationship between these market phenomena are investigated. We examine incentives, barriers, displacement and replacement for a panel data set of 23 Dutch shoptypes for the 1981-1988 period. Results indicate that profit as a ratio of modal income, growth of consumer spending and growing unemployment are important incentives to enter and disincentives to exit. Requirements of floorspace and specific requirements of professional skills appear to reduce entry rates. We find evidence for entry and exit to interact but not to be simultaneously determined. The implication being that entry (exit) has a separate influence on exit (entry) next to market incentives and entry and exit barriers. Measures of market dynamics suggest that shoptypes with specific legal requirements to start a retail venture are less 'dynamic' than shoptypes with only limited requirements.
4.1 Introduction

The selection process in which markets choose between established and entrant firms has a vital contribution to the transformation of industries over time. Industries with low birth and death rates are likely to be more vulnerable to an inadequate allocation of resources, limited innovativeness, and some form of formal or tacit collusion (Geroski and Jacquemin (1985)). The continuous flows of entry and exit represent a changing pool of potentially strong competitors. Beesley and Hamilton (1984) describe them as the seedbed of new activities from which will emerge the successful businesses and industries of the future. High barriers to entry and exit may therefore be serious impediments to dynamic market efficiency.

Entry and exit rates of firms differ strongly across industries and over time (Dunne et al. (1988), Van Herck (1984)). In empirical studies this variation in the intensity of the selection process by which incumbents are displaced by new entrants is explained by variation in profitability and market growth and by variation in the height of entry and exit barriers. A wide variety of entry barriers are supposed and found to influence entry decisions. A survey of a total of nineteen different market entry barriers can be found in Karakaya and Stahl (1989). Most studies incorporate the effects of profitability and market growth on rates of entry and exit, but the empirical evidence is mixed (see Section 4.2).

Entry and exit are interrelated market phenomena. Incentives to enter, like high profitability and strong market growth, are also disincentives to exit. Many entry barriers can also be considered exit barriers and vice versa (Eaton and Lipsey (1980), Shapiro and Khemani (1987)). Entry and exit may however not only be influenced by the same overall market conditions, but could also respond directly to each other (Rosenbaum and Lamort (1992)). Entering firms may replace and/or displace exiting firms. Replacement occurs when exit causes entry. Exit may enhance opportunities for (potential) entrepreneurs to enter. Displacement occurs when entry causes exit. Entry may adversely affect survival chances of incumbent firms. Turnover of firms may be strongly influenced by the extent of these replacement and displacement effects.

The aim of this chapter is to extend and supplement empirical evidence on the processes of entry and exit as build up since Mansfield (1962). There are three contributions in this study to the investigation of entry and exit. First, it examines the causal interrelationship of replacement and displacement using an extensive and rich panel data set at a low level of aggregation. Second, it is one of the first to explain determinants of entry and exit in a non-manufacturing industry, viz. retailing. The retail sector has an important contribution to the economy. It accounts for about 13% of total labour force in the Dutch private sector in 1988 (Bode (1990)). The total number of economically active enterprises in retailing (92,000) in the Netherlands is twice of that
in manufacturing (46,700) (EIM (1991)). Third, it stands out in that it concentrates on a comparatively simple entrepreneurial activity. The setting up or closing down of a venture in retailing generally demands much less capital, time and knowledge investment than in manufacturing. This is due to the more limited size of retail ventures and the more clearly structured retail market environment resulting from the limited impact of innovations, complex network-dependencies and international competition. Hence, retailing can be regarded as a promising testing area for establishing the influence of market incentives on entry and exit rates.

In Section 4.2 we examine patterns of firm entry and exit in Dutch retailing and discuss previous findings in the empirical entry and exit literature. We provide a survey of the effects of profitability and market growth on entry and exit as found in 44 studies. Section 4.3 is used for a discussion of the determinants of entry and exit that are used in this study. In Section 4.4 we pay special attention to the problems related to the estimation of displacement and replacement effects. In Section 4.5 we present the data and empirical results. The average replacement and displacement periods and the persistence of profits for two groups of shoptypes are derived in Section 4.6. The first group consists of shoptypes with specific legal requirements to start a retail venture while the second group of shoptypes has only general requirements. Section 4.7 is used for concluding remarks.

4.2 Patterns of entry, exit and their determinants

The average annual entry and exit rates for Dutch shoptypes (retail industries) for the 1981-1988 period can be found in Table 4.1. We compare these figures with those presented by Dunne et al. (1988) for U.S. manufacturing for the 1963-1982 period. Patterns and determinants of entry and exit in this sector have been studied intensively in the past decades. The average annual entry and exit rate are 0.083 and 0.091 in Dutch retailing, while they are 0.081 and 0.074 in U.S. manufacturing. The lowest and highest average entry rates in Dutch shoptypes are 0.035 for tobacco shops and 0.154 for florists. The corresponding figures for U.S. manufacturing industries (at two-digit

---

19 Employment shares of the retail trade in the private sector do not differ much across countries in Western Europe. The total number of employees in the retail sector in 1988 in France, Germany, the Netherlands and the United Kingdom were (percentages of total employment in the private sector between brackets): 1,639,000 (12.1), 2,254,000 (11.5), 527,700 (12.6) and 2,615,000 (12.8), respectively (Vogelesang et al. (1992)).

20 Throughout this study entry rate (ENTRY) is defined as the total number of firms entering from period t-1 to t divided by the total number of firms in period t-1. The exit rate (EXIT) is the total number of firms leaving from period t-1 to t divided by the total number of firms in period t-1.

21 The figures for Dutch retailing relate to the 23 shoptypes discussed in the present chapter.
level) are 0.043 for tobacco (SIC 21) and 0.121 for instruments (SIC 38). The lowest and highest average exit rates in Dutch retailing are 0.052 for bicycles and 0.138 for furnishing stores with mixed assortment. The lowest average exit in U.S. manufacturing is again found in tobacco (SIC 21) at 0.047 and highest average exit in apparel (SIC 23) at 0.095.

Table 4.1: Entry, exit, profitability and demand growth in shoptypes

<table>
<thead>
<tr>
<th>Shoptype</th>
<th>ENTRY</th>
<th>EXIT</th>
<th>PMI₁</th>
<th>DCS₁</th>
</tr>
</thead>
<tbody>
<tr>
<td>supermarkets with butcher's shop</td>
<td>0.079</td>
<td>0.098</td>
<td>1.63</td>
<td>0.006</td>
</tr>
<tr>
<td>butchers</td>
<td>0.071</td>
<td>0.079</td>
<td>1.09</td>
<td>-0.000</td>
</tr>
<tr>
<td>greengrocers</td>
<td>0.083</td>
<td>0.094</td>
<td>0.88</td>
<td>0.034</td>
</tr>
<tr>
<td>dairy shops</td>
<td>0.051</td>
<td>0.081</td>
<td>0.75</td>
<td>0.014</td>
</tr>
<tr>
<td>fish shops</td>
<td>0.113</td>
<td>0.107</td>
<td>0.85</td>
<td>0.005</td>
</tr>
<tr>
<td>bakers</td>
<td>0.060</td>
<td>0.062</td>
<td>1.60</td>
<td>0.011</td>
</tr>
<tr>
<td>confectioners</td>
<td>0.092</td>
<td>0.107</td>
<td>1.27</td>
<td>0.014</td>
</tr>
<tr>
<td>tobacco shops</td>
<td>0.035</td>
<td>0.070</td>
<td>0.84</td>
<td>-0.035</td>
</tr>
<tr>
<td>liquor stores</td>
<td>0.085</td>
<td>0.111</td>
<td>0.54</td>
<td>-0.010</td>
</tr>
<tr>
<td>textiles mens wear</td>
<td>0.140</td>
<td>0.125</td>
<td>1.82</td>
<td>0.002</td>
</tr>
<tr>
<td>shoes</td>
<td>0.071</td>
<td>0.073</td>
<td>1.26</td>
<td>0.019</td>
</tr>
<tr>
<td>household goods</td>
<td>0.087</td>
<td>0.089</td>
<td>0.85</td>
<td>0.004</td>
</tr>
<tr>
<td>furnishing stores, mainly furniture</td>
<td>0.100</td>
<td>0.105</td>
<td>1.92</td>
<td>-0.015</td>
</tr>
<tr>
<td>furnishing stores, mainly clothing material</td>
<td>0.140</td>
<td>0.125</td>
<td>1.41</td>
<td>-0.014</td>
</tr>
<tr>
<td>furnishing stores, mixed assortment</td>
<td>0.109</td>
<td>0.138</td>
<td>1.91</td>
<td>-0.015</td>
</tr>
<tr>
<td>paint, glass, wall-paper</td>
<td>0.049</td>
<td>0.082</td>
<td>1.09</td>
<td>0.006</td>
</tr>
<tr>
<td>ironmongers/do-it-yourself shops</td>
<td>0.052</td>
<td>0.066</td>
<td>1.02</td>
<td>0.006</td>
</tr>
<tr>
<td>bicycles</td>
<td>0.041</td>
<td>0.052</td>
<td>0.88</td>
<td>-0.020</td>
</tr>
<tr>
<td>photographer's shops</td>
<td>0.082</td>
<td>0.079</td>
<td>1.27</td>
<td>0.062</td>
</tr>
<tr>
<td>watches, gold, silver</td>
<td>0.080</td>
<td>0.083</td>
<td>1.30</td>
<td>-0.018</td>
</tr>
<tr>
<td>druggists</td>
<td>0.053</td>
<td>0.060</td>
<td>1.17</td>
<td>0.051</td>
</tr>
<tr>
<td>florists</td>
<td>0.154</td>
<td>0.132</td>
<td>0.67</td>
<td>0.026</td>
</tr>
<tr>
<td>pet shops</td>
<td>0.089</td>
<td>0.086</td>
<td>0.61</td>
<td>0.006</td>
</tr>
<tr>
<td>mean</td>
<td>0.083</td>
<td>0.091</td>
<td>1.16</td>
<td>0.006</td>
</tr>
</tbody>
</table>

Note: The variables ENTRY, EXIT, PMI and DCS stand for the shoptype's entry rate, exit rate, average profit (divided by modal income) and the growth of consumer spending, respectively. Figures are averages over the 1981-1988 period.

Entry and exit rates in Dutch retailing have a correlation coefficient of 0.78 (over one year periods, see Table 4.2). Dunne, Roberts and Samuelson also found entry and exit to be highly correlated. They found a coefficient of 0.27 over five year periods. One may
expect correlation between average entry and exit rates to be somewhat higher in retailing than in manufacturing. This is a consequence of the lower probability of survival of new small firms in retailing (38.4 per cent for a ten year-period, U.S. data) versus those in manufacturing (46.9 per cent) (Phillips and Kirchhoff (1989)).

Table 4.2: Correlations of entry, exit, profitability and demand growth

<table>
<thead>
<tr>
<th></th>
<th>ENTRY</th>
<th>EXIT</th>
<th>PMI₁</th>
<th>PRI₁</th>
<th>DCS₁</th>
</tr>
</thead>
<tbody>
<tr>
<td>ENTRY</td>
<td>0.617*</td>
<td>0.783*</td>
<td>0.219*</td>
<td>-0.119</td>
<td>0.070</td>
</tr>
<tr>
<td>EXIT</td>
<td></td>
<td></td>
<td>0.122</td>
<td>-0.072</td>
<td>-0.133</td>
</tr>
<tr>
<td>PMI₁</td>
<td>0.180*</td>
<td>-0.148*</td>
<td></td>
<td>-0.231*</td>
<td>0.182*</td>
</tr>
<tr>
<td>PRI₁</td>
<td>0.039</td>
<td>-0.012</td>
<td>0.185*</td>
<td></td>
<td>0.154*</td>
</tr>
<tr>
<td>DCS₁</td>
<td>0.108</td>
<td>-0.155*</td>
<td>0.422*</td>
<td></td>
<td>0.144</td>
</tr>
</tbody>
</table>

Note: Correlations to the left of the diagonal are adjusted for industry effects. The per shotype mean of the variable is subtracted before the correlations are estimated. Correlations to the right of the diagonal are not adjusted for industry effects. Estimates which are significant at a 5%-significance level are denoted by *. The variable PRI stands for average profit (divided by equity). See also note to Table 4.1.

Empirical studies of the determinants of entry and exit have received growing attention since Orr’s influential paper (Orr (1974)). Recent summaries of the empirical evidence can be found in Siegfried and Evans (1994) and Geroski (1995). Table 4.3 presents a total of 44 studies with characteristics of the data set used (country, number of industries, period, measure of entry or exit). All these studies are strictly confined to manufacturing industries, with the exception of Highfield and Smiley (1987), where a small proportion of non-manufacturing industries is incorporated into the data set. There has been considerable progress in the empirical entry and exit literature since Mansfield’s 1962 article. First, data for several countries and for highly disaggregated industries have been used. Some of the more recent studies also use panel data. The use of panel data is of great importance in examining a process as highly dynamic as entry and exit (Schmalensee (1989)). Second, different types of entry and exit have been examined. One may expect that incentives and barriers differ between diversifying and specialist entry, between domestic and foreign entry and between entry by small and by large firms. Third, the empirical relevance of a large number of different entry and exit barriers has been investigated. Fourth, displacement and replacement effects tend to be incorporated in recent studies. See Geroski (1991b, 1995) for a more general exposition of the causes and effects of entry and exit.

---

22This has also led to important progress in the theoretical line of research (Hopenhayn (1992) and Jovanovic and Lach (1989)).
<table>
<thead>
<tr>
<th>Study</th>
<th>Country</th>
<th>Ind</th>
<th>Period</th>
<th>Mea</th>
<th>ProfGrow</th>
<th>Remarks</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mansfield (1962)</td>
<td>USA</td>
<td>4</td>
<td>1916-59</td>
<td>GRE</td>
<td>++</td>
<td>n</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td><strong>GRX - n</strong></td>
</tr>
<tr>
<td>Marcus (1967)</td>
<td>USA</td>
<td>17</td>
<td>1951-55</td>
<td>GRX</td>
<td>-</td>
<td>n</td>
</tr>
<tr>
<td>Orr (1974)</td>
<td>Canada</td>
<td>71</td>
<td>1964-67</td>
<td>NNE</td>
<td>0</td>
<td>0/+</td>
</tr>
<tr>
<td>Duetsch (1975)</td>
<td>USA</td>
<td>307</td>
<td>1958-67</td>
<td>NRE</td>
<td>0/+</td>
<td>++</td>
</tr>
<tr>
<td>Gorecki (1975)</td>
<td>UK</td>
<td>44</td>
<td>1958-63</td>
<td>NRE</td>
<td>n</td>
<td>++ Diversifying entry</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>**NRE n ++ Specialist entry</td>
</tr>
<tr>
<td>Gorecki (1976)</td>
<td>Canada</td>
<td>62</td>
<td>1964-67</td>
<td>NNE</td>
<td>0</td>
<td>0 Domest ic entry</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>**NNE n ++ Foreign entry</td>
</tr>
<tr>
<td>Hirschey (1981)</td>
<td>USA</td>
<td>87</td>
<td>1947-72</td>
<td>NRE</td>
<td>0/-</td>
<td>++</td>
</tr>
<tr>
<td>Masson and Shaanan (1982)</td>
<td>USA</td>
<td>37</td>
<td>1958-63</td>
<td>GSE</td>
<td>+</td>
<td>0</td>
</tr>
<tr>
<td>Creedy and Johnson (1983)</td>
<td>UK</td>
<td>16</td>
<td>1966-77</td>
<td>GRE</td>
<td>+</td>
<td>n</td>
</tr>
<tr>
<td>Shapiro (1983)</td>
<td>Canada</td>
<td>131</td>
<td>1972-76</td>
<td>GNE</td>
<td>+</td>
<td>+ Domestic entry</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>**GNE + 0 Foreign entry</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>**GNX 0/+ 0 Domestic exit</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>**GNX 0 0 Foreign exit</td>
</tr>
<tr>
<td>Duetsch (1984)</td>
<td>USA</td>
<td>95</td>
<td>1963-72</td>
<td>NNE</td>
<td>+</td>
<td>++</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>**GRE n +</td>
</tr>
<tr>
<td>Hilke (1984)</td>
<td>USA</td>
<td>16</td>
<td>1950-66</td>
<td>GSE</td>
<td>0</td>
<td>++</td>
</tr>
<tr>
<td>Hamilton (1985)</td>
<td>Scotland</td>
<td>86</td>
<td>1976-80</td>
<td>GRE</td>
<td>0</td>
<td>+ Independent entry</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>**GRE 0 + Dependent entry</td>
</tr>
<tr>
<td>Kessides (1986)</td>
<td>USA</td>
<td>262</td>
<td>1972-77</td>
<td>NNE</td>
<td>*</td>
<td>++</td>
</tr>
<tr>
<td>Khemani and Shapiro (1986)</td>
<td>Canada</td>
<td>143</td>
<td>1972-76</td>
<td>GNE</td>
<td>++</td>
<td>++</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>**GRE 0 +</td>
</tr>
<tr>
<td>MacDonald (1986)</td>
<td>USA</td>
<td>46</td>
<td>1976-82</td>
<td>GEE</td>
<td>0</td>
<td>++ Specialist entry</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>**GEX n 0 Specialist exit</td>
</tr>
<tr>
<td>Masson and Shaanan (1986)</td>
<td>USA</td>
<td>26</td>
<td>1958-63</td>
<td>GSE</td>
<td>+</td>
<td>++</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>**GNE 0 0 Foreign exit</td>
</tr>
<tr>
<td>Highfield and Smiley (1987)</td>
<td>USA</td>
<td>60</td>
<td>1976-81</td>
<td>GRE</td>
<td>++</td>
<td>++</td>
</tr>
<tr>
<td>Lieberman (1987)</td>
<td>USA</td>
<td>38</td>
<td>1952-82</td>
<td>GDE</td>
<td>n</td>
<td>++</td>
</tr>
<tr>
<td>Masson and Shaanan (1987)</td>
<td>USA</td>
<td>43</td>
<td>1964-67</td>
<td>NNE</td>
<td>0/+</td>
<td>0</td>
</tr>
<tr>
<td>Schwalbach (1987)</td>
<td>Germany</td>
<td>122</td>
<td>1977-82</td>
<td>GSE</td>
<td>++</td>
<td>++ Diversifying entry</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>**GNX 0 0 Displacement</td>
</tr>
</tbody>
</table>
| Khemani and Shapiro (1988)    | Canada  | 143 | 1972-76| GNE | ++       | ++ Special  
|                               |         |     |        |     |          | entry + Diversifying entry       |
|                               |         |     |        |     |          |                                  |
| Acs and Audretsch (1989b)     | USA     | 247 | 1978-80| NRE | 0/+      | ++ Small entry                   |
| Acs and Audretsch (1989c)     | USA     | 238 | 1976-82| GRE | 0        | ++                               |

continued on next page
<table>
<thead>
<tr>
<th>Reference</th>
<th>Country</th>
<th>Year Range</th>
<th>Measure</th>
<th>Scale</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chappell et al. (1990)</td>
<td>USA</td>
<td>1972-77</td>
<td>NNE</td>
<td>0/++</td>
<td>++</td>
</tr>
<tr>
<td>Jeong and Masson (1990)</td>
<td>Korea</td>
<td>1976-81</td>
<td>NRE</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Kessides (1990)</td>
<td>USA</td>
<td>1972-82</td>
<td>NNE</td>
<td>*</td>
<td>++</td>
</tr>
<tr>
<td>Dunne and Roberts (1991)</td>
<td>USA</td>
<td>1963-82</td>
<td>GRE</td>
<td>++</td>
<td>++</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>GSE</td>
<td>0</td>
<td>++</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>GRX</td>
<td>+</td>
<td>--</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>GSX</td>
<td>+</td>
<td>--</td>
</tr>
<tr>
<td>Flynn (1991)</td>
<td>USA</td>
<td>1978-84</td>
<td>GNX</td>
<td>--</td>
<td>-</td>
</tr>
<tr>
<td>Geroski (1991a)</td>
<td>UK</td>
<td>1983-84</td>
<td>NSE</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>Mata (1991)</td>
<td>Portugal</td>
<td>1982-86</td>
<td>GNE</td>
<td>+</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>GNE</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Schwalbach (1991)</td>
<td>Germany</td>
<td>1983-85</td>
<td>GRE</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Sleuwaegen and Dehandschutter (1991)</td>
<td>Belgium</td>
<td>1980-84</td>
<td>GRE</td>
<td>+</td>
<td>++</td>
</tr>
<tr>
<td>Von der Fehr (1991)</td>
<td>Norway</td>
<td>1981-85</td>
<td>GNE</td>
<td>++</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>GNE</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Yamawaki (1991)</td>
<td>Japan</td>
<td>1980-84</td>
<td>NRE</td>
<td>0/+</td>
<td>++</td>
</tr>
<tr>
<td>Rosenbaum and Lamort (1992)</td>
<td>USA</td>
<td>1972-82</td>
<td>GRE</td>
<td>++</td>
<td>++</td>
</tr>
<tr>
<td>Mata (1993)</td>
<td>Portugal</td>
<td>1982-86</td>
<td>GNE</td>
<td>0</td>
<td>+</td>
</tr>
<tr>
<td>Rosenbaum (1993)</td>
<td>USA</td>
<td>1972-82</td>
<td>NRE</td>
<td>++</td>
<td>++</td>
</tr>
<tr>
<td>Audretsch and Acs (1994)</td>
<td>USA</td>
<td>1976-86</td>
<td>GNE</td>
<td>n</td>
<td>0</td>
</tr>
<tr>
<td>Santarelli and Sterlacchini (1994)</td>
<td>Italy</td>
<td>1986-89</td>
<td>GRE</td>
<td>0</td>
<td>++</td>
</tr>
<tr>
<td>Wagner (1994a)</td>
<td>Germany</td>
<td>1979-88</td>
<td>GEE</td>
<td>0/-</td>
<td>-/++</td>
</tr>
</tbody>
</table>

Note:
Measures of entry and exit
- GDE = Gross Dummy Entry (1 in case of entry else 0)
- GEE = Gross Employment share of Entry
- GEX = Gross Employment share of eXit
- GNE = Gross Number of Entering firms
- GNX = Gross Number of eXiting firms
- GRE = Gross Rate of Entry
- GRX = Gross Rate of eXit
- GSE = Gross Sales (or production) share of Entry
- GSX = Gross Sales (or production) share of eXit
- NNE = Net Number of Entering firms
- NRE = Net Rate of Entry
- NSE = Net Sales share of Entry

*Kessides (1986,1990) estimates a non-linear specification, leading to industry-dependent effects of profitability on entry.
Most studies incorporate two incentives to enter (or disincentives to exit): a measure of profitability and a measure of market growth. Table 4.3 presents the reported significance of the effects of profitability and growth on entry and exit in the 44 studies. This overview provides an adequate benchmark against which to compare our results on the effects of profitability and market growth on entry and exit in the retail sector. A summary of the findings is presented in Table 4.4. In 20 out of 35 empirical results a significant positive effect of (a measure of) profitability on (a measure of) gross entry is found. A significant negative effect is only found by Wagner (1994a) when using a reweighted regression approach. In 28 out of 37 empirical results a significant positive effect of (a measure of) market growth on (a measure of) gross entry is found. No study reports a significant negative effect. The results for gross exit are somewhat less cut. Only 4 out of 10 results show a significant negative effect of profitability on gross exit, while 3 results show the opposite. The results for market growth are more straightforward: 5 out of 9 results are negative and significant, while no study finds a significant positive effect. Some studies use net entry (gross entry minus gross exit) as dependent variable. Significant positive effects of (a measure of) profitability and market growth on (a measure of) net entry are found in 8 out of 13 and 13 out of 18 empirical results, respectively. In both cases only one result shows a significant negative effect.

Table 4.4: Summary of empirical studies listed in Table 4.3

<table>
<thead>
<tr>
<th>Effect of profitability</th>
<th>Effect of growth</th>
</tr>
</thead>
<tbody>
<tr>
<td>- - 0 + ++ total</td>
<td>- - 0 + ++ total</td>
</tr>
<tr>
<td>Gross Entry</td>
<td>1 0 14 11 9 35</td>
</tr>
<tr>
<td>Gross Exit</td>
<td>1 3 3 2 1 10</td>
</tr>
<tr>
<td>Net Entry</td>
<td>0 1 4 6 2 13</td>
</tr>
</tbody>
</table>

Note: See note to Table 4.3. In case the study has presented evidence for both a significant and a non-significant effect (seven cases for the effect of profitability and one case for the effect of growth), we have chosen for the significant one.

4.3 The determinants of entry and exit in retailing

In this chapter we incorporate two measures of profitability in the entry and exit equations. The first is profit divided by modal income (PMI). Most (potential) entrepreneurs in retailing are small independents. About 88,600 out of 92,000 retail firms in the Netherlands have less than 10 employees, while 3,200 have 10 to 100 employees and 200 have more than 100 employees (EIM (1991)). As a small independent’s profit generally equals his income (before taxes), we may consider PMI to be a good indicator for the attractiveness to enter a certain shop type. A more traditional measure of profitability is profit as a ratio of equity (PRI). This corresponds to usual
measures of return on investment. The effect of demand growth on entry and exit is investigated by the variable DCS: the growth rate of real consumer expenditure on goods primarily sold in the shop-type. The variables PMI,PRI and DCS are one period lagged to adjust for the time period necessary for (potential) entrepreneurs to react to incentives. See Table 4.2 for the correlations between these three measures and the entry and exit rates.

We also incorporate UN, the level of unemployment (in 1,000,000 persons), and the change in this variable (DUN).23 A high level of unemployment or an increase in unemployment may have a positive effect on entry. Workers who are or become unemployed may consider becoming self-employed in the retail sector. Retailing is a comparably simple entrepreneurial activity which requires only a limited financial investment. Also, it may have a negative effect on exit. High levels of unemployment or strong increases in this level may deter shopkeepers to exit because of the unfavourable conditions on the labour market. Being or becoming unemployed can therefore be interpreted as incentive to enter and disincentive to exit. However, one has to be careful when interpreting the effect of the (change in the) level of unemployment because unemployment data do not always provide an adequate picture of the tension on the labour market and because the great majority of unemployed never consider a new firm start-up or let alone reach the stage of starting their own retail business. The variables UN and DUN are one period lagged in the entry equation. We assume that shopkeepers' exit decisions are affected by contemporaneous labour market conditions but that there is a lag between being or becoming unemployed and starting a retail venture.

Three indicators of the level of entry and exit barriers are considered. First, the degree of franchising, DF, is included. This variable is defined by the number of franchisees divided by the total number of firms. Entry by an independent firm may be expected to be less attractive in a shop-type in which franchising has proved a successful instrument of market penetration. The importance of the degree of franchising as entry barrier may however be limited because the viability of market niches is seldomly affected by an increase in the degree of franchising. An important feature of franchising is a high degree of similarity of the shops and franchisees therefore may lack the specialization needed for successful exploitation of market niches. The extent to which franchising may be considered to be an exit barrier is also uncertain. Small independents will be confronted by competition from franchisees. The franchiser may however provide independents who threaten to succumb to this competitive pressure the opportunity of becoming franchisees as well. That is, firms which

23 See Hudson (1989) and Storey (1991) for investigations into the effect of unemployment on entry and exit.
suffer from the competition by franchisees may very well profit from the competition between franchisers to obtain new outlets. A second indicator of the level of entry and exit barriers is small store presence, SSP. A large small store presence indicates that entry and exit barriers are low because a high sales volume does not seem to be prerequisite for entrepreneurial activity. We define SSP as the ratio of sales by firms with less than 10 employees to total sales. A positive effect of small store presence on entry and on exit is expected. Acs and Audretsch (1989b) however find for U.S. manufacturing that small firms do not tend to enter those industries in which there is already a considerable presence of small firms. They claim that 'flexible specialization' may have enabled small firms to enter and exist in markets where they previously would have experienced severe scale disadvantages. A third indicator of the level of entry and exit barriers is floorspace requirement. Retail business which needs extensive floorspace, FS, may be more difficult to set up. High floorspace requirement indicates high financial investments and problems in finding an appropriate establishment. This constitutes an important entry barrier in retailing. It may also be considered an exit barrier because the amount of floorspace is often closely linked to high investments in inventory and long tenancy agreements.24 We measure floorspace in 10,000 square metres. The variables DF, SSP and FS are all one period lagged in the entry equation. We assume that a decision to enter is based on market characteristics of one year before, while a decision to exit is affected by contemporaneous market characteristics.

4.4 Effects and estimation of displacement and replacement

Some recent studies, like Rosenbaum and Lamort (1992) and Sleuwaegen and Dehandschutter (1991), consider the effects of displacement and replacement. Displacement is the positive effect of entry on exit. There are two possible causes for this effect. First, new entry may lead to stronger competitive pressures which lead to exit (entry 'forces' exit). Second, independents who seriously consider to exit may wait for new entrants to take over their business. Replacement is the positive effect of exit on entry. There are again two possible causes. First, exit may provide additional market room for (potential) entrepreneurs to enter (exit 'allows' entry). Second, entrepreneurs may wait to enter until they are offered the opportunity to take over a shop (and its clientele).

Displacement implies that entry is an explanatory variable in the exit equation, while replacement implies that exit is an explanatory variable in the entry equation. In case exit (entry) can be treated as exogenous to the entry (exit) equation we use the

Seemingly Unrelated Relations (SUR) estimation technique. If however entry and exit are simultaneously determined, i.e. exit (entry) is endogenous to the entry (exit) equation, we choose a simultaneous equations estimation technique like Three Stage Least Squares (3SLS). The question now is whether exit (entry) can be treated as exogenous versus endogenous in the entry (exit) equation. In case of exogeneity the equations can be interpreted as reduced form equations. This implies that the entry and exit rates cannot be expressed as functions of market incentives and barriers only.

We test the null hypothesis of exogeneity of the exit (entry) rate in the entry (exit) equation in two ways.\textsuperscript{25} Using the Hausman approach (Godfrey (1988), sec. 1.6 and MacKinnon (1992), sec. 9) we estimate the model using Three Stage Least Squares (3SLS) and using SUR. Both estimation techniques are consistent under the null hypothesis of exogeneity, but SUR is the more efficient one. SUR is however not consistent under the alternative hypothesis of endogeneity, while 3SLS is. The Hausman test statistic is defined as \((\hat{\alpha}_{3SLS} - \hat{\alpha}_{SUR})' (S_{3SLS} - S_{SUR}^{-1} (\hat{\alpha}_{3SLS} - \hat{\alpha}_{SUR})\), where \(\hat{\alpha}_{3SLS}\) and \(\hat{\alpha}_{SUR}\) stand for the parameter vector estimators using 3SLS and SUR, respectively. The matrices \(S_{3SLS}\) and \(S_{SUR}\) are the corresponding estimators of the covariance matrices.\textsuperscript{26} The Hausman test statistic is asymptotically \(\chi^2\)-distributed under the null hypothesis with degrees of freedom equal to the number of elements of the parameter vector. Rosenbaum and Lamort (1992) use the Spencer-Berk test statistic (Kmenta (1986), sec. 13.6).\textsuperscript{27} This test involves calculating exit's (entry's) expected value by regressing exit (entry) on all the predetermined variables in the system and by evaluating the estimated t-value of the coefficient of this expected value when included next to the observed exit (entry) using OLS estimation. Under the null hypothesis of exogeneity this value is t-distributed with degrees of freedom equal to the number of observations minus the number of variables in the model.

We also incorporate the one period lagged entry and exit rates in the model. The lagged exit rate (EXIT\(_{t-1}\)) affects the entry rate when replacement is not immediate but taking place after one time period. We also expect the lagged entry rate (ENTRY\(_{t-1}\)) to have a positive effect on the entry rate. Gort and Konakayama (1982) argue that perceptions of profit opportunities are positively related to the successful experience of

\textsuperscript{25}MacKinnon (1992, p.126) claims that one should not interpret these tests as tests on exogeneity. One really tests whether possible endogeneity causes SUR estimates to be inconsistent.

\textsuperscript{26}The procedure runs into difficulties when \(S_{3SLS} - S_{SUR}\) is not positive definite, which is often the case in practice. We did not encounter this problem in any of the tests.

\textsuperscript{27}Rosenbaum and Lamort (1992) report an insignificant Spencer-Berk exogeneity test statistic. This implies that exit (entry) can be treated as exogenous to the entry (exit) equation in their sample. Their final conclusion is however incorrect: they claim that entry and exit do not appear to be causally related but that they are just part of the same market process. Exogeneity does however not imply that entry does not cause exit and vice versa. It implies that entry and exit are not simultaneously determined.
others in the market. Entrepreneurial activity may very well tend to be autocorrelated: the incidence of entry may stimulate (potential) entrepreneurs to consider entry as well. Gort and Konakatama refer to this as the demonstration effect. Similar arguments for the incorporation of the lagged entry and exit rate in the exit equation can be used.28

The empirical model (fixed shop type effects included) to be investigated is:

(4.1) \[ \text{ENTRY}_{it} = a_0 + a_1 \text{PMI}_{it-1} + a_2 \text{PRI}_{it-1} + a_3 \text{DCS}_{it-1} + a_4 \text{UN}_{it-1} + a_5 \text{DUN}_{it-1} + a_6 \text{DF}_{it-1} + a_7 \text{SSP}_{it-1} + a_8 \text{FS}_{it-1} + a_9 \text{EXIT}_{it} + a_{10} \text{ENTRY}_{it-1} + \epsilon_{it} \]

(4.2) \[ \text{EXIT}_{it} = b_0 + b_1 \text{PMI}_{it-1} + b_2 \text{PRI}_{it-1} + b_3 \text{DCS}_{it-1} + b_4 \text{UN}_{it-1} + b_5 \text{DUN}_{it-1} + b_6 \text{DF}_{it} + b_7 \text{SSP}_{it} + b_8 \text{FS}_{it} + b_9 \text{ENTRY}_{it} + b_{10} \text{ENTRY}_{it-1} + \epsilon_{2it} \]

where:

ENTRY = Entry rate;
EXIT = Exit rate;
PMI = Profit divided by modal income;
PRI = Profit divided by equity;
DCS = Change in real consumer spending
UN = Number of unemployed;
DUN = Change in the number of unemployed;
DF = Degree of franchising;
SSP = Small store presence;
FS = Floorspace requirement.

4.5 Data and empirical results

In this study a data set of 23 Dutch shootypes for the period 1981-1988 is used. This implies a total of 184 data-points. The source of the data is an ongoing panel of independent, mainly small Dutch retailers called Bedrijfssignaleringssysteem (interfirm

28In their study of the interaction of entry and exit rates, Johnson and Parker (1994) also investigate displacement, replacement and demonstration effects. They use a different terminology for these effects, viz. competition and multiplier effects. They apply a vector autoregression approach without exogenous variables to entry and exit rates in the United Kingdom in the 1980s and find support for the existence of both these effects. Kleijweg and Lever (1994, 1996) study the determinants of entry and exit in Dutch manufacturing and find evidence for replacement and displacement effects.
comparison system) which is operated by the EIM Small Business Research and Consultancy (EIM) in Zoetermeer.\textsuperscript{29} On average a data-point is computed using observations from about seventy individual retail stores. The consumer spending and modal income data are from the Central Bureau of Statistics (CBS) in Voorburg,\textsuperscript{30} unemployment data were retrieved from the UN Monthly Bulletin of Statistics and data on entry and exit of retail establishments were gathered by the Central Registration Office (CRK) in The Hague. Data on the total number of franchisees were available only for 1980, 1982 and 1988. We used an interpolation for the intermediate years.

Table 4.5 presents the empirical results with fixed effects (shoptype dummies) assumed in all models (use is made of the SAS-module SYSLIN). The fixed effects correct for shoptype-specific aspects which are not incorporated into the entry and exit equations. The first two columns contain the estimation results when displacement and replacement effects are not taken into account ($a_9 = a_{10} = a_{11} = b_9 = b_{10} = b_{11} = 0$). We incorporate the displacement and replacement effect in the next two columns by considering exit (entry) to be endogenous to the entry (exit) equation ($a_{10} = a_{11} = b_{10} = b_{11} = 0$). We proceed to test whether SUR estimation is viable by using the Hausman and Spencer-Berk test statistics. Both statistics support the null hypothesis of the entry rate to be exogenous to the exit equation and vice versa. The Hausman statistics are 10.204 for the entry equation (p-value is 0.33) and 4.994 for the exit equation (p-value is 0.83), while the Spencer-Berk statistics are 0.564 (p-value is 0.57) and -1.377 (p-value is 0.17), respectively. Columns five and six contain the results when the model is re-estimated using SUR (again $a_{10} = a_{11} = b_{10} = b_{11} = 0$). In the next columns we present estimation results when the one period lagged entry and exit rates are incorporated into the model. The Hausman and Spencer-Berk test statistics again point to (pair-wise) exogeneity of the entry and exit rate. The Hausman statistics are 10.530 for the entry equation (p-value is 0.48) and 2.430 for the exit equation (p-value is 0.99), while the Spencer-Berk statistics are 0.889 (p-value is 0.37) and -1.265 (p-value is 0.20), respectively. Therefore, we also present estimation results using SUR in the last two columns.

\textsuperscript{29}In Carree and Thunk (1994) and in Chapter 3 a more extensive data set is used of 36 shoptypes covering a longer time period. However, the limited availability of data on the degree of franchising and of small store presence confines our data set to 23 shoptypes for the 1981-1988 period.

\textsuperscript{30}There were no data available on small store presence for three shoptypes: "bakers", "confectioners" and "pet shops". These shoptypes consist of small firms (less than 10 employees) for nearly 100%. Small store presence is taken to be unity in every period for these shoptypes. Data on small store presence were available for four shoptypes for 1988 only: "greengrocers", "florists", "fish shops", and "tobacco shops". In these shoptypes small firms accounted for at least for 85% of the sales in that year. Small store presence in the years 1980-87 is taken to be equal to the 1988-level.
### Table 4.5: Empirical results for entry and exit models

<table>
<thead>
<tr>
<th>Variable</th>
<th>SUR ENTRY</th>
<th>SUR EXIT</th>
<th>3SLS ENTRY</th>
<th>3SLS EXIT</th>
<th>SUR ENTRY</th>
<th>SUR EXIT</th>
</tr>
</thead>
<tbody>
<tr>
<td>PMI&lt;sub&gt;1&lt;/sub&gt;</td>
<td>0.019&lt;sup&gt;+&lt;/sup&gt; -0.012</td>
<td>0.031&lt;sup&gt;+&lt;/sup&gt; -0.012</td>
<td>0.028&lt;sup&gt;-&lt;/sup&gt; -0.019&lt;sup&gt;-&lt;/sup&gt;</td>
<td>0.029&lt;sup&gt;-&lt;/sup&gt; -0.006</td>
<td>0.026&lt;sup&gt;-&lt;/sup&gt; -0.017&lt;sup&gt;-&lt;/sup&gt;</td>
<td></td>
</tr>
<tr>
<td>(2.4)</td>
<td>(2.0)</td>
<td>(4.6)</td>
<td>(1.7)</td>
<td>(4.7)</td>
<td>(3.9)</td>
<td>(4.5)</td>
</tr>
<tr>
<td>PRI&lt;sub&gt;1&lt;/sub&gt;</td>
<td>0.003</td>
<td>0.006</td>
<td>-0.009</td>
<td>0.004</td>
<td>-0.008</td>
<td>0.009</td>
</tr>
<tr>
<td>(0.2)</td>
<td>(0.4)</td>
<td>(0.6)</td>
<td>(0.3)</td>
<td>(0.5)</td>
<td>(0.7)</td>
<td>(0.8)</td>
</tr>
<tr>
<td>DCS&lt;sub&gt;1&lt;/sub&gt;</td>
<td>0.026</td>
<td>-0.042</td>
<td>0.088&lt;sup&gt;-&lt;/sup&gt; -0.045</td>
<td>0.087&lt;sup&gt;-&lt;/sup&gt; -0.080&lt;sup&gt;-&lt;/sup&gt;</td>
<td>0.088&lt;sup&gt;-&lt;/sup&gt; -0.054</td>
<td>0.085&lt;sup&gt;-&lt;/sup&gt; -0.082&lt;sup&gt;-&lt;/sup&gt;</td>
</tr>
<tr>
<td>(0.7)</td>
<td>(1.3)</td>
<td>(2.7)</td>
<td>(1.3)</td>
<td>(3.0)</td>
<td>(3.1)</td>
<td>(2.5)</td>
</tr>
<tr>
<td>UN&lt;sub&gt;(t)&lt;/sub&gt;</td>
<td>0.010</td>
<td>0.028&lt;sup&gt;-&lt;/sup&gt;</td>
<td>-0.014</td>
<td>0.022&lt;sup&gt;-&lt;/sup&gt;</td>
<td>-0.017&lt;sup&gt;-&lt;/sup&gt;</td>
<td>0.013</td>
</tr>
<tr>
<td>(1.1)</td>
<td>(2.8)</td>
<td>(1.6)</td>
<td>(2.2)</td>
<td>(2.3)</td>
<td>(1.7)</td>
<td>(1.7)</td>
</tr>
<tr>
<td>DUN&lt;sub&gt;(t)&lt;/sub&gt;</td>
<td>0.043&lt;sup&gt;-&lt;/sup&gt;</td>
<td>-0.011</td>
<td>0.032&lt;sup&gt;-&lt;/sup&gt;</td>
<td>-0.018</td>
<td>0.023&lt;sup&gt;-&lt;/sup&gt;</td>
<td>-0.025&lt;sup&gt;-&lt;/sup&gt;</td>
</tr>
<tr>
<td>(2.9)</td>
<td>(0.9)</td>
<td>(2.3)</td>
<td>(1.4)</td>
<td>(2.1)</td>
<td>(2.4)</td>
<td>(2.5)</td>
</tr>
<tr>
<td>DF&lt;sub&gt;(t)&lt;/sub&gt;</td>
<td>0.195&lt;sup&gt;-&lt;/sup&gt;</td>
<td>0.141&lt;sup&gt;-&lt;/sup&gt;</td>
<td>0.042</td>
<td>0.101</td>
<td>0.024</td>
<td>0.019</td>
</tr>
<tr>
<td>(2.6)</td>
<td>(2.7)</td>
<td>(0.7)</td>
<td>(1.7)</td>
<td>(0.4)</td>
<td>(0.4)</td>
<td>(1.0)</td>
</tr>
<tr>
<td>SSP&lt;sub&gt;(t)&lt;/sub&gt;</td>
<td>-0.106</td>
<td>0.036</td>
<td>-0.138</td>
<td>0.049</td>
<td>-0.160&lt;sup&gt;-&lt;/sup&gt;</td>
<td>0.130&lt;sup&gt;-&lt;/sup&gt;</td>
</tr>
<tr>
<td>(1.1)</td>
<td>(5.5)</td>
<td>(1.8)</td>
<td>(0.7)</td>
<td>(2.2)</td>
<td>(2.4)</td>
<td>(1.1)</td>
</tr>
<tr>
<td>FS&lt;sub&gt;(t)&lt;/sub&gt;</td>
<td>-0.888&lt;sup&gt;-&lt;/sup&gt;</td>
<td>-0.309&lt;sup&gt;-&lt;/sup&gt;</td>
<td>-0.654&lt;sup&gt;-&lt;/sup&gt;</td>
<td>-0.464&lt;sup&gt;-&lt;/sup&gt;</td>
<td>-0.379&lt;sup&gt;-&lt;/sup&gt;</td>
<td>0.023</td>
</tr>
<tr>
<td>(4.2)</td>
<td>(2.0)</td>
<td>(3.1)</td>
<td>(1.5)</td>
<td>(2.4)</td>
<td>(0.2)</td>
<td>(2.6)</td>
</tr>
<tr>
<td>EXIT &lt;sup&gt;*&lt;/sup&gt;</td>
<td>1.080&lt;sup&gt;+&lt;/sup&gt;</td>
<td>1.108&lt;sup&gt;+&lt;/sup&gt;</td>
<td>0.992&lt;sup&gt;+&lt;/sup&gt;</td>
<td>1.031&lt;sup&gt;+&lt;/sup&gt;</td>
<td>0.720&lt;sup&gt;+&lt;/sup&gt;</td>
<td>0.040</td>
</tr>
<tr>
<td>(5.1)</td>
<td>(17.5)</td>
<td>(4.3)</td>
<td>(14.9)</td>
<td>(0.9)</td>
<td>(16.0)</td>
<td>(0.1)</td>
</tr>
</tbody>
</table>

**Note:** All models are estimated with fixed effects. Numbers in parentheses are t-statistics. Coefficients which are significant at a 5%-significance level are denoted by <sup>+</sup>. The test statistics \( F_{FE} \), \( \chi^2_{Hausman} \), and \( t_{Spencer-Berk} \) stand for the F-test on the hypothesis of the absence of shoptype specific effects (fixed effects) and the Hausman \( \chi^2 \)-test and Spencer-Berk t-test on the hypothesis of exogeneity of ENTRY in the exit equation and of EXIT in the entry equation. The coefficient \( r \) stands for the estimated correlation between the observed entry (exit) rate and the estimated entry (exit) rate. The variables UN, DUN, DF, SSP and FS are one period lagged only for the entry equation.

The SUR estimation results give strong support for the positive effects of profitability and demand growth on entry and the negative effects of these variables on exit. Entry and exit rates in retailing seem to be more sensitive to market incentives than in manufacturing when considering the mixed results for manufacturing industries (see
Section 4.2). The surplus of profit over modal income (PMI, \(_t\)) and the growth rate of real consumer spending for goods primarily sold in the shootype (DCS, \(_t\)) have positive and significant effects on entry and negative and significant effects on exit. See Table 4.1 for the average values of PMI, \(_t\) and DCS, \(_t\) in the 23 shootypes. Return on investment (PRI, \(_t\)) has no effect on either entry or exit rates. This indicates that (potential) shopkeepers place higher value on the reward for their labour input than on the reward for their capital investments when considering entry or exit.

The level of unemployment (UN) does not have the expected positive effect on the entry rate nor the negative effect on the exit rate. This may be a consequence of the level of unemployment acting as a business cycle indicator next to the shootype-specific market growth (DCS). The change in the level of unemployment (DUN) does however have the expected effects. This variable may be a more reliable indicator of opportunities on the labour market during the 1980s than the level of unemployment because of the consistent high number of structurally unemployed in the Netherlands during that period. Workers who have only recently become unemployed may also be more able and motivated to use self-employment as an alternative to being unemployed than people who have had no working experience for a longer time period. The results show that an increase in the number of unemployed appears to lead to more entry while it seems to prevent exit.

The degree of franchising (DF) has no significant effects on entry or exit when the displacement and replacement effects are included in the model. The positive effect of the degree of franchising on the entry and exit rate when displacement and replacement are not taken into consideration, does indicate that part of the replacement and displacement process originates from the growing market penetration of franchisers. Small store presence (SSP) has a significant effect on the entry and exit rate only in case of SUR-estimation with lagged dependent variables excluded. The presence of many small stores appears not to be attractive to new entrants, which is in accordance with results found by Acs and Audretsch (1989a) for U.S. manufacturing. At the same time we do find that the exit rate is higher, ceteris paribus, in shootypes dominated by small stores. The net exit in shootypes dominated by small stores leads to the general decline of small store presence in retailing.\(^{31}\) The floorspace requirement (FS) has a significant negative effect on entry, while it has no significant effect on exit. One can conclude that floorspace requirement (or more generally investment requirement) obstructs entry.

The effects of many other entry and exit barriers are adjusted for by the inclusion of

---

\(^{31}\) See Carree, Potjes and Thurik (1993) for a study into the determinants of the declining small store presence found in Japanese retailing. See Nooteboom (1986a) for a general discussion of the 'ousting of smallness' in retailing.
fixed effects. The value and significance of the shoptype dummy-coefficients therefore may provide some additional information on the (relative) ease of entry and exit over the shoptypes. One can divide the 23 shoptypes into 9 shoptypes with relatively high (legal) requirements of professional skills and 14 shoptypes with only limited requirements. The first group consists of "supermarkets with butcher's shop", "butchers", "bakers", "confectioners", "liquor stores", "bicycles", "photographer's shops", "watches, gold, silver" and "druggists". In case of SUR estimation without displacement and replacement effects the mean dummy in the entry equation is 0.122 for this first group, while it is 0.185 for the other 14 shoptypes (t-test on equality of means: -2.21). For the mean dummy in the exit equation the figures are 0.043 and 0.076 (t-test: -1.57). In case of SUR estimation with lagged dependent variables included the mean dummy in the entry equation is 0.046 for the first group and 0.064 for the second group (t-test: -1.56). The respective mean dummies in the exit equation are -0.027 and -0.028 (t-test: 0.07). That is, (legal) requirements of professional skills appear to reduce entry rates (and possibly exit rates) and therefore may slow down adjustment processes in retailing. Carree, Fris and Thurik (1993) provide a complete overview of the dummies for the 23 shoptypes. They use a model similar to equations (4.1) and (4.2) but without UN and the replacement and displacement variables. They find values of the t-test on equality of means of -2.1 for the entry equation and -1.8 for the exit equation. They conclude that the recent abolishment (January 1996) of many specific requirements to start an enterprise in Dutch retailing is likely to lead to more turbulence (entry plus exit) and to a somewhat higher number of firms (through increased net entry).

The coefficient of the exit rate incorporated in the entry equation is very close to unity, independent of whether 3SLS or SUR is used as estimation technique. Rosenbaum and Lamort (1992) also found this result for U.S. manufacturing using the SUR estimation technique. If no information on incentives or barriers would be available one could consider the exit rate to be a good predictor of the entry rate in a shoptype. The extent to which the entry rate exceeds the exit rate, i.e. the extent to which net entry is positive, depends on the attractiveness of the shoptype and barriers impeding new entrants to occupy the market room created by exit. There is support for the demonstration effect: the lagged entry rate has a significant positive effect on the present period's entry rate.

The treatment of the entry rate as endogenous or exogenous variable, has an important impact on the effect of this variable in the exit equation: the coefficient becomes significant in case SUR is used, while it is not significant using the 3SLS estimation technique. The Hausman and Spencer-Berk test statistics provide support for the use of the SUR estimation technique. This is evidence for entry to influence exit

---

while they are not simultaneously determined. The lagged exit rate has a significant positive effect on the present period's exit rate. This could again be related to a demonstration effect. Firms that change their range of activities, shifting to another shoptype, may be followed by competitors.

4.6 Average replacement and displacement periods and persistence of profits

In Carree, Fris and Thurik (1993) it is shown that the abolishment of specific requirements to start a retail venture may not only lead to more entry and exit but also to shorter replacement and displacement periods. It is also shown that it may lead to less persistence of profits. In this section we report on these two alternative measures of market dynamics. The same groups of shoptypes as introduced in Section 4.5 will be included, one with relatively high legal requirements (9 shoptypes) and one with only limited legal requirements (14 shoptypes).

The average replacement and displacement periods are computed as follows. In equation (4.1) with \( a_9 = a_{10} = a_{11} = 0 \) we add \( EXIT_{t-1} \) through \( EXIT_{t-5} \). The regression coefficients of these five variables measure the replacement effect.\(^{33}\) The effect of \( EXIT_{t-k} \) on \( ENTRY_{t} \) is replacement taking place after a period of length \( k \). The average replacement period is computed as the weighted sum of the five replacement effects. In Table 4.6 the average replacement periods are computed for the two groups of shops. For example, the 9 shoptypes with specific legal requirements have an estimated replacement period of \( 1 \times 0.085 + 2 \times 0.052 + 3 \times 0.675 + 4 \times 0.515 + 5 \times 0.327 = 2.639 \) years. The replacement period is an indication of the time needed before exit of a firm, which cannot be related to an unfavourable market environment (e.g. low profits or low growth of demand), leads to new entry. The same procedure is followed to compute the average displacement period. In equation (4.2) with \( b_9 = b_{10} = b_{11} = 0 \) we add \( ENTRY_{t-1} \) through \( ENTRY_{t-5} \). The regression coefficients of these five variables measure the displacement effect.\(^{34}\) The effect of \( ENTRY_{t-k} \) on \( EXIT_{t} \) is displacement taking place after a period of length \( k \). The average displacement period is computed as the weighted sum of the five displacement effects. In the last two columns of Table 4.6 the average displacement periods are computed for the two groups of shops.

Table 4.6 shows that both the average replacement period and the average displacement period in the 9 shoptypes with specific legal requirements are almost two years longer than these periods in the other 14 shoptypes. The additional requirements

\(^{33}\)The sum of the coefficients is restricted to unity. This implies a complete replacement within 5 years. The use of a restricted least squares estimator alleviates problems of multicollinearity which result from incorporating the five subsequent exit rates (Judge et al. (1982), p.626).

\(^{34}\)The sum of the coefficients is again restricted to unity.
seem to lead to potential entrepreneurs needing more time to fill up market room and incumbents staying longer in the market despite increased competition.

Table 4.6: Estimation of average replacement and displacement period (1984-1988)

<table>
<thead>
<tr>
<th>Effect on</th>
<th>Entry 9 shootypes</th>
<th>Exit 9 shootypes</th>
<th>Entry 14 shootypes</th>
<th>Exit 14 shootypes</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 year</td>
<td>0.085 (0.4)</td>
<td>0.317 (2.2)</td>
<td>1.085 (10.2)</td>
<td>0.714 (9.3)</td>
</tr>
<tr>
<td>2 years</td>
<td>0.052 (0.2)</td>
<td>-0.252 (1.4)</td>
<td>-0.004 (0.0)</td>
<td>0.077 (1.1)</td>
</tr>
<tr>
<td>3 years</td>
<td>0.675 (2.3)</td>
<td>0.097 (0.6)</td>
<td>-0.012 (0.1)</td>
<td>0.129 (1.8)</td>
</tr>
<tr>
<td>4 years</td>
<td>0.515 (2.1)</td>
<td>0.630 (4.1)</td>
<td>-0.209 (2.4)</td>
<td>0.018 (0.3)</td>
</tr>
<tr>
<td>5 years</td>
<td>-0.327 (1.1)</td>
<td>0.207 (1.4)</td>
<td>0.140 (2.0)</td>
<td>0.061 (1.0)</td>
</tr>
</tbody>
</table>

Replacement per. 2,639
Displacement per. 3,659

Note: t-values between brackets. The sum of the coefficients in each column is restricted to unity. 'k years' stands for the effect of the entry (exit) rate on the exit (entry) rate k periods later.

Another measure of market dynamics is the speed at which profits converge to the equilibrium level. In a dynamic market setting excess profits will lead to increased competition (possibly from new entrants) causing profits to decline. Profits below the equilibrium level will lead to termination or adoption of inefficient activities (possibly complete firms). The speed at which profits converge to the equilibrium level can be estimated by a simple error-correction model (see also Chapter 3). The endogenous variable in this regression is ΔPMi,t and the exogenous variables are dummies for each shootype, PMi,t-1 and DCSi,t. Incorporating dummies allows for the equilibrium profit level to differ across shootypes. The growth in demand is incorporated to correct for changes in demand. The regression results for the two groups of shootypes can be found in Table 4.7. The results show that the speed of adjustment of profit is somewhat larger (but not significantly so) for the shootypes without specific requirements.

The three measures of market dynamics each suggest that the 9 shootypes with specific legal requirements to start a retail venture are less 'dynamic' than the 14 shootypes with only limited requirements. The last group of shootypes has higher entry and exit rates, shorter replacement and displacement periods and less persistence of profits. Not each difference has been found to be significant but the combination of measures leaves little doubt about which market settings are more dynamic.

---

35See Mueller (1977,1990). A similar model is used by Kleijweg (1993) for Dutch manufacturing industries. His estimates for the effect of excess profit on the change in profit are -0.340 for small firms (10-100 employees) and -0.275 for large firms (more than 100 employees). The speed of adjustment is slower for large firms when compared to small firms. This may be a consequence of the lower barriers to enter as a small firm than as a large firm.
Table 4.7: The change in profitability explained by excess profits and growth of demand

<table>
<thead>
<tr>
<th>Dependent: ΔPMI&lt;sub&gt;it&lt;/sub&gt;</th>
<th>9 shootypes</th>
<th>14 shootypes</th>
</tr>
</thead>
<tbody>
<tr>
<td>PMI&lt;sub&gt;i,t-1&lt;/sub&gt;</td>
<td>-0.300 (3.1)</td>
<td>-0.386 (4.9)</td>
</tr>
<tr>
<td>DCS&lt;sub&gt;it&lt;/sub&gt;</td>
<td>1.266 (1.8)</td>
<td>1.785 (4.6)</td>
</tr>
</tbody>
</table>

Note: t-values between brackets. The first 9 shootypes have relatively high legal requirements to start a retail venture while the second 14 shootypes have only limited legal requirements. Shootype dummies are included in both regressions.

4.7 Conclusions

The determinants of entry and exit in retailing are the subject of research in this chapter. Retailing is an interesting testing area for these determinants not only because of its importance in the economy but also because of its presumed sensitivity to market incentives due to the limited amount of investments and skills needed when compared to other sectors. We find lagged profitability and lagged market growth in this sector to have a significant positive effect on entry and a significant negative effect on exit. Growing unemployment also seems to stimulate entry and to obstruct exit. Evidence is found for shootypes dominated by small stores to have lower entry rates and higher exit rates. The variable describing the extent of floorspace requirement has a significant negative effect on the entry rate suggesting that it acts as an impediment to new firm entry. We also find evidence for the entry rate to be lower and replacement and displacement periods to be longer in shootypes with relatively high requirements of professional skills. This provides support for the decision made by the Dutch Ministry of Economic Affairs to abolish a wide variety of legal requirements in retailing in order to improve dynamic market efficiency.

There is strong statistical support for the incorporation of exit (entry) as determinant of entry (exit). Shootypes with high entry rates also have high exit rates. Such high turnover may however not only result from low entry (exit) barriers, but also from entrants replacing or displacing exiting firms. We find evidence for these displacement and replacement processes. There is however no evidence of simultaneity of the entry and exit rates. This implies that equations (4.1) and (4.2) can be interpreted as reduced form equations. As a consequence, the entry and exit rates cannot be expressed as functions of market incentives and barriers only. Studies which do not take into account the separate influence of entry on exit, i.e. displacement, and the separate influence of exit on entry, i.e. replacement, will therefore suffer from an omitted variable bias.
<table>
<thead>
<tr>
<th>Year</th>
<th>1-year</th>
<th>2-year</th>
<th>3-year</th>
<th>4-year</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.663 (0.4)</td>
<td>0.662 (0.7)</td>
<td>0.675 (2.3)</td>
<td>0.615 (2.1)</td>
<td></td>
</tr>
<tr>
<td>1.085 (0.2)</td>
<td>0.304 (0.0)</td>
<td>0.202 (0.1)</td>
<td>0.186 (0.3)</td>
<td></td>
</tr>
<tr>
<td>0.271</td>
<td>0.077</td>
<td>0.027</td>
<td>0.018</td>
<td></td>
</tr>
</tbody>
</table>

**Table 4.** Average number of days off work due to injury or illness by duration of injury (1056-19203)
CHAPTER 5

THE CARRYING CAPACITY AND ENTRY AND EXIT FLOWS IN RETAILING

This study introduces a new and more uniform theory explaining entry and exit rates. According to our model emphasizing the carrying capacity of a market, net entry is a reaction to a disequilibrium situation. This is a situation in which the number of firms in the market is unequal to the carrying capacity of that market. Using oligopoly theory we derive an expression for the carrying capacity and also some testable implications. We investigate the speed of adjustment towards equilibrium, the effect of changes in consumer demand on the carrying capacity and the relative importance of entry and exit in the adjustment process. The theoretical model is tested using a panel data set of 22 retail industries for the 1981-1988 period.
5.1 Introduction

The process of entry and exit of firms serves as an important source of structural change in industries. Industries with relatively low birth and death rates are alleged to be vulnerable to misallocation of resources, limited innovativeness, and formal or tacit collusion (Geroski and Jacquemin (1985)). The continuous flows of entry and exit represent a changing pool of competitors, viz. the seedbed of new activities from which will emerge new and successful businesses and industries (Beesley and Hamilton (1984)).

Entry and exit rates vary strongly across industries (Dunne et al. (1988)). Empirical studies have proposed and examined a large number of factors which explain these differences. These factors pertain to four broad categories. The first category is the industry's environment and comprises the basic exogenous demand and cost conditions. Some of these environmental characteristics may restrict entry and exit rates during the entire life span of the industry. Examples are extent of consumer loyalty, laws affecting start-ups and possession of strategic raw materials. The second category is the stage of the industry life cycle. Industries are often subject to large waves of entry in the early stage of their life cycles followed by a 'shake-out' period in which many young inefficient firms exit while almost no firms enter (Gort and Klepper (1982), Klepper and Graddy (1990), Klepper and Miller (1995)). Over the life span of an industry the average growth rate of demand usually declines while the number and height of entry barriers rises. Some of these barriers which become more important at each subsequent stage of the industry life cycle are cost advantages of incumbents, capital requirements and advertising intensity (Karakaya and Stahl (1989)). The third category is the strategic behaviour of incumbents. Incumbent firms may limit price or install excess capacity in order to forestall or regulate entry. Highly concentrated industries could, as a result, exhibit lower entry rates than comparable industries with a low degree of concentration (Bunch and Smiley (1992), Masson and Shaanan (1986)). The last factor is the business cycle. During periods of a temporarily higher growth rate of demand entrants are attracted by increases in profitability due to the inability of incumbents to expend their capacity fast enough (Hause and Du Rietz (1984)). Entry and exit are interrelated in that they are both affected by the factors mentioned above. Entry and exit may however also be causally related: exit may cause entry and entry may cause exit. Entering firms may replace or displace exiting firms (see Chapter 4,

---

36The recently developed resource-partitioning theory predicts exactly the opposite: high levels of market concentration are thought to lead to higher founding rates of firms (Swaminathan (1995)). The theory predicts that high concentration allows specialists to draw on peripheral resources without entering into direct competition with generalists.
Love (1996), Rosenbaum and Lamort (1992)).

Most empirical studies, with some exceptions like Kessides (1986, 1990) and Bresnahan and Reiss (1990), do not derive testable implications of a theoretical model of industrial structure but instead use an ad-hoc specification for the rates of entry and exit. As a consequence, much of the 'knowledge' about the processes of entry and exit is scattered (Schmalensee (1989), p. 997). In this study we propose a new and more uniform approach. Each industry is assumed to consist of a number of separate and relatively homogeneous markets for each of which an equilibrium number of firms exists. We then model entry and exit at the separate market level as adjustments to a disequilibrium situation. The number of firms in the market is said to have attained the equilibrium value when all entrepreneurs earn some critical level of profits. This equilibrium number of firms is called the carrying capacity. This term was first used in population ecology work (Hannan and Freeman (1977)). It refers to the numbers of an organizational form that can be sustained in a particular environment in isolation from other populations (Hannan and Carroll (1992), p. 29). In case the actual number of firms in a market is larger than the carrying capacity, there is need for exit in the current or a subsequent period. However, if the actual number of firms is smaller than the carrying capacity this implies room for entry. The carrying capacity depends on the amount of consumer demand. In case of a growing demand the carrying capacity will increase as well.

In this study we derive an expression for the carrying capacity which we apply to retail industries (shoptypes). These industries consist of many local markets with only a handful establishments. The retail markets can usually be characterized by a high degree of product homogeneity, entrepreneurs facing similar cost functions and demand conditions, and low entry and exit barriers. These elements of market structure allow to make some strong assumptions which facilitate the derivation of an equilibrium number of independent firms in a market. Our model may therefore be particularly applicable to traditional retailing where innovation does not play an important role. An error correction model is used to estimate the effect of changes in consumer demand on the carrying capacity, the speed of adjustment of the actual number of firms to the carrying capacity and the relative importance of entry and exit in the adjustment process. For this a panel data set at a low level of aggregation for the retail sector is used. This study is one of few to investigate the determinants of the selection process of entry and exit over time in a non-manufacturing industry.

The plan of this chapter is as follows. In Section 5.2 an expression for the carrying capacity is derived using oligopoly theory. Section 5.3 is used to model the adjustment process towards the equilibrium number of firms. In Section 5.4 the data are presented and some characteristics of the retail sector are discussed. In Section 5.5 two implications of the oligopoly model are tested for Dutch retailing. The empirical results
for the error correction model are presented in Section 5.6, and Section 5.7 is used for summary and discussion.

5.2 The carrying capacity

In this section the carrying capacity of a market with oligopolistic competition is derived. This carrying capacity will be used as a benchmark for the actual number of firms. The derivation of the carrying capacity is analogous to that of the equilibrium number of firms in the homogeneous good exogenous sunk costs case discussed by Sutton (1991, pp. 30-32). Consider a market with $N$ firms producing a homogeneous good. Each of these firms is supposed to choose output, $q_j$, and to have identical cost functions, $C(q_j)$, and conjectural variations, $\phi$. Firm $j$ maximizes its profit, $\pi_j = p q_j - C(q_j)$, as follows:

$$
(5.1) \quad \frac{\partial \pi_j}{\partial q_j} = p \left(1 + \frac{Q}{p} \frac{\partial p}{\partial Q} \frac{q_j}{Q} \frac{\partial Q}{\partial q_j} \right) - \frac{\partial C}{\partial q_j} = 0, \quad j = 1, \ldots, N,
$$

where $Q = \sum_{j=1}^{N} q_j$ is total market output and $p$ is the market price. The term $\partial Q/\partial q_j$ is equal to the change in total market output resulting from a change in the output of firm $j$. It can be rewritten as

$$
(5.2) \quad \frac{\partial Q}{\partial q_j} = 1 + \sum_{k \neq j} \frac{\partial q_k}{\partial q_j} = 1 + \phi_j, \quad j = 1, \ldots, N,
$$

where $\phi_j$ are the conjectural variations. They are assumed to be identical across firms and equal to $\phi$. Note that $\phi$ is zero for a Cournot oligopoly. Introducing $e = \frac{\partial Q}{\partial p} p$ as price elasticity of demand we have from (5.1) that

$$
(5.3) \quad p \left(1 + \frac{1}{e} \frac{q_j}{Q} \right) = \frac{\partial C}{\partial q_j}, \quad j = 1, \ldots, N.
$$

Because marginal costs are equal for each firm, the production levels are also equal: $q_j = Q/N$. Taking a linear cost function, $C(q_j) = \alpha + \beta q_j$, the following market price can be derived:

$$
(5.4) \quad p = \beta \left(1 + \frac{1}{N e} \right)^{-1}.
$$
We shall return to the justification of linearity of the cost function when discussing the application of the model to the retail sector. The price elasticity of demand $\epsilon$ is usually assumed to be below zero. This implies that $N\epsilon < -1 - \phi$ and $\phi > -1$ in order to have prices higher than marginal costs. Profits for each firm, $\pi$, can be expressed as follows:

$$\tag{5.5} \pi = \frac{\beta Q \epsilon}{N \epsilon + 1 + \phi} - \alpha - \frac{\beta Q}{N}. $$

The demand elasticity of profit, $\eta_Q$, and the number of firms elasticity of profit, $\eta_N$, can now be determined:

$$\tag{5.6} \eta_Q = \frac{\partial \pi}{\partial Q} \frac{Q}{\pi} = \frac{\beta Q (1 + \phi)}{\beta Q (1 + \phi) + \alpha N (N \epsilon + 1 + \phi)} \quad \text{and} \quad \eta_Q = \frac{2N \epsilon + 1 + \phi}{N \epsilon + 1 + \phi}, $$

$$\tag{5.7} \eta_N = \frac{\partial \pi}{\partial N} \frac{N}{\pi} = \frac{-\beta Q (1 + \phi) (2N \epsilon + 1 + \phi)}{\beta Q (1 + \phi) (N \epsilon + 1 + \phi) + \alpha N (N \epsilon + 1 + \phi)^2}. $$

Note that $\eta_N$ must be more than twice $\eta_Q$ in absolute terms because $N \epsilon < -1 - \phi$. The question now is how many firms this market can carry. Assume that there is an exogenously given critical profit level, $\pi^*$. This level is determined by alternative opportunities for the owners of the firm to receive income.$^{37}$ The carrying capacity is that number of firms at which this "reservation wage", $\pi^*$, is equal to the firms' profit, $\pi$, in the industry. Solving for $N$ in equation (5.5), one finds that$^{38}$

$$\tag{5.8} N^* = \frac{1}{2\epsilon} \left( -1 - \phi - \sqrt{(1 + \phi)^2 - 4\beta Q \epsilon (1 + \phi) / (\pi^* + \alpha)} \right). $$

The carrying capacity, $N^*$, enlarges when total market demand, $Q$, increases, while it shrinks when the critical level of profits, $\pi^*$, increases. The carrying capacity is lower in markets with high constant costs, $\alpha$, and low variable costs, $\beta$, when compared to markets with low constant and high variable costs. Low price elasticity of demand, $\epsilon$, and high conjectural variation, $\phi$, lead to a less competitive environment and therefore to a higher carrying capacity. The market is said to be in a state of equilibrium when the...

$^{37}$De Wit and Van Winden (1990) find that the probability of becoming self-employed is higher, the greater is the difference between profit in case of self-employment and wage in case of employment.

$^{38}$Note that conjectural variations are not constrained to be 'consistent' (Kamien and Schwartz (1983)).
actual number of firms is equal to the carrying capacity.

The demand elasticity of the carrying capacity, $\nu_Q$, is derived as follows:

$$
(5.9) \quad \nu_Q = \frac{\partial N^*}{\partial Q} \frac{Q}{N^*} = \frac{1}{2} - \frac{1 + \phi}{2(1 + \phi)^2 - 4(Q(1 + \phi)(1 + \phi))} = \frac{N^*e + 1 + \phi}{2N^*e + 1 + \phi}.
$$

An important implication of the oligopoly model is that $\nu_Q$ must be less than 0.5 because $N^*e < -1 - \phi$. From equations (5.7) and (5.9) it is easy to derive that $\nu_Q = -\eta_\alpha/\eta_N$ in the equilibrium situation. It can also be derived that in this situation equation (5.6) simplifies to $\eta_\alpha = (\pi^* + \alpha)/\pi^*$. 

5.3 Entry, exit and disequilibrium

The net entry of firms is assumed to be a reaction to a state of disequilibrium. This reasoning is in line with the Kirznerian notion of entrepreneurial activity originating from the existence of disequilibria characterized by the existence of profit opportunities (Ikeda (1990), Kirzner (1973, 1979)). We define disequilibrium as the discrepancy between the carrying capacity in a market in a certain period and the actual number of firms in that market. This notion of disequilibrium is strongly connected to the existence of profit opportunities. One derives from equations (5.5) and (5.8) using Taylor-expansion that

$$
(5.10) \quad \frac{\pi_i - \pi_i^*}{\pi_i^*} = \frac{1}{\nu_Q} \frac{N_i - N_i^*}{N_i^*}.
$$

when evaluated in $N_i = N_i^*$ where $i$ is the market index. Equation (5.10) shows the relation between two discrepancies: that between the carrying capacity and the actual number of firms in a market on the one hand and that between the exogenously given critical level of profits and the actual level of profits in a market on the other. Entrepreneurs are inclined not to accept a disequilibrium situation expected to arise in period $t$ because of the profit opportunities involved, but are hampered by costs of

---

38 Schmalensee (1992, p. 126) also derives a free-entry equilibrium number of firms. His expression also implies that a doubling of market demand less than doubles the equilibrium number of firms. Bresnahan and Reiss (1990) estimate the size of the market needed to support two automobile dealers to be larger than two times the size of the market needed to support a monopoly dealer. From the oligopoly model one would expect this size to be at least three times the size of the market needed to support a monopolist.
entering (or exiting) immediately in period $t$. These costs are assumed to rise quadratically because of congestion when many entrepreneurs want to enter or exit a market. That is, entrepreneurs minimize the following loss function (see Gilbert (1986) for similar analyses):

$$
(5.11) \quad \Lambda_{it} = \lambda_1 (E_{t-1}(\ln(N_{it}^*)) - \ln(N_{it}))^2 + \lambda_2 (\Delta \ln(N_{it}))^2 \quad \lambda_1, \lambda_2 > 0,
$$

where $E_{t-1}$ is the expectancy operator using information up till period $t-1$. Minimizing equation (5.11) with respect to $N_{it}$ yields

$$
(5.12) \quad \Delta \ln(N_{it}) = -\frac{\lambda_1}{\lambda_1 + \lambda_2} (E_{t-1}(\ln(N_{it}^*)) - \ln(N_{it-1}))
$$

We now assume that the expected carrying capacity in the succeeding period is equal to the capacity in the current period, i.e $E_{t-1}(N_{it}^*) = \ln(N_{it-1}^*)$. In the Appendix we discuss three alternatives to this assumption. The main conclusions of the study are not altered in any of these alternative cases. Using $\gamma = \lambda_1 / (\lambda_1 + \lambda_2)$ we obtain equation (5.13):

$$
(5.13) \quad \Delta \ln(N_{it}) = \gamma (\ln(N_{it-1}^*) - \ln(N_{it-1})) \quad \gamma > 0
$$

From this equation it is clear that net entry serves as an adjustment mechanism to a disequilibrium situation. The adjustment process towards the carrying capacity in the present study corresponds to that used in population ecology work (Hannan and Freeman (1977), p. 941).

We substitute for $N_{it}^*$ to arrive at a testable relation of the oligopoly model. Assuming the growth rate of the carrying capacity to be determined completely by changes in consumer demand, we obtain:40

$$
(5.14) \quad N_{it}^* = (1 + g_d) N_{it-1}^* = \left(1 + \nu \frac{\Delta Q_{it}}{Q_{it-1}}\right) N_{it-1}^*
$$

---

40In this study we concentrate on the effect of market demand on the carrying capacity. That is, we assume for simplicity that the price elasticity of demand, $e$, the conjectural variation, $\phi$, the parameters of the cost function, $\alpha$ and $\beta$, and the exogenously given critical level of profits, $\pi^*$, are constant. In retailing it is likely that the price elasticity of demand depends upon the average (regional) household income, that the critical level of profits depends upon average (regional) wages and that the parameters of the cost function depend upon average (regional) labour costs. From the beginning (1980) to the end (1988) of the period of investigation the general real wage index and the real wage index in retailing in the Netherlands declined just a few per cent.
We can rewrite equation (5.13) using equation (5.14) as follows:

\[(5.15) \quad \Delta \Delta \ln(N_{it}) = -\gamma(\ln(1 + g_{i,t-1}) - \Delta \ln(N_{i,t-1}))\]

Considering that \(\ln(1 + x) = x\) when \(x\) is close to zero and adding a disturbance term assumed to be independently and identically distributed over the markets and over the years, we have the following relation to be estimated:

\[(5.16) \quad \Delta \frac{N_{it}}{N_{i,t-1}} = v_{it} \left( \frac{\Delta Q_{i,t-1}}{Q_{i,t-2}} - \frac{\Delta N_{i,t-1}}{N_{i,t-2}} \right) + e_{it}\]

That is, the change in the net entry rate, \(\Delta N_{it}/N_{i,t-1}\), is equal to the adjustment rate, \(\gamma\), times the development over time of the level of disequilibrium plus a disturbance term.

The change in the net entry rate is the result of a change in the gross entry rate and a change in the gross exit rate. Gross entry and exit rates show substantial and persistent differences across industries (Dunne et al. (1988)). Especially in the retail sector, gross entry and exit rates are very stable over the years. This is a consequence of little change over time in the relative ease of entry and exit and of a continuously large pool of potential entrants for an industry. Each year a fairly constant flow of entrepreneurs would like to enter or to exit (retail) industries. This is partly caused by entrepreneurs retiring and being replaced by younger entrepreneurs. We assume that deviating from the stable gross entry and exit rate levels comes at a cost. This allows us to investigate the relative importance of entry and exit in the adjustment process. We extend the loss function \(\Lambda_{it}\) by considering loss occurring from the gross entry rate (\(ent_{it}\)) and the gross exit rate (\(ext_{it}\)) to be different from an industry-specific constant, denoted by \(d_{1i}\) and \(d_{2i}\), respectively:

\[(5.17) \quad \Lambda_{it} = \lambda_{1}(E_{i,t-1}(\ln(N_{it}))) - \ln(N_{i,t-1}) - ent_{it} + ext_{it})^2 + \lambda_{2}(ent_{it} - ext_{it})^2 +
\]
\[+ \lambda_{3}(ent_{it} - d_{1i})^2 + \lambda_{4}(ext_{it} - d_{2i})^2 \quad \lambda_{1}, \lambda_{2}, \lambda_{3}, \lambda_{4} > 0\]

Note that \(ent_{it} - ext_{it} = \Delta N_{it}/N_{i,t-1} = \Delta \ln(N_{it})\). In case \(\lambda_{3} = \lambda_{4}\), entry and exit differ in their sensitivity to a state of disequilibrium and, hence, in their relative importance in the adjustment process to a state of equilibrium. Solving for \(ent_{it}\) and \(ext_{it}\) from the first-

\[\footnote{Van Praag and Van Ophem (1995) find for example that there are almost seven times more individuals who wish to switch to self-employment than the actual number of switchers.}\]
order conditions gives the following two equations:

\[ (5.18) \quad \text{ent}_{ii} = \frac{\lambda_3(\lambda_1 + \lambda_2 + \lambda_4) d_{1i} + \lambda_4(\lambda_1 + \lambda_2) d_{2i} + \lambda_4 \lambda_3 E_{i-1}(\ln(N_n^i) - \ln(N_{i-1}^n))}{(\lambda_1 + \lambda_2)(\lambda_3 + \lambda_4) + \lambda_3 \lambda_4} \]

\[ (5.19) \quad \text{ext}_{ii} = \frac{\lambda_3(\lambda_1 + \lambda_2) d_{1i} + \lambda_4(\lambda_1 + \lambda_2 + \lambda_4) d_{2i} - \lambda_4 \lambda_3 E_{i-1}(\ln(N_n^i) - \ln(N_{i-1}^n))}{(\lambda_1 + \lambda_2)(\lambda_3 + \lambda_4) + \lambda_3 \lambda_4} \]

Using (5.14), taking the first difference and adding disturbance terms, we have the following two equations to be estimated:

\[ (5.20) \quad \Delta \text{ent}_{ii} = \gamma_q \left( v_q \frac{\Delta Q_{i,t-1}}{Q_{i,t-2}} - \frac{\Delta N_{i,t-1}}{N_{i,t-2}} \right) + e_{Eit} \]

\[ (5.21) \quad \Delta \text{ext}_{ii} = \gamma_x \left( v_q \frac{\Delta Q_{i,t-1}}{Q_{i,t-2}} - \frac{\Delta N_{i,t-1}}{N_{i,t-2}} \right) + e_{Xit} \]

There is a clear connection between the adjustment rates \( v_q \) and \( v_x \) and the parameters \( \lambda_3 \) and \( \lambda_4 \) of the loss function: \(-v_q/v_x = \lambda_4/\lambda_3\). In case this ratio exceeds one, the entry rate is more sensitive to a state of disequilibrium than the exit rate. The disturbance terms \( e_{Eit} \) and \( e_{Xit} \) probably are positively correlated because fluctuations in takeover activity are not incorporated in our model. The correlation between the disturbance terms gives an indication of the importance of replacement and displacement in an industry within the same period. Displacement and replacement which take more than one period are accounted for by the adjustment process. An entry which leads to an increase of the actual number of firms in excess of the carrying capacity is bound to lead to an exit in a subsequent period provided the carrying capacity does not grow. An exit reducing the actual number of firms below the carrying capacity paves the way for an entry in a subsequent period.

Equations (5.16), (5.20) and (5.21) should be estimated at the market level or, equivalently, at the level of a distinct organizational population. The firms which produce and sell a homogeneous good on the same geographical market are assumed to constitute a single population effectively isolated from interaction with other populations.
Only on this level of analysis the concept of a carrying capacity is useful.\textsuperscript{42} In our empirical application we will use aggregated data to estimate the parameters of the model and then use equations (5.7) and (5.9) to estimate the average (equilibrium) number of firms in the separate markets. A condition for using aggregated data to estimate parameters at the population (market) level is that the populations do not interact. In this case, we can change from market to industry index. This is viable for retail industries because firms which trade on distinct local retail markets usually barely compete. The parameters of the models when applied to the industry level are then weighted averages of these same parameters at the market level. We use aggregated data for two reasons.

First, the number of firms in a market is an integer and will generally be small. In most periods the net entry rate will be zero and in the periods when it is not, the difference from zero is necessarily large. This complicates the estimation of the parameters at a very disaggregated level of analysis. Second, in most applications, as in ours, only aggregated industry level data are available, for example because of the lack of clear market boundaries. In manufacturing this implies that several separate markets each with a limited number of competitors producing a homogeneous good are compressed into one manufacturing industry usually at best at the four-digit level. In distribution this implies that several geographically separated markets each with a limited number of local competitors are compressed into one retailing industry usually at best at the regional or state level. In both cases populations which differ with respect to either the good produced or the geographical scope are aggregated.

5.4 Data and characteristics of the retail sector

In this section we start with a short discussion of our data set for Dutch retailing. Subsequently, some remarks will be made on characteristics of retailing which are important to keep in mind when applying the model. For the retail sector we seek to estimate the effect of demand change on the carrying capacity, the speed of adjustment to an equilibrium situation, and the relative importance of entry and exit in this adjustment process. We use a panel data set for the Dutch retail trade. Despite its economic importance retailing has received only limited attention from industrial economists. The retail sector has a significant contribution to the economy. It accounts

\textsuperscript{42}See Hannan and Carroll (1992, ch. 7) for a discussion of the appropriate level of analysis of organizational populations for the U.S. brewing industry. In their survey of the influence of economics on sociology, Baron and Hannan (1994, pp. 1128-30) pay attention to the similarities between recent ecological work and research on the evolution of markets. See also Boone and Van Witteloostuijn (1995) and Van Witteloostuijn (1997b) for potentials for cross-fertilization between behavioural and economic models in the study of markets and organizations.
for about 23% of the total number of economically active enterprises and for about 13% of total labour force in the Dutch private sector in 1988 (Bode (1990)). These figures are in line with those of other countries of the European Union (EIM (1995)). Retailing is also prominent when we consider the vividity of entry and exit movements. On average annual entry and exit rates are almost 10% for the Dutch retail trade (see Table 5.1). This indicates that sunk entry and exit costs in Dutch retailing are low.

Our data are available per shoptype (retail industry) in which establishments sell relatively homogeneous goods. The source of the data on profits is an ongoing panel of independent, mainly small Dutch retailers. This panel is operated by the EIM Small Business Research and Consultancy in Zoetermeer. Each year and for each shoptype a questionnaire has been sent to a sample of shopkeepers and an average number of about seventy, averaged over the 1981-88 period and the 22 shoptypes in Table 5.1, have completed and returned it. The source of data on total consumer demand for product packages sold in the shoptypes is the Central Bureau of Statistics in Voorburg. The Central Registration Office (CRK) in The Hague provided data of entry and exit rates of establishments. These data may suffer from changes in counting procedures which could lead to "outlying" observations. We use the Jarque-Bera statistic to test for normality of the residuals of the regression equations. In case normality is rejected at a reasonable level of significance we also estimate the model without the "outlying" observations.\(^{43}\) For 22 shoptypes data are available in each of the three data sources for the period 1981 through 1988 (176 data points). The correlation between the entry and exit rates for this panel data set is 0.77. The average entry and exit rates over the 1981-88 period can be found in Table 5.1.

Retail markets differ from manufacturing markets. Michael Porter highlighted the most important elements of the structure of retailing as follows: "In contrast with typically low national concentration ratios for a given retail outlet class, the concentration of retail establishments in the relevant retail market is often high. Two to five retail establishments commonly make up such a market. The limited geographic extent of the market and the magnitude of demand within this market area impose a strong constraint on the maximum number of retailers. The equivalent constraint on a national manufacturing industry is much weaker, so that the number of sellers in a retail market is typically smaller than that in even a "tight oligopoly" in manufacturing. ... Locational proximity and substantial similarity of product lines promote the structural symmetry of

\(^{43}\)Wagner (1994a) suggested to apply both ordinary and reweighted least squares to regression equations with the entry (or exit) rate as the dependent variable. A weighting procedure is used to limit the influence of a small number of "outlying" observations on the regression results. The "outlying" observations which are removed in estimating the parameters of equations (5.16), (5.20) and (5.21) show large shifts in entry and exit rate which indicate changing counting procedures. Further inquiry at the CRK did not reveal what these procedures have been through each of the years of our sample.
competing retailers. Local demand trends, important input costs and other key structural market conditions are likely more similar than among manufacturers. Retail firms can quickly and accurately detect strategy changes by competitors, and the possibilities for secret changes are minimal. ... A straightforward application of oligopoly theory suggests, then, that mutual dependence recognized among retail competitors in a given retail market will be higher than in an equally concentrated manufacturing industry because detection and retaliation lags are low. ... It follows that competition may lack vigor and the chances for tacit agreement will be high." (Porter (1976), p. 13/14).

Table 5.1: Average yearly entry and exit rates in Dutch shop types (1981-1988)

<table>
<thead>
<tr>
<th>Shop type</th>
<th>ent</th>
<th>ext</th>
<th>Shop type</th>
<th>ent</th>
<th>ext</th>
</tr>
</thead>
<tbody>
<tr>
<td>bicycle stores</td>
<td>0.041</td>
<td>0.052</td>
<td>pet shops</td>
<td>0.089</td>
<td>0.086</td>
</tr>
<tr>
<td>tobacco shops</td>
<td>0.035</td>
<td>0.070</td>
<td>household goods</td>
<td>0.087</td>
<td>0.089</td>
</tr>
<tr>
<td>druggists</td>
<td>0.053</td>
<td>0.060</td>
<td>greengrocers</td>
<td>0.083</td>
<td>0.094</td>
</tr>
<tr>
<td>do-it-yourself shops</td>
<td>0.052</td>
<td>0.066</td>
<td>supermarkets</td>
<td>0.079</td>
<td>0.098</td>
</tr>
<tr>
<td>bakers</td>
<td>0.060</td>
<td>0.062</td>
<td>liquor stores</td>
<td>0.085</td>
<td>0.111</td>
</tr>
<tr>
<td>paint, glass, wall-paper</td>
<td>0.049</td>
<td>0.082</td>
<td>confectioners</td>
<td>0.092</td>
<td>0.107</td>
</tr>
<tr>
<td>dairy shops</td>
<td>0.051</td>
<td>0.081</td>
<td>furniture stores</td>
<td>0.100</td>
<td>0.105</td>
</tr>
<tr>
<td>shoe stores</td>
<td>0.071</td>
<td>0.073</td>
<td>fish shops</td>
<td>0.113</td>
<td>0.107</td>
</tr>
<tr>
<td>butchers</td>
<td>0.071</td>
<td>0.079</td>
<td>furnishing stores</td>
<td>0.109</td>
<td>0.138</td>
</tr>
<tr>
<td>photographer's shops</td>
<td>0.082</td>
<td>0.079</td>
<td>textiles mens wear</td>
<td>0.140</td>
<td>0.125</td>
</tr>
<tr>
<td>jewelleries</td>
<td>0.080</td>
<td>0.083</td>
<td>florists</td>
<td>0.154</td>
<td>0.132</td>
</tr>
</tbody>
</table>

Note: The entry and exit rate figures are yearly averages over the 1981-1988 period. The number of shop types is one less than in Chapter 4. The shop type 'furnishing stores, mainly clothing material' is left out.

The oligopoly model predicts that the market price level decreases as the number of independent firms increases (see equation (5.4)). This is due to the assumption of $\phi$ to be in excess of -1, at which value the model would become equivalent to Bertrand competition. From Porter’s remarks we would certainly expect $\phi$ to be well above -1 in retailing. Weiss (1989, chapter 9) provides a survey of empirical studies dealing with the relation between price and market concentration for U.S. food and gasoline retailing. All these studies indicate a clear positive effect of market concentration on prices.\footnote{A recent exception, however, is Newmark (1990) who finds that retail grocery prices are not significantly related to concentration.} For example, it is found for a data set covering 32 U.S. SMSAs in 1974 that supermarket price levels decline about 15% when going from a monopoly to a completely unconcentrated market. In the rest of the study we will assume Cournot competition.
(φ=0) although the empirical studies for (US) retailing only indicate a value of φ larger than -1.45

Because of the geographically dispersed demand the shoptypes can be divided into a large number of local oligopolies usually consisting of only a handful of establishments selling a relatively homogeneous good. This is a precondition for the use of the oligopoly model. However, we made two other important assumptions. First, we assumed a linear cost function. The linearity of the cost function for retail establishments is theoretically supported by Nooteboom (1982) and Frenk et al. (1991). See also Dean (1973), Douglas (1973), Nooteboom (1982) and Thurik (1984) for empirical support. Second, we assumed firms to have identical cost functions and conjectural variations and as a consequence to have equal production levels. Porter's description of the structure of retailing suggests that retail establishments are indeed confronted with very similar conditions. The average relative sizes, measured in number of employees, of entering and exiting firms in Dutch retailing compared to that of incumbents are about 0.55 and 0.70, respectively (Klomp and Thurik (1995)).46 This too provides at least some support for not discriminating in size between entering, exiting and incumbent firms.

5.5 Two implications of the oligopoly model

In this section we test two implications of the oligopoly model derived from the elasticities of profit equations (5.6) and (5.7). Therefore, we need an estimate of the cost function in the various shoptypes. Data of average costs and sales of (three) size classes (indexed by g) are available for 14 shoptypes from the EIM 1988 questionnaire. The estimation results for the linear cost function for these shoptypes in 1988 are similar across shoptypes.47 The estimates for α range between -32442 and -5715 with a mean of -21738 and the estimates for β range between 0.903 and 0.942 with a mean of 0.927. In the analysis we use the estimates when the data of the 14 shoptypes are pooled. The estimation results for the 14 shoptypes over three size classes can be found in equation (5.22). The results shown here imply that the cost-sales ratio is

45Aiginger (1996) finds empirical data on Austrian manufacturing firms to be to some extent compatible with and favourable to the basic Cournot model. Both market share and price elasticity seem to affect the profit margin of firms in the way predicted by Cournot.

46Dunne et al. (1988) report that relative sizes of entering and exiting firms in 2-digit U.S. manufacturing industries are much smaller, between about 0.11 and 0.52.

47Our definition of costs includes purchase value of products, wages for employees, rent, interest payments over debt, interest loss over equity, depreciation, transportation, energy, promotion and other costs. It does not include a reward for the entrepreneur.
increasing slightly with the size of the store (t-values between brackets).\textsuperscript{48}

\[(5.22) \quad C_{gi} = -19553 + 0.924 Q_{gi} \quad \quad R^2 = 0.9985 \quad \quad N = 42 \]

\[(4.4) \quad (165.9) \]

We proceed as follows. The demand elasticity \( \eta_Q \) and number of firms elasticity \( \eta_N \) are estimated for the panel data set. Two implications of the oligopoly model from equations (5.6) and (5.7), \( \eta_N < -2\eta_Q \) and \( \eta_Q < 1 \) (because \( \alpha < 0 \)), are then examined. From the discussion below equation (5.9) we know that \( \eta_Q = (\pi' + \alpha)/\pi' \) in equilibrium. We do not know the value of \( \pi' \) but it should be higher than the modal wage to compensate for the additional risk and working hours when compared to working as an employee. In 1988 the modal wage in the Netherlands was 46440 guilders.\textsuperscript{49} We therefore expect the demand elasticity of profit in equilibrium to exceed \((46440-19553)/46440 = 0.58 \). However, note that the estimate of \( \alpha \) was based upon data of a subset of 14 shootypes.

The estimates for the two elasticities, \( \eta_Q \) and \( \eta_N \), for the Dutch retail sector are based upon 22 shootypes for the 1981-88 period. The effect of the relative change in consumer demand for the products primarily sold in the shootype (\( \Delta Q_{lt}/Q_{lt-1} \)) on the relative change in average profit in the shootype (\( \Delta \pi_{lt}/\pi_{lt-1} \)) is taken as estimate for \( \eta_Q \).\textsuperscript{50} Similarly, the estimate for \( \eta_N \) is equal to the effect of the relative change in the number of firms in the shootype (\( \Delta N_{lt}/N_{lt-1} \)) on the relative change in average profit. That is, we estimate the 'total differential' equation:

\[(5.23) \quad \frac{\Delta \pi_{lt}}{\pi_{lt-1}} = \zeta + \eta_Q \frac{\Delta Q_{lt}}{Q_{lt-1}} + \eta_N \frac{\Delta N_{lt}}{N_{lt-1}} + \xi_{it} \]

The least squares estimates of \( \eta_Q \) and \( \eta_N \) can be found in Table 5.2. From the Jarque-Bera statistic in the first column of the table it is clear that the residuals are highly non-

\textsuperscript{48}Equation (5.22) appears not to suffer from heteroskedasticity. See the following regression results:

\[\frac{C_{g}}{Q_{g}} = 0.923 - 19204 (1/Q_{g}) \quad \quad R^2 = 0.4837 \]

\[(130.0) \quad (5.1) \]

\textsuperscript{49}For the 22 shootypes in our data set the average profit is equal to 69017 Dutch guilders in 1988. The average reward for entrepreneurial activity in Dutch retailing is therefore about 50% more than for the activities of an (modal) employee. This is not unreasonable as starting a retail venture is relatively simple when compared to starting ventures in most other sectors of the economy. See also Carree and Thurik (1994).

\textsuperscript{50}Both the consumer demand and the average profit are divided by a general consumer price index to correct for growth in these variables simply due to inflation.
normally distributed. This is caused by one outlier (liquor stores, 1981) with an estimated residual of +1.87, while the second largest estimated residual is +0.64. After removing this observation non-normality cannot be rejected anymore. The estimates of \( \eta_q \) and \( \eta_n \) are 0.71 and -1.87, respectively. The estimate for \( \eta_q \) is below unity, while the estimate for \( \eta_n \) is -2.64 times that of \( \eta_q \). We conclude that reactions of profits on changes in consumer demand and number of firms are in line with oligopolistic behaviour.

From equation (5.7) we estimate \( \kappa_c \) to be -2.55 on average in Dutch retailing (in case of Cournot oligopoly: \( \phi = 0 \)). Considering that the price elasticities of demand in retailing are roughly between -0.4 and -0.9 (Zeelenberg (1986), p. 116), the average market sizes of the local oligopolies in retailing are estimated to be between 3 and 7 firms. This can be compared to the estimated equilibrium market sizes derived in the next section. In case the number of firms is equal to the carrying capacity \( v_q \) equals 0.378.\(^{51}\)

Table 5.2: Estimates of the determinants of \( \Delta \pi_t / \pi_{t-1} \) using equations (5.23) and (5.24)

<table>
<thead>
<tr>
<th></th>
<th>( \zeta )</th>
<th>( \eta_q )</th>
<th>( \eta_n )</th>
<th>( \eta_{nf} )</th>
<th>( \eta_{nx} )</th>
<th>JB</th>
<th>R²</th>
<th>( X^2_{\eta_{nf} - \eta_{nx}} )</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.035</td>
<td>0.815</td>
<td>-2.216</td>
<td>-2.206</td>
<td>2334.76</td>
<td>0.066</td>
<td>0.21</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(1.7)</td>
<td>(2.2)</td>
<td>(2.8)</td>
<td>(2.8)</td>
<td>(2.5)</td>
<td>(0.000)</td>
<td>(0.21)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.029</td>
<td>0.708</td>
<td>-1.872</td>
<td>1.859</td>
<td>6.21</td>
<td>(0.045)</td>
<td>0.071</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(1.7)</td>
<td>(2.4)</td>
<td>(2.9)</td>
<td>(2.9)</td>
<td>(2.7)</td>
<td>(0.000)</td>
<td>(0.071)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.010</td>
<td>0.801</td>
<td>2.490</td>
<td>2.225</td>
<td>2372.27</td>
<td>0.067</td>
<td>0.49</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.2)</td>
<td>(2.2)</td>
<td>(2.5)</td>
<td>(2.7)</td>
<td>(0.000)</td>
<td>(0.037)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>-0.004</td>
<td>0.689</td>
<td>2.225</td>
<td>2.225</td>
<td>6.60</td>
<td>0.074</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note: Absolute t-values between brackets. JB stands for the Jarque-Bera test statistic on normality of the residuals (p-value between brackets). In the second and fourth column one outlier has been removed to guarantee (near-)normality.

\(^{51}\)The quality of this estimate of \( v_q \) in an equilibrium situation is not much affected by a strong correlation between the estimates of \( \eta_q \) and \( \eta_n \). This correlation is -0.117.
We may also consider the effects of entry and exit separately. The results are presented in the last two columns of Table 5.2 both with and without the outlying observation. The equation to be estimated reads:

\[
(5.24) \quad \frac{\Delta n_{it}}{n_{i,t-1}} = \zeta + n_{t} \frac{\Delta Q_{it}}{Q_{i,t-1}} + n_{N_e} ent_{it} + n_{N_x} ext_{it} + \theta_{it}
\]

The variables \(ent_{it}\) and \(ext_{it}\) stand for gross entry and exit rates, respectively. Of course, equation (5.23) is a special case of equation (5.24), where \(n_{N_e} = - n_{N_x}\). It appears that an exit has a somewhat larger effect on changes in average profits than an entry. This may be a consequence of exiting firms to be somewhat larger in size on average. However, note that the difference of the magnitude of the effects is not statistically significant. Entry and exit appear to be symmetric in their effect on changes in the average profit in a shopotype.

5.6 Empirical results for entry and exit flows in retailing

In this section the empirical results of equations (5.16), (5.20) and (5.21) are presented. This provides us with insight in the effect of demand growth on carrying capacity and the speed of adjustment in the Dutch retail trade. We study the relative importance of gross entry and exit in the adjustment process by estimating their sensitivities to a state of disequilibrium. We will also investigate whether the estimates differ between shopotypes with generally small shops and shopotypes with somewhat larger shops on average. The results for the least squares estimation of equation (5.16) can be found in Table 5.3.

Only after removing four outlying observations (liquor stores, 1987; furnishing stores, 1985; confectioners, 1985; photographer's shops, 1983), the Jarque-Bera test on normality could not be rejected at a 1%-significance level. However, estimation results barely changed as a result of this. See the first and second column of Table 5.3, respectively. The estimates for \(\gamma\) and \(\nu_{0}\) are 0.387 and 0.303, respectively. The estimate of the demand elasticity is below the value expected at equilibrium of 0.378. This implies that the equilibrium size of markets is smaller than the actual size of markets in Dutch retailing. Because \(N^* c\) is estimated to be -1.77 using equation (5.9), we expect equilibrium market sizes in retailing to be between 2 and 5 firms (in case of Cournot oligopoly: \(\phi = 0\)). In the previous section we found the actual market sizes to be somewhat larger. This corresponds to the decrease in the total number of firms in retailing in the period 1981 through 1988 (Nootbooom (1986a)).
Table 5.3: Estimates of adjustment rate and demand elasticity using equation (5.16)

<table>
<thead>
<tr>
<th>( Y )</th>
<th>0.381 (6.5)</th>
<th>0.387 (7.9)</th>
<th>0.391 (8.0)</th>
</tr>
</thead>
<tbody>
<tr>
<td>( Y_S )</td>
<td>0.453 (6.9)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( Y_L )</td>
<td>0.322 (5.4)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( V_Q )</td>
<td>0.301 (3.5)</td>
<td>0.303 (4.2)</td>
<td>0.322 (4.4)</td>
</tr>
<tr>
<td>( V_{Q_S} )</td>
<td></td>
<td></td>
<td>0.397 (3.8)</td>
</tr>
<tr>
<td>( V_{Q_L} )</td>
<td></td>
<td></td>
<td>0.235 (2.8)</td>
</tr>
<tr>
<td>( JB )</td>
<td>142.25 (0.000)</td>
<td>2.21 (0.332)</td>
<td>1.86 (0.395)</td>
</tr>
<tr>
<td>( R^2 )</td>
<td>0.255</td>
<td>0.324</td>
<td>0.334</td>
</tr>
<tr>
<td>( X^2_{Y_S-Y_L} )</td>
<td></td>
<td></td>
<td>2.70</td>
</tr>
<tr>
<td>( X^2_{V_{Q_S}-V_{Q_L}} )</td>
<td></td>
<td></td>
<td>1.82</td>
</tr>
</tbody>
</table>

Note: Absolute t-values between brackets. JB stands for the Jarque-Bera test statistic on normality of the residuals (p-value between brackets). With the exception of the first column four outliers have been removed to guarantee (near-)normal residuals.

Equation (5.16) is estimated assuming that all shoptypes have equally sized equilibrium market sizes. Some shoptypes may however have smaller equilibrium market sizes due to higher entry and exit barriers or important scale advantages. These shoptypes are expected to have a lower adjustment rate and a lower demand elasticity \( v_Q \). A variable which may effectively distinguish between shoptypes with relatively high and relatively low barriers and scale advantages is average floorspace in a shoptype. For example, tobacco shops and dairy shops had an average floorspace of only 103 and 138 square metres in 1988. On the other hand, furnishing stores with mainly furniture and furnishing stores had an average of 3319 and 2634 square metres in that year. The average floorspace in the data set is 558 square metres at a standard deviation of 671. In the third and fourth column of Table 5.3 we divide our sample into two equally sized subsamples: one sample with average floorspace smaller than 309 square metres and one sample with average floorspace larger than 309 square metres. The results show that the speed of adjustment is about 40% higher in shoptypes with small floorspace. The
difference is however not statistically significant. The elasticity of demand may also be estimated separately for the two groups. The fourth column of the table shows that this elasticity is higher for shootypes with little floorspace. The estimate of \( v_Q \) in this sample of (very) low barriers and scale advantages is even larger than the value expected at equilibrium of 0.378. The difference between the elasticities is however again not significant.

We now turn to the relative importance of the entry and exit rate in the adjustment process. The equations (5.20) and (5.21) are estimated with Iterative SUR with demand elasticity \( v_Q \) assumed to be equal in the two equations. The estimation results can be found in Table 5.4. Both the residuals of the gross entry and gross exit rate equation were non-normally distributed. Only after removing seven outlying observations (furniture stores, 1988; furnishing stores, 1985; florists, 1982 and 1985; butchers, 1988 and fish shops, 1982 and 1985), the Jarque-Bera test on normality could not be rejected for both equations at a 1% significance level. From the second column of Table 5.4 we find estimates of \( v_E, v_X \) and \( v_Q \) of 0.227, -0.137 and 0.266, respectively. These results show a higher adjustment rate for entry than for exit, but not significantly so. The correlation between the residuals of equation (5.20) and (5.21) is 0.220 and positive as expected.

We have also investigated the difference between shootypes with relatively little floorspace and much floorspace using the same distinction as in Table 5.3. Results in the third column (after removing outlying observations) show that adjustment for a state of disequilibrium in shootypes with small floorspace is completely dominated by changes in the entry rate. A state of disequilibrium seems not to have an effect on the exit rate. This suggests that in case the number of firms is higher than the carrying capacity a negative net entry rate is caused by a decline in the gross entry rate rather than by an increase in the gross exit rate. The results for the shootypes with the larger shops are quite different. There seems to be no difference in the ability of the entry or exit rate to generate adjustment in the direction of equilibrium for these shootypes.

The relative importance of entry and exit in the adjustment process to a state of equilibrium can be estimated as follows. Using the estimates 0.227 and -0.137 for \( v_E \) and \( v_X \), we are able to derive an estimate of the ratio \( \lambda_4 / \lambda_3 \) of 1.66. The penalty on the deviation of the entry rate from the desired level is somewhat smaller than that of the exit rate. Some evidence for the larger flexibility of the entry rate can also be found by simply considering the correlation between the net entry rate, \( \text{ent}_{it} - \text{ext}_{it} \), and the deviation of the gross entry and exit rates from their averages over time, \( \text{ent}_{it} - \text{ent}_{i} \) and \( \text{ext}_{it} - \text{ext}_{i} \). These correlations are 0.50 and -0.17, respectively. An increase in the number of firms in a shootype is therefore more likely to be accompanied by a relatively high entry rate than by a relatively low exit rate.
Table 5.4: Estimates of adjustment rates of gross entry and exit rates using equations (5.20) and (5.21)

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Estimation</th>
<th>Standard Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>$v_E$</td>
<td>0.206</td>
<td>0.227</td>
</tr>
<tr>
<td></td>
<td>(3.7)</td>
<td>(5.5)</td>
</tr>
<tr>
<td>$v_{Es}$</td>
<td></td>
<td>0.332</td>
</tr>
<tr>
<td></td>
<td>(5.8)</td>
<td></td>
</tr>
<tr>
<td>$v_{El}$</td>
<td></td>
<td>0.133</td>
</tr>
<tr>
<td></td>
<td>(2.6)</td>
<td></td>
</tr>
<tr>
<td>$v_X$</td>
<td>-0.170</td>
<td>-0.137</td>
</tr>
<tr>
<td></td>
<td>(3.6)</td>
<td>(3.7)</td>
</tr>
<tr>
<td>$v_{Xs}$</td>
<td></td>
<td>-0.061</td>
</tr>
<tr>
<td></td>
<td>(1.2)</td>
<td></td>
</tr>
<tr>
<td>$v_{XL}$</td>
<td></td>
<td>-0.197</td>
</tr>
<tr>
<td></td>
<td>(3.9)</td>
<td></td>
</tr>
<tr>
<td>$v_Q$</td>
<td>0.314</td>
<td>0.266</td>
</tr>
<tr>
<td></td>
<td>(3.6)</td>
<td>(3.3)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(3.7)</td>
</tr>
<tr>
<td>JBE</td>
<td>489.91</td>
<td>2.97</td>
</tr>
<tr>
<td></td>
<td>(0.000)</td>
<td>(0.226)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.538)</td>
</tr>
<tr>
<td>JBX</td>
<td>352.38</td>
<td>5.17</td>
</tr>
<tr>
<td></td>
<td>(0.000)</td>
<td>(0.075)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.375)</td>
</tr>
<tr>
<td>$R^2_E$</td>
<td>0.051</td>
<td>0.098</td>
</tr>
<tr>
<td>$R^2_X$</td>
<td>0.066</td>
<td>0.070</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.091</td>
</tr>
<tr>
<td>corr($\hat{\epsilon}_E$, $\hat{\epsilon}_X$)</td>
<td>0.540</td>
<td>0.220</td>
</tr>
<tr>
<td>$x^2_{vE-vX}$</td>
<td>0.17</td>
<td>2.39</td>
</tr>
<tr>
<td>$x^2_{vEs-vXs}$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$x^2_{vEl-vXl}$</td>
<td></td>
<td>0.75</td>
</tr>
</tbody>
</table>

Note: Absolute t-values between brackets. With the exception of the first column seven outliers have been removed to guarantee (near-)normal residuals. A star (*) means significant at a 5% significance level. JBE and JBX stand for the Jarque-Bera test statistics on normality of the residuals in the entry and exit equations, respectively (p-values between brackets).

5.7 Summary and discussion

Entry and exit are reactions to (the lack of) profit opportunities. In this study we use this simple notion to propose a new and more uniform approach to the study of entry and exit processes. Following Kirzner, these opportunities may be considered market
"errors". In this study we formalize this line of reasoning by using an error correction mechanism to explain the extent of entry and exit flows in retailing. This is done using the notion of carrying capacity. The carrying capacity is that number of firms at which there is no incentive for net entry (entry minus exit). This is a special case of the general meaning of carrying capacity in the population ecology literature, which is the capacity of the environment to support a certain population size. In case the actual number of firms in an industry is below the carrying capacity, an increase in the net entry rate adjusts for that disequilibrium. This increase may be the result of an increase in the entry rate or a decrease in the exit rate. The present study develops an expression for the carrying capacity from oligopoly theory. Obviously, changes in consumer demand are an important determinant of changes in this capacity. This study tests the theoretical model for Dutch retailing and gives estimates of the effect of demand change on the carrying capacity, of the speed of adjustment to an equilibrium situation, and of the relative importance of entry and exit in this adjustment process. A panel data set of 22 retail industries for the 1981-88 period is used for this purpose.

Estimation results do not contradict the use of the oligopoly model at local market levels which are typical for the retail sector. The demand elasticity of the carrying capacity in retailing is estimated to be about 0.3. This estimate is comparable to estimates of the effect of demand growth on net entry found in studies on U.S. manufacturing (Acs and Audretsch (1989b) and Hirschey (1981)). The speed of the adjustment process to a situation where the total number of firms is equal to the carrying capacity is estimated to be about 40% per year in retailing. This corresponds to complete adjustment in about 5 years, which is consistent with empirical evidence found by Levy (1987) for U.S. manufacturing. The role of change in the entry rate as a means of adjusting to disequilibrium for shoetypes with mainly small stores was found to be much more important than the role of the change in the exit rate. MacDonald (1986) also finds for 46 U.S. manufacturing industries over the period 1976-82 that the effect of growth on the entry rate is much larger than on the exit rate. The role of industry profitability and growth appear to play a more important role, on average, for entry decisions than for exit decisions. Although Dutch retailing and U.S. manufacturing differ, empirical results for these two sectors are, by and large, very much in line.

5.A Appendix

In Section 5.3 we assume that the expected carrying capacity is equal to the current carrying capacity, i.e. \( E_t \cdot \ln(N_t^*) \ln(N_t) = \ln(N_{t-1}) \). As a consequence, the adjustment process towards equilibrium corresponds to that used in population ecology work. However, the assumption implies that entrepreneurs do not expect consumer demand to grow. In this appendix we discuss some alternative adjustment processes which result from the
expected growth in consumer demand to depend on the current growth in consumer demand.

Denote the expected growth rate of the carrying capacity from period \( t-1 \) to period \( t \) by \( E_{t-1}g_{it} \). The expected carrying capacity is then equal to

\[
(A1) \quad E_{t-1}\ln(N_{it}) = \ln((1 + E_{t-1}g_{it})N_{it-1})
\]

The expected growth rate of the carrying capacity is assumed to be determined by a shoptype-specific element and the growth rate in the current period, i.e. 
\( E_{t-1}g_{it} = \delta_{ol} + \delta_{i}g_{lt-1} \). Using \( g_{it} = v_{Q} \frac{\Delta Q_{lt}}{Q_{lt-1}} \), a generalization to equation (5.16) becomes:

\[
(A2) \quad \frac{\Delta N_{it}}{N_{it-1}} = \gamma \left( v_{Q} \left( 1 + \delta_{i} \right) \frac{\Delta Q_{lt-1}}{Q_{lt-2}} - \delta_{i} \frac{\Delta Q_{lt-2}}{Q_{lt-3}} \right) - \frac{\Delta N_{lt-1}}{N_{lt-2}} + \epsilon_{it}
\]

We discuss four alternatives for choosing the value of \( \delta_{i} \):

(i) The expected growth in consumer demand does not depend on the current growth in consumer demand. That is, \( \delta_{i} = 0 \). In this case equation (A2) is identical to equation (5.16).

(ii) The expected growth in consumer demand equals the current growth in consumer demand. That is, \( \delta_{i} = 1 \).

(iii) Entrepreneurs use their previous experience of the relation between the growth in consumer demand in two consecutive years. The estimation results for the 22 shoptypes over the 1981-88 period with and without shoptype-specific constants are as follows (t-values between brackets):

\[
(A3) \quad \frac{\Delta Q_{lt}}{Q_{lt-1}} = \delta_{ol} + 0.346 \frac{\Delta Q_{lt-1}}{Q_{lt-2}} \quad R^{2} = 0.250
\]

\[
(A4) \quad \frac{\Delta Q_{lt}}{Q_{lt-1}} = 0.006 + 0.425 \frac{\Delta Q_{lt-1}}{Q_{lt-2}} \quad R^{2} = 0.193
\]

The hypothesis of equal shoptype-specific constants (\( \delta_{ol} = \delta_{o} \)) cannot be rejected at a 10%-significance level (\( \chi^{2}(21) = 12.88 \)). Therefore, we choose \( \delta_{i} \) equal to 0.425.
(iv) Estimate \( \delta_1 \) directly from equation (A2). The way in which entrepreneurs expect consumer demand to change is thus derived from the way in which they react to this expected change in demand by entering and/or exiting the shop type.

In Table A.1 we show the estimation results for equation (A2) for these four different alternatives. We use data for the 22 shop types over the 1982-88 period (154 data points). In the lower part of the table we show the results in case the four outliers (liquor stores, 1987; furnishing stores, 1985; confectioners, 1985; photographer's shops, 1983) are removed from the sample.

From Table A.1 we find that a positive value of \( \delta_1 \) may lead to a lower estimated adjustment parameter \( \gamma \) and a lower estimated demand elasticity \( v_q \) when compared to the case of \( \delta_1 = 0 \). The results also show that the variance explained is highest for values of \( \delta_1 \) between about 0.25 and 0.5. However, the large standard error of the estimated value of \( \delta_1 \) in case of alternative (iv) indicates that the data cannot discriminate between the four alternatives.

Table A.1: Estimation results with four alternative choices for expected demand growth

<table>
<thead>
<tr>
<th>Alternative</th>
<th>(i)</th>
<th>(ii)</th>
<th>(iii)</th>
<th>(iv)</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \gamma )</td>
<td>0.399</td>
<td>0.370</td>
<td>0.381</td>
<td>0.380</td>
</tr>
<tr>
<td>(6.2)</td>
<td>(5.7)</td>
<td>(6.0)</td>
<td>(5.8)</td>
<td></td>
</tr>
<tr>
<td>( v_q )</td>
<td>0.310</td>
<td>0.188</td>
<td>0.261</td>
<td>0.258</td>
</tr>
<tr>
<td>(3.3)</td>
<td>(3.1)</td>
<td>(3.3)</td>
<td>(2.5)</td>
<td></td>
</tr>
<tr>
<td>( \delta_1 )</td>
<td>0</td>
<td>1</td>
<td>0.425</td>
<td>0.452</td>
</tr>
<tr>
<td>R²</td>
<td>0.264</td>
<td>0.268</td>
<td>0.272</td>
<td>0.272</td>
</tr>
</tbody>
</table>

| \( \gamma \) | 0.408 | 0.380 | 0.389 | 0.394 |
| (7.6) | (7.0) | (7.3) | (7.2) |
| \( v_q \) | 0.313 | 0.180 | 0.256 | 0.276 |
| (4.1) | (3.7) | (4.0) | (3.3) |
| \( \delta_1 \) | 0 | 1 | 0.425 | 0.293 |
| R² | 0.339 | 0.334 | 0.344 | 0.344 |

Note: Absolute t-values between brackets. In the upper part of the table results for the complete data set (154 data points) are reported while in the lower part of the table results are reported after removing four outlying observations.
CHAPTER 6

SMALL FIRMS AND ECONOMIC GROWTH

Economic activity in manufacturing industries moved away from large towards small firms in relative terms in many OECD-countries during the last two decades. However, the speed of this industrial transformation process has varied considerably across countries and industries. This study investigates the consequence of lagging behind in this restructuring process in manufacturing. A sample of 14 manufacturing industries in 13 European countries has been constructed for this purpose. It is found that, on average, the employment share of large firms in 1990 has a negative effect on growth of output in the subsequent four-year period. This provides support for specific policies introduced during the 1980s in European countries stimulating small enterprises. A second sample consists of 71 industries in Dutch manufacturing. Results indicate that industries in which small firms grew faster than their larger counterparts during the 1978-87 period experienced more labour productivity growth than industries in which larger firms have been able to maintain or expand their share of industry employment.
6.1 Introduction

There is substantial evidence that economic activity moved away from large firms to small firms in the 1970s and 1980s. Acs and Audretsch (1993b) and Carlsson (1992) provide an overview of evidence concerning manufacturing industries in countries in varying stages of economic development. Carlsson mentions two explanations for this shift. The first deals with fundamental changes occurring in the world economy from the 1970s onwards. These changes relate to the intensification of global competition, the increase in the degree of uncertainty and the growth in market fragmentation. The second deals with changes in the character of technological progress. He shows that flexible automation has various effects resulting in a shift from large to smaller firms. The pervasiveness of changes in the environment, i.e. the world economy, and in the direction of technological progress result in a structural shift which affects the economies of all industrialised countries. Piore and Sable (1984) also argue that the market instability resulted in the demise of mass production and promoted flexible specialization. This fundamental change in the path of technological development led to the occurrence of vast diseconomies of scale.

This shift away from large firms is not confined to manufacturing industries. Brock and Evans (1989) show that this trend has been economy-wide at least for the United States. They provide us with four more reasons as to why this shift has occurred: the increase of labour supply; changes in consumer tastes; relaxation of (entry) regulations; and the fact that we are in a period of creative destruction. Loveman and Sengenberger (1991) stress the influence of two trends of industrial restructuring: that of decentralisation and vertical disintegration of large companies and that of the formation of new business communities. Furthermore they emphasize the role of public and private policies promoting the small business sector. De Koning en Snijders (1992) provide an overview of the various public policies in countries of the European Union which have been introduced during the 1980s.

The consequences of the shift of economic activity from large to smaller businesses have been discussed by Acs (1992). His claims are that small firms play an important role in the economy serving as agents of change by their entrepreneurial activity, being the source of considerable innovative activity, stimulating industry evolution and creating an important share of the newly generated jobs. Baumol (1990) amply deals with the role of entrepreneurial activities. The role of smallness in the process of innovative activities is investigated extensively in Acs and Audretsch (1990b), Audretsch (1995)

---

52 Entrepreneurship is not necessarily confined to small firms. Large enterprises seek to promote entrepreneurship by creating largely independent business units. Burgeleman (1994) discusses possibilities for large firm entrepreneurship.
and Cohen and Klepper (1992a, 1996a). The discussion of the relation between the role of small firms and industry dynamics is spread out: examples can be found in Audretsch (1993, 1995). The role of small firms in the job creation process is controversial. See Davis et al. (1996) and Chapter 7 for a recent discussion.

In this chapter we concentrate on the question whether the size class structure of industries has affected their economic performance, more specifically their economic growth. The list of consequences as mentioned above, suggests such a relationship. Clearly, the question neglects to a large extent the various roles of the intermediary variables between the size class structure and economic growth. Some preliminary results for the relation between changes in the firm-size distribution and economic growth are presented by Thurik (1995, 1996). The analysis shows a positive effect of an increase in the economy-wide share of small firms on growth in GDP. The interpretation of this result is somewhat difficult because it is not clear whether changes in the economy-wide share of small firms mainly result from changes in the sectoral composition or from downscaling in the specific industries. Schmitz (1989) presents an endogenous growth model which relates entrepreneurial activity and economic growth. He shows that an increase of the proportion of entrepreneurs in the working force leads to an increase in long-run economic growth. His model also implies that the equilibrium fraction of entrepreneurs is lower than the social optimal level, providing a rationale for policies stimulating entrepreneurial activity. The size class structure of an industry and the proportion of entrepreneurs in its working force are strongly related. The current study therefore also provides a preliminary empirical test of the Schmitz-model at the industry level. Recent studies on the relation between industry structure and performance are Nickell (1996) and Nickell et al. (1997) who present evidence that competition, as measured by increased number of competitors, has a positive effect on the rate of total factor productivity growth.

The empirical analyses in this chapter are preceded by a tentative model of the effect of different market settings on industry structure and performance. The model is a pure selection model, in which entrepreneurs are assumed not to learn or innovate. Despite the simplicity of the model assumptions the predicted relationship between the number of entrepreneurs and economic performance is far from simple. It depends upon the particular group of market participants. Section 6.3 discusses the data set of large firm presence in European manufacturing. Empirical results of the effect of the degree of large firm presence on real output growth are provided in Section 6.4. The second empirical examination can be found in Section 6.5. The effect of changes in market structure on labour productivity in Dutch manufacturing is subject of the section. Section 6.6 provides some concluding remarks on the role of smallness in promoting economic growth in modern manufacturing industries.
6.2 Competitive selection with differing rates of entry and adjustment

Consider a population of $N$ persons who choose between being an entrepreneur and being an employee. Each person has an 'entrepreneurial ability' $e_i$ which can be used in combination with $L_i$ employees earning a wage $w$ to produce an output of $e_i L_i^\beta$ with $0<\beta<1$.\textsuperscript{53} Taking the price of the good to be unity total profit will be $\pi_i = e_i L_i^\beta - w L_i$. From the first order condition it is easy to find that the optimal level of labour input and profit are equal to

$$L_i^* = \left( \frac{\beta e_i}{w} \right)^{\frac{1}{1-\beta}}$$

$$\pi_i^* = (1-\beta) e_i \left( \frac{\beta e_i}{w} \right)^{\frac{\beta}{1-\beta}}$$

Entrepreneurs exit the industry in case the optimal level of profit is below the wage level $w$. These entrepreneurs have no long term prospect of their venture being profitable. That is, firms remain incumbent if

$$e_i > \frac{w}{\beta^\beta (1-\beta)^{1-\beta}}$$

The extent of entry and mobility are built into the model in the following way. Each period every person has a probability $p_i$ of (re-)entering the industry. These probabilities may be equal (to $p$) across persons or may depend on their ability as $p$ times the ratio of $e_i$ over the mean of the abilities in the population. In the second case entrepreneurs are assumed aware to some extent of their own chances of success in the market. The amount of labour entrepreneurs use is assumed to adjust gradually to the optimal level:

$$L_i^* = L_{i,t-1} + \lambda (L_{i,t}^* - L_{i,t-1})$$

The equilibrium wage level is derived from the condition that the demand and supply of labour are identical. Let $M$ be the total number of entrepreneurs and $\Theta$ the set of entrepreneurs, then we have:

\textsuperscript{53} Some other models in which heterogeneous ability is an essential feature are Lucas (1978), Gifford (1993) and Klepper (1996).
\[
N - M = \sum_{i \in \Theta} L_{it} \iff N - M = (1 - \lambda) \sum_{i \in \Theta} L_{Ii+1} + \lambda \sum_{i \in \Theta} L_{it} \iff w = \beta \left( \frac{\lambda \sum_{i \in \Theta} e_i^{1 - \beta}}{N - M - (1 - \lambda) \sum_{i \in \Theta} L_{Ii+1}} \right)^{1 - \beta}
\]

Competitive selection proceeds through entry, exit and mobility of firms. The intensity of the competitive selection process in this model depends on the probability of entry, \(p\), and the adjustment rate, \(\lambda\). We use some simple simulations to show the effect of this intensity on the market performance, measured by total market output. We first introduce the market settings which are identical across the scenarios and then we discuss the specific assumptions about the particular group of market participants. The abilities are assumed to be distributed as a linear combination of two independent lognormally distributed variables \(X_1\) and \(X_2\) with the same mean and variance. The weighting factor is \(\alpha_i\). That is, \(e_i = \alpha_i x_{1i} + (1 - \alpha_i) x_{2i}\), where \(x_{1i}\) and \(x_{2i}\) are the realizations of \(X_1\) and \(X_2\) for person \(i\). Note that the mean of the abilities is independent of \(\alpha\), while the variance is at a minimum when \(\alpha_i = 0.5\) and at a maximum when \(\alpha_i\) is either zero or one. The variables \(X_1\) and \(X_2\) stand for two ‘underlying’ abilities which combine into the entrepreneurial ability necessary to be successful in the industry.\(^{54}\) Gifford (1993) also distinguishes between two kinds of ability: ability to innovate (also used in Klepper (1996)) and ability to manage (also used in Lucas (1978)).

In the rest of this section we will discuss several scenarios. In all of these scenarios the mean and variance of the lognormal distribution are \(e^{0.5}\) and \((e^{0.25} - 1)/e\). In the first period there is a probability of 0.1 for each person to be one of the initial entrepreneurs. The value of \(\beta\) is taken as 0.8. Changing the value of \(\beta\) over time would certainly have an impact on the firm-size distribution. However, in this model we concentrate on selection instead of on changes in production technologies or organizational forms. The number of persons \(N\) is fixed at 10,000. Changing these values of the market settings within reasonable borders does not affect the basic conclusions that follow.

The scenarios

We will show five scenarios out of eight which are derived by choosing three times between two alternative assumptions. The three other possible scenarios are left out because they do not provide additional insights. The first choice is between:

---

\(^{54}\)The impact of (characteristics of) top managers on organizational outcomes is subject of a large literature. Boone et al. (1996) provide a short review, an integrative framework and empirical test for the Flemish furniture industry consisting of many small firms.
(a) Nobody enters or leaves the population.
(b) Each period 200 persons (2%) are removed from the sample and 200 new ones are added. The ordering in which this is executed is determined before the simulation.

We first choose between having the same group of market participants throughout the entire simulation period or having this group 'refreshed' each period as some retire while others start their working career. The first option is viable only when the firm 'inherits' the ability of the founder. Successive entrepreneurs leading the firm are then assumed to be influenced strongly by their predecessors. See also Cohen and Klepper (1992b) and Klepper (1996) who suggest that R&D-related capabilities are firm specific. The second choice is between:

(c) The weighting factor $\alpha$, is 0.75 and constant in time.
(d) The weighting factor $\alpha$, starts at 0.5, increases with constant steps to 0.9 fifty periods later, then decreases again with the same steps to 0.1 one hundred periods later, etc.

Taking a constant $\alpha$, implies that entrepreneurial ability is also constant over time. Changes in the production process or market environment may shift the relevant abilities somewhat over time. In the second option $\alpha$, changes with 0.008 each period. It implies that the importance of $X_1$ and $X_2$ and therefore the set of individuals with the highest abilities also changes over time. The second alternative in both the first and second choice creates continuous incentives for displacement and replacement. Entrepreneurs who retire are replaced and entrepreneurs whose abilities decrease below a certain level due to a changing weighting factor are displaced. The third choice is between:

(e) Each person has an identical probability, $p$, of entry.
(f) A person's probability of entry is equal to $pe^0.5$, i.e. $p$ times $e^0.5$ divided by the mean of the lognormal distributions $e^0.5$.

It is likely but not certain that the probability of entry is a positive function of entrepreneurial ability. The question is one of perception. Do individuals with above (below) average ability perceive this correctly? If so, then the second option seems applicable. If not, then the first option might be more realistic. Table 6.1 shows the five scenarios for the adjustment rate equal to 0.15 and 0.05 and the probability of entry equal to 0.03 and 0.01.

The following conclusions can be drawn from these simulations. First, in case of an
An unstable environment, i.e., changing population and abilities, the performance is considerably better when the rate of mobility and entry are higher. This is a consequence of the constant need of selecting the most able entrepreneurs. In a stable environment, the rate of entry may turn out to be inefficiently high after the selection process has done most of its work and the best entrepreneurs run the firms. The extent to which the number of entrepreneurs and economic performance are positively related depends therefore on the stability of the business environment. This is in line with the findings of Mills and Schumann (1985) and Das et al. (1993) who conclude that small firm presence is positively related to fluctuations in demand.

**Table 6.1: Simulation results for five scenarios**

<table>
<thead>
<tr>
<th>Scenario</th>
<th>λ</th>
<th>p</th>
<th>Number of entrepreneurs t=10</th>
<th>t=25</th>
<th>t=100</th>
<th>t=300</th>
<th>Output index t=10</th>
<th>t=25</th>
<th>t=100</th>
<th>t=300</th>
</tr>
</thead>
<tbody>
<tr>
<td>(a)(c)(e)</td>
<td>0.15</td>
<td>0.03</td>
<td>701</td>
<td>810</td>
<td>816</td>
<td>802</td>
<td>117</td>
<td>130</td>
<td>143</td>
<td>144</td>
</tr>
<tr>
<td></td>
<td>0.15</td>
<td>0.01</td>
<td>429</td>
<td>494</td>
<td>567</td>
<td>624</td>
<td>111</td>
<td>119</td>
<td>138</td>
<td>145</td>
</tr>
<tr>
<td></td>
<td>0.05</td>
<td>0.03</td>
<td>702</td>
<td>808</td>
<td>799</td>
<td>778</td>
<td>112</td>
<td>125</td>
<td>143</td>
<td>144</td>
</tr>
<tr>
<td></td>
<td>0.05</td>
<td>0.01</td>
<td>409</td>
<td>479</td>
<td>544</td>
<td>611</td>
<td>109</td>
<td>117</td>
<td>137</td>
<td>145</td>
</tr>
<tr>
<td>(a)(c)(f)</td>
<td>0.15</td>
<td>0.03</td>
<td>676</td>
<td>750</td>
<td>802</td>
<td>773</td>
<td>129</td>
<td>142</td>
<td>144</td>
<td>145</td>
</tr>
<tr>
<td></td>
<td>0.15</td>
<td>0.01</td>
<td>413</td>
<td>534</td>
<td>590</td>
<td>619</td>
<td>119</td>
<td>130</td>
<td>145</td>
<td>146</td>
</tr>
<tr>
<td></td>
<td>0.05</td>
<td>0.03</td>
<td>673</td>
<td>741</td>
<td>785</td>
<td>756</td>
<td>121</td>
<td>137</td>
<td>144</td>
<td>144</td>
</tr>
<tr>
<td></td>
<td>0.05</td>
<td>0.01</td>
<td>397</td>
<td>521</td>
<td>575</td>
<td>607</td>
<td>115</td>
<td>126</td>
<td>145</td>
<td>146</td>
</tr>
<tr>
<td>(b)(c)(e)</td>
<td>0.15</td>
<td>0.03</td>
<td>623</td>
<td>736</td>
<td>816</td>
<td>678</td>
<td>120</td>
<td>126</td>
<td>126</td>
<td>133</td>
</tr>
<tr>
<td></td>
<td>0.15</td>
<td>0.01</td>
<td>387</td>
<td>445</td>
<td>489</td>
<td>388</td>
<td>110</td>
<td>112</td>
<td>111</td>
<td>115</td>
</tr>
<tr>
<td></td>
<td>0.05</td>
<td>0.03</td>
<td>623</td>
<td>775</td>
<td>919</td>
<td>728</td>
<td>116</td>
<td>124</td>
<td>125</td>
<td>131</td>
</tr>
<tr>
<td></td>
<td>0.05</td>
<td>0.01</td>
<td>393</td>
<td>465</td>
<td>556</td>
<td>405</td>
<td>108</td>
<td>111</td>
<td>110</td>
<td>113</td>
</tr>
<tr>
<td>(a)(d)(e)</td>
<td>0.15</td>
<td>0.03</td>
<td>807</td>
<td>846</td>
<td>845</td>
<td>835</td>
<td>120</td>
<td>140</td>
<td>140</td>
<td>140</td>
</tr>
<tr>
<td></td>
<td>0.15</td>
<td>0.01</td>
<td>523</td>
<td>539</td>
<td>586</td>
<td>557</td>
<td>115</td>
<td>129</td>
<td>133</td>
<td>137</td>
</tr>
<tr>
<td></td>
<td>0.05</td>
<td>0.03</td>
<td>785</td>
<td>825</td>
<td>817</td>
<td>820</td>
<td>115</td>
<td>135</td>
<td>137</td>
<td>137</td>
</tr>
<tr>
<td></td>
<td>0.05</td>
<td>0.01</td>
<td>487</td>
<td>536</td>
<td>579</td>
<td>556</td>
<td>112</td>
<td>126</td>
<td>131</td>
<td>135</td>
</tr>
<tr>
<td>(b)(d)(f)</td>
<td>0.15</td>
<td>0.03</td>
<td>686</td>
<td>725</td>
<td>799</td>
<td>760</td>
<td>127</td>
<td>143</td>
<td>138</td>
<td>138</td>
</tr>
<tr>
<td></td>
<td>0.15</td>
<td>0.01</td>
<td>435</td>
<td>464</td>
<td>497</td>
<td>439</td>
<td>117</td>
<td>129</td>
<td>127</td>
<td>127</td>
</tr>
<tr>
<td></td>
<td>0.05</td>
<td>0.03</td>
<td>680</td>
<td>772</td>
<td>912</td>
<td>826</td>
<td>121</td>
<td>139</td>
<td>136</td>
<td>134</td>
</tr>
<tr>
<td></td>
<td>0.05</td>
<td>0.01</td>
<td>433</td>
<td>493</td>
<td>569</td>
<td>483</td>
<td>114</td>
<td>127</td>
<td>125</td>
<td>124</td>
</tr>
</tbody>
</table>

Note: The total number of market participants is 10,000. The output index is 100 in the first year.
Second, while higher mobility always leads to a better economic performance, this need not be the case for a higher rate of entry. Especially in the case when the probability of entry depends upon the entrepreneurial ability the simulation results of the first two scenarios in Table 6.1 show that the entry rate can be too high. An inefficiently large number of small firms enter in the long run in the market setting with \( p = 0.03 \).

Third, a higher number of firms need not be positively correlated with market performance even when the probability of entry is constant. Interestingly enough the number of firms for the low mobility scenarios is higher than that for high mobility scenarios in case the population changes from period to period. The reverse is the case when the population does not alter over time. In case of a changing population the optimal market structure is also constantly changing. An inefficiently large number of small firms survive in that market setting due to more able entrepreneurs not being capable of quick responses to a changing optimal firm size.

The conclusions show that there is not a simple relation between the number of firms and economic performance. However, it is likely that in a period of creative destruction a positive relation between these variables can be found. To test for this we simulate a period of creative destruction by abruptly changing the value of \( \alpha \) from 0.75 to 0.25 in period 301 for two scenarios. The results can be found in Table 6.2.

Table 6.2: Simulation results for a period of creative destruction

<table>
<thead>
<tr>
<th>Scenario</th>
<th>( \lambda )</th>
<th>( p )</th>
<th>Number of entrepreneurs</th>
<th>Output Index</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>t=300</td>
<td>t=305</td>
</tr>
<tr>
<td>(a)(c)(e)</td>
<td>0.15</td>
<td>0.03</td>
<td>802</td>
<td>647</td>
</tr>
<tr>
<td></td>
<td>0.15</td>
<td>0.01</td>
<td>624</td>
<td>407</td>
</tr>
<tr>
<td></td>
<td>0.05</td>
<td>0.03</td>
<td>778</td>
<td>634</td>
</tr>
<tr>
<td></td>
<td>0.05</td>
<td>0.01</td>
<td>611</td>
<td>400</td>
</tr>
<tr>
<td>(b)(c)(f)</td>
<td>0.15</td>
<td>0.03</td>
<td>678</td>
<td>533</td>
</tr>
<tr>
<td></td>
<td>0.15</td>
<td>0.01</td>
<td>361</td>
<td>291</td>
</tr>
<tr>
<td></td>
<td>0.05</td>
<td>0.03</td>
<td>722</td>
<td>573</td>
</tr>
<tr>
<td></td>
<td>0.05</td>
<td>0.01</td>
<td>394</td>
<td>315</td>
</tr>
</tbody>
</table>

Note: The total number of market participants is 10,000. The output index is 100 in the first year.

The simulation results show that a fast recovery in the number of firms after the shakeout due to a sudden change in abilities is beneficial for the recovery process. This process can be seen to have been about twice as fast in the case of \( \lambda = 0.15 \) and \( p = 0.03 \) when compared to that of \( \lambda = 0.05 \) and \( p = 0.01 \).

The simulations results which we have discussed above show that the qualitative assumptions about the 'pool' of market participants may affect the evolution of market
structure and performance considerably. It makes clear that an increase in the number of firms may be due to higher mobility barriers or to a too high level of entry and therefore possibly be negatively correlated with performance. It may of course be the reverse case: an increase in the number of firms may be due to lower mobility barriers or an increase of the rate of entry to more efficient levels. The question which case is relevant in which industry and which period remains an empirical one. We will turn to two empirical examinations in the next sections. The presented model of competitive selection is only tentative. Alternative analyses, also considering the role of technological learning, can be found in Eliasson (1984), Winter (1984) and Dosi et al. (1995).

6.3 Large firm presence in European manufacturing

Considerable data material is available showing that the size class structure of firms is changing. The most impressive and also the most cited is that of the 500 largest American firms, the so-called Fortune 500. Their employment share dropped from 20 per cent in 1970 to somewhat more than 10 per cent now. European data zooming in on the size distribution of firms was not available until recently in a systematic manner. However, Eurostat has begun to publish yearly summaries of the firm-size distribution of (potential) EU-members at the two-digit level for the entire business sector. In this study we will use data from the Third Edition of this summary, entitled Enterprises in Europe. The efforts of Eurostat are currently being supplemented by the European Network of SME Research (ENSR), a cooperation of 16 European institutes. This organisation publishes a yearly report of the structure and the developments of the small business sectors in the countries of the European Union. See EIM (1995).

Table 6.3 shows the development of the share of large business (firms with more than 500 employees) in total employment in some European countries, Canada, Japan and the U.S. in the late 1980s and early 1990s. On average, most countries have experienced a decreasing share of large firms. The table does not show whether these decreases are due to sectoral shifts or whether they are due to downsizing within industries. They are probably due to a combination of downsizing in the manufacturing sector and a decreasing share of this sector in the total economy.

In the next section we investigate the effect of differences in the size class structure of firms on the growth of industrial output. This will be done for a sample of 14 manufacturing industries in 13 European countries for a recent period (1990-1994). The share of large firms is calculated from Eurostat (1994). Not all data of industries and countries in this Eurostat report are used. Some countries are not incorporated because they provide establishment data instead of enterprise data. We also do not take industries into consideration where the total number of employees is below 10,000.
Finally, Eurostat sometimes does not provide employment data due to reasons of confidentiality. Two measures of the share of large firms are calculated. The first is the employment share of enterprises with 100 or more employees, MFP (Medium-sized and large Firm Presence) for short. For this variable there is a total of 144 observations. The second is the employment share of enterprises with 500 or more employees, LFP (Large Firm Presence) for short. For this variable there are 130 observations. The correlation between the two measures MFP and LFP for the 130 observations is 0.929. Growth in total production from 1990 to 1993 and from 1990 to 1994 is measured by the production indices of the industry in 1993 and 1994 with base year 1990. The primary sources for the indices are Eurostat (1996) and OECD (1996).

Table 6.3: The employment share of large firms, LFP, in 1988 and 1991.\textsuperscript{55}

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Belgium</td>
<td>35.7</td>
<td>34.9</td>
</tr>
<tr>
<td>Denmark</td>
<td>22.9</td>
<td>21.0</td>
</tr>
<tr>
<td>Finland</td>
<td>38.6</td>
<td>37.2</td>
</tr>
<tr>
<td>France</td>
<td>35.3</td>
<td>33.7</td>
</tr>
<tr>
<td>West Germany</td>
<td>36.1</td>
<td>37.2</td>
</tr>
<tr>
<td>Portugal</td>
<td>23.5</td>
<td>21.0</td>
</tr>
<tr>
<td>United Kingdom</td>
<td>35.2</td>
<td>33.8</td>
</tr>
<tr>
<td>Canada</td>
<td>38.8</td>
<td>34.6</td>
</tr>
<tr>
<td>Japan</td>
<td>26.7</td>
<td>27.6</td>
</tr>
<tr>
<td>United States</td>
<td>50.2</td>
<td>43.1</td>
</tr>
</tbody>
</table>

Note: begin of period is 1989 for Finland and end of period is 1990 for France and West Germany and 1992 for Japan. Source: OECD (1994).

Data are available for 13 countries (number of industries incorporated): Belgium (11), Denmark (10), Finland (7), France (14), Germany (13), Italy (14), the Netherlands (13), Norway (6), Portugal (12), Spain (14), Sweden (8), Switzerland (8) and the United Kingdom (14). All data refer to the year 1990 except for Italy (1989) and Switzerland (1991). The five countries with total employment in the industries incorporated above one million persons are Germany (7.6 million), United Kingdom (4.9 million), Italy (4.2 million), France (4.0 million) and Spain (2.4 million). Total employment in the 144 industries equals 27.3 million persons. Table 6.4 shows how these are distributed over the 14 two-digit level manufacturing industries. The next three columns of the table

\textsuperscript{55}Zoltan Acs argued during the Workshop 'The Impact of Technological Change on Firm and Industry Performance', 29-30 August 1997, that the figure provided by the OECD (1994) for the U.S. in 1991 is incorrect. The figure for LFP should be much closer to 50% in that year.
show the average medium-sized and large firm presence, MFP, and the average production indices. The right hand column shows the average correlation between MFP and the production indices for 1993 and for 1994. The non-weighted average of these correlations is -0.074. On average medium-sized and large firm presence and growth of production appear to be negatively related, but the differences across industries are large. The correlations range from -0.55 to +0.66. In the rest of this chapter we focus on the average effect of MFP and LFP on growth of production. The large range in correlations indicates that the effect may differ quite strongly from one specific industry to another.

Table 6.4: Some descriptive measures.

<table>
<thead>
<tr>
<th>NACE Description</th>
<th>OBS</th>
<th>EMPL</th>
<th>MFP</th>
<th>P93</th>
<th>P94</th>
<th>COR</th>
</tr>
</thead>
<tbody>
<tr>
<td>21/22 Basic metals</td>
<td>8</td>
<td>907</td>
<td>0.87</td>
<td>91.6</td>
<td>99.4</td>
<td>-0.48</td>
</tr>
<tr>
<td>24 Non-metallic mineral products</td>
<td>10</td>
<td>1,081</td>
<td>0.55</td>
<td>88.0</td>
<td>94.0</td>
<td>-0.18</td>
</tr>
<tr>
<td>25/26 Chemicals</td>
<td>12</td>
<td>2,047</td>
<td>0.82</td>
<td>100.5</td>
<td>107.7</td>
<td>0.41</td>
</tr>
<tr>
<td>31 Metal articles</td>
<td>9</td>
<td>2,972</td>
<td>0.39</td>
<td>89.0</td>
<td>93.8</td>
<td>-0.07</td>
</tr>
<tr>
<td>32 Mechanical engineering</td>
<td>11</td>
<td>3,146</td>
<td>0.58</td>
<td>87.5</td>
<td>93.1</td>
<td>0.02</td>
</tr>
<tr>
<td>34 Electrical engineering</td>
<td>10</td>
<td>2,949</td>
<td>0.74</td>
<td>99.1</td>
<td>107.4</td>
<td>0.09</td>
</tr>
<tr>
<td>35 Motor vehicles</td>
<td>8</td>
<td>1,885</td>
<td>0.89</td>
<td>85.9</td>
<td>96.5</td>
<td>0.10</td>
</tr>
<tr>
<td>37 Instrument engineering</td>
<td>7</td>
<td>492</td>
<td>0.54</td>
<td>92.0</td>
<td>95.7</td>
<td>0.66</td>
</tr>
<tr>
<td>41/42 Food, drink and tobacco</td>
<td>13</td>
<td>3,177</td>
<td>0.58</td>
<td>102.5</td>
<td>104.1</td>
<td>-0.11</td>
</tr>
<tr>
<td>43 Textiles</td>
<td>10</td>
<td>1,410</td>
<td>0.61</td>
<td>87.0</td>
<td>91.5</td>
<td>-0.48</td>
</tr>
<tr>
<td>45 Footwear and clothing</td>
<td>11</td>
<td>1,872</td>
<td>0.38</td>
<td>87.2</td>
<td>87.7</td>
<td>-0.49</td>
</tr>
<tr>
<td>46 Wood and wooden products</td>
<td>13</td>
<td>1,759</td>
<td>0.27</td>
<td>95.8</td>
<td>102.0</td>
<td>-0.09</td>
</tr>
<tr>
<td>47 Paper, publishing and printing</td>
<td>11</td>
<td>2,381</td>
<td>0.57</td>
<td>100.4</td>
<td>105.9</td>
<td>-0.55</td>
</tr>
<tr>
<td>48 Rubber and plastics</td>
<td>11</td>
<td>1,269</td>
<td>0.56</td>
<td>95.7</td>
<td>101.8</td>
<td>0.15</td>
</tr>
<tr>
<td>all</td>
<td>144</td>
<td>27,348</td>
<td>0.59</td>
<td>93.5</td>
<td>99.0</td>
<td></td>
</tr>
</tbody>
</table>

Note: OBS stands for the number of countries for which data are available on MFP and production indices for both 1993 and 1994. EMPL stands for the total employment in the industries for the countries for which data are available (in thousands). MFP stands for average medium and large firm presence (the share of firms with more than 100 employees in total employment). P93 and P94 are the average production indices in 1993 and 1994, respectively. The production index in base year 1990 is equal to 100. COR is the average of the correlations between MFP and P93 and between MFP and P94.

6.4 Empirical Results

To test for the effect of the share of (medium-sized and) large firms on growth of production we use the following equations:
(6.1) \( P_{ij} = a_i + b_j + cMFP_{ij} + e_{1ij} \) and

(6.2) \( P_{ij} = a_i + b_j + cLFP_{ij} + e_{2ij} \)

where \( i \) refers to industry and \( j \) to country. The variable \( P_{ij} \) is the production index of industry \( i \) in country \( j \) in 1993 or 1994. The variables \( a_i \) and \( b_j \) are industry and country dummies, respectively. The variables \( e_{1ij} \) and \( e_{2ij} \) are residuals assumed to be i.i.d. It is necessary to incorporate industry dummies because a certain level of large firm presence considered relatively high in one industry may be considered relatively low in another. Choosing a specific period over which to evaluate economic growth is crucial. If the period is too long then the size class structure of the industry may change considerably during the period of observation. If the period is too short then the effect of the size class structure may be overshadowed by the business cycle influence on industry output. We consider two periods, 1990-93 and 1990-94. In 1993 most European manufacturing industries experienced a period of recession. The average production index in our sample in that year was 93.5. The year 1994 disclosed a strong recovery for most industries and the average production index rose again to almost the same level as in 1990.

In Table 6.5 we present least squares estimation results of equations (6.1) and (6.2) in case only industry dummies are incorporated, i.e. \( b_j = 0 \) for all \( j \). We also present the results weighted by employment. This implies that countries and/or industries with a large number of employees have a stronger impact on the regression results. The first four rows of the table show that the effect of MFP or LFP on growth of production is only significant in case we use weighted least squares. There is little difference between equations (6.1) and (6.2) in the percentage of variance explained. The interpretation of the coefficients in the table is straightforward. For example, the weighted least squares results in the first two rows of Table 6.5 imply that an increase in MFP by 0.1 leads to a decrease in output growth by one per cent for the 1990-93 period and two per cent for the 1990-94 period. That is, industries not only appear to be more affected by the recession in case medium-sized and large firms had a larger employment share, they also tend to recover slower from this recession.

In the last four rows of Table 6.5 we present the results in case Spain and Portugal are omitted from the sample. These two countries are in a stage of economic development different from the other countries: the GDP per capita in the two countries is about two-thirds of that in the other countries and the large firm presence in industries is usually much lower than in the same industries in the other countries. It is not unlikely that many firms in Spanish and Portuguese manufacturing have a sub-optimal scale. Small firm presence may only have a positive effect on output growth in a certain stage of organizational and technological development in which scale economies have become less important. Spain and Portugal which joined the European Union only recently, probably have not reached this stage yet. The estimation results in Table 6.5 seem to confirm this. The effect of MFP
and LFP on the growth of output is stronger when Spain and Portugal are left out. The
effect is now also significant for the unweighted least squares estimation results for
equation (6.1).

Table 6.5: Estimation results with industry dummies (t-values in parentheses).

<table>
<thead>
<tr>
<th>Measure</th>
<th>Year</th>
<th>Countries</th>
<th>Obs.</th>
<th>Unweighted</th>
<th>$R^2$</th>
<th>Weighted</th>
<th>$R^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>MFP</td>
<td>1993</td>
<td>All</td>
<td>144</td>
<td>-6.66</td>
<td>0.35</td>
<td>-12.34</td>
<td>0.30</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(1.2)</td>
<td></td>
<td>(3.5)</td>
<td></td>
</tr>
<tr>
<td>MFP</td>
<td>1994</td>
<td>All</td>
<td>144</td>
<td>-7.81</td>
<td>0.29</td>
<td>-21.23</td>
<td>0.19</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(1.1)</td>
<td></td>
<td>(4.5)</td>
<td></td>
</tr>
<tr>
<td>LFP</td>
<td>1993</td>
<td>All</td>
<td>130</td>
<td>-0.18</td>
<td>0.34</td>
<td>-11.77</td>
<td>0.27</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(0.0)</td>
<td></td>
<td>(3.1)</td>
<td></td>
</tr>
<tr>
<td>LFP</td>
<td>1994</td>
<td>All</td>
<td>130</td>
<td>-1.01</td>
<td>0.27</td>
<td>-19.08</td>
<td>0.16</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(0.1)</td>
<td></td>
<td>(3.8)</td>
<td></td>
</tr>
<tr>
<td>MFP</td>
<td>1993</td>
<td>excl. Spain and Portugal</td>
<td>118</td>
<td>-17.51</td>
<td>0.42</td>
<td>-17.93</td>
<td>0.35</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(2.7)</td>
<td></td>
<td>(4.6)</td>
<td></td>
</tr>
<tr>
<td>MFP</td>
<td>1994</td>
<td>excl. Spain and Portugal</td>
<td>118</td>
<td>-19.83</td>
<td>0.33</td>
<td>-26.01</td>
<td>0.23</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(2.3)</td>
<td></td>
<td>(4.9)</td>
<td></td>
</tr>
<tr>
<td>LFP</td>
<td>1993</td>
<td>excl. Spain and Portugal</td>
<td>104</td>
<td>-7.53</td>
<td>0.40</td>
<td>-16.19</td>
<td>0.30</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(1.2)</td>
<td></td>
<td>(3.8)</td>
<td></td>
</tr>
<tr>
<td>LFP</td>
<td>1994</td>
<td>excl. Spain and Portugal</td>
<td>104</td>
<td>-10.87</td>
<td>0.31</td>
<td>-22.40</td>
<td>0.20</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(1.3)</td>
<td></td>
<td>(3.8)</td>
<td></td>
</tr>
</tbody>
</table>

Note: MFP is medium and large firm presence and LFP is large firm presence. Industry dummies are
incorporated in each regression. In the last four rows Spain and Portugal are excluded.

In Table 6.6 we present least squares estimation results when country dummies are
incorporated. We only present the results for MFP because they do not differ much from
those for LFP. The reason for incorporating country dummies is to correct for country-
specific events in the 1990-93 and 1990-94 periods. One such event was the collapse of
the Finnish-Russian trade relationship leading to a strong recession in the Finnish
economy. The general conclusion remains that industries with a higher medium and large
firm share in 1990 have shown less growth of output in the subsequent years. Results are
again somewhat stronger in case Spain and Portugal are omitted from the sample. Despite
the general conclusion of a positive effect, on average, of small firm presence in European
two-digit manufacturing industries on growth of real output, there are some issues to be
resolved in future research. First, the effect appears to differ across industries (see Table
6.4) and to depend upon the stage of economic development of countries (and, as a result,
industries in these countries). Second, the empirical results presented are based upon data
on a high level of aggregation. Research on a lower aggregation is to be recommended.
Table 6.6: Estimation results with country and industry dummies (t-values in parentheses).

<table>
<thead>
<tr>
<th>Measure</th>
<th>Year</th>
<th>Countries</th>
<th>Obs.</th>
<th>Unweighted</th>
<th>R²</th>
<th>Weighted</th>
<th>R²</th>
</tr>
</thead>
<tbody>
<tr>
<td>MFP</td>
<td>1993</td>
<td>All</td>
<td>144</td>
<td>-5.85</td>
<td>0.43</td>
<td>-14.74</td>
<td>0.35</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(0.7)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MFP</td>
<td>1994</td>
<td>All</td>
<td>144</td>
<td>-3.50</td>
<td>0.42</td>
<td>-18.58</td>
<td>0.30</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(0.3)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MFP</td>
<td>1993</td>
<td>excl. Spain and Portugal</td>
<td>118</td>
<td>-16.10</td>
<td>0.49</td>
<td>-16.88</td>
<td>0.40</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(1.7)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MFP</td>
<td>1994</td>
<td>excl. Spain and Portugal</td>
<td>118</td>
<td>-14.71</td>
<td>0.45</td>
<td>-21.23</td>
<td>0.34</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(1.2)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note: MFP is medium firm presence and LFP is large firm presence. Country and industry dummies are incorporated in each regression. In the last two rows Spain and Portugal are excluded.

6.5 Changes in market structure and productivity in Dutch manufacturing

The share of small firms in most manufacturing industries has increased since the 1970s. This may have been the result of, for example, downsizing of large firms, entry of new firms using advanced technologies and introducing new products and of flexible specialization of small firms. The speed and intensity of these developments have not been equal across industries and across countries as demonstrated in Table 6.3. Prominent examples of downsizing are IBM which has been reported to have its workforce reduced from about 400,000 employees in 1987 to about 200,000 employees in 1995 and General Motors which cut employment from about 800,000 in 1979 to about 450,000 in the early 1990s (The Economist (Dec. 21 1996)).

Smaller firms gaining market share can be positively related to economic performance in two ways. First, a decrease in market concentration may lead to more competition and hence an improved performance. Second, the increase in the market share of small firms may point at a fast and intensive process of introducing new products and technologies. It may be interpreted as a measure of industry flexibility which is likely to be positively related with economic performance. In this section we investigate both the effect of the change in C4 and the effect of the change in the share of small firms in employment on the growth of labour productivity. We use data from the DUMA file compiled by EIM (1996) which has information on 71 third-digit manufacturing industries for the 1978-91 period. The dependent variable is the relative change in real output per employee in the period 1987-91, DPROD. We use three explanatory variables, all measured as changes over the 1978-

---

56 Van Witteloostuijn (1997a) compares the severe cuts by large companies with anorexia. He argues that the short-term maximization of shareholders' value pursued by large companies may have important negative effects on chances of survival in the long term.
The first is the difference between the ratio of the total number of employees in the four largest firms and total employment of all firms with 20 or more employees in 1987 and the same ratio in 1978, DC4. The second is the difference between the ratio of total employment in firms with less than 50 employees divided by total employment in all firms in 1987 and the same ratio in 1978, DSSE. The third is the lagged dependent variable: the relative change in real output per employee in the period 1978-87, DPRODL. We note that the variables DC4 and DSSE are computed using data from two distinct data sets both constructed by the Netherlands Central Bureau of Statistics. Firms with less than 20 employees are absent in the first data set, while they have been included in the second.

The causal relation between the structure of an industry and its average productivity growth may be in either or even both directions. In our analysis we stress the effect of changes in market structure on productivity growth. That is, we have a structuralist view (Shepherd (1990) and Nickell (1996)). Adherents of the “Chicago School” view, introduced at the University of Chicago by George Stigler in the 1950s, stress the reverse causal relation. They argue that the performance of firms in the industry influences its structure. We incorporate the lagged dependent variable in our regression equation to prevent any confusion on the causality issue. Such confusion might arise in case DPRODL is strongly correlated with both DPROD and with either DC4 or DSSE.

The average number of firms with more than 20 employees in the 71 industries in 1991 is 88.7. The number ranges from 9 to 483 and only 19 industries have more than 100 firms with more than 20 employees. The average total number of employees in the 71 industries in firms with 20 or more employees in 1991 is 10,800. Total industry employment in firms with 20 or more employees ranges from 600 to 102,500. No less than 17 industries have an employment in firms with 20 or more employees below 3000. The wide spread in the ‘size’ of industries, measured either in terms of number of firms or employment makes it recommendable also to consider weighted least squares results. We use total employment in firms with 20 or more employees, EMP, and the total number of firms with 20 or more employees, N, as weights. These two weights have a correlation coefficient of only 0.15 and it is therefore possible that the choice of weight affects the results considerably.

Table 6.7 shows the regression results for the set of 71 manufacturing industries. The effect of the change of concentration or the change in small firm presence on labour productivity growth is not significant using ordinary least squares. However, using either employment or the number of firms as weight, the changes in industry structure show a significant effect on productivity growth. An increase of the concentration ratio during the 1978-87 period leads to less growth of productivity over the 1987-91 period while an increase of the share of small business leads to more productivity growth. These results appear to confirm the hypothesis that industries in which small firms have gained larger shares of the market turn out to be the same industries that show strong growth of productivity.
Table 6.7: Least squares estimation results (dependent variable: DPROD)

<table>
<thead>
<tr>
<th>Independents</th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
<th>(6)</th>
</tr>
</thead>
<tbody>
<tr>
<td>const</td>
<td>0.119</td>
<td>0.120</td>
<td>0.113</td>
<td>0.115</td>
<td>0.073</td>
<td>0.090</td>
</tr>
<tr>
<td></td>
<td>(5.1)</td>
<td>(4.9)</td>
<td>(5.8)</td>
<td>(5.1)</td>
<td>(2.7)</td>
<td>(4.1)</td>
</tr>
<tr>
<td>DPRODL</td>
<td>-0.011</td>
<td>0.088</td>
<td>-0.007</td>
<td>0.001</td>
<td>0.277</td>
<td>0.118</td>
</tr>
<tr>
<td></td>
<td>(0.2)</td>
<td>(1.1)</td>
<td>(0.1)</td>
<td>(0.0)</td>
<td>(3.6)</td>
<td>(1.5)</td>
</tr>
<tr>
<td>DC4</td>
<td>-0.103</td>
<td>-0.773</td>
<td>-0.758</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.6)</td>
<td>(5.6)</td>
<td>(3.5)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DSSSE</td>
<td></td>
<td></td>
<td></td>
<td>-0.313</td>
<td>1.926</td>
<td>0.618</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(1.5)</td>
<td>(5.1)</td>
</tr>
</tbody>
</table>

Weight EMP N EMP N
R² weighted 0.006 0.825 0.640 0.032 0.814 0.600

Note: Ordinary least squares results are displayed in columns (1) and (4). Least squares results weighted with employment can be found in columns (2) and (5). Least squares results weighted with number of firms can be found in columns (3) and (6).

The average firm size in manufacturing industries varies from large-scale foundries and manufacturers of chemical products to small-scale firms in the leather or knitting industry. Clearly, an increase in the share of small firms in an industry with large firms may have a different effect than such an increase in an industry already with many small firms. Acs and Audretsch (1987) find that small firms have an innovative advantage in industries which are composed of a relatively high proportion of large firms. In a later study (Acs and Audretsch (1989b)) they find that industries with a considerable small firm presence are less attractive to new entrants. This suggests that more small firms promote performance in industries dominated by some larger counterparts. We investigate this hypothesis by dividing our sample into 35 industries with an average firm size in 1991 larger than 100 employees and 36 industries with an average firm size smaller than 100 employees. The estimation results can be found in Table 6.8.

The estimation results show that there is no strong effect of the change in the concentration ratio or the change in small firm presence on productivity growth in industries with on average small firms. These industries seem not to benefit from an increasing small firms' market share. The effect of an increased importance of small firms on productivity growth is confirmed to be positive for industries with high average firm sizes. However, the effect is significant only in case of weighted least squares. Our small sample sizes prevent a rigorous analysis, but the results suggest some support for the hypothesis that manufacturing industries with on average large firms benefit more from more 'smallness', either through downsizing, new firms or flexible specialization by small firms, than
manufacturing industries with on average smaller firms.

Table 6.8: Estimation results for the two sub-sets (dependent variable: DPROD)

<table>
<thead>
<tr>
<th>Indep.</th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
<th>(6)</th>
<th>(7)</th>
<th>(8)</th>
</tr>
</thead>
<tbody>
<tr>
<td>const</td>
<td>0.147</td>
<td>0.123</td>
<td>0.145</td>
<td>0.120</td>
<td>0.137</td>
<td>0.123</td>
<td>0.105</td>
<td>0.101</td>
</tr>
<tr>
<td></td>
<td>(3.7)</td>
<td>(4.6)</td>
<td>(3.6)</td>
<td>(4.4)</td>
<td>(3.5)</td>
<td>(4.9)</td>
<td>(2.9)</td>
<td>(3.5)</td>
</tr>
<tr>
<td>DPRODL</td>
<td>-0.001</td>
<td>-0.156</td>
<td>0.056</td>
<td>-0.077</td>
<td>0.023</td>
<td>-0.144</td>
<td>0.178</td>
<td>-0.016</td>
</tr>
<tr>
<td></td>
<td>(0.0)</td>
<td>(0.6)</td>
<td>(0.5)</td>
<td>(0.7)</td>
<td>(0.2)</td>
<td>(1.6)</td>
<td>(1.8)</td>
<td>(0.1)</td>
</tr>
<tr>
<td>DC4</td>
<td>-0.307</td>
<td>0.148</td>
<td>-0.709</td>
<td>-0.587</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(1.4)</td>
<td>(0.6)</td>
<td>(3.9)</td>
<td>(1.2)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DSSE</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.198</td>
<td>-0.325</td>
<td>3.006</td>
<td>0.622</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(0.4)</td>
<td>(1.6)</td>
<td>(5.2)</td>
<td>(2.1)</td>
</tr>
</tbody>
</table>

Weight

<table>
<thead>
<tr>
<th>EMP</th>
<th>EMP</th>
</tr>
</thead>
<tbody>
<tr>
<td>EMP</td>
<td>EMP</td>
</tr>
</tbody>
</table>

Av. size

Large  Small  Large  Small  Large  Small  Large  Small

R²_weighted

0.059  0.110  0.847  0.637  0.006  0.168  0.879  0.666

Note: Regressions using data on the 35 industries with an average firm size exceeding 100 in 1991 are indicated by 'Large'. The indication 'Small' is used for the 36 industries with an average firm size in 1991 below 100. Least squares results weighted with employment are displayed in columns (3), (4), (7) and (8).

6.6 Smallness and economic growth

The causes and consequences of the shift in economic activity from large to small firms have been on top of the research agenda since the early 1980s. In the present contribution we supplement the work of the pioneers in this field by investigating whether a higher share of small business at the start of the 1990s has led to higher output growth in the subsequent three or four years in European manufacturing. Additionally, we have studied whether the increased presence of small firms has affected the development of productivity in Dutch manufacturing. It would give some support for the intuition of the 'small business economists' that the changes in industrial structure have had some real effects on economic performance. Our results indeed indicate that an industry with a low large firm presence relative to the same industries in other countries has performed better in terms of growth of output. This suggests that lagging behind in the industrial restructuring process has come at a cost of lower economic growth. Countries which have been most active in improving the business environment for the small business sector in the 1980s may very well have reaped the fruits of this policy. The findings are in line with the endogenous growth model of Schmitz (1989). The results presented in Section 6.5 on labour productivity
in Dutch manufacturing are in line with empirical results for U.K. enterprises as provided by Nickell (1996). Nickell studies the effect of competition on the development of productivity of firms. He finds that an increased number of competitors is associated with higher rates of total factor productivity growth.

"Small business has to save us" is a slogan often heard from European politicians and representatives of social and institutional groups. They fear for a further rise of the already unacceptably high level of unemployment caused by the sheer endless series of efficiency and cost-cutting operations of the public and large business sectors. They hope that employment can be fought by stimulating smallness. There are several reasons which may warrant their hopes and slogans. Firstly, stimulating smallness, whatever it may be, lifts the dependency on possibly sluggish and transient resources like scale, scope and experience, and intensifies the dependency on resources like adjustment and effectiveness. The latter resources are likely to be more robust against uncertainty and change than the former. Secondly, stimulation of smallness means stimulation of labour intensity and hence employment by definition. See Loveman and Sengenberger (1991). Thirdly, stimulation of smallness implies an increase in the variety of the range of products and services offered. This not only paves the way for a competitive selection process, and a process with different innovative approaches (Cohen and Klepper (1992a)) but may also satisfy a fragmented and differentiated demand. Finally, Murphy et al. (1991) argue that stimulating talented people to become entrepreneurs instead of rentseekers will benefit growth. In many Western countries, rent seeking rewards talent more than entrepreneurship does, resulting in stagnation.

In the United States small firms replaced large firms not just in terms of generating almost all of the 18 million new jobs created in the 1980s, but also in terms of much of the innovative activity that has driven the growth of new industries and renewed international competitiveness. Meanwhile, throughout Europe, job layoffs and down-sizing of large firms, often in traditional moderate-technology industries, have been common phenomena. The empirical results in this study suggest that a policy of stimulating small firms, or more generally entrepreneurship, may be an effective way of combatting the current decrease in competitiveness of European industry. The results also show that the effectiveness of such a policy may differ significantly across industries.
CHAPTER 7

SMALL BUSINESS, JOB CREATION AND TESTING FOR CONVERGENCE

The effect of firm size on firm growth is the subject of this study. A heated debate has developed concerning the job creation prowess of small firms. Davis et al. (1996) present some apparently convincing evidence that small firms have far less potential to create new (lasting) jobs than assumed by some economists and many politicians. We discuss some of their arguments, and in particular the problem of 'regression fallacy'. We conclude that a possible 'regression-to-the-mean' bias cannot explain the empirical findings of a positive relation between firm size and growth. In the second part of the study we develop two statistics which can be used to test the hypothesis that firm size affects firm growth even when corrected for a possible 'regression-to-the-mean' bias. The statistics test whether the variance of a distribution decreases significantly over time. They are applied to U.S. firm data used in Evans (1987b).
7.1 Introduction

Birch (1981) was the first who claimed that small firms are major contributors of new jobs. Studies in other countries and using more reliable data sets have supported this initial claim. Some economists, most recently Davis et al. (1996), have criticized the conclusion that small firms outperform large firms in terms of job creation. They argue that studies relating job creation to the initial firm size suffer from some methodological flaws. The most important problem appears to be the so-called 'regression fallacy' or 'regression-to-the-mean' bias (see also Friedman (1992) and Kleijweg and Nieuwenhuijsen (1996, 1997)).

A heated debate has developed concerning the job creation prowess of small businesses. Davis et al. claim to dissect the myth that small businesses would be the fountainhead of job creation. They argue that large firms created more jobs than small business did in the U.S. manufacturing industries over the 1972-1988 period. However, related to their employment share small firms created more jobs than their large counterparts. To be more precise, large firms, defined as firms with at least 500 employees, created only 53 per cent of the new jobs while their employment share is 65 per cent. Besides, large firms destroyed 56 per cent of the jobs, which is higher than their share of 53 per cent in job creation. Davis et al. also present as a chief finding that survival rates for new and existing manufacturing jobs increase sharply with employer size. However, this is not surprising and consistent with the common finding in the literature that survival rates increase with firm size. The relevant question of course is not whether small establishments have higher exit rates but whether the higher growth rates of successful small firms are large enough to compensate for their higher discontinuance rates. From Table III of Dunne et al. (1989) and from Cooper et al. (1989) it is indeed clear that this is the case: when taking the growth rate of failures to be –100%, small firms on average still grow more strongly than large firms do.

Another chief finding would be that smaller manufacturing firms and plants exhibit sharply higher gross rates of job creation but not higher net rates. Davis et al. reach this conclusion using the average plant size measure to correct for the 'regression-to-the-mean' bias. We will show in this chapter that this correction may be inappropriate. Moreover, when

57 These observations are in line with Carlsson (1989a) where a decrease in the Fortune 500 employment share in total manufacturing is observed for the 1975-1985 period. This share was still 78.7 per cent in 1975, while it dropped to 72.5 per cent ten years later.

58 See for instance Freeman et al. (1983), Evans (1987b), Dunne et al. (1989) and Audretsch and Mahmood (1995). It is a common practice to correct for the selection bias which occurs in the growth-size regression as a consequence of small firms failing more frequently than large firms. However, Evans (1987b) finds barely any change in the regression results when using such a correction. Moreover, a consequence of such a correction is that entry and exit of firms are treated asymmetrically. Therefore, one might wonder whether such a selectivity bias correction is appropriate.
using the regression-to-the-mean correction proposed by Davis et al., Baldwin and Picot (1995) still find a higher net job creation rate of firms in the small size classes for Canada. Konings (1995) not only finds that gross rates of job creation decrease with firm size, but also that gross job destruction rates increase with firm size in the U.K. during the 1980s. This of course implies that the net job creation rates in the U.K. have been a lot higher for small firms when compared to their larger counterparts.

In Section 7.2 we pay considerable attention to the 'regression fallacy', a possible source of overestimation of the small firm job creation prowess. In Section 7.3 we discuss some recent developments in the testing of convergence of productivities of countries. This literature has grown considerably in the last three years. The test statistics developed in this literature also appear relevant for studying convergence of firm sizes. In case small firms grow faster than their larger counterparts then this implies convergence. Lichtenberg (1994) shows that this is only a weak form of convergence: the so-called $\beta$-convergence. He suggests to study a stronger form of convergence: the so-called $\alpha$-convergence (see also Friedman (1992)). The test of $\alpha$-convergence is in fact a means to circumvent the 'regression fallacy'.

In his recent article Lichtenberg proposes a test of the convergence hypothesis that the variance of productivity across countries decreases over time. He argues that the ratio of the variance in the first and the variance in the last period of the time series is F-distributed but overlooks the dependency between these two variances. As a consequence, probabilities of committing a type II error of incorrectly rejecting the convergence hypothesis are large. This problem manifests most strongly in short time periods. Lichtenberg, for example, rejects the convergence hypothesis for a data set of 22 OECD-countries over the 1960-85 period. Using two alternative test statistics, we claim that there is strong empirical evidence for convergence in that time period. These test statistics are also applied on the convergence of firm sizes in Section 7.4.

7.2 Regression fallacy

We now turn to the regression fallacy. The interpretation of this fallacy in the (employment) growth-size relationship is that large firms may be large because of a stochastic, transitory increase in employment. Because this increase is temporary, firms that are large in a certain period are (on average) likely to have a decrease in the next period. The reverse holds for small firms. The observation that - measured from the base-year or initial period - small firms show on average higher growth rates than large ones is then the regression...

---

59 The more common finding is that both job creation and job destruction rates are higher in small firms than in large ones. This seems also to have been the case for the Dutch manufacturing sector over the period 1978-1991 (Broersma and Gautier (1997)).
fallacy. This observation in favour of small firms is claimed to be caused by stochastic, temporal shocks in employment and not by small firms outperforming large firms in terms of job creation. Davis et al. make a plea to use the average size as the independent variable in the growth-size relationship to avoid the 'regression-to-the-mean' bias.

In their Section 7 Davis et al. suggest that longitudinal studies of the relationship between growth and (employer) size which are corrected for the regression fallacy should be preferred above those which are not. We disagree with this general suggestion. Davis et al. suggest that firms fluctuate around their own long run size. We doubt whether the concept of a long run size is useful. First, firms are continuously confronted with environmental changes. Such changes conflict with a long run firm size which has to remain constant over time. Second, for young firms - which are generally small because firms enter the market at small sizes - it may be quite impossible to give a proper definition of its long run size. Especially, young and small firms will, if they survive, grow faster than their old and large counterparts.

Table 7.1 provides an illustration of a case for which we believe that a correction for the 'regression-to-the-mean' bias is inadequate. The conclusion from the first part of the table seems to be clear. Firm 1 grows, employment in firm 2 is stable and firm 3 declines. If we regress the (logarithmic) growth for the entire period on initial size then we would find - in accordance with Table 7.1 - a negative regression coefficient. However, if we replace initial size by average size there would be no relation between growth and size at all.

The main difference between the first part of Table 7.1 and the second part of the table (Box 4 in Davis et al.) is that in the first part of the table firms show a persistent growth or decline, while the firms in the second part show temporary changes in the level of employment. We believe that allowing for the possibility for (at least some) firms to grow persistently is the key element for the decision whether or not to correct for the 'regression-

---

60 Also in the theoretical literature on industry dynamics the existence of a unique equilibrium firm size is debated. For instance, Jovanovic (1982) assumed agents to be of intrinsic, unchanging 'types' which are revealed when participating in the industry. In this sense his model is a dynamic extension of the model of Lucas (1978). Firm sizes have an equilibrium value in these models. A recent, alternative, approach to industry dynamics is developed by Ericson and Pakes (1995). They do not assume a passive learning process of an already fixed 'type' but allow firms to explore and exploit speculative ideas. As a result, firms do not have an equilibrium size and even are bound to disappear from the industry in the long run.


63 We set up Table 7.1 in the same fashion as Box 4 in Davis et al. Thus, we intend to show that the illustration of the regression fallacy by Davis et al. may not be generalized.
to-the-mean' bias. With the difference between persistent or temporal growth rates in mind, we observe that in the literature two 'schools' may be identified in the formulation of the growth-size model. Adherents of the first school - who support the assumption that firms fluctuate around their own long run size and that deviations from this size are temporal - are for instance Davis et al. and Leonard. The analysis in Leonard (1986) starts with equation (7.1):

\[ \ln S_{it} = X_i \beta + \epsilon_{it} \]

where \( S_{it} \) is the size of firm \( i \) in period \( t \), \( X_i \) is the vector of firm characteristics given optimal scale, i.e., a vector of time-invariant firm characteristics and \( \epsilon_{it} \) is a random error that may include measurement error and which is i.i.d. \( N(0, \sigma^2) \).

Table 7.1 Illustration of an undesirable effect of the correction for regression fallacy

<table>
<thead>
<tr>
<th>Year 1 Employment</th>
<th>Firm 1</th>
<th>Firm 2</th>
<th>Firm 3</th>
<th>Firm 1</th>
<th>Firm 2</th>
<th>Firm 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>400</td>
<td>500</td>
<td>600</td>
<td></td>
<td>450</td>
<td>550</td>
<td>600</td>
</tr>
<tr>
<td>500</td>
<td>500</td>
<td>500</td>
<td></td>
<td>550</td>
<td>450</td>
<td>600</td>
</tr>
<tr>
<td>600</td>
<td>500</td>
<td>400</td>
<td></td>
<td>450</td>
<td>550</td>
<td>600</td>
</tr>
<tr>
<td>Year 2 Logarithmic Growth</td>
<td>0.223</td>
<td>0</td>
<td>-0.182</td>
<td>0</td>
<td>0.201</td>
<td>-0.201</td>
</tr>
<tr>
<td>Year 3 Logarithmic Growth</td>
<td>0.182</td>
<td>0</td>
<td>-0.223</td>
<td>-0.201</td>
<td>0.201</td>
<td>0</td>
</tr>
</tbody>
</table>

Note: The first three columns contain a case of persistent growth of firms. The second three columns are based upon Box 4 in Davis et al. They contain a case of temporary changes in firm sizes.

The expected growth conditional on initial size (\( \ln S_{it-1} \)) equals \(-\epsilon_{it-1}\). Thus, compared to their expected sizes, large firms are expected to shrink and small firms are expected to grow. A regression of firm growth rates on initial firm sizes will therefore suffer from the 'regression-to-the-mean' bias. Adherents of the first school assume firm size to fluctuate randomly around the expected value of \( X_i \beta \) in equation (7.1). Followers of the second school analyse whether the firm growth rates fluctuate randomly, i.e. whether Gibrat's law is valid. In Evans (1987b) the following regression framework is presented:

---

\(^{64}\)The use of average size instead of initial size also introduces confusion about the causal relationship. The (exogenous) size variable does not precede the period over which the (endogenous) growth variable is measured.

\(^{65}\)Leonard (1986) allows firms to adjust their stock of human capital employed over more than one period of time. However, this extension does not change the basic argument of the 'regression-to-the-mean' bias.
(7.2) \((\ln S_{it} - \ln S_{i,t-d})/d = \ln G(A_{i,t-d}, S_{i,t-d}) + \eta_{it}\)

where \(d\) is the number of years between the beginning and the end of the observation period, \(A\) is firm age and \(\eta_{it}\) is a random error. Vairiyam and Kraybill (1992) and Mata (1994) for example use the same approach.\(^{66}\)

Evans (1987b) reports that firm growth decreases with firm age and firm size.\(^{67}\) Adherents of the first school would say that the negative relation between firm size and growth may be just the result of 'regression-to-the-mean' bias. However, as illustrated in Table 7.1, in our opinion the correction for the 'regression-to-the-mean' bias is only useful if firms show temporary changes in employment, but if firms grow or shrink persistently then the correction for the 'regression-to-the-mean' bias, as suggested by Davis et al., is inappropriate. Moreover, there is ample empirical evidence to doubt that equation (7.1) describes the dynamics of firm size. We provide three illustrations.

First, even if the 'regression-to-the-mean' bias is relevant but nevertheless the regression framework of equation (7.2) has been used, it is possible to derive a theoretical probability limit value for the regression coefficient for the size variable. This value equals \(-\alpha_1/(d \sigma^2 + d \text{var}(X, \beta))\), in a two-period model where a linear form \(\ln G = \alpha_0 + \alpha_1 \ln S_{i,t-d}\) in equation (7.2) is chosen. We find, for example, in Evans (1987b) that the regression coefficient for firms younger than 46 years of age is more negative than this theoretical value. We conclude that the growth-size relation found in Evans (1987b) cannot be generated by a 'regression-to-the-mean' model as in equation (7.1). This also holds for some older studies, like Prais (1976), where even a significantly positive effect of initial size on growth is found up till the 1960s. Such a positive coefficient can in no case be the consequence of equation (7.1).\(^{68}\)

Second, Boeri and Cramer (1992) claim for German establishments in the period 1977-1990 that the growth-size relation is not the result of a 'regression-to-the-mean' bias only. They model growth of incumbents as a first-order Markov process and reject the hypothesis

---

\(^{66}\)A study of the relation between size, age and growth of firms for a developing country (Ivory Coast) using the Evans-framework can be found in Goedhuys and Slevwaegen (1997).

\(^{67}\)It is sometimes claimed that especially new firms have above-average growth rates of employment when corrected for their larger rates of exit (Birley (1986), Klomp and Thurik (1997)).

\(^{68}\)The effect of firm size on growth was found by Prais (1976) to be positive till the late 1960s. Since then, however, most studies have reported a negative effect. It appears therefore that in some periods large firms grow faster while in other periods small ones do. These empirical results are consistent with Gartner and Shane (1995) who report that the number of firms per capita decreases during the 1950s and 1960s, remains stable during the 1970s and starts to increase quite considerably during the 1980s. It is also consistent with Acs et al. (1991) who show that, on average, self-employment rates in the OECD countries decreased in the 1965-1977 period and increased afterwards.
of stationarity of the Markovian transition matrix. Stationarity is needed if the growth-size relation would just be the result of a 'regression-to-the-mean' bias.

Third, in several previous growth-size studies a temporal analysis on firm growth has been carried out.\textsuperscript{69} A positive, but usually small, autocorrelation is generally found in such studies. From equation (7.1) one would expect this autocorrelation to be negative.

The two test statistics which are derived in the next section can be used to check whether a constant variance model like equation (7.1) might adequately summarize the data generating process. They are designed to test for decreasing variance (convergence) of productivities across countries, but are also useful to test for decreasing variance of firm sizes.

7.3 Testing the convergence hypothesis

7.3.1 Convergence of productivity

The Solow growth model and endogenous growth models give opposite predictions about the development over time of productivity differences across countries. A convergence of productivities is considered empirical evidence in favour of the Solow growth model while it is considered difficult to reconcile with endogenous growth theories. Recently, a discussion has emerged about the precise way to test for convergence (Bernard and Durlauf (1996), Den Haan (1995), Evans (1997), Islam (1995) and Oxley and Greasley (1995)). A common test of the convergence hypothesis has been to investigate whether poor countries grow faster than rich countries. Lichtenberg (1994) criticizes this practice of testing convergence. He emphasizes the difference between convergence and mean reversion, which we have discussed in the previous section, and shows that a negative effect of initial productivity on its growth rate does not automatically imply convergence. That is, Lichtenberg claims that empirical studies have \textit{overestimated} the rates of convergence.\textsuperscript{70} He argues that convergence is equivalent to a decrease over time in the variance of productivity across countries. In more technical terms, Lichtenberg, like Friedman (1992), asserts that research should focus on $\sigma$-convergence instead of $\beta$-convergence and he shows that $\beta$-convergence is a necessary but not a sufficient condition for $\sigma$-convergence. Therefore, he suggests to use a ratio of variances statistic to test the convergence hypothesis which he claims to be F-distributed. In this comment we show that this claim is incorrect and we discuss a likelihood ratio test statistic and an adjusted ratio

\textsuperscript{69}See, for instance, Boeri (1989), Wagner (1992) and Klomp and Thurik (1994) for some references.

\textsuperscript{70}Other researchers have recently claimed that the rates of convergence have been \textit{underestimated} (Evans (1997), Islam (1995)).
of variances test statistic which may be used to test the convergence hypothesis. We compare the performance of the test statistics in a simulation experiment. Finally, the test statistics are applied to a data set of 22 OECD countries over the 1950-1994 period and some subperiods. The results indicate that the variance of productivity across these countries has decreased significantly since 1950 and the alternative test statistics both reverse the earlier Lichtenberg finding of non-convergence for the OECD-countries for the 1960-85 period. The newly developed test statistics are applied to firm data in Section 7.4.

7.3.2 Two new test statistics

Lichtenberg proposes a test of the convergence hypothesis that the variance of productivity across countries decreases over time. If \( y_{it} = \ln(Y_{it}) \), where \( Y_{it} \) is the productivity in country \( i \) at time \( t \) and \( \sigma_i^2 = \frac{1}{N} \sum (Y_{it} - \bar{Y}_i)^2 \) is the variance of \( Y_{it} \) across countries, then Lichtenberg claims that \( \sigma_T^2 \) is \( F(N-2, N-2) \)-distributed in case productivities do not converge over time, where \( N \) is the number of countries and \( T \) is the end of the period of investigation. We do not agree with this claim and show that this test procedure is biased towards finding no convergence. Assume that productivities are determined by the following autoregressive process:

\[
(7.3) \quad y_{it} = \rho y_{i,t-1} + \nu_{it}, \quad t = 2, \ldots, T \quad i = 1, \ldots, N,
\]

where the intercept is suppressed. The \( y_{it} \) are supposed to be i.i.d. \( N(\mu, \sigma_i^2) \) and to be independent of the \( \nu_{it} \) which are i.i.d. \( N(0, \sigma_i^2) \). The null hypothesis of no convergence is equivalent to the parameter restriction \( \rho^2 = 1 - \sigma_T^2 / \sigma_i^2 \). Productivities converge over time in case \( \rho^2 < 1 - \sigma_T^2 / \sigma_i^2 \). From equation (7.3) we derive Lichtenberg's equation (4):

\[
(7.4) \quad y_{it} = \pi y_{i,t} + u_i, \quad i = 1, \ldots, N,
\]

where \( \pi = \rho^{T-1} \) and \( u_i = \sum_{t=2}^{T} \rho^{T-t} \nu_{it} \). The case of no convergence is equivalent with \( \pi^2 = 1 - \sigma_T^2 / \sigma_i^2 \). It is clear that \( \bar{o}_T^2 \) and \( \bar{o}_1^2 \) are not independently distributed if \( \pi = 0 \). Therefore, Lichtenberg's claim that \( T_1 = \bar{o}_1^2 / \bar{o}_T^2 \) would be F-distributed is incorrect in the common case of \( \pi > 0 \). The deviation of the test statistic from a F-distribution is stronger, the larger is \( \pi \). The implication of incorrectly using critical values of a F\((N-2, N-2)\)-distribution
is that probabilities of committing a type I error are smaller than the significance level.\footnote{Lichtenberg also incorrectly takes the degrees of freedom of both the nominator and denominator of his ratio of variance statistic to be \(N-2\) instead of \(N-1\) although this will not affect empirical results substantially if \(N\) is larger than about 15.} This is a consequence of the larger variability of \(\hat{\sigma}_1^2 / \hat{\sigma}_T^2\) (when \(\sigma_1^2 = \sigma_T^2\)) in case \(\hat{\sigma}_1^2\) and \(\hat{\sigma}_T^2\) are supposed to be uncorrelated when compared to the case of positively correlated \(\hat{\sigma}_1^2\) and \(\hat{\sigma}_T^2\). The main problem is of course not the low probabilities of committing a type I error but the attendant high probabilities of committing a type II error.

We propose two alternative test statistics of the hypothesis that the variances in the first and in the last period are equal. Asymptotic distributions for both test statistics will be given while small sample performances will be investigated in the next section. The first test statistic, \(T_2\), is derived using the likelihood ratio principle. The second test statistic, \(T_3\), is found by deriving the correct (asymptotic) distribution of Lichtenberg's \(T_1\)-statistic.

We first derive a likelihood ratio test statistic which is a function of \(\hat{\sigma}_1^2\) and \(\hat{\sigma}_T^2\), like Lichtenberg's statistic, and of the covariance of productivities in the first and last period, \(\hat{\sigma}_{1T} = \sum (y_{it} - \hat{y}_t)(y_{iT} - \hat{y}_T) / N\). The productivities in the first and last period have a bivariate normal distribution:

\[
\begin{pmatrix}
    y_{1t} \\
    y_{iT}
\end{pmatrix} \sim N
\begin{pmatrix}
    \mu_1 \\
    \mu_T
\end{pmatrix}
\begin{pmatrix}
    \sigma_1^2 & \sigma_{1T} \\
    \sigma_{1T} & \sigma_T^2
\end{pmatrix}
\]

The null hypothesis of no convergence is equivalent to \(\sigma_1^2 = \sigma_T^2 = \sigma^2\) and the maximum likelihood estimator of \(\sigma^2\) is \(\hat{\sigma}^2 = \frac{1}{2} \hat{\sigma}_1^2 + \frac{1}{2} \hat{\sigma}_T^2\). The value of the log likelihood function in case of the null hypothesis of equal variances, \(L_0\), and in case of no parameter restriction, \(L_A\), are:

\[
L_0 = -\frac{1}{2} N \text{ln}(4\pi^2) + \text{ln}(\hat{\sigma}^4 - \hat{\sigma}_{1T}^2) + \text{ln}(\hat{\sigma}^2(\hat{\sigma}_1^2 + \hat{\sigma}_T^2) - 2\hat{\sigma}_{1T}^2) / (\hat{\sigma}^4 - \hat{\sigma}_{1T}^2)
\]

\[
L_A = -\frac{1}{2} N \text{ln}(4\pi^2) + \text{ln}(\hat{\sigma}_1^2 \hat{\sigma}_T^2 - \hat{\sigma}_{1T}^2) + 2.
\]

If we substitute \(\hat{\sigma}^2 = \frac{1}{2} \hat{\sigma}_1^2 + \frac{1}{2} \hat{\sigma}_T^2\), then two times the difference between \(L_A\) and \(L_0\) equals:

\[
2(L_A - L_0) = N \text{ln}((\hat{\sigma}_1^2 + \hat{\sigma}_T^2)^2 / 4 - \hat{\sigma}_{1T}^2) - \text{ln}(\hat{\sigma}_1^2 \hat{\sigma}_T^2 - \hat{\sigma}_{1T}^2)) = N \text{ln}(1 + \frac{1}{4} \frac{(\hat{\sigma}_1^2 - \hat{\sigma}_T^2)^2}{\hat{\sigma}_1^2 \hat{\sigma}_T^2 - \hat{\sigma}_{1T}^2}).
\]
This statistic has a limiting $\chi^2(1)$-distribution using the standard asymptotic property of the likelihood ratio test. From Morrison (1978, p. 250) we find that the $\chi^2$-approximation is improved if we replace $N$ in equation (7.8) by $N-2.5$. Hence, we define our test statistic as:

(7.9) \[ T_2 = (N-2.5) \ln(1 + \frac{1}{4} \frac{(\hat{\sigma}_1^2 - \hat{\sigma}_T^2)^2}{\hat{\sigma}_1^2 \hat{\sigma}_T^2 - \hat{\sigma}_{1u}^2}). \]

The second way of testing the equality of variances $\sigma_1^2$ and $\sigma_T^2$ is as follows. From equation (7.4) it can be derived that

(7.10) \[ \hat{\sigma}_T^2 = \pi^2 \hat{\sigma}_1^2 + 2m \hat{\sigma}_{1u} + \hat{\sigma}_u^2. \]

The elements of the covariance matrix of $y_{1i}$ and $u_i$ have asymptotically a trivariate normal distribution (Wesselman (1987, p.20)):

(7.11) \[ \sqrt{N} \begin{pmatrix} \hat{\sigma}_1^2 - \sigma_1^2 \\ \hat{\sigma}_{1u} \\ \hat{\sigma}_u^2 - \sigma_u^2 \end{pmatrix} \sim N \left( 0, \begin{pmatrix} 2\sigma_1^4 & 0 & 0 \\ 0 & 2\sigma_u^4 & 0 \\ 0 & 0 & 2\sigma_u^4 \end{pmatrix} \right). \]

An approximate distribution of the ratio of variances, $T_f$, under the hypothesis $\pi^2 = 1 - \sigma_u^2 / \sigma_1^2$ can now be derived using the "delta method" (Wesselman (1987, p. 22)):

(7.12) \[ \sqrt{N} \frac{\hat{\sigma}_1^2}{\hat{\sigma}_T^2} - 1 \sim N \left( 0, \begin{pmatrix} \frac{\partial (\hat{\sigma}_1^2 / \hat{\sigma}_T^2) / \partial \hat{\sigma}_1^2}{\partial (\hat{\sigma}_1^2 / \hat{\sigma}_T^2) / \partial \hat{\sigma}_T^2} \\ 0 & \frac{\partial (\hat{\sigma}_1^2 / \hat{\sigma}_T^2) / \partial \hat{\sigma}_{1u}}{\partial (\hat{\sigma}_1^2 / \hat{\sigma}_T^2) / \partial \hat{\sigma}_u^2} \\ \frac{\partial (\hat{\sigma}_1^2 / \hat{\sigma}_T^2) / \partial \hat{\sigma}_u^2}{\partial (\hat{\sigma}_1^2 / \hat{\sigma}_T^2) / \partial \hat{\sigma}_u^2} \end{pmatrix} \right) = N(0, 4 - 4\pi^2). \]

By substituting the least squares estimate $\hat{\pi}$ for the parameter $\pi$ we derive an adjusted ratio of variances test statistic, $T_3$, which has asymptotically a standard normal

\[ \text{Note that } \hat{\sigma}_{1T} \text{ can simply be computed as } \hat{\sigma}_{1T} = \hat{\pi} \hat{\sigma}_1^2, \text{ where } \hat{\pi} \text{ is the least squares estimate of } \pi \text{ in equation (7.4).} \]
distribution.$^{73}$

\[
T_3 = \frac{\sqrt{N}(\bar{\sigma}^2 / \sigma^2_1 - 1)}{2\sqrt{1 - \bar{\eta}^2}}
\]

7.3.3 Simulation experiment

In order to compare the small sample performance of the three test statistics $T_1$, $T_2$, and $T_3$, we use a Monte Carlo simulation experiment with 20,000 replications for in total 90 elements of the parameter space $(N, T, \sigma^2, \rho)$. In each of the experiments we fix $\mu$ at zero and $\sigma_1$ at unity. We do not incorporate an intercept in equation (7.3) although we allow for it in the least squares estimation. The test statistics are compared on basis of the nine $(N, T)$-combinations from the set (\{10,25,100\},\{10,25,40\}). Empirical analyses on convergences of productivities are usually based upon a low number of countrie or regions and a low number of (yearly) time observations. Asymptotic distributions of the test statistics may not be very useful in these settings.

Table 7.2 shows the results of the simulation experiment for the test statistic $T_1$, while Table 7.3 and 7.4 shows the corresponding results for the likelihood ratio test statistic $T_2$ and the adjusted ratio of variances test statistic $T_3$, respectively. For each element of the parameter space we compute the percentage of the replications which gave a test statistic larger than the value which corresponds to the theoretical 5%-significance level. In case $\rho^2 = 1 - \sigma^2_v / \sigma^2_1$ ($H_0$) this rejection frequency corresponds to the significance level while in case $\rho^2 < 1 - \sigma^2_v / \sigma^2_1$ it corresponds to the power of the test. We concentrate on values of $\rho$ close to one because these are most common in empirical work concerning convergence.

Lichtenberg’s $T_1$-statistic has a simulated significance level close to the theoretical significance level only in experiments with $T=40$ and $\rho = 0.98$. This was to be expected because $T_1$ will be approximately F-distributed only in cases in which $\pi = \rho^{T-1}$ is close to zero. The likelihood ratio test statistic has simulated significance levels quite close to the theoretical significance level for all experiments. The simulated significance levels of the adjusted ratio of variances test statistic deviate somewhat stronger from the theoretical level. They are below the theoretical level in case $\pi = \rho^{T-1}$ is close to one and they are above the theoretical level in case $\pi = \rho^{T-1}$ is close to zero. The speed of convergence of $T_3$ to a standard normal distribution appears to be quite low.

\[73\] If $|\bar{\eta}| > 1$ then $T_3$ cannot be determined. When this case occurred in the simulation experiment we decided for not accepting convergence.
Table 7.2: Performance of ratio of variances test statistic $T_1$

<table>
<thead>
<tr>
<th>Hyp</th>
<th>$\sigma^2_v$</th>
<th>$\rho$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$H_0$</td>
<td>0.0100</td>
<td>0.995</td>
</tr>
<tr>
<td></td>
<td>0.99</td>
<td>0.99</td>
</tr>
<tr>
<td></td>
<td>0.985</td>
<td>0.98</td>
</tr>
<tr>
<td></td>
<td>0.98</td>
<td>0.98</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>N,T</th>
<th>10,10</th>
<th>10,25</th>
<th>10,40</th>
<th>25,10</th>
<th>25,25</th>
<th>25,40</th>
<th>100,10</th>
<th>100,25</th>
<th>100,40</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.0</td>
<td>0.0</td>
<td>0.1</td>
<td>0.3</td>
<td>0.0</td>
<td>0.1</td>
<td>0.3</td>
<td>0.0</td>
<td>0.0</td>
<td>0.2</td>
</tr>
<tr>
<td>0.0199</td>
<td>0.99</td>
<td>0.1</td>
<td>0.5</td>
<td>1.3</td>
<td>0.0</td>
<td>0.5</td>
<td>1.1</td>
<td>0.0</td>
<td>0.2</td>
</tr>
<tr>
<td>0.985</td>
<td>0.98</td>
<td>0.2</td>
<td>3.9</td>
<td>10.0</td>
<td>0.1</td>
<td>11.0</td>
<td>26.1</td>
<td>1.6</td>
<td>56.0</td>
</tr>
<tr>
<td>0.0298</td>
<td>0.985</td>
<td>0.1</td>
<td>1.0</td>
<td>2.1</td>
<td>0.1</td>
<td>1.0</td>
<td>2.2</td>
<td>0.0</td>
<td>0.7</td>
</tr>
<tr>
<td>0.98</td>
<td>0.3</td>
<td>2.4</td>
<td>5.0</td>
<td>0.2</td>
<td>4.1</td>
<td>8.5</td>
<td>0.4</td>
<td>10.9</td>
<td></td>
</tr>
<tr>
<td>0.0396</td>
<td>0.98</td>
<td>0.3</td>
<td>1.5</td>
<td>2.6</td>
<td>0.1</td>
<td>1.6</td>
<td>2.9</td>
<td>0.1</td>
<td>1.3</td>
</tr>
</tbody>
</table>

Note: The table shows the percentage of 20,000 replications in which the statistic exceeds the $F_{0.95}(N-2,N-2)$ level.

Table 7.3: Performance of likelihood ratio test statistic $T_2$

<table>
<thead>
<tr>
<th>Hyp</th>
<th>$\sigma^2_v$</th>
<th>$\rho$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$H_0$</td>
<td>0.0100</td>
<td>0.995</td>
</tr>
<tr>
<td></td>
<td>0.99</td>
<td>0.99</td>
</tr>
<tr>
<td></td>
<td>0.985</td>
<td>0.98</td>
</tr>
<tr>
<td></td>
<td>0.98</td>
<td>0.98</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>N,T</th>
<th>10,10</th>
<th>10,25</th>
<th>10,40</th>
<th>25,10</th>
<th>25,25</th>
<th>25,40</th>
<th>100,10</th>
<th>100,25</th>
<th>100,40</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.8</td>
<td>4.7</td>
<td>4.8</td>
<td>5.0</td>
<td>5.2</td>
<td>5.3</td>
<td>4.8</td>
<td>4.8</td>
<td>5.0</td>
<td></td>
</tr>
<tr>
<td>0.0199</td>
<td>5.1</td>
<td>4.9</td>
<td>5.0</td>
<td>4.9</td>
<td>5.2</td>
<td>5.2</td>
<td>5.1</td>
<td>5.0</td>
<td></td>
</tr>
<tr>
<td>0.985</td>
<td>5.7</td>
<td>6.9</td>
<td>7.2</td>
<td>7.5</td>
<td>10.8</td>
<td>12.2</td>
<td>16.3</td>
<td>29.8</td>
<td></td>
</tr>
<tr>
<td>0.98</td>
<td>8.0</td>
<td>11.2</td>
<td>13.1</td>
<td>14.7</td>
<td>25.5</td>
<td>30.7</td>
<td>47.7</td>
<td>77.2</td>
<td></td>
</tr>
<tr>
<td>0.0298</td>
<td>5.0</td>
<td>4.9</td>
<td>4.9</td>
<td>5.0</td>
<td>5.0</td>
<td>4.8</td>
<td>5.1</td>
<td>4.6</td>
<td></td>
</tr>
<tr>
<td>0.98</td>
<td>5.3</td>
<td>6.1</td>
<td>6.4</td>
<td>6.6</td>
<td>8.5</td>
<td>9.2</td>
<td>12.0</td>
<td>19.6</td>
<td></td>
</tr>
<tr>
<td>0.0396</td>
<td>5.3</td>
<td>4.9</td>
<td>4.7</td>
<td>4.9</td>
<td>4.9</td>
<td>4.8</td>
<td>4.9</td>
<td>5.0</td>
<td></td>
</tr>
</tbody>
</table>

Note: The table shows the percentage of 20,000 replications in which the statistic exceeds the $\chi^2_{0.95}(1)$ level.
Table 7.4: Performance of adjusted ratio of variances test statistic $T_3$

<table>
<thead>
<tr>
<th>Hyp</th>
<th>$\sigma^2_\nu$</th>
<th>$\rho$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.0100</td>
<td>0.995</td>
</tr>
<tr>
<td>$H_0$</td>
<td>0.99</td>
<td>0.9</td>
</tr>
<tr>
<td></td>
<td>2.3</td>
<td>14.6</td>
</tr>
<tr>
<td></td>
<td>0.985</td>
<td>5.1</td>
</tr>
<tr>
<td></td>
<td>0.98</td>
<td>10.5</td>
</tr>
<tr>
<td></td>
<td>0.0199</td>
<td>0.99</td>
</tr>
<tr>
<td>$H_0$</td>
<td>3.1</td>
<td>8.0</td>
</tr>
<tr>
<td></td>
<td>0.985</td>
<td>5.1</td>
</tr>
<tr>
<td></td>
<td>0.98</td>
<td>9.1</td>
</tr>
<tr>
<td></td>
<td>0.0298</td>
<td>0.985</td>
</tr>
<tr>
<td></td>
<td>5.0</td>
<td>10.3</td>
</tr>
<tr>
<td></td>
<td>0.98</td>
<td>7.8</td>
</tr>
<tr>
<td></td>
<td>0.0396</td>
<td>0.98</td>
</tr>
<tr>
<td>$H_0$</td>
<td>6.8</td>
<td>11.9</td>
</tr>
<tr>
<td></td>
<td>0.4</td>
<td>4.7</td>
</tr>
<tr>
<td></td>
<td>25,10</td>
<td>25,25</td>
</tr>
<tr>
<td></td>
<td>25,40</td>
<td>100,10</td>
</tr>
<tr>
<td></td>
<td>100,25</td>
<td>100,40</td>
</tr>
</tbody>
</table>

Note: The table shows the percentage of 20,000 replications in which the statistic exceeds the $N_{05}(0.1)$-level.

If we consider the power of the tests, then it can simply be seen that both the likelihood ratio test statistic and the adjusted ratio of variances test statistic outperform Lichtenberg's test statistic $T_1$ for all experiments. Whether to prefer test statistic $T_2$ above test statistic $T_3$ depends upon the value of $\pi = \rho^{T^{-1}}$ and the sample size $N$. A clear advantage of the likelihood ratio test statistic is that it appears to have a distribution close to the asymptotic $\chi^2(1)$-distribution already for small values of $N$. However, in those cases in which the adjusted ratio of variances test statistic has a simulated significance level close to the theoretical level it often has a much higher power than the likelihood ratio test statistic.

7.3.4 Empirical results for 22 OECD-countries

We computed the statistics for a data set of GDP per capita for 22 OECD countries for the 1950-94 period (source: Tables D-1a and D-1b of Maddison (1995)). We did the same for the period 1960-85 which was used by Lichtenberg. The results can be found in Table 7.5. For the period 1950-94 all three statistics indicate that the variance of productivities has decreased indeed. In 1950 it was 0.3160 while it was only 0.1215 in 1994. For the period 1960-85 the $T_1$-statistic suggests that there has not been convergence of GDP per capita while the other two test statistics report convergence. It is clear from Table 7.2 that the use of the $T_1$-statistic for short time periods has a large probability of committing a type II error.
We conclude that there has also been convergence for the period 1960 through 1985, in which the variance decreased from 0.2320 to 0.1424.

Table 7.5: Empirical results for the 22 OECD-countries

<table>
<thead>
<tr>
<th>Period</th>
<th>$T_1$</th>
<th>$T_2$</th>
<th>$T_3$</th>
<th>$\hat{\rho}$</th>
<th>$\hat{\sigma}_\rho^2$</th>
<th>$\hat{\nu}$</th>
<th>$\hat{\sigma}_\nu^2$</th>
<th>$\hat{\sigma}_1^2$</th>
<th>$\hat{\sigma}_T^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1950-94</td>
<td>2.60`</td>
<td>8.04`</td>
<td>4.19`</td>
<td>0.9822</td>
<td>0.0016</td>
<td>0.446</td>
<td>0.096</td>
<td>0.00076</td>
<td>0.3160</td>
</tr>
<tr>
<td>1960-85</td>
<td>1.63</td>
<td>4.79`</td>
<td>2.05`</td>
<td>0.9794</td>
<td>0.0026</td>
<td>0.693</td>
<td>0.082</td>
<td>0.00076</td>
<td>0.2320</td>
</tr>
<tr>
<td>1950-61</td>
<td>1.39</td>
<td>5.03`</td>
<td>1.55`</td>
<td>0.9833</td>
<td>0.0039</td>
<td>0.807</td>
<td>0.058</td>
<td>0.00096</td>
<td>0.3160</td>
</tr>
<tr>
<td>1961-72</td>
<td>1.53</td>
<td>6.15`</td>
<td>1.89`</td>
<td>0.9812</td>
<td>0.0040</td>
<td>0.758</td>
<td>0.063</td>
<td>0.00076</td>
<td>0.2272</td>
</tr>
<tr>
<td>1972-83</td>
<td>1.09</td>
<td>0.94`</td>
<td>0.62`</td>
<td>0.9907</td>
<td>0.0045</td>
<td>0.941</td>
<td>0.041</td>
<td>0.00067</td>
<td>0.1488</td>
</tr>
<tr>
<td>1983-94</td>
<td>1.13</td>
<td>1.51`</td>
<td>0.76`</td>
<td>0.9891</td>
<td>0.0040</td>
<td>0.922</td>
<td>0.044</td>
<td>0.00050</td>
<td>0.1367</td>
</tr>
</tbody>
</table>

Note: *The estimate is significant at the 5%-significance level. The critical values corresponding to this level of significance are 2.12 for the $T_1$-statistic, 3.84 for the $T_2$-statistic and 1.645 for the $T_3$-statistic.*

We also examined some shorter time periods, viz. the time period 1950-94 divided into four subperiods of 12 years each. For such short time periods the $T_1$-statistic is likely to be insignificant and the results in Table 7.5 confirm this. For example, while the likelihood ratio test statistic and the adjusted ratio of variances test statistic indicate that there has been convergence in the 1961-72 period, the $T_1$-statistic never indicates convergence for the shorter time periods. Overall, the pattern seems to correspond well with that the formulation of equation (7.3). From this equation we find that the variance $\sigma_t^2$ of $y_t$ is determined as follows:

\[(7.14) \ \sigma_t^2 = \rho^2 \sigma_{T-1}^2 + \sigma_\nu^2, \quad t = 2, ..., T\]

In case $\rho^2 < 1 - \sigma_\nu^2 / \sigma_1^2$ the variance decreases over time but the decrease becomes less severe over time and the variance converges to $\sigma_\nu^2 / (1 - \rho^2)$. From Table 7.5 one can compute that the estimates of the value of this limiting value lie between 0.019 (period 1960-85) and 0.036 (period 1972-83). A similar pattern is found by Den Haan (1995) for 49 states from 1940 till 1990. The dispersion of per capita income in his sample has become much smaller over this time period but seems to have settled down in the 1970s and 1980s.

7.3.5 Conclusion

The question whether productivities converge across countries is considered important evidence in comparing the Solow growth model and the endogenous growth models.
Lichtenberg (1994) claims that the ratio of the variance in the first period and the variance in the last period can be used as a test statistic of the convergence hypothesis. He argues that this statistic is $F(N-2,N-2)$-distributed, which we show to be an incorrect claim as the variances are not independently distributed. Using a simulation experiment we show that the test procedure proposed by Lichtenberg leads to a low probability of accepting the hypothesis of convergence regardless whether it is true or not. As a consequence, one is biased towards finding empirical support for the endogenous growth models, especially in short time periods. We discuss two alternative test statistics which outperform Lichtenberg's statistic for 'short' time periods.

We apply the three test statistics to a data set of 22 OECD-countries for the time periods 1950-94 and 1960-85. For the period 1950-94 all statistics indicate convergence. For the period 1960-85, as used by Lichtenberg, only his test procedure indicates no convergence. We claim that this is due to the low power of his test statistic. Overall, the results indicate that the degree of convergence across OECD-countries has diminished since 1950. Bernard and Jones (1996a, 1996b, 1996c) investigate which sectors drive the aggregate convergence in OECD-countries. They show that manufacturing shows little evidence of productivity convergence while especially the service sector contributes to the aggregate productivity convergence. These findings warrant a shift in interest from the productivity of the entire economy to that of individual industries.

7.4 Convergence of firm sizes

The number of objects in growth-size relationships of firms is usually very large while the number of time periods is quite limited when compared to country data. However, tests on convergence on an international level are usually also applicable on an industry or sector level. In this section we examine the $c$-convergence of firm sizes which were used by David Evans in his well-known study (Evans (1987b)). He examines the growth of employment over the 1976-82 period of 17,400 U.S. manufacturing firms. These firms are divided into four age-groups: 0-6, 7-20, 21-45 and 46+ years of age. Evans finds firm size to have a significantly negative effect on growth (at the sample mean) for each of these age groups. This $\beta$-convergence does not imply that the variances of firm sizes in different age groups decline. In case of the results being subject to 'regression-to-the-mean', variances are expected not to change over time. We pursue this question by using summary statistics as provided by Evans. Table 7.6 shows summary statistics for the four age groups.

The summary statistics show that the standard deviation of firm sizes has decreased for firms in the youngest age-groups, while it has increased for the group of firms which were 46 years of age or older in 1976. This implies that $c$-convergence of firm sizes is rejected for the oldest firms in the sample. The test statistic $T_2$ equals 16.1, 1327.1 and 27.6 for the age-groups 0-6, 7-20 and 21-45, respectively. The test statistic $T_3$ equals 3.92,
32.54 and 4.88 for these three age-groups. Both indicate that there was indeed \( \alpha \)-convergence of firm sizes for firms younger than 46 years. This is in agreement with the regression results of Evans (1987b) being irreconcilable with the 'regression-to-the-mean' model (7.1), as reported in Section 7.2.

| Table 7.6: Summary statistics for four age-groups in Evans (1987b) |
|-----------------------|--------|--------|--------|--------|
|                      | 0-6    | 7-20   | 21-45  | 46+    |
| \( \bar{\alpha} \)   | 1.461  | 1.581  | 1.622  | 1.705  |
| \( \bar{\alpha}_T \)  | 1.409  | 1.250  | 1.570  | 1.748  |
| \( \bar{\alpha}_{TT} \) | 1.6555 | 1.7317 | 2.2662 | 2.5756 |
| \( \bar{n} \)        | 0.7756 | 0.6928 | 0.8614 | 0.8860 |
| \( N \)              | 4343   | 6124   | 5413   | 1520   |

Note: The figures for \( \bar{\alpha} \), \( \bar{\alpha}_T \), and \( \bar{n} \) are computed using data provided in Section IV in Evans (1987b).

The very question of \( \alpha \)-convergence of firm sizes is important in the discussion of the role of small firms in the job generation process. However, the average job creation prowess of small and large firms in a specific industry, sector or country may as such not be directly relevant for policy makers. The comparison of the job generation process across industries and countries is probably much more relevant. Potential policy improvements may be derived from foreign experiences in case (new or) small firms generate many new jobs in a specific industry in other countries while they do not in the home country. The OECD Industry Committee's Working Party on SMEs has analysed the effectiveness of several SME policies which have been implemented recently (OECD (1995b)). The Working Party identified four broad areas, i.e. financing, business environment, management capability and access to markets. The contribution of small firms to the economy can sometimes be facilitated by government action in these areas.
CHAPTER 8

THE LIFE CYCLE OF A COMPETITIVE INDUSTRY

This study presents a model of industry evolution focusing on firm dynamics over a long time period. We apply this model to the U.S. tire industry over the period 1913-1973. The model explains the path of output, price level and number of producers using a recursive system of equations. It presents an explanation of the nonmonotonicity in firm numbers found in many young industries. Cost of production and the numbers and relative sizes of producers determine the price level. In turn, the cost of production is affected by input prices and the extent of learning-by-doing. The price level and exogenous shifts in the demand function determine the amount of sales. Finally, price-cost margins and growth rates of the amount of sales determine the net entry rate of producers. We show that the rapid shakeout of the number of firms in the U.S. tire industry has not been the result of any dominant innovation. Instead, it is claimed to have been caused by a continuous learning-by-doing process combined with a strong and persistent price competition in the tire industry followed by a strong decline in demand for tires.
8.1. Introduction

A recent literature has emerged focusing on industry evolution, or the dynamic patterns that industries and firms follow as they systematically evolve over time. This literature is important because of the insights provided about how industries change, why they change, and the consequences of industrial change. The research on the evolution of industry received three important contributions in the year 1982. Boyan Jovanovic published the first formal model of industry evolution, Richard Nelson and Sidney Winter presented their influential book on the causes and effects of this phenomenon, and Michael Gort and Steven Klepper published their careful analysis on the stages of the product life cycle. Our knowledge about industry dynamics grew considerably in the last decade, see e.g. Carlsson (1989b), Chandler (1990), Dosi (1988), Geroski and Masson (1987) and Mueller (1990). Important theoretical progress was made by Ericson and Pakes (1995) and Klepper (1996). However, we still lack a model of industry evolution which can adequately explain a particular data set of a specific industry over a long time period. Jovanovic and MacDonald (1994a) were the first to attempt such a model explaining patterns of number of firms, output and prices over time.

The Jovanovic-MacDonald model has one important limitation. It assumes that one innovation followed by one refinement is responsible for the nonmonotonicity in firm numbers. This implies that the model is applicable only for those industries having experienced only two technologies in their entire history, i.e. an initial low-tech and a subsequent high-tech phase. Most industries, and also the automobile tire industry, have an ongoing history of significant improvements in product quality and productivity. It seldom occurs that one innovation is dominant. This is certainly true for the tire industry. In his study on the U.S. tire industry, Nelson (1987) observed that "... there was no single technical breakthrough that unlocked the industry's potential for high-speed production; nor was any individual or firm of overriding importance. The advent of mass production was a cumulative process resulting from a vast number of successive small changes" (p. 331-2, italics added). Reynolds (1938) described a similar process: "The great improvement in tire quality during the past 30 years is undoubtedly due to constant repetition of [the] cycle of invention and imitation" (p. 463).

In the present study, we model this cumulative process based upon learning-by-doing. The analysis by Jovanovic and MacDonald of the automobile tire industry has three further shortcomings. First, they use a demand function in which the increase in the quality of

---

74Case studies on the evolution of specific industries are quite helpful in determining which elements are crucial in modelling industry evolution. Two examples of recent case studies of U.S. consumer goods industries are Swaminathan (1995) on the wine industry and Elzinga and Mills (1996) on the disposable diaper industry.
automobile tires over the period 1910-30 is left out. They do also not take into account the obvious relation between the number of motor vehicles (newly produced or already registered) and the demand for automobile tires. As a consequence, their estimate of the price elasticity of demand for automobile tires is biased. Second, they do not incorporate the natural rubber prices in their model. The price of natural rubber has declined strongly over the 1910-32 period. This partly explains the decline in the price of tires in the same period (Orton (1927), Reynolds (1938) and French (1991)). Third, they assume the number of firms to be independent of the changes in demand for automobile tires. However, many new firms entered the tire industry during a period in which the growth of demand for tires was extremely rapid. The subsequent expansion of (replacement) demand during the 1920s was less than expected by these new entrants and helped to undermine several frail firms (French, ch. 5). The impact of the Great Depression hastened the shakeout in the tire industry (French, ch. 6). Output of tires and number of firms in 1932 both were only half of those in 1928. By neglecting demand as a determinant of net entry, Jovanovic and MacDonald also overlook the stylized fact of industries with higher growth rates on average also to have higher net entry rates (see Section 4.2).

In this chapter we develop a model of the demand for automobile tires, the price of tires and the net entry rate of firms producing tires. The predicted values of these variables are more in line with the data of the U.S. tire industry over the period 1913-1973 when compared to the predicted values found by Jovanovic and MacDonald. In our model the shakeout in the number of producers is a consequence of a continuous process of decreases in the profit margin per tire. Small producers can only survive in case this margin exceeds a certain critical value. We claim that the strong and persistent price competition in the U.S. tire industry in the 1920s followed by a strong decline in demand for tires during the Great Depression made the shakeout of the number of producers so rapid (Reynolds (1938)).

The remainder of this chapter is organized as follows. In the next section we describe our model of industry evolution. This model consists of three equations. The first equation relates the demand for automobile tires to the number of motor vehicles, the price index of tires and a quality index. The second equation describes the decomposition of the price index of tires into a competition effect and a marginal cost effect. The net entry rate is related to the one-period lagged profit margin and growth of demand for automobile tires. Because of this lag the model is a system of recursive equations. From the second and third equation one may also derive the equilibrium number of firms in the market. In Section 8.3 we present the empirical results for the U.S. automobile tire industry over the 1913-1973 period. The estimation results are used to simulate the industry evolution in Section 8.4. Section 8.5 is used for a conclusion.
8.2 The evolution of prices, output and number of firms

In this section we will discuss the model which explains the time paths of automobile tire prices, total output of tires and number of firms producing tires over a period of several decades. The model consists of three equations and contains the following variables:

\[ Q_t = \text{Output of millions of tires in period } t \]
\[ P_t = \text{Real price index of tires (1967=1) in period } t \]
\[ M_t = \text{Output of millions of motor vehicles in period } t \]
\[ S_t = \text{Number of motor vehicles registered for one year or longer in millions in period } t \]
\[ R_t = \text{Real price index of natural rubber (1967=1) in period } t \]
\[ N_t = \text{Number of firms producing tires in period } t \]

The first equation relates the output of automobile tires, \( Q_t \), to the output of motor vehicles, \( M_t \), the number of motor vehicles registered for one year or longer, \( S_t \), and the price index of tires, \( P_t \). The demand for automobile tires can be decomposed into demand for original equipment and renewal purchases. The demand for original equipment depends upon the output of motor vehicles. It is not likely that this demand is influenced by the price of tires because this price is low compared to that of the rest of the motor vehicle. The replacement demand depends upon the number of motor vehicles which have already been on the road for one year or more. Because the replacement of tires can often be delayed, this demand is assumed to be negatively affected by the price of tires. The higher the price of tires, the longer drivers will wait to replace them. A simple estimate of the average number of tires drivers replace per year is \( (Q_t - 5M_t)/S_t \). That is, we subtract the demand for original equipment from the total number of tires produced and then divide it by the number of registered motor vehicles. We assume here that every new motor vehicle needs five tires and that replacement does not occur within a year after a motor vehicle is sold. Five tires per new motor vehicle is somewhat low because trucks and buses are also incorporated in the number of motor vehicles and usually have more than five tires.

Jovanovic and MacDonald use a simple demand function: \( Q_t = c_0 P_t^{-c_1} \). They find estimates for \( c_0 \) of 31.358 and for \( c_1 \) of 0.763. It implies that a 1% increase in price leads to a 0.8% decrease in the demand for tires. However, Jovanovic and MacDonald ignore not only the impact of the number of new and old motor vehicles on the demand for tires, but also the improvement in the quality of both tires and roads. Therefore, \( c_1 \) cannot be interpreted as a price elasticity of the demand for automobile tires. The plot of the rate of replacement variable \( (Q_t - 5M_t)/S_t \) shows a marked decline from about 5 in the period
1910-17 until around 1.5 in the period 1930-73 (1942-46 excluded).\footnote{We exclude the Second World War period 1942-46 here and in the rest of the chapter because tire and car output were restricted.} In this period the price index also declines from about 5.7 to about 1.1. The main reason why total demand increases in this period is not because of lower prices but because of the strong increase in the number of motor vehicles. The downward movement of the rate of replacement of tires is a consequence of large improvements in the quality of both roads and tires. The average life of a tire increased from six months in 1910 to more than two years in the late 1930s (Reynolds (1938)). The average life did not increase much during the first decades after World War II. It was the introduction of the radial tire in the 1970s which increased the average life further. We claim that the price elasticity found by Jovanovic and MacDonald is incorrect and that this may bias other results. Therefore an alternative demand function is proposed which is a linear function of the number of motor vehicles produced and registered:

\begin{equation}
Q_t = a_0 M_t + (a_1 + a_2 QUAL_t + a_3 \log(P_t)) S_t + \epsilon_t^Q.
\end{equation}

Because the quality improvement in tires starts in the beginning of this century, we introduce a quality index, $QUAL_t$, which equals "year - 1930" for the period 1913-30 and 0 afterwards. We assume that the price effect on the rate of replacement can be represented best by the log of the price index. From equation (1) one then finds a price elasticity of demand of $a_3 S_t / Q_t$. We expect $a_3$ to be somewhat larger than five because each new motor vehicle needs at least five tires.\footnote{Note that the number of passenger cars as a percentage of total motor vehicle factory sales was between 74 and 91 per cent from the First World War up to 1973 (U.S. Department of Commerce (1975)). As a consequence, the number of motor trucks and buses as a percentage of total sales has been between 9 and 26 per cent.} A positive value of $a_1$, in excess of the minimum of $a_2 QUAL_t + a_3 \log(P_t)$ is expected because of positive replacement demand and $a_2$ and $a_3$ are expected to be negative because both a higher quality and a higher price lead to a lower rate of replacement of tires.

The price level of a good is determined by two factors, viz. the level of costs for producing such a good and the degree of competition prevailing in the market. We decompose these two elements in our price equation. The degree of competition is modelled as a non-linear function of the total number of producers. The level of costs of producing tires is assumed to consist of three parts. The first part refers to costs which are constant and unavoidable such as transportation costs. The second part is the cost of rubber. The third part are costs which diminish as a result of learning-by-doing. We will first discuss this last part.
Learning-by-doing is essential for understanding technological development in most industries. As a result from experience with the production process, firms are able to save on production costs. Workers are found to increase their skills in specific tasks through repetition, the routing and handling of materials and planning of required maintenance are steadily improved over time, employees and tasks become better matched (Bahk and Gort (1993)).\textsuperscript{77} These learning processes will be most forceful in the first years after the development of the production process when many elements of the process are still to be optimized. That is, the learning rate declines with the age of the production process or the cumulative output produced. Jovanovic and MacDonald assume that technological know-how can be kept entirely proprietary. In their model firms cannot imitate their more efficient rivals but can only acquire know-how by chance. Their analysis indicates that total production costs of high-tech firms were only about one per cent of that of low-tech firms at the same level of production. We doubt that any such cost differences have ever existed in the industry and more generally that technological know-how was completely internalized within the firm (Ghemawat and Spence (1985) and Irwin and Klenow (1994)).

The second equation relates the real price index of tires, \( P_t \), to learning economies, the price of rubber, \( R_t \), and the number of firms, \( N_t \). We start by considering the extent of the learning economies in the U.S. automobile tire industry.\textsuperscript{78} The estimate of the learning rate will be used in our subsequent analysis. A common specification for estimating learning economies is the regression of the log of the price on the log of cumulative output (Irwin and Klenow (1994)). The reason for this is the lack of data on production costs. Of course, this procedure is only legitimate if the price and marginal costs are highly correlated. The regression results over the period 1913-73 are as follows (standard error between brackets, DW is the Durbin-Watson statistic):

\[
(8.2) \log(P_t) = 2.928 - 0.371 \log\left(\sum_{s=1}^{t-1} Q_s\right) \quad R^2 = 0.902 \quad DW = 0.25
\]

(0.115) (0.017)

The low Durbin-Watson statistic indicates that the residuals possibly have an autocorrelation coefficient of unity. This is confirmed by the adjusted Dickey-Fuller test statistic (with a constant) of -2.32. The hypothesis of a unit root in the residuals cannot be rejected. We therefore estimate the equation in first differences to find (standard errors between

\textsuperscript{77}See Lucas (1993) for empirical evidence on the productivity effects of learning by doing in the production of nearly 2,500 Liberty Ships in 14 shipyards from 1941 through 1944. The reductions in manhours per ship with each doubling of cumulative output ranged from 12 to 24 per cent.

\textsuperscript{78}In this chapter we focus on learning-by-doing at the industry level. Learning-by-doing at the firm level would entail firms having a head start also to be among the least cost producers in later stages of the industry life cycle. The model which is developed in the next chapter displays a similar kind of first-mover advantages.
brackets):

\[
\frac{\Delta \log(P_i)}{0.080} = -0.312 \Delta \log \left( \sum_{s=1}^{t-1} Q_s \right)
\]

\[
R^2 = 0.090 \quad DW = 1.64
\]

This equation implies that with each doubling of cumulative output costs fall with 19.4%. This rate is close to that found in the semiconductor industry (Irwin and Klenow (1994), Table 1). It is noted that from 1914 to 1929 the automobile tire industry led all U.S. manufacturing in terms of growth in output per man-hour (French, p. 52). An alternative measure of learning-by-doing is the age of the industry. The relation using this measure is also estimated in first differences (standard errors between brackets):

\[
(8.4) \Delta \log(P_i) = -1.134 \Delta \log(t-1905.85)
\]

\[
R^2 = 0.127 \quad DW = 1.57
\]

According to this equation the leaning-by-doing in the industry and hence the industry itself started at the end of 1905. This corresponds well with the Thomas Register of American Manufacturers which gives 1906 as the earliest date of positive output (Jovanovic and MacDonald, p. 335). From here we will assume that the index \( t \) is the year minus 1905. We use the age of the industry as measure of learning-by-doing in the empirical analysis. However, the main conclusions are not affected when using cumulative output as a measure.

Another important determinant of the cost of tires is the price of natural rubber. For example, the very strong decrease in the prices of tires in the early 1920s was triggered by the collapse of the natural rubber price in that period (French, p. 40). The role of natural rubber declined after World War II when synthetic rubber was introduced commercially. The simple correlation coefficient between the rubber price index and the tire price index is 0.90 for the period 1913-73, while it is 0.89 for the period 1913-41 and only 0.74 for the after-war period 1947-73. Rubber prices fluctuated heavily during the 1920s and 1930s. The rubber price index almost tripled during the London rubber boom of 1925 and as a consequence the tire price index rose for the first time since the First World War. During the Great Depression rubber prices collapsed and a pound of rubber could be bought for as low a price as two cents in 1932 (Coates (1987), p. 255). In that same year the tire price index reached its prewar minimum.

The weighted average of marginal costs for producers of tires is assumed to be a linear combination of constant unavoidable costs, \( b_1 \), of the cost of rubber, \( b_2 R_t \), and of costs subject to learning economies, \( b_3 t^{-b_4} \). In sum, our approximation for the average marginal production costs is \( b_1 + b_2 R_t + b_3 t^{-b_4} \), where \( b_4 \) is fixed at 1.13 in the empirical analysis in order to avoid high correlation between the estimates of \( b_3 \) and \( b_4 \).

Tire prices are not only influenced by the cost of tires but also by the degree of
competition. In times of severe competition the prices of tires are closer to the cost of tires than in times when there is some degree of collusion in the industry. We take the degree of competition to be a function of the number of firms in the industry. Tire prices are expected to be influenced negatively by the number of firms but decreasingly so \((\partial P_t/\partial N_t < 0, \partial^2 P_t/\partial N_t^2 > 0)\). Cowling and Waterson (1976) derive for a Cournot oligopoly the following relation between price \((\text{PRICE})\), a weighted average of marginal costs \((\text{MC})\), the price elasticity of demand \((\text{ELAS})\) and the Herfindahl index \((\text{HERF})\):

\[ (8.5) \text{PRICE} = (1+\text{HERF/ELAS})^{-1}\text{MC} \]

We adopt this price relation and approximate the Herfindahl index by \(N^{-k}\), where \(k > 0\). Approximating the degree of competition by the number of competitors in the market is also predominant in the organization ecology work of Hannan and Carroll (1992). Two options, one with a price elasticity derived from equation (8.1) and one with a constant price elasticity in time, \(e\), are as follows:

\[ (8.6a) P_t = \left(1 + \frac{N_t^{-b_0}}{\delta_3 S/Q_t} \right)^{-1} \left( b_1 + b_2 R_t + b_3 t^{-1.13} \right) + e_t \]

\[ (8.6b) P_t = \left(1 + N_t^{-b_0/e} \right)^{-1} \left( b_1 + b_2 R_t + b_3 t^{-1.13} \right) + e_t \]

The advantage of having the elasticity constant in time is that we do not have simultaneity between price, \(P_t\), and output, \(Q_t\). That is, the model is recursive and the equations can be estimated separately.

The last equation relates the net entry rate, \((N_t - N_{t-1})/N_{t-1}\), to expected profit opportunities. The expectation of profit opportunities is supposed to be influenced by three factors. The first factor is the profit margin of incumbents in the preceding period derived above. A high average profit margin in the current period indicates a disequilibrium situation of a too low number of firms operating in the market. Potential entrepreneurs react to this disequilibrium situation by entering. If we confine ourselves to this one factor, this would imply the following 'basic form' for the net entry rate \((\text{NER})\) equation.\(^{79}\)

\[ \text{NER} = \nu (\text{PRICE}/\text{MC} - w) \]

---

\(^{79}\)This equation can be derived as follows. Assume that entrepreneurs have the option to operate a plant at fixed cost \(F\) capable of producing \(q_t\) tires. This option is profitable in case \(\pi = (\text{PRICE} - \text{MC})q_t - F > 0\). If the net entry rate is a linear function of profit \(\pi\) minus the opportunity costs of the entrepreneur, then we have equation (8.7). Geroski and Murtin (1991) use an alternative incentive for entry in their equation (2). It implies an equation like \(\text{NER} = \nu (\text{PRICE}/\text{MC} - w)\), where \(\nu\) and \(w\) have the same kind of interpretation as in equation (8.7). We have also estimated an equation of this type. However, the estimated fit of this equation was worse when compared to that of the 'basic form' equation (8.7).
\[(8.7) NER = \nu (\text{PRICE-MC-w})\]

The parameter \(\nu\) is the extent to which firms entering and exiting the market react to a difference between the actual average profit margin and their critical profit margin, \(w\). The second factor is the growth of demand for tires in the preceding period. It may be more costly for incumbents to adjust their capacity than it is for new firms to enter the industry (Hause and Du Rietz (1984)). A growing demand may also increase the number of viable market niches. We use both the growth rate of demand, \(\Delta Q_{t-1}/Q_{t-1}\), and the change in this growth rate, \(\Delta (\Delta Q_{t-1}/Q_{t-1})\), as determinants of the net entry rate. The third factor is the number of firms already in the market. In case this number is high this is a disincentive for new firms to enter as most market niches will already have been filled. Acs and Audretsch (1989b) and Rosenbaum (1993) also find that net entry rates tend to be lower in industries in which there is already a considerable presence of small firms. An important market niche in the tire industry has been the production of truck and bus tires (French, p. 49). Specialist tire production may provide better margins than the average profit margin in the industry. Since the 1970s small firms have increasingly concentrated on niches in specialist sectors and some larger firms have abandoned specialist tire production (French, p. 114-6). We use the logarithm of the number of firms in our empirical analysis to prevent the large number of very small entrants during the early 1920s to have a too high impact on the results. Adding the second and third factor to equation (8.7), we have

\[(8.8) \frac{\Delta N_{t}}{N_{t-1}} = c_0 (P_{t-1} - c_1 - c_2 R_{t-1} - c_3 (t-1)^{1.13}) + c_4 \frac{\Delta Q_{t-1}}{Q_{t-2}} + c_5 \frac{\Delta Q_{t-1}}{Q_{t-2}} + c_6 \log(N_{t-1}) + \epsilon'_t\]

Nearly all the firms which enter and exit the market are small (MacDonald (1986)). This implies that it must be profitable to produce tires already at a small scale of production in order for small producers to enter the market or for small incumbents not to exit. Small firms usually cannot survive the financial pressure of a low or even negative profit margin for a long time. We claim that the rapid shakeout of firms in the 1920s and early 1930s was at least partly due to an unexpected low profit margin in the tire industry. Reynolds (1938, p. 464) reports that the average rate of profit in tire manufacturing was only about half of that in all manufacturing in this period. In the period 1920-35 there was only one year (1925) in which the percentage profit in tire manufacturing was higher than that for the entire manufacturing sector.

Among the factors which caused the strong competition and hence low profit margins were the violent fluctuations of rubber prices, the rise of the large retailer and the establishment of company stores, the shakeout in the number of car producers, decreasing
growth rates of demand for automobile tires, and excess capacity (Reynolds (1938)).\footnote{The shakeout in the number of car producers was in the same period as the shakeout in the number of tire producers. Klepper and Simons (1993) report a decrease in the number of car producers from 175 in 1921 to 55 in 1925. Utterback and Suárez (1993) use a somewhat different data source and find a peak of 75 producers in 1923 followed by a rapid fall to only about 15 in the 1930s. The simultaneous decreases in the number of car and tire producers seems to be directly in line with Galbraith’s ‘Countervailing Power’ argument (Schumacher (1991)). It is certainly so that most small and new producers were not able to compete with the established contracts between large tire and car producers. However, the share of the replacement market of automobile tires in total tire sales was about 70% during the 1920s (Reynolds (1938), Table I). The development of the car industry into a highly concentrated industry is therefore likely to have affected prices of only about 30% of the tire sales.} The strong decline of the rubber prices in the 1920-21 recession gave new and small firms a temporary advantage because they could buy at current low price levels while larger firms had to work off expensive inventories (French, p. 40). The London rubber boom of 1925 had the reverse effect. Small firms faced high rubber costs while their larger rivals were still using cheap rubber (French, p. 52-3).

Equations (8.5) and (8.7) imply an equilibrium number of firms. In case this number is at the equilibrium value, we have of course NER=0. This implies that

\[(8.9) \text{ HERF/ELAS} = -(w/MC)/(1+w/MC)\]

It is apparent that a decrease in marginal costs over time leads to a higher Herfindahl index or, equivalently, to a smaller number of firms supported by the market. The equilibrium number of firms in the model is therefore bound to decrease in an industry in which learning economies are important. In the end only a few very large producers may survive. This scenario has become real in the tire industry which is now dominated by only a handful of producers worldwide. The actual number of firms may of course deviate from the equilibrium number of firms. This deviation explains the typical pattern in many industries of the number of firms to increase and subsequently to decrease. The reason for this is as follows. In the first years of the industry the actual number of firms rises slowly to the equilibrium value which is still quite high at that date. The actual number may even overshoot this equilibrium value due to the lag between perceiving the incentives and starting a production unit. Because the equilibrium number of firms is declining steadily as a result of learning economies the actual number of firms is also bound to decline. The extent of this decline depends of course upon the difference between the actual and equilibrium number of producers.

Our model of the evolution of the tire industry is a system of recursive equations in case the price elasticity of demand is assumed to be constant in time. The tire price, \(P_{t-1}\), demand for tires, \(Q_{t-1}\), and number of firms, \(N_{t-1}\), in the preceding period determine the number of firms, \(N_t\), in the current period. This number of firms, \(N_t\), determines the tire
price, \( P_t \), in the current period and this price again determines the demand for tires, \( Q_t \). Because the model is recursive we are able to estimate each equation separately.

### 8.3 Empirical results

The data are available for the period 1913-1973 and include a United States price index of tires, the output of tires and the number of U.S. firms producing tires. The Second World War period 1942-46 is left out of the analysis because tire and car output were restricted. In Figure 8.1 the price index of tires is displayed. The tire price index declined strongly from 7.7 in 1913 to 1.1 in the early 1930s and then rose slowly to 1.4 in the years before the war. After the war the price index started from a low point of 1.0 and then rose slowly until 1.3 in the mid 1950s, followed by a steady decline to 0.9 in the early 1970s. The output of automobile tires can be found in Figure 8.2. The output rose strongly from 6 million tires in 1913 to 78 million tires in 1928. During the Great Depression this number fell to 40 million tires but then rose gradually to 120 million tires in 1960. The rise in the output then accelerated in the 1960s and in 1968 more than 200 million automobile tires were produced. The nonmonotonicity in the firm numbers in the tires industry can be seen in Figure 8.3. The number of producers had only been 10 in 1906 but had already increased to 74 by 1913. The increase in the number continued until 1922 when there were 275 firms. The number of firms then dropped very fast to about 120 before the Great Depression and even to 62 in 1933.\(^{81}\) This was followed by a monotonic decline in the number of firms to around 30 in the 1970s.

#### Table 8.1: Summary statistics

<table>
<thead>
<tr>
<th>Description</th>
<th>Symbol</th>
<th>Mean</th>
<th>Stdev</th>
<th>Growth</th>
<th>Mean</th>
<th>Stdev</th>
</tr>
</thead>
<tbody>
<tr>
<td>Output of tires</td>
<td>Q</td>
<td>90.07</td>
<td>58.96</td>
<td>8.4%</td>
<td>19.4%</td>
<td></td>
</tr>
<tr>
<td>Price index tires</td>
<td>P</td>
<td>1.82</td>
<td>1.45</td>
<td>-3.5%</td>
<td>9.1%</td>
<td></td>
</tr>
<tr>
<td>Output of motor vehicles</td>
<td>M</td>
<td>5.55</td>
<td>3.14</td>
<td>10.1%</td>
<td>28.8%</td>
<td></td>
</tr>
<tr>
<td>Motor vehicles registered</td>
<td>S</td>
<td>42.04</td>
<td>31.15</td>
<td>10.1%</td>
<td>13.5%</td>
<td></td>
</tr>
<tr>
<td>Price index rubber</td>
<td>R</td>
<td>2.05</td>
<td>1.88</td>
<td>2.2%</td>
<td>49.7%</td>
<td></td>
</tr>
<tr>
<td>Number of firms</td>
<td>N</td>
<td>80.45</td>
<td>60.57</td>
<td>-1.9%</td>
<td>11.8%</td>
<td></td>
</tr>
</tbody>
</table>

Note: 'Stdev' stands for the standard deviation of the variables and of the percentage growth of the variables, respectively.

---

\(^{81}\)See Fricke (1982) for a study on the remarkable struggle to survive of the McCreary Tire & Rubber Company during the 1930s. This company was the smallest of the tire manufacturers which survived the Great Depression.
Figure 8.1: The price index of tires (1967=1)

Figure 8.2: The output of tires (in millions)

Figure 8.3: The number of firms

Figure 8.4: The actual and predicted net entry rate
Summary statistics of the variables can be found in Table 8.1. The sources of the data are Table A1 of Jovanovic and MacDonald (1994a) for the data on the price index of tires, the output of tires and the number of firms, U.S. Department of Commerce (1975, 1989) for the data on the number of motor vehicles produced and registered, and Orton (1927), Coates (1987) and Citibase for the data on the natural rubber price. The high volatility of the rubber price (before the Second World War) is clearly visible in the high standard deviation of the percentage yearly growth of the price index.

We first examine our claim that the shakeout has been caused by a strong price competition in the tire industry followed by a strong decline in automobile tire demand. In Table 8.2 we report the tire and rubber price index, the demand for tires and the number of producers for five successive periods. In the 1915-22 period demand for tires grew strongly and the decline of tire prices was less than that of rubber prices. As a consequence, the number of firms increased at an average rate of 15% per year. The 1923-29 period is characterized by fierce competition: tire prices declined more than in the previous time period while rubber prices rose. The growth rate of demand was also on average nine percentage points lower. The net entry rate was -11% in that period. In the 1930-33 period price competition was still strong (French, ch. 6). The cost of rubber did collapse but was only a fraction of tire production costs in this time period. More important, demand for tires declined 15% every year during the Great Depression. In this time period the tire industry suffered its most profound reorganization (French, p. 60). The number of firms also decreased at a yearly rate of 15%.

Table 8.2: Changes in Price Indices, Demand and Number of Firms

<table>
<thead>
<tr>
<th>Period</th>
<th>$% \Delta P_{t-1}$</th>
<th>$% \Delta R_{t-1}$</th>
<th>$% \Delta Q_{t-1}$</th>
<th>$% \Delta N_{t}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1915-22</td>
<td>-0.084</td>
<td>-0.219</td>
<td>0.248</td>
<td>0.149</td>
</tr>
<tr>
<td>1923-29</td>
<td>-0.125</td>
<td>0.142</td>
<td>0.158</td>
<td>-0.110</td>
</tr>
<tr>
<td>1930-33</td>
<td>-0.049</td>
<td>-0.342</td>
<td>-0.150</td>
<td>-0.151</td>
</tr>
<tr>
<td>1934-41</td>
<td>0.028</td>
<td>0.462</td>
<td>0.063</td>
<td>-0.011</td>
</tr>
<tr>
<td>1948-73</td>
<td>-0.015</td>
<td>-0.022</td>
<td>0.063</td>
<td>-0.009</td>
</tr>
</tbody>
</table>

The regression results of the three equations (8.1), (8.6) and (8.8) are presented in Table 8.3. We have estimated the output equation (8.1) using the entire sample, using the period before the war and the period after the war. The estimate of the effect of the output of motor vehicles on the number of tires is somewhat above 5. This is reasonable as new trucks and buses usually have more than five wheels. The effect of motor vehicles already registered for one year or more is a constant of about 1.5 minus the quality effect and minus the price effect. Both the quality and price effects are highly significant. The higher the quality of the tires (or roads) the lower is the replacement demand per registered motor vehicle. The
higher is the price of the tires, the more costly it is to replace tires and hence the lower is the rate of replacement. The ratio of registered motor vehicles to the output of tires \((S_t/Q_t)\) rises from about 0.1 at the beginning of the sample period to 0.5 in the early 1930s and remains around that value until the end of the sample period. This implies that also the price elasticity of demand became somewhat stronger over the first decades of the sample and stabilized afterwards. We use -0.3, the mean of -0.1 and -0.5, as an estimate of the price elasticity \(e\) in case this elasticity is assumed constant over time as in equation (8.6b).

### Table 8.3: Empirical results of the three-equation evolution model

<table>
<thead>
<tr>
<th>Output equation (8.1)</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1913-73</td>
<td>1913-41</td>
<td>1947-73</td>
<td></td>
</tr>
<tr>
<td>(a_0)</td>
<td>5.642</td>
<td>5.880</td>
<td>5.000</td>
</tr>
<tr>
<td>(0.993)</td>
<td>(1.044)</td>
<td>(1.490)</td>
<td></td>
</tr>
<tr>
<td>(a_1)</td>
<td>1.368</td>
<td>1.625</td>
<td>1.429</td>
</tr>
<tr>
<td>(0.118)</td>
<td>(0.188)</td>
<td>(0.175)</td>
<td></td>
</tr>
<tr>
<td>(a_2)</td>
<td>-0.325</td>
<td>-0.317</td>
<td></td>
</tr>
<tr>
<td>(0.046)</td>
<td>(0.051)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(a_3)</td>
<td>-1.031</td>
<td>-1.365</td>
<td>-1.197</td>
</tr>
<tr>
<td>(0.180)</td>
<td>(0.408)</td>
<td>(0.255)</td>
<td></td>
</tr>
<tr>
<td>(R^2)</td>
<td>0.980</td>
<td>0.940</td>
<td>0.958</td>
</tr>
<tr>
<td>DW</td>
<td>1.29</td>
<td>1.38</td>
<td>1.43</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Price equation (8.6)</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>(b_0)</td>
<td>1.986</td>
<td>1.158</td>
<td>0.737</td>
<td>0.807</td>
<td>0.715</td>
<td>0.786</td>
<td>0.621</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(15.84)</td>
<td>(1.052)</td>
<td>(0.373)</td>
<td>(0.657)</td>
<td>(0.324)</td>
<td>(0.235)</td>
<td>(0.246)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(b_1)</td>
<td>0.038</td>
<td>0.011</td>
<td>-0.028</td>
<td>0.442</td>
<td>0.090</td>
<td>-0.006</td>
<td>-0.119</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(0.097)</td>
<td>(0.145)</td>
<td>(0.264)</td>
<td>(4.517)</td>
<td>(2.455)</td>
<td>(0.102)</td>
<td>(1.093)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(b_2)</td>
<td>0.183</td>
<td>0.172</td>
<td>0.129</td>
<td>0.052</td>
<td>0.048</td>
<td>0.134</td>
<td>0.053</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(0.080)</td>
<td>(0.077)</td>
<td>(0.070)</td>
<td>(0.049)</td>
<td>(0.035)</td>
<td>(0.062)</td>
<td>(0.035)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(b_3)</td>
<td>59.000</td>
<td>59.364</td>
<td>54.429</td>
<td>61.858</td>
<td>101.177</td>
<td>55.057</td>
<td>51.892</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(6.679)</td>
<td>(6.476)</td>
<td>(5.872)</td>
<td>(28.56)</td>
<td>(28.03)</td>
<td>(5.095)</td>
<td>(21.95)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(R^2)</td>
<td>0.930</td>
<td>0.930</td>
<td>0.769</td>
<td>0.443</td>
<td>0.442</td>
<td>0.952</td>
<td>0.691</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DW</td>
<td>0.33</td>
<td>0.33</td>
<td>0.22</td>
<td>1.69</td>
<td>1.72</td>
<td>0.35</td>
<td>1.52</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

continued on next page
Net entry equation (8.8)

<table>
<thead>
<tr>
<th></th>
<th>1914-73</th>
<th>1914-73</th>
<th>1914-73</th>
<th>1914-73</th>
<th>1914-73</th>
</tr>
</thead>
<tbody>
<tr>
<td>$c_0$</td>
<td>0.132</td>
<td>0.133</td>
<td>0.139</td>
<td>0.108</td>
<td>0.116</td>
</tr>
<tr>
<td></td>
<td>(0.030)</td>
<td>(0.030)</td>
<td>(0.029)</td>
<td>(0.030)</td>
<td>(0.030)</td>
</tr>
<tr>
<td>$c_1$</td>
<td>0.404</td>
<td>0.402</td>
<td>0.392</td>
<td>-2.376</td>
<td>-1.941</td>
</tr>
<tr>
<td></td>
<td>(0.197)</td>
<td>(0.194)</td>
<td>(0.180)</td>
<td>(1.575)</td>
<td>(1.377)</td>
</tr>
<tr>
<td>$c_2$</td>
<td>0.082</td>
<td>0.117</td>
<td>0.076</td>
<td>0.308</td>
<td>0.224</td>
</tr>
<tr>
<td></td>
<td>(0.096)</td>
<td>(0.095)</td>
<td>(0.088)</td>
<td>(0.158)</td>
<td>(0.134)</td>
</tr>
<tr>
<td>$c_3$</td>
<td>58.674</td>
<td>58.319</td>
<td>59.277</td>
<td>11.917</td>
<td>19.495</td>
</tr>
<tr>
<td></td>
<td>(11.67)</td>
<td>(11.49)</td>
<td>(10.69)</td>
<td>(26.97)</td>
<td>(23.52)</td>
</tr>
<tr>
<td>$c_4$</td>
<td>0.103</td>
<td>0.113</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.077)</td>
<td>(0.074)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$c_5$</td>
<td></td>
<td></td>
<td>0.089</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(0.047)</td>
<td></td>
<td>(0.046)</td>
</tr>
<tr>
<td>$c_6$</td>
<td></td>
<td></td>
<td>-0.090</td>
<td></td>
<td>-0.081</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(0.037)</td>
<td></td>
<td>(0.036)</td>
</tr>
<tr>
<td>$R^2$</td>
<td>0.396</td>
<td>0.417</td>
<td>0.443</td>
<td>0.472</td>
<td>0.487</td>
</tr>
<tr>
<td>DW</td>
<td>2.23</td>
<td>2.19</td>
<td>2.19</td>
<td>2.27</td>
<td>2.25</td>
</tr>
</tbody>
</table>

Note: Standard errors between brackets. DW is the Durbin-Watson statistic. The symbol $P_e$ stands for the use of the constant price elasticity of -0.3 (equation (8.6b)). In the columns with $ΔP_e$ the equation is estimated in first differences. In the last two columns of the price equation a dummy for the years 1917 and 1918 is added.

The estimation of the price equation requires the decision whether the data in the first years of the sample period should be incorporated or not. During the First World War the prices of tires were subject to regulation and did not decrease much despite the increase in output. In the first column of Table 8.3 we use equation (8.6a) and include the data of the 1913-1918 period. In all other columns equation (8.6b) and a constant elasticity of -0.3 is used. Because of the presence of a unit root in the residuals, we also estimated equation (8.6b) in first differences. In the second and third column of Table 8.3 results are shown including data for the 1913-1918 period and excluding this period. Both for equation (8.6a) and (8.6b) for the 1913-1973 period the estimated value of $k$ is above the theoretical maximum of unity and is estimated imprecisely.\(^{62}\) In the fourth and fifth column the results using first differences are reported. The price index of tires shows an unexpected lack of decrease especially in the last two years of the First World War, i.e. 1917 and 1918. We introduce a dummy which equals one for the years 1917 and 1918 and zero in other years.

\(^{62}\)See the discussion below equation (8.5).
In the last two columns this dummy is added to equation (8.6b). In the sixth column the estimated parameter of this dummy is 1.399 (standard error is 0.291) and in the last column it is 0.822 (standard error is 0.130).

The estimation results provide support for a negative effect of the number of firms on the price. The power to which this number is raised lies around -0.7. This points at a Herfindahl index of around 0.09 at the end of the observation period and of around 0.06 during the early 1930s. This is lower than we expected. The market shares of unit tires sales of the four leading companies (Goodyear, Firestone, U.S. Rubber and Goodrich) were about 30, 15, 19 and 8 per cent in 1933 (French, p. 47). The market shares in original equipment sales of the leading companies in 1970 (Goodyear, Firestone, Uniroyal and Goodrich) were 32, 27, 18 and 16 per cent, spanning almost the entire market (French, p. 111). However, despite the dominance of the big four tire manufacturers the industry was very competitive before the World War II (Reynolds (1938)). This may be a reason for the estimate of the Herfindahl index to be biased downwards. Furthermore, one should also take the large standard error of the estimate of the power into consideration. The constant in the marginal cost part of the equation \( b_1 \) is insignificant. This indicates that the costs of tires are predominantly composed of rubber costs and costs which are subject to learning-by-doing. Both the effects of the natural rubber price \( b_2 \) and costs which decrease through a learning-by-doing process \( b_3 \) are positive although the effect of the rubber price is not or barely significantly different from zero.

The net entry equation (8.8) was first estimated in its basic form of equation (8.7). The results can be found in the first column. The results for the marginal cost approximation are similar to those found for the price equation (8.6b) with data for the entire 1914-73 period. In the second column we introduce the growth rate of demand into the model. The effect of the growth rate on the net entry rate is positive but insignificant. However, the effect of the change in this growth rate of demand is significant, as can be seen in the third column of the table. Incumbents may expect the growth rate of demand to remain constant and adjust their capacity accordingly. As a consequence, a higher growth rate provides room for entrants while a lower growth rate leads to excess capacity. The latter phenomenon occurred in the late 1920s when there was large plant construction (Reynolds (1938), p. 465). In the last two columns the logarithm of the number of firms is added as an indicator of the extent to which market niches are already filled. The effect of is negative, as expected, and significant. This confirms earlier empirical findings that markets where there is already a high small firm presence on average have lower net entry rates. However, the results are now less comparable with those found for the price equation. In the simulation experiment we will not include \( \log(N_{t-1}) \) into the net entry equation.

We investigate the fit of the number of firms series by comparing the actual and fitted net entry rate time series. The regression equation where the change in the growth rate of demand and the logarithm of the number of firms are incorporated is used for this purpose.
The solid line in Figure 8.4 is the series of actual net entry rates, while the dashed line is the series of predicted rates. Note that the 1942-47 period has been left out. Both the actual and predicted net entry rates change from positive to negative in 1923. Therefore, the regression results correctly predict the peak of the number of firms in 1922. The sharply higher net entry rate in 1922 was not predicted by the model. One may wonder whether the bulk of entrants in the early 1920s made a well-considered decision. Many of these entrepreneurs had very brief careers in tire manufacturing (French, p. 47). The predicted net entry rate in the Jovanovic-MacDonald model is quite unrealistic. It is zero all of the time except for one period of massive entry and one period of massive exit (cf. their figure 3a). Our predicted net entry rates seem to follow the actual rates closely.

8.4 Simulated industry evolution

Using the three equation we simulate the industry evolution starting from the year 1914. We use the following three estimated equations:

\[
\frac{\Delta N_t}{N_{t-1}} = 0.139 (P_{t-1} - 0.392 - 0.076 R_{t-1} - 59.277 (t-1)^{-1.13}) + 0.096 \frac{\Delta Q_{t-1}}{Q_{t-2}}.
\]

\[
P_t = (1 - N_t^{-0.786} / 0.3)^{-1} (-0.006 + 0.134 R_t + 55.057 t^{-1.13});
\]

\[
Q_t = 5.642 M_t + (1.368 - 0.325 QUAL_t - 1.031 \log(P_t)) S_t.
\]

The equations are put according to the order in which the variables are determined. Equation (8.8') uses data on output of the three previous periods and the price level in the previous period to determine the number of firms \( N_t \). Equation (8.6b') uses this number of firms to determine the price level. For the years 1917 and 1918 the estimated value of the dummy-parameter 1.399 is added to (8.6b) in the simulation experiment. Equation (8.1') uses this price level to determine the output of tires (in millions). Next to data on the rubber prices, the sales and stock of cars and the quality of tires, we require the output level in the years 1911 through 1913 and the number of firms and the price level in the year 1913. A simple computer program of the recursive system can then simulate the path of the output, price level and number of firms from 1914 through 1973.

Table 8.4 shows the simulation results for some selected years. Although the simulated series are quite close to the actual series from 1930 on, they clearly differ during the 1920s. The simulated price level is much lower than the actual price level in the years 1919-1921 leading to lower net entry rates than which have been recorded in reality. The simulated

---

83Klepper and Miller (1995) develop a theory of industry shakeouts which is based on this idea of 'overshooting'.

137
peak in the number of producers is 131 in the year 1919, three years before the actual peak in this number. After 1921 the simulated price level is much higher than the actual price level leading to only a small decrease in the number of producers when compared to the actual data. In the last column of Table 8.4 we have replaced equation (8.6b') with the actual price data. That is, we simulate output and number of firms using the actual data of prices. The simulated series of number of firms has a peak of 167 in the number of producers in the year 1922 now. The peak year corresponds to that in the real data but, again, the shakeout is less severe than the real data show. The model cannot adequately predict the massive entry of small firms in the early 1920s. However, most other elements of the evolution of the tire industry are predicted reasonably well.

Table 8.4: Simulation results

| Year | $Q_{sim}$ | $P_{sim}$ | $N_{sim}$ | $N_{sim}|P$ |
|------|-----------|-----------|-----------|-------------|
| 1917 | 26        | 5.66      | 133       | 93          |
| 1918 | 23        | 5.57      | 143       | 112         |
| 1919 | 33        | 4.43      | 144       | 131         |
| 1920 | 32        | 4.25      | 181       | 129         |
| 1921 | 29        | 3.66      | 199       | 124         |
| 1922 | 41        | 2.52      | 275       | 118         |
| 1923 | 45        | 2.35      | 246       | 114         |
| 1924 | 52        | 1.98      | 213       | 112         |
| 1930 | 51        | 1.13      | 96        | 94          | 124          |
| 1940 | 59        | 1.35      | 52        | 65          | 81           |
| 1950 | 93        | 1.10      | 46        | 56          | 68           |
| 1960 | 120       | 1.09      | 44        | 43          | 49           |
| 1970 | 190       | 0.94      | 30        | 31          | 35           |
| 1973 | 223       | 0.84      | 38        | 29          | 32           |

Note: The subscript 'sim' indicates simulated series. The first year of simulation is 1914. In the last column the number of firms is simulated with actual price data instead of simulated price data.

8.5 Conclusion

Jovanovic and MacDonald (1994a) recently presented an explanation of the nonmonotonicity in firm numbers which is found in many young industries. They claim that the number of firms rises at first due to innovation opportunities but that subsequent failure to innovate then leads to a gradual or catastrophic shakeout. In this chapter, we provide an alternative model of the life cycle of a competitive industry. The model provides a more accurate fit of the number of firms series when applied to the data of the automobile tire industry over the
period 1913-1973, as used by Jovanovic and MacDonald. It is especially applicable to industries in which a single innovation is not so dominant as their model assumes. We think that this is the case in most industries over a period of several decades. We have also built a more realistic demand function for automobile tires than the one used in the Jovanovic-MacDonald article. We have argued that an important part of the decline in tire prices in the 1920s and early 1930s has been caused by the downward trend of natural rubber prices in that period. Jovanovic and MacDonald do not consider this important variable in their model. We claim that the rapid shakeout of the number of firms in the U.S. tire industry has not been the result of any dominant innovation. Instead, it has been caused by a strong and persistent price competition in the tire industry followed by a strong decline in demand for tires during the Great Depression.

Our model consists of three equations. The first explains the output of tires using the number of motor vehicles, the quality of tires (and roads) and the price index of tires. The second equation relates the price index of tires to the competition in the market, measured by the Herfindahl index, to the natural rubber price index and to costs which decrease due to learning-by-doing. The last equation explains the net entry rate from the prevailing profit margin in the preceding period, the growth rate of demand for tires, and from the presence of small firms in the industry. The model predicts that the equilibrium number of firms decreases as marginal costs decline such as in the case of important learning economies.

The model is developed especially for explaining patterns of output, prices and number of firms over time in the tire industry. However, it can also be used to explain these same patterns in other industries with a relatively homogeneous product. Our model does require some more industry-specific information with regard to the demand function and marginal costs than the Jovanovic-MacDonald model. We claim that this extra information is essential for predicting the way in which output, prices and the number of firms evolve over time in an industry. It may provide answers as to why some industries have no shakeouts, some have a slow and others have a rapid shakeout (Gort and Klepper (1982), Klepper and Graddy (1990), Klepper and Simons (1993) and Klepper and Miller (1995)).
8.3 Conclusion

Jerez-Spinola and MacDonald (1994a) recently presented an explanation of the unpredictable
ity in firm behaviors which is based in many writing industries. They claim that the behav-
iors of firms are at least due to innovation opportunities but that subsequent behavior is more
then leads to a gradual or catastrophic breakdown. In this chapter, we present an alternative
model of the life cycle of a competitive industry. The model predicts a more accurate fit of
the number of firms since which agreed to the data of the automobile industry that the
CHAPTER 9

A FIRST-MOVER ADVANTAGE MODEL

In this chapter a new model of industry evolution is developed. This structural model entails first-mover advantages strong enough to have early entrants to be among the most successful firms throughout the industry life cycle. Elements of industry evolution like the rate of cost decreases and survival rates across cohorts are predicted by the model. We present some preliminary results of the new model to explain entry and exit rates during the evolution of the U.S. tire industry. The estimation process of the parameters of the model is performed using the simulated annealing algorithm. The proposed model is an alternative to the model developed in Chapter 8.
9.1 Introduction

The research into the developments of entry and exit rates over the evolution of an industry has benefitted from some important contributions in the last years. Examples are Agarwal and Gort (1996) who show how entry and exit rates depend on the stage of development of the market in the cycle from birth to maturity, Klepper (1996) who emphasizes differences in firm innovative capabilities and the importance of firm size in appropriating returns from innovation and Klepper and Miller (1995) who explain the shakeout of producers which occurs in many industries from an 'overshooting’ model. These authors are primarily interested in the explanation of the increase of the entry rate in the first years of the evolution of an industry followed by a period of decreasing entry rates and increasing exit rates again followed by a period of decreasing exit rates and low entry rates. The consequence of this time pattern of the entry and exit rates is that the number of producers shows a peak after which it decreases considerably. This fast decline is called the 'shakeout period' (Gort and Klepper (1982)).

The development of the number of firms which occurs in so many industries is again examined using a new theoretical model in Section 9.2. This is a first-mover advantage model in which early entrants may achieve a headstart by pursuing process innovation.\textsuperscript{84} The model predicts not only the entry and exit patterns over the evolution of the industry, but also the cohort of the exiting firms. It shows which firms are most likely to exit during the shakeout period and during the periods of slow decline in the number of producers thereafter. In Section 9.3 we present some preliminary results of the first-mover advantage model for a unique data set of the U.S. tire industry. This enables the reader to compare the predictions with the results presented in Chapter 8. A relatively new method of estimation is used in this section, the so-called method of simulated annealing. The research presented in the coming two sections is a first report on a combined research project together with Steven Klepper of Carnegie Mellon University.

9.2 A first-mover advantage model

Firms come and go during the evolution of an industry. Using a conceptually simple model we seek to explain these patterns of entry and exit over the life-cycle of an industry. We assume that there is a well-defined industry with firms producing one homogeneous good and a linear demand function which is constant over time. All firms are assumed to be price-takers. Supply and demand of the good are assumed to be equal in each period and

\textsuperscript{84}Szymanski et al. (1995) perform a meta-analysis and find that, on average, earlier entry is associated with greater market share. They discuss marketing mix elements which may account for this finding.

142
the price which results from this equilibrium condition is denoted by \( p_t \). Firms differ in two respects. The first difference between firms is the period in which they entered the industry, viz., their cohort. The sooner firms enter the more time they have to grow into a large firm. The second difference is the R&D capability which is assumed to be exogenously determined (Cohen and Klepper (1992b)). The higher the R&D capability the more cost savings firms can expect from their R&D efforts. We will concentrate on process R&D and the result of R&D efforts is a decrease in per unit costs of production.

A firm which entered in period \( k \) and has an R&D capability \( \beta_i \) is called a type \( j \) firm of cohort \( k \). The cost per unit, output and R&D expenditure of a type \( j \) firm of cohort \( k \) in period \( t \) are denoted by \( c_{jkt} \), \( q_{jkt} \) and \( r_{jkt} \), respectively. In case a firm does not perform R&D, for instance when \( \beta_j \) equals zero, then it can produce at a unit cost of \( c_i \). The best practice in producing the good is assumed to become public knowledge at the end of each year. This implies that \( c_i \) equals the lowest cost per unit of all firms in period \( t-1 \). All firms are assumed to maximize profits in each separate year. They do not take future profits or the behaviour of competitors into consideration when deciding how much to produce and how much to spend on R&D. Firms enter when they can make a profit and they exit when they make a loss. The number of potential entrants in each period is assumed to depend on the number of incumbents in the previous period. Their R&D abilities are drawn from a random distribution which does not change over time.

A firm of type \( j \) and cohort \( k \) maximizes its profit

\[
\pi_{jkt} = (p_t - c_{jkt}) q_{jkt} - r_{jkt} - f(q_{jkt}) (\Delta q_{jkt})^2
\]

where \( c_{jkt} = c_i - \delta_i \ln(r_{jkt}) \) where \( \delta_i = \beta_i \gamma_i \) for \( j = 1, \ldots, J \) and \( \beta_1 < \ldots < \beta_J \) and where \( \gamma_i \) is an index of technological opportunity. That is, profit equals sales minus production costs, R&D expenditure and adjustment costs. The function \( f \) has \( f(x) > 0 \) and \( f'(x) < 0 \) for all \( x > 0 \). The index of technological opportunity should generally decline over time and be so low as to guarantee that costs per unit cannot become negative. Maximizing with respect to \( r_{jkt} \) gives \( r_{jkt} = \delta_i q_{jkt} \). That is, R&D expenditures are proportional to the output of the firm for firms of the same type \( j \). We substitute this into the profit equation and as a consequence the first order condition for output \( q_{jkt} \) is:

\[
(9.1) \quad q_{jkt} = q_{jkt-1} + \frac{\delta_i \ln(q_{jkt}) + p_t - c_i - \delta_i \ln(\delta_i)}{2f(q_{jkt-1})}
\]

Maximum profit can therefore be written as a quadratic function in \( q_{jkt} \):

\[
(9.2) \quad \pi_{jkt} = f(q_{jkt-1})q_{jkt}^2 - \delta_i q_{jkt} - f(q_{jkt-1})q_{jkt-1}^2
\]

We now define \( x_{jkt-1} = \sqrt{\delta_i^2 + 4f(q_{jkt-1})q_{jkt-1}^2} \). The participation constraint, \( \pi_{jkt} > 0 \), is then
equivalent to $q_{jkt} > (\delta_{jt} + x_{jkt-1} + 1)/(2f(q_{jkt-1}))$. The critical participation price, $p_{jkt}$, at which $\pi_{jkt} = 0$, of a firm of type $j$ and cohort $k$ is therefore equal to:

$$(9.3) \quad p_{jkt} = c_t - \delta_{jt} \ln(\delta_{jt}) + \delta_{jt} x_{jkt-1} - 2f(q_{jkt-1}) q_{jkt-1} - \delta_{jt} \ln \left( \frac{\delta_{jt} x_{jkt-1}}{2f(q_{jkt-1})} \right)$$

This simplifies to $p_{jkt} = c_t$ for firms with $\beta_j = \delta_{jt} = 0$. For entrants the condition is also more simple: $p_{jkt} = c_t - 2\delta_{jt} \ln(\delta_{jt}) + 2\delta_{jt} + \delta_{jt} \ln(f(0))$. The critical participation price declines with $q_{jkt-1}$ if the function $f$ has $f'(x) x + f(x) > 0$ for all $x > 0$. For optimal profit we have the inequality $\pi_{jkt} \leq (p_t - c_t + \delta_{jt} \ln(\delta_{jt}) - \delta_{jt}) q_{jkt}$, implying that $p_t - c_t + \delta_{jt} \ln(\delta_{jt}) + \delta_{jt} \ln(q_{jkt}) \geq \delta_{jt}$ for participating firms. As a consequence, $\Delta q_{jkt} \geq 0$. That is, firms either grow or exit.

Firms which have an optimal R&D expenditure below unity increase their profit by choosing not to perform R&D at all and to produce at the unit cost of $c_t$. The decision process is adapted to include the choice not to perform R&D and to have the profit function equal to $\pi_{jkt} = (p_t - c_t) q_{jkt} - f(q_{jkt-1}) (\Delta q_{jkt})^2$. The optimal output in this case is $q_{jkt} = q_{jkt-1} + (p_t - c_t)/(2f(q_{jkt-1}))$. Firms compare the maximum profit when performing R&D with the optimal profit when not performing R&D and choose the best option.

We introduce an approximation for the optimal $q_{jkt}$ which makes it linear in price $p_t$. This approximation is not used in the estimation or simulation stages but may illustrate the way in which the industry is predicted to evolve over time.

$$(9.4) \quad q_{jkt} = q_{jkt-1} + \frac{1}{2f(q_{jkt-1})} \left( p_t - c_t + \delta_{jt} \ln(\delta_{jt}) + \delta_{jt} \ln \left( \frac{\delta_{jt}}{2f(q_{jkt-1})} \right) \right)$$

This is a first order Taylor approximation in the point $q_{jkt} = q_{jkt-1} + \delta_{jt}/(2f(q_{jkt-1}))$, a quantity participating firms will at least produce. When we assume a linear demand function, $q_t^D = A - B p_t$, then this approximation implies a simple expression for the equilibrium price given the number of firms of each type and cohort. If $n_{jkt}$ is the number of firms of type $j$ and cohort $k$ then the equilibrium price $p_t^e$ can be solved from $\sum_{j,k} n_{jkt} q_{jkt} = A - B p_t$:

$$(9.5) \quad p_t^e = \frac{A - \sum_{j,k} n_{jkt} q_{jkt-1} + \sum_{j,k} \frac{n_{jkt}}{2f(q_{jkt-1})} \left( c_t - \delta_{jt} \ln(\delta_{jt}) - \delta_{jt} \ln \left( q_{jkt-1} + \frac{\delta_{jt}}{2f(q_{jkt-1})} \right) \right)}{B + \sum_{j,k} \frac{n_{jkt}}{2f(q_{jkt-1})}}$$

144
When there is no exit, i.e. $\Delta n_{jkt} \geq 0$ for all $j$ and $k$, then equation (9.5) can be rewritten using

$$
\sum_{j,k} n_{jkt} R_{jkt-1} = A - B \rho_{t-1},
$$

(9.6) $\rho_t^* = \frac{B \rho_{t-1}}{B + \sum_{j,k} \frac{n_{jkt}}{2f(q_{jkt-1})}} + \frac{\sum_{j,k} \frac{n_{jkt}}{2f(q_{jkt-1})} \left( c_t - \delta_j \ln(\delta_j) - \frac{\delta_j \ln \left( q_{jkt-1} + \frac{\delta_j}{2f(q_{jkt-1})} \right)}{2f(q_{jkt-1})} \right)}{B + \sum_{j,k} \frac{n_{jkt}}{2f(q_{jkt-1})}}$

That is, in periods when no firms exit (usually when $\rho_t$ exceeds $c_t$) the equilibrium price is equal to a linear combination of the price in the previous period and a weighted average of the per unit production costs of the firms in the industry. As the number of firms and their output grows over time the equilibrium price will be determined more and more by the per unit production costs. The difference between $\rho_t$ and $c_t$ will become less and less until this difference becomes negative and firms which do not perform R&D start to exit. This is the ‘shakeout period’. Many small firms which produce too less to profit from R&D effort exit. The firms which exit first are the ones which have entered the latest and/or which have the lowest R&D capabilities. Firms which remain in the industry are the ones which have entered in the first few years after the foundation of the industry, hence the ‘first-mover advantage’, and/or which have high R&D capabilities.

In the next section we present some empirical results of the model for the U.S. tire industry as discussed in Chapter 8. A unique data source, collected and revised by Steven Klepper and Kenneth Simons, of the entry and exit figures per cohort is used to estimate the parameters of the first-mover advantage model (see Klepper and Simons (1993) for more information on the data sources).\(^{65}\) However, we need to make some additional assumptions first. We assume that the function $f(x)$ equals $\theta/(1+x)$. This introduces $\theta$ as a parameter to indicate the importance of adjustment costs. We assume that the index of technological opportunity $\gamma_t$ equal is to $c_t$.\(^{66}\) The number of potential entrants is taken to be $\alpha$ times the number of firms in the previous period ($N_{t-1}$) with a minimum of one. The R&D capability of each entrant is determined as follows. The value $\beta_j$ is assumed to be

\(^{65}\)Not all data were available for each cohort separately in the first period of the industry evolution. In case the exact year of entry is unknown then the figures for the period (of two or three years) of entry are divided across the years.

\(^{66}\)The R&D intensity of firms, $f_{jt}$ divided by $\rho_t q_{jt}$, is equal to $\beta_j$ times $c_t/\rho_t$ in this specific case. The R&D intensity will be lowest during the early years of the industry evolution and rise to somewhat above $\beta_j$ after the shakeout period. The average value of $\beta_j$ can therefore be seen as a reasonable proxy for the average R&D intensity of the industry at hand.
equal to \( j \) times the value \( \beta \). Hence, the lowest capability is \( \beta \) while the highest capability is \( J \beta \). The type \( j \) of the entrant is then determined from a binomial distribution with probability \( P_\beta \) and number of possibilities \( J \). The exact procedure is as follows. Assume that \( X \) has a \( Bin(J - 1, P_\beta) \)-distribution. The number of potential entrants of type \( j \) is then equal to the integer part of \( \alpha P[X + 1 = j]N_{t-1} \). In case of non-integer parts the number of potential entrants of the types with the highest non-integer parts are increased by one until the total number of potential entrants equals \( \alpha N_{t-1} \). The unknown parameters of the model are: \( A, B, c, \theta, \beta, P_\beta, J \) and \( \alpha \).

### 9.3 Some preliminary results

In this section we use a subsample of the entry and exit data of firms in the U.S. tire industry. The first year of the sample is 1905 and the last year is 1959. The year of the 'shakeout' was 1922, as can been seen in Chapter 8. After 1922 not many firms entered the industry anymore. From the two first cohorts, of 1905 and 1906, 32% survived until the year 1934. From the next eight cohorts, of 1907 through 1914, only 7% survived until the year 1934 while this figure is 9% for the last eight cohorts before the shakeout period. It appears that survival is negatively related to the date of entry, especially when one takes into consideration that the number of years between the date of entry and 1934 decreases as one goes from the early to the late entrants. Table 9.1 shows the survival percentages after five and ten years for the six groups of three successive cohorts up till the shakeout period 1922. The results show that these survival percentages tend to decline with the date of entry with the exception of the cohort 1908 through 1910. We note that the sample of firms entering in these years was quite small. The general pattern of low survival rates for late entrants is in line with the general predictions of the first-mover advantage model.

<table>
<thead>
<tr>
<th>Cohorts</th>
<th>1905/07</th>
<th>1908/10</th>
<th>1911/13</th>
<th>1914/16</th>
<th>1917/19</th>
<th>1920/22</th>
</tr>
</thead>
<tbody>
<tr>
<td>Survival after 5 yr</td>
<td>74%</td>
<td>48%</td>
<td>60%</td>
<td>71%</td>
<td>62%</td>
<td>31%</td>
</tr>
<tr>
<td>Survival after 10 yr</td>
<td>39%</td>
<td>19%</td>
<td>40%</td>
<td>36%</td>
<td>26%</td>
<td>12%</td>
</tr>
</tbody>
</table>

The data \( N_{k,t} \) of the number of firms of cohort \( k \) in period \( t \) have the following matrix form:

\[
\begin{array}{cccccc}
N_{1,1} & N_{1,2} & N_{1,3} & \cdots & \cdots & N_{1,T} \\
0 & N_{2,2} & N_{2,3} & \cdots & \cdots & N_{2,T} \\
0 & 0 & N_{3,3} & \cdots & \cdots & N_{3,T} \\
\cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\
0 & 0 & 0 & \cdots & \cdots & N_{T,T}
\end{array}
\]

This is a survival matrix and it is obvious that \( N_{k,t+1} \geq N_{k,t} \). We want to find values of the
parameters at which the predicted entry and exit rates are as 'close' as possible to those which can be derived from the real survival matrix. The number of entrants in cohort $k$ is simply $N_{k,k}$. The number of exits from cohort $k$ from period $t$ to period $t+1$ equals $N_{k,t} - N_{k,t+1}$. As criterion of 'closeness' we choose an expected squared error criterion. If $P_{k,t}$ is the predicted number of firms of cohort $k$ in period $t$ then the Neyman chi-square criterion is:

$$\chi^2 = \eta \sum_k \sum_{t>k} \frac{(N_{k,t} - N_{k,t+1} - P_{k,t+1} + P_{k,t})^2}{N_{k,t} - N_{k,t+1}} + (1 - \eta) \sum_k \frac{(N_{k,k} - P_{k,k})^2}{N_{k,k}}$$

$$0 \leq \eta \leq 1$$

The first part of the criterion compares predicted and actual exits and the second part compares predicted and actual entries. The value of $\eta$ determines the relative weight of these two parts. In the current analysis we fix it at 0.5. A problem with the chi-square criterion is that the cells of exits and entries should contain number of firms preferably larger than about 5. This implies that cohorts and periods may have to be aggregated. For our data set of the U.S. tire industry we have chosen the following aggregation. The total number of entrants over the entire time period is counted. The first cell for entry consists of the first cohorts of which the number of entrants add to at least 10% of that total. The first cell for exit consists of the number of firms of these cohorts exiting within 4 years. The second and third cell are the same for exit between 5 and 10 years and between 11 and 20 years. The fourth cell is the cell of surviving firms after 20 years. This procedure is repeated for the other cohorts. Table 9.2 shows the cells for the current data set.

**Table 9.2: Observed entry and exit patterns**

<table>
<thead>
<tr>
<th>group</th>
<th>until</th>
<th>entry</th>
<th>exit(1-4)</th>
<th>exit(5-10)</th>
<th>exit(11-20)</th>
<th>surv(20+)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1910</td>
<td>65</td>
<td>24</td>
<td>22</td>
<td>5</td>
<td>14</td>
</tr>
<tr>
<td>2</td>
<td>1914</td>
<td>93</td>
<td>31</td>
<td>32</td>
<td>21</td>
<td>9</td>
</tr>
<tr>
<td>3</td>
<td>1917</td>
<td>80</td>
<td>25</td>
<td>32</td>
<td>14</td>
<td>9</td>
</tr>
<tr>
<td>4</td>
<td>1920</td>
<td>84</td>
<td>51</td>
<td>22</td>
<td>9</td>
<td>2</td>
</tr>
<tr>
<td>5</td>
<td>1922</td>
<td>143</td>
<td>98</td>
<td>33</td>
<td>6</td>
<td>6</td>
</tr>
<tr>
<td>6</td>
<td>1925</td>
<td>68</td>
<td>37</td>
<td>24</td>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td>7</td>
<td>1952</td>
<td>61</td>
<td>28</td>
<td>18</td>
<td>9</td>
<td>6</td>
</tr>
</tbody>
</table>

Note: The small group of 9 firms which entered after 1952 are not taken into consideration. The second column contains the period up till which the cohorts are aggregated. The third column contains the total number of entrants in those aggregated cohorts. The fourth, fifth and sixth columns contain the number of firms exiting between 1 and 4, between 5 and 10 and between 11 and 20 years after entry. The last column contains the number of firms which survive after 20 years.

---

87 See Berkson (1980) for other chi-square criteria and a discussion of the principles of minimum chi-square estimation.
It is important to keep in mind that function (9.7) is integer valued while the parameters are real (except for $J$). One consequence is that different parameter vectors may give the same value of the criterion. A second consequence is that an extremely small change in the parameter vector may lead to a large change in the criterion value. The choice of this criterion is also likely to lead to local optima given the large number of elements in the parameter vector and the large size of the survival matrix.

The estimation of the values of the parameters goes as follows. First, we fix the value of $J$ equal to 25. Choosing this value rather large leaves enough room for the $\beta$ parameter to get an adequate binomial distribution. Second, we fix the constant $A$ in the demand function equal to 1000. The parameters $B$ and $\gamma$ are therefore the principal determinants of the amount of demand for the good. This leaves six parameters to be estimated. Each function evaluation requires a simulation of the entire industry evolution given these parameters.

Because of the complicated function we choose a robust algorithm introduced in the early 1980s in thermodynamics (Kirkpatrick et al. (1983)). The importance of this algorithm in difficult optimization problems in econometrics is shown by Goffe et al. (1994). We use a slightly revised version of the simulated annealing algorithm which they present in their Appendix B. The advantages of the simulated annealing method are that it can escape from local optima and that it has little demands about the function to be optimized. In fact, it needs not even be continuous. The main disadvantage are the computing costs.

The simulated annealing algorithm is a global optimization method which has the following procedure of finding an optimum. An initial (high) value for $T$ ('temperature' in statistical mechanics) and starting values for the parameters are chosen. A new parameter vector is chosen randomly in the 'neighbourhood' of the first parameter vector. In case this vector has a lower function value then the procedure accepts the new vector. In case this vector has a higher value then the procedure decides to accept the new vector anyhow if a random draw from a uniform $[0,1]$ distribution is below $e^{(F_1 - F_2)/T}$, where $F_1$ is the function value corresponding to the original parameter vector and $F_2$ is the function value corresponding to the new vector (this is the so-called Metropolis criterion). The larger $T$ and the smaller the difference between the two function values, the larger the probability that the new vector is accepted even though it has a higher function value. The procedure updates the best value of the optimum and the corresponding parameter vector each time a superior point is found. After a certain number of random draws of the parameter vector, or after a certain condition is fulfilled, the value of $T$ is reduced and the procedure starts at the best current parameter vector and goes through the same random search again. The procedure ends when lowering the value of $T$ does not seem to lead to new optima.

The probability of escaping local optima depends on the initial value of $T$, the number of random draws at each level of $T$ and on the speed at which $T$ is reduced. A high initial
value of $T$, many random draws and a slow reduction of $T$ increase the probability of finding the global optimum but also increase computing costs. We have chosen initial values of $T$ of about 500, about 200 draws at each level of $T$ and a standard reduction of 20% of $T$ after these draws. The lowest sum of squared errors found was 260.38. Values of $B$, $c_r$, $\theta$, $\beta$, $\alpha$ and $R_T$ which give an industry evolution that has a survival matrix with is 'closest' to the real survival matrix in terms of the sum of squared errors are 22.79, 37.52, 4.282, 0.0026, 0.196 and 0.096.

The estimates suggest that the R&D intensity in the U.S. tire industry has a maximum of around 6%, but that the average value is much lower: around 1%. The number of potential entrants is estimated to be around 20% of the number of incumbents in the previous period. The probability in the binomial distribution of R&D capabilities is quite low (around 0.1) and corresponds to the study on R&D intensity distributions by Cohen and Klepper (1992b). The predicted number of firms entering and exiting the industry for the period 1905 until 1952 can be found in Table 9.3.

Table 9.3: Predicted entry and exit patterns

<table>
<thead>
<tr>
<th>group</th>
<th>until</th>
<th>entry</th>
<th>exit(1-4)</th>
<th>exit(5-10)</th>
<th>exit(11-20)</th>
<th>surv(20+)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1910</td>
<td>32</td>
<td>0</td>
<td>0</td>
<td>9</td>
<td>23</td>
</tr>
<tr>
<td>2</td>
<td>1914</td>
<td>33</td>
<td>0</td>
<td>0</td>
<td>33</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>1917</td>
<td>47</td>
<td>0</td>
<td>34</td>
<td>13</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>1920</td>
<td>79</td>
<td>32</td>
<td>47</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>5</td>
<td>1922</td>
<td>82</td>
<td>82</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>6</td>
<td>1925</td>
<td>53</td>
<td>53</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>7</td>
<td>1952</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

Note: No firms are predicted to enter after 1952. See note to Table 9.2.

The predicted and observed number and cohorts of firms exiting show the same general pattern of early exit in later cohorts and late or no exit in earlier cohorts. This pattern is much more clear for the predicted numbers than for the observed number, but this should not be very surprising. In early years of the industry evolution no firm needs to exit because even without performing R&D remaining in the industry is profitable. In later years of the industry evolution no firms can profitably enter as the large incumbents are able to produce at much lower unit costs due to their large R&D expenditures.

We conclude this discussion of the first-mover advantage model by considering some of the possible extenstions to the model. A first extension may be to have the demand function change over time. In Chapter 8 we derived that demand for tires increased substantially due to the increase in the number of vehicles. A second extension may be to introduce niche markets in which small firms may survive despite increase in the overall concentration in the industry. Such an extension would be in line with the resource
partitioning theory (Swaminathan (1995)). A third extension would be to correct for the fact that many incumbents remain in the industry while exit is the most rational option. These entrepreneurs expect improved market conditions. However, in our structural model these hopes are not fulfilled. These and several other research questions on this first-mover advantage model provide an interesting agenda for research in the near future.
CHAPTER 10

EPILOGUE

The aim of this book is to increase our understanding of the processes of market dynamics and market evolution. It presents several new dynamic models which are empirically examined. Special attention is paid to the processes of entry and exit, the role of small firms in modern economies and the evolution of prices, output and number of firms. The choice of these subjects is inspired by the increased shares of small firms in many manufacturing industries and the increased degree of volatility in terms of diffusion of new technologies and shorter product life cycles in most parts of the private sector.

In this epilogue some central aspects of a general theory of the development of small business presence are presented. These aspects are directly related to the investigations of the role of small firms presented throughout this book. The aspects, which are presented in Propositions 10.1 through 10.4, concentrate on small entrepreneurial firms. That is, we do not take 'mom-and-pop' businesses reluctant to innovate into consideration. These businesses play some role in providing employment but fail to play a significant role in affecting economic progress. Furthermore, it is noted that certain activities would entail very high transaction costs when executed by different firms (Teece (1980), Williamson (1989)). Such activities which are usually characterized by some form of asset specificity are less costly executed within the same legal entity. Small firms are not likely to replace their larger counterparts in industries with high asset specificity and operations on a large scale like many industries in the chemical sector.

The first proposition is related to the first-mover advantage discussed in Chapter 9. This model provides an explanation of the development of the number of firms entering and exiting an industry, especially in industries in which R&D plays an important role. The model shows, in accordance with figures in many industries (Gort and Klepper (1982), Klepper and Graddy (1990) and Agarwal and Gort (1996)), that the share of small firms begins to decrease some years after the initial stage of the industry life cycle. After some decades only large firms remain in the industry together with a fringe of small firms in market niches. Advantages to large size in R&D within well-defined technological

88 Szymanski et al. (1995) show that there may also be a first-mover advantage in consumer goods industries.

89 Large firms may also acquire small firms which have been successful in the high-risk development stage of creating new products and production technologies (Granstrand and Sjölander (1990)). These small firms may not have the financial means to successfully exploit their innovations. See also proposition 10.3.
trajectories are an important reason of this trend towards concentration.\textsuperscript{90}

Proposition 10.1: The share of small firms on a specific market decreases over time after the initial stage of the industry life cycle in case no new technological trajectories emerge.

A second aspect is the recurrent introduction of new products and hence new markets. An important reason for this development is the increase in disposable income. This increase and the related process of individualization in Western countries leads to an increasing need for diversity, i.e. new products. Jackson (1984) shows how the number of commodities expands with total expenditure both in the aggregate and for commodity groups.\textsuperscript{91} The positive effect of market growth on (net) entry documented in Chapter 4 may partly result from increased variety in demand.\textsuperscript{92}

Proposition 10.2: Increases in wealth lead, possibly indirectly through increased individualization, to more variety in demand and, hence, to more markets.

Increasing variety of demand may affect the structure of supply, but the same holds for the continuing growth of the knowledge base.\textsuperscript{93} University research, e.g. in materials science and chemistry, is very important for innovativeness in many industries. New technological trajectories may emerge from scientific progress and small firms are likely to play an important role in these new developments. The advantages of size will re-emerge during later stages of the trajectories but during a change in the basic technological paradigm these may not be present. The third aspect of the theory is the effect of the growth in the

\textsuperscript{90}Industries in which R&D plays a less significant role may also have periods in which there is little (organizational) innovation. These industries are also likely to show an increasing share of large firms. Examples can be found in service industries.

\textsuperscript{91}Jackson shows that the average number of commodities purchased within the commodity group 'Food' is below 3 for the low income groups and about 4 for the high income groups. For 'Recreation' the differences are much larger. Low income groups purchase less than 4 different commodities while high income groups purchase more than 10 different commodities. Differences are even larger for 'Gifts'. High income groups purchase about 7 different kinds of gifts while low income groups purchase only about 2.

\textsuperscript{92}This implies that entry of small entrepreneurial firms in East-European countries may be hampered not only due to lack of 'entrepreneurial experience' and venture capital but also through low average income levels.

\textsuperscript{93}Dosi (1988) defines the knowledge base as "the set of information inputs, knowledge, and capabilities that inventors draw on when looking for innovative solutions" (p.1126).
knowledge base on the technological trajectories in an industry.

Proposition 10.3: Growth in the knowledge base causes new technological trajectories to emerge, which opens up possibilities for new, small firms.

The last aspect deals with the advantages and disadvantages of diversification. An increase in the number of markets might not lead to an increase in the number of firms in case the existing firms could diversify into the new markets and leave little chance for new entrants. However, there is a limit to diversification. These limits have been discussed in Chapter 2. Jovanovic (1993) discusses several reasons for diversification. They include gaining market power, avoiding risk, having access to funds, making products compatible, reaping efficiency gains or pursuing managerial goals. Many of the motives to diversify (or acquire or merge) do not improve the market position of the company. In periods in which the disadvantages of diversification are underestimated firms are therefore likely to grow too large.

Proposition 10.4: There are advantages and disadvantages to diversification. Their perception by managers and the motives of managers to diversify or to concentrate on ‘core business’ may change over time.

Among these four aspects the advantages in size within trajectories and increasing attention for advantages of diversification (economies of scope) lead to a decreasing share of small firms. The increasing demand for variety and the changes in the knowledge base leading to new trajectories (creative destruction) and increasing attention for disadvantages of diversification lead to an increasing share of small firms. These four propositions may explain why the century long trend to ever larger firms in manufacturing reversed during the 1970s. Chandler (1990) stresses the importance of investment in production, distribution, and management needed to exploit economies of scale and scope during the period after the second industrial revolution of the second half of the 19th century. It was a period of relatively well-defined technological trajectories, of a stable demand and of seemingly clear advantages of diversification. Probably the largest difference between the 1950s and 1960s on the one hand and the 1970s and 1980s on the other was instability and insecurity. Demand turned out to become less predictable and less uniform, new technological developments made many ‘old ways’ of production and R&D obsolete, and unexpected disadvantages of diversification brought many
conglomerates into large difficulties.94

The theory suggested above implies that periods in which the share of small firms declines are periods of stability in the technological trajectories and the variety of the demand of consumers. Periods of increasing variety of customers' tastes and of quick developments in the knowledge base leading to new technological trajectories are periods in which small firms may prosper. The 'management fashion' with regard to diversification will play some role too. In the 1990s the fashion is downsizing, rightsizing and going back to core competences. This development may go one or several steps too far, just like the conglomerate merger wave of the late 1960s went one or several steps too far in the other direction.

---

94 There were other factors, like the oil crises, flexibility in the rates of exchange and the increasing international competition, primarily from Japanese companies.
MARKTSPROCESSEN (SUMMARY IN DUTCH)

Marktdynamiek en de evolutie van bedrijfstakken zijn centrale onderwerpen in de economische wetenschap. Dit boek heeft als doel deze processen beter in kaart te brengen door nauwkeurig oorzaken en gevolgen van een veranderende bedrijfstakstructuur te onderzoeken. Daarvoor zijn nieuwe modellen ontwikkeld welke vervolgens empirisch zijn getoetst. Speciale aandacht wordt besteed aan de toe- en uittreding van bedrijven, de rol van het midden- en kleinbedrijf (MKB) in moderne economieën en de ontwikkeling van prijzen, productie, het aantal bedrijven en hun grootteklasseverdeling over een tijdsspanne van meerdere decennia. De keuze van deze onderwerpen is geïnspireerd door een tweetal ontwikkelingen die hebben plaatsgevonden sinds de jaren '60. Ten eerste is het aandeel van het MKB in veel, voornamelijk industriële, bedrijfstakken niet verder gedaald of zelfs toegenomen. Ten tweede is de dynamiek van het bedrijfsleven over het algemeen sterk toegenomen.

In Hoofdstuk 1 wordt aangegeven hoe de houding van economen en beleidsmakers ten aanzien van kleine bedrijven is veranderd. 'Ondernemerschap' is weer helemaal terug op de agenda en het flexibel inspelen op de snel wijzigende vraag en produktietechnologieën is van levensbelang in veel sectoren van de economie. In het hoofdstuk worden verder kort enige oorzaken en gevolgen van de groeiende aanwezigheid van kleine bedrijven besproken. Figuur 1.1 op pagina 11 laat de belangrijkste verbanden zien die in het boek worden onderzocht. De publicaties waarop de hoofdstukken van dit boek zijn gebaseerd staan vermeld aan het einde van Hoofdstuk 1.

In Hoofdstuk 2 worden statistische gegevens over ontwikkelingen van het aandeel van kleine en middelgrote ondernemingen besproken. Hierbij wordt gekeken naar Nederlandse en Amerikaanse gegevens en wordt een onderscheid gemaakt tussen ontwikkelingen in de dienstensector en de industrie. Een zevental redenen voor de toename van het aandeel van het MKB worden daarna uitgebreid geanalyseerd. Deze zijn Schumpeter's 'creatieve vernietiging', technologische ontwikkelingen, de wereldwijde integratie van markten, veranderingen in de omvang en de samenstelling van de beroepsbevolking, toegenomen variatie in de vraag, deregulering en privatisering en, als laatste, de terugkeer naar kernactiviteiten door veel grote concerns.

In Hoofdstuk 3 wordt er overgegaan naar een belangrijk element van marktdynamiek, namelijk de manier waarop winstgevendheid en netto toetreding op elkaar reageren. Netto toetreding is het verschil tussen het aantal toe- en uittredende bedrijven. Hoge winstgevendheid maakt toetreding aantrekkelijk en uittreding juist onaantrekkelijk. Deze positieve netto toetreding heeft dan in zijn algemeenheid weer een drukkend effect op de winstgevendheid door de toegenomen concurrentie. In een simultaan vergelijkingen model worden de wederzijdse effecten van winstgevendheid en netto toetreding op elkaar


Een theoretisch model van de lokale oligopoliestructuur in de detailhandel wordt ontwikkeld in Hoofdstuk 5. Er wordt hierbij gebruik gemaakt van het begrip 'draagcapaciteit', afkomstig uit de populatie-ecologie. Dit is het maximum aantal bedrijven dat in evenwicht in een lokale markt aanwezig kan zijn. De gedachte is dat afwijkingen van dit evenwichtsaantal leiden tot toe- dan wel uittreding. Ook dit model wordt toegepast op gegevens van Nederlandse winkeltypen. Met behulp van geschatte parameters kan afgeleid worden hoe snel de aanpassing richting het evenwicht plaatsvindt en hoe groot het effect van veranderingen in de hoeveelheid gevraagde goederen op de 'draagcapaciteit' is. Een schatting van de gemiddelde 'draagcapaciteit' in de Nederlandse detailhandel kan ook worden gemaakt en ligt zo tussen de twee en vijf winkels per regio (meestal dorp of wijk). De schatting van het werkelijk aantal winkels per regio lag eind jaren '80 nog hoger dan dit aantal.

De gevolgen van (de groei van) het werkgelegenheidsaandeel van het MKB staan centraal in Hoofdstuk 6. Er wordt een eenvoudig selectiemodel gepresenteerd dat inzicht biedt in factoren die de relatie tussen de structuur en prestaties van een bedrijfstak beïnvloeden. Het hoofdstuk bevat twee empirische analyses. De eerste gaat over het effect van het aandeel van het MKB in bedrijfstakken in Europese landen op de ontwikkelingen van de geleverde produktie. De tweede analyse gaat over het effect van
de groei van het aandeel van het MKB in Nederlandse bedrijfstakken op de arbeidsproductiviteitsontwikkeling in deze bedrijfstakken. Bij beide analyses worden er aanwijzingen gevonden dat (de groei van) het aandeel van het MKB gemiddeld genomen een positief effect heeft gehad op de prestaties van een bedrijfstak. Het komt echter ook duidelijk naar voren dat dit van bedrijfstak tot bedrijfstak sterk kan verschillen.

In Hoofdstuk 7 staat de relatie tussen de grootte (en de leeftijd) van een onderneming en haar groei centraal. De vraag is of kleine ondernemingen sneller groeien dan grote ondernemingen. Hierover is veel discussie ontstaan, vooral vanuit de gedachte dat er ‘regressie-naar-het-gemiddelde’ zou plaatsvinden. In het hoofdstuk wordt aangetoond dat het waargenomen verband tussen groei en grootte meer weerspiegelt dan dat van bedrijven die zich om hun ‘evenwichtsgrootte’ heen bewegen. Er wordt hierbij gebruik gemaakt van twee nieuw ontwikkelde toetsgrootheden op convergentie. Deze worden toegepast op gegevens van BNP per inwoner voor OESO-landen en op Amerikaanse bedrijfsgroepen. In beide gevallen worden er sterke aanwijzingen voor convergentie gevonden.

In Hoofdstuk 8 wordt de nadruk verlegd van marktdynamiek naar de evolutie van bedrijfstakken over een periode van meerdere decennia. Feitelijk is het hoofdstuk een case-study van de Amerikaanse autobandenindustrie. Er wordt een recursief model ontwikkeld met als variabelen produktie, prijsniveau en netto toetreding van producenten. Het gemiddeld prijsniveau wordt bepaald door het aantal producenten (concurrentie), de prijs van rubber op de wereldmarkt en de mate waarin in de bedrijfstak ervaring is opgedaan met de produktie van banden. Het model wordt gebruikt om de oorzaken te analyseren van de snelle terugloop van het aantal bandenproducenten in de jaren ‘20 en ‘30. Het recursief model model maakt het mogelijk voorspellingen te doen op wat langere termijn van de structuur en prestaties van een bedrijfstak.

In Hoofdstuk 9 wordt een ander theoretisch model van bedrijfsvolgorde beschreven. Dit model laat zien hoe vroeg toegelaten producenten meer geld aan R&D zullen besteden en aldus een permanente voorsprong kunnen opbouwen. De meeste van de later toegelaten ondernemingen zullen niet stand kunnen houden en vanaf een bepaald moment zullen er zelfs geen nieuwe bedrijven meer toetreden. De parameters van dit model worden geschat met behulp van toe- en uitgedeeldegegevens per cohort van Amerikaanse banden-producenten. De voorspellingen van dit structureel model zijn veelbelovend en verdere verbeteringen en toepassingen op andere bedrijfstakken zijn mogelijk.

In de Epiloog wordt kort ingegaan op een aantal belangrijke elementen die ten grondslag kunnen liggen aan een theorie over de ontwikkeling van het aandeel van het MKB. Er wordt een viertal algemene stellingen gepresenteerd. De eerste stelling geeft aan dat het aandeel van het MKB in een bepaalde markt de neiging heeft te dalen op het moment dat er geen radicale technologische vernieuwingen plaatsvinden. De tweede
REFERENCES


Bosch, L.H.M. (1990), Beloningsverhoudingen en Bedrijfgrootte, EIM (Zoetermeer).


Brock, W.A. and D.S. Evans (1986), The Economics of Small Business: Their Role and Regulation in the U.S. Economy, Holmes and Meier (New York).


Cooper, A.C., C.Y. Woo and W.C. Dunkelberg (1989), Entrepreneurship and the Initial


Gilbert, C.L. (1986), Professor Hendry's Econometric Methodology, *Oxford Bulletin of
Granstrand, O. and S. Sjölander (1990), The Acquisition of Technology and Small Firms by Large Firms, Journal of Economic Behavior and Organization 13, 367-386.


Keren, M. and D. Levhari (1983), The Internal Organization of the Firm and the Shape of


Lichtenberg, F.R. (1994), Testing the Convergence Hypothesis, Review of Economics and
Statistics 76, 576-579.


Meredith, J. (1987), The Strategic Advantages of New Manufacturing Technologies for


Kleinbedrijf, Wetenschappelijke Raad voor het Regeringsbeleid (Den Haag).


Oi, W.Y. (1983), Heterogeneous Firms and the Organization of Production, Economic Inquiry 21, 147-171.


University Press (Cambridge).


Economy 75, 123-138.
Witteloostijn, A. van (1997a), De Anorexiastrategie als Maatschappelijke Kwaal, in Redes gehouden op de 21e Dies Natalis van de Universiteit Maastricht 10 januari, Universiteit Maastricht.
Zwan, A. van der (1981), Uit de Schaduw van de Groten, ESB 66, 976-979.
The Tinbergen Institute is the Netherlands Research Institute and Graduate School for Economics and Business, which was founded in 1987 by the Faculties of Economics and Econometrics of the Erasmus University in Rotterdam, the University of Amsterdam and the Free University in Amsterdam. The Institute is named after the late Professor Jan Tinbergen, Dutch Nobel Prize laureate in economics in 1969. The Tinbergen Institute is located in Amsterdam and Rotterdam. If available, trade editions of the books which are published in the Tinbergen Institute Research Series can be ordered through Thesis Publishers, P.O. Box 14791, 1001 LG Amsterdam, the Netherlands, phone: +3120 6255429; fax: +3120 6203395. The following books recently appeared in this series:

120. T.R.P.J. KROES, Financial intermediation and monetary transmission.
121. T.H.F. CHEUK, Exotic options.
122. K. TEPLA, Accounting for the quality of work-life. Putting price tags on sources of job (dis)satisfaction.
123. R.H.M. EMMERINK, Information and pricing in road transport: theoretical and applied models.
124. T.J.S. OFFERMAN, Beliefs and decision rules in public good games - theory and experiments.
125. F. MERCURIO, Claim pricing and hedging under market imperfections.
126. A.W.M. ODÉ, Migrant workers in the Dutch labour market today.
129. G. RUSSO, Firms' recruitment behaviour: an empirical investigation of the use of search channels, the rate of arrival of applicants, and the spatial radius of search.
130. J.N. VAN OMMEREN, Commuting and relocation of jobs and residences: a search perspective.
131. F.N. GOUWELEEUW, A general approach for computing loss probabilities in finite-buffer queues.
132. M.R.H. MANDJES, Rare event analysis of communication networks.
133. E.J. VISSER, Local sources of competitiveness: spatial clustering and organisation dynamics in small-scale clothing in Lima, Peru.
134. M.H.H. HUYSMAN, Dynamics of organizational learning.
135. J.K. GOEREE, Applications of discrete choice models.
136. N.J. VAN DER LIJN, Consumer markets and welfare under central planning.
139. M.P.E. MARTENS, Interaction between financial markets.
140. L.W.N. VAN DER TORRE, Reasoning about obligations. Defeasibility in preference-based deontic logic.
143. R. SLOOF, Game-theoretic models of the political influence of interest groups.
144. J.N.M. WIT, Dynamics and information aggregation in elections and markets.
145. N.T. BUDINA, Essays on consistency of fiscal and monetary policy in Eastern Europe.
146. F. LIPPI, Central Bank independence and credibility: essays on delegation arrangements for monetary policy.
147. H. HOEK, Variable trends: a Bayesian perspective.
149. C.S.J. JIE-A-JOEN, Strategic trade policy, multinational firms, and vertically related industries.
150. R.G.J. DEN HERTOG, Empirical essays on growth, uncertainty and prediction.
151. M. GELDERMAN, Success of management support systems: a review of the literature and an empirical investigation.
152. M.A. VAN DER ENDE, On the values of jobs and specific training.
153. M. KLPWIJK, Rural industry clusters in central Java, Indonesia; an empirical assessment of their role in rural industrialization.
154. R.J.G. VENNIKER, Transition processes from constrained to Walrasian equilibrium.
156. J.F. STURM, Primal-dual interior point approach to semidefinite programming.
157. R. FREILING, Models and techniques for integrating vehicle and crew scheduling.
158. S. BORGHAERF, Is more control better? The effects of organizational control on salespersons’ job performance, job attitudes and sales organization effectiveness.
159. C. VINKENBURG, Managerial behavior and effectiveness: Determinants, measurement issues and gender differences.
160. P.A. GAUTIER, The flow approach to the labor market.
162. Y. GAO, Management support of information system project planning.
166. J.P.M. SIJM, Food security and policy interventions in sub-Saharan Africa. Lessons from the past two decades.
167. W.A. LETTERIE, Partisan politics, persuasion and information.
168. R. VAN DIJK, Corporate finance policy and equity investment: panel data analyses.