# Differentiated Customers and Firms

DISSERTATION

zur Erlangung des akademischen Grades
doctor rerum politicarum
(dr. rer. pol.)
im Fach Wirtschaftswissenschaften

eingereicht an der
Wirtschaftswissenschaftliche Fakultät
Humboldt-Universität zu Berlin

von
Dipl.-Kfm., Dipl.-Inf.Steffen Brenner
geboren am 7.8.1971 in Berlin

Präsident der Humboldt-Universität zu Berlin:
Prof. Dr. Jürgen Mlynek

Wirtschaftswissenschaftliche Fakultät:
Prof. Michael C. Burda, Ph.D.

Gutachter:

1. Prof. Dr. Joachim Schwalbach
2. Prof. Lars-Hendrik Röller, Ph.D.

eingereicht am: 3. Januar 2002
Tag der mündlichen Prüfung: 29. April 2002

## Zusammenfassung

Die Dissertation behandelt die Thematik der Unternehmensheterogenität. Die Arbeit besteht aus einem theoretischen und einem empirischen Teil. Im theoretischen Teil wird untersucht, wie sich Asymmetrien auf der Ebene der Konsumenten auf das Unternehmensverhalten auswirken. Im empirischen Teil wird die Frage behandelt, in welchem Maße die Performance von Firmen innerhalb der gleichen Industrie differiert.

Kapitel 1 führt in die Literatur zur Unternehmensheterogenität ein. Kapitel 2 gibt einen Überblick über die aktuelle Literatur zu den Modellen der Horizontalen Produktdifferenzierung. In Kapitel 3 wird das Hotelling-Duopolmodell nach d'Aspremont, Gabszewicz und Thisse (1979) in Bezug auf die Anzahl der Spieler generalisiert. Es kann gezeigt werden, dass die Anzahl der Spieler einen Einfluss auf das Niveau der Differenzierung besitzt. Das Prinzip der Maximalen Differenzierung im Duopol verliert seine Gültigkeit für den Oligopolfall. In Kapitel 4 wird das Investitionsverhalten in neue Disrtibutionstechnologien in Märkten mit Wechselkosten der Konsumenten mit Hilfe eines Realoptionsmodells analysiert. Kapitel 5 umfasst den empirischen Teil der Arbeit. Hier wird statistisch untersucht, ob innerhalb der Industrien Firmen langfristig eine ähnliche Performance aufweisen (Performancegruppen) und welchen Anteil der Varianz der Firmen durch diese gruppierten Unternehmen erklärt wird.

*Schlagwörter*: Firmenheterogenität, Horizontale Produktdifferenzierung, Realoptionen, Wechselkosten, Strategische Gruppen

## Abstract

The dissertation considers the heterogenity of firms. It consists of a theoretical and an empirical part. In the theoretical part it is examined to what extent asymmetries between customers have an impact on firm behavior. In the empirical part, the author analyzes the level of heterogeneity of firm performance within industries.

Chapter 1 introduces the reader into the topic of firm heterogeneity. Chapter 2 provides an overview over the recent literature on horizontal product differentiation. In Chapter 3 the Hotelling duopoly model à la d'Aspremont, Gabszewicz and Thisse (1979) is generalized with respect to the number of firms. It is shown that the number of firms has an impact on the level of product differentiation. The Principle of Maximum Differentiation valid for the duopoly does not hold for the oligoply case. In Chapter 4, the optimal investment in a new distribution technology in markets with consumer switching costs is investigated using a real option model. Chapter 5 corresponds to the empirical part of the dissertation. It is studied if there are firms within industries with a similar long-run performance (performance groups) and how much of the total variance of the firm profits are explained by these firm groups.

*Key words*: firm heterogeneity, horizontal product differentiation, real options, switching costs, strategic groups

# Contents

# List of Figures

# List of Tables

# Acknowledgments

The present study comprises a variety of theoretical and empirical approaches to the topic of differentiation. The chapters are related to economic theory, industrial organization and strategic management. On the one hand this reflects the diversity of my personal interests. More importantly, however, they are the result of my development during the time at the Department of Economics and Business Administration of the Humboldt-University Berlin.

Special thanks are due to my supervisor, Prof. Dr. Joachim Schwalbach. His support and advice was invaluable. I am also thankful to Prof. Lars-Hendrik Röller, Ph.D. He gave me very valuable comments as my second supervisor.

Truly, I profited from my colleagues at the Institute of Management. My ideas were shaped by discussions with Anja Schwerk, Clemens Oberhammer, Su Qi, and others. I am also thankful to Brigitte Erlinghagen, Anna-Lena Bujarek and the students working in the institute who all were helping me one way or another.

The courses, workshops and seminars of the Ph.D.-program "Applied Microeconomics" at Humboldt-University and Free University Berlin have deepened my understanding of economics and econometrics. I am especially grateful to the organizers, Prof. Dr. Elmar Wolfstetter and Prof. Dr. Ulrich Kamecke. Furthermore, I wish to thank all the professors and students giving me their time to discuss my work.

As a member of the Sonderforschungsbereich 373 "Quantification and Simulation of Economic Processes" (Project C4: Dynamics of Competition) I benefited a lot from its stimulating atmosphere. I wish to emphasize the cooperation with Dr. Bernd Droge and Prof. Dr. Olaf Bunke whose outcome is part of my dissertation.

As important step stones on my way to complete the Ph.D. I view the participation at several summer schools and conferences, among them the Summer Schools of the EEA in Toulouse 1999 and Buch/Ammersee in 2001, and the EEA Conference 2001 in Lausanne. I had the opportunity to discuss my work with H. Bester, P. Baake, J. Laffont, J. Tirole, P. Geroski, M. Ivaldi, P. Rey, S. Martin, and V. Lambson.

I am indebted to my whole family and all my friends. In particular, I am grateful to the encouragement and support of Claudia and her tender love and care.

# Chapter 1

# Introduction

Seven decades ago, Coase (1937) asked the provocative question of why firms exist. Now, with the theory of the firm as a developed field of economics it is time to ask why firms are different. It is a common observation that even firms from the same industry are different in many ways. They produce goods of different characteristics, apply different production technologies, choose a different time schedule for their actions, pursue different strategies, and so forth.[1] At first sight, this might be puzzling since one would expect that competition induces a process of selection after which only those firms sustain which sell the best products, implement the optimal technology, and so forth. From the positive perspective this brings up the question of why there is such a large firm heterogeneity, while from the normative perspective it may be asked to which extent firms should be different. Addressing these questions is a very complex task and should, thus, take into consideration a variety of approaches.

In this dissertation, we will take advantage of and contribute to developments in several fields of economics and management science in order to obtain a clearer picture of this issue. In particular, we will base our analyses on oligopoly theory, industrial organization, and strategic management. However, such a study can never be comprehensive because firm heterogeneity is a very general phenomenon. Rather, it is shown how different theoretical and empirical methods can be applied in order to explain some observations and regularities associated with firm heterogeneity.

In the following two subsections we review the economic and management literature on firm heterogeneity. In subsection 1.3 an overview over the subsequent chapters of this work are given.

[1]There is rich empirical evidence that firms differ persistently with respect to size, strategy, and performance (Hatten and Schendel 1977, Cubbin and Geroski 1987 and 1990, Mueller 1986 and 1990, Pakes 1987, Schmalensee 1987, Schwalbach, Grasshoff, and Mahmood 1989, Dhawan 2001).

## 1.1 Contributions from microeconomics and industrial organization

One possible way to structure the heterogeneity literature in economics is to look at how the models introduce firm asymmetries: either exogenously or endogenously. Largely, we will concentrate on papers which explain firm asymmetry endogenously. In the following we will distinguish between papers from microeconomics and industrial organization since these two fields contributed most of the work on this topic.[2] Let us first elaborate on how the observation of differentiated firms is reflected in economic theory.

In microeconomic theory, a large portion of research is governed by perfect competition and the paradigm of symmetric firms. Firms usually are allowed to differ only along few dimensions such as the market share or the production costs. One extreme example in which these overly restrictive assumptions regarding the firms lead to an unconvincing result is the so-called Bertrand paradox (Bertrand 1883). The Bertrand paradox says that even given very few firms competing in a market, perfect competition will prevail with zero profits and prices at marginal costs.

There are several attempts to reconcile theory and reality regarding the Bertrand paradox. One of the most famous examples introduces differentiation between firms along the spatial domain.[3] This line of research was initiated by the seminal paper of Hotelling (1929). He generalized Bertrand's model by allowing for different firms' locations in geographic space. Later, this model was more often interpreted as a model of product differentiation.[4]

Competition in Hotelling games is local, i.e. firms compete for customers with their direct neighbors along a spatial domain. The equilibrium results of Hotelling games are driven by a basic trade-off between two opposite effects of relocation: the positive demand effect which leads firms to locate close to their direct neighbors in order to attract a higher demand, and the negative price effect which results from the increased local competition. The relative strengths of those effects decide upon which equilibrium pattern is established: either the maximum, the minimum, or an in-between differentiation solution, given an equilibrium exists.

Thus, there are two basic sources of differentiation in this model. The first is that customer locations or preferences for certain product characteristics are heterogenous. This gives firms space to differentiate from each other. The second one

---

[2]Of course, the research programs of microeconomics and industrial organization are not distinct from each other but share several fields such as oligopoly theory. The following classification originates in the authors view on this subject and hence, can be seen as a bit arbitrary.

[3]Others accomplish this by introducing capacity constraints or repeated interaction between firms. A recent paper solves this problem by introducing sluggish consumers in an evolutionary model (Hehenkamp 2001).

[4]Models of spatial competition might be of some broader interest in business policy. For example, Tang and Thomas (1992: p. 325) suggest that the implications of the model "can be extended to non-spatial competition based upon reliability, availability, customer credit, etc."

relates to competitive forces such as the threat of Bertrand-like price competition and the attraction of demand. Thus, incentives and disincentives to differentiate are established.

The Hotelling model provided the basis for the so-called "address branch" of the differentiation literature. In line with the Lancasterian tradition, product differentiation can be divided into horizontal and vertical product differentiation. Hotelling models belong to the horizontal product differentiation line. Gabszewicz and Thisse (1979, 1986), Mussa and Rosen (1978), and Shaked and Sutton (1982, 1983, 1987) develop the counterpart of vertical product differentiation. Horizontal and vertical differentiation are distinguished from each other by whether or not there is a general agreement of all customers about the ranking of two products. Vertical product differentiation prevails if this condition is fulfilled. In the vertical case, firms can differentiate from each other by offering different qualities of the product.

Another kind of differentiated consumers is associated with switching costs. Switching costs occur in markets where consumers cannot change their supplier after repeatedly purchasing a product without incurring extra costs (Klemperer 1995 surveys this literature). Because firms have a (limited) monopoly power over their attached consumers, we observe in those markets prices above marginal costs, super-normal profits, and first-mover advantages used as entry deterrents.

Differentiation and firm heterogeneity has also attracted interest in the research community of industrial organization. Traditionally, research in industrial organization has focused on the characteristics of industries in order to explain the market outcome. This has become known as the structure-conduct paradigm (Bain 1956, Mason 1939). From firm differences it was largely abstracted. Also later developments such as the theory of contestable markets relate industry characteristics to firm performance (Baumol, Panzar, and Willig 1982). Demsetz (1974) was one of the first to link profitability heterogeneity to efficiency heterogeneity on the firm-level. With the emergence of game theory, the so-called new industrial organization has established taking differences of firms into account in a more structured way.

Recently, Röller and Sinclair-Desgagné (1996, 1997) consider firm heterogeneity in an extended two-stage Cournot model of two markets. Basically, they explain how firm asymmetry evolves if there is asymmetry already in the beginning. In their first paper, they introduce a parameter of "capability" into the model. This parameter is associated with the ability to produce efficiently for a certain market. In the second paper, the setup is generalized by industry asymmetry in the form of different demand elasticities. They show that both kinds of asymmetry may increase over time and thus, are able to explain increasing differences in market shares and profitability. However, in their particular model industry asymmetry is a more important determinant of heterogeneity. Mills and Smith (1996) examine a duopoly model of technology choice and identify conditions under which otherwise identical firms would choose different technologies in equilibrium. Hermalin (1994) detects that heterogeneity of firms may be due to a lack of convexity in the organizational design. Productive efficiency heterogeneity under industry dynamics was considered by Pakes and McGuire (1994) and Ericson and Pakes (1995).

Of course, many other sources of heterogeneity could be added, such as asymmetry of information, technology diffusion, and increasing returns to scale of production. A complete survey would be beyond the scope of this work.[5] Finally, it is worth mentioning that recently a complementary line of research to the previous one emerged which addresses the value of diversity (Nehring and Puppe 2001).

## 1.2 Contributions from strategic management and organizational theory

The topic of differentiation has also been discussed in strategic management. Here, attention is payed to the question of where to optimally locate a company in the strategic space in order to gain a competitive advantage. Should a company imitate strategies of competitors and hence, choose a similar position as their rivals, or should it differentiate as much as possible? To illustrate the tension of the manager when it comes to the strategy he should pursue, Ghemawat (1999) used the metaphor of a landscape. The space of the landscape represents strategic dimensions and the hills and valleys represent clusters and gaps of profitability, each one corresponding to a certain business model. Ghemawat (1999: p. 20) underlined that "[t]he central challenge of strategy is to guide a business to a relatively high point on this landscape." Differentiation from competitors would be profitable since in this case the company could capitalize on the certain business model as a monopolist. However, since every company would try to climb up on a hill, rivals may be attracted. Additionally, there may be several forces which counteract the effort of a company to change its position such as organizational inertia, uncertainty, and entry barriers. In contrast, it can sometimes be possible to reshape the business landscape to one's own benefit by approaching a competitor, for example if the competitor offers complementary products or if there are spill-overs.

Since strategic management is an interdisciplinary field several contributions mentioned in this section could also be assigned to industrial organization or even to microeconomics. However, most of the papers surveyed in this section are published in journals devoted to management issues. They have in common addressing the question of how a manager can achieve a competitive advantage over rival firms, by strategic conformity or by strategic differentiation? Pure models from economics and the corresponding equilibrium concepts are sometimes not appropriate to solve this set of problems since they are too unrealistic and require too much rationality from the agents. Because of this, strategic management relies on a variety of methods from economics, sociology, psychology, and other fields of science.

Several theories from organizational theory and strategic management relate similarity/conformity to firm performance. However, their implications are equivocal.

[5] See also the complementary survey of Röller and Sinclair-Desgagné (1996). Another approach to this topic can be found in the dissertation by Lewis (1998).

Rationales are given for both propositions: the one which favors differentiation and the one which supports the conformity hypothesis. Chen and Hambrick (1995) and also Rumelt, Schendel, and Teece (1994) emphasize the high priority which resolving this puzzle should possess in business policy research.

Strategic heterogeneity within an industry was first considered within the framework of strategic groups (for surveys see Barney and Hoskisson 1990, Thomas and Venkatraman 1988, McGee and Thomas 1986, and Chapter 5). A strategic group comprises firms pursuing a similar set of strategies. It was hypothesized that firms within certain groups are isolated from competition by mobility barriers (Caves and Porter 1977) and hence, these firms may earn super-normal profits. Later, strategic group researchers also considered cognitive concepts (Reger and Huff 1993, Porac and Baden-Fuller 1989) and the level of strategic interaction (Dranove, Peteraf, and Shanley 1998, among others) as explanations of a similar within-group performance. Another explanation why firms with a similar strategy may be more profitable refers to tacit collusion (Porter 1979). The argument for this claim is that firms' cooperative behavior can be easier maintained the more similar firms are. Further support for the conformity proposition is provided by the new institutional theories (for example, Scott 1995, DiMaggio and Powell 1983, Haveman 1993, Hambrick and d'Aveni 1992). Basically, firms should be similar according to those theories because otherwise their strategy might not be viewed as legitimate by the institutional and organizational environment.

In contrast, it was argued that high similarity between firms would lead to an increased competition and thus to lower profits (Baum and Mezias 1992, Baum and Singh 1994, among others). The theory of organizational ecology and that of organizational learning (Miles, Snow, and Sharfman 1993) emphasize the benefits within an industry of a diversity of organizations (Hannan and Freeman 1989, Hatten and Hatten 1987). Diversity maintains a high level of experimentation regarding different opportunities and market niches from which the whole industry could gain through spill-overs.

To summarize the previous paragraphs, there are many counteracting forces of strategic conformity having an effect on performance. What recommendations should companies be given regarding the extent of differentiation in face of these forces? Two recent papers published in the *Strategic Management Journal* address this issue without taking notice of each other. They also come to opposite conclusions.

The first paper from Dooley, Fowler, and Miller (1996) introduces the concept of strategic variety and examines its impact on performance. Strategic variety is an industry-level concept in contrast to strategic conformity which represents a firm-level construct. Nevertheless, both can be compared since strategic variety is just defined "as the strategic distance between firms in an industry" (Dooley, Fowler, and Miller 1996: p. 194). Hence, strategic variety aggregates at the industry-level what strategic conformity measures at the firm-level.

Dooley, Fowler, and Miller emphasize that in terms of industry profits there is a trade-off between pursuing homogeneous and heterogenous strategies. The firms in an industry may benefit from following the same strategy since then collusion can be enforced more easily. On the other side, in an industry with greater variety firms perform well because the competition they face is soft. They conclude that industries with in-between variety face a tougher competition but are unable to collude. Hence, they claim that it is the middle ground in the strategic space which should be avoided.

The other paper stems from Deephouse (1999) who builds a theory of strategic balance. To support the differentiation hypothesis he cites theories which relate similarity to increased competition. For the conformity proposition he refers to new institutional theories. Deephouse suggests that firms gain a competitive advantage if they differentiate as much as legitimately possible. In such a position, firms benefit from a relaxed competition but do not face legitimation challenges.

Obviously, the proposition of Dooley, Fowler, and Miller (1996: 293-294) that "the situation to be avoided is not heterogeneity, or homogeneity, but the middle ground between the two extremes" clashes with Deephouse's prescription that "firms seeking competitive advantage should be as different as legitimately possible" (1999, p. 148). Nevertheless, both papers include empirical evidence confirming the respective hypotheses. Partially the dissent between Deephouse and Dooley, Fowler, and Miller can be traced back to different assumptions. While Deephouse neglects the possibility of collusion, Dooley, Fowler, and Miller allow for collusion. However, they are ignorant of the pressure to be a legitimate firm. Nevertheless, some methodological problems of their approach may also have caused the opposing results.[6]

Being more specific on the levels of production and product design strategies, Porter (1985) developed the so-called generic strategies. He describes the tension of a firm to either focus on the cost-leadership or on (product) differentiation and emphasizes that firms should avoid to get "stuck in the middle". Although this concept became very influential in the field of strategic management it was not successful in describing real firm behavior (Campbell-Hunt 2000).

To conclude, there are many approaches to the issue of strategic conformity from the several fields of strategic management. However, their results are often contradictory. It is necessary to explore which are really the dominating forces of differentiation and how they interact with each other in a more structured way.

## 1.3 Overview over the chapters

After having reviewed the literature related to firm heterogeneity let us turn to the contribution of this piece of research. The study consists of a theoretical and an

[6]For example, they implicitly claim that the theory they are building applies to many kinds of strategic domains. Actually, one can expect that some strategies might not be dominated by legitimacy or collusion challenges. Thus, it is plausible to suggest that theory building should take place at the level of specific strategies instead of the most abstract level. A similar argument was made by Priem and Butler (2001) towards the resource-based view.

empirical part. The theoretical part covers the Chapters 2, 3, and 4, while the empirical part corresponds to Chapter 5. In the theoretical part, firm asymmetry is explained by asymmetric customers. Customers are assumed to differ regarding their preferences for goods of different characteristics or regarding their supplier switching costs. In the empirical part, we analyze the homogeneity of profit rates within industries.

Chapter 2 introduces the reader to the topic of horizontal product differentiation. Its main purpose is to survey the literature in order to extract the determinants of differentiation along the dimension(s) of product characteristics. Economists often speak of a general 'principle of differentiation' (e.g. Tirole 1990: p. 278), usually meaning that firms want to avoid unbridled price competition as it would follow by the Bertrand logic if firms were not differentiated. Of course, this statement is vague and does not take into account the different nuances of differentiation. The expanding literature on Hotelling models, however, allows for a kind of a meta-analysis of the model features which support and counteract the centrifugal forces of a tough price competition. We find that the relative strength of those forces depends heavily on a variety of model features, in particular on how the price competition is set up, on the demand elasticity, on the incentives to collude, on the distribution of customers, on cooperation and uncertainty, and on the number of firms and dimensions.

This survey is distinguished from other surveys (Gabszewicz and Thisse 1992, Lancaster 1990, Waterson 1989) by its focus on the determinants of horizontal product differentiation. Other issues such as the specification of the transport cost function, the relationship to other modeling approaches and welfare implications are better treated in different surveys. Furthermore, there is no other recent survey of this topic which takes into account the great share of the Hotelling literature which has only recently been published.

In Chapter 3, the case of Hotelling games with more than two competitors is treated. We consider the interval model with quadratic transport costs. Competition occurs in two stages: the location choice stage and the price setting stage. It is shown that a price equilibrium exists in every feasible subgame. For the location choices, we expect maximum differentiation since this pattern prevails in the duopoly case and in the similar multi-firms model along the circular domain. Maximum differentiation is defined as maximizing the minimum distance of each firm towards its direct neighbor. However, it can be shown that maximum differentiation is never an equilibrium if the number of firms is greater than two. Also the opposite pattern of minimum differentiation is not a stable spatial configuration. We use a numerical method to calculate subgame perfect equilibria for games with up to nine firms. They are characterized by a U-shaped price structure, inside locations of the corner firms, and an intermediate level of differentiation.

This part of the work could be viewed as a convex combination of Economides (1989) who treats the case of quadratic transport costs in the circular market, and Economides (1993) who treats the case of linear transport costs on the linear market. Interestingly, the equilibrium results are not similar to either of both cases. The former is related to maximum differentiation while the latter is related to minimum differentiation. However, for our case we find that in-between differentiation dominates.

In Chapter 4, we intend to explain different firm behavior by focusing on customers which are differentiated with respect to switching costs. This work is inspired by the emergence of new distribution technologies such as the Internet. These technologies can have two effects on switching costs. The first one is that it raises another barrier for the consumers to switch between companies using different technologies since they are required to be equipped with some complementary devices and with the knowledge to use them. However, costs of switching between firms applying the new technology may be lower compared to the case of two old technology firms since, for example, geographic space may not be relevant anymore and, thus, no great effort is necessary to buy from another company. Consequently, the question arises why and at which time firms would like to introduce the new technology. One of the main results of the previous literature on switching costs is that firms may benefit from the presence of these costs by setting prices higher than marginal costs and consequently, they earn super-normal profits. In our model, firms are able to delay the introduction of this technology because of switching costs. Realistically, we assume that demand is uncertain. We use the framework of real option theory to analyze the impact of various model parameters on the optimal time to invest.

This part of the work extends the literature on real options (Dixit and Pindyck 1994, Trigeorgis 1998) by taking into account consumer switching costs and a different competitive environment depending on the distribution technology. There are several applications of the real option method which consider the optimal time to invest in a new technology given a stochastic demand schedule (e.g., Schwartz and Zozaya-Gorostiza 2000, Kogut and Kulatilaka 1994). Our model is distinguished from this work by considering explicitly a decreasing stochastic demand schedule on the old technology market, an increasing deterministic demand schedule on the new technology market, and a tougher competitive environment at the new technology market associated with lower prices compared to the old technology market. Making these restrictive assumptions seems to be plausible in light of the Internet example as an important new distribution technology.

This work is also related to the switching costs literature (Klemperer 1995). The models in this line of research are usually two-stage or multi-period games of two players. Although our model does not allow for strategic interaction because of analytical tractability it has the merit of a continuous time model. Consequently, it is the more appropriate model for investigating on the optimal time to adopt a new technology.

Chapter 5 turns the concentration away from differentiated customers towards differentiated firms. The question of whether firms should be differentiated is related to the research on strategic groups. It is a common observation that within many industries, there are groups of firms pursuing a similar set of strategies. Many researchers were trying in a rather *ad hoc* fashion to examine the relationship between group membership and performance. Because of methodological problems, this work was subject to much criticism. In this chapter, we are not primarily interested in finding the link between group membership and performance. Instead, we investigate the question of whether groups can potentially explain a share of corporate performance. For this purpose we construct a statistical method which decomposes the variance of the firms' returns by taking into account performance groups. Performance groups comprise firms from the same industry with a similar performance over a longer period of time including strategic groups as a subset. By applying this method to a German data set we find that a considerable share of firms can be assigned to performance groups.

Thus, we complement the literature on the relative importance of firm, industry, and other effects (Schmalensee 1987, Rumelt 1991, McGahan and Porter 1997, among others) by considering the intermediate concept of performance groups. The results indicate that previous estimations of the relative importance of the firm and industry effects may be upward biased because the group-level effects were ignored. Further, this work also addresses the recent debate about whether strategic groups research is useful and how it should be carried out (Barney and Hoskisson 1990). It presents a method of selecting performance groups of which the strategic groups are a subset which is advantageous in several aspects to the one proposed by Wiggins and Ruefli (1994).

# Chapter 2

# Determinants of product differentiation

This chapter reviews the Hotelling literature of product differentiation. The purpose of this work is to examine the impact of the market structure on price competition and equilibrium differentiation. The existence of a general 'principle of differentiation' is rejected. In contrast, differentiation depends on a number of market parameters such as the costs of disutility, the demand elasticity, the number of firms, the density of consumers, and so forth. It is argued that the analysis of the predictive validity of this research stream by experimental and empirical studies is overdue.

## 2.1 Introduction

Differentiation is often viewed as a necessary condition to gain competitive advantage over rival firms. Following a different strategy is perceived as providing the opportunity to obtain unique or at least superior access to resources and customers. McMillan and McGrath (1997: p. 133) have recognized that

> "Most profitable strategies are built on differentiation: offering customers something they value that competitors don't have."

Although differentiation was not the target of economic theorizing itself,[1] researchers paid much attention on differentiation along the dimensions of product characteristics. During the past decades, product differentiation has become a well developed field of industrial economics. The amount of optimal product variety and the proximity of different brands in equilibrium are central questions addressed in this area. In order to model consumer and firm behavior, two prevalent research streams evolved: the class of spatial models in the spirit of Hotelling (1929) and Lancaster (1979), and the class of non-spatial models in the spirit of Chamberlin (1933), Spence (1976), and Dixit and Stiglitz (1977). In particular, the Hotelling branch of this literature has attracted considerable interest in the previous years. Consequently, this chapter surveys this part of the literature on product differentiation.[2]

What is the aim of this survey? Hotelling models were criticized for delivering rather unrobust results. Hotelling's main proposition is that firms agglomerate in the market center. A seemingly slight modification of the model led to the opposite result of extreme differentiation (D'Aspremont, Gabszewicz, and Thisse 1979).[3] Both findings were exaggeratedly called 'principles of differentiation'. Obviously, such a universal principle cannot be derived from the literature. Also in reality, we find both: agglomeration and differentiation. The former occurs, for example, in the form of shopping malls, while the latter can be associated with a great variety of differently designed consumer goods. Hence, differentiation seems to be related to different parameters of the model in a complex way. We aim at showing how the amount of differentiation is determined by some parameters of the market structure.

Competition in Hotelling games is local, i.e. firms compete for customers with their direct neighbors along a spatial domain. The equilibrium results of Hotelling

[1] One exception from the business literature is Deephouse (1999).

[2] We concentrate on horizontal product differentiation which implies that there is no general agreement between consumers how to rank the products according to their preferences. Vertical product differentiation, however, is distinct from it because such a unanimous agreement about a ranking exists (Gabszewicz and Thisse 1979). For an analysis of how vertical differentiation is related to horizontal differentiation see Anglin (1992).

[3] Although the result of Hotelling has shown to be wrong (d'Aspremont, Gabszwewicz, and Thisse 1979), there were several successful attempts to restore the minimum differentiation equilibrium (e.g. Stahl 1982) .

games are driven by a basic trade-off between two opposite effects of relocation: the positive (short-run) demand effect which leads firms to locate close to their direct neighbors in order to attract a high demand, and the negative (long-run) price effect which results from the increased local competition. The relative strength of those effects decides upon which equilibrium pattern is established: either the maximum, the minimum, or an in-between differentiation solution (given an equilibrium solution exists). Minimum differentiation subsumes every spatial configuration in which firms choose the same location in equilibrium. In the maximum differentiation solution, firms maximize the respective distances to their neighbors. The in-between equilibria consists of all the remaining formations. The model of spatial competition mutates into a model of product differentiation just by reinterpreting the domain of the model and the underlying utility function.[4]

Hotelling's model was reviewed and modified many times such that now a huge body of literature exists from which we can draw. Given the purpose of this work and in order to keep the number of articles manageable, we restrict the paper selection using the following criteria. First, only those papers are considered which focus strongly on the original Hotelling model, i.e. which change only few features of the model setup. Otherwise, it might be hardly possible to disentangle the various effects which lead to different equilibrium solutions. In particular, we almost completely neglect articles which do not model both the location and price setting stage. Furthermore, we concentrate on the sequential game of location and price choice, not allowing for any kind of Stackelberg games (Anderson 1987, Harter 1996). Second, we concentrate on the issue of product differentiation. Papers which are of more interest in fields like public finance (e.g. Hohaus, Konrad, and Thum 1994) or political theory (e.g. Weber 1998) are ignored in our review. The dominant equilibrium concept is the subgame perfect Nash equilibrium in pure strategies. However, keep in mind that pure strategy equilibria sometimes cease to exist whereas mixed strategy equilibria do not (Osborn and Pitchik 1987).

This survey may be of some general value in industrial economics and related fields. To the best knowledge of the author there is no other recent survey of this topic (Gabszewicz and Thisse 1992, Lancaster 1990, Waterson 1989). About one half of the cited papers were published after 1990. This survey is distinguished from other surveys by its focus on the determinants of horizontal product differentiation. Other issues such as the specification of the transport cost function, the relationship to other modeling approaches and welfare implications are better treated in different surveys. Further, it seems that the subtleties of the model are not always fully acknowledged because often it is concentrated on the focal results of maximum or minimum differentiation. As we will see they only represent equilibrium solutions for a subset of all the considered models. Finally, models of spatial competition may be of some broader interest in business policy. For example, Tang and Thomas (1992: p. 325) suggest that the implications of the model "can be extended to non-spatial

[4]Lancaster (1966, 1971) provides a full formal structure for this approach.

competition based upon reliability, availability, customer credit, etc."

Reviewing the literature shows that the relative strength of the centripetal and the centrifugal forces in the model depends on a variety of features of the considered market, in particular on how the price competition is set up, on the demand elasticity, on the incentives to collude, on the distribution of customers, on uncertainty, and on the number of firms and dimensions.

The organization of the chapter is as follows. In the next section the reference model will be presented. In Section 2.3 the literature survey provides the reader with the determinants of differentiation. Finally, Section 2.4 concludes the chapter.

## 2.2 The model

In this section the basic model is described. The games considered in the following survey can be derived from it by generalizing or restricting the model parameters appropriately. Since the model is a formal, mathematical one it is open to different interpretations. Most common are interpretations which go along spatial and product differentiation. As intimated already, there might exist other strategic domains for which this model can reasonably be applied. However, in order to avoid confusion we stick to the conventional interpretation of product differentiation.

There are $n$ firms producing a good at zero marginal cost. The products are horizontally differentiated. We consider a domain of product characteristics which is projected onto the unit interval, i.e. each value from the $[0, 1]$-interval represents an amount of the product characteristic. In some cases we allow for a unit circumference representation. There is an infinite number of customers $i$ whose preferences for the product characteristics $w_i$ are distributed along the unit interval with distribution $F(w_i)$.

The model is a two-stage game of complete information. In the first stage, firms $i = 1, .., n$ choose product characteristics $\bar{x} = (x_1, x_2, .., x_n) \in [0, 1]^n$ simultaneously before posting prices $\bar{p} = (p_1, p_2, .., p_n) \in R^n$ in the second stage of the game. This reflects the idea that producing a product requires a long-term commitment while setting the price can be done instantaneously. Without loss of generality, we assume $0 \leq x_1 \leq ... \leq x_n \leq 1$. Finally, customers decide if and which product to buy. This decision is determined by the preference $w_i$ of the customer, the available products, and their respective prices according to the following indirect utility function:

$$u(k_j, w_j, p_i, x_i) = \left\{ \begin{array}{l|l} 0 & q_{ij} = 0 \\ k_j - p_i - d(x_i - w_j) & q_{ij} = 1 \end{array} \right.$$

where $q_{ij}$ is the quantity of product $i$ consumer $j$ buys. In general, $q_{ij} \in [0, 1]$, i.e. consumer $i$ buys no more than one unit of the product $j$. Further, consumers only buy a product from one firm. $p_i$ is the price charged for the product of firm $i$. $d$ represents any distance function which is monotonously increasing. This function measures the disutility which is related to the difference between the amount of the characteristics of the most preferred product $w_j$ and the considered product $x_j$. These "distaste" costs are referred to as transport costs in this chapter. An outside option is accounted for by reservation price $k_j > 0$.[5]

The model exhibits a coordination problem. Assuming that equilibria exist in which firms do not choose the same location it remains unclear which firm should locate at which of the different positions. Of course, one solution would be to consider mixed strategies, i.e. firms would be allowed to follow different strategies in equilibrium, each with a certain probability. Since this requires high analytical capabilities the equilibrium concept might be implausible for real situations. To circumvent this problem, we assume that the number assigned to a firm corresponds to its rank in the spatial ordering, i.e. firm 1 is at the extreme left and firm $n$ is at the extreme right.

We focus on Nash equilibria as the only equilibrium concept. Other equilibrium concepts such as adaptation, evolution, etc were suggested to be more relevant in reality since they require less capabilities (Camerer 1991). However, they are widely ignored in the literature being surveyed and extending the analysis would be beyond the scope of this work.

### 2.2.1 Determinants of differentiation

As we will see below, the considered models allow for a great variety of equilibrium outcomes in prices and locations which range from minimum to maximum differentiation and from soft to tough price competition. For a subset of model parameters, the games are plagued with the non-existence of equilibria. In these situations, we cannot predict prices or locations. Consequently, we do not suggest the existence of very general relationships of differentiation. Rather, we show how differentiation is affected by a variety of parameters of the game. This overview is, of course, restricted by the availability of results on the various modifications of the model. By far, recent research on product differentiation does not provide results on every possible case of the model. Among other reasons, the complexity which the model poses on the equilibrium concept often increases dramatically with incremental generalizations. However, we try to identify the drivers and the inhibitors of differentiation for an important subset of the class of all models. We start with the determinants of price competition.

### 2.2.2 Price competition

The competition in prices is affected by two model features: the transportation costs and the general pricing strategy of the firms. The former directly influences

[5] The problem of non-existence of the price equilibrium occurs if it is assumed that income is sufficiently heterogenous such that some customers cannot afford to buy a differentiated good (Peitz 1999). Throughout the paper, we neglect this case.

the "toughness" of price competition. The latter limits the ability of firms to spatially discriminate customers. Let us first concentrate on the impact transportation costs have on price competition. They play a central role in research on product differentiation because the way in which they are modeled decides on how much firms differentiate their products in equilibrium. Furthermore, existence of price equilibrium and thus the existence of subgame perfect equilibrium for the whole game depend on transport costs.

Transport costs influence the price competition between two firms by the amount of customers which a firms is able to withdraw from (lose to) its neighbor by decreasing (increasing) its price by one unit. We refer to this as the *degree of price competition*. More formally, this measure is represented as the first price derivative of the demand function. Note that it may depend on the locations of the firms. Table 1 lists the papers which are related to modifications of the transport costs. The first five articles are devoted to the interval model and the last two papers consider the circumference domain. The models assume a duopoly with uniformly distributed customers and inelastic demand. The third row of Table 1 shows the functional form of the transport cost, while the fourth row indicates if a price equilibrium exists for this class of cost functions. In the class of interval models, this is only the case for quadratic and a subset of convex non-linear functions considered by Economides (1986). However, convexity is not sufficient to ensure the existence of the price equilibrium (see fifth and sixth row). Rather, it is the non-quasi-concavity of the profit function which prevents the equilibrium existence.

Further, one could suspect that the degree of price competition is related to the existence of a price equilibrium. Both cases which provide the existence of an equilibrium are associated with an infinitely high degree of price competition if firms are very close (see last column of Table 1). In contrast, in the setup examined by Gabszewicz and Thisse (1986) and Anderson (1988) price competition has an upper bound even if firms locate at the same position. However, as proved by Anderson (1988) configurations in which firms are located very close and in which firms are located very distant imply an equilibrium. For a region in-between, the equilibrium breaks down. Hence, there is no clear relationship between the degree of price competition and the existence of a price equilibrium for the interval model.

Unfortunately, previous analysis is restricted only to rather simple functional forms. Although it would be nice to have some theorems on the conditions of non-existence of the price equilibrium, to establish them "seems to be a hopeless task." (Gabszewicz and Thisse 1986: p. 167). Thus, for a great variety of transport functions no price equilibrium and hence, no subgame perfect equilibrium exist. Furthermore, the degree of price competition is not linked to the existence of price equilibria via a simple relationship.

Interestingly, in the circumference model the price equilibrium does not exist in the case of linear transport costs (Kats 1995) but if costs are linear quadratic (de Frutos, Hamoudi, and Jarque 1999). Hence, convex and concave functions may allow price equilibria. De Frutos, Hamoudi, and Jarque (1999) underlined that if

| Article | domain | cost function | price equilib. exists? | equilib. locations | level of price competition[6] |
|---|---|---|---|---|---|
| Hotelling (1929) | interval | $d(x) = ax$<br>$a > 0$ | no | - | $\frac{1}{2a}$ |
| d'Aspremont, Gabszewicz, Thisse (1979) | interval | $d(x) = bx^2$,<br>$b > 0$ | yes | max. diff. | $\frac{1}{2bx}$ |
| Economides (1986) | interval | $d(x) = bx^\alpha$,<br>$b > 0$,<br>$1 \leq \alpha \leq 2$ | if $1.26 < \alpha \leq \frac{5}{3}$<br>if $\frac{5}{3} < \alpha \leq 2$<br>not otherwise | interior<br>max. diff.<br>- | $\frac{1}{\alpha b\left(y_2^{\alpha-1}+y_1^{\alpha-1}\right)}$, where $y_i = \lvert D_1(p) - x_i \rvert$ |
| Gabszewicz, Thisse (1986) | interval | $d(x) = ax + bx^2$, | no | - | $\frac{1}{2a+2bx}$ |
| Anderson (1988) | | | | | |
| Kats (1995) | circular | $dx = ax$,<br>$a > 0$ | no | - | $\frac{1}{a}$ |
| de Frutos, Hamoudi, Jarque (1999) | circular | $d(x) = ax + bx^2, a \geq 0$,<br>$b \in R$ | if $a = 0$ or<br>if $a = -b$<br>not otherwise | max. diff.<br>max. diff.<br>- | $\frac{1}{a+bx}$ |

Table 2.1: Papers on horizontal product differentiation and transport costs

linear costs are assumed firms perceive prices and locations as strategic substitutes whereas otherwise, they are perceived as strategic complements. In the linear case, this leads to situations in which it becomes profitable to undercut a rival's price if firms' locations are sufficiently distant. However, in the non-linear case prices increase with distance such that it would not be profitable to deviate form this price equilibrium even if firms are distant.

Concentrating on the cases where equilibrium existence is of no concern which are all covered by Economides (1986) we see that the degree of price competition and the pattern of differentiation are related in the way intuitively expected. Hence, given a price equilibrium exists for the duopoly Hotelling model with uniformly distributed consumer preferences on the unit interval, then the higher the degree of price competition the stronger is strategic differentiation.

However, it is worth mentioning that differentiation does not range from minimum to maximum differentiation but rather from in-between to maximum differentiation. From this result, minimum differentiation does not seem to be a likely outcome other than a random result in the case without equilibrium existence. We will see below that minimum differentiation can be obtained by relaxing different model parameters. For the circular model de Frutos, Hamoudi, and Jarque (1999) show that the principle of maximum differentiation is dominating because firms

[6] $D_i(p)$ is the demand of firm $i$ given prices $p = (p_1, p_2)$.

always would like to differentiate unless linear transport costs are assumed.

In the previously mentioned papers, transport costs were taken to be exogenous. However, several examples like PCs designed to fulfill a wide range of functions or cosmetic articles ("all in one") suggest that firms try to reduce the disutility of consumers incurred by the distance of their preferred good and the offered good measured in the space of product characteristics. Von Ungern-Sternberg (1988) and Hendel and de Figueiredo (1997) tackle this problem for the circular domain. In their model, firms first choose to enter and then a location. Subsequently, they set transport costs and prices simultaneously (von Ungern-Sternberg 1988) or sequentially (Hendel and de Figueiredo 1997). If the transport costs are close to zero, the 'general purposeness' of the product is said to be high. Otherwise, the product design is said to be focused. Transport costs are assumed to be linear with a variable slope parameter.

In the simultaneous choice model, firms set transport costs at a rather low level (lower than socially optimal) which increases price competition. In the sequential setup of Hendel and de Figueiredo (1997), however, the choice of the focus introduces a strategic effect on pricing. The impact strategic focussing has on price competition is not straightforward. In the duopoly case where changing the focus is costless, firms have an incentive to increase their focus in order to relax price competition. For the case $n > 2$ this does not hold anymore. Here firms are interacting with two different neighbors which results in external effects of the attempt to soften price competition. Consequently, firms increase the general purposeness of their products leading to prices at marginal costs. Given that entry only occurs at fixed cost there is no equilibrium with more than two firms. Hence, despite of free entry, in equilibrium there are only two firms in the market. However, if changing the focus is associated with costs the results change for $n > 2$. In this case, the focus will be increased by the firms leading to positive prices and hence, possible oligopoly equilibria. Thus, the level of transport cost and consequently, the level of price competition depend on the costs of altering the transport costs and on the possibility to act strategically on them.

A final note should be made regarding the plausibility of different functional forms of transport costs. Lancaster (1979, Chapter 2) argues that linear transport costs are inconsistent with plausibly shaped indifference curves. In contrast, quadratic costs are not inconsistent. Thus, a justification for the frequent assumption of this particular functional form is provided.

### 2.2.3 Pricing strategies

The lion's share of papers on Hotelling models assumes mill pricing, i.e. firms sell the products at a fob price (*free on board*) which is not related to the location of the consumers. Another possibility would be to charge spatially discriminatory prices. Here, for example, firms charge lower prices from customers which are more distant. Applied to the domain of product differentiation there exist two possible interpretations. The obvious one is that firms compensate the disutility customers incur from buying a product whose characteristics are not aligned to their preferences by adjusting the price accordingly. That is, firms would charge different prices depending on the location of customers in the characteristics space. Another interpretation is that firms redesign the goods to be sold (which is assumed to be done at zero cost), pay the cost of transport and charge a uniform total price. Both price strategies can be observed in reality: discriminatory and non-discriminatory price-setting. An industry where the two forms occur simultaneously is the book retail industry. Books may be either bought in the book shop or by one of the numerous mail-order or internet book retailers at a similar compound price (including transportation).

The impact of different price regimes on spatial competition was considered by Hoover (1937), Eaton and Wooders (1985), Thisse and Vives (1988), Norman and Thisse (1996), Tabuchi (1999), and Zhang (1995). From the competition point of view, the impact of the choice between mill pricing and discriminatory pricing is not straightforward. First, the opportunity to discriminate on prices increases the firm's flexibility. A lot of articles show the profit enhancing effect of monopolistic price discrimination (Tirole 1990). However, in the short-run firms would prefer mill pricing. This is because mill pricing offers the opportunity to commit to a set of prices while in discriminatory pricing it is possible to cut prices at each location separately. Gabszewicz and Thisse (1992) point out that under discriminatory pricing, there is a separate Bertrand price game at each location of the interval. Hence, price competition is tougher under discriminatory pricing resulting in lower equilibrium prices (Thisse and Vives 1988). Moreover, it was shown for a general set of conditions that if prices and price regimes are chosen simultaneously, discrimination occurs as an equilibrium. Additionally, under more rigid assumptions the sequential subgame of first choosing the price regime and then posting a price simultaneously delivers the same result (Thisse and Vives 1988).

The case in which marginal costs of production differ between discriminatory and mill pricing was considered by Tabuchi (1999). Here mill pricing results as an equilibrium only if marginal costs of discriminatory pricing are much larger than marginal costs of mill pricing. However, taking into account the long-run effect of discriminatory pricing on entry it appears that less firms enter the market. The higher concentration of firms is accompanied by higher profits (Norman and Thisse 1996). Consequently, discriminatory pricing can be applied as an effective deterrent of entry. However, if entry is blockaded exogenously, incumbent firms would generate greater profits using mill pricing.

Zhang (1995) examines the effect of price-matching policies on the Hotelling game. If a firm follows a price-matching policy it is committed to match or beat the prices of its competitors. The game has three stages: (1) location choice, (2) decision about the price-matching policy, and (3) setting the price. Here again, an instrument which seems very competitive leads to a softer price competition via the strategic effect of entry deterrence. Consequently, although assuming quadratic transport costs minimum differentiation prevails as an equilibrium.

### 2.2.4 Elasticity of demand

The standard Hotelling model makes the rather unrealistic assumption that demand is inelastic. Relaxing this assumption may have two effects. First, equilibrium prices can be expected to decrease if demand elasticity is sufficiently high because then it is worthwhile for the firms to attract new demand by decreasing their prices. Second, price competition between the firms may be cooled because the firms' focus will become increasingly local. Since transport costs are payed by the customers, the real prices are lower for customers close to the firm which are easily attracted even if demand elasticity is high. Consequently, this would lower the incentives to differentiate to the maximum.

Smithies (1941), Salop (1979), Böckem (1994, 1996), and Hinloopen and Marrewijk (1999) contribute to model generalizations in this respect. Smithies (1941: p. 432) investigated a setting in which the consumers are individually endowed with a linear inverse demand function. He argued informally, not being provided with modern game theoretic tools, that firms would locate "closer to the center than to the quartiles." Customers in the hinterland of each firm are not necessarily attracted by the respective firms if demand is elastic. Salop (1979) concentrates on the price stage given symmetric locations of firms on the circumference and shows existence of the price equilibrium if there is an outside good. Böckem (1994, 1996) models a continuum of consumers whose reservation price $k \in [0, 1]$ is distributed uniformly over the interval. Assuming a duopoly with quadratic transport costs it is shown that a price equilibrium exists and further, that firms locate between maximum and minimum differentiation. The numeric solution of equilibrium locations is [0.272,0.728]. Indeed, it is the effect that it pays for the firms to be close to the demand which attracts them towards the market center.

The study of Hinloopen and Marrewijk (1999) examines a similar setup where transport costs are linear and the reservation price is constant across consumers. Given the reservation price is sufficiently high, the original Hotelling result holds in which no price equilibrium exists. If the reservation price is low, firms become local monopolists which leads to a continuum of equilibrium locations including maximum and intermediate differentiation. However, reservation prices in-between imply symmetric equilibrium locations where the distance between the firms is between one fourth and one half of the market. For a range of reservation prices there exists a negative relationship between this value and the amount of differentiation. Thus, summarizing we conclude that given a price equilibrium exists for the duopoly Hotelling model with uniformly distributed consumer preferences on the unit interval then the higher the elasticity of demand the less firms will differentiate.

### 2.2.5 Collusion

The games usually considered in spatial competition assume non-cooperative behavior of firms. However, through frequent interaction firms might partially build up a cooperative attitude towards their competitors resulting in collusive agreements. Firms may collude on both, locations and prices. In the duopoly examined by d'Aspremont, Gabszewicz, and Thisse (1979) this would lead to profit maximizing (and socially optimal) locations at the quartiles. Jehiel (1992) and Friedman and Thisse (1993), however, argue that firms will likely collude on prices only since locations are fixed once for all. Following the collusive agreement implies bearing the risk of being cheated. If a rival deviates from the collusive outcome, revenge could be taken only via prices which may not be credible.

For the case that firms are allowed to collude only in prices, Jehiel (1992) and Friedman and Thisse (1993) consider a game of location choice in the first stage followed by an infinitely repeated price game. While Jehiel applies Nash bargaining within the cartel, Friedman and Thisse use a profit-sharing rule corresponding to the respective ratio of profits without collusion to determine the collusive outcome. Without allowing for money transfers between the two competitors in both models minimum differentiation results, i.e. both firms locate at the market center. This result may be expected since through collusion the threat of fierce price competition preventing the firms to approach each other disappears. Moreover, taking the same position has the advantage that the threat of cutting the price has the greatest effect on the rivals' profits and thus, making it more probable for collusion to sustain. However, Jehiel introduces *ex post* money transfers between the firms which leads to cases where firms' distance is not greater than one half. In this case, it becomes inessential to be well located in order to punish the rivals for deviating. Hence, firms are more attracted by the profit maximizing locations at the quartiles.[7]

Cartel stability of spatial models of product differentiation was investigated by Ross (1992) and Rothschild (1997). Locations are assumed to be equidistant on the linear market without boundaries. The distance between two direct neighbors is interpreted as a measure of product differentiation. In the case of inelastic demand the result is clear-cut: Here the cartel stability is monotonously increasing in the measure of product differentiation (Ross 1992). Introducing demand elasticity, however, complicates the result. In this case, the impact of the amount of differentiation could have any direction, depending on the other model parameters (Rothschild 1997). The impact of collusion in the location choice stage (holding prices fixed) was examined theoretically by Huck, Knoblauch, and Müller (2000). Interestingly, it was shown that collusion is stable only if the number of colluding firms is large.

### 2.2.6 The number of firms

Let us now address the question of how a firm's choices of location and price is affected by the number of firms. In general, the analytical tractability of Hotelling models decreases with the number of firms. This is because a firm's actions depend on actions of every other firm. Further, in the interval model asymmetry arises

[7]This holds when a sufficiently low finite reservation price is assumed.

because it matters how many neighbors a firm has on each side. Hence, there is only a limited number of contributions to the multiple firms model.

Examining the model on the circular domain Salop (1979) proves the existence of the price stage equilibrium given an equidistant location setting. However, an equilibrium does not exist for every feasible subgame. Again, as in the duopoly this problem can be circumvented by assuming quadratic transport cost. This assures the existence of a price equilibrium in every subgame and further, the equilibrium existence of the whole game (Economides 1989). The locations in this equilibrium can be interpreted as following the principle of maximum differentiation. That is, firms locate equidistantly in order to maximize the minimum distance to their direct neighbors. Hence, in the circular domain the number of firms does not have a significant effect on the relative amount of differentiation.

In the linear market, the failing of a price equilibrium of Hotelling's original model could be remedied by allowing more than two firms to enter (Economides 1993). However, as there are strong incentives for the firms to agglomerate at the market center, a location equilibrium could not be established. Agglomeration is not stable since it is associated with Bertrand competition and deviating increases profits. In (non-equilibrium) equidistant configurations the corner firms are provided with some kind of market power which enables them to charge higher prices than their competitors. This is attributable to the fact that corner firms only have rivals on one side and hence, price competition is lower than for inside firms. The price structure is U-shaped, i.e. prices decrease towards the center firms.

A similar setup with quadratic transport costs was analyzed by Brenner (2001 and Chapter 3). He showed that a price equilibrium exists for every feasible subgame and further, that the principle of maximum differentiation does not carry over to the multiple firms case. For games with up to nine players equilibria are calculated which are characterized by a U-shaped price structure. Moreover, corner firms locate inside the interval which manifests their market power. However, if more than three firms are in the market the equilibrium differentiation pattern does not much deviate from the socially optimal pattern.

In summary, allowing more than two firms to compete has the following effects. The price competition is not endangered although it seems to be softened. This effect arises because there exists an externality of a price change. By responding to an action of one neighbor a firm also has to take into account the response of the second neighbor. Consequently, the maximum differentiation equilibrium in the linear market with quadratic transport costs is destroyed. Furthermore, given a market boundary the corner firms enjoy a market power from possessing a hinterland which is reflected in higher prices and higher profits than their competitors if $n > 3$.

### 2.2.7 Customer distributions

In the standard model customers are distributed uniformly over the unit interval or the circumference. This was sometimes viewed as one of the most unrealistic assumptions of the model. In the domain of product characteristics one usually finds customers' preferences clustered around some brands while in geographic space there is often an agglomeration of inhabitants at the business center of a town. Intuitively, we might expect that differentiation decreases as the density at the interval center increases while price competition gets fiercer. In the extreme case where all the density is concentrated in one point the Bertrand paradigm results.

Several attempts have been made to relax the uniform distribution assumption. Shilony (1981) has proven that the problem of equilibrium non-existence in Hotelling's original model could not be solved by allowing more general density functions. Interestingly, however, the locations which provide price equilibria still imply a tendency to agglomerate. Assuming quadratic transport costs, Neven (1986) has shown that for concave symmetric distributions a price equilibrium exists and further, that the whole game has an equilibrium. For distributions which are not too concave maximum differentiation holds. From a certain degree of concavity onward, however, firms choose inside locations which are not more than one eighth away from the interval boundaries.

In their frequently cited paper, Caplin and Nalebuff (1991) have shown that price equilibria exist for the broader class of log-concave distributions.[8] This has led to the question of whether equilibrium existence holds for the whole game. Tabuchi and Thisse (1995) found that symmetric locational equilibria do not exist for the subclass of triangular density functions. It was suspected that the non-differentiability of the density at the median point would be the reason for this. However, Baake and Oechssler (1997) have proven that it is not the non-differentiability but rather the steepness of the density at the center which prevents the equilibrium existence.

The most general approach to this problem so far was presented by Anderson, Goeree, and Ramer (1997). Considering a setup of quadratic transport costs and unrestricted locations on the real line they have shown for the class of log-concave distributions that symmetric location equilibria exist only if the distribution of consumers is not too concave at the center. Otherwise, asymmetric equilibria may appear given the distribution is not too asymmetric. By positioning asymmetrically firms shift the marginal consumer away from the very competitive region of high density to a more remote area in order to relax the competition.

Some marketing scholars may argue that the distribution of consumers is not completely exogenous. Rather, it can be influenced by firms via advertising. How this affects the price competition was examined by Bloch and Manceau (1999). In their model, firms are allowed to spend money on advertising which leads to a shift of consumers' preferences towards the advertised product. Locations are fixed at the interval boundaries. The two main findings of their work are that first, the price for the advertised product increases and second, there are incentives to equalize the distribution. The latter means that only that firm engages in advertising which is not favored by the initial distribution. Further, it pays to advertise until the uniform

[8]These are distributions which are concave after log-transformation.

distribution is reached. The uniform distribution is preferred by the firms since it minimizes the price competition for the considered set of distributions.

### 2.2.8 Uncertainty

Uncertainty by the firms about an auxiliary dimension of product characteristics was introduced by Rhee, de Palma, Fornell, and Thisse (1992). In their model of linear transport costs, a variety of possible outcomes exist. If consumers exhibit sufficient heterogeneity along the unobservable attribute, minimum differentiation results. In contrast, the level of differentiation increases along the observable attribute as the uncertainty about the other dimension decreases. Moreover, increased uncertainty makes the competition more and more irrelevant leading to increasing prices.

### 2.2.9 Multiple dimensions

Finally, it is considered if adding further dimensions of differentiation influences the outcome of the game. Neven and Thisse (1990) and Tabuchi (1994) challenged the generality of the principle of maximum differentiation which holds for the analogous one-dimensional game by adding a second dimension. In the case of Neven and Thisse (1990) this is a dimension of vertical product differentiation while Tabuchi (1994) introduces another horizontal product differentiation dimension. Interestingly, their equilibrium solutions are rather similar. In the first case of vertical product differentiation, maximum differentiation prevails at one dimension while agglomeration is observed at the other. Along which dimension distance is maximized depends on the relative length of those domains. For the case of a two-dimensional plane of horizontal product differentiation (a square) minimum differentiation results along one and maximum differentiation along the other dimension.

These findings give rise to the question of whether introducing further dimensions along which firms are able to differentiate from rivals leads to more or less differentiation. Irmen and Thisse (1998) show that along all but one dimension firms agglomerate. Dimensions are weighted differently according to their relative importance. The dimension which is weighted maximally is used to differentiate. Hence, in this case the principle of minimum differentiation possesses more plausibility than its counterpart. Even for the three-firms case it could be shown that maximum differentiation is never an equilibrium. The main reason for this result is that differentiation along one dimension suffices to relax price competition.

## 2.3 Conclusions

In a preceding survey, Waterson (1989: p. 24) suggested that "with product differentiation, anything can happen". Further twelve years of research confirmed his view regarding the non-existence of something like a general 'principle of product differentiation'. This survey reveals that differentiation - either in geographic or product space - depends delicately on parameters of the market structure. With respect to the degree of price competition, the existing literature suggests that it becomes tougher

- as the transport costs become more convex,
- as discriminatory pricing or price-matching is available as a strategy,
- as the demand elasticity becomes lower,
- the fewer firms are competing,
- the less concentrated the density of consumers is in the center of the market,
- as firm's uncertainty increases regarding the heterogeneity of consumers' preferences of a second dimension of product characteristics,
- if advertising tools are available which shift consumers towards the advertised product.

Furthermore, price competition vanishes, of course, if firms are able to collude in prices. If the number of dimensions increases, price competition seems to become more and more irrelevant since differentiation occurs only along one single dimension. Endogenous transport costs, however, are inconclusive with respect to price competition.

The degree of price competition affects the location choices via the strategic effect. If price competition becomes tougher, differentiation is fostered, and vice versa. Furthermore, there are effects of the model parameters which directly influence profits of the location choice. In particular, differentiation is increased if the density of customers becomes more concentrated. Furthermore, there is a tendency to agglomerate when collusion is permitted in order to punish a deviating rival maximally.

One of the points of concern about the Hotelling literature is the strong imbalance between the number of theoretical papers on the one side and the number of empirical and experimental papers on the other. Notable exceptions are the empirical studies of Feenstra and Levinsohn (1995), Thomas and Weigelt (2000), and Pinkse, Slade, and Brett (2001). The former two estimate a highly localized discrete-choice model in a space of several dimensions of product characteristics. The latter discriminate between theories of localized and global competition. Simplified versions of the Hotelling game were subject to experimental research. The impact of collusion in the location choice stage was examined by Brown-Kruse, Cronshaw, and Schenk (1993) and Brown-Kruse and Schenk (2000). In laboratory experiments, individuals have chosen to locate at the center if communication was possible. Otherwise, there was a tendency towards the profit maximizing point at the quartiles. Huck, Müller, and Vriend (2000) investigated how individuals behave in a four-players game without price competition. In the Nash equilibrium, two players locate at each of the quartiles. The experimental results, however, suggest that because of best-response dynamics there is a tendency to locate at the center of the market. An experiment with location and price choice was carried out by Barreda, García, Georgantzís, Andaluz, and Gil (2000). They found that differentiation is smaller than predicted by economic theory.

Further research in the Hotelling tradition could strongly benefit from extending this branch of literature empirically and experimentally. Thus, it would be possible to test if the model assumptions and the equilibrium concepts are appropriate and if its conclusions better describe reality than competing theories. Otherwise, this literature remains purely conceptual without having any predictive power. Our review may facilitate to test some predictions of the model. On the theoretical side, it may be interesting to investigate the relationship to other theories. The paper by Anderson and de Palma (2000) who synthesize a unifying framework of local and global competition, seems to be quite promising in this respect.

# Chapter 3

# Hotelling games with three, four, and more players

This chapter extends the standard Hotelling model with quadratic transport costs to the multi-firm case. Considering locational equilibria we show that neither holds the Principle of Maximum Differentiation - as in the duopoly model - nor does the Principle of Minimum Differentiation - as in the multiple firms game with linear transport cost. Subgame perfect equilibria for games with up to nine players are characterized by a U-shaped price structure and interior corner firms locations. In equilibrium the level of differentiation is almost at the socially optimal level if the number of firms is larger than three. Otherwise, there is too little differentiation.

## 3.1 Introduction

The literature on Horizontal Product Differentiation focuses on the extent to which competing firms should give their products a similar design interpreted as a location in the space of product characteristics. For the two-stage model where firms choose locations in the first stage and set prices in the second stage, Hotelling (1929) acclaimed the Principle of Minimum Differentiation. According to this principle firms approach each other as closely as possible and share the market equally. D'Aspremont, Gabszewicz, and Thisse (1979) corrected him by showing that in his model neither this strategy nor any other location choices were subgame perfect since they fail to imply an equilibrium in prices for each subgame. By altering the utility function from a linear to a quadratic form, resulting in a tougher second stage price competition, the Principle of Maximum Differentiation could be established where firms maximize the distance to the opposite player.[1]

Influenced by this result some researchers believed that "this [maximum] differentiation behavior could be fairly general." (Neven 1985: 322). However, subsequent results indicate that relaxing certain model assumptions shifts the balance away from the centrifugal towards the centripetal forces within the model leading to less differentiation. For example, Economides (1986) considers a whole family of utility functions which lie between the linear and the quadratic form. He shows that some equilibria exist where differentiation is not at maximum. Böckem (1994), Hinloopen and van Marrewijk (1999), and Wang and Yang (1999), among others, generalize the model on the demand side. More general distributions of the consumers are introduced into the model by Neven (1986) and Tabuchi and Thisse (1995). Others analyze markets without boundaries (Lambertini 1994) , the impact of demand uncertainty (Balvers and Szerb 1996), the introduction of information exchange through communication (Mai and Peng 1999), and endogenous household locations (Fujita and Thisse 1986). In all those cases, introducing flexibility into the model may destroy the equilibrium of maximum differentiation and may lead to in-between solutions.[2]

However, the major part of the literature concentrates on duopoly markets only. Widely ignored is the question of how the number of firms affects the equilibrium outcome. Exceptions comprise de Palma, Ginsburgh, and Thisse (1987) who treat the three-firm case in a probabilistic framework, and Lancaster (1979), Salop (1979), Novshek (1980), and Economides (1989, 1993) who do not restrict the number of firms. Linear utility models with multiple firms located on a circumference were analyzed by Salop (1979) and Economides (1989). It has been shown by Salop that in contrast to its unit interval duopoly counterpart considered by Hotelling, price equilibria exist in the symmetric subgame where symmetry refers to an overall equal

[1] Note that the quadratic term in the utility function measures a loss of utility.

[2] This kind of equilibrium even appears in the model considered by D'Aspremont, Gabszewicz, and Thisse (1979) if one permits mixed strategy Nash equilibria (Bester, de Palma, Leininger, Thomas, v. Thadden 1996).

price and equidistance of succeeding firms. Nevertheless, a perfect equilibrium could not be found for every subgame. Economides (1993) elaborated on the similar $n$-firms interval model with linear utility. In contrast to the circular model it supports a noncooperative equilibrium in every price subgame but fails to imply an equilibrium for the stage of the location choices.

Within the class of multi-firm models with a quadratic utility function only the circular model was considered. Economides (1989) proved the existence of a price equilibrium for each pattern of locations and further the existence of a subgame perfect equilibrium with equidistantly located firms. This equilibrium can be interpreted as the Principle of Maximum Differentiation since firms try to maximize the minimum distance to each adjacent competitor. The circular model is seen as an approximation to the interval model which is sometimes less favored because it exhibits the interval limit problem. This introduces asymmetry into the model since it usually matters if a firm is an inside or a corner firm. Furthermore, the circumference might be appropriate as a representation of characteristics such as color but it fails when it comes to other characteristics which are ordered like a convex set of real numbers (height and weight, for example). This chapter fills the research gap by analyzing the multi-firm interval model with quadratic utility. Hence, this part of the work could be viewed as a convex combination of Economides (1989) who treats case of quadratic transport costs on the circular domain and Economides (1993) who treats the case of linear transport costs on the linear domain. Its main objective is to examine if the results deduced from these cases are sufficiently general to be extended to our case of the quadratic transport costs interval model.

Essentially, behavior of Hotelling firms is driven by a trade-off between the short-run and the long-run effects of relocation. In the short-run, firms may attract new customers by moving towards a competitor's position. The strategic effect is a lower price for both the aggressor and the stationary firm because of the increased competition. Which effect dominates depends on the model setup.

In the Hotelling game with two firms and a quadratic utility function in equilibrium firms locate at the interval borders. By differentiating, players avoid the tough price competition of the second stage. One may expect a similar behavior when the number of firms increases, i.e. the corner firms are located at the interval boundaries and the remaining firms are equidistant in order to maximize the minimum distance to their neighbors. One of the main results of the chapter is that the Principle of Maximum Differentiation does not hold for the interval model. Considering this symmetric configuration we show that the corner firms have incentives to move inwards. Furthermore, weakening this principle by allowing corner firms to squeeze their (equidistantly located) inside competitors is not a valid equilibrium strategy either if $n > 3$. At the other extreme we can exclude Minimum Differentiation in the sense that firms do not have incentives to move towards the central firm(s) in every given locational pattern as is the case in the related linear utility model. These results are very interesting because they correspond neither to the equilibrium in the circular domain (Economides 1989), nor to the linear domain with linear transport costs (Economides 1993), nor to our model setting with only two firms. Hence, it is shown that the number of firms plays an important role for the extent to which firms are differentiating from each other.

Due to numerical difficulties, explicit location and price equilibria could only be computed for a maximum number of nine firms. In those cases we obtain symmetric equilibria where corner firms move considerably towards the center firm(s) and prices are U-shaped. The maximum differentiation results from the circular model are destroyed because of the monopoly power that benefits the corner firms. In the linear utility case this leads to a strong tendency towards the market center which prevents the existence of locational equilibria (Economides 1993). In the model considered here, the greater curvature of the utility function secures the existence of locational equilibria.

The chapter is organized as follows. In the next section the model is described. In Section 3.3 it is demonstrated that a price equilibrium exists for all locational patterns. In Section 3.4 it is elaborated on symmetric price equilibria. Analyses of equidistant locational configurations and explicit perfect equilibria for games with up to nine players follow in Section 3.5 and 3.6, respectively, which represent the core of this work. Section 3.7 concludes the chapter. Some of the proofs are contained in Appendix A.

## 3.2 The model

We examine a generalized Hotelling-game with quadratic utility of customers. While d'Aspremont, Gabszewicz, and Thisse (1979) consider a duopoly model we allow for an arbitrary (but fixed) number of firms. The game proceeds in two stages. In the first stage $n$ firms choose locations $\bar{x} = (x_1, x_2, ..., x_n)$ on the unit interval $[0, 1]$. At the second stage, prices $\bar{p} = (p_1, p_2, ..., p_n)$ are fixed simultaneously by the firms. The interval can be seen as a street in which firms represent shops. The firms produce a homogeneous product and sell one unit to each consumer. Consumers living in the street are equally distributed over the interval. Firms use the same constant returns to scale production technology. Marginal costs are normalized to zero.

Since products are homogeneous only transport costs matter for the decision from which firm to buy. Consumers are endowed with the following utility function which is separable with respect to the products:

$$u_w(x_j, p_j) = k - p_j - (x_j - w)^2$$

where $x_j$ represents the location of firm $j$. Consumer utility $u_w$ has a peak where its location $w$ and the firm's location coincide. The term $-(x_j - w)^2$ can be interpreted as the quadratic disutility which consumers incur through the distance of transport. $k > 0$ is the reservation price. Only if $k$ exceeds the sum of price and transport costs does the consumer buy.

We are looking for perfect Nash equilibria in pure strategies and assume the coordination problem away. Thus, firms are exogenously assigned numbers which represent their position in the spatial order. We neglect them not only in view of the analytical challenges they introduce but also because mixed strategies are not played by people in complex situations (Rapoport and Amaldoss 2000). Nevertheless, ignoring the coordination problem in the multi-firm setting can be seen as a shortcoming of the model since no plausible explanations exist why and how such a formation could appear, in particular if the number of firms is large.

## 3.3 Price equilibrium

In order to solve for a subgame perfect equilibrium, we first consider the last stage of the game. At this stage locations are already chosen and prices are to be posted simultaneously and non-cooperatively. To establish results on the existence and uniqueness of the price equilibrium we start by examining the demand of the firms.

The purchasing decision of a consumer is determined by its position and reservation price, the position of the firms, and the prices they charge. In the duopoly case when the reservation price constraint is binding, i.e., if $u_w$ of some individuals located between the shops becomes negative for the price-location combination of both firms, these customers will not buy any products. Then firms become local monopolists. Let us consider the case where the reservation price is sufficiently high such that every consumer buys a unit of the product. Assuming further that each firm chooses a different location, the duopoly demand can easily be calculated by finding the consumer who is indifferent between buying from the left or from the right firm, i.e. where $u_w(x_1, p_1) = u_w(x_2, p_2)$. This individual divides the set of consumers into two convex subsets where the left subset will be supplied by the left firm and vice versa. In the multi-firm case, demand for an inside firm $j$ $(1 < j < n)$ is not necessarily the set between the marginal consumers with respect to firm $j-1$ and $j+1$. This holds because the direct neighbors might attract no demand at all by charging prohibitively high relative prices. In this case firm $j$ competes with two of the more distant firms. However, this could never be an equilibrium because a firm with no demand would be better off lowering its price until it attracts some customers.

Assuming each firm has a positive market share, demand can be expressed as:

$$D_{1,0}(\bar{p}|\bar{x}) = \tfrac{p_2-p_1}{2(x_2-x_1)} + \tfrac{x_1+x_2}{2},$$
$$D_{j,0}(\bar{p}|\bar{x}) = \tfrac{p_{j+1}-p_j}{2(x_{j+1}-x_j)} - \tfrac{p_j-p_{j-1}}{2(x_j-x_{j-1})} + \tfrac{x_{j+1}-x_{j-1}}{2}, \text{ if } 1 < j < n,$$
$$D_{n,0}(\bar{p}|\bar{x}) = 1 - \tfrac{p_n-p_{n-1}}{2(x_n-x_{n-1})} - \tfrac{x_n+x_{n-1}}{2}.$$

Obviously, demand $D_{j,0}$ is linear in price $p_j$.[3] Starting from a price $p_j^{prohib}$ at

[3] The first subscript is the number of the firm while the second subscript denotes the number of neighbors whose demand is totally withdrawn by firm $j$.

which the marginal consumer is indifferent between buying from firms $j-1$, $j$, and $j+1$, by successively decreasing its price an inside firm $j$ will reach a value $p_j'$ at which it drives one of the neighbors out of business and starts to compete with the next firm. Then the above relationship does not hold anymore. If for example the whole demand of firm $j+1$ would have been withdrawn by firm $j$, i.e. for the marginal consumer with $u_w(p_{j+2}, x_{j+2}) = u_w(p_j, x_j)$ inequality $u_w(p_j, x_j) > u_w(p_{j+1}, x_{j+1})$ holds, the corresponding demand relationship would be:

$$D_{j,1}(\bar{p}|\bar{x}) = \frac{p_{j+2}-p_j}{2(x_{j+2}-x_j)} - \frac{p_j-p_{j-1}}{2(x_j-x_{j-1})} + \frac{x_{j+1}-x_{j-1}}{2}, \text{ if } 1 < j < n.$$

This subsequent piece of the demand curve is linear too but less steep. Decreasing price $p_j$ further leads to the undercutting of more and more competitors until the firm has attracted all customers or $p_j = 0$. Figure 1 shows the resulting demand curve. Undercutting of neighbors leads to kinks in this curve at prices $p_j'$, $p_j''$, and $p_j'''$ where the whole demand of the respective neighbors is withdrawn. The resulting demand function of firm $j$ corresponds to the lower envelope of all $D_{j,l}$ $(l = n-1)$. Obviously it is concave. Similarly, one obtains a concave demand curve for the corner firms. Hence, we have established:

**Proposition 1** *The demand function $D_j$ is concave with respect to the firm's price $p_j$.*

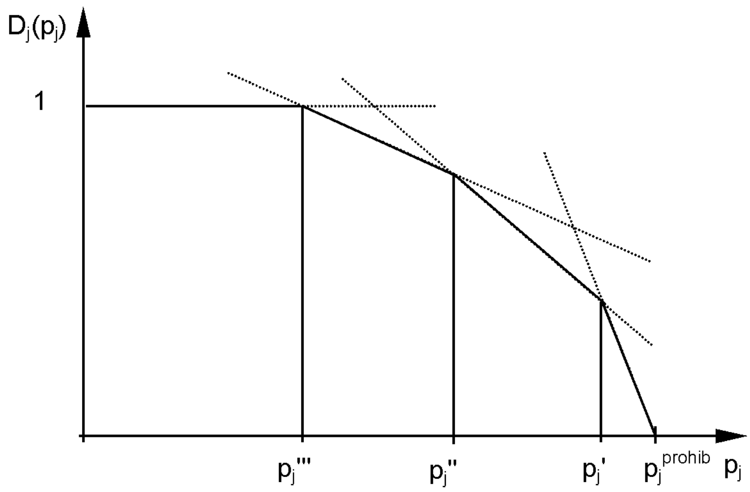


Figure 3.1: Demand curve for an inside firm with three neighbors

Now we can apply Economides' (1989) reasoning for the circular model. By applying Kakutani's fixed point theorem we are able to prove the existence of a non-cooperative price equilibrium (Friedman 1977). Moreover, uniqueness of this equilibrium can be shown. Economides' proof for the circular model is easily extended to the interval model.

Note that up to this point configurations were neglected where some firms choose the same location. If this is the case then these firms would price their products at marginal cost because it would always pay off to undercut the rival's price to attract some demand given prices are higher than marginal costs. This logic follows the well known Bertrand result. Together with the above considerations this leads to

**Proposition 2** *For every given pattern of locations there exists a unique equilibrium of the price setting stage.*

Given each firm has a positive market share maximizing the profit function $\Pi_j(\bar{p}|\bar{x}) = p_j D_{j,0}(\bar{p}|\bar{x})$ with respect to $p_j$ leads to the following price reaction functions:

$$
\begin{aligned}
&p_1^*(\bar{p}|\bar{x}) = \left[p_2 + (x_1 + x_2)(x_2 - x_1)\right]/2, \\
&p_j^*(\bar{p}|\bar{x}) = \tfrac{p_{j+1}(x_j - x_{j-1})}{2(x_{j+1} - x_{j-1})} + \tfrac{p_{j-1}(x_{j+1} - x_j)}{2(x_{j+1} - x_{j-1})} + \tfrac{(x_{j+1} - x_j)(x_j - x_{j-1})}{2}, \text{ if } 1 < j < n, \\
&p_n^*(\bar{p}|\bar{x}) = \left[p_{n-1} + (2 - x_{n-1} - x_n)(x_n - x_{n-1})\right]/2.
\end{aligned}
$$

Obviously, the system of prices is identified since the number of independent equations equals the number of variables. Let us concentrate on symmetric configurations where corner firms locate at $\alpha\delta$ and $1 - \alpha\delta$, respectively and inside firms' positions are at equal distance $\delta = \frac{1}{2\alpha + n - 1}$ away from their direct neighbors. In the case of $\alpha = 0$ this configuration corresponds to the Principle of Maximum Differentiation in the sense that firms maximize the minimum distance to their direct neighbors.

Resulting prices can be expressed as:

$$
\begin{aligned}
&p_1^*(\bar{p}|(\alpha\delta, \alpha\delta + \delta, \alpha\delta + 2\delta, .., 1 - \alpha\delta)) = \left[p_2 + \delta^2 + 2\alpha\delta^2\right]/2, \\
&p_j^*(\bar{p}|(\alpha\delta, \alpha\delta + \delta, \alpha\delta + 2\delta, .., 1 - \alpha\delta)) = \left[\tfrac{p_{j+1}}{2} + \tfrac{p_{j-1}}{2} + \delta^2\right]/2, \text{ if } 1 < j < n, \\
&p_n^*(\bar{p}|(\alpha\delta, \alpha\delta + \delta, \alpha\delta + 2\delta, .., 1 - \alpha\delta)) = \left[p_{n-1} + \delta^2 + 2\alpha\delta^2\right]/2.
\end{aligned}
$$

These are equilibrium prices only if it would not be profitable for any firm $j$ to drive others out of the market. If this would be the case then we should insert a demand term $D_{j,l}(\bar{p}|\bar{x})$, $l > 0$, into the above profit function instead of $D_{j,0}(\bar{p}|\bar{x})$. However, for the symmetric configurations as considered here we can show that driving neighbors out of business is not a credible strategy because this would require to set a price below zero. Thus, we conclude

**Proposition 3** *Given the symmetric equidistant market structure* $\bar{x} = (\alpha\delta, \alpha\delta + \delta, \alpha\delta + 2\delta, .., 1 - \alpha\delta)$ *where* $\alpha \geq 0$ *the equations* $p_1^*(\bar{p}|\bar{x}) = \left[p_2 + \delta^2 + 2\alpha\delta^2\right]/2$, $p_j^*(\bar{p}|\bar{x}) = \left[\frac{p_{j+1}}{2} + \frac{p_{j-1}}{2} + \delta^2\right]/2$, *if* $1 < j < n$, *and* $p_n^*(\bar{p}|\bar{x}) = \left[p_{n-1} + \delta^2 + 2\alpha\delta^2\right]/2$ *represent equilibrium prices.*

**Proof.** See Appendix A. ■

Most of our proofs rely on a specific representation of the equilibrium prices in the equidistant configuration. For example, considering the implicit price function of an inside firm $j$, $p_j^* = \left[\frac{p_{j+1}}{2} + \frac{p_{j-1}}{2} + \delta^2\right]/2$, Lemma 2 in the Appendix tells us that $p_j^*$ can also be stated in the following two ways:

$$p_j = a_j p_{j-1} + (1 - a_j)\,\delta^2 + b_j \alpha \delta^2$$
$$p_j = c_j p_{j+1} + (1 - c_j)\,\delta^2 + d_j \alpha \delta^2$$

The parameters of the first equation incorporate prices and positions of firms which are located on the left of firm $j$ while the parameters of the second equation capture those located on the right side. As shown in Lemma 2 upper and lower boundaries can be calculated for the parameters which only depend on the number of firms on the right respectively on the left and on the size of $\alpha$. Coefficient $\alpha$ corresponds to the ratio of the distances between the corner firm and its interval boundary and the distance between two direct neighbors. Substituting $p_{j-1}$ and $p_{j+1}$ in the equation of $p_j^*$ and transforming readily leads to an equation without prices on the right hand side. For this price we can calculate upper and lower boundaries as well which are sufficiently close for our proofs to work.

## 3.4 Symmetric price equilibria

Before we turn to the first stage of the game, firms' pricing behavior is analyzed for some symmetric configurations as they might be candidates for a subgame perfect equilibrium. Let us first consider the symmetric configuration where corner firms locate at the interval boundaries and inside firms have equal distance $\delta = \frac{1}{n-1}$ to their direct neighbors. This configuration corresponds to the Principle of Maximum Differentiation in the sense that firms maximize the minimum distance to their direct neighbors. Resulting prices can be expressed as:

$$p_j^*(\bar{p}|(0, \delta, 2\delta, .., 1)) = \left[p_{j+1} + \delta^2\right]/2, \text{ if } j = 1,$$
$$p_j^*\left(\bar{p}|(0, \delta, 2\delta, .., 1)\right) = \left[\frac{p_{j+1}}{2} + \frac{p_{j-1}}{2} + \delta^2\right]/2, \text{ if } 1 < j < n,$$
$$p_j^*\left(\bar{p}|(0, \delta, 2\delta, .., 1)\right) = \left[p_{j-1} + \delta^2\right]/2, \text{ if } j = n.$$

It is easily tested that a solution for this system would be $p_j = \frac{1}{(n-1)^2}$ $(j = 1, .., n)$. Hence, there is a uniform price level which decreases quadratically with the number of firms.

Possibly, corner firms possess some market power since they compete only with rivals on one side. Then they might take advantage of this fact by moving inwards. Let us elaborate on the equilibrium prices in the resulting configuration where corner firms are located at distance $\alpha\delta < 0.5$ from the interval limits and inside firms are located at a distance $\delta$ away from their direct neighbors. Note that firms are equidistantly located again. In this case we obtain a system of price equations which differs from the previous one only for firm 1 and firm $n$:

$p_j^*(\bar{p}|(\alpha\delta, \delta+\alpha, 2\delta+\alpha, .., 1-\alpha\delta)) = [p_{j+1} + \delta^2 + 2\alpha\delta^2]/2$, if $j = 1$,
$p_j^*(\bar{p}|(\alpha\delta, \delta+\alpha, 2\delta+\alpha, .., 1-\alpha\delta)) = [p_{j-1} + \delta^2 + 2\alpha\delta^2]/2$, if $j = n$.

Considering the Hotelling model with linear transport costs, Economides (1993) shows that such a symmetric configuration leads to a U-shaped curve of prices. As stated in Proposition 4 this is the case in our model, too.

**Proposition 4** *Given a symmetric and equidistant market structure $\bar{x} = (\alpha\delta, \delta + \alpha, 2\delta + \alpha, .., 1 - \alpha\delta)$ the equilibrium prices are the higher the closer a firm's location is to one of the border firms with respect to the spatial order of firms.*

**Proof.** See Appendix A. ■

We find that indeed corner firms are able to charge higher prices than their competitors in the interior. This is attributable to the fact that they only have to compete on one side. Corner firms are not as much affected by price decisions of their competitors as inside firms since their distance to other firms is maximal with respect to the given spatial ordering. At the other extreme, the center firm is closest to its rivals in this respect and therefore in a position in which it is difficult to raise its prices. Not only corner firms benefit from the advantage of being remote to competition but also their closer neighbors. The corner firm's market power decreases continuously towards the central firms.

Having characterized equilibrium price structures on the basis of some simple locational patterns, we are now ready to turn to the first stage of the game.

## 3.5 Symmetric locational patterns

In this section we are concerned about the existence of symmetric locational equilibria given the price reaction functions of the last stage. Symmetry in this context means that the configuration of locations in the left half of the interval is mirrored in the right half. Attention is restricted to locational patterns with equal distance between direct neighbors. We show that in contrast to the circular model no symmetric equidistant locational equilibrium can be sustained in the interval model when the number of firms is larger than three. Thus, *a fortiori* the Principle of Maximum Differentiation is not valid in this case. Moreover, we show that firms do not have an incentive to move towards the middle firm(s) regardless of how firms are distributed along the interval. This destroys the possibility for the (strong) Principle of Minimum Differentiation to hold.

Following Economides (1989) we argue that two firms would not choose the same location since this would drive prices down to the level of the marginal costs. Deviating from this position increases profits. Let us elaborate on the symmetric configuration where corner firms are located at the interval boundaries and inside firms have equidistant positions. It is a candidate equilibrium since it corresponds to the Principle of Maximum Differentiation. This principle holds for the duopoly and might thus be also an equilibrium outcome of the multi-firm case.

In order to represent an equilibrium the symmetric choices of locations have to maximize the profit function $\Pi_j(p^*(\bar{x}), x_j|\bar{x}_{-j})$ of each firm $j$ where $\bar{x}_{-j}$ are the symmetric locations of all firms except firm $j$ and $p^*$ are the equilibrium prices of the last stage. Changes of profits as a result of a marginal movement arise from two different effects. The first one is called the demand effect. Relocating firms could gain (lose) additional demand by approaching (moving away from) one of their neighbor firms. The second effect is called the strategic price effect and is induced by the second stage price competition. Since prices are fixed sequentially after the choice of the location, rivals can react on an approach by lowering their prices.

The incentives to deviate from the given configuration might be greatest for the corner firms since their demand effect is positive.[4] Indeed, according to the first order conditions which turn out to be negative for the right and positive for the left firm while the second order conditions are negative, there are incentives for the corner firms to move inwards. It can be shown that the symmetric market structure $\bar{x} = (0, \delta, 2\delta, .., 1)$ where corner firms are located at the interval boundaries and the inside firms are located equidistantly does not support a perfect equilibrium. Hence, we conclude that:

**Proposition 5** *The Principle of Maximum Differentiation does not hold in the multi-firm unit interval Hotelling game with quadratic disutility.*

**Proof.** See Appendix A. ■

As expected the corner firms have a substantial interest to change their locations from the boundaries towards their competitors' positions which is due to the direct demand effect overcompensating the price effect. This establishes a market power of the corner firms which drives them to squeeze their rivals in-between. Another potential symmetric equilibrium might be suggested which accounts for the market power of the corner firms. In this configuration the distance between direct neighbors equals $\delta$. Corner firms are located $\alpha\delta$ $(0 < \alpha)$ away from the interval boundaries. Hence, the two corner firms squeeze all inside firms which themselves maximize the minimum distance to their neighbors.[5] Let us first consider the corner players. We have shown that their profits increase when they deviate from the maximum differentiation pattern. This does not generally hold if one allows the corner players to choose a position inside the interval (without changing their relative position). Concluding from the following lemma the incentives to relocate decrease proportionally with $\alpha$ such that for a certain value of $\alpha$ the first order conditions of the corner

[4]For the symmetric pattern considered here one can easily show that the demand effect is zero for every inside firm since all prices are equal.

[5]In the case that $\alpha = \frac{1}{2}$ this would correspond to the welfare maximizing configuration.

firms become zero while the second order conditions are negative provided that this location choice represents a local maximum.[6]

**Lemma 1** *Given the symmetric equidistant market structure* $\bar{x} = (\alpha\delta, \alpha\delta + \delta, \alpha\delta + 2\delta, .., 1 - \alpha\delta)$ *where* $\alpha > 0$ *the profit maximizing first order condition of firms* $j = 1$ $(j = n)$*,* $\frac{\partial \Pi_j(p^*(\bar{x}), x_j|\bar{x}_{-j})}{\partial x_j}$ *is positive (negative) if* $\alpha < \frac{(3-2a_3)(7-2a_3)}{(7-2a_3)^2-(1-2a_3)(1+b_3)}$*, zero if* $\alpha = \frac{(3-2a_3)(7-2a_3)}{(7-2a_3)^2-(1-2a_3)(1+b_3)}$*, and negative (positive) otherwise, where* $a_n = \frac{1}{2}$*,* $a_{i-1} = \frac{1}{4-a_i}$ *and* $b_n = 1$*,* $b_{n-1} = \frac{b_n}{4-a_n}$*. Furthermore,* $\Pi_j(p^*(\bar{x}), x_j|\bar{x}_{-j})$ *is concave in* $x_j$*.*

**Proof.** See Appendix A. ■

Lemma 1 states that in the symmetric configuration corner firms use their market power to move inwards as long as the price reduction from the increased second stage competition does not exceed the gains of new demand. Given the parameter boundaries of Lemma 2 the critical value of $\alpha$ can be calculated as one third in the case of three firms. If the number of firms goes to infinity it approaches 0.385. Parameter $\alpha$ can be interpreted as an indicator of market power possessed by the corner firms. It reflects the best response of corner firms to an equidistant configuration. Although the market power increases with the number of firms it remains on a moderate level.

From Lemma 1 we can construct an equilibrium if the number of firms equals three (just insert $a_3 = \frac{1}{2}$ and $b_3 = 1$ into the above $\alpha-$term). The resulting value of $\alpha$ equals one third, i.e. in this case the distance between the corner firms and their respective interval boundary is one third of the distance between the corner firm and the market center. Concavity of the profit function with respect to the locations is easily established.[7] Together with the existence of the price equilibrium (see Section 3.3) it can be concluded that the corresponding location choices represent a subgame perfect equilibrium.

**Proposition 6** *If the number of firms equals three* $\bar{x} = \left(\frac{1}{8}, \frac{1}{2}, \frac{7}{8}\right)$ *represents subgame perfect equilibrium location choices.*

If the number of firms is larger than three, the value of $\alpha$ for which the first order conditions of the corner players become zero should also provide no incentive for the inside firms to relocate, i.e. at this value their respective first order conditions should be zero, too, provided the profit function is quasi-concave. However, calculating the first order conditions of the inside firms for $n = 4$ leads to a negative value of $\alpha$ and

[6] The conditions could also be stated in terms of $c_{n-2}$ and $d_{n-2}$. Since they are equal to $a_3$ and $b_3$, respectively, by construction this would not make a difference.

[7] One can show that the profit function of firm $j$ is concave in each of the three intervals in which the unit interval is divided by the two rivals' locations. Then it remains to calculate the profits of jumping to another position in the spatial formation (leapfrogging). However, it is obvious that leapfrogging would not be profitable as the sales potential is small but the threat of increased price competition is huge. Proofs are available from the author.

to values of $\alpha$ larger than one if $n$ is greater than five.[8] Taking into account that the optimal level of $\alpha$ for a corner firm is not larger than 0.385 this means that if $n \geq 5$ inside firms are more attracted by the central firm(s) than corner firms. This can be attributed to the weakness of the central firms to respond to price changes of their neighbors as described above. Hence, we conclude

**Theorem 1** *No symmetric equidistant market structure* $\bar{x} = (\alpha\delta, \alpha\delta+\delta, \alpha\delta+2\delta, .., 1-\alpha\delta)$ *where* $\alpha \geq 0$ *is supported in the multi-firm unit interval Hotelling game if the number of firms is larger than three.*

Up to this point we have shown that the strong Principle of Maximum Differentiation does not represent equilibrium behavior in oligopoly. Even the weak version of this principle where only corner firms do not differentiate maximally is an equilibrium if at least four firms are competing. In the following we will focus on the opposite strategy, the Principle of Minimum Differentiation. Examining the unit interval multi-firm Hotelling model with linear transport costs, Economides (1993) has shown that this principle is prevalent in its strong version. The strong version states that whatever spatial pattern firms start there are always incentives to move towards the central firm.

However, in this section we have shown that the two outside-right and the two outside-left firms would not move inwards in any (symmetric equidistant) configuration because this would not maximize their respective profit functions. In other words, there exist values of $\alpha \in R$ for which the first order conditions of the quasi-concave profit functions become zero. Thus we conclude

**Theorem 2** *The strong Principle of Minimum Differentiation does not hold in the multi-firm unit interval Hotelling game with quadratic disutility.*

We have shown that some prominent locational structures can be excluded from the set of potential equilibrium strategies. Unfortunately, we still do not know which if any equilibria appear in the case of more than three players. This question is addressed in the following section.

## 3.6 Locational equilibria for three to nine players

In order to obtain explicit locational equilibrium solutions $\bar{x}^* = (x_1^*, x_2^*, ..., x_n^*)$ for the multi-firm case one first needs the explicit price reaction functions of the last stage of the game. As derived in Section 3.3 the price of a firm depends on its

[8] It can be shown that $\alpha = \frac{(26-7a_4)}{(2-6b_4-a_4)}$ makes the first order condition of the second or the second to last player, respectively, zero where $2-\sqrt{3} < a_4 \leq \frac{1}{2}$ and $b_4 \leq 1$. Obviously, the latter constraints permit negative solutions for $\alpha$. This would be the case if the number of firms is four because then $a_4 = \frac{1}{2}$ and $b_4 = 1$ (see Lemma 2). Increasing the number of firms leads to decreasing values of $a_4$ and $b_4$, such that $\alpha$ becomes larger than one in the case of five competitors.

location, the location of its direct neighbors, and the prices they charge. Hence, prices depend indirectly on all other prices and locations. Transforming this system of implicit price functions into a system of functions with only locations on the right hand side is a difficult task. In the previous section this was accomplished by restricting the location choices to symmetric equidistant configurations. Without imposing this structural constraint the equation system could not be solved when the number of firms is large. Hence only those cases are considered where the market consists of not more than nine firms.

Equilibria were found by calculating the first and second order conditions of the profit function providing that the solution represents a local profit maximum. To exclude an instability of the pattern with respect to larger steps of relocation we made sure that it would not pay off for any firm to change the position along the interval limited by its direct neighbors' locations given the location of every other firm. In particular, it is shown numerically that $\frac{d^2\Pi_j(p^*(x_j,x^*_{-j}),x_j|x^*_{-j})}{dx_j^2} < 0$ for $x_j \in \left[x^*_{j-1}, x^*_{j+1}\right]$, if $1 < j < n$, and for $x_1 \in [0, x^*_2]$ and $x_n \in \left[x^*_{n-1}, 1\right]$, otherwise.[9] Notice that the second derivative of the profit function is continuous in the considered interval and hence, the sign is easy to ascertain if the nulls of this function are known. Furthermore, leap-frogging has to be considered. Hence, we examined if it is profitable for any firm to change the location with respect to the spatial order, given that the locations of the competitors remain fixed. This was accomplished by searching for each firm $j$ the profit maximizing location within each of the neighbors' intervals and compare this value with the equilibrium profits. It appeared that leap-frogging profits are far below equilibrium profits. The computations were run using the Mathematica software package.[10]

The solutions for the three- to nine-player cases are shown in Figure 2. As we may expect from the results of the previous section the corner firms move considerably towards the inside firms. In the case of three competitors the distance to the interval boundary is 0.125. Adding firms to the market leads to corner firm locations which are closer to the interval end points in absolute terms. Thus, it seems that competition for the corner firms becomes tougher as they are pushed towards the borders. However, relating the absolute distance between the corner firm and its respective interval boundary to the average distance between the direct neighbors shows a reverse picture. The relative position of the corner firms will be strengthened the more players are in the game. In the four-player case this ratio is about one half as opposed to one third in the three firms case. The oligopoly cases of five to nine players lead to only weakly increasing values of about 0.53. Thus apparently the pressure for the corner firms to differentiate decreases in relative terms with the

[9]In some instances the boundaries of the region of concavity are close to the respective interval boundaries $\left[x^*_{j-1}, x^*_{j+1}\right]$. Hence, due to a certain inexactness of the numerical results, it can not be excluded that profits are not concave very close to these boundaries. However, if the location of firm $j$ is close to one of its neighbors' locations the equilibrium price $p^*_j(x_j)$ becomes sufficiently small such that the respective profit never exceeds the equilibrium profit.

[10]The programs are available from the author on request.

number of firms since price competition is softened at the same time.

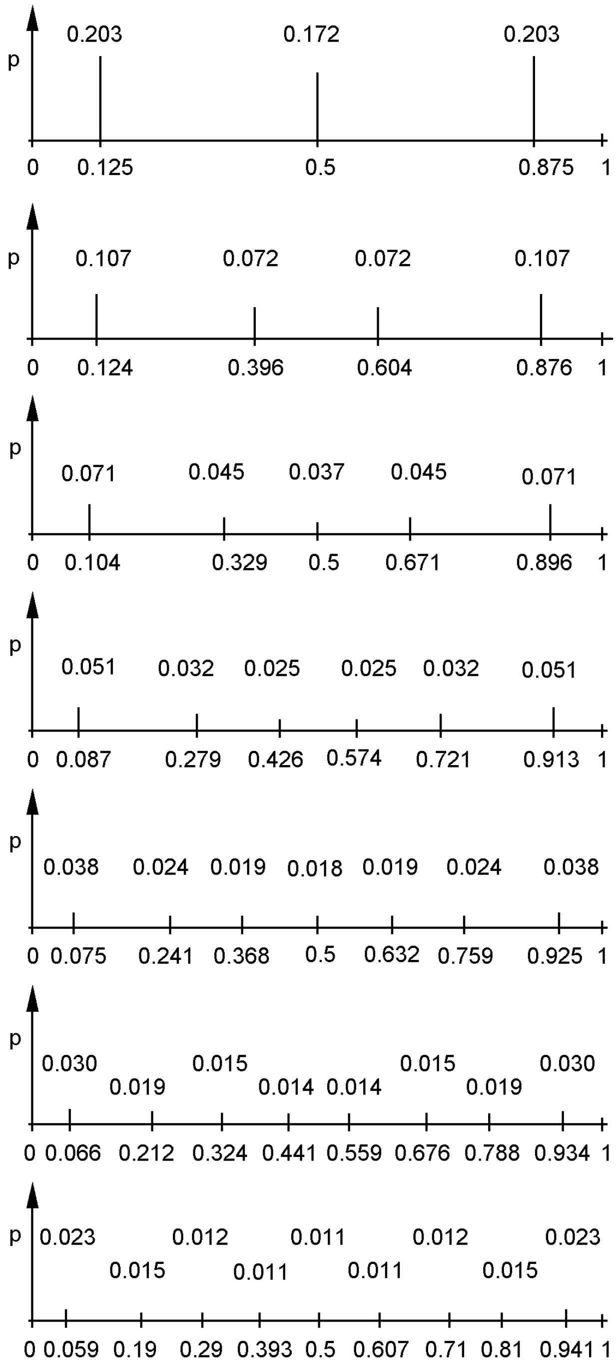


Figure 3.2: Equilibrium outcome in pure strategies of locations and prices for the three to nine firms Hotelling game (from top to bottom)

| firms | Firm 1 | Firm 2 | Firm 3 | Firm 4 | Firm 5 | Firm 6 | Firm 7 | Firm 8 | Firm 9 |
|---|---|---|---|---|---|---|---|---|---|
| three | 0.0550 | 0.0787 | 0.0550 | - | - | - | - | - | - |
| four | 0.0209 | 0.0219 | 0.0219 | 0.0209 | - | - | - | - | - |
| five | 0.0113 | 0.0105 | 0.008 | 0.0105 | 0.0113 | - | - | - | - |
| six | 0.0068 | 0.0063 | 0.0043 | 0.0043 | 0.0063 | 0.0068 | - | - | - |
| seven | 0.0044 | 0.0041 | 0.0028 | 0.0025 | 0.0028 | 0.0041 | 0.0044 | - | - |
| eight | 0.0030 | 0.0028 | 0.0024 | 0.0017 | 0.0017 | 0.0024 | 0.0028 | 0.0030 | - |
| nine | 0.0022 | 0.0020 | 0.0013 | 0.0012 | 0.0012 | 0.0012 | 0.0013 | 0.0020 | 0.0022 |

Table 3.1: Equilibrium profits

Viewed from the welfare perspective this is an interesting result. The socially optimal outcome requires that all succeeding firms have equal distance. Corner firms however locate one half of this stretch away from the interval boundaries since this pattern would minimize transport costs. Although it was shown that this configuration is not supported as an equilibrium in this model the patterns of four and more players come rather close to the optimal solution. In games with two and three players firms could increase welfare by choosing locations closer to each other than in equilibrium. In games with more than four firms there seems to be a (slight) tendency of too few differentiation of the corner firms.

Surprisingly, the market outcome becomes more and more similar to the optimal social solution if the number of firms is large. This is attributable to the change in the price competition. Asymmetry in price competition is large between firms close to the boundaries. This applies in particular to the firms 1, 2, and $n-1$, $n$, respectively. However, as the asymmetry decreases between firms near the market center the incentives to differentiate become similar leading to similar distances between them.

As shown in Table 2 equilibrium profits decrease with the number of firms. This is due to the increase of price competition induced by additional competitors. More interestingly, the relative increase of market power enjoyed by the corner firms following an increase of the number of rivals is also reflected in the firms' profits. In the three-firm case the equilibrium profit of the center firm considerably exceeds the corner firms' profits. In this case a higher number of consumers attracted by the center firm compensates for the lower price. When four players compete, profits are almost equally distributed while in the oligopoly with five or more players central firms earn much less than their corner rivals. Although the center firm finds itself in an uncomfortable position there is no higher profitable market niche to relocate. It cannot be excluded that this might change as the number of firms grows. Positions between the boundary and the corner firm or between the corner and its neighbor firm may provide a region large enough to be more profitable than the center location if the number of firms exceeds nine. Then subgame perfect equilibria may not be identified by first and second order conditions alone which makes the search for them more difficult or even destroys the possibility for a perfect equilibrium.

## 3.7 Conclusions

We analyzed the multi-firm unit interval Hotelling model assuming quadratic transport costs. Firms choose locations in the first stage and prices in the second stage of the game. Existence and uniqueness of the short run price game is established. In contrast to the results for the similar circular model (Economides 1989) and the duopoly interval (d'Aspremont, Gabszewicz, and Thisse 1979) model it has been shown that the Principle of Maximum Differentiation does not hold. In the corresponding configuration where corner firms are located at the interval boundaries and adjacent firms are located equidistantly corner firms would benefit from moving marginally towards the market center. Other symmetric equidistant configurations do not represent perfect equilibria either. Equally the (strong) Principle of Minimum Differentiation does not hold as it is prevalent in the multi-firm interval model with linear transport costs (Economides 1993). Hence, if a perfect equilibrium exists it must correspond to an in-between differentiation configuration. Consequently, it is shown that the number of firms plays an important role for the extent to which firms are differentiating from each other.

Explicit symmetric perfect equilibria are calculated for games with up to nine firms. They are characterized by an U-shaped price structure and corner firms which are located away from the interval boundaries. The distance between the corner firms and their direct neighbors is larger than the distances between the inside firms. This follows because of an asymmetric price competition. Furthermore, it seems that corner firms are in a favorable position since they are able to squeeze the market to a certain extent and to charge the highest prices. In fact the profits exceed their rivals' profits only when $n \geq 5$. If the number is smaller then they suffer from having a demand limit at one side. Increasing the number of firms leads to a change in the stage of price competition. Center firms' positions are weakened because more rivals positioned close by increase the price competition by which corner firms are less affected. The resulting increase of "relative" market power together with an increase of overall competition leads to lower absolute but higher relative profits of the corner firms. Further, the competitive advantage of the corner firms is partially transferred to their neighbors which is reflected in higher prices between neighbors towards the boundaries.

Viewed from the welfare perspective we find too much differentiation if the number of competing firms is small ($n \leq 3$). Increasing this number shifts relatively more power to the corner firms which use this power to squeeze their inside rival. This leads to an approximately optimal level of differentiation in the equilibrium case of more than three players.

How does the number of firms influence the equilibrium outcome of Hotelling games? Considering our results and reviewing Salop (1979) and Economides (1989, 1993), it seems that increasing the number of firms has three effects. First, price equilibrium is not endangered. For the quadratic transport costs model, a Nash equilibrium exists for the duopoly as well as for the oligopoly. Assuming linear transport costs price equilibria exist for the multi-firm case in symmetric configurations while it does not exist in the duopoly case. Second, the threat of (second stage) competition is reduced if a market boundary is present. For the multi-firm interval models under linear transport costs this leads to the strong Principle of Minimum Differentiation while it destroys the Principle of Maximum Differentiation under quadratic transport costs. Third, given a market boundary corner firms enjoy a greater market power as their inside competitors. This is reflected by a U-shaped price structure in symmetric and equilibrium configurations and larger than average distances between corner firms and their direct neighbors.

## 3.8 Appendix A

**Proof of Proposition 3.** Since prices maximize the profit function if each firm has a positive market share we only have to show that driving neighbor firms out of the market does not pay off. For this purpose we show that given a symmetric and equidistant market structure $\bar{x} = (\alpha\delta, \alpha\delta + \delta, \alpha\delta + 2\delta, .., 1 - \alpha\delta)$, $\alpha \geq 0$, the price $p'_j$ $(1 < j \leq n)$ for which $D_{j-1,0}(p^*_{-j}, p'_j|\bar{x}) = 0$ is negative. Likewise the price $p'_j$ $(1 \leq j < n)$ for which $D_{j+1,0}(p^*_{-j}, p'_j|\bar{x}) = 0$ is negative.

If firm $j's$ neighbors are inside firms their demand in the equilibrium configuration can be expressed as $D_{j-1,0}\left(p^*_{-j}, p_j|\bar{x}\right) = \frac{p^*_{j+1}-p_j}{2\delta} - \frac{p_j-p^*_{j-1}}{2\delta} + \delta$ and $D_{j+1,0}\left(p^*_{-j}, p_j|\bar{x}\right) = \frac{p^*_{j+1}-p_j}{2\delta} - \frac{p_j-p*_{j-1}}{2\delta} + \delta$, respectively (see Section 3.3). Setting these equations to zero and solving for $p_j$ leads to $p_j = 2p^*_{j-1} - p^*_{j-2} - 2\delta^2$ and $p_j = 2p^*_{j+1} - p^*_{j+2} - 2\delta^2$. By applying Lemma 2 (see below) it is easy to show that the equilibrium prices cannot be larger than $\delta^2$ assuming each firm has a positive market share. Of course, the best response prices will not increase if firm $j$ lowers its price. Then it is clear that $p_j$ must be negative. If firm $j's$ neighbor is a corner firm the demand of the corner firm in the equilibrium configuration can be expressed as $D_{j-1,0}\left(p^*_{-j}, p_2|\bar{x}\right) = \frac{p_j-p^*_l}{2\delta} + \frac{\delta}{2}$, $l \in \{0, n\}$. Setting this equation to zero and solving for $p_j$ leads to $p_j = p^*_l - \delta^2$. In this case the same argumentation applies as before. ■

**Proof of Proposition 4.** Let us consider two arbitrarily chosen succeeding firms $i$ and $i+1$ $(1 \leq i \leq n-1,\ n > 2)$ where $n$ is the number of firms. If firm $i$ is the left corner firm $(i = 1)$ then we obtain the following implicit price reaction functions (see Section 3.4): $p^*_1 = \frac{p_2}{2} + \frac{\delta^2}{2} + \alpha\delta^2$ and $p^*_2 = \frac{p_1}{4} + \frac{p_3}{4} + \frac{\delta^2}{2}$. According to Lemma 2 the price function of firm 3 can be represented as: $p^*_3 = a_3 p_2 + (1 - a_3)\delta^2 + b_3\alpha\delta^2$ where $2 - \sqrt{3} < a_3 \leq \frac{1}{2}$, and $0 < b_3 \leq 1$. Solving this system for $p^*_1$ and $p^*_2$ and subtracting leads to $p^*_2 - p^*_1 = \frac{\delta^2\alpha(-6+2a_3+b_3)}{(7-2a_3)}$, which is less than zero. Similarly we can show that $p^*_n - p^*_{n-1} > 0$.

If firm $i$ is one of the inside firms we obtain the following system of price functions according to Lemma 2:

$p_i = c_i p_{i+1} + (1 - c_i)\delta^2 + d_i\alpha\delta^2$

$p_{i+1} = a_{i+1}p_i + (1 - a_{i+1})\delta^2 + b_{i+1}\alpha\delta^2$,

where $2 - \sqrt{3} < a_{i+1} \leq \frac{1}{2}$, $2 - \sqrt{3} < c_i \leq \frac{1}{2}$ and $0 < b_{i+1} \leq 1$, $0 < d_{i+1} \leq 1$. After some calculations, we obtain

(*) $$p_i - p_{i+1} = \delta^2\alpha\frac{(1-a_{i+1})d_i - b_{i+1}(1-c_i)}{1-a_{i+1}c_i}$$

Now let us assume that firm $i$ is located on the left part of the interval, i.e. $i < \frac{n}{2}$. In this case it follows from Lemma 3 (which is proved below) that $c_i > a_{i+1}$ and $b_{i+1} < d_i$ and hence (*) is positive. Similarly this difference would be negative if $i \geq \frac{n+1}{2}$. Further, $p_i$ and $p_{i+1}$ are equal if firm $i$ and $i + 1$ are the two central firms. ■

**Proof of Proposition 5.** The profit maximizing first order condition of firm $j = 1$ is $\frac{\partial\Pi_1(p^*(\bar{x}),x_1|\bar{x}_{-1})}{\partial x_1} = \frac{\partial D_1}{\partial x_1} + \frac{\partial D_1}{\partial p_2}\frac{\partial p^*_2}{\partial x_1} = \frac{1}{2} + \frac{p^*_2-p^*_1}{2(x_2-x_1)^2} + \frac{1}{2(x_2-x_1)}\frac{\partial p^*_2}{\partial x_1} \overset{!}{=} 0$. Solving this equation requires explicit price reaction functions. By taking advantage of Lemma

2 we only consider implicit functions for $p_1^*$ and $p_2^*$. Price $p_3^*$ is represented in the following way where $a_3$, and $b_3$ have the properties according to Lemma 2.

$p_1^* = \frac{1}{2}p_2 + \frac{(x_1+x_2)(x_2-x_1)}{2}$

$p_2^* = \frac{p_3(x_2-x_1)}{2(x_3-x_1)} + \frac{p_1(x_3-x_2)}{2(x_3-x_1)} + \frac{(x_3-x_2)(x_2-x_1)}{2}$

$p_3^* = a_3p_2 + (1-a_3)\delta^2 + b_3\alpha\delta^2$

Note that corner firms are located at the interval boundaries and hence $\alpha = 0$. This system can easily be translated into a system of explicit price reaction functions. After some calculations we obtain $\frac{\partial\Pi_1(p^*(\bar{x}),x_1|\bar{x}_{-1})}{\partial x_1} = \frac{1}{2} - \frac{2}{[7-2a_3]}$ and $\frac{\partial^2\Pi_1(p^*(\bar{x}),x_1|\bar{x}_{-1})}{\partial x_1^2} = -\frac{6(5-2a_3)}{\delta(7-2a_3)^2}$. From Lemma 2 we know that $2-\sqrt{3} < a_3 \leq \frac{1}{2}$ and hence $\frac{\partial\Pi_1(p^*(\bar{x}),x_1|\bar{x}_{-1})}{\partial x_1} > 0$ and $\frac{\partial^2\Pi_1(p^*(\bar{x}),x_1|\bar{x}_{-1})}{\partial x_1^2} < 0$. ■

**Proof of Lemma 1.** We proceed as in the previous proof which leads us to the first order condition of firm $i = 1$ : $\frac{\partial\Pi_1(p^*(\bar{x}),x_1|\bar{x}_{-1})}{\partial x_1} = \frac{1}{2} - \frac{2}{(7-2a_3)} + \frac{\alpha}{2}\left(\frac{(1-2a_3)(b_3+1)}{(7-2a_3)^2} - 1\right)$ where $2-\sqrt{3} < a_3 \leq \frac{1}{2}$ and $0 < b_3 \leq 1$. It is easy to show that the first order condition becomes zero at $\alpha = \frac{(3-2a_3)(7-2a_3)}{(7-2a_3)^2-(1-2a_3)(1+b_3)}$ and for all values of $\alpha$ below (above) the first order condition would be positive (negative).

Now consider the second order condition $\frac{\partial^2\Pi_1(p^*(\bar{x}),x_1|\bar{x}_{-1})}{\partial x_1^2} =$

$$\underbrace{-\frac{(23-10a_3)}{\delta(7-2a_3)^2}}_{<0}\underbrace{-\frac{\alpha\left(339-284a_3+80a_3^2-8a_3^3-4b_3+10a_3b_3-4a_3^2b_3\right)}{\delta(7-2a_3)^3}}_{<0} < 0$$

It is obvious that all the terms are negative and hence, the second order condition is negative irrespective of $\alpha$. Further, the profit function is concave and thus quasi-concave in $x_1$. The proof for the opposite corner firm $i = n$ works analogously. ■

**Lemma 2** *If corner firms' distance to their respective interval boundary is $\alpha\delta$ ($\alpha \geq 0$) and the distance between direct neighbors equals $\delta$ then the price reaction function of firm $j$ ($1 < j \leq n$) can be represented as $p_j = a_jp_{j-1} + (1-a_j)\delta^2 + b_j\alpha\delta^2$, where $a_n = \frac{1}{2}$, $a_{j-1} = \frac{1}{4-a_j}$, and $b_n = 1$, $b_{j-1} = \frac{b_j}{4-a_j}$. Further, for every $i$, $0 \leq i \leq n$ parameters are constrained as follows: $2-\sqrt{3} < a_j \leq \frac{1}{2}$ and $0 < b_j \leq 1$. Equivalently, price reaction function of firm $j$ ($1 \leq j < n$) can be represented as $p_j = c_jp_{j+1} + (1-c_j)\delta^2 + d_j\alpha\delta^2$, where $c_1 = \frac{1}{2}$, $c_{j+1} = \frac{1}{4-c_j}$, and $d_1 = 1$, $d_{j+1} = \frac{d_j}{4-c_j}$. For every $0 \leq i \leq n$ parameters are constrained as follows: $2-\sqrt{3} < c_j \leq \frac{1}{2}$ and $0 < d_j \leq 1$.*

**Proof.** The first part of Lemma 2 will be proven by complete induction. The proof of the last part can be completed analogously.

As shown in Section 3.3 the price reaction function of firm $j$ is $p_j^*(\bar{p}|(\alpha\delta, \alpha\delta + \delta, \alpha\delta + 2\delta, .., 1-\alpha\delta)) = \left[p_{j-1} + \delta^2\right]/2 + \alpha\delta^2$, if $j = n$. Hence, Lemma 2 holds in this case.

*Induction hypothesis:* We will show that whenever Lemma 2 holds for firm $j$ it holds for firm $j-1$, too.

*Induction proof:* For an inside firm $j-1$ the price reaction function derived from the first order conditions is $p^*_{j-1}(\bar{p}|(\alpha\delta, \alpha\delta+\delta, \alpha\delta+2\delta, .., 1-\alpha\delta)) = \left[\frac{p_j}{2} + \frac{p_{j-2}}{2} + \delta^2\right]/2$. Given $p_j = a_j p_{j-1} + (1-a_j)\delta^2 + b_j\alpha\delta^2$ and inserting this expression into the former equation leads to $p_{j-1} = a_{j-1}p_{j-2} + (1-a_{j-1})\delta^2 + b_{j-1}\alpha\delta^2$ where $a_{j-1} = \frac{1}{4-a_j}$, and $b_{j-1} = \frac{b_j}{4-a_j}$. If $2-\sqrt{3} < a_j \leq \frac{1}{2}$, and $0 < b_j \leq 1$ hold, one can easily check that also $2-\sqrt{3} < a_{j-1} \leq \frac{1}{2}$, and $0 < b_{j-1} \leq 1$ hold. ■

**Lemma 3** *Given* $a_n = \frac{1}{2}$ *and* $b_n = 1$ *then (i) the sequences* $a_i = \frac{1}{4-a_{i+1}}$ *and* $b_i = \frac{b_{i+1}}{4-a_{i+1}}$ *are increasing in* $i$ $(i = 1, .., n)$*. Equivalently, given* $c_1 = \frac{1}{2}$ *and* $d_1 = 1$ *then (ii) the sequences* $c_i = \frac{1}{4-c_{i-1}}$ *and* $d_i = \frac{d_{i-1}}{4-c_{i-1}}$ *are decreasing in* $i$ $(i = 1, .., n)$*.*

**Proof.** (i) Given $a_n = \frac{1}{2}$ it is easily shown by complete induction that $2-\sqrt{3} < a_i < 2+\sqrt{3}$ $(i = 1, .., n-1)$. Then it is clear that $a_i < a_{i+1}$. Further, it follows immediately that $b_i = \frac{b_{i+1}}{4-a_{i+1}} < b_{i+1}$ given the constraints of $a_i$. Proof of (ii) works analogously. ■

# Chapter 4

# Uncertain and dynamic consumer switching costs

In this chapter we analyze the optimal timing of an investment in a new distribution technology such as the Internet. It is a common observation that competition increases and prices decrease if competition is transferred from traditional ways of distribution to the Internet. Hence, firms may like to postpone the introduction of this technology in order to avoid tough rivalry. One way to accomplish this would be to take advantage of customers locked in the company because of switching costs. We focus on the costs customer incur if they switch from the old to the new distribution technology. They are distinguished from the usually considered switching costs (Klemperer 1995) because they may differ across customers, they are temporally decreasing because the Internet technology diffuses, and they may be uncertain.

We show that the optimal time to invest under the given switching costs depends negatively on the price gap between the old and the new technology markets and the market growth of the demand for new technology services or products. In contrast to many real option models, we find that there is no monotonous relationship between uncertainty and the investment timing. Given intermediate values of uncertainty firms would invest earlier compared to cases of very high and very low uncertainty. We also find that firms with more loyal customers, i.e. firms whose customers remain faithful for a longer time, would switch later to the new technology.

## 4.1 Introduction

One of the central findings of the literature on switching costs is that there is a trade-off between attracting new valuable repeat-purchasers by charging a low price and reaping already attached customers by charging a high price.[1] Hence, this line of research is able to explain what may cause the firm's pricing strategies and consequently, why prices and profits are higher if switching costs are present. Further insights were given into how firms use prices as entry deterrents and what the welfare implications of this are. Largely neglected, however, is the question of how a new distribution technology could reshape the barrier for customers to switch from the firm they are buying from to one of its competitors and how this would affect the introduction time of this technology. In this chapter, we will argue that the new distribution technology may affect the amount and dynamics of consumer switching costs in a stochastic way. These, in turn, will influence the optimal time to invest in the new technology.

The Internet provides the infrastructure for new kinds of distribution channels. In many industries, managers face the question of whether and when to make use of this opportunity. The advantage of the computer network is, among others, its broad geographic reach and its high efficiency. Hence, new customers can be attracted and the costs of production are to be decreased (for example, because of decreased inventory in the retail industry). Furthermore, new valuable services could be developed for Internet customers.

Also the customers potentially benefit from this new technology since services may be cheaper and of better quality. In retail banking, for example, it could be observed that prices for several kinds of banking and brokerage services strongly decreased after introducing Internet facilities. However, since the Internet requires the customers to be equipped with the respective devices, software and knowledge, additional costs are imposed on them if they already buy from an old technology firm: the costs of switching to the new distribution technology. The decision of becoming an Internet user and incur the respective costs may not be influenced by the availability of a single application (e.g. online banking) but by a basket of different services. One could safely assume that the number of Internet users follows a common technology diffusion process and is exogenous to the decision of a firm to offer a new service. Thus, the technology related switching costs are decreasing over time.

From the management perspective, one of the most focal questions is when it would be appropriate to introduce the new distribution technology. Why should timing matter? First, firms would be eager to keep their market share. If the new online offer is broadly accepted in the population as a more convenient way of carrying out the transactions, it would become necessary for a firm to offer this kind of service, too. An early mover may prevent potential rivals to enter the market. Moreover, it is reasonable to assume competitive parameters changing if new

[1] An excellent survey on the topic of switching costs is provided by Klemperer (1995).

distribution technologies are available. The Internet leads to a higher transparency and thus, possibly, to a higher degree of competition. Hence, firms may like to delay the introduction of this technology. Here the switching costs come into play: Are firms able to delay the investment in new technology in order to skim their attached customers? If switching costs are high, the answer may be yes. If switching costs are decreasing, however, there is a trade-off between losing customers with low switching costs and increasing the degree of competition by introducing the new technology. An additional incentive to switch early may be growth prospects of the electronic market. Finally, firms may not know the consumer switching costs exactly. This uncertainty about demand could also affect the optimal time for investment.

We develop a dynamic model of continuous time in which two down-stream firms face the decision of investing in a new distribution technology. On the consumer side, there are complementary investments necessary in order to access the services or products provided through the new distribution channel. In the case of the Internet, these investments can be related to technical devices such as computers, modems, and software. Further, a certain amount of effort is necessary to learn how these devices work. These investments represent switching costs from old technology to new technology firms. Additionally, there might be the usually considered physical switching costs and the brand loyalty as psychological switching costs.[2] We assume that the amount of switching costs differ between the individuals because of a different endowment of the new devices and because individuals may differ in their ability to make use of them. However, the mean of the switching costs is decreasing over time since the new technology diffuses and becomes more easy to handle. This induces more and more customers recently locked in the old technology to migrate to a new technology company. The number of switching customers is uncertain. The firm can stop the process of migration by investing in the new technology.

Applying the net present value technique to decide whether to invest in the project may not be optimal since it ignores the value of delaying the investment until the conditions of the project are more favorable. Questions regarding the investment under uncertainty are naturally treated in the framework of real options (Dixit and Pindyck 1994, Trigeorgis 1998). The real option theory focuses on the analysis of the optimal time for investment in and the value of a risky project whose investment can be delayed and is at least partially irreversible. From this literature we borrow the view that certain investments should be interpreted as options. In our model, we consider the option to change the firm's distribution technology within some time interval. If the firm invests early, it may secure its base of customers and take advantage of demand growth for the new technology services. Furthermore, since uncertainty is related to the switching costs it can thus avoid a risky demand schedule. However, if market conditions are less favorable after introducing the new technology then the firm may sacrifice some future market shares by skimming its locked in customers. Note that the considered option has no date of maturity like

[2]Klemperer (1995) provides an overview over the different categories of switching costs.

many financial options.

Generally, the investment in new technology, in particular in information technology (IT) has attracted much interest by real option researchers. Schwartz and Zozaya-Gorostiza (2000) present a methodology to evaluate IT projects. They distinguish between IT development and acquisition projects and consider a decay of the costs of IT assets over time. A particular kind of IT investment - the investment in platforms - was considered by Kogut and Kulatilaka (1994) and Perotti and Rossetto (2001). Platforms are understood as means to accomplish a higher flexibility of the production process, as strategic entry options, or just simply as growth options. Clearly, this literature is related to our work since the new distribution technology can also be viewed as a platform in the latter sense. Other papers on growth options like Kulatilaka and Perotti (1999) and Kulatilaka and Perotti (1998) emphasize the positive effect of these options on receiving greater opportunities and analyze how advantages of a long-term commitment to an investment strategy may dominate the real options logic of staged investments. McGrath (1997) presents a model which extends the work of Dixit and Pindyck (1994) and others on real option and strategic management theory of technology choice. She introduces the view that technology investments as strategic investments may even decrease uncertainty. A different approach was taken by Wang (2001) who considers endogenous technology change in a general equilibrium setting and overcomes a technical problem related to many real option applications.[3] Our model is distinguished from the previous work on real options by considering explicitly a decreasing stochastic demand schedule on the old technology market, an increasing deterministic demand schedule on the new technology market, and a tougher competitive environment at the new technology market associated with lower prices. Making these restrictive assumptions seems to be plausible in light of the Internet example as an important new distribution technology. Hence, previous work on real option theory provides the appropriate context to discuss our problem but fails to take into account switching costs as an important characteristic of real markets. Another more technical issue in which our model differs from many others is the number of stochastic sources. In more simple models there is usually only one source of uncertainty. In our case, the stochastic demand process which results implicitly from the (not explicitly modeled) switching costs is different between firms but it is allowed to be correlated. As a result, the problem cannot be solved analytically but by a new numerical method presented by Kamrad and Ritchken (1991). This method was recently applied by Chi (2000) for the valuation of shares of a joint venture.

Our model also differs from the approaches of the industrial economics literature on switching costs. Most importantly, because we do not focus on issues like optimal

[3] This problem is associated with the so-called "spanning assumption" (Dixit and Pindyck, 1994): The underlying asset of the option or an asset whose value is correlated with it must be traded in every point of time spanning the uncertainty. If there is no such asset, usually a discount rate is fixed exogenously. However, this rate may be dependent on the technology choice and thus be endogenous.

pricing or entry deterrence but on the optimal investment timing and because of analytical tractability, we take prices and demand as exogenously given. Like Farrell and Shapiro (1988), we model the dynamics of the game by using a continuous time approach.[4] Similar to Gabszewicz, Pepall, and Thisse (1992) we assume that switching costs are heterogeneous across the customers. However, they consider them as the effort of learning-by-using. In contrast, we interpret the switching costs more broadly as including also physical investments. We drop the standard assumption of time constant switching costs by replacing it by the assumption of temporal decreasing costs of switching induced by technology diffusion.[5] Moreover, we introduce the choice of the firms of the adoption of a new technology which itself has an impact on switching costs to the extent that they are technology dependent. Beggs (1989) also considers technology choice, but not on the firm but on the consumer level.

In summary, neither the literature on switching costs nor the previous work on investment under uncertainty is able to explain the optimal time to change to a new distribution technology when switching costs are present. We first investigate which model parameters influence the timing decision of the investment if demand is not uncertain. Then we analyze the stochastic case. Most interestingly, we find that increasing the uncertainty does not necessarily lead to a delay of the investment, but rather that there is no monotonous but a U-shaped relationship between uncertainty and the investment timing. Given intermediate values of uncertainty, firms would invest earlier compared to cases of very high and very low uncertainty.

In the following section we will provide some additional motivation for our model which is then presented in Section 4.3. Subsequently, we consider the case in which uncertainty is absent while Section 4.5 treats the case of stochastic demand. Subsection 4.5.1 is devoted to the calculation of the option values. In Subsection 4.5.2 we compute the optimal time to invest under uncertainty for some simulations. Finally, the chapter is concluded in Section 4.6.

## 4.2 Motivation

Switching costs are frequently described as the costs which customers incur if they change their bank accounts. During the past decade, the banking industry has witnessed great technological changes. Through the emergence of computer networks it suddenly became possible to run a bank without necessarily open branch banks. The concept of "online banking" emerged. Online banking is a new distribution channel which is associated with financial platforms in the Internet. Hence, in

[4] For two-period models see, for example, Klemperer (1987a,b), Basu and Bell (1991), Padilla (1992), and Banerjee and Summers (1987). A multi-period model of discrete time was presented by Beggs and Klemperer (1992), To (1996), and Padilla (1995).

[5] In another context, the declining switching costs could also be interpreted as a kind of social learning.

order to motivate our model the retail banking industry seems to be an appropriate candidate.

With the emergence of the Internet, banks were faced with a multitude of challenges. There were new opportunities and a potential for growth as the geographic reach of customers could be increased dramatically and products better suiting the customer needs could be distributed more easily. Furthermore, there appeared a potential of cost reduction by implementing the new processing and distribution technologies since the financial transactions could be processed automatically requiring less investment in real estates and human capital. However, threats were connected with the new technology since new competitors were attracted by the opportunities and the competitive environment was possibly due to a change.

The new distribution technology may be of advantage to customers since it could lower their costs of a financial transaction. Given that customers already make use of the Internet, some of them may prefer the new technology over the traditional one since it is more convenient, more flexible, and maybe cheaper. However, the emergence of a new technology also imposes additional costs to the customers. If a customer previously wished to change its bank account there were some immediate expenses related to the closure of the current account and the opening of the new one. Also non-physical costs, known as brand loyalty, may play a role.[6] Now, if a customer switches from an old technology to a new technology firm there are additional switching costs which are technology related. These costs create a barrier between the two submarkets characterized by a different technology.

Customers differ when it comes to their computer equipment and the knowledge of how to use it. For some customers it is very easy to switch. In the given example, these customers are already experienced users and are provided with the necessary devices for Internet access (PC, modem, etc.). In contrast, customers with higher switching costs would have to invest in learning and in equipment in order to be enable to use this facility. Hence, these switching costs are not stable over time. Rather, they are dependent on the number of customers already using the new distribution technology which in turn depends on the speed of the diffusion of the Internet. Usually, the process of new technology adaptation resembles a diffusion process (Geroski 2000). However, it is important to note that switching costs between two banks providing online services may be lower than between two conventional banks since the competitor is only "a mouse-click away".

From the bank's perspective the question arises then if and when to implement the new technology. Determinants of this decision may be actual and potential competition on both submarkets. The Internet may lead to a higher market transparency and may thus be more competitive.[7] Of course, also the size of the consumer

[6]It is a frequent observation that customers are loyal towards brands or companies making them less sensitive to better quality and lower priced substitutes.

[7]William Harrison, the president of US bank Chase Manhattan, emphasizes on the 1999 Bank Administration Institute Retail Delivery conference that "[t]he Internet has created a massive power shift to the buyer...Prices are more easily compared - choice has dramatically increased -

switching costs is relevant for this problem as it might relax the competitive pressure if its value is high. Since we argued that these costs are decreasing parallel to the diffusion of the Internet technology, also the speed of this process may influence the timing. In contrast, the growth of demand for new technology products may take effect in the opposite direction. The impact of uncertainty on the time of investment is not clear. Real option theory suggests that as uncertainty increases so does the optimal time to invest in order to avoid unfavorable outcomes. In our case, however, it might be the other way around since a risky demand schedule can be terminated by the technology investment.

## 4.3 The model

Consider the following setup. In the market, initially, there are two incumbent firms, denoted as firm 1 and firm 2, selling a homogeneous product. The product is produced at constant average costs which are normalized to zero. The incumbents are threatened by a potential competitor, firm 3, ready to enter the market. The distribution technology by which the product is delivered to the customers plays a crucial role in our model. In the beginning, the firms apply the traditional technology. The incumbent firms are free to switch to the new technology at any point of time by investing the amount of $I$. Once the investment is made, it cannot be recovered. In order to focus on the decision of the incumbents about when to switch to the new technology and to keep the analysis simple, we assume throughout the chapter that firm 3 enters the market at time $t = 0$ delivering the product by the new technology.

Regarding the customers, they only prefer the new technology if they are enabled to make use of it, which means they have to be equipped with some technical devices. We assume that the number of customers who already possess such devices follows a diffusion process. Hence, the dynamics of the switching costs between the two technologies are a consequence of a technology diffusion process.

Let us call the submarkets as $m1$ (old technology) and $m2$ (new technology). As soon as the services offered on $m2$ are available, some customers immediately choose the new offer. These are the customers with low switching costs. However, some customers stay with their supplier. We refer to this fraction of customers, however motivated, as loyal customers. The parameter of consumer loyalty $\alpha_i$, $i \in \{1, 2\}$, measures the relative amount of consumers buying previously from the incumbent who stay with their firm in spite of new technology offers for an incremental period of time. The loyalty parameters determine by how much the old technology submarkets shrink in an increment of time due to the availability of new technology services.

There are two distinct cases to be considered for the evolution of demand of firm $i$ at time $t$. In the first case, the incumbent firm does not switch to the

and the physical limits of comparison shopping are gone."

new technology. Here the respective demand decreases according to the following differential equation:

$$\frac{dq_i^{m1}(t)}{q_i^{m1}(t)} = -\alpha_i dt + \sigma_i dz,\ i \in \{1,2\}\,, \tag{1}$$

where $q_i^{m1}(t)$ is the demand quantity at time $t$ on the old technology market attracted by firm $i$, $i \in \{1,2\}$. Equation (1) corresponds to a Geometric Wiener process with drift $-\alpha_i$ and variance $\sigma_i$. $dz$ is the increment of a standard Wiener process with mean zero and standard deviation one. The consumers which are migrating move from firm $i$ to firm 3.

In order to fully specify the dynamics of demand we have to state what happens in the moment when one firm changes the technology. At this moment, firm $i$ can stop the migration process. The size of $m2$ is assumed to be increasing at the rate $\beta$. Hence, given the time of switching as $T_i$, demand of firm $i$ on $m2$ can be expressed as:

$$\frac{dq_i^{m2}(t)}{q_i^{m2}(t)} = \beta dt,\ q_i^{m2}(T_i) = q_i^{m1}(T_i)g(T_i), \tag{2}$$

where $g(t)$ is a factor of growth of this submarket with the initial level $g(0)$. For analytical convenience, we assume it to evolve exponentially as $g(t) = g(0)e^{\beta t}$. Of course, $q_i^{m2}(t) = 0$ if $t < T_i$.

We call the market prices for one unit of the product $p^{m1}$ and $p^{m2}$, respectively. Note that these prices are net of marginal costs and, thus, represent the mark-ups on both submarkets. Since the marginal costs are lower on market $m2$, we assume $p^{m1} \geq p^{m2}$. This is also reasonable because there may be different competitive conditions on both submarkets. Traditional distribution technologies are likely to occur together with a low price transparency. Hence, they imply a soft price competition. In contrast, the new technology firms face a tough price competition since facilities such as the Internet lead to a high market transparency. Apart from this, prices are assumed to be constant over time.

We exclude the possibility of a switching firm to differentiate between its customers, i.e. after the new technology is implemented the new price applies to all of its customers.

## 4.4 The deterministic case

Let us first consider the case in which demand is not a stochastic variable. It is assumed for simplicity that the customers migrating from firm 1 and firm 2 all switch to firm 3.[8] The profit function of firm $i$, $i \in \{1,2\}$, over an infinite time horizon can then be stated as

[8] Otherwise, the analysis would become much more complex. If some of the customers of the non-switching firm migrate to the switching firm this would increase the incentives to switch early. However, an equilibrium might not exist if both firms prefer to be first-movers. This assumption is

$$\Pi_i\left(T_i\right)=\int_0^{T_i} p^{m1} q_i^{m1}(t) e^{-rt} dt+\int_{T_i}^{\infty} g(t) p^{m2} q_i^{m1}(T_i) e^{-rt} dt-I(T_i) e^{-rT_i} \tag{3}$$

where $T_i$ denotes the time at which firm $i$ implements the new technology and $r$ is the rate of discount. The first term of (3) corresponds to the profits made before switching while the second and the third terms represent the profits after switching net of the discounted value of the investment. According to the differential equations (1) and (2) demand can be expressed by the exponential function $q_i^{m1}(0)e^{-\alpha_i t}$. Moreover, it is assumed that the market for new technology services increases exponentially at growth rate $\beta$ and the real value of the investment $I(t) = I(0)e^{rt}$ remains constant over time. Then the profit function becomes:

$$\Pi_i\left(T_i\right)=\int_0^{T_i} p^{m1} q_i^{m1}(t) e^{-rt} dt+\int_{T_i}^{\infty} g(0) p^{m2} q_i^{m1}(T_i) e^{(\beta-r)t} dt-I(0) \tag{4}$$

In order to obtain a regular solution, we assume that $0 < \beta < r$. We can easily find the optimal time to switch by solving the following maximization problem (see Appendix B):

$$T_i^* = \arg\max \Pi_i\left(T\right) = \max\left[0, \frac{1}{\alpha_i} \ln \frac{p^{m1}(r-\beta)}{g(0)p^{m2}(r+\alpha_i-\beta)}\right] \tag{5}$$

If $\alpha_1 = \alpha_2$ and $q_1^{m1} = q_2^{m1}$ then, of course, the optimal time to switch is identical for both firms.

As can be expected from this kind of model, the optimal time of implementing the new technology depends negatively on the amount of marginally switching customers $\alpha_i$. Hence, the more loyal the customers of a firm are the later the company switches to the new technology. Furthermore, the optimal point in time is later, the lower (higher) the price on the submarket $m1$ ($m2$) and the lower the coefficient of demand growth $g$. The switching time depends positively on the discount rate and negatively on the growth rate $\beta$ of the submarket $m2$.

Interestingly, the initial number of customers of firm $i$, $q_i^{m1}(0)$, has no impact on $T_i^*$. Hence, a company which is initially endowed with a greater market share does not exploit its customers for a longer period than their rivals do. This contrasts with other models of switching costs (Klemperer 1995). This result may be induced by the assumption that the investment necessary to implement the new technology grows with the discount rate. If we assume $I(t)$ to be nominally constant, for example, then we may obtain a positive relationship between $T_i^*$ and $q_i^{m1}(0)$.

---

also plausible in the case of the banking industry. A survey by the *Sunday Times* (Oct. 10 1999) revealed that only seventeen percent of bank customers say their banks represent good value. Hence, a great portion of them is likely to switch if at all to an entrant.

| Probabilities | $q_1^{m1}$ | $q_2^{m1}$ |
|---|---|---|
| $P_1 = P(u_1u_2)$ | $v_1$ | $v_2$ |
| $P_2 = P(u_1d_2)$ | $v_1$ | $-v_2$ |
| $P_3 = P(d_1u_2)$ | $-v_1$ | $v_2$ |
| $P_4 = P(d_1d_2)$ | $-v_1$ | $-v_2$ |
| $P_5 = P(h_1h_2)$ | 0 | 0 |

Table 4.1: The size of a demand jump in an incremental time given the respective probabilities

## 4.5 The stochastic case

In our model, demand is affected by uncertainty (see equation (1)). Uncertainty arises because firms do not know exactly the number of loyal customers at each point in time. This uncertainty about future demand might influence the decision about when to implement the new technology. Usually, in models with real options firms delay investments under uncertainty in order to eliminate the down-side risk (McDonald and Siegel 1986). In our case, uncertainty might accelerate the investment because this is how risk can be eliminated.

Unfortunately, it is not possible to solve the system of differential equations in the stochastic case analytically. The mathematics of option theory is even difficult for less complex problems and does not provide any tools for our case. Hence, we follow the standard procedure to approximate the stochastic processes and solve the problem via the techniques of dynamic programming (Dixit and Pindyck 1994).

The demand variables $\ln q_1^{m1}(t)$ and $\ln q_2^{m1}(t)$ are chosen to follow a Geometric Wiener Process with drift $\mu_i = -\alpha_i - \sigma_i^2/2$, where $-\alpha_i$ is the rate of decline if no uncertainty is present and $\sigma_i^2$ is the instantaneous variance. Then, we obtain:

$$\ln q_i^{m1}(t + \Delta t) = \ln q_i^{m1} + \zeta_i(t) \tag{6}$$

where $\zeta_i(t)$, $i \in \{1, 2\}$, are normal random variables with mean $\mu_i \Delta t$, and variance $\sigma_i^2 \Delta t$. We allow for correlation between the stochastic parameters. The parameter of correlation between the variables is called $\rho$.

The stochastic processes are approximated by discrete trinomial distributions as suggested by Kamrad and Ritchken (1991).[9] Modeling the processes by a trinomial distribution approximation implies that in each time increment $\Delta t$, the variables can jump upwards ($u_i$), downwards ($d_i$) , or remain constant ($h_i$). The joint distribution of the stochastic variables $\alpha_i$ for each time increment $\Delta t$ is shown in Table 3. The variables $v_i$ denote the jump parameters which measure the size of positive or negative changes in each time increment. Of course, the probabilities sum up to one.

[9] A similar methodology was applied by Chi (2000) who examines the option to acquire or divest a joint ventures with two sources of uncertainty.

Following Kamrad and Ritchken (1991), we parametrize the jump parameters by $v_i = \lambda\sigma_i\sqrt{\Delta t}$, $i \in \{1,2\}$, where $\lambda \geq 1$ is a coefficient which directly influences the probabilities. If $\lambda = 1$, the probability of a horizontal jump is zero. The higher the value of $\lambda$, the higher the probability of a horizontal jump. The resulting probabilities can be expressed as follows:

$$
\begin{aligned}
P_1 &= \tfrac{1}{4}\left\{\tfrac{1}{\lambda^2} + \tfrac{\sqrt{\Delta t}}{\lambda}\left(\tfrac{\mu_1}{\sigma_1} + \tfrac{\mu_2}{\sigma_2}\right) + \tfrac{\rho}{\lambda^2}\right\} \\
P_2 &= \tfrac{1}{4}\left\{\tfrac{1}{\lambda^2} + \tfrac{\sqrt{\Delta t}}{\lambda}\left(\tfrac{\mu_1}{\sigma_1} - \tfrac{\mu_2}{\sigma_2}\right) - \tfrac{\rho}{\lambda^2}\right\} \\
P_3 &= \tfrac{1}{4}\left\{\tfrac{1}{\lambda^2} + \tfrac{\sqrt{\Delta t}}{\lambda}\left(-\tfrac{\mu_1}{\sigma_1} - \tfrac{\mu_2}{\sigma_2}\right) + \tfrac{\rho}{\lambda^2}\right\} \\
P_4 &= \tfrac{1}{4}\left\{\tfrac{1}{\lambda^2} + \tfrac{\sqrt{\Delta t}}{\lambda}\left(-\tfrac{\mu_1}{\sigma_1} + \tfrac{\mu_2}{\sigma_2}\right) - \tfrac{\rho}{\lambda^2}\right\} \\
P_5 &= 1 - \tfrac{1}{\lambda^2}
\end{aligned}
$$

The trinomial approximations used in the next section are based on those probabilities.

### 4.5.1 The option values

Basically, there are two strategies available for each of the incumbent firms. The first would be to apply the old technology over the whole period of time. By doing so, an erosion of their initial consumer base may occur because unloyal consumers switch to firms offering products and services on the new technology submarket. In the end, the firm might lose all of its customers. Firms may prevent this by implementing the new technology. This decision can be made at some time $t \in [0,T]$ throughout the whole period of planning. Since the investment is irreversible and in the market there is demand uncertainty, it may pay off to wait until circumstances turn out to be more favorable. Another reason for postponing the investment is the profit of reaping the customers on the less competitive submarket $m1$. In our model, firms might as well speed up their investment in order to avoid a risky demand schedule and to get access to market $m2$ with its increasing volume.

The optimal time to invest is naturally addressed in the framework of real options. Let us consider the option to invest in the new technology.

The initial value for firm $i$ of investing in a platform at time $T_i$, $V_i(T_i)$, can be expressed in the following way:

$$
V_i(T_i) = p^{m2}\int_{T_i}^{T} g(0)q_i^{m2}(T_i)e^{(\beta-r)t}dt - p^{m1}\sum_{t=T_i}^{T} q_i^{m1}(t)e^{-rt\Delta t}dt - I(0) \qquad (7)
$$

The discounted value of the project consists of three parts. The left term in (7) measures the value of the gains from switching to the new technology at time $T_i$. Since uncertainty is resolved after switching, this value can be computed by standard algebra. In contrast, the opportunity costs of switching are affected by stochastic demand on the submarket $m1$. The right profit stream corresponds to the forgone profits by skimming the customers with high switching costs on the submarket $m1$. The costs of implementing the new technology are represented by the third term in (7). Since its nominal value increases with the discount rate $r$, its real value remains constant.

We assume that firms behave rationally in the way that they maximize the stream of profits over the whole period of consideration. This requires the firms to calculate for each point in time if the value of exercising the option is greater than the value of upholding the decision. The option value of implementing the new technology corresponds to the following dynamic maximization problem:

$$J_i(t) = \max \begin{bmatrix} V_i(t) \\ J_i(t + \Delta t)e^{-r\Delta t} \end{bmatrix} \tag{8}$$

Solving this equation lead us to the path of values of the optimal investment decision, each made at time $t$. Possibly, it does not pay to invest in the new technology right in the beginning because the losses of an increased price competition plus the costs of installment exceed the gains from growth. Then it would be optimal to wait with the investment. Note that (8) does not allow for negative values. If changing the technology could lead to losses then it would always be optimal to wait to the end. Finally, the option value to defer the investment from time $t = 0$ to some later time can be easily calculated as $\max[0, J_i(0)]$.

### 4.5.2 The optimal time to invest

In this section the simulation results are reported. In each case we have approximated the stochastic process of demand by one thousand values. The simulation program is based on the software package *Mathematica 4.0* and can be found in Appendix C. Simulations were run for different values of the stochastic process, i.e. for different values of the drift and the variance parameters.

Table 4 gives a clear picture of how the drift of demand influences the time to invest. In our numerical example, we consider a period of planning which consists of $T = 100$ subperiods. We find that in some cases if the decrease in the own customer base is not too large, no firm would optimally introduce the new technology. We also find that increasing the speed of demand erosion, *ceteris paribus*, tends to lead to an earlier adoption of the technology.

In Table 4, there are several panes according to different levels of correlation of demand change. In the case of perfect correlation and identical drift shown in the lower pane there is no difference in the switching time of both firms. Here the decrease of the switching costs is an industry-wide process. However, given the firms are asymmetric there must be some firm-specific difference in the ties between the customers and the company. They might result from different customer loyalty which may be induced by a different level of marketing. Another interpretation is that the firms' customer bases are heterogenous. For example, one company may

| | | $\mu_2$ | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | $\mu_1$ | -0.01 | | -0.02 | | -0.03 | | -0.04 | | -0.05 | |
| | -0.01 | 96.64 | 100 | | | | | | | | |
| | -0.02 | 86.97 | 99.16 | 89.78 | 89.17 | | | | | | |
| $\rho$=0.0 | -0.03 | 36.95 | 98.34 | 38.39 | 89.18 | 34.02 | 38.27 | | | | |
| | -0.04 | 5.47 | 99.09 | 8.20 | 82.13 | 6.74 | 33.50 | 6.80 | 6.58 | | |
| | -0.05 | 2.38 | 98.45 | 2.61 | 86.38 | 3.02 | 39.93 | 2.39 | 6.27 | 2.64 | 2.13 |
| | -0.01 | 100 | 100 | | | | | | | | |
| | -0.02 | 87.64 | 100 | 91.75 | 87.71 | | | | | | |
| $\rho$=0.2 | -0.03 | 40.39 | 100 | 32.57 | 86.99 | 36.71 | 37.97 | | | | |
| | -0.04 | 9.20 | 100 | 5.90 | 85.47 | 4.65 | 31.93 | 9.02 | 6.49 | | |
| | -0.05 | 2.58 | 100 | 1.87 | 87.21 | 2.08 | 42.26 | 2.03 | 5.36 | 2.22 | 2.96 |
| | -0.01 | 100 | 100 | | | | | | | | |
| | -0.02 | 89.75 | 99.00 | 87.07 | 86.63 | | | | | | |
| $\rho$=0.4 | -0.03 | 37.64 | 100 | 41.20 | 93.66 | 35.52 | 35.68 | | | | |
| | -0.04 | 4.90 | 100 | 8.95 | 89.82 | 5.71 | 38.00 | 7.21 | 5.89 | | |
| | -0.05 | 3.00 | 100 | 1.77 | 87.25 | 1.86 | 39.10 | 1.98 | 8.270 | 1.94 | 1.88 |
| | -0.01 | 99.36 | 100 | | | | | | | | |
| | -0.02 | 90.39 | 100 | 87.34 | 85.19 | | | | | | |
| $\rho$=0.6 | -0.03 | 36.77 | 100 | 34.92 | 83.89 | 38.396 | 39.49 | | | | |
| | -0.04 | 5.09 | 100 | 7.65 | 87.49 | 8.033 | 45.31 | 9.45 | 8.69 | | |
| | -0.05 | 3.67 | 99.16 | 1.98 | 81.58 | 2.014 | 40.69 | 2.09 | 7.59 | 1.99 | 2.51 |
| | -0.01 | 100 | 100 | | | | | | | | |
| | -0.02 | 91.22 | 99.26 | 91.66 | 93.64 | | | | | | |
| $\rho$=0.8 | -0.03 | 38.06 | 100 | 42.08 | 92.11 | 36.35 | 39.44 | | | | |
| | -0.04 | - | - | 9.67 | 86.90 | 8.07 | 38.55 | 7.61 | 10.50 | | |
| | -0.05 | - | - | - | - | 2.72 | 41.15 | 2.25 | 9.65 | 2.01 | 2.47 |
| | -0.01 | 100 | 100 | | | | | | | | |
| | -0.02 | - | - | 91.75 | 91.75 | | | | | | |
| $\rho$=1.0 | -0.03 | - | - | - | - | 46.14 | 46.14 | | | | |
| | -0.04 | - | - | - | - | - | - | 8.80 | 8.80 | | |
| | -0.05 | - | - | - | - | - | - | - | - | 2.40 | 2.40 |

Table 4.2: The optimal time to switch under different parameters of correlation and drift; the left columns correspond to firm 1; for the simulations, the following values are chosen: Invest = 1, q1 = 1.0, q2 = 1.0, pr1 = 2.8, pr2 = 1, lamda = 1.2, beta = 0.03, dt = 0.1, sig1 = 0.05, sig2 = 0.05, r = 0.01, g = 2

| | | $\sigma_2$ | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | $\sigma_1$ | 0.03 | | 0.23 | | 0.43 | | 0.63 | | 0.83 | | 2.00 | |
| $\rho$=0.0 | 0.03 | 58.98 | 57.14 | | | | | | | | | | |
| | 0.23 | 24.57 | 61.05 | 24.66 | 26.96 | | | | | | | | |
| | 0.43 | 29.38 | 49.54 | 24.71 | 24.53 | 22.88 | 32.44 | | | | | | |
| | 0.63 | 32.49 | 63.40 | 35.50 | 24.73 | 33.95 | 34.26 | 30.50 | 37.26 | | | | |
| | 0.83 | 37.25 | 55.34 | 32.08 | 22.57 | 33.10 | 30.62 | 34.25 | 32.21 | 33.52 | 29.03 | | |
| | 2.00 | 43.16 | 58.31 | 40.76 | 24.96 | 44.24 | 24.28 | 36.99 | 37.97 | 42.88 | 31.29 | 42.31 | 40.11 |
| $\rho$=0.2 | 0.03 | 61.43 | 61.68 | | | | | | | | | | |
| | 0.23 | 26.12 | 62.06 | 22.12 | 24.56 | | | | | | | | |
| | 0.43 | 30.90 | 53.54 | 26.63 | 27.16 | 28.15 | 28.02 | | | | | | |
| | 0.63 | 37.98 | 60.72 | 30.12 | 25.26 | 39.17 | 29.01 | 29.83 | 34.64 | | | | |
| | 0.83 | 39.13 | 51.65 | 34.22 | 24.57 | 32.58 | 24.18 | 35.83 | 38.62 | 35.83 | 38.62 | | |
| | 2.00 | 37.36 | 59.42 | 41.82 | 25.86 | 36.38 | 28.45 | 44.69 | 33.59 | 39.15 | 33.69 | 40.31 | 45.91 |
| $\rho$=0.4 | 0.03 | 49.00 | 59.43 | | | | | | | | | | |
| | 0.23 | 19.58 | 51.45 | 21.52 | 28.74 | | | | | | | | |
| | 0.43 | 31.82 | 51.31 | 25.31 | 17.44 | 28.75 | 29.23 | | | | | | |
| | 0.63 | 30.64 | 52.44 | 33.74 | 26.05 | 38.37 | 29.87 | 34.33 | 30.15 | | | | |
| | 0.83 | 36.65 | 60.92 | 36.65 | 33.05 | 35.59 | 34.44 | 35.00 | 32.84 | 35.90 | 29.05 | | |
| | 2.00 | 44.35 | 52.36 | 40.76 | 23.80 | 36.33 | 28.10 | 36.99 | 35.37 | 42.25 | 35.62 | 42.59 | 41.59 |
| $\rho$=0.6 | 0.03 | 50.40 | 58.39 | | | | | | | | | | |
| | 0.23 | 27.28 | 61.82 | 27.58 | 25.43 | | | | | | | | |
| | 0.43 | 26.04 | 58.64 | 28.55 | 24.94 | 29.90 | 28.71 | | | | | | |
| | 0.63 | 33.92 | 54.13 | 32.01 | 28.89 | 31.46 | 28.70 | 34.43 | 34.82 | | | | |
| | 0.83 | 35.04 | 53.84 | 34.15 | 24.05 | 40.92 | 30.11 | 33.54 | 27.24 | 31.04 | 30.51 | | |
| | 2.00 | 43.56 | 51.62 | 46.20 | 28.40 | 45.20 | 25.89 | 45.14 | 34.65 | 46.66 | 39.48 | 45.15 | 43.95 |
| $\rho$=0.8 | 0.03 | 61.09 | 63.01 | | | | | | | | | | |
| | 0.23 | - | - | 27.61 | 25.97 | | | | | | | | |
| | 0.43 | - | - | 39.23 | 32.10 | 34.39 | 35.12 | | | | | | |
| | 0.63 | - | - | 32.43 | 26.45 | 36.45 | 33.78 | 32.48 | 37.59 | | | | |
| | 0.83 | - | - | 39.46 | 24.32 | 32.23 | 32.20 | 34.18 | 36.41 | 39.13 | 38.65 | | |
| | 2.00 | - | - | 42.03 | 28.91 | 44.68 | 32.01 | 38.23 | 36.07 | 41.07 | 37.48 | 43.05 | 43.57 |
| $\rho$=1.0 | 0.03 | 58.77 | 58.77 | | | | | | | | | | |
| | 0.23 | - | - | 23.22 | 23.22 | | | | | | | | |
| | 0.43 | - | - | - | - | 24.31 | 24.31 | | | | | | |
| | 0.63 | - | - | - | - | - | - | 36.42 | 36.42 | | | | |
| | 0.83 | - | - | - | - | - | - | - | - | 37.75 | 37.75 | | |
| | 2.00 | - | - | - | - | - | - | - | - | - | - | 43.97 | 43.97 |

Table 4.3: The optimal time to switch under different parameters of correlation and uncertainty; the left columns correspond to firm 1; for the simulations, the following values are chosen: Invest = 1, q1 = 1.0, q2 = 1.0, pr1 = 2.8, pr2 = 1, lamda = 1.2, beta = 0.03, dt = 0.1, mju1 =-0.03, mju2 =-0.03, r = 0.01, g = 2

have older or less educated customers with higher technology related switching costs than another one.

Neither the level of correlation between the two demand processes nor the impact of a rivals's change in drift seem to have a significant impact on the investment timing. This might be attributed to the sample size. However, the complexity of the calculations command to restrict the number of simulations.

Regarding the level of uncertainty about the firm-specific demand decrease the simulation results are not as clear-cut as for the drift parameters. From the lowest pane of Table 5 which represents the case of completely symmetric firms, we conclude that there is no monotonous relationship between uncertainty and the switching time but a convex U-shaped one.[10] This result is not easy to explain. For moderate levels of uncertainty - the left part of the U-shaped curve - uncertainty is negatively related to the exercise time of the option. This corresponds to the effect commonly found in the real option literature that firms take advantage of temporal flexibility by choosing to invest only when the outcome is favorable. In most of the models this leads to a delay of the investment since this is how firms can choose between favorable and unfavorable outcomes. In contrast, in our model firms can avoid unfavorable outcomes by investing early. If the demand stays relatively high firms would not switch as before. However, if demand drops quickly firms would switch earlier leading on average to an earlier optimal time to invest. On the other side, if uncertainty is very high this may dominate the deterministic downward trend of the process inducing a later optimal switching to the new technology.

Commenting on the upper panes of Table 5, the U-shaped investment time-variance relationship is supported. No direct impact of the correlation parameter $\rho$ can be observed.

## 4.6 Conclusions

In this chapter, we combine the literature of switching costs and investment under uncertainty. In particular, we consider heterogenous and temporally decreasing switching costs. This case seems to be of great relevance since it appears in association with many applications of the Internet as a new distribution channel. Potentially, the new technology raises switching costs between firms using a different technology but decreases switching costs between firms using the same technology. Furthermore, these technology related switching costs are assumed to be decreasing over time. In the early days of this computer network, there were few people equipped and enabled to make use of it, and hence, make use of services provided by the Internet. The number of Internet users has increased steadily following the usual technology diffusion pattern. Hence, more and more companies decided to

[10]Using the same methodology in a different application, Chi (2000) found a similar relationship between uncertainty and the option values. However, in his case this is caused by the value enhancing effect of different expectations about the uncertainty of two assets.

offer Internet-based services.

Our model helps to understand what are determinants of the optimal time of introducing a new distribution technology. We show that the appropriate time to exercise the option on the new distribution technology depends negatively on the price gap between both markets and the market growth of the demand for new technology services or products. In contrast to many real option models, we find that firms would bring forward or postpone the investment as demand uncertainty increases, depending on the present level of uncertainty. Given intermediate values of uncertainty firms would invest earlier compared to cases of very high and very low uncertainty. And we find that firms with more loyal customers, i.e. firms whose customers remain faithful for a longer period of time, would switch later to the new technology.

Of course, our conclusions hinge on several assumptions which might not always be valid. One of the central assumptions is that firms cannot differentiate between the customers regarding the technology they use. If they could, they would not lose from immediately switching because then they could skim the locked in customers and at the same time take advantage of the growth prospects provided by the new distribution technology. Sometimes, firms try to separate by running their businesses by different subsidiaries. Another critical point is that the customers who are switching first may have different characteristics than the later switching individuals. If they are, for example, better educated and more wealthy, there may be an additional incentive for the firm to switch early.

This work is distinguished from the previous real options literature. In this line of research, it is common to study the optimal time to invest in a risky asset given the investment is at least partially irreversible. In many applications, the demand schedule is just modeled as an unrestricted Geometric Brownian motion. In our model, it is a Geometric Brownian motion with negative drift for the old technology period and deterministically increasing for the new technology period. This is mainly because we focus on the uncertainty of the consumer switching costs and those impact on the decision to invest in the new technology. In contrast to the conventional real option models we find a non-monotonous relationship between uncertainty and the optimal to exercise the option.

Our model also differs from the switching cost models such as described by Klemperer (1995) and others. Previous work was based on models of strategic interaction in which time was represented by stages of a game.[11] This is appropriate if questions are addressed such as what is the level of prices compared to the situation without costs of switching and how can entry be deterred by strategic pricing. However, if it is asked for the optimal time of an action a continuous time approach may be favorable. Moreover, it is reasonable to account for demand uncertainty since we know from real option theory we learned that this may affect the investment time. In order to benefit from the merits of both lines of research (real option theory and

[11] One exception to this is the paper by Farrell and Shapiro (1988).

the switching cost literature), one should amalgamate those models, if possible.

However, due to problems of analytical tractability, it is difficult to amalgamate models of oligopolistic competition, uncertainty, and dynamic switching costs. The chapter represents a first attempt to accomplish such an amalgamation. Clearly, it could be useful to extend and modify it in different respects. For example, it would be interesting to introduce strategic interaction between firms. A first step would be to introduce first-mover advantages in the duopoly by allowing the first-mover to benefit from the migrating customers of the late-mover. This would introduce the possibility of preemption. More complicated may be the introduction of models of oligopolistic competition à la Cournot, for example. Furthermore, we may gain new insights by considering other sources of uncertainty such as demand uncertainty after introducing the new technology or uncertainty about the project costs (e.g. Schwartz and Zozaya-Gorostiza 2000). It also remains for further research to investigate empirically if the predictions of our model concerning the investment timing are valid. The retail banking industry is possibly a good example since banks have chosen very different time points for the introduction of their Internet facilities.

## 4.7 Appendix B

**Proof of (5).** Solving the first derivative of the profit function $\frac{d\Pi_i(T_i)}{dT_i} = -\frac{q_i^{m1}(0)e^{-(\alpha_i+r)T_i}}{(r-\beta)}\left[g(0)p^{m2}e^{\beta T_i}\left(r+\alpha_i-\beta\right)-p^{m1}\left(r-\beta\right)\right]$ with respect to the time of switching leads to $T_i' = \frac{1}{\alpha_i}\ln\frac{p^{m1}(r-\beta)}{gp^{m2}(r+\alpha_i-\beta)}$. For this value the second derivative of the profit function

$$\frac{d^2\Pi_i(T_i)}{dT_i^2} = \frac{q_i^{m1}(0)e^{-(\alpha_i+r)T_i}}{(r-\beta)}\left[g(0)p^{m2}e^{\beta T_i}\left(r+\alpha_i-\beta\right)^2-p^{m1}\left(r+\alpha_i\right)\left(r-\beta\right)\right]$$

becomes negative: $-\beta p^{m1}q_i^{m1}(0)\left(\frac{p^{m1}(r-\beta)}{(g(0)p^{m2}(r+\alpha_i-\beta))}\right)^{\frac{r-\alpha_i}{\beta}}$. With the given constraints for $\alpha_i$ and $\beta$ it is clear that if the profit function is concave at $T_i'$ it is also concave for values smaller than $T_i'$. However, for some $T_i$ large enough $\frac{d^2\Pi_i(T_i)}{dT_i^2}$ turns positive. Nevertheless, if $T_i' > 0$ then there could never be a profit maximum to the right of $T_i'$ because then the first derivative of the profit function is clearly negative. Consequently, if $T_i' < 0$ the profit maximum corresponds to the corner solution $T_i^* = 0$. ■

## 4.8 Appendix C

The following Mathematica 4.0 program is a simulation procedure which calculates the optimal time to invest in the technology for 100 different random paths of demand . Each path consists of 1000 consecutive values. Based on these results, we calculate the average number of the optimal time to invest.

%First, the probabilities for the discrete jumps p1,..,p5 and the size of these jumps v1 and v2 are computed according to Section 4.5.

```
p1=0.25*(1/lamda^2+(dt^0.5)/lamda*(mju1/sig1+mju2/sig2)+rho/lamda^2)
p2=0.25*(1/lamda^2+(dt^0.5)/lamda*(mju1/sig1-mju2/sig2)-rho/lamda^2)
p3=0.25*(1/lamda^2+(dt^0.5)/lamda*(-mju1/sig1-mju2/sig2)+rho/lamda^2)
p4=0.25*(1/lamda^2+(dt^0.5)/lamda*(-mju1/sig1+mju2/sig2)-rho/lamda^2)
p5=1-1/lamda^2

v1=lamda*sig1*dt^0.5
v2=lamda*sig2*dt^0.5

< <Statistics‘MultiDiscreteDistributions‘
```

% In the next two loops the lists l3 and l4 are produced which consist of 100 random schedules of (logarithmic) demand. For convenience, these numbers are already discounted by r.

```
j=0;l3={};l4={};
While[j<100,i=0;f1=Log[q1];f2=Log[q2];l1={};l2={};
    While[i<1000,
        rd=(Random[MultinomialDistribution[1,{p1,p2,p3,p4,p5}]]);
        f1=f1+rd.{v1,v1,-v1,-v1,0};
        f2=f2+rd.{v2,-v2,v2,-v2,0};
        l1=Append[l1,E ^{f1-i*dt*r}];
        l2=Append[l2,E ^{f2-i*dt*r}];
        i++];
    l3=Append[l3,l1];
    l4=Append[l4,l2];
    j++]
```

% Variables V1 and V2 correspond to the value function (6). q1T1 and q2T2 are the demand on market m1 at time T1 and T2, respectively. q1t and q2t are the accumulated demand on market m1 after time T1 and T2, respectively.

```
V1=pr2*Integrate[g*q1T1*E^((beta-r)*t),{t,T1,T}]-pr1*q1t-Invest
V2=pr2*Integrate[g*q2T2*E^((beta-r)*t),{t,T2,T}]-pr1*q2t-Invest
```

% Here, the lists lv1 and lv2 of values V1 and V2 are produced corresponding to the stochastic demand schedules.

```
lv1={};j=0;T=100;tmpl={};
While[j<100,tmpl=Part[l3,j+1];i=0;hl={};
    While[i<1000,T1=j*dt;
        q1T1=First[tmpl];
        q1t=Apply[Plus,tmpl];
        tmpl=Delete[tmpl,1];
        hl=Append[hl,V1];
        i++];
    lv1=Append[lv1,hl];
    j++]

lv2={};j=0;T=100;tmpl={};
While[j<100,tmpl=Part[l4,j+1];i=0;hl={};
    While[i<1000,T2=j*dt;
        q2T2=First[tmpl];
        q2t=Apply[Plus,tmpl];
        tmpl=Delete[tmpl,1];
        hl=Append[hl,V2];
        i++];
    lv2=Append[lv2,hl];
    j++]
```

% Based on the lists lv1 and lv2 we calculate for each case the optimal time to invest applying the principle of dynamic programming. The lists timel1 and timel2 contain the optimal time points.

```
timel1={};j=0;tmpl={};T=100;
While[i<100,tmpl=Part[lv1,j+1];i=0;vopt=-1000000;
    While[i<1000;curr=Last[tmpl];
        If[curr≥vopt,vopt=curr;top1=T-i*dt,vopt=vopt];
        tmpl=Delete[tmpl,1000-i];
        i++];
    timel1=Append[timel1,top1];
    j++]

timel2={};j=0;tmpl={};T=100;
While[i<100,tmpl=Part[lv2,j+1];i=0;vopt=-1000000;
    While[i<1000;curr=Last[tmpl];
        If[curr≥vopt,vopt=curr;top2=T-i*dt,vopt=vopt];
        tmpl=Delete[tmpl,1000-i];
        i++];
```

```
timel2=Append[timel2,top2];
j++]
```

# Chapter 5

# The relative importance of group-level effects on the performance of German companies

We examine the impact of performance groups on the estimation of the relative importance of firm, industry and other effects on corporate performance. Performance groups comprise firms from the same industry with a similar performance over a longer period of time. We present a statistical method which improves the procedure of variance decomposition by allowing firm effects and the interacting effects of firms and time to be unified into the group effects. Applied to a German data set of 219 companies observed over a period of eleven years (1987-1997) it appears that the majority of the firms can be ascribed to performance groups. The variance proportion of the group effects is about one half of the non-grouped firm effects. They explain about 17.9 percent of the total variance of the returns.

## 5.1 Introduction

This chapter builds upon the recent literature on the relative importance of firm, industry and other effects on corporate performance. The debate which provides the motivation for this line of research was initiated by the paper of Schmalensee (1985). His study is descriptive and does not aim at discriminating between theories. Rather, it addresses the question of which paradigm is potentially the most fruitful to deliver a consistent theory of corporate performance. Mainly, the prevalent paradigms of the following disciplines were considered to be competing in this arena: the traditional Industrial Organization (IO), the modern theoretical Industrial Organization, and the Strategic Management. The traditional IO emphasizes the structural characteristics of industries like growth and the degree of concentration as determinants of corporate performance. The modern theoretical IO, however, focuses on market shares and other partially firm-related concepts while the Strategic Management scholars consider the specific resources of a firm to be important performance predictors.

In order to assess the relative importance of the several research agendas it was analyzed how much of the variance of firm performance could be attributed to industry effects, to the market share, and to firm specific effects.[1] Using a cross-section of the 1975 FTC LB data on the manufacturing sector Schmalensee showed that with about 20 percent of the total variance of the firm performance the industry effects explained more than firm and market share effects. This was interpreted as supporting the view that industry is important but not the only influence on corporate performance.

The article provoked much criticism towards the choice of the data set and the estimation procedure. Rumelt (1991) extended Schmalensee's sample for another three years of data (1974-1977) which allowed him to distinguish not only between industry and corporate-parent effects but also between the error term and the effects of the business units. Further, he allowed business units to enter his data set which did not meet the size criterion of Schmalensee. In this case, the effects of the lines of businesses explaining about 45 percent of the firm performance variance dominated clearly the transient and intransient industry effects which accounted for about nine to sixteen percent. The corporate parent effects were shown to be small in comparison.

Subsequent studies were based on much larger data sets, more sophisticated estimation methods, and a greater variety of performance measures (Hansen and Wernerfelt 1989, Roquebert, Phillips, and Westfall 1996, McGahan and Porter 1997, McGahan 1999, Brush, Bromiley, and Hendrickx 1999, Bunke, Droge, and Schwalbach 2000). If we take stock of the estimation results of those studies the following stylized facts can be put forward: Industry matters, but to a much larger extent business units do. Temporal effects are consistently found to be rather small. With

[1] If data on business units are available, the firm effects are divided into corporate parent and business unit effects.

respect to corporate parent effects the results are equivocal. Another approach was used by Cubbin and Geroski (1987) who considered the dynamics of firm's profitability. They found a large degree of heterogeneity within the industries in the sense that about fifty percent of the companies profitability changes could not be attributed to industry-wide dynamic factors.

Schmalensee and his successors may be criticized for assessing two rather extreme views on economic firm behavior for their ability to predict profit rates: One which ignores firm-specific sources of profit variation versus a view which neglects performance determinants on the industry level. Intermediate concepts introducing some homogeneity between firms and some heterogeneity within an industry, respectively, were not considered.

The most popular concept in IO and Strategic Management of this kind is the concept of strategic groups. Strategic groups are defined as collections of firms whose performance is influenced by group characteristics after controlling for firm and industry effects (Dranove, Peteraf, and Shanley 1998 ).[2] In the literature, the relationship between strategic group membership and firm performance is a central issue. From the twenty papers surveyed by Thomas and Venkatraman (1988) fifteen are committed to the analysis of this relationship. Thus, one might be well advised to discuss the strategic group concept within the debate on the relative importance of firm and industry effects on firm performance.

Considering group effects could make a big difference for the estimation of the relative importance of the various effects. Assuming that the relationship exists between strategic group membership and performance at least for some industries, it might turn out that industry effects disappear when allowing for group effects because high performing industries consist of some well protected high performing groups. On the other side, it might be the case that within industries heterogeneity is attributable to some differently performing groups and thus, firm effects are upward biased if groups are neglected.

Up to now it is still not clear which relative impact strategic groups have on the performance of a firm. The literature on the performance effects of strategic groups cannot provide an answer to this question. Those studies are inherently limited in their ability to deliver such a decomposition of the profit rate variance because to accomplish that large scale data sets are necessary.[3]

The present chapter aims at filling this research gap by extending the study of Bunke, Droge, and Schwalbach (2001) who neglected group phenomena. However,

[2]Previous studies on strategic groups (Hunt 1972, Porter 1979) focussed on clusterings along the relevant strategic dimensions of an industry. However, Dranove, Peteraf, and Shanley (1998) argue that by doing so one is not able to distinguish between firm and group effects of performance.

[3]The majority of the empirical studies testing theories of strategic groups are industry specific, i.e. particular industries were considered where background knowledge was used to extract the relevant strategic dimensions of the industry. Then clusters of firms were detected by different methods. Finally, it was tested if the average profitability differed significantly between or within groups (for surveys see Barney and Hoskisson 1990, Thomas and Venkatraman 1988, McGee and Thomas 1986).

because of several reasons which are discussed in the following section we do not assess the relative importance of strategic groups directly. Instead, we use the concept of "performance groups" introduced by Wiggins and Ruefli (1995). Besides strategic groups, they comprise every clustering of equally performing firms within an industry. The data set consists of 219 German companies observed over a period of eleven years (1987-1997). The companies cover a wide range of industries from non-financial sectors. The performance measure is returns on sales. Applying a variety of methods, we find that a large fraction of the firms have a similar performance after controlling for other effects. Hence, they can be grouped into performance groups. This supports the hypotheses that there may be concepts between the level of firms and the level of the industry which are able to explain a considerable share of the firm performance.

The study proceeds as follows. The next section gives a short overview over the literature on strategic groups and introduces the concept of performance groups. Then, we present the statistical method in Section 5.3. The data set and the empirical results are reported in Section 5.4 before the chapter is concluded.

## 5.2 Performance groups

In the literature, there are several explanations for the appearance of strategic groups. The first one refers to the existence of mobility barriers within an industry (Caves and Porter 1977). This concept is an extension of the entry barriers of Bain (1956). Essentially, mobility barriers prevent firms from freely changing group membership, thus protecting higher performing groups from potential competition. The second, more recent concept, focuses on the perceptions of the managers whose cognition tend to simplify industries by mapping it into groups of firms (Porac and Baden-Fuller 1989, Fombrun and Zajac 1987, Reger and Huff 1993, Lant and Baum 1995, Hodgkinson 1997, Osborne, Stubbart, and Ramaprasad 2001).[4] A third concept combines the level of strategic interaction between firms with the mobility barriers concept (Cool and Dierickx 1993, Peteraf 1993, Dranove, Peteraf, and Shanley 1998).

From models of the first stream of research it can directly be deduced that performance consequences exist for group membership (Porter 1980). In the latter class of models, group membership may have consequences on intermediate outcomes such as the level of rivalry (Smith, Grimm, and Wally 1997, Cool and Dierickx 1993), the group's reputation (Ferguson, Deephouse, and Ferguson 2000), and the groups's identity (Peteraf and Shanley 1997) which may result in performance differentials between groups.

However, there is only weak empirical evidence suggesting the group membership-performance relationship holds (Barney and Hoskisson 1990, Thomas and Venkatraman 1988, Cool and Dierickx 1993). Furthermore, Barney and Hoskisson conjec-

[4]Bogner and Thomas (1993) integrate both views into one theoretical framework.

tured that due to methodological problems both basic assertions of strategic group research are untested: first, that strategic groups exist, and second, that strategic group membership has performance implications. This made obvious the critical state of strategic group research. The main points of concern are the lack of theory about when strategic groups will emerge, the application of clustering algorithms which virtually always produce groups of firms whatever data set is explored, and the misleading interpretation of performance differences between groups as a result of mobility barriers (Barney and Hoskisson 1990). Hence, if studies detect group effects of performance it remains unclear if those effects are due to the existence of mobility barriers between strategic groups, cognitive mapping mechanisms, or if they are just statistical artifacts.[5]

Wiggins and Ruefli (1995, p. 1636) circumvent these problems by not selecting strategic groups *ex ante* as usual but by referring to "performance groups" which are defined as "set[s] of firms whose performance levels are statistically indistinguishable from those of other firms in the group but are distinguishable from the performance levels of firms in other performance groups." If performance groups are detected in a short period, then it is tested if those groups are stable over a longer time period. This reflects the idea that group membership requires temporal stability. When stable performance groups are discovered, further research is required to explore the reasons for their existence. Of course, performance groups and strategic groups are not necessarily congruent. However, the existence of performance groups is a necessary condition for the existence of differently performing strategic groups.[6] If performance groups are not detected, this can be interpreted as a case against the existence of strategic groups. Indeed, Wiggins and Ruefli did not find evidence for the existence of strategic groups in their analysis of five industries.

We follow Wiggins and Ruefli by considering performance groups instead of strategic groups in our analysis. Our statistical procedure is different. In search for performance groups, Wiggins and Ruefli discriminate between firm clusters of different profit distributions by applying iteratively Kolmogorov-Smirnov tests. The drawback of this procedure is that nothing can be said about its reliability. The error of the grouping is the result from the errors made at each single step of the procedure. Hence, it is not clear how large the error is that was made during the whole procedure, since the significance level ($\alpha = .05$) used for the tests at each step can only be regarded as some tuning parameter. Significance tests testing the equality of average returns also seem to be inadequate statistical methods from the point of view that firms forming a performance group should not necessarily have exactly identical average performances, the performance differences having to be small in some not well-defined sense. Furthermore, the considered tests are non-parametric in nature but rely on rather few observations for each firm. Moreover,

[5]Nonetheless, there have been efforts to develop cluster analysis further as a reliable method in Strategic Management (Ketchen and Shook 1996).

[6]Note that we concentrate on horizontal groups of firms from the same industry. Vertical interindustry groups are neglected in this study.

these observations, although obtained for consecutive years, are treated as independent replications, which does not seem to be always reasonable. Finally, as in any procedure based on iterative testing it remains unclear how the obtained grouping takes into account the objective of the analysis. Confirming the resulting groups by some discriminant function analysis may not be sufficient to completely overcome these drawbacks. In contrast, our method considers simultaneously (almost) every possible grouping of firms within each industry. Then, the grouping of firms with similar performance is chosen which is optimal with respect to some criterion presented later. This criterion reflects clearly the objective of the analysis and, hence, the procedure provides a result which is optimal (in a certain way). The next section describes our method in more detail.

## 5.3 Methods

As performance measure we observe the returns on sales $r_{ikt}$ of firm $k$'s activity in industry $i$ at time $t$. Throughout this chapter we will assume that the returns are uncorrelated and have variances which may depend on the firm and industry, but not on time. Hence, the expectation and variances of the returns may be denoted by

$$E(r_{ikt}) = \mu_{ikt} \quad \text{and} \quad Var(r_{ikt}) = \sigma^2_{ik} \ , \tag{5.1}$$

respectively, with $i = 1, \ldots, m;\ \ k = 1, \ldots, n_i;\ \ t = 1, \ldots, T;\ \ \sum_i n_i = n$ and $N = nT$.

If the impact of certain effects on the performance is investigated, one uses typically analysis of variance models which decompose the expected returns as follows:

$$\mu_{ikt} = \mu + \alpha_i + \phi_{ik} + \gamma_t + \delta_{it} + \nu_{ikt} \ , \tag{5.2}$$

where $\mu$ is the average return of firms of all industries over the whole time period, the terms $\alpha_i$, $\phi_{ik}$ and $\gamma_t$ denote the effects of industry $i$, of firm $k$ within industry $i$ and of year $t$, respectively, and $\delta_{it}$ and $\nu_{ikt}$ represent time-dependent effects, i.e., the interaction between industry $i$ and year $t$ as well as the interaction between firm $k$ and year $t$. The identification of the parameters in (5.2) requires certain parameter constraints such as

$$\begin{aligned} \sum_i w_{i.}\alpha_i &= \sum_k w_{ik}\phi_{ik} = \gamma_. = \sum_i w_{i.}\delta_{it} = \delta_{i.} = \nu_{ik.} = \sum_k w_{.k}\nu_{ikt} = 0 \\ i &= 1, \ldots, m;\ \ k = 1, \ldots, n_i;\ \ t = 1, \ldots, T \ , \end{aligned} \tag{5.3}$$

where $w_{ik} \geq 0$ are some time-independent weights. Taking $w_{ik} \equiv 1$ in (5.3) provides the "usual" parameter constraints. Here, and in the remaining part of the paper, we use the usual ANOVA notation. That is, if a suffix is replaced by a dot, variables are summed over the values of that suffix, e.g. $\gamma_. = \sum\limits_i \gamma_i$ . The average over the values of a suffix is denoted by an additional upper bar, e.g. $\bar{\gamma}_. = \frac{1}{T}\sum\limits_i \gamma_i$ .

Without any additional model assumptions, (5.2) describes a saturated model. Most analyses are based on smaller models, that is, on models with fewer "effective" parameters than the number of observations $N$. Such models may be obtained by deleting some effects, compare Bunke, Droge, and Schwalbach (2001) where, for example, a model without interactions $\nu_{ikt}$ was considered:

$$\mu_{ikt} = \mu + \alpha_i + \phi_{ik} + \gamma_t + \delta_{it} \quad . \tag{5.4}$$

However, the analyses differ not only in their assumptions on the expected returns. There are also several models for the variances $\sigma_{ik}^2$ in (5.1) imaginable. Very popular is, for example, the homogeneous variance model assuming a common variance for the returns of all firms over the whole time period:

$$\sigma_{ikt}^2 \equiv \sigma^2 \quad . \tag{5.5}$$

Alternatively, one could use a homogeneous variance model within each industry allowing the returns' variances to depend on the industries, but not on the specific firm:

$$\sigma_{ikt}^2 = \sigma_i^2 \quad . \tag{5.6}$$

The objective of our analysis is to find groups of firms within each industry with a similar performance over the whole time period. Let $M_i = \{1, \ldots, n_i\}$ denote the firms of industry $i$. Then a grouping of firms within the industries may be described by a partitioning of $M_i$ into $g_i$ disjoint groups $M_{il}$ $(l = 1, \ldots, g_i \ , \ 1 \leq g_i \leq n_i)$:

$$M_i = M_{i1} \cup \ldots \cup M_{ig_i} \quad , \quad i = 1, \ldots, m \quad . \tag{5.7}$$

The ideal assumption of identical performance of the firms within the groups leads thus to a submodel of (5.1) by setting

$$\mu_{ikt} = \mu_{ik't} \quad \text{if} \quad k, k' \in M_{il} \quad (l = 1, \ldots, g_i; i = 1, \ldots, m) \quad . \tag{5.8}$$

The return of a firm would then be predicted by a weighted average of the returns of all firms which belong to the same group:

$$\hat{\mu}_{ikt}^g = \sum_{j \in M_{il}} \frac{w_{ij}}{\sum_{j \in M_{il}} w_{ij}} r_{ijt} \quad \text{for all } \ k \in M_{il} \quad . \tag{5.9}$$

Note that (5.8) may also be described as submodel of (5.2) by

$$\phi_{ik} = \phi_{ik'} \ \text{ and } \ \nu_{ikt} = \nu_{ik't} \quad \text{if} \quad k, k' \in M_{il} \quad (l = 1, \ldots, g_i; i = 1, \ldots, m) \quad . \tag{5.10}$$

Now, if we more realistically assume that the firms belonging to the same group have similar but not necessarily identical average returns, we still may use the weighted group mean (9) as a prediction and be more accurate than using the returns as predictions of the expected returns. The latter predictions correspond to the trivial partition into groups each containing a single firm. It seems sensible to form performance groups, grouping firms together in such a way that the prediction of the expected return using the model determined by the grouping (that is using (9)) is as accurate as possible. Equivalently, the estimation of the dependence of the return on the industry, the firm and time will be as accurate as possible. The performance of a model such as a grouping of the firms may therefore be assessed by the weighted mean squared error of prediction (MSEP)

$$R_g = \sum_{i,k,t} \frac{w_{ik}}{T w_{..}} E(\hat{\mu}^g_{ikt} - \mu_{ikt})^2 + \sum_{i,k} \frac{w_{ik}}{w_{..}} \sigma^2_{ik} \ , \tag{5.11}$$

where $\hat{\mu}^g_{ikt}$ denotes the estimate of the expected returns under the model associated with the grouping $g$, cp. (5.9). Assuming a specific model for the variances $\sigma^2_{ik}$, we would always use the inverse of these variances as weights, i.e.:

$$w_{ik} = \sigma^{-2}_{ik} \ , \tag{5.12}$$

since this leads to generalized least squares estimators of the expected returns, which possess certain optimality properties.

To select an appropriate grouping of the firms, one would ideally try to minimize the MSEP over all possible groupings. Unfortunately, this is impossible since the MSEP depends on the unknown expected returns and variances. Therefore one resorts in practice to data-driven methods such as minimizing some convenient estimator of the MSEP. Now, for given variances $\sigma^2_{ik}$ and a given grouping $g$, an unbiased "estimate" of the MSEP (5.11) could be calculated which depends, however, on the unknown variances. Replacing in this formula the variances by some estimates $\tilde{\sigma}^2_{ik}$ based on the assumed variance model, we finally obtain the following criterion for comparing the competing groupings of firms:

$$\hat{R}_g = RSS^w_g + \frac{2\sum_i g_i}{\sum_{i,k} \tilde{\sigma}^{-2}_{ik}} \ , \tag{5.13}$$

where

$$RSS^w_g = \sum_{i,k,t} \frac{w_{ik}}{T w_{..}} (r_{ikt} - \hat{\mu}^g_{ikt})^2$$

denotes the weighted residual sum of squares for the grouping $g$ using the weights $w_{ik}/Tw_{..} = \tilde{\sigma}^{-2}_{ik}/(T\sum_{i,k}\tilde{\sigma}^{-2}_{ik})$. Note that we have also to replace the unknown weights in (5.9) in the same way to arrive at "reasonable" estimates (weighted least squares estimates, WLSE) of the expected returns. To reduce bias effects due to an inadequate modelling of the expected returns, we will use the following model-independent variance estimates:

$$\hat{\sigma}^2_{ik} = \frac{1}{2(T-1)} \sum_{t=2}^{T} (r_{ikt} - r_{ik,t-1})^2 \ . \tag{5.14}$$

Consequently, the variance estimates under the submodels (5.5) and (5.6) are given by

$$\tilde{\sigma}^2_{ik} = \frac{1}{n}\sum_{i,k}\hat{\sigma}^2_{ik} := \hat{\sigma}^2 \quad \text{and} \quad \tilde{\sigma}^2_{ik} = \frac{1}{n_i}\sum_{k}\hat{\sigma}^2_{ik} := \hat{\sigma}^2_i \ , \ \text{respectively.} \qquad (5.15)$$

## 5.4 Data and empirical results

### 5.4.1 Data set and exploratory data analysis

The empirical analysis is based on a panel data set of German companies provided by the Kienbaum Consultants International GmbH. The sample consists of $n = 219$ firms and covers a wide range of $m = 26$ industries from non-financial sectors. Originally, it includes more than 1700 large companies. However, eliminating the firms whose profits or sales are not observed over the whole period of time and excluding all the financial companies reduces the set enormously. No information is available about whether the companies are diversified or not. Each company is assigned to a single industry. The performance measure is returns on sales, defined as the ratio of accounting profits to sales. For each firm, the returns on sales are available over a period of $T = 11$ years (1987-1997). The distribution of the firms over the industries is shown in Table 5.1.

Like any statistical method, our procedure depends on certain assumptions. Therefore we carried out some exploratory data analysis to reveal the features of the data set under study. In particular, we tried to answer the following questions: Do the data contain errors or outliers? Since the data were collected over time, is there any evidence of serial correlation? Do the data have a nearly constant variance? Can the analysis be improved by some convenient data transformation?

A first impression of the data is provided by the box plots of the returns for all firms. A detailed inspection of extreme or outlying values led to the conclusion that the raw data set contained errors probably introduced at the point of data collection. In most cases it was not possible to reconstruct the correct values, so that we eventually omitted the data of 18 firms from the original data set of 237 firms.

The distribution of the observational errors $\varepsilon_{ikt} = r_{ikt} - \mu_{ikt}$ is roughly described by the distribution of the residuals $\hat{\varepsilon}_{ikt} = r_{ikt} - \hat{\mu}_{ikt}$, where $\hat{\mu}_{ikt}$ is obtained by fitting some model for the expected returns $\mu_{ikt}$. But the residuals are not independent nor do they have constant variance, even if both conditions are fulfilled by the errors.

To examine possible serial correlations, or dependencies, we tested for each firm $ik$, whether the autocorrelation function $\rho_{ik}$ of the errors $\varepsilon_{ikt}$ at time lag 1 vanishes. The tests are based on estimates of the autocorrelations $\rho_{ik}$ calculated under model (5.4) as well as under the simple model,

$$\mu_{ikt} = \mu_{ik} \ , \qquad (5.16)$$

which considers the returns of a firm over the years as replicated observations. It turned out that, among the 219 firms, only 28 (under (5.4)) and 32 (under (5.16)), respectively, possess coefficients $\rho_{ik}$ which differ significantly from 0 at level $\alpha = 0.05$. Consequently, it seems plausible to assume uncorrelated returns or observational errors.

Box plots as well as plots of standardized residuals under (5.4) against fitted values indicate that the variances of the returns depend on the firms. In case of replicated observations (5.16) heteroscedasticity is easy to detect and there exist simple formal tests such as the Cochran and the Bartlett tests under the assumption of normally distributed errors. Thus, as a formal quantity for checking (5.5), i.e., the homogeneity of the error variances, we use, in analogy to Cochran's test, the statistic

$$G = \max_{i,k} \frac{\hat{\sigma}_{ik}^2}{\sum_{i,k} \hat{\sigma}_{ik}^2} \ , \tag{5.17}$$

where $\hat{\sigma}_{ik}^2$ is given by (5.14), and compare it with the related critical value for Cochran's test. Cochran's test is based on a statistic, $\tilde{G}$ say, which in (5.17) replaces the variance estimates $\hat{\sigma}_{ik}^2$ by $s_{ik}^2 = (T-1)^{-1} \sum_t (r_{ikt} - \bar{r}_{ik\cdot})^2$, and the corresponding critical value is calculated under (5.16) assuming normally distributed errors. Naturally, this critical value is not the correct one when using $G$, since the variance estimates (5.14) are not $\chi^2$-distributed as in the "Cochran"-case of replicated observations; but it turns out to be a reasonable approximation and so the resulting test may serve as exploratory data analysis tool. For our data set, the hypothesis of homogeneous variances was rejected at significance level $\alpha = 0.01$ based on both statistics $G$ and $\tilde{G}$.

Similarly, we have performed tests for checking (5.6), i.e. the variance homogeneity within each industry, using statistics $G_i$ defined as $G$ in (5.17) but with the summation and maximization over $k$ only. At significance level $\alpha = 0.05$ the hypothesis of a homogeneous variance of the firms within an industry was always rejected except for five industries. Under the replication model (5.16), one would use the variance estimates $s_{ik}^2$ instead of $\hat{\sigma}_{ik}^2$, leading to Cochran-statistics $\tilde{G}_i$. On the basis of these statistics, the homogeneity hypotheses would always be rejected except for two industries.

If heteroscedasticity is detected, then ordinary least squares (OLS) methods cannot be used. Points for which the variance is comparatively large should be downweighted when models for the expected returns are fitted to the data. This may be accomplished by using WLSE with weights depending on the variances instead of OLS estimates. However, in general this requires estimation of the variances since these variances will be unknown. Therefore it is *not* clear whether WLS with estimated variances is superior to OLS or not! Nevertheless, our analysis will be based on WLS with weights

$$w_{ik} = \hat{\sigma}_{ik}^{-2} \ , \tag{5.18}$$

since the variances differ significantly such that neither (5.5) nor (5.6) can be assumed. Naturally, improvements of the procedure are imaginable by searching for an appropriate variance model with less than $n$ parameters, which is different from (5.5) and (5.6); but this is beyond the scope of this chapter.

Finally, Box-Cox transformations may be seen as another approach to correct for both nonnormality and heteroscedasticity. We tried the seven (modified) Box-Cox transformations described in Bunke, Droge, and Schwalbach (2001) and found that the identical transformation is optimal for both models (5.4) and (5.16). All investigations in the remaining part of this chapter will therefore deal with the original, untransformed data.

### 5.4.2 Performance groups under heteroscedasticity

As explained in the previous subsection, we allow different variances for the returns of different firms, i.e., we assume (5.1). Consequently, we use the WLSE based on weights (5.18) for estimating the effects and expected returns. The optimal model or performance group, $\hat{g}$ say, is then defined as the minimizer of the criterion $\hat{R}_g$ over all possible groupings $g$ of firms, where $\hat{R}_g$ is defined by (5.13) with $\tilde{\sigma}^2_{ik} = \hat{\sigma}^2_{ik}$, cp. (5.14).[7] It turns out that the optimal grouping of the 219 firms within the 26 industries consists of 113 groups. The number of groups within the different industries is presented in Table 5.1. Note that 56 of these groups contain only one firm.

Table 5.1: *For each industry, number of firms and number of groups under the optimal model $\hat{g}$.*

| Industry ($i$) | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| No. of firms ($n_i$) | 10 | 5 | 9 | 15 | 7 | 4 | 3 | 5 | 4 | 9 | 10 | 3 | 3 |
| No. of groups under $\hat{g}$ | 5 | 2 | 4 | 9 | 6 | 4 | 2 | 2 | 3 | 3 | 4 | 2 | 3 |
| Industry ($i$) | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 | 24 | 25 | 26 |
| No. of firms ($n_i$) | 12 | 4 | 7 | 16 | 5 | 4 | 3 | 9 | 6 | 5 | 49 | 8 | 4 |
| No. of groups under $\hat{g}$ | 7 | 2 | 3 | 6 | 3 | 4 | 2 | 5 | 4 | 2 | 19 | 4 | 3 |

Table 5.2 summarizes some additional results. It shows also how the weighted variance proportions of some effects is influenced by the optimal grouping. Here, $\tilde{r}_1$, $\tilde{r}_2$, $\tilde{r}_3$, $\tilde{r}_{13}$ and $\tilde{r}_{23}$ denote the empirical weighted variance proportions of the industries, firms, years, industry-year interactions and firm-year interactions, respectively.

[7] Actually, the implemented procedure does not examine all possible partitionings of firms within the industries. Instead, because of numerical feasability, it proceeds stepwise, starting by taking each of the $n$ firms as a group. Then, among all possible pairs of groups within any industry, we join that pair to a single group, which leads to the largest reduction of the estimated risk (5.13). This process is continued until no further decrease of the estimated risk can be achieved and leads to a suboptimal grouping.

Their definition as measure for the impact of the different factors or effects on the performance may be found in Bunke, Droge, and Schwalbach (2001).[8]

Table 5.2: *Some results of optimal grouping under (5.1) using WLS.*

| Model $g$ | Dimension of $g$ | MSEP $10^4 \cdot \hat{R}_g$ | Weighted variance proportions $\tilde{r}_2$ | $\tilde{r}_{23}$ | $\tilde{r}_2 + \tilde{r}_{23}$ |
|---|---|---|---|---|---|
| Saturated model (5.2) | $N = 2409$ | 0.335819 | 0.449305 | 0.117614 | 0.566919 |
| Optimal model $\hat{g}$ | 1243 | 0.258688 | 0.443392 | 0.086713 | 0.530105 |
| Reduction (in %) | 48.40 | 22.97 | 1.32 | 26.27 | 6.49 |

A grouping of firms leads to a replacement of the firm effects and the firm-year interactions by firm group effects and firm group-year interactions, respectively, when modelling the expected returns. Hence, the grouping of firms has only an influence on the variance proportions of the firms and the firm-year interactions. The other empirical weighted variance proportions remain unchanged, and for our data set we obtain:

$$\tilde{r}_1 = 0.396843 \ , \quad \tilde{r}_3 = 0.006797 \ , \quad \tilde{r}_{13} = 0.029440 \ .$$

Similar to Bunke, Droge, and Schwalbach (2001), we could conclude that the industry effects are dominated by the firm effects. This holds for both the permanent effects ($\tilde{r}_1 < \tilde{r}_2$) and when adding the time-dependent effects to the permanent effects ($\tilde{r}_1 + \tilde{r}_{13} < \tilde{r}_2 + \tilde{r}_{23}$), and it remains also true after an optimal grouping of the firms. Despite optimal grouping the percentage of performance variance explained by the permanent and time-dependent firm effects remains nearly unchanged (53.0% instead of 56.7% before the grouping), although the corresponding model dimension is drastically reduced by 48.8 % . Note that about 35.1 % of the performance variance is explained by the 53 "single-firm-groups", whereas the remaining 60 groups with 163 firms in all explain 17.9 % of that variance.

Recall that our procedure for finding performance groups does not rely on the assumption of normally distributed observations. However, some formal tests such as those described in Section 3 for checking variance homogeneity would require such an assumption (at least approximately). To check whether the observational errors are normally distributed, one should use the standardized residuals,

$$e_{ikt} = \frac{\hat{\varepsilon}_{ikt}}{\tilde{\sigma}_{ik}\sqrt{1 - h_{ikt}}} \ . \qquad (5.19)$$

[8]For example, the weighted variance proportion of the industry effects is defined by $\tilde{r}_1 = \tilde{s}_1^2/\tilde{s}^2$, where $\tilde{s}^2 = (Tw_{..})^{-1}\sum_{i,k,t} w_{ik}(r_{ikt} - \hat{\mu})^2$ and $\tilde{s}_1^2 = w_{..}^{-1}\sum_i w_{ik}\hat{\alpha}_i^2$ are the weighted empirical variances of the returns and the industry effects, respectively, and $\hat{\mu}$, $\hat{\alpha}_i$ denote the WLSE of the effects $\mu$, $\alpha_i$.

Here, $h_{ikt}$ denotes the diagonal element $ikt$ of the hat matrix associated with the model under consideration. Several diagnostic plots (plots of standardized residuals against fitted values, normal QQplots and histograms for standardized residuals) as well as estimated skewness (0.005) and kurtosis (5.787) of the standardized residuals after optimal grouping suggest that a normal approximation to the error distribution would work.

### 5.4.3 Firm groups under alternative aims and short summary

Here we consider two additional approaches for the definition of performance groups which correspond to different models for the expected returns. That is, the competing models are no longer given by (5.8). We continue to assume heteroscedastic variances as in (5.1).

First we aim at finding groups of firms within each industry, which show a similar behavior of their returns over the time, but which have possibly different levels of performance, i.e., possibly different averages of returns. For this, we start with model (5.4) and introduce additionally firm-year interactions $\nu_{ikt}$ as in (5.2). If two firms, $(ik)$ and $(ik')$ say, interact with the years in a similar way, then they will enter the same group. Hence, the model for the expected returns will assume the same interactions with the years for both firms, but not the same firm effects! That is, the competing models for the expected returns may be described by the set of all possible partitions (5.7) such that additionally to (5.2) the following constraints hold:

$$\nu_{ikt} = \nu_{ik't} \quad \text{if} \quad k, k' \in M_{il} \quad (l = 1, \ldots, g_i; i = 1, \ldots, m) \ . \tag{5.20}$$

As before, a model selection criterion may be derived as an appropriate estimate of the risk (5.11). With the notations of the previous section and $\hat{\mu}^g_{ikt}$ being the WLSE of the expected returns calculated under the assumption (20), this leads to

$$\tilde{R}_g = RSS^w_g + \frac{2[n + (T-1)\sum_i g_i]}{T \sum_{i,k} \tilde{\sigma}^{-2}_{ik}} \ . \tag{5.21}$$

The minimizer of (5.21) with respect to the possible groupings $g$ will be denoted by $\tilde{g}$. For our data set, the optimal grouping $\tilde{g}$ classifies the 219 firms into 80 groups. It is not surprising that this number is smaller than that of $\hat{g}$, because it is now more likely that firms are considered to behave similarly.

Another aim could be the search for groups of firms with approximately the same time-independent firm effects, neglecting completely the year-firm interactions. That is, the competing models for the expected returns are given by (5.16) and assuming additionally

$$\mu_{ik} = \mu_{ik'} \quad \text{if} \quad k, k' \in M_{il} \quad (l = 1, \ldots, g_i; i = 1, \ldots, m) \ . \tag{5.22}$$

In this case, the competing models (partitions $g$) can be compared by the following criterion:

$$\bar{R}_g = RSS_g^w + \frac{2\sum_i g_i}{T\sum_{i,k}\tilde{\sigma}_{ik}^{-2}} \ , \qquad (5.23)$$

whose minimizer over $g$ will be denoted by $\bar{g}$. Here, $\hat{\mu}_{ikt}^g$ is the WLSE of the expected returns assuming (22). Note that the optimal grouping (model) would remain unchanged by assuming any submodel of (2), which contains at least firm effects $\phi_{ik}$ ($i = 1, \ldots, m$; $k = 1, \ldots, n_i$), and selecting among the partitions (5.7) with

$$\phi_{ik} = \phi_{ik'} \quad \text{if} \quad k, k' \in M_{il} \quad (l = 1, \ldots, g_i; i = 1, \ldots, m) \ . \qquad (5.24)$$

For our data set, the optimal grouping $\bar{g}$ classifies the 219 firms into 117 groups.

Obviously, the obtained optimal groupings depend on the different aims of the analysis. Table 5.3 summarizes some results. For the sake of completeness, it contains also the results of subsection 5.4.2 as well as those for some models for the expected returns such as

$$\mu_{ikt} := \mu + \alpha_i + \phi_{ik} + \gamma_t \ , \qquad (5.25)$$

which have not been considered until now.

Table 5.3: *Estimated risk for some models for the expected returns assuming heteroscedastic variances.*

| Model $g$ | Dimension of $g$ | $10^4\times$ Estimated risk |
|---|---|---|
| (5.2) | 2409 | 0.335819 |
| (5.4) | 479 | 0.339645 |
| (5.25) | 229 | 0.373098 |
| (5.16) | 219 | 0.387474 |
| *Optimal under:* | | |
| (5.2), (5.10); i.e. $\hat{g}$ | 1243 | 0.258688 |
| (5.2), (5.20); i.e. $\tilde{g}$ | 1019 | 0.238697 |
| (5.4), (5.24) | 377 | 0.327375 |
| (5.25), (5.24) | 127 | 0.360828 |
| (5.16), (5.22); i.e. $\bar{g}$ | 117 | 0.375204 |

Recall that all considered models contain the same industry effects, so that they have the following associated empirical weighted variance proportion: $\tilde{r}_1 = 0.396843$. Moreover, the presented estimated risks were always calculated as almost unbiased estimates of the MSEP (5.11) under the condition that the variances are given by (5.14), that is, they were calculated according to (5.13), (5.21), (5.23) or a corresponding formula for other models. We observe that even some optimal models may be outperformed by the saturated model with respect to the estimated risk when this "largest model" doesn't belong to the class of competing models.

The optimal grouping model $\tilde{g}$ under (5.2) and (5.20) with 80 groups of firms appears most convenient for predicting future returns when we compare all candidates considered in Table 5.3. The second choice would be model $\hat{g}$, which was obtained in Subsection 5.4.2 as optimal solution under (5.2) and (5.10). This model provides a grouping of the firms into 113 groups with both a similar time-dependent and a similar permanent behavior of their returns. Note that another grouping could be preferred to the optimal one if its estimated risk is close to the optimal risk $\hat{R}_{\hat{g}}$ and if it provides fewer or easily interpretable groups. This is reasonable, since our procedure is based on *estimates* of the risk. Thus any appealing grouping in our stepwise search, $g^*$ say, fulfilling a rule of thumb like

$$\hat{R}_{g^*} < (1+\delta)\hat{R}_{\hat{g}} \ , \quad \text{with some small } \delta > 0 \text{ such as } 0.1 \ ,$$

could be our first choice. But such an approach is not discussed in more detail, since the economic interpretation of specific groupings is not addressed in this chapter.

Generally, time-dependent industry and firm effects seem to be important for describing the dependence of the returns on some effects. Models such as (5.16) neglect this fact by treating the data observed over time as independent replications and may thus not serve as an appropriate basis for statistical analysis. Naturally, there is some hope to improve the prediction quality of the models by considering some variations such as allowing an additional grouping of the years, which could drastically reduce the dimension of models containing, for example, interactions between industries (and/or firms) and years, without having a substantial effect on the fit.

Finally, one could also try to find optimal groups of firms with industries by a simultaneous selection of (grouping) models for the expected returns and of (again grouping) models for the variances by use of an appropriate criterion such as cross-validation, which can be defined without having some estimates of the variances. But this is beyond the scope of this chapter. The most convenient way of analyzing the data in our setting is probably to assume just the rather general model (5.1) of heteroscedastic variances.

## 5.5 Conclusions

This chapter extends the literature on the relative importance of firm, industry, and other effects on firm performance by examining the effect of performance groups. The concept of performance groups was introduced by Wiggins and Ruefli (1995) to investigate necessary conditions for the existence of strategic groups. Using a variety of methods, we found that, in contrast to Wiggins and Ruefli, performance groups exist in almost every industry of our data set. In particular, we found that the majority of firms can be grouped with respect to a criterion which measures the ability of a grouping and of the corresponding model to predict the returns of a firm. The performance groups explain about 17.9 percent of the performance variance. However, about 35.1 percent of this variance is explained by non-grouped firms. Because of the splitting of the firm effects into group and single-firm effects, now the largest impact is associated with the industry effects of about 39.7 percent. It is worth noting that the grouped model uses much less parameters than the saturated model (about one half) but does not explain much less of the firm effects (53.0 percent versus 56.7 percent).

How can these results be interpreted? Of course, the study is descriptive in nature and thus, no structural causalities can be uncovered. Nonetheless, the results suggest that firm-focussed concepts from Strategic Management and industry-focussed concepts from IO do not tell the whole story about corporate performance. Hence, the respective literature could be fertilized by considering intermediate group-level concepts. However, a note of caution seems to be appropriate as Dranove, Peteraf, and Shanley (1998) point out that real evidence of group effects can only be found if data on group characteristics are available. Possibly, our group effects are spurious in the sense that they just result from some aggregated firm specific characteristics and not from genuine group characteristics.

What other reasons are possible for firms having apparently similar levels of performance within an industry? First, our data set offers a segmentation of firms into industries which might be too coarse. Comparing with segmentations like the SIC-3 and SIC-4 code, our classification covers relatively large bundles of industry segments. On the other side, we only considered groups which lasted for the whole period of time. Peteraf and Shanley (1997) suggest that the periods of group membership vary and that groups may be more important in unstable industries. However, addressing these questions remains for further research.

In particular, it would be interesting to investigate if our results (which are to a certain extent opposing to Wiggins' and Ruefli's) can be reproduced with different sets of data and different measures of performance. Another possible research avenue would be to further elaborate on the constituent characteristics of strategic groups and other possible group concepts which are correlated with the performance groups. The empirical framework of Dranove, Peteraf, and Shanley (1998) is a possible starting point.

# Chapter 6

# Summary and concluding remarks

This study considers sources of firm heterogeneity theoretically and empirically. In the first part of this work we consider the impact of differentiated customers on firm behavior. Customers are either assumed to be differentiated with respect to their taste for different characteristics of a product or with respect to the costs of switching to a different supplier. In the last part of the work, we are concerned with the question of whether there is homogeneity of the profit rates of firms within an industry in the long-run.

One important line of explanation of firm heterogeneity in the economic literature is associated with product differentiation. Within this research area many recent contributions are related to Hotelling-like models of horizontal product differentiation. In Chapter 2, we survey this stream of the literature. The main purpose of this work is to extract from the literature the basic determinants of differentiation. Many scholars suggested that there exists a general "principle of differentiation" (Neven 1985: 322, Tirole 1990: 278). However, this survey shows that the extremes of maximum or minimum differentiation proposed by some economists are rather an exception to the rule.

We find that there is the trade-off between attracting customers by approaching a rival's location in the product space and the resulting increase in price competition. By having a direct impact on the "toughness" of the price competition, the relative size of this trade-off has been shown to depend on a number of model features such as the convexity of the transport costs, the demand elasticity, the concentration of customers in the market center, et cetera. Apart from the direct effect of differentiation via the price competition, there is a direct effect of density of consumers and, of course, when collusion is permitted. The predictive power of these models has still to be shown by empirical and experimental studies.

This survey complements similar work because it focuses on the determinants of horizontal differentiation rather than solely on pricing regimes (Gabszewicz and Thisse 1992), on the whole class of 'address' models (Waterson 1989), or on general modeling issues (Lancaster 1990). Furthermore, it has the advantage of including the large number of recently published papers on this topic.

In Chapter 3, the multi-firm version of the Hotelling model proposed by d'Aspremont, Gabszewicz, and Thisse (1979) is considered. It is a common experience of game theorists that the number of players may influence the equilibrium outcome. Nevertheless, up to now, only the less complex circular case (Economides 1989) and the corresponding case of linear transport costs (Economides 1993) were considered allowing more than two players to compete. Chapter 3 fills this research gap.

The model has two stages of location choice on the unit interval and price choice under quadratic transport costs. We show that the equilibrium of maximum differentiation is destroyed by allowing more than two firms to compete. In particular, we find that given maximum differentiation corner firms benefit from moving marginally towards the market center. This result can be attributed to the market power of the corner firms. In equilibrium, these firms take advantage from possessing a hinterland with customers for whom they can easily compete. The market power not only materializes in inside locations but also in higher-than-rivals' prices. The maximum differentiation behavior in the duopoly is a special case since the superior position of a corner firm is offset by the competition with the equally well positioned rival. However, if there is an inside firm, the corner firms' relative position becomes stronger. Concerning the price structure this result is consistent with the analysis of Economides (1993) for the linear utility case where U-shaped price patterns occur for equidistant (non-equilibrium) configurations. However, regarding the optimal locations our results differ from the interval market with linear transport costs implying (strong) minimum differentiation (Economides 1993) and the circular market with quadratic transport costs implying maximum differentiation (Economides 1989).

In Chapter 4, we consider the determinants of temporal differentiation of firms, i.e. we address the question of why firms would choose a different time for a certain action. In the model, a firm is faced with the question of whether and when to invest in a technology which would make distribution of its product more efficient (in a broad sense) and may provide the opportunity to grow. At the same time, however, it would also make competition more intense. Since this technology is assumed to impose additional (temporally decreasing) switching costs to the customers, firms may be interested in introducing this technology not right away but later in order to skim their attached customers for a certain while. We introduce the view that the investment in such a technology corresponds to a (real) option. We show that the optimal time to exercise this option depends on the prices of both submarkets (divided not by the products but by the distribution technologies) and the growth for new technology services. Uncertainty has a non-monotonic impact on the optimal investment time. Compared to the reference level of intermediate uncertainty, a firm would invest later if uncertainty increases or decreases strongly. Of course, that firm invests later whose customers are more loyal.

This Chapter represents an extension to the existing literature on real options and switching costs. Our model is distinguished from the work on real options by considering explicitly a decreasing stochastic demand schedule on the old technology market, an increasing deterministic demand schedule on the new technology market, and a tougher competitive environment at the new technology market associated with lower prices compared to the old technology market. This work is also related to the switching costs literature (Klemperer 1995). The models in this line of research are usually two-stage or multi-period games of two players. Although our model does not allow for strategic interaction because of analytical tractability it has the merit of a continuous time model. Consequently, it is the more appropriate model for investigating on the optimal time to adopt a new technology.

Chapter 5 studies the empirical side of firm heterogeneity. There exists an interesting literature which aims at assessing the relative importance of the industry, the management, and other factors for explaining the performance of a company. A neglected issue in this literature is if there are intermediate concepts between the level of industries and companies such as the concept of strategic groups which are able to predict a share of a company's returns. Hence, it is ignored if there is homogeneity of the firms' performance below the industry level. We present a new statistical method of variance decomposition which takes possible groups of firms into account having a similar level of performance over a longer period of time. For a longitudinal data set consisting of large German companies, it is shown that indeed, there is a large number of firms which can be assigned to such equally performing groups. This work overcomes several shortcomings of a statistical procedure of detecting performance groups proposed by Wiggins and Ruefli (1995). It complements the literature on the relative importance of firm, industry, and other effects which previously have neglected group-level performance effects.

Investigating the questions of why firms differentiate and to which extent they should be different (which are not necessarily different questions) would make up a huge research program. Our work has contributed to the literature by providing several pieces of a mosaic. Together with complementary research this would present a more complete picture of firm heterogeneity. Theoretical explanations are given for the facts that firms offer different products and choose a different time to switch to a new technology. Empirically, it is shown that within industries there is a certain degree of homogeneity of the firms' profitability.

The work is based on different theoretical and empirical methods. They have proven useful providing ground for the recommendation that further research in this field should be based on a multitude of methods. Notwithstanding the different perspectives of microeconomics, industrial organization, and strategic management it seems that the issues they address are at least overlapping and thus, the suggestion for more interaction between these fields is well grounded.

# Bibliography


Anderson, S. P. (1987), 'Spatial competition and price leadership', *International Journal of Industrial Organization* **5**, 369–398.

Anderson, S. P. (1988), 'Equilibrium existence in the linear model of spatial competition', *Economica* **55**, 479–491.

Anderson, S. P. & de Palma, A. (2000), 'From local to global competition', *European Economic Review* **44**, 423–448.

Anderson, S. P., Goeree, J. K. & Ramer, R. (1997), 'Location, location, location', *Journal of Economic Theory* **77**, 102–127.

Anglin, P. (1992), 'The relationship between models of horizontal and vertical differentiation', *Bulletin of Economic Research* **44**(1), 1–20.

Baake, P. & Oechssler, J. (1997), 'Product differentiation and the intensity of price competition', *Zeitschrift für Wirtschafts- und Sozialwissenschaften - Verein für Socialpolitik* **117**, 247–256.

Bain, J. S. (1956), *Barriers to New Competition*, Harvard University Press, Cambridge, MA.

Balvers, R. & Szerb, L. J. (1996), 'Location in the Hotelling duopoly model with demand uncertainty', *European Economic Review* **40**, 1453–1461.

Banerjee, A. & Summers, L. H. (1987), 'On frequent-flyer programs and other loyalty inducing arrangements', *Harvard University Working Paper* .

Barney, J. B. & Hoskisson, R. E. (1990), 'Strategic groups: Untested assertions and research proposals', *Managerial and Decision Economics* **11**, 187–198.

Barreda, I., García, A., Georgantzís, N., Andaluz, J. & Gil, A. (2000), 'Price competition and product differentiation: Experimental evidence', *LINEEX Working Papers* **11**, 1–41.

Basu, K. & Bell, C. (1991), 'Fragmented duopoly: Theory and applications to backward agriculture', *Journal of Development Economics* **36**, 145–165.

Baum, J. A. C. & Mezias, S. J. (1992), 'Localized competition and organizational failure in the Manhattan hotel industry, 1898-1990', *Administrative Science Quarterly* **37**, 580–604.

Baum, J. A. C. & Singh, J. V. (1994), 'Organizational niche overlap and the dynamics of organizational mortality', *American Journal of Sociology* **100**, 346–380.

Baumol, W., Panzar, J. & Willig, R. (1982), *Contestable Markets and the Theory of Market Structure*, Harcourt Brace Jovanovich, New York.

Beggs, A. (1989), 'A note on switching costs and technology choice', *Journal of Industrial Economics* **37**, 437–439.

Beggs, A. & Klemperer, P. (1992), 'Multiperiod competition with switching costs', *Econometrica* **60**, 651–666.

Bertrand, J. (1883), 'Théorie mathématique de la richesse sociale', *Journal de Savants* pp. 499–508.

Bester, H., de Palma, A., Leininger, W., Thomas, J. & Thadden, E.-L. v. (1996), 'A noncooperative analysis of Hotelling's location game', *Games and Economic Behavior* **12**, 165–186.

Bloch, F. & Manceau, D. (1999), 'Persuasive advertising in Hotelling's model of product differentiation', *International Journal of Industrial Organization* **17**, 557–574.

Böckem, S. (1994), 'A generalized model of horizontal product differentiation', *Journal of Industrial Economics* **42**, 287–298.

Böckem, S. (1996), 'Corrigendum to: A generalized model of horizontal product differentiation', *Mimeo* .

Bogner, W. C. & Thomas, H. (1996), 'A longitudinal study of the competitive positions and entry paths of European firms in the US pharmaceutical market', *Strategic Management Journal* **17**, 85–107.

Brenner, S. (2001), 'Hotelling games with three, four, and more players', *Discussion Paper, SFB 373, Humboldt-University Berlin* **23**, 1–27.

Brown-Kruse, J., Cronshaw, M. B. & Schenk, D. J. (1993), 'Theory and experiments on spatial competition', *Economic Inquiry* **31**, 139–165.

Brown-Kruse, J. & Schenk, D. J. (2000), 'Location, cooperation and communication: An experimental examination', *International Journal of Industrial Organization* **18**, 59–80.

Brush, T. H., Bromiley, P. & Hendrickx, M. (1999), 'The relative influence of industry and corporation on business segment performance: An alternative estimate', *Strategic Management Journal* **20**, 519–547.

Bunke, O., Droge, B. & Schwalbach, J. (2001), 'Die relative Bedeutung von Firmen- und Industriezweigeffekten für den Unternehmenserfolg', *Zeitschrift für Betriebswirtschaft (forthcoming)* .

Camerer, C. F. (1991), 'Does strategy research need game theory?', *Strategic Management Journal* **12**, 137–152.

Campbell-Hunt, C. (2000), 'What have we learned about generic competitive strategy: A meta-analysis', *Strategic Management Journal* **21**, 127–154.

Caplin, A. & Nalebuff, B. (1991), 'Aggregation and imperfect competition: On existence of equilibrium', *Econometrica* **59**, 25–59.

Caves, R. & Porter, M. E. (1977), 'From entry barriers to mobility barriers', *Quarterly Journal of Economics* **91**, 241–261.

Chamberlin, E. H. (1933), *The Theory of Monopolistic Competition*, Harvard University Press, Boston.

Chen, M. & Hambrick, D. C. (1995), 'Spead, stealth, and selective attack: How small firms differ from large firms in competitive behavior', *Academy of Management Journal* **38**, 453–482.

Chi, T. (2000), 'Option to acquire or divest a joint venture', *Strategic Management Journal* **21**, 665–687.

Coase, R. H. (1937), 'The nature of the firm', *Economica* **4**, 386–405.

Cool, K. O. & Dierickx, I. (1993), 'Rivalry, strategic groups, and firm profitability', *Strategic Management Journal* **14**, 47–59.

Cubbin, J. & Geroski, P. (1987), 'The convergence of profits in the long run: Inter-firm and inter-industry comparisons', *Journal of Industrial Economics* **35**, 427–442.

Cubbin, J. & Geroski, P. (1990), The persistence of profits in the United Kingdom, *in* D. C. Mueller, ed., 'The Dynamics of Company Profits', Cambridge University Press, Cambridge, England.

D'Aspremont, C. J., Gabszewicz, J. & Thisse, J.-F. (1979), 'On Hotelling's stability in competition', *Econometrica* **47**, 1145–1150.

de Frutos, M. A., Hamoudi, H. & Jarque, X. (1999), 'Equilibrium existence in the circle model with linear quadratic transport cost', *Regional Science and Urban Economics* **29**, 605–615.

de Palma, A., Ginsburgh, V. & Thisse, J.-F. (1987), 'On existence of location equilibria in the 3-firm Hotelling problem', *Journal of Industrial Economics* **36**, 245–252.

Deephouse, D. L. (1999), 'To be different, or to be the same? It's a question (and theory) of strategic balance', *Strategic Management Journal* **20**(2), 147–166.

Demsetz, H. (1974), Two systems of belief about monopoly, *in* H. J. Goldschmid, H. M. Mann & J. F. Weston, eds, 'Industrial Concentration: The New Learning', Little, Brown, Boston.

Dhawan, R. (2001), 'Firm size and productivity differential: Theory and evidence from a panel of US firms', *Journal of Economic Behaviour and Organization* **44**, 269–293.

DiMaggio, P. J. & Powell, W. W. (1983), 'The iron cage revisited: Institutional isomorphism and collective rationality in organizational fields', *American Sociological Review* **48**, 147–160.

Dixit, A. K. & Pindyck, R. S. (1994), *Investment under Uncertainty*, Princeton University Press, New Jersey.

Dixit, A. K. & Stiglitz, J. E. (1977), 'Monopolistic competition and optimal product diversity', *American Economic Review* **67**, 297–308.

Dooley, R. S., Fowler, D. M. & Miller, A. (1996), 'The benefits of strategic homogeneity and strategic heterogeneity: Theoretical and empirical evidence resolving past differences', *Strategic Management Journal* **17**, 293–305.

Dranove, D., Peteraf, M. & Shanley, M. (1998), 'Do strategic groups exist? An economic framework for analysis', *Strategic Management Journal* **19**, 1029–1044.

Eaton, B. C. & Wooders, M. H. (1985), 'Sophisticated entry in a model of spatial competition', *RAND Journal of Economics* **16**, 282–297.

Economides, N. (1986), 'Minimal and maximal product differentiation in Hotelling's duopoly', *Economics Letters* **21**, 67–71.

Economides, N. (1989), 'Symmetric equilibrium existence and optimality in differentiated product markets', *Journal of Economic Theory* **47**, 178–194.

Economides, N. (1993), 'Hotelling's "Main Street" with more than two competitors', *Journal of Regional Science* **33**(3), 303–319.

Farrell, J. & Shapiro, C. (1988), 'Dynamic competition with switching costs', *RAND Journal of Economics* **19**, 123–137.

Feenstra, R. C. & Levinsohn, J. A. (1995), 'Estimating markups and market conduct with multidimensional product attributes', *Review of Economic Studies* **62**, 19–52.

Ferguson, T. D., Deephouse, D. L. & Ferguson, W. L. (2000), 'Do strategic groups differ in reputation?', *Strategic Management Journal* **21**, 1195–1214.

Fombrun, C. J. & Zajac, E. J. (1987), 'Structural and perceptual influences on intraindustry stratification', *Academy of Management Journal* **30**, 33–50.

Friedman, J. W. (1977), *Oligopoly and the Theory of Games*, North-Holland, Amsterdam.

Friedman, J. W. & Thisse, J.-F. (1993), 'Partial collusion fosters minimum product differentiation', *RAND Journal of Economics* **24**, 631–645.

Fujita, M. & Thisse, J.-F. (1986), 'Spatial competition with a land market: Hotelling and von Thunen unified', *Review of Economic Studies* **53**, 819–841.

Gabszewicz, J. J. & Thisse, J.-F. (1979), 'Price competition, quality and income disparaties', *Journal of Economic Theory* **22**, 340–359.

Gabszewicz, J. J. & Thisse, J.-F. (1986), 'On the nature of competition with differentiated products', *Economic Journal* **96**, 160–172.

Gabszewicz, J. J. & Thisse, J.-F. (1992), Location, *in* R. J. Aumann & S. Hart, eds, 'Handbook of Game Theory', Elsevier Science Publishers, Amsterdam, pp. 282–304.

Gabszewicz, J., Pepall, L. & Thisse, J.-F. (1992), 'Sequential entry with brand loyalty caused by consumer learning-by-using', *Journal of Industrial Economics* **40**, 397–416.

Geroski, P. (2000), 'Models of technology diffusion', *Research Policy* **29**, 603–625.

Ghemawat, P. (1999), *Strategy and the Business Landscape: Text and Cases*, Addison-Wesley, Reading, MA.

Hambrick, D. C. & D'Aveni, R. A. (1992), 'Top management team deterioration as part of the downward spiral of large bankruptcies', *Management Science* **38**, 1445–1466.

Hannan, M. T. & Freeman, J. H. (1989), *Organizational Ecology*, Harvard University Press, Cambridge, MA.

Hansen, G. S. & Wernerfelt, B. (1989), 'Determinants of firm performance: The relative importance of economic and organizational factors', *Strategic Management Journal* **10**, 399–411.

Harter, J. F. R. (1996), 'Hotelling's competition with demand location uncertainty', *International Journal of Industrial Organization* **15**, 327–334.

Hatten, K. & Hatten, M. L. (1987), 'Strategic groups, asymmetrical mobility barriers and contestability', *Strategic Management Journal* **8**(4), 329–342.

Hatten, K. J. & Schendel, D. E. (1977), 'Heterogeneity within an industry: Firm conduct in the U.S. brewing industry, 1952-1971', *Journal of Industrial Economics* **26**, 97–113.

Haveman, H. A. (1993), 'Follow the leader: Mimetic isomorphism and entry into new markets', *Administrative Science Quarterly* **38**, 593–627.

Hehenkamp, B. (2001), 'Sluggish consumers: An evolutionary solution to the Bertrand paradox', *Games and Economic Behavior (forthcoming)* .

Hendel, I. & de Figueiredo, J. N. (1997), 'Product differentiation and endogeneous disutility', *International Journal of Industrial Organization* **16**, 63–79.

Hermalin, B. E. (1994), 'Heterogeneity in organizational form: Why otherwise identical firms choose different incentives for their managers', *RAND Journal of Economics* **25**, 518–537.

Hinloopen, J. & Marrewijk, C. V. (1999), 'On the limits and possibilities of the principle of minimum differentiation', *International Journal of Industrial Organization* **17**, 735–750.

Hodgkinson, G. P. (1997), 'The cognitive analysis of competitive structures: A review and critique', *Human Relations* **50**(6), 625–654.

Hohaus, B., Konrad, K. A. & Thum, M. (1994), 'Too much conformity?: A Hotelling model of local public goods supply', *Economics Letters* **44**(3), 295–299.

Hoover, E. M. (1937), 'Spatial price discrimination', *Review of Economic Studies* **4**, 182–191.

Hotelling, H. (1929), 'Stability in competition', *The Economic Journal* **3**, 41–57.

Huck, S., Knoblauch, V. & Müller, W. (2000), 'On the profitability of collusion in location games', *Discussion Paper, SFB 373, Humboldt-University Berlin* (23), 1–8.

Huck, S., Müller, W. & Vriend, N. (2001), 'The east end, the west end, and king's cross: On clustering in the four-player Hotelling game', *Economic Inquiry (forthcoming)* .

Hunt, M. (1972), Competition in the Major Home Appliance Industry, 1960-1970, PhD thesis, Harvard University.

Irmen, A. & Thisse, J.-F. (1998), 'Competition in multi-characteristics spaces: Hotelling was almost right', *Journal of Economic Theory* **78**, 76–102.

Jehiel, P. (1992), 'Product differentiation and price collusion', *International Journal of Industrial Organization* **10**, 633–643.

Kamrad, B. & Ritchken, P. (1991), 'Multinomial approximating models for options with k state variables', *Management Science* **37**(12), 1640–1653.

Ketchen, D. J. & Shook, C. L. (1996), 'The application of cluster analysis in strategic management research: An analysis and critique', *Strategic Management Journal* **17**, 441–458.

Klemperer, P. (1987a), 'Markets with consumer switching costs', *Quarterly Journal of Economics* **102**, 375–394.

Klemperer, P. (1987b), 'The competitiveness of markets with switching costs', *RAND Journal of Economics* **18**, 138–150.

Klemperer, P. (1995), 'Competition when consumers have switching costs: An overview with applications to industrial organization, macroeconomics, and international trade', *Review of Economic Studies* **62**, 515–539.

Kogut, B. & Kulatilaka, N. (1994), 'Options thinking and platform investment: Investing in opportunity', *California Management Review* **36**, 52–71.

Kulatilaka, N. & Perotti, E. C. (1998), 'Strategic growth options', *Management Science* **44**(8), 1021–1033.

Kulatilaka, N. & Perotti, E. C. (2000), Time-to-market advantage as a Stackelberg growth option, *in* E. Schwartz & L. Trigeorgis, eds, 'Innovation and Strategy: New Developments and Applications in Real Options', Oxford University Press, Oxford, forthcoming.

Lambertini, L. (1994), 'Equilibrium locations in the unconstrained Hotelling game', *Economic Notes* **23**(3), 438–446.

Lancaster, K. (1966), 'A new approach to consumer theory', *Journal of Political Economy* **74**, 132–157.

Lancaster, K. (1971), *Consumer Demand: A New Approach*, Columbia University Press, New York.

Lancaster, K. (1979), *Variety, Equity, and Efficiency*, Columbia University Press, New York.

Lancaster, K. (1990), 'The economics of product variety: A survey', *Marketing Science* **9**(3), 189–206.

Lant, T. K. & Baum, J. A. C. (1995), Cognitive sources of socially constructed competitive groups: Examples from the Manhattan hotel industry, *in* W. R. Scott & S. Christensen, eds, 'The Institutional Construction of Organizations: International and Longitudinal Studies', Sage, Thousand Oaks.

Lewis, A. (1998), Essays on Imperfect Competition and Firm Heterogeneity, PhD thesis, European University Institute, Florence.

MacMillan, I. C. & McGrath, R. M. (1997), 'Discovering new points of differentiation', *Harvard Business Review* **8**, 133–145.

Mai, C.-C. & Peng, S.-K. (1999), 'Cooperation vs. competition in a spatial model', *Regional Science and Urban Economics* **29**, 463–472.

Mason, E. S. (1939), 'Price and production policies of large-scale enterprises', *American Economic Review* **29**, 61–74.

McDonald, R. & Siegel, D. (1986), 'The value of waiting to invest', *Quarterly Journal of Economics* **101**, 707–727.

McGahan, A. M. (1999), 'The performance of US corporations: 1981-1994', *Journal of Industrial Economics* **47**, 373–398.

McGahan, A. M. & Porter, M. E. (1997), 'How much does industry matter, really?', *Strategic Management Journal* **18**, 15–30.

McGee, J. & Thomas, H. (1986), 'Strategic groups: Theory, research and taxonomy', *Strategic Management Journal* **7**, 141–160.

McGrath, R. G. (1997), 'A real options logic for initiating technology positioning investments', *Academy of Management Review* **22**, 974–996.

Miles, G., Snow, C. C. & Sharfman, M. P. (1993), 'Industry variety and performance', *Strategic Management Journal* **14**(3), 163–177.

Mueller, D. C. (1986), *Profits in the Long Run*, Cambridge University Press, Cambridge, MA.

Mueller, D. C., ed. (1990), *The Dynamics of Company Profits: An International Comparison*, Cambridge University Press, Cambridge, MA.

Mussa, M. & Rosen, S. (1978), 'Monopoly and product quality', *Journal of Economic Theory* **18**, 301–317.

Nehring, K. & Puppe, C. (2001), 'A theory of diversity', *Econometrica (forthcoming)* .

Neven, D. (1985), 'Two stage (perfect) equilibrium in Hotelling's model', *Journal of Industrial Economics* **33**, 317–325.

Neven, D. J. (1986), 'On Hotelling's competition with non-uniform customer distributions', *Economics Letters* **21**(2), 121–126.

Neven, D. & Thisse, J.-F. (1990), On quality and variety competition, *in* J. J. Gabszewicz, J. Richard & L. Wolsey, eds, 'Economic Decision Making: Games, Econometrics and Optimisation. Contributions in Honour of J. Drèze', Elsevier Science Publishers, Amsterdam.

Norman, G. & Thisse, J.-F. (1996), 'Product variety and welfare under discriminatory and mill pricing policies', *Economic Journal* **106**, 76–91.

Novshek, W. (1980), 'Equilibrium in simple spatial (or differentiated product) models', *Journal of Economic Theory* **22**, 313–326.

Osborne, J. D., Stubbart, C. I. & Ramaprasad, A. (2001), 'Strategic groups and competitive enactment: A study of dynamic relationships between mental models and performance', *Strategic Management Journal* **22**, 435–454.

Osborne, M. J. & Pitchik, C. (1987), 'Equilibrium in Hotelling's model of spatial competition', *Econometrica* **55**, 911–923.

Padilla, A. J. (1992), 'Mixed pricing in oligopoly with consumer switching costs', *International Journal of Industrial Organization* **10**, 393–412.

Padilla, A. J. (1995), 'Revisiting dynamic duopoly with consumer switching costs', *Journal of Economic Theory* **67**, 520–530.

Pakes, A. (1987), 'Mueller's profits in the long run', *RAND Journal of Economics* **18**, 319–332.

Peitz, M. (1999), 'A difficulty with the address models of product differentiation', *Economic Theory* **14**, 717–727.

Perotti, E. & Rossetto, S. (2001), 'Internet portals as portfolios of entry option', *Paper presented at the American Finance Association, New Orleans* .

Peteraf, M. (1993), 'Intraindustry structure and response towards rivals', *Journal of Managerial and Decision Economics* **14**, 519–528.

Peteraf, M. & Shanley, M. (1997), 'Getting to know you: A theory of strategic group identity', *Strategic Management Journal, Summer Special Issue* **18**, 165–186.

Pinkse, J., Slade, M. E. & Brett, C. (2001), 'Spatial price competition: A semiparametric approach', *Econometrica (forthcoming)* .

Porac, J. F., Thomas, H. & Baden-Fuller, C. (1989), 'Competitive groups as cognitive communities: The case of Scottish knitwear manufacturers', *Journal of Management Studies* **15**, 397–416.

Porter, M. E. (1979), 'The structure within industries and companies' performance', *Review of Economics and Statistics* pp. 214–227.

Porter, M. E. (1980), *Competitive Strategy: Techniques for Analyzing Industries and Competitors*, Free Press, New York.

Porter, M. E. (1985), *Competitive Advantage*, Free Press, New York.

Priem, R. L. & Butler, J. E. (2001), 'Is the resource-based "view" a useful perspective for strategic management research?', *Academy of Management Review* **26**(1), 22–40.

Rapoport, A. & Amaldoss, W. (2000), 'Mixed strategies and iterative elimination of strongly dominated strategies: An experimental investigation of states of knowledge', *Journal of Economic Behavior and Organization* **42**(4), 483–521.

Reger, R. K. & Huff, A. S. (1993), 'Strategic groups: A cognitive perspective', *Strategic Management Journal* **14**, 103–123.

Rhee, B.-D., de Palma, A., Fornell, C. & Thisse, J.-F. (1992), 'Restoring the principle of minimum differentiation in product positioning', *Journal of Economics and Management Strategy* **1**, 475–506.

Roquebert, J. A., Phillips, R. L. & Westfall, P. A. (1996), 'Markets vs. management: What 'drives' profitability?', *Strategic Management Journal* **17**, 653–664.

Ross, T. W. (1992), 'Cartel stability and product differentiation', *International Journal of Industrial Organization* **10**, 1–13.

Rothschild, R. (1997), 'Product differentiation and cartel stability: Chamberlin versus Hotelling', *The Annals of Regional Science* **31**, 259–271.

Rumelt, R. P. (1991), 'How much does industry matter?', *Strategic Management Journal* **12**, 167–185.

Rumelt, R. P., Schendel, D. E. & Teece, D. J. (1994), Fundamental issues in strategy, *in* R. P. Rumelt, D. E. Schendel & D. J. Teece, eds, 'Fundamental Issues in Strategy: A Reserach Agenda', Harvard Business School Press, pp. 9–47.

Salop, S. (1979), 'Monopolistic competition with outside goods', *Bell Journal of Economics* **10**, 141–156.

Schmalensee, R. (1985), 'Do markets differ much?', *American Economic Review* **75**, 341–351.

Schmalensee, R. (1987), 'Collusion vs. differential efficiency: Testing alternative hypothesis', *Journal of Industrial Economics* **35**, 399–425.

Schwalbach, J., Grasshoff, U. & Mahmood, T. (1989), 'The dynamics of corporate profits', *European Economic Review* **33**, 1625–1639.

Schwartz, E. & Zozaya-Gorostiza, C. (2000), 'Valuation of information technology investments as real options', *Finance Working Paper #6-00, Anderson Graduate School of Management, University of California at Los Angeles* .

Scott, W. R. (1995), *Institutions and Organizations*, Sage, Thousand Oaks, CA.

Shaked, A. & Sutton, J. (1982), 'Relaxing price competition through product differentiation', *Review of Economic Studies* **49**, 3–13.

Shaked, A. & Sutton, J. (1983), 'Natural oligopolies', *Econometrica* **51**, 1469–1483.

Shaked, A. & Sutton, J. (1987), 'Product differentiation and industrial structure', *Journal of Industrial Economics* **36**, 131–146.

Shilony, Y. (1981), 'Hotelling's competition with general customer distributions', *Economics Letters* **8**, 39–45.

Sinclair-Desgagné, B. & Röller, L.-H. (1996), 'On the heterogeneity of firms', *European Economic Review* **40**, 531–539.

Sinclair-Desgagné, B. & Röller, L.-H. (1997), 'Heterogeneity in a duopoly', *Mimeo* .

Smith, K. G., Grimm, C. M. & Wally, S. (1997), 'Strategic groups and rivalrous firm behavior: Towards a reconciliation', *Strategic Management Journal* **18**, 149–157.

Smithies, A. (1941), 'Optimum location in spatial competition', *Journal of Political Economy* **49**, 423–439.

Spence, A. M. (1976), 'Product selection, fixed costs and monopolistic competition', *Review of Economic Studies* **43**, 217–235.

Stahl, K. (1982), 'Differentiated products, consumer search, and locational oligopoly', *Journal of Industrial Economics* **31**, 97–114.

Tabuchi, T. (1994), 'Two-stage two-dimensional spatial competition between two firms', *Regional Science and Urban Economics* **24**, 207–227.

Tabuchi, T. (1999), 'Pricing policy in spatial competition', *Regional Science and Urban Economics* **29**, 617–631.

Tabuchi, T. & Thisse, J.-F. (1995), 'Asymmetric equilibria in spatial competition', *International Journal of Industrial Organization* **13**, 213–227.

Tang, M. & Thomas, H. (1992), 'The concept of strategic groups: Theoretical construct or analytical convenience', *Managerial and Decision Economics* **13**, 323–329.

Thisse, J.-F. & Vives, X. (1988), 'On the strategic choice of spatial price policy', *American Economic Review* **78**, 122–137.

Thomas, H. & Venkatraman, N. (1988), 'Research on strategic groups: Progress and prognosis', *Journal of Management Studies* **25**, 537–555.

Thomas, L. & Weigelt, K. (2000), 'Product location choice and firm capabilities: Evidence from the U.S. automobile industry', *Strategic Management Journal* **21**, 897–909.

Tirole, J. (1990), *The Theory of Industrial Organization*, MIT Press, Cambridge MA.

To, T. (1996), 'Multiperiod competition with switching costs: An overlapping generations formulation', *Journal of Industrial Economics* **44**, 81–87.

Trigeorgis, L. (1998), *Real Options: Managerial Flexibility and Strategy in Resource Allocation*, MIT Press, Cambridge, MA.

von Ungern-Sternberg, T. (1988), 'Monopolistic competition and general purpose products', *Review of Economic Studies* **55**, 231–246.

Wang, T. (2001), 'Equilibrium with new investment opprotunities', *Journal of Economic Dynamics and Control* **25**, 1751–1773.

Waterson, M. (1989), 'Models of product differentiation', *Bulletin of Economic Research* **41**(1), 1–27.

Weber, S. (1997), 'Entry deterrence in electoral spatial competition', *Social Choice and Welfare* **15**, 31–56.

Wiggins, R. R. & Ruefli, T. W. (1995), 'Necessary conditions for the predictive validity of strategic groups: Analysis without reliance on clustering techniques', *Academy of Management Journal* **38**, 1635–1656.

Zhang, Z. J. (1995), 'Price matching policy and the principle of minimum differentiation', *Journal of Industrial Economics* **43**, 287–299.

## Erklärung

Hiermit erkläre ich an Eides statt, die vorliegende Arbeit selbständig und ohne Benutzung anderer als der angegebenen Hilfsmittel angefertigt zu haben. Die Personen, von denen ich Unterstützung erhalten habe, sind im Vorwort erwähnt.

Die Arbeit wurde in gleicher oder ähnlicher Form keiner anderen Prüfungsbehörde vorgelegt.

Ich bezeuge durch meine Unterschrift, dass meine Angaben über die bei der Abfassung meiner Dissertation benutzten Hilfsmittel, über die mir zuteil gewordene Hilfe sowie über frühere Begutachtungen meiner Dissertation in jeder Hinsicht der Wahrheit entsprechen.

Berlin, den 2. Januar 2002

Steffen Brenner