

A Thesis Submitted for the Degree of PhD at the University of Warwick

Permanent WRAP URL:

<http://wrap.warwick.ac.uk/106901/>

Copyright and reuse:

This thesis is made available online and is protected by original copyright.

Please scroll down to view the document itself.

Please refer to the repository record for this item for information to help you to cite it.

Our policy information is available from the repository home page.

For more information, please contact the WRAP Team at: wrap@warwick.ac.uk

INDUSTRIAL ECONOMICS STUDIES IN INSURANCE
MARKETS

by

Otto Iisakki Toivanen

Thesis submitted for a degree of Doctor of Philosophy to
the University of Warwick Department of Economics

August 1994

CONTENTS

I	INTRODUCTION	1
I.1	The objectives of the thesis	1
I.2	Literature overview	3
I.3	Brief outline of the three main chapters	9
I.3.1	Persistence of profits, strategic groups and the effect of mergers on competition: the Finnish non-life insurance market	9
I.3.2	Informationally asymmetric markets and organizational form	11
I.3.3	Oligopolistic services and cost function estimation	13
I.4	Organization of the thesis	15
II	PERSISTENCE OF PROFITS, STRATEGIC GROUPS AND THE EFFECT OF MERGERS ON COMPETITION: THE FINNISH NON-LIFE INSURANCE MARKET	17
II.1	Introduction	17
II.2	Theory and method	19
II.3	The market	29
II.4	Strategic groups and profits	35
II.5	Hypotheses and estimation results	43
II.6	Summary	49
	Appendix I: econometric restrictions and specification tests	52
	A.I.1 econometric restrictions	52
	A.I.2 specification tests	54
	Appendix II: regulation of the Finnish insurance market	57
	Appendix III: how to calculate the profits of an insurance firm, form economic entities and exclude firms from the sample	61

III	INFORMATIONALLY ASYMMETRIC MARKETS AND ORGANIZATIONAL FORM	69
III.1	Introduction	69
III.2	Adverse selection and moral hazard in competitive insurance markets	76
III.3	Vertical strategies and monopoly	92
III.4	Vertical strategies and oligopoly	104
III.5	Oligopolistic vertical strategies with adverse selection and moral hazard	127
III.6	Other interpretations	131
	III.6.a Labour market, education, and recruitment policy	132
	III.6.b Internal organization of creditors	133
	III.6.c Internal organization and personnel policy	134
	III.6.d organization of regulatory institutions	135
III.7	Conclusions	136
IV	OLIGOPOLISTIC SERVICES AND COST FUNCTION ESTIMATION	141
IV.1	Introduction	141
IV.2	A model of branch networks	145
IV.3	The data, the market and the empirical model	157
IV.4	Empirical results and comparisons to earlier studies	167
IV.5	Conclusions	176
	Appendix: proofs of the propositions	180
V	SUMMARY	188
	REFERENCES	194

LIST OF FIGURES AND TABLES

table II.1	submarkets of the Finnish non-life insurance market	31
figure II.1	Herfindahl-index, direct non-life insurance	33
table II.2	Descriptive statistics of profits in the Finnish non-life insurance industry	40
table II.3	Persistence of profit equation (3b)	47
table A.I.1	Restrictions	54
table A.I.2	F-tests	55
table A.I.3	F-tests	55
table A.I.4	F-tests	56
table A.III.1	Profit calculation of insurance firms	63
figure III.1	Moral hazard	79
figure III.2	Moral hazard and perfect competition	83
figure III.3	Adverse selection and moral hazard with perfect competition	87
figure III.4	Non-existence of equilibrium	89
figure III.5	Welfare in competitive markets	91
figure III.6	Vertically integrated monopoly	94
figure III.7	Direct-selling monopoly	100
figure III.8	Threats in oligopoly	111
figure III.9	Welfare in oligopoly	120
figure III.10	Oligopoly with moral hazard and adverse selection	128
figure IV.1	Two-city Hotelling	150
table IV.1	Lines of insurance in the Finnish non-life insurance market (1991 data)	159
table IV.2	Key information on firms in the Finnish non-life insurance market	160
table IV.3	Descriptive statistics of estimation variables	169
table IV.4	Estimation results for the quadratic cost function	170
table IV.5	Estimates of economies of scale and scope	171
table IV.6	Estimates of the cost of branch proliferation	175

ACKNOWLEDGEMENTS

It is a great pleasure to thank people who have in one way or another helped me during this process, and I am lucky to have several persons to thank. My two supervisors, Dennis Leech and Mike Waterson have not only greatly improved the contents of the thesis, but have also given moral support whenever that was needed. I am glad to be able to call them not only my supervisors, but also my friends. This is equally true in the case of Paul Stoneman, whose research assistant I have been, even if he hasn't been directly involved with my thesis. Many people apart from those already mentioned have given comments to some or all of the chapters, and the following list is not exhaustive: Keith Cowling, Tor Eriksson, Morten Hviid, Norman Ireland, Ismo Linnosmaa, Harri Lonka, Tapani Myllymäki, Paavo Okko and Veikko Reinikainen. Of these, I would like to thank Tor Eriksson and Paavo Okko, supervisors of my licentiate thesis, for their encouragement and support. The same thanks go to Seppo Honkapohja, my undergraduate professor, who at later stages of my studies has given me his unconditional support.

On par with the aforementioned people come my various office- and flatmates and other friends at Warwick; they made the time I was writing my thesis extremely enjoyable. My friends back home deserve thanks, too. I would also like to extend thanks to my brothers and sisters, my parents, and Kati and Mirva. Without you all this would never have happened.

Last, but not least, I would like to thank those institutions that provided the financing of my post-graduate studies: Yrjö Jahnesson Foundation, Tapiola insurance group, the Association of Finnish Insurers and Turku School of Economics and Business Administration Sampo Fund. Thank you for your support.

DECLARATION

Chapter III of the thesis "Persistence of profits, strategic groups and the effect of mergers on competition: The Finnish non-life insurance market" is partly based on my licentiate thesis, represented at the Turku School of Economics and Business Administration in 1992. The chapter is published as Warwick University economics department working paper no. 9422.

TO LIFE

SUMMARY

This thesis consists of three essays each studying insurance markets from a different perspective. The first studies competition in the domestic Finnish non-life insurance market using a persistence of profits model, where it is assumed that firms use competitors' past profits as signals of attractiveness of given submarkets. The firms were divided into two strategic groups. The existence of these groups, the effects of two mergers, and the level of competition were tested for. It emerged that the groups compete hard against each other, that fringe firms compete more with the leader group than with each other, that leaders' either follow some kind of tacit collusion strategy or compete very aggressively against each other, and that the mergers lead to a tightening of competition. The second essay is theoretical. The question asked is: does it pay for an insurance firm to acquire information of its customers' type and level of effort. Adverse selection and moral hazard analyses are combined, using geometric tools. Welfare analysis is central in this essay. Decision rules are derived for a monopoly to become vertically integrated. It is shown that in oligopoly it is possible to have an equilibrium where firms use asymmetric vertical strategies. Welfare effects of vertical integration prove to be ambiguous. The model has several other applications, eg. job market, organization of regulatory institutions. In the third essay it is argued that oligopolistic firms do not necessarily minimize costs when maximizing profits, and that this affects cost function specification and estimation. A cost function is constrained so that it can be estimated even though the number of products is large. The proposed specification gives a better fit than traditional specifications, and the quantitative and qualitative results are very different. The costs of branch proliferation are calculated, and the lowest mean for five biggest firms is 37% of total operating costs.

I INTRODUCTION

I.1 The objectives of the thesis

Insurance as an economic phenomenon is of tremendous importance. In recent decades economists have come to understand that a lot of economic activity is coloured at least partly by an insurance objective. In markets as far apart as financial markets and labour markets, it is now acknowledged that insurance plays a major role. Within this broad category of insurance, the actual insurance markets are in an important position. It can be argued that without well functioning insurance markets a lot of economic activity would not exist, and thus the importance of insurance markets (usually only a few percentage points of GDP in any given country, insurance centres apart) stretches wide over the actual market borders.

In the economic literature concerning insurance markets themselves, the two largest strands are the design of an optimal insurance policy (eg. Raviv 1979) and the effects of asymmetric information (see the collection edited by Dionne&Harrington 1992). Whilst these research directions are of undoubted importance in uncovering the mechanisms of insurance, they are of only indirect interest in this study, because most of these literatures assume competitive markets. In this study, the emphasis is firmly in imperfectly competitive markets. As pointed out above (and uncovered in the above literature), the way an insurance market functions, and thus the degree of competition therein, has not only direct effects, but also indirect

ones in that a lot of socially valuable economic activity is not carried out if there is no appropriate insurance available at an affordable price level.

Despite the above, there is a clear lack of research in the area of how imperfectly competitive insurance markets work. The main objective of this thesis is to partly fill that gap, with both theoretical and empirical work. In the process, my aim is to use the tools created in recent years in industrial economics, and to intelligently apply them to the theoretical and empirical environment that insurance as a phenomenon and the Finnish non-life insurance market as a source of data provide. In industrial economics, the behaviour of imperfectly competitive markets is studied. Questions asked are e.g. how likely is entry, can collusion be sustained, is there excess or inadequate R&D, do the firms enjoy supernormal profits and if so, then why? As is common in economics, welfare issues connected with the above-type questions are often raised. The current literature is influenced by work in the theory of the firm and regulation and in fact many people work on all of these areas. Industrial economics has a strong empirical past and after the surge of theoretical literature during the last ten years or so, empirical work, especially various dynamic models of competition, are getting more attention.

As industrial economics has expanded it has slowly begun to tackle some specific problems of financial markets. Whether theoretical (e.g. financial intermediation) or empirical (e.g. estimations of scale and scope) work, the

interest has never the less centred very much on banking. Insurance on the other hand has by and large remained untouched by industrial economists. This is not to say that there are no studies, but that the importance of the industry is not reflected by the number of studies. There are understandable reasons for this lack of interest: the product is badly defined (what is an insurance policy, how do you measure output?), firms are operating on a multitude of markets (as we can interpret every line of insurance as a distinct market) and theoretical models suffer under the various informational effects. In many (all?) countries the situation is worsened (from the point of view of an researcher) by some type/several types of regulation.

1.2 Literature overview

As each chapter includes a discussion of the general (as opposed to insurance) literature relevant for that chapter, I will here concentrate mainly on insurance literature. The seminal study in the area of industrial economics and insurance is from Joskow (1973). He studies the U.S. property-liability industry using the structure-conduct-performance-framework. He studied scale economies, the efficiency of the market and the effects of regulation. According to Joskow the U.S. insurance market is competitive with low levels of concentration and relatively easy entry. Regulation distorts competition and protects ineffective distribution systems. There are (almost) no economies of scale. Cummins and

interest has never the less centred very much on banking. Insurance on the other hand has by and large remained untouched by industrial economists. This is not to say that there are no studies, but that the importance of the industry is not reflected by the number of studies. There are understandable reasons for this lack of interest: the product is badly defined (what is an insurance policy, how do you measure output?), firms are operating on a multitude of markets (as we can interpret every line of insurance as a distinct market) and theoretical models suffer under the various informational effects. In many (all?) countries the situation is worsened (from the point of view of a researcher) by some type/several types of regulation.

1.2 Literature overview

As each chapter includes a discussion of the general (as opposed to insurance) literature relevant for that chapter, I will here concentrate mainly on insurance literature. The seminal study in the area of industrial economics and insurance is from Joskow (1973). He studies the U.S. property-liability industry using the structure-conduct-performance-framework. He studied scale economies, the efficiency of the market and the effects of regulation. According to Joskow the U.S. insurance market is competitive with low levels of concentration and relatively easy entry. Regulation distorts competition and protects ineffective distribution systems. There are (almost) no economies of scale. Cummins and

Harrington (1987) studied the effects of rate regulation in the U.S. property-liability market using cross section data (as did Joskow, although he did run separate estimations on data from several years). Their results can largely be interpreted in favour of regulation, as it resulted in higher loss-ratios¹ and thus gave consumers more value for their money. Firm size had a negative impact on loss-ratios. Loss-ratios do not directly tell the profitability of an insurance firm, but they tell how much of the premiums flow back to the customers as incurred claims. To find out the profitability of an insurance firm, its investment income has to be added to its earnings from the "pure" insurance activity. Finsinger and Schmid (1991) study prices, distribution channels and regulation in Europe. According to their study regulation is costly and tends to decrease market share variability, thus stabilising the market. Concentration has a similar, though smaller effect. Tied distribution channels (compared to untied brokers) provide less information to customers and are more prevalent in regulated markets. Cummins and Vanderhei (1979) compare different distribution systems. Their results are that exclusive agents are more efficient than independent ones (these results are in line with Joskow 1973), the difference being between 15 and 23%. These results are not sensitive to loss adjustment expenses, indicating that the difference comes from operating expenses. Zweifel and Ghermi (1990) also compare exclusive and independent agents, using Swiss data. Their findings indicate that exclusive agencies have significantly higher expense ratios that can not be attributed to risk

¹ loss-ratio is defined as follows: (claims incurred)/premiums

selection. Their results are thus orthogonal to findings based on U.S. data.

One of the common variables used in empirical studies is the ownership form of firms, since there is a clear division between stock owned and mutual firms in insurance. Mayers and Smith (1982) concentrate on ownership effects, again using U.S. data. I will here concentrate only on their results when comparing stock owned and mutual firms, although they had two more ownership forms in their estimations. The reason for my choice is that in the Finnish context these are the two relevant ownership forms. They hypothesized that mutuals would concentrate on lines of business that are not vulnerable to managerial discretion, since in mutuals it is more difficult to control managers than it is in stock-owned firms. Because of this effect mutuals should also concentrate more on specific lines of insurance, thus easing the controlling of management. For the same reason mutuals should be geographically more concentrated than stock-owned firms. They found out that stock-companies are less concentrated geographically and by line of insurance. When controlling for size no distinction can be made, however. Thus there is only partial support for their hypothesis concerning geographical concentration. The by-line concentration hypothesis, however, seems to get support.

A class of their own are specific studies of economies of scale and/or scope. In contrast to the banking industry, there are only a few studies using modern econometric techniques. Studies conducted by Joskow (1973),

Allen (1974), Skogh (1982) and others concentrate explicitly on scale economies and besides this use cost functions that currently are referred to as primitive, such as linear or Cobb-Douglas-functions. The only study that to my knowledge uses a flexible cost function to estimate both economies of scale and scope in insurance is due to Suret (1991). He uses a translog cost function on Canadian data. His results show that there are significant economies of scale for firms with assets between 40 and 100\$ million. He finds no evidence on economies of scope. Suret used data from three years (1986-1988) and claims as a measure of output. He estimated the cost function separately for every year. The problem is that three years can prove too short a period since claims can fluctuate a lot. A more stable measure of output would be desirable.

The theoretical industrial economics literature is even more scarce. Insurance has received lot of attention among economists studying effects of information on markets. Moral hazard (Shavell 1979) and adverse selection (Rotschild&Stiglitz 1976) literatures were very much based on an insurance context. Most of the theoretical literature is concerned with regulation and no wonder, regulation has such a major impact on insurance markets. The industrial economics regulation literature (see e.g. Joskow and Rose 1989 for a survey) differs from traditional insurance regulation literature in that it assumes a non-competitive market. This insurance regulation literature (e.g. MacMinn&Witt 1987) often assumes a competitive market and places the firms then under a regulatory

constraint(s). Models of insurance firms also seem often to deviate from the classical profit-maximizing assumption and instead have firms maximizing utility, appropriately defined. As an example of the of non-profit-maximizing literature, Ang and Lai (1987) study the pricing behaviour of a mean-variance-maximizing insurance firm without discussing too much the choice of the optimand. There is to my knowledge no genuine theoretical industrial organization literature on insurance. Most research on insurance firm behaviour takes its tools from the financial literature, for obvious reasons. These models are then, however, unable to capture differences in the behaviour of firms operating in competitive insurance markets vis-a-vis firms that operate in imperfectly competitive insurance markets.

To review the economics literature on Finnish insurance markets is an easy task since it is almost non-existent. Apart from few papers on forecasting in insurance (Salo 1980), the only economic papers on Finnish insurance are Valkonen's study on insurance firms' investment behaviour (Valkonen 1990) and my study (Toivanen 1992a). Valkonen's paper addresses insurance from a finance-point of view, not a competitive point of view, although these are, of course, related. My study, conducted for the Office of Free Competition, covers all submarkets and because it is the first research on the area made for the Office, it concentrates very much on the institutional setting of the market and the effects it has on competition. It does not contain any econometric modelling. In addition to these I should

mention my unpublished licentiate thesis (Toivanen 1992b). It will form the basis of the first chapter. Since its main empirical contents will become clear later, I only mention it here.

As a summary of the existing literature it can be noted that there is a clear need for modern research on insurance using the tools of industrial economics and thus addressing questions directly related to competition. The earlier studies either lack this point of view or then, as is the case of studies of economies of scale, use outdated empirical tools. A special caveat is that all the studies use cross-section data. This will lead to erroneous results since it is clear to anybody familiar with the industry that yearly data fluctuate due to stochasticity of losses. Theoretical work is even more rare than empirical.

What is needed is research that addresses the problems side-stepped previously. In empirical work, models that take into account the difficulties mentioned in the beginning of this section (multiproduct industry, product differentiation, heterogenous customers, price discrimination) have to be used. This will in many cases mean that models relying on straight forward customer- and firm-optimization have to be abandoned since they are not able to capture the above mentioned features. Panel data should be used. If it is not used, the reason for that should be explicitly stated. Theoretical work should take use of modern tools whenever possible and

create models based on profit maximization² of firms.

I.3 Brief outline of the three main chapters

I.3.1 Persistence of profits, strategic groups and the effect of mergers on competition: the Finnish non-life insurance market

Persistence of profits models (Mueller 1990) measure the signalling effect of past profits. If a company has higher than average profits, this should attract competition and increased competition should result in smaller profits. So the idea behind these models is very simple. An even more appealing feature is that persistence of profits models allow for product differentiation, price discrimination and other market imperfections that make the use of more rigorously derived models difficult. Since it is inherently dynamic in nature, it suits very well the insurance industry. I use the framework that I created in my licentiate thesis (Toivanen 1992b³) for the model developed by Geroski (Geroski 1990). The Geroski model differentiates between two sources of competitive pressure: the group of the firm under study and other groups. In Geroski's original formulation

² here I have to allow for variation if I consider different ownership forms. A mutual firm might better be modelled as a cooperative firm, maximizing the welfare of its customers who at the same time are its owners, too.

³The econometric modelling used in the licentiate thesis differs substantially from that used here. The data sets are also somewhat different.

these groups were industries, but I applied the model to strategic groups and divided the market into two strategic groups, the "leaders" and the "fringe". The model measures the differences in the behaviour of these groups, but as panel data methods are used, the model allows for intra-group differences, too.

The theory of strategic groups applied entry barriers to individual industries (Caves and Porter 1977). The authors claimed that there can be significant profit differences within industries and that these could be at least partly attributed to within-the-industry entry barriers, which they labelled mobility barriers. This claim has since been verified (Newman 1977, Porter 1978, Oster 1981). I used this theory to form the two above-mentioned groups, the leaders and the fringe.

In this essay I develop my previous work and study the effect of two mergers within the leaders group in the early 80's. At the time, Aura and Pohja Groups merged to what now is the Tapiola Group and the Fennia Group joined forces with Yrittäjän Vakuutus. Now I study what was the effect of these two mergers on competition, measured by the persistence of profits model. This is accomplished by augmenting the original model and then using statistical tests to discriminate between the different models. I use data from the period 1970-1991. The standard persistent of profits model is thus augmented in several ways: instead of using inter-industry data and doing firm-level time series based analysis, I concentrate

on a single industry and the emphasis is on strategic groups, not firms. In addition, I am not aware of other empirical research studying the effects of mergers on competition in insurance⁴.

I.3.2 Informationally asymmetric markets and organizational form

Insurance markets are one of the established textbook examples of markets plagued by asymmetric and imperfect information. While the effects of these have been extensively studied in a competitive market environment, far less work has been done in an imperfectly competitive setting, the notable exception being the monopoly study of Stiglitz (1977). The first paper utilizing the type of model I build on is that of Rothschild and Stiglitz (1976), who showed that the only possible equilibrium is a separating one, and that none might exist. They concentrated on adverse selection and left moral hazard (see eg. Shavell 1979) out of the analysis. In this paper, I combine an analysis of moral hazard with that of adverse selection. Most studies assume that the firm(s) cannot find out the true characteristics of customers, or the effort they exert. While there are papers studying the relaxation of this assumption (Guasch&Weiss 1980, Nalebuff&Scharfstein 1987⁵), they use labour market models, with (perfectly) competitive firms. Here, my emphasis is very different: there is imperfect competition

⁴ as an example of studies on other industries, Barton&Sherman 1984 study microfilm producers.

⁵ I should probably notify the reader that I became aware of this literature only after having constructed my model.

(monopoly of duopoly), and firms can either hire an agent (labelled vertical integration) or resort to self-selection, as in earlier models⁶. This decision is discussed using mainly geometrical arguments relying on the fact that the axes of the figures and the 45°-degree line measure not only the customer's wealth, but also the firms' profits. I make the simplifying assumption that the agent is able to find out the true type of any customer. In the monopoly case, this enables the monopolist to capture all the surplus from trade whereas a direct-selling monopolist, relying on self-selection, has (possibly) to surrender some of the surplus to its customers. In both the monopoly and the oligopoly models there is firm-intern vertical product differentiation in the case of the vertically integrated firm. In the oligopoly model, there is vertical product differentiation also between the firms, but only in one part of the market. The welfare effects of vertical integration are also discussed. As far as I know, the motivation behind vertical integration in this model, acquiring knowledge of customer characteristics, is new to the literature.

This is the only purely theoretical paper of the thesis. Although the model produces several testable results, my opinion is that an empirical model should incorporate some of the aspects left out of the analysis here, the most notable ones being horizontal product differentiation and regulation.

⁶ The combined moral hazard and adverse selection analysis is introduced in a model of competitive markets, however.

I.3.3 Oligopolistic retailing and cost function estimation

In this paper I estimate economies of scale and scope in insurance. As I reported in the literature survey, most previous research has used outdated models, including linear and Cobb-Douglas cost functions. Only Suret (1991) has employed a flexible functional form, which have superior properties compared to those functional forms used in previous studies of insurance. A flexible cost function allows the measurement of U-shaped cost curves and the simultaneous estimation of economies of scale and scope. As (Finnish) insurance firms are multiproduct firms, the simultaneous estimation of both of these effects is the only sensible research strategy. Omitting either one would give erroneous results for the one studied.

The most widely used flexible cost function, the translog function, has one unfortunate feature that makes us unable to use it. It does not allow zero production of any single product. To avoid this problem we use a quadratic cost function (see Baumol, Panzar & Willig 1982), that has similar flexibility (ie. allows U-shaped cost curves) as the translog, but does not collapse if some product is not produced.

An additional problem for this kind of studies in insurance is that there is no broad consensus among economists on how to measure the output of an insurance firm. Both premiums and claims have been used. The former

has the difficulty that it is a valid measure of output only when the market is in long-term equilibrium (Skogh 1982), a condition that is hard to verify. The latter presents the problem that it is highly volatile and thus cross-section (or even short panel data) estimations are sure to give erroneous results. Using a Hotelling type (see eg. Tirole 1988) model I show that in addition to these problems, if the firms engage in product differentiation in an oligopoly setting, premium income is a biased measure of output. The problems created by product differentiation are present in all services markets, not only in insurance.

In tackling this problem, I am going to take a new way and use as the measure of output the number of policies sold. This information is available from Finnish insurance statistics to all lines of insurance but reinsurance and foreign insurance. The rationale behind this is that in my view an insurance firm should be viewed as a "black box", using inputs (labour, capital) to produce a stochastic process which entails implicitly both profits and claims. An insurance company does not know the exact claim it has to pay for a given policy, only the expectation of this. The claims process follows statistical laws (Pentikäinen et. al. 1989) and is inherently different from the production process that created it. Thus the economies of scale and scope of the stochastic process (including investments) should be studied separately from the production processes' economies of scale and scope. The problem with the number of policies as a measure of output is that the a given type of policies sold are not of the

same size. This should not be too big a problem, though, since the production costs of e.g. different sized auto-policies are probably very near each other, ie. the production costs (not including claims of course) are not very dependent on the value of the policy. A similar measure of output is in wide use in banking studies (e.g. Kolarik&Zardkoochi (1989) use the number of bills and advances in their estimations).

There are inherent problems in using normal cost function techniques to services industries. The basic argument brought forward in this chapter is that in services, the unit where production happens is the branch, and thus services firms are multiplant firms. Furthermore, opening a new branch increases the market power of a firm. I claim, and then show theoretically (and empirically) that the firms weigh two effects against each other: the increase in market power due to additional branches and the (possible) increase in average costs due to the same thing. This can create a situation (depending on the characteristics of the market) where there are diseconomies of scale at the firm level and economies of scale at the branch level.

I.4 Organization of the thesis

As the title of the thesis probably reveals, the thesis consists of work that have little direct links with each other. The three substantive chapters - chapters 2 to 4 - are independent of each other. All of them have their

own introduction, literature review (placed into the introduction) and conclusion or summary sections, and they can be read in any order. The ordering of the chapters within the thesis reflects more the process of work than anything else: the chapters are in chronological order. Because of this self-sustainability, there is no need for an ordinary conclusions or summary chapter. Despite this, one is provided. There (ch. 5), I very briefly highlight the main results - as I have here tried to highlight the main objectives of the thesis - and discuss directions for future work. These are not necessarily directly linked to the chapters presented here in a technical sense, but without doubt the work done for this thesis has been used as a source of inspiration in drawing up these future plans.

II STRATEGIC GROUPS, THE PERSISTENCE OF PROFITS AND THE EFFECT OF MERGERS ON COMPETITION: THE FINNISH DOMESTIC NON-LIFE INSURANCE MARKET

II.1 Introduction

Multi-market, oligopolistic, product-differentiating firms that actively price discriminate between their customers are not among the easiest modelling targets. The insurance firms belongs to this category as every line of insurance can be interpreted as an independent market, all firms are active in product differentiation¹, and price discrimination is the bread and butter of the industry as it tries to differentiate between high- and low-risk customers². The market under study here, the Finnish non-life insurance market, is very concentrated. In addition to this profits are disturbed by stochastic shocks; i.e. claims fluctuate from period to period.

1 This is at least partly due to signalling. As the risk-averse customers do not have perfect information on the quality of the product they are buying, firms can enter signalling games in order to assure customers that in the case of an accident they will pay compensation to the full amount of damage. This is, however, harder than normally since only a fraction of customers have an accident and thus the normal reputation effects are not as effective than with, say, durable goods.

2 The standard Rothschild-Stiglitz (1976) result can be interpreted as price discrimination as the firms are not willing to sell the same product to all customers at the same price. This is the motivation behind the firms' provision of products that take into account the self-selection constraints. Bond and Crocker (1991) have shown that insurance firms can use prior information to categorize their customers and thus avoid adverse selection and moral hazard problems.

The aim of this chapter is to measure competition or competitive pressure in the Finnish non-life insurance market using a persistence of profits model. This type of models has earlier been used with data from several industries (Mueller 1977, 1986, 1990, Cubbin&Geroski 1987). As I will later argue, the model is probably better suited to industry (or small sample of industries) studies than to its original use. I divide the market according to the theory of strategic groups (Caves&Porter 1977) and then estimate competition within each group and between the groups. In addition to this, the effects on competition of two simultaneous mergers within the strategic group "leaders" is studied. Most papers that study mergers' effects on competition seem to use stock market data (eg. Prager 1992). Here, the effects are traced from firms' behaviour³.

In the second section of the chapter I review the earlier literature, present the theory of strategic groups and the econometric model. The third section contains a description of the Finnish non-life insurance market. The fourth section discusses the data and presents the way the firms were divided into two groups. The hypotheses are presented and tested in section five. Section six concludes the chapter.

³ As an example of a paper using non-stock market data, Kim&Singal 1993 measure price changes in the airline industry.

II.2 Theory and method

When limiting the view on entry barriers it has been standard to assume that incumbent firms are all alike except for their size. When we relax this assumption we at the same moment must somehow categorize the firms in an industry. This can be done with the help of strategic groups. Caves&Porter (1977) define an industry as a group of competitors producing substitutes that are close enough that the behaviour of any firm affects each of the others either directly or indirectly. An industry can then be viewed as consisting of groups of firms that follow similar strategies and have the same key decision variables. These groups are then called strategic groups. Intra-industry mobility and its barriers rest on the same theoretical foundations as entry and exit barriers. An industry may have only a single strategic group and a group can consist of only one firm.

The existence of strategic groups in an industry changes the way we have to study it. Instead of thinking of entry barriers to the industry we must think of barriers of entry to different strategic groups. It is probably the case that entry into the strategic group of full-line, nation-wide insurers is more difficult than to the strategic group of small, narrow-line insurers. Thus, when interested in the causes of a firm's profitability, we must distinguish between market, group and firm effects. By this I mean that there are variables such as industry growth that affect all firms in a given industry and variables such as the amount of vertical integration that make

different groups within an industry respond differently to exogenous variables.

Instead of talking about entry into an industry and how it affects competition we must talk about combined entry and mobility into a strategic group. Thus different strategic groups may well be in different positions concerning potential competition. Different groups have different barriers that protect them against entry and different barriers that protect them against intra-industry mobility. A potential entrant must then choose not only whether to enter the industry or not, but if entering, then into which group. We can rank the groups within an industry according to their mobility barriers. The groups lower in the ranking may benefit from the "protective umbrella" of barriers to entry created by groups ranking higher. It is also likely that those groups ranking lower will suffer more from tighter competition than higher ranking groups.

At least some empirical studies of strategic groups have been made (Porter 1979, Newman 1978, Oster 1981). All these studies supported the theory, finding that profit differences within industries are greater than those between industries. Caves and Ghemawat (1992) have attempted to identify mobility barriers by using a data-base that contains information not only on firms but on their conjectures on their rivals. All the previous empirical studies on strategic groups or mobility barriers use cross-section data and attempt only to test whether there are strategic groups and if there are, how their profits differ. In this study I go a step further and

estimate how the two identified strategic groups compete, both within and between the groups.

Within the period under study two simultaneous mergers happen within the leaders group (the groups are defined in section 3). As long as the industry is not maximizing industry profits or the competition is of the Bertrand type, theoretical models usually give a prediction that a decline in the number of firms, everything else held equal, raises the (average) industry profits⁴. The theory of strategic groups implies that mobility barriers can be asymmetric in the sense that it can be possible for the "higher" positioned group's firms to compete more effectively with the firms in the "lower" ranking groups than vice versa. This would imply, then, that the mergers had different effects on the different groups. The leaders group is possibly better positioned to collude and the fringe might be under more competitive threat from the leaders group.

One aim of this study is to measure the competitiveness of the Finnish non-life market. As pointed out earlier, the market consists of several submarkets and, in addition, can be divided along geographical criteria. In order to be able to make inferences of the competitiveness of the whole market, an approach that allows for these features was needed. An additional problem in comparison to the traditional IO approaches is the

⁴ As some theoretical literature suggests, the profits of merging firms can decline (see Salant, Schwitzer and Reynolds 1983).

apparent product differentiation (see e.g. Mueller 1977) found in the market. Because of the product it was felt that a dynamic model was an essential prerequisite for the study. In my opinion, all these preconditions are filled by the persistence of profits approach. The persistence of profits approach has so far been used in an inter-industry context. But as any bigger sample of industries will contain firms in other industries that in reality do not pose a source of competition, real or potential, to each other, the results of such estimations are probably biased upwards (ie. underestimate the competitive pressures). As some recent theoretical research (e.g. Lambson 1992) points out, in the presence of sunk costs, even perfectly competitive markets can have different profit levels. The implication is that better results can be obtained by carefully picking the industries that are included in the estimations.

There are several ways to model the persistence of profits or competition over time, but the central feature of all these models is that the profits under or over the long-term average should gradually adjust towards this average. The main problem in modelling this process is that many forces affecting the competition are latent. It is difficult to measure the effects of potential competition, for example. The mere threat of an entry can be enough to discipline the operating firms. These effects should be captured in the model, if we want to measure the true effects of competition. I use a model developed by Geroski (1990; see also Cubbin&Geroski 1987, Geroski&Masson 1987 and Geroski, Masson&Shanaan 1987). The idea of the model is to capture the various effects of actual and potential

competition in two terms, which are called "entry" and "mobility". It is assumed (in the original model) that entry is attracted by higher than average industry profits and higher than average firm-level profits. Mobility is created when a firm's competitors enter its market segment after realising that the firm makes above average profits. Profits are affected through these two forces. Depending on firm (and industry) characteristics, there can exist a level of profits that is unaffected by competition. These persistent profits can be due to productivity differences or strategic behaviour. Starting from simple autoregressive equations (of order 1) modelling the way higher than average profits attract entry and mobility, Geroski ends up with equation (1):

$$(1) \quad p_t = \alpha_0 + \alpha_1 F_{i,t-1} + \alpha_2 G_{i,t-1} + \omega_{i,t}$$

where

p_t = the difference between profits of firm i and the industry average in period t

α_0 = the level of persistent profits, measured as a profit difference to the industry average

F = the difference between profits of i and its group in period $t-1$, ie. the within group profit difference

G = the difference between the average profits of firm i 's group I and the industry average in period $t-1$, ie. the between groups profit difference

$\omega_{i,t}$ = zero-mean, white-noise residual

$F_{i,t-1}$ attracts competition into firm i 's market segment from other firms in the same strategic group, whereas $G_{l,t-1}$ attracts mobility into the group l , where firm i operates. It is thus assumed that the previous period's profit differences can act as a proxy for the various signals that attract competition, ie. that the profits follow an autoregressive process of order 1. The constant α_0 has an expected value of zero. When interpreting the results, it should be remembered that the variables are profit differences, not profit levels. Zero persistent profits mean that a firm/group/industry has neither super- nor supranormal profits.

The novelty in this study is not in the model used, but the way in which I use it. Instead of aiming to measure inter-industry differences, I am able to concentrate on intra-industry differences in the profit-adjustment process. Thus what in Geroski's model is industry, is in this study a strategic group and what in the original model is the whole economy (or the industries included in the study) is here the industry. A further innovation is the way the model is estimated. Geroski (and all the other previous studies on the persistence of profitability and mobility barriers) estimates them on firm level data, using time-series methods. I use panel data methods and can thus obtain a single equation containing both strategic groups (plus firm level effects besides that). I am thus able to compare directly the effect of inter-group rivalry on different groups as well as persistent profit levels and the intensity of the intra-group rivalry. Concentrating on group-level differences also relaxes requirements placed on data. Firm-level time-series estimations would require a longer

observation period, and would still not allow easy inclusion of firms which exit early. The benefits of the adopted approach are not entirely without costs, however. Each firm in a group is restricted to having same slope coefficients, and firm-level differences are captured by the firm-effect of the two-way fixed effect model. Thus, here I am more interested in the behaviour of the groups than in the behaviour of individual firms, as the previous authors have been. There are some econometric restrictions to the model, too. They are discussed in more detail in the appendix, and shortly in the text.

Although the model was created to estimate competition in an inter-industry sample, nothing prevents its use in the current purpose. The model only presupposes that the firms in the sample can meaningfully be divided into two groups. Another possible point of criticism is the applicability of the model to study insurance markets. This is not a problem either. The signalling power of short-run profits in the insurance market is likely to be smaller than in other markets due to acknowledged stochastic fluctuations. This does not mean that they do not have any signalling power and these fluctuations should be captured in the error term and in the period effects (period specific deviations from the constant term) of the two-way fixed effects model used in the estimations. The profits would have no signalling power if they were purely random and we could not infer any information on the next period's profit from those of the previous periods. If this were the case, there would hardly be any market in insurance because the risks would not be manageable. The firms

offering insurance have to be able to predict with some degree of accuracy the amount of losses they will have to incur in order to provide such insurance.

I augmented the original model so that the new model included both groups and allowed me to get different coefficients and a different constant term for the two groups. I achieved this by naming the fringe the "basic" group and introducing a dummy for the members of the leader group, and the according interaction terms. This model (2) is my basic model.

$$(2) \quad p_t = \alpha_0 + \alpha_1 F_{i,t-1} + \alpha_2 G_{1,t-1} + \beta_1 DF_{i,t-1} + \beta_2 DG_{1,t-1} + \varepsilon_{i,t}$$

where

D = a dummy, value 1 for members of the leader group, 0 otherwise

DF = an interaction term between D and F

DG = similarly between D and G

$\varepsilon_{i,t}$ = zero-mean, white-noise residual

The dummy D has to be excluded from the estimations, since there is no firm-wise variation in it. The firm-wise (and group-wise) differences in the levels of persistent profits are measured by the firm-wise deviations from the constant in the two-way fixed-effects estimations. To be able to study the effects of the two leader-group mergers, equation (2) has to be augmented further. Luckily, the mergers took place in the same year, 1983, so I only need to divide the data into two subsets. I assume that the year

1984 is the first year when the new firms operated, ie. that neither of the two pairs of firms changed their behaviour (significantly) prior to the merger. The augmentation is done by introducing a dummy M_t . It gets the value 1 for $t=1984, \dots, 1991$ and zero otherwise. In addition, I formed interaction terms between M_t and the other variables. The test on merger effects will then be a series of F-tests, restricting the coefficients of M_t and the interaction terms to zero one at a time and then different combinations of them, ending with an F-test on restricting them all to zero. The estimated equation (3) will thus encompass equation (2) and all the other possible combinations of restrictions and has the following form:

$$(3) \quad p_t = \alpha_0 + \alpha_1 F_{i,t-1} + \alpha_2 G_{i,t-1} + \beta_1 DF_{i,t-1} + \beta_2 DG_{i,t-1} + \gamma_0 M_t + \gamma_1 MF_{i,t-1} + \gamma_2 MG_{i,t-1} + \delta_0 MD_t + \delta_1 MDF_{i,t-1} + \delta_2 MDG_{i,t-1} + v_{i,t}$$

The estimations shed light on whether the mergers had an effect on the average persistent profits of each group, on the degree of intra-group and inter-group competition. The division of the sample into two subperiods also gives the opportunity to make statements on the degree of competition, since without these merger-estimations no appropriate scale would have existed. The previous persistence of profits-studies give only limited guidance on this question, since they all estimate competition in and between a big number of industries. Also, it should be noted that various industry-related factors probably make it inappropriate to compare a financial market and manufacturing industries. As an example of such effects I could mention solvency regulation that imposes a (long-term)

lower-bound constraint on the profits of insurance companies, because growth of the firm (=premium income) has to be accompanied by a growth in assets to maintain the minimum solvency requirements.

As the observation period is relatively long (1970-1991) and incorporates considerable fluctuations in overall economic activity, it is very much possible that the level of persistent profits fluctuates within the observation period. Also, it is possible that there are within-group differences in the level of persistent profits. To allow for these, I used a two-way fixed effects model in the estimations. The firm-wise fixed effects measure the deviations from the group-mean of persistent profits and the period effects the period-wise deviations from this same mean. The constant term in equation (3) is thus divided into three: the overall constant, the firm-wise deviations from it and the period-wise deviations from it (see Hsiao 1986).

One of the econometric restrictions is easily explained, and it is more a question of definition than a restriction that results from the way the model is built. As the sample consists only of firms in the same industry, the overall constant term, shared by all firms, must be zero (=insignificantly different from zero). If it were positive, then this would mean that the firms on average had persistently higher profits than the market average, a clear impossibility. The other one is not as intuitive: it turns out that β_2 has to be (insignificantly different from) zero, too. The latter restriction's economic interpretation is that the model does not allow for asymmetric competition between the two groups. These constraints are

not imposed on the empirical model, but both the constant and β_2 are allowed to take any value. It turns out, however, that the results are in line with these restrictions, as will become clear in section 5.

II.3 The market

The market under study is a part of the Finnish insurance market, and the period under study is 1970-1991. A supply-side definition of the market is used: those firms that actually do, or are potentially capable, of competing in non-life insurance are included. Since entry into the non-life market is regulated (see later), this definition leads me to include those firms who have a licence to provide one or more lines of non-life insurance (again, see later), and to exclude all others. This definition allows me to include the diverse submarkets that exist under the non-life insurance heading. The whole Finnish insurance market can be divided into three pieces: pension insurance provided by law⁵ (52.6 % of premiums in 1989), other pension insurance and life insurance (8.1 %) and non-life insurance (39.3%). This study concentrates on non-life insurance, but some parts of it will be left out. Since one of the aims of this study is to evaluate how competitive the Finnish insurance market is, a market had to be identified. To achieve this, both foreign direct insurance and all reinsurance were left out and only domestic direct lines were studied. Reinsurance was subtracted because I

⁵ In most other countries, this type of insurance is totally in the hands of the state. Thus the relative importance of the non-life submarket is bigger than the share of premiums indicates.

wanted to estimate the competitiveness of the direct insurance market which is the market where "normal" customers operate. Besides this, the reinsurance market is much more affected by international competition than the direct insurance market, which by and large is totally isolated from foreign competition (with the exception of big customer companies in the recent years. They have been able to establish their own captives abroad or otherwise increase the potential competition). Non-life insurance was chosen since it has the biggest number of firms, relatively little regulation and it is the submarket where most free consumer choices are made.

Isolating pension insurances and life assurance from other insurance is not difficult since, due to legal restrictions, different companies are active in the pension insurance and life assurance markets on the one hand and on the non-life insurance market on the other. Reinsurance and direct international insurance proved a little more troublesome since they are sold by the same companies as domestic direct insurance. The Official Statistics, however, contain almost all the information needed by the line of insurance and thus I was able to separate these lines with reasonable accuracy. The piece of the Finnish insurance business under study is thus only a relatively large slice of the whole market. The actual size of the part of the non-life market under study was 10 billion FIM in 1989. The part of the non-life market that is studied here is however not a homogeneous piece. It consists of several submarkets, which are listed in table 1, which also gives information on the relative size of these lines. From now on I

will refer to this part of the non-life market as "the non-life market". Two of the three biggest of these lines are regulated. They are the compulsory (third party) motor and worker's compensation insurances. I will go through the main points of this regulation here⁶. The two most important features are:

SUBMARKETS OF THE FINNISH NON-LIFE INSURANCE MARKET	
line of insurance	share of total premiums
statutory workmen's compensation	19.81%
other accident	5.40%
compulsory motor	23.62%
motor vehicle	14.14%
hull	1.82%
cargo	3.63%
fire and other combined property	23.32%
loss of profits	2.00%
forest	0.30%
third party	3.06%
credit	1.56%
other	1.34%
foreign direct ins. of all above	11.5%

table 1

⁶ There is a separate appendix describing the regulatory environment in more detail.

- (1) the regulatory body has great powers when it comes to entry. Entry can be denied without explanation. Foreign firms are barred from the compulsory lines.
- 2) the prices are regulated. Earlier they were the same for all the firms, but recently firms have been allowed to seek discounts from the calculated price. The regulatory price includes a return for equity.

Besides this by-line regulation the regulatory body sets limits to firms' solvencies. Although the non-life and other insurance lines are separated and different legal entities operate in different submarkets, it is easy to form groups of firms which sell all the lines. During the period of research the market has been dominated by groups which all provide all the lines of insurance. Some of the smaller firms have formed such groups too, but not all. Entry to the non-life market has been relatively little, and only one foreign firm has entered during the observation period (in 1989).

The Finnish non-life market has been very stable and highly concentrated for the last two decades as can be seen from the figure 1. It displays the Herfindahl index. The only considerable jump in the Herfindahl-index is in 1983/84, when two mergers occurred. The concentration jumped from a level equivalent to 7 equally sized firms (1400 points) to a level of 6 equally sized firms (1600 points). During the 70's there was some movement in the market, but during the 80's there was only one entrant to the market, in 1989. In 1990 there were two additional entries of which

HERFINDAHL INDEX
direct domestic non-life insurance

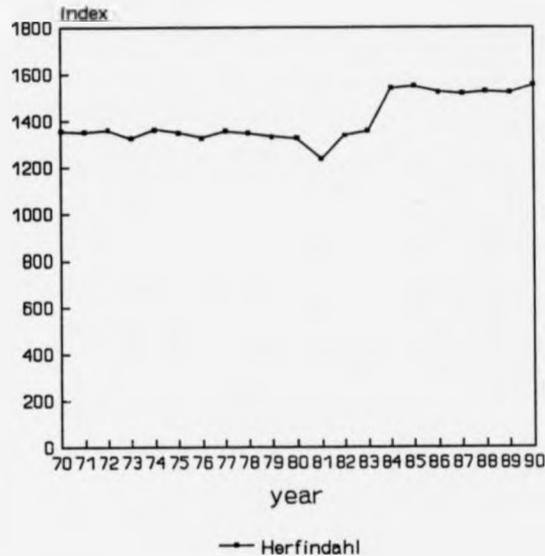


figure 1

only one started to do business (but stopped doing business after only a year), the other started to do business in 1991. These, however, are not taken into consideration in the estimates made, since the persistence of profits model needs observations from at least two periods. It should be noted that all exit has happened through mergers. The number of the firms active in the market has thus been relatively stable but slowly declining⁷.

⁷ A small number of fringe firms had to be excluded from the estimations. The reasons for excluding them were: 1) having only small direct non-life activities, 2) being in the market effectively only for one period 3) being a so-called captive insurer. According to industry sources, the captive firms exist mainly for tax- etc. reasons and do not affect market behaviour, 4) insurance associations and the two foreign firms were excluded because of their minor importance and lack of adequate data.

All but the two mergers, which caused the jump in the Herfindahl-index are assumed to have a negligible effect on market behaviour.

The Finnish insurance market in general is well protected from outside competition and from the entry in general. All entrants need a licence, which is given by the line of business. Foreign firms cannot get a licence for compulsory insurances. Foreign firms need to establish a subsidiary or deposit a given sum in Finland in order to be able to enter. As the regulatory body can in effect decide whether to allow the entry or not when it comes to compulsory lines, it in effect controls the entry to the market as a whole. Since all compulsory lines make 70% of all direct premiums and 45% of the direct premiums paid in the non-life market, declining entry to these markets forces the entrant (or current fringe firm) to the fringe. As both the industry representatives and the regulators have stated, insurance is often bought in "packages", not by shopping separately for every individual insurance. Thus the policy of regulators not to allow all firms to sell compulsory lines of insurance can amount to creating (artificial), possibly vertical, product differentiation in the market.

Many small domestic firms have access to a broad line of products by contracts with the big insurance groups. Although this makes them able to offer all or most types of insurance, their strategic position is still different from the big insurance groups. One of the peculiarities of the Finnish market is the (until recently) complete absence of brokers. This is an additional feature that heightens the entry barriers, since a comprehensive

distribution network is likely to be the biggest sunk cost in the market and sunk costs are a deterrent to entry (see e.g. Sutton 1991).

II.4 Strategic groups and profits

Using the theory of strategic groups it is fairly easy to divide the Finnish non-life market into two groups, which I will subsequently call "the leaders" and "the fringe". As a basis for this division I used the following criteria:

- a wide (own) distribution net⁸
- membership of a group providing other lines of insurance in addition to non-life lines
- breadth of supply of the non-life products
- market share
- a licence to provide both compulsory lines of non-life insurance

These criteria were chosen to approximate differences in the firms' strategies, based on (more or less) sunk investments. In order for a firm to be placed in the leader group, I required that it fulfills at least four of the five above criteria. With regard to the first criterion, there is a clear threshold in the industry: Five (seven before the mergers in 1983) firms

⁸ By a wide distribution net I mean that the firm's branch network covers the whole country.

have a branch network of over 30 branches, whereas the next biggest branch network consists of only 17 branches (1989 figures). These seven firms all provide the whole range of life-, non-life- and pension insurances, together with an eighth firm that differs from the seven in that it has only one branch. Some small firms also offer the whole product range, but usually in alliance with one of the bigger firms, not independently. The above mentioned eight firms also provide all lines of non-life insurance, as categorized in table 1. None of their rivals does that.

Size provides, together with the branch network, the clearest division among the firms: the eight firms that provide all non-life lines and the whole range of other than non-life lines are bigger than the rest. There are differences among the eight firms, the smallest having a market share of ca. 5% and the biggest covering round a quarter of the market. These eight firms also hold licences to sell both compulsory lines of non-life insurance, together with a number of smaller firms.

The criteria divide the market clearly into two groups. The so-called leader group consists of 8 (pre-merger, ie. 1970-1983), respectively 6 (post-merger, ie. 1984-1991) firms that all but one satisfy all five criteria, and the eighth satisfies all other criteria but the branch-network one. Although there are significant differences between these leader-group firms, they are clearly more homogenous than the rest of the firms. The fringe consists of firms that are very different: some sell a comparatively wide range of products through their own distribution nets, some are industry captives, some are

specialised into selling only a few lines of insurance. It is very likely that some of these firms compete more against the leaders than against each other. Despite this, because of the mobility barriers, they are fringe firms in their chosen submarket(s). The structure of the non-life market resembles closely the structure studied in price-leadership models, where the dominant incumbent is replaced by a group of colluding incumbents.

The tight entry regulations and barriers are likely to make the pressure of potential competition negligible. If such threats are felt, they should be stronger in the fringe, since foreign entry (and domestic entry with an equally high probability) happens into the fringe. By the same logic the competitive pressures of the leaders come from among themselves and from the fringe. In the econometric model I treat the market as if there was no threat of entry. This assumption is necessary in order to enable me to use data from the Finnish non-life market only. If this assumption creates bias in the results, it should be greater in the fringe estimations than in the estimations for the leaders. The other reason for neglecting the entry threats is that besides other insurance markets, it proved difficult to find out the potential sources of entry. The Finnish banking industry could seem an appropriate choice, but the banks⁹ have been prohibited to enter the insurance market. The effects of by-line regulation are to some part

⁹ Holding companies which would own both banks and insurance firms have been allowed. No such firms have existed during the period of study. One is about to be created, but it will concentrate on life- and pension insurance in its insurance activities.

entry barriers, effectively blocking a large part of the total insurance market from foreign entry. By-line regulation is probably limiting also competition between existing firms, since those firms without a licence to provide compulsory lines have a disadvantage with respect to customers who prefer to buy their insurances as a package.

The aim in calculating the profits was to get as near as possible to the true yearly result of an insurance firm¹⁰. In addition, I tried to trace out a way of measuring profits that would resemble the economic models of insurance, where the basic formula is very simple:

+ PREMIUMS
+ INVESTMENT (NET)INCOME
- CLAIMS INCURRED
- OPERATING COSTS
PROFIT

It turned out that a profit measure used by the industry, the so-called stochastic result, is very close to the definition above. As the firms used in this study are not the legal entities, some markets are not included and since I had to clean out the effect of some accounting conventions, the

10 Almost all needed information is published in Finnish Official Statistics: Insurance, and the rest was helpfully provided by the Insurance Office of the Ministry of Social Affairs and Health.

profits used in this study could not be picked up from the financial statements directly but had to be calculated¹¹. The firms in this study are different from legal entities since several insurance firms have insurance subsidiaries, and these have been summed together with their parent companies. It is not always clear when a merger has de facto happened, since it seems that in many cases the merging firms were operating closely together before the legal merger took place. I have followed the only clearly verifiable strategy and summed two firms together only after the other has been bought by the other, or the firms have merged. Some of the firms are stock, some mutual firms, and a few change ownership form (from mutual to stock) during the observation period. Two measures of profit were used. The rate of return on equity (ROE¹²) and the price-cost-margin (PCM¹³) were both calculated without the smoothing effect of the so-called equalization reserve. The equalization reserve is designed to smooth out the changes in claims paid. When it is left out, the true profit of the period can be calculated. These profit measures were then used to calculate the profits from the whole portfolio, which includes the investment income and the insurance portfolio, from which the investment income has been subtracted and which thus displays the profit that comes from the "pure" insurance activity. Table 2 presents information on profitability and its variability in the Finnish

11 There is a separate appendix explaining in detail how the profits were calculated, which firms were excluded from the estimations, which firms were counted as a single entity etc.

12 By ROE I mean profits divided by equity capital, appropriately defined.

13 $PCM = \text{PROFIT} / \text{PREMIUMS}$

DESCRIPTIVE STATISTICS OF PROFITS IN THE FINNISH NON-LIFE INSURANCE INDUSTRY						
1970 - 1991 weighted averages						
profit measure	period	strategic group	whole portfolio		insurance portfolio	
			avg.	std.	avg.	std.
ROE	70-83	fringe	9.11	175.05	-45.24	191.09
		lead.	58.42	67.66	-38.97	85.15
		ind.	24.08	150.76	-43.34	166.23
	84-91	fringe	13.12	142.67	-53.86	127.55
		lead.	30.27	67.17	-37.08	63.16
		ind.	18.26	124.91	-48.82	112.18
	70-91	fringe	9.84	178.33	-51.17	187.16
		lead.	50.01	69.48	-39.96	81.11
		ind.	21.99	153.76	-47.78	162.55
PCM	70-83	fringe	14.06	30.66	-3.71	23.74
		lead.	10.02	9.18	-4.17	9.07
		ind.	12.84	26.09	-3.85	20.44
	84-91	fringe	19.46	31.18	-4.28	29.89
		lead.	8.64	22.26	-7.49	21.40
		ind.	16.21	28.80	-5.24	27.62
	70-91	fringe	15.01	31.88	-4.03	26.69
		lead.	9.89	15.45	-4.95	14.70
		ind.	13.47	27.95	-4.31	23.71

table 2

non-life insurance market. The numbers in the table are a weighted average of the firm-wise averages over the 22-year-period of study. The standard deviations presented are the weighted averages of the firm-wise standard deviations calculated. I used the number of periods that a firm

i was in the market as weights. As the table shows, the profits are fairly high, with the highest ROE measures over 50%¹⁴. The standard deviations reveal that the fringe firms' average standard deviations are higher than the leaders', and that the standard deviations of the insurance portfolio are higher than those of the whole portfolio, due to the use of hedging possibilities. The investment income figures are unfortunately suspected of having a large bias, since they have been taken from the statements of income at face value and the Finnish accounting regulations allow considerable spread between the true returns and the returns shown. To be more specific, firms are allowed to accumulate so called hidden reserves since they do not have to report the possible rise of the value of their investments. If the value of investments goes below those reported, the change has to be reported. Thus these figures cannot be viewed as totally accurate and as firms have considerable hidden reserves, the figures are likely to be near a lower bound of estimates. The insurance portfolios earn a negative return on average. The two profit measures put the two groups in different ordering for both portfolios. The difference between ROE and PCM figures can be explained by differences in the relative amount of equity in the two groups. This, and the fact that some of the firms are mutuals and thus are not so concerned about the return on equity, lead me

14 The profits are calculated before depreciation and taxes. Depreciation plays a minor role in insurance, however. This is apparent from the following calculation: the total depreciations of the industry in 1989 accounted for 1.90% of the premiums. If this is taken as a representative figure, it would lower the long term (=1970-1991) ROE to 8.59% for the fringe and 40.4% for the leaders.

to choose PCM as the profit variable for the estimations. Although the mutuals probably do not maximize profits, they still need a nonnegative PCM for the whole portfolio in order to stay in the market¹⁵. Despite these qualifications, the ROE measures have their own interest in a market as regulated as the Finnish non-life insurance market. The whole portfolio figures are suspect, since the reporting of investment income (=the rise in value of assets) is dependent on the firms' decisions. If the fringe firms have a relatively larger share of their investments in assets that yield returns but no rise in value (bonds etc. as contrasted with shares and real estate) than the leaders, then this could lead to the figures presented. This difference is clearly evident when the firms themselves report their liabilities to capital ratios. Using book values for investments, these ratios are mostly in the region of 50%, but using true (market) values, these are doubled (or even quadrupled), and the difference seems to be bigger for the larger firms. No such discrepancy can exist in the insurance portfolio figures, however. It should also be remembered that these figures display the profitability of the direct, domestic insurance and the total profitability of the industry or a particular firm can be different. The persistent high profitability of the market can be viewed as evidence of entry barriers, either structural or regulatory.

15 Mutuals might need a positive PCM due to regulations in order to keep their so-called solvency-ratios on a required level also when facing a growing market.

In the estimations I used the profits from the insurance portfolios. The rationale behind this was that it is the the profit of the insurance portfolio that is a good indicator of profit opportunities in a given submarket. Also, investment income, which is the difference between the total portfolio and the insurance portfolio profits, feeds into the profit levels of the insurance portfolio. More efficient investors can lower their insurance portfolio price-cost-margins and attain the same whole portfolio profits as their competitors and thus send a exit/no entry signal to the competitors who are less efficient investors. So differences in investment performance are not neglected, but they affect the profit levels of insurance portfolios and also the attractiveness of different market niches and thus the signalling effect of profit differences indirectly.

II.5 Hypotheses and estimation results

In light of the description of the market given in the previous section, it seems that the market is divided into two different groups and that there are mobility barriers between these groups. The model (3), displayed here again,

$$(3) \quad p_t = \alpha_0 + \alpha_1 F_{i,t-1} + \alpha_2 G_{i,t-1} + \beta_1 DF_{i,t-1} + \beta_2 DG_{i,t-1} + \gamma_0 M_t + \gamma_1 MF_{i,t-1} \\ + \gamma_2 MG_{i,t-1} + \delta_0 MD_t + \delta_1 MDF_{i,t-1} + \delta_2 MDG_{i,t-1} + v_{i,t}$$

was estimated on the firm-wise profit data from 1970-1991 that formed an unbalanced panel due to firms exiting and entering. The following hypotheses will be tested:

a. hypotheses concerning competition

H₁: The leaders, protected by mobility barriers, will have higher or at least as high persistent profits as the fringe. The group-wise level of persistent profits (as a deviation from the industry average) is measured as the group-wise average of firm-level fixed effects in the two-way estimations. The persistent profits of the fringe should be negligible or negative.

H₂ : The leaders are in a better position to collude than the fringe. The coefficient of *DF* should be significant, but its sign is ambiguous, since it is not clear what kind of a coefficient represents collusive behaviour (it could be negative, indicating some kind of a punishment strategy, for example). The fringe should be under more competitive pressure than the leaders, indicating a low value for α_1 . As stated earlier, it can be that the fringe firms do not so much compete with each other as with the leaders. This would lead to a high value of α_1 . Since it is expected that these two forces balance each other, a positive value for α_1 is hypothesized.

H₃: The leaders do not collude with the fringe, but the firms of the two groups compete against each other. Thus, *G* is expected to get a small or insignificant coefficient. Due to modelling restrictions, *DG* should get an insignificant coefficient.

b. hypotheses concerning the effects of mergers

H₄: Both groups should benefit from the mergers in that the level of their persistent profits rise, or do not decline. This effect could be stronger for

the leaders. The prediction is, then, that both γ_0 and δ_0 , the coefficients of M and MD , get a non-negative sign.

H_5 : Since the two mergers are both within the leaders group, they should have a minor effect on the fringe group's intra-group competition. If, on the other hand, the mergers make it more difficult for the fringe to compete against the leader group, then the mergers should lead to a tightening of intra-group competition in the fringe. This means that γ_1 should be insignificantly different from zero, or have a small negative value. The case of the leaders is more difficult. It is not clear what value the intra-group coefficient should have if the firms collude, neither the direction of change. The effect should be at least as big as for the fringe, however, indicating a δ_1 equal to zero, or having the same sign as γ_1 .

H_6 : Because the mobility barriers can be asymmetric, the mergers should have no effect on the competitive pressure the fringe firms exert on the leaders group. This results in a predicted value of zero for γ_2 , the coefficient of MG . Neither should the mergers have any great impact in the way the leader firms compete with the fringe group, so an insignificant coefficient is expected for δ_2 , the coefficient of MDF , too.

The estimations¹⁶ were started by estimating the unrestricted model (3). Next I estimated the basic model (2) and tested it with an F-test against (3). Then I added, one at a time, one of the merger-variables (M , MF , MG , MD , MDF , MDG) to (2). These six models were then tested against (3). The

16 All estimations were done with LIMDEP 6.0 (Greene 1991).

model that survived was (3b). Then I proceeded into augmenting this with other merger variables and testing these augmented models against the restricted version. It was tested against (2), too. As it turned out (see appendix I at the end of the chapter for the F-tests), this model survived against all other models, since no other merger variable could be added to (3b), neither could *MF* be taken out without worsening the fit. The results for the estimations of the final equation¹⁷

$$(3b) \quad p_t = \alpha_0 + \alpha_1 F_{i,t-1} + \alpha_2 G_{i,t-1} + \beta_0 D + \beta_1 DF_{i,t-1} + \beta_2 DG_{i,t-1} + \gamma_1 MF_{i,t-1} + \tau_{i,t}$$

are presented in table 3¹⁸.

The following results were obtained:

H_1 : The group-wise average deviations (weighted by the number of periods each firm is present in the market) from the common constant turned out to be insignificantly different from zero for both groups, as was the overall constant α_0 . The insignificance of the overall constant is in line with the restrictions that the adopted approach places on the econometric model (the value of α_0 is -0.096 and its probability value 0.924). There are

17 The equation (3b) was tested for the assumption of first (and higher) order autocorrelation by regressing the residuals on lagged residuals (up to four lags) and original variables. F-tests clearly rejected any form of (up to fourth order) autocorrelation.

18 The model assumes first order autocorrelation. That this is a valid assumption was checked by retaining the errors, and regressing them on lagged values using 2SLS. As the results take a lot of space, and higher order autocorrelation was rejected, the results are not presented here.

PERSISTENCE OF PROFITS EQUATION (3b)	
TWO-WAY FIXED EFFECTS MODEL	
(3b) $p_t = \alpha_0 + \alpha_1 F_{1,t-1} + \alpha_2 G_{1,t-1} + \beta_1 DF_{1,t-1} + \beta_2 DG_{1,t-1} + \gamma_1 MF_{1,t-1} + \tau_{i,t}$	
variables	value
α_0	-0.096
std. error	1.010
α_1	0.543***
std. error	0.058
α_2	-0.726
std. error	0.953
β_1	-0.645***
std. error	0.156
β_2	1.032
std. error	1.356
γ_1	-0.198**
std. error	0.085
$\alpha_1 + \beta_1$	-0.099***
F-test	54.295
$\alpha_1 + \beta_1 + \gamma_1$	-0.300***
F-test	59.344
$\alpha_1 + \gamma_1$	0.338***
F-test	57.507
test statistics	
R ²	0.442
adj. R ²	0.373
F-stat.	6.368
log-L.	-2205.787
Hausmann test (fixed vs. random effects) df. = 5	8.198 prob. value 0.146
est. autocorr. coeff.	-0.034
*** = sign. at 1% level ** = 5 % level * = 10% level	

table 3

thus no differences in the level of persistent profits at the groups level, although there are differences at firm level.

Here, it should be remembered that the differences in overall average profits, as displayed in table 2, are quite substantial.

H_2 : The within-groups competition is different in the two groups. In the fringe, intra-group competitive pressure is small, reflected by a value of 0.543 for α_1 , indicating that the fringe firms have chosen different submarkets. The leaders have a small negative coefficient $(\alpha_1 + \beta_1)^{19}$ of -0.102, which in my opinion points to the possibility of collusive punishment strategies. It must be remembered that a negative coefficient means not only that above average profits disappear immediately, but also that below average profits are allowed to rise above the group average. The result could be, however, consistent with tight competition as well, and therefore a precise interpretation of the result cannot be given.

H_3 : The between-groups competition seems tough, as G gets an insignificant coefficient. Also DG gets - as it should - an insignificant coefficient (the value of DG 's coefficient β_2 is 1.032 and its probability value 0.447).

b. hypotheses concerning the effects of mergers

H_4 : The mergers proved to have no effect on persistent profits in either group, as exclusion of both M and MD from the final model indicates.

19 I have performed F-tests for the significance of the combined coefficients needed for the testing of some of the hypotheses. These summations are shown in table 3, with F-test-values in place of standard errors.

H₅: The intra-group competition was affected, and to the same degree in both groups, as *MDF* was dropped from the final version of equation (3). The coefficient of *MF*, γ_1 got a value of -0.198, indicating a considerable tightening of the within group competition in the fringe (the coefficient goes from 0.543 of *F* to 0.345 of *F+MF*) and (if we stick to the punishment-interpretation of a negative coefficient), an increase in the severity of punishment in the leader group, as the leader coefficient goes from -0.102 (*F+DF*) to -0.300 (*F+DF+MF*).

H₆: Here, values of zero were predicted and insignificant values obtained. These resulted in the exclusion of both *MG* and *MDG* from the final model. This is in accordance with the hypothesis.

II.6 Summary

This chapter employs the theory of strategic groups and a model of persistence of profits. This combination suits especially well the market under study, the Finnish domestic non-life industry. The first part of the setting sits well because the market can be divided into two distinctively different groups, the other because the market consists of several submarkets, and because product differentiation, regulation and price discrimination escape more formal models.

In order to be able to use the persistence of profits model, it was assumed that there is no outside threat of competition. While this certainly is a simplification, I hope that the data on the market has convinced the reader

that it is not so far from the truth. There is some theoretical backing to use this type of models with a more narrowly defined set of industries than has been the case so far. Great effort was put into calculating the true yearly profits, as these were not directly available. The accounting profits had to be altered to take into account the exclusion of foreign insurance and reinsurance, the merging of legally separate but functionally same firms, and to calculate the economic (as opposed to accounting) profits of these created firms.

The market was divided into two strategic groups using five criteria. The pre-tax profits turned out to be quite high. Six hypotheses were formed, three of them relating to the within and between groups competition as such, and three to the effects of two simultaneous mergers on competition. As it turned out, all hypotheses were confirmed by the data. It should be remembered that a zero constant does not indicate that profits are low, but is necessitated by the structure of the model. An industry cannot have persistently higher profits than its own average. The fringe firms do not compete tightly with each other, but more with the leader group. The interpretation of the leaders' within group competition coefficient is not unambiguous: it might be that they compete aggressively with each other. But there is also the possibility that the leaders are engaged in some kind of a collusive punishment strategy. This is to my knowledge the first time that a negative coefficient is reported in this kind of a study. The hypotheses of asymmetric mobility barriers was not confirmed: the leaders do not have higher persistent profits. Both groups exert a considerable

competitive threat upon each other, demonstrated by an insignificant coefficient for all inter-group competition variables. The mergers had an effect on competition. They did, however, not result in higher persistent profits. The effect was a considerable tightening of intra-group competition within the fringe, and a change into tighter competition or bigger punishments in the leader group. As was expected, the mergers had no effect on inter-group competition. An additional thing to note is the extremely good performance of the model: it implies that despite fluctuations in claims incurred, competitors are able to deduce the true (non-stochastic) profitability of the different lines of insurance and market niches of their competitors, and react to these.

APPENDIX I

ECONOMETRIC RESTRICTIONS AND SPECIFICATION TESTS

A1.1 Econometric restrictions

The model is based on profit differentials, and the following notation is used:

p_t = difference between a firm's profits and the industry average profit

$p_{i,t}$ = difference between group i 's average profit and industry average profit, i
= L,F

n_i = number of firms in group i , $i = L,F$

n = number of firms in the industry

Equation (1) implies for a fringe firm (suppressing the error term) that

$$(A.1) \quad p_t = \alpha_0 + \alpha_1(p_{t-1} - p_{F,t-1}) + \alpha_2 p_{F,t-1}$$

If this is summed over all firms in the fringe, we get

$$(A.2) \quad \sum_F p_t = n_F \alpha_0 + \alpha_1 (\sum_F p_{t-1} - n_F p_{F,t-1}) + \alpha_2 n_F p_{F,t-1}$$

The parenthesis disappears by the definition of the variables. Dividing by n_F we get

$$(A.3) \quad p_{F,t} = \alpha_0 + \alpha_2 p_{F,t-1}$$

Going through the same process for the leader group we get

$$(A.4) \quad p_{L,t} = \alpha_0 + (\alpha_2 + \beta_2)p_{L,t-1}$$

If these are multiplied each by the number of firms in the group in equation and added up, the result is

$$(A.5) \quad n\alpha_0 + n_L\beta_2 p_{L,t-1} = 0$$

Since this has to hold for all t , we know that α_0 is zero by definition, n_L is positive and that $p_{L,t-1}$ is nonzero and not constant, then β_2 has to be zero. It is the coefficient of DG , the variable measuring the asymmetry in inter-group competition. The restriction means that the model does not allow for asymmetric inter-group competition. Furthermore, by solving for the long-run levels of $p_{i,t}$ (=assuming that $p_{i,t} = p_{i,t-1}$) from equations (A.3) and (A.4), it is clear that these are zero because of the value of α_0 . Thus the model assumes that in the long-run, the profit differences disappear. This implies that in the long run, firm-level profits differences vanish as well.

A.I.2 Specification tests

The unrestricted model is:

$$(3) p_t = \alpha_0 + \alpha_1 F_{i,t-1} + \alpha_2 G_{i,t-1} + \beta_1 DF_{i,t-1} + \beta_2 DG_{i,t-1} + \gamma_0 M_t + \gamma_1 MF_{i,t-1} + \gamma_2 MG_{i,t-1} \\ + \delta_0 MD + \delta_1 MDF_{i,t-1} + \delta_2 MDG_{i,t-1} + \pi_{i,t}$$

The following models were estimated:

Table A.I.1

model no. restrictions

model no.	restrictions	unrestricted merger variables
I	unrestricted (3)	-
II	$\Sigma \gamma_j + \Sigma \delta_0 = 0, j=0, \dots, 2$	M
III	$\gamma_1 + \gamma_2 + \delta_0 + \delta_1 + \delta_2 = 0$	MF
IV	$\gamma_0 + \gamma_2 + \delta_0 + \delta_1 + \delta_2 = 0$	MG
V	$\gamma_0 + \gamma_1 + \delta_0 + \delta_1 + \delta_2 = 0$	MD
VI	$\gamma_0 + \gamma_1 + \gamma_2 + \delta_1 + \delta_2 = 0$	MDF
VII	$\gamma_0 + \gamma_1 + \gamma_2 + \delta_0 + \delta_2 = 0$	MDG
VIII	$\gamma_0 + \gamma_1 + \gamma_2 + \delta_0 + \delta_1 = 0$	MF, M
IX	$\gamma_2 + \delta_0 + \delta_1 + \delta_2 = 0$	MF, MG
X	$\gamma_0 + \delta_0 + \delta_1 + \delta_2 = 0$	MF, MD
XI	$\gamma_0 + \gamma_2 + \delta_1 + \delta_2 = 0$	MF, MDF
XII	$\gamma_0 + \gamma_2 + \delta_0 + \delta_2 = 0$	MF, MDG
XIII	$\gamma_0 + \gamma_2 + \delta_0 + \delta_1 = 0$	

The unrestricted model I was tested against all other models, with the following results: (d.f. of the denominator are 436²⁰)

table A.I.2

model	d.f. of numerator	F-test value	reject H_0
II	6	0.995	no
III	5	0.136	no
IV	5	1.194	no
V	5	1.113	no
VI	5	1.192	no
VII	5	1.038	no
VIII	5	1.122	no
IX	4	0.169	no
X	4	0.045	no
XI	4	0.170	no
XII	4	0.135	no
XIII	4	0.035	no

the critical points at 5% level for $F[x, \infty]$ are

x	critical point
6	2.10
5	2.21
4	2.37

Then, the less restricted models with one merger variable were tested against the most restricted (II) model. Critical point of the F-test is 2.60.

table A.I.3

restricted model	d.f. of denominator		
II	441		
unrestricted model	d.f. of numerator	F-test value	reject H_0
III	1	0.001	no
IV	1	5.342	yes
V	1	0.407	no
VI	1	0.011	no
VII	1	0.778	no
VIII	1	0.360	no

¹⁹ the d.f for the denominators were taken out of the F-test results that LIMDEP produces. This was by far the easiest way to obtain them, since the panel is unbalanced.

Of these, the restrictions of II get accepted in all other cases but model IV (with merger variable MF). Next, I tested whether model IV can be augmented with any of the other merger variables without worsening the fit of the model. Critical point is 2.60.

table A.I.4

restricted model	d.f. of denominator		
IV	440		
unrestricted model	d.f. of numerator	F-test value	reject H_0
IX	1	0.005	no
X	1	0.504	no
XI	1	0.001	no
XII	1	0.143	no
XIII	1	0.547	no

The restrictions of model IV are accepted against all augmented models. Thus model III, presented in equation (3b) is the final equation.

$$(3b) p_t = \alpha_0 + \alpha_1 F_{t-1} + \alpha_2 G_{t-1} + \beta_1 DF_{t-1} + \beta_2 DG_{t-1} + \gamma_1 MF_t$$

APPENDIX II

REGULATION OF THE FINNISH NON-LIFE INSURANCE MARKET

As was pointed out in section II.3, the non-life insurance market is only a piece of the Finnish insurance market, albeit a relatively big one. Clearly the biggest market is the market for pension insurance provided by law. This type of insurance is in other countries mostly taken care of by the state and in Finland it is part of the social insurance as opposed to private insurance. Pension insurance provided by law, as well as other compulsory insurances, is under the rate of return regulation. I will not go into details here, but I only mention the main point from the point of view of this study: the state has great freedom when considering entry (licence) to this market and foreign firms are and will be barred from entry. This means that the firms that do not have a licence are handicapped since they can not provide a full line of policies, but must rely on cooperation with some pension insurance firm. The biggest ones of these (which control almost the whole market) are tied to the insurance groups that form the leader firms in the non-life market.

The entry to the non-life market is licenced and foreign firms can not get a licence for the compulsory lines of insurance (which are workmen's compensation and motor insurance)²¹. In order to get a licence the applying firm must display adequate funds, and its entry must not hamper the sound development of the industry, whatever that means. The licence may be a general licence or it may be restricted by line and/or geographical factors or by

21 This is going to change due to the implementation of EEA.

possible clientele. The capital requirements are low and they as such do not form a barrier to entry. The biggest regulatory barrier is the possibility of the state to consider whether or not it is necessary to grant a licence for the compulsory lines. Historically, there were several foreign firms active in Finland around the turn of the century, but during the research period only two foreign firms were active and both of these had a very marginal role to the extent that they were excluded from this study (the principal reason being that I did not get enough information of their balance sheets and financial reports). One of these is specialized on insuring Russian property in Finland, the other withdrew its licence in the end of the 80's. A new foreign insurance firm entered in 1989 but its premiums are in the order of 30 million FIM compared to a market total of 10 billion FIM (in 1989).

The two compulsory non-life insurance lines are both rate of return regulated. This means that the prices are calculated by the regulators on the basis of the data from the previous period and a forecast of the market. Earlier all the companies had the same price, but quite recently, discounts have been allowed (and have been applied) after a firm has proved that its costs are so low or its investment returns so high that a discount is financially possible. The rate of return calculated on equity has been under 10% and varying. Anyway it is lower than the profits really earned by the firms (see section 3).

Besides the regulations by line of insurance of the compulsory insurances, the firms are relatively free to act as they wish (in the non-life sector). However, they have to report their policy conditions to the regulator. In addition to the regulation by line of insurance, the main form of regulation is solvency

regulation. This is intimately tied with the calculation of reserves (see appendix B). The firms are not allowed to transact any other business than insurance and, accordingly, there are limits to the ownership of shares in other than insurance companies. In accordance with the limits on ownership, there are limits to the forms of investments the insurance firms are allowed to make. The possibilities to invest in foreign currencies are limited (although here some relaxation has occurred) and the investments must be safe and guarantee a sure return (examples of such could be obligations issued by the state or bonds of municipalities).

The supervision of the industry has three main features (OECD report 1991):

- firstly, the supervision is organized so that the same regulatory body (the Insurance department of the Ministry of Social Affairs and Health) is responsible for the supervision of both social and private insurance.
- secondly, the solvency regulation is based on risk theory and
- thirdly, the ministry is not limited to the regulatory role, but can be (and has been) active in e.g. technical matters of the industry.

It is the duty of the regulator to "safeguard sound and effective economic competition against detrimental restrictive practices and to promote competition". The starting point of the Finnish insurance regulation was, as in other countries, to limit the possibilities of insolvency and to see that no "detrimental" competition emerged. The modern competition promoting ideas have only come through in the late 80's and until now it can be said that the

regulators have acted according to the principal motives that created the regulatory system in the first place.

One interesting feature in the Finnish supervision system is the so-called advisory body. It consists of five experts. They make statements on new potential entrants (who apply for a licence) and on new technical bases for actuarial calculations. These advisors have normally been members of insurance firms active in the market. This means that they get to say their opinion and at the least review information of their the plans of actual or potential rivals. The problem is that almost all the insurance knowledge is in the firms or in the regulatory body and thus the experts are usually from the firms active in the market. The effects on competition cannot be positive.

A large part of the legislation is concerned with the possible insolvency or end of business of an insurance firm, but these aspects are of a limited interest in this study and, accordingly, I will not review them here.

APPENDIX III

HOW TO CALCULATE THE PROFITS OF AN INSURANCE FIRM, FORM
ECONOMIC ENTITIES, AND TO EXCLUDE FIRMS FROM THE SAMPLE

In this appendix I will explain the way the profits in this study have been calculated and how they differ from the profits displayed in the statement of income.

First thing I did was to subtract other than the domestic non-life insurance lines. This meant that foreign direct insurance and both foreign and domestic reinsurance were subtracted. In the case of direct foreign insurance there was not data for all the years, but this should be a minor problem, since vast majority of the foreign business is reinsurance. Luckily, the Official Statistics have almost all the information also by the line, so it was not a problem to subtract, say premium income. On some areas, e.g. business expenses and investments, there was no data available at all or not for all the years. In most cases, I subtracted an amount according to the relative size of the premiums received. This means that if the subtracted lines constituted 30% of firm i's premiums in year t , I subtracted the same portion (30%) from the investment income and other comparable items. The business expenses were a little more problematic, since reinsurance is vastly cheaper to produce than direct insurance. I calculated the so-called expense ratio for the years 70-75 for reinsurance (as the data on business expenses was available for these years). This ratio is normally between 15 and 25 %, but in reinsurance it was never over 5%. When the share of reinsurance in most cases is well below 50%, I decided to take the business expenses at face value and not subtract the an estimate of the business expenses

of the subtracted lines at all. This meant that I had to discard three firms which mainly had reinsurance and only occasionally direct insurance, as they got hugely underestimated profits. This was because my formula included 100% of their expenses but, say, only 5% of their premium income. None of these firms was in the market more than for five periods and their market shares were very low (about 1% or less) so in my opinion no great harm was done.

I will now display the official statement of income and the version I used, which tries to resemble the formula presented here (section 3) as closely as possible.

The formula suggested by the economic theory is:

+ PREMIUMS
+ INVESTMENT INCOME
- CLAIMS INCURRED
- EXPENSES
PROFIT

The official statement of income (this is the current version, in the early 70's it was a little different) and my version are as presented in table B.1. As can be seen from the table, I have adjusted the period result to the fact that only a part of the firm may be included into the estimations. In addition, I have taken into my calculations neither the depreciation nor the provisions or other expenses/revenues. The reason for leaving the depreciation out is that the depreciations shown in the financial statement

	THE OFFICIAL STATEMENT OF INCOME	THE VERSION I USED
1.	premium income: direct insurance, (dom. & foreign) reinsurance	premium income: direct domestic insurance
2.	credit losses on premiums	credit losses on premiums \times A
3.	investment income: +revenues -expenses + revaluations	investment income \times A: +revenues -expenses
4.	change in premium reserve	change in premium reserve - B
5.	claims incurred	claims incurred - B
6.	change in claims reserve	change in claims reserve - B - C
7.	total of reinsurer's share	total of reinsurer's share \times A
8.	salaries and comissions	salaries and comissions
9.	transfer to pension foundation	
10.	other social expenses	other social expenses
11.	other operating expenses	other operating expenses
12.	depreciation	
13.	other revenues	
14.	other expenses	
15.	provisions	
16.	direct taxes	
17.	net profit	net profit
where		
A =	the relative share of direct domestic premiums of total premiums	
B =	the same item designated to subtracted lines (as displayed in the official statistics)	
C =	the appropriate amount (=change \times A) of the change in the equalization reserve (hidden in the change of the claims reserve). Cleaned out in order not to smooth the period's result.	

table B.1

are arbitrary and do not necessarily reflect the economic depreciations of the firms. Thus I decided that by leaving them out altogether, my profit would not be any more biased than the official one, and assuming approximately the same sized relative fixed investments, there would occur no bias at all in the relative profits. As footnote 12 in section 3 indicated, depreciation plays a relatively minor role in insurance, amounting to about 2% of the premiums (using 1989

figures). The other expenses and incomes are hardly a signal that the competition monitors very intensively. Transfers to pension foundation, provisions and revaluations were left out for the same reasons as depreciation: they cannot be viewed as unbiased economic figures, but rather they are or can be used to adjust the profit.

The items controlling much of the financial statement and especially the balance sheet are reserves. There are three types of reserves, which I will explain here. The first reserve is the premium reserve. It is made because the policy periods and the financial period of the firm are often different (otherwise than in the economic models). It is

equal to the capitalized value of the payments anticipated from the future occurrences of contingencies insured by the insurance contracts in force and of other anticipated expenses resulting from these insurances, reduced by the capitalized value of future premiums and increased by the capitalized value of the liability that may arise from insurance policies terminated before the expiration of the period of insurance agreed upon (OECD report 1991).

The second reserve is the claims reserve, which is made because all the contingencies arising during the financial period are not paid for during the period. If you crash your car on Dec. 31st, the loss to the insurance firm whose financial period ends the same day will be reported in the claims reserve. The claims reserve is

equivalent to the amount of incurred but outstanding claims and other expenditures related thereto and includes an equalization amount, calculated according to risk theory, to provide for years with a high loss frequency (OECD report 1991).

The claims reserve incorporates the third reserve, the equalization reserve. Its function is to smooth out the fluctuations in paid claims. This is achieved by using a moving (10 year) average as a point of comparison. If the incurred claims of the period exceed this point they are adjusted downwards and vice versa. In the long run the equalization reserve should not affect the profit. This change in reserve has been subtracted from part of the profits calculated for this study. These are called unsmoothed profits. The smoothed profits then include the change in the equalization reserve. The problem from the point of view of this study is that the equalization reserve changes were available only as firm-wise aggregates: Thus I was forced to multiply them by A . This is a possible source of bias since the insurance lines left out of this study may very well have had differently fluctuating claims than the lines in this study. By taking the premium-based share of the change I miss this possible change in different directions. I have checked this for source of bias for 1989, a year for which I have data on by-line changes in the equalization reserve, and the errors in the change of equalization reserve I use were normally very small, a few percentage points. Another problem is that this effect cannot be assumed to be homogenous over the firms as can reasonably accurately be done with depreciation. The bias hereby induced is probably quite small, since the lines left out are relatively small.

Going back to the measures of profit, there were two of them. The rate of return on the equity capital and price-cost margin. ROE was calculated by multiplying the equity capital with A (the share of direct premiums to total premiums) and the assumption behind this is that the riskiness of different lines (to be accurate: the riskiness of the lines left out and those included in the study) is of the same magnitude thus needing the same amount of equity. The equity was further adjusted since it is comprised, among others, of the profit/loss of the previous period. This was corrected to be the profit from the calculations made for this study. Thus the equities used are not book-value equities. As the profits were not smoothed with the equalization reserve (in all cases), a few times the equity was in danger of turning negative. To avoid this, in these cases I decided to use the book-value equity subtracted by the official profit from the previous period. Using this formula I guaranteed that the equities were positive even when the normal equity formula used would have produced a negative equity. The reason for adjusting the equities in the first place is that as the official profit is not an adequate measure of the financial result, it should not be allowed to affect the base on which the return is calculated. It can be thought that the owners can calculate the true profit as well and thus calculate their true investment (in the form of retained dividends) in the firm. PCM was calculated by dividing the profit by total direct premiums.

Another problem posed by the legal details is that some firms have subsidiaries or in the case of mutuals, have paid the guarantee capital of another mutual firm. In these cases the firms were added together. As it was hard to obtain data when the de facto merger had in all cases taken place, I took a conservative stand: if there was no clear evidence that two firms acted together, they were

treated as separate firms. Anyhow, the forming of economic entities instead of relying on legal ones means that the number of the firms in the market is different when applying these different criteria.

As mentioned in the text, some firms were excluded from the estimations. A problem is that the results can prove to be volatile in the sense that leaving out or including one firm can alter the results in a significant way. The reason for this is that if a firm is present t periods and is left out for some reason, this affects all the variables²² in all those t periods, since the variables are differences and leaving one firm out changes the industry average as well as the average of that group. This would not be a problem if it were clear which firms to leave out, but there is some room for speculation here. I have chosen to leave out firms belonging to the following four categories:

1) the firms do not practice direct insurance on a big scale, but are largely reinsurers. As I include all the operating costs into my profit calculation, these firms get too low profits, since, say, only 5% of their premiums, but 100% of their operating costs are included. There are four such firms and none of these is active for more than four periods.

2) so-called captive firms (owned by manufacturing firms), that concentrate in insuring the risks of their owners. According to industry representatives, these firms do not actively participate in the market, but exist largely for tax (etc.) reasons. There are three such firms, each active over the whole period of study.

²² but one: the other groups' within group profit difference variable-values are not affected

3) firms present less than two periods (2 firms), or present for two periods, but active only in one (1 firm). There is one firm in the latter group and the reason for eliminating it from the sample is that its second period losses are so big that they affect in a significant way the industry and group averages. The firm has hardly any premiums in the second period of its existence and this is the reason why I think it is correct to exclude it.

4) insurance associations and foreign firms (2 firms) were excluded, partly because of lack of adequate data. An additional problem with the associations was whether to treat them as a single unit or separate units. They are legally separate, but as they are operating in different geographical markets and cooperate, they could as well be treated as a single unit. They can have a non-insignificant market share in some lines in some local markets, but on national level their influence is small, although increasing. The foreign firms, as explained in section 3, play a very limited role and excluding them should do no harm to the estimations.

The averages of the groups that I used in calculating the profit differences were normal averages. I experimented with calculating different smoothed averages, since the small number of firms makes the average vulnerable to big outliers, and these can thus affect the variable values of all firms. It proved, however, that the normal averages performed well and thus no smoothing was required.

III INFORMATIONALLY ASYMMETRIC MARKETS AND ORGANIZATIONAL FORM

III.1 Introduction

Adverse selection models and principal-agent analysis are two prime examples of core results and methods of information economics from the last two decades. In adverse selection models, the firm cannot distinguish between different (in the insurance context: high- and low risk) customers, and has to give them (or at least the high-risk ones) an incentive to reveal their type. Adverse selection can lead to a breaking down of the market (Akerlof 1970) and in any case has severe consequences for the functioning of a market as well as for the welfare implications of market exchange. This can happen in the job market (Spence 1974) as well as in the insurance market (Rothschild&Stiglitz, 1976), which are the established examples of this phenomenon. The existence and characteristics of an equilibrium have been subject of much research (Miyazaki 1977, Wilson 1977, Jaynes 1978, Riley 1979, Dasgupta&Maskin 1986a,b). The reason for the adverse implications of asymmetric information is clear: if the one party cannot find out all the relevant characteristics of the other, it has to pay to get this information. Moral hazard, often analyzed in the principal-agent framework, is the other prime example. There, the insurance firm (or principal) cannot monitor the activities of the customer (agent), and these activities affect the probabilities of different states of the world. Moral hazard, too, has been thoroughly analyzed: see eg. Holmstrom (1979, 1982),

Grossman and Hart (1983) and Arnott and Stiglitz (1988). Moral hazard leads to inefficient solutions, because the principal (or firm) has to pay the agent (customer) to get her to act in the principal's interest, and the payment is bigger than would be with perfect information. These two analyses, adverse selection and moral hazard, are not often combined, and especially, there is to the author's knowledge no combined analysis in the original insurance framework. Here, such an analysis is provided, using essentially the same geometric tools that are common in pure adverse selection models¹. In the standard models it is implicitly or explicitly assumed that there is no other mechanism available that could render the needed information, but self-selection mechanisms. For some situations and commodities this can be so; not for others. In the latter cases the question is essentially to find the cheapest (ie. profit-maximizing) way to gather the needed information.

This chapter has two main objectives: the first one is to analyze an insurance market where both adverse selection and moral hazard prevail. The introduction of moral hazard into an adverse selection model has consequences that one might expect: full insurance is not (always) available. It turns out that the merger of these two problems is pretty straight forward. To give a taste of what is to come, think of a market with

¹ For a moral hazard analysis that at least partly relies on geometric arguments, see Arnott and Stiglitz (1988). The space they use is, however, different from the one used here. Rasmusen (1989) uses the same space as is used in this paper, but does not fully develop the analysis. Geometry is especially helpful when analyzing imperfectly competitive markets.

just one type of customers: these can either exert a fixed level of effort, or exert none. This means, then, that they in effect can choose between two types, which have different probabilities of accident and different levels of wealth (as it can be assumed that effort decreases wealth). As in pure adverse selection models, also here indifference curves can be drawn for different types. In an adverse selection model they would be labelled high- and low-risk, here they are labelled effort and no effort. Furthermore, these indifference curves are well-behaved (although the actual indifference curve of a customer, being a combination of those parts of the effort and no effort indifference curves that dominate the other, is not), and facilitate analysis considerably.

The second main objective is to study the organizational form of an insurance firm. It has been shown that in the insurance market, the use of customer specific information, both exogenous (Crocker&Snow 1986) and endogenous (Bond&Crocker 1991), such as age or living habits can alleviate the asymmetric information problem and bring the market closer to the first-best solution. Such categorization is widely used by insurance firms. It could be argued that such methods can be used to divide the customers into finer and finer groups, but as long as all the relevant characteristics are not revealed, the firm has to resort to self-selection within each of these groups. Such categorization can work well when customers are (within each group) nearly homogenous, but if this is not the case, it might be that relying on categorization is not satisfactory. Also,

exogenous and/or endogenous categorization can be inaccurate: the observable variables are noisy measures of the probability of an accident. An example of such a situation is industrial fire insurance. Even with all the characteristics of the firm (that can be transmitted and are ex post verifiable without the use of an agent from the part of the insurance firm) revealed, there can be big differences within each subgroup in, say, the way the staff has been trained to handle fire, and thus the information given does not pin down the accident probability accurately enough. Also, it might be that ex post it is difficult or too costly to verify the level of effort that has been exerted in fire prevention. These unidentified differences can lead to unused profit opportunities: the information could be collected and the insurance firm would not have to rely on self-selection, but it could sell a more profitable policy to its customer. The profit opportunity arises since a risk-averse customer is willing to pay for a reduction in risk.

One way to gather such information is to hire an agent. Throughout this chapter, I will not be concerned about the contract between the firm and the agent, but will simply assume that there is a given cost to use an agent and a fixed cost in establishing a vertically integrated structure (eg. setting up branches, training staff). The reason for this approach is that the standard principal-agent problem has been thoroughly analyzed (see eg. Hart&Holmstrom 1987) and bringing it into the current analysis would complicate the model without giving any additional insights. Brokers as an

intermediary are very different from either branches or salesmen that deal exclusively with one firm's products. They in effect claim to be agents of the customers and thus the principal-agent relationship is between the customer and the intermediary, not between the firm and the intermediary. I am excluding brokers from the current analysis because the change in the principal-agent relationship calls for a different modelling framework than the one used here.

In this chapter, hiring an agent is equated to being vertically integrated; the lack of vertical structure means that the firm is relying on self-selection mechanisms. Thus this chapter does not deal with vertical relationships as such, but with the question whether or not it is profitable to add an additional layer into the organization. A real life example of these two different strategies can be found in insurance markets. Most established insurers rely on some kind of a vertical structure: either they have their own branch network, or they sell through brokers. Salesmen with different types of contracts are commonly used. Some new (mainly life-) insurance firms rely instead only on telephone selling as their distribution method. This amounts to having no vertical structure in this chapter's classification. The latter firms can use categorization to some extent, but after that they have to rely on self-selection mechanisms. The benefit of direct selling is, of course, that the firm avoids the fixed costs of establishing a vertically integrated structure and the costs of hiring an agent to screen the customers. The same kind of activities can be identified in the job market,

too. For some jobs applicants have to go through rigorous, time-consuming and costly (for the employer) tests, which are often run by specialists, hired for the purpose. This is clearly the same kind of activity as that of an insurance firm representative going to check the fire alarm installations and general conditions of a manufacturing plant before making a policy offer. There are some papers studying this matter in a labour market framework, eg. Nalebuff & Scharfstein (1987). They show that testing may restore the competitive equilibrium.

It seems that for example in the insurance market one of the main objectives in having a vertically integrated structure is its capability to collect information. This is a very different motivation than those normally dictating the form of vertical relationships, such as market power, delegation, control etc. (for a recent survey, see Waterson 1993). Sometimes the agent's better knowledge of market conditions, whether employee, franchisee or contract partner, is cited as a reason for some type of contracts, though (see Rey & Tirole 1986). This information gathering as a motive is even clearer in the above job market example, which is a pure case of hiring an agent for a special task, namely information gathering. In this case it probably cannot be called vertical integration, though. Some earlier papers have studied informationally motivated vertical integration: Arrow (1975) studied the effects of unknown upstream costs, Carlton (1979) downstream demand uncertainty and Crocker (1983) private information about production costs in a bilateral monopoly, but none of

these papers addresses the question about gathering information of customer characteristics. As will become clear in the fourth section of this chapter, there is a connection between this chapter's model and models of vertical product differentiation. Bolton and Bonanno (1988) have studied vertical restraints in a model of vertical product differentiation.

To show the analysis of simultaneous adverse selection and moral hazard in a familiar framework, and to concentrate on it, I start with models of competitive markets, ie. markets where firms make zero expected profits. The next section shows how to treat moral hazard geometrically, and how to combine this analysis with the standard Rothschild-Stiglitz adverse selection model. The organizational form is a question that cannot be treated in the competitive markets framework, and thus after the second section, I shift to imperfect competition. The third section presents the combined adverse selection and moral hazard problem in an insurance framework, and the implications that a monopolistic structure have on the outcome. The decision of the monopolist whether to become vertically integrated or to rely on self-selection is then analyzed. The fourth section extends the monopoly model into an oligopoly framework, but concentrates on the pure adverse selection case. There, the conditions under which different vertical strategies can be observed are calculated and (some of) the equilibria analyzed. The effects of asymmetric vertical strategies on product differentiation are also discussed. The fifth section extends the analysis of the fourth to include moral hazard. Throughout

these sections, welfare effects of different levels of asymmetric information are considered and compared. The sixth section discusses other applications of the model(s), and the seventh concludes.

III.2 Adverse selection and moral hazard in competitive insurance markets

The usual insurance analysis, common nowadays in graduate textbooks, separates adverse selection and moral hazard and uses geometric methods only with the former, whereas the latter is often discussed in a principal-agent framework. Here, both are combined, but geometric methods similar to those of standard adverse selection references are used.

The first step in the analysis is to show how moral hazard can be represented using geometric arguments. Let's assume that there is just one type of customers, whose probability of accident is known to all market participants. Thus there is no adverse selection. The moral hazard problem of an insurance firm is the following: it is assumed that the customer can affect the probability of accident, but in a way that is costly for the customer and unobservable to the firm. The cost is usually (and in the following as well) labelled "effort", to stress the possibility that it includes non-monetary costs, like driving carefully. To keep the analysis clear, I will assume that there is only one level of effort that the customer can exert, the other option being "no effort". If effort is exerted, the probability of an accident is lowered to a level that everybody knows. If the firm gives full

insurance, then the customer has no incentive to exert effort, thus increasing the probability of accident and resulting in negative expected profits for the firm. To avoid this, the firm has to design a contract that gives the customer a higher expected utility if she exerts effort than if she does not. This constraint on the contract design is called the incentive compatibility (IC) constraint. There are two states of the world: accident (W_A , y-axis) and no accident (W_{NA} , x-axis). Thus the expected utility V is a weighted average of the utility of wealth U in the two states of the world, where weights are the probabilities of no accident and an accident, respectively. Point O depicts the initial point where the customer finds herself without insurance and without effort. The customer's expected utility is:

$$(1) \quad V = (1-p)U(W) + pU(W-L)$$

I will use the following notation:

W = the initial wealth of all (potential) customers

p_i = probability of accident for customer type i , $i \in \{H, L\}$, $p_H > p_L$

p_i^e = probability of accident for customer type i if effort of level e is exerted, $p_i > p_i^e > 0$

α_i = the price customer of type i pays in case of no accident

β_i = the net payment customer of type i gets in case of accident

e = the cost of effort, $e > 0$. As a superscript it indicates that effort is

exerted (see the notation for probabilities of accident)

N_i = the expected number (or proportion) of customers of type i

V = von Neumann-Morgenstern expected utility function of all customers

U = von Neumann-Morgenstern utility of money function of all customers

L = monetary value of an accident (loss)

If the customer decides to exert effort of level e , her expected utility becomes²:

$$(2) \quad V = (1-p^e)U(W-e) + p^eU(W-L-e)$$

Thus, in figure 1, starting from point O , there is a line with a 45° angle to the x-axis, on which the customer lies if effort is exerted. The angle of the line follows from the assumption, standard in moral hazard analysis, that effort is exerted before the state of the world is revealed. Effort has thus the same effect on the wealth of the customer in both states of the world, as can be seen from eq. (2), and this gives the 45° -line. Point O' is the point where the customer lies if the effort level is e , as is assumed. The indifference curve (solid line) going through point O is the one on which the customer lies if she does not exert any effort. If she exerts effort, however, the shape of the indifference curve changes: the customer can, so

² Here, I make the assumption that the utility function is well-behaved w.r.t to effort. As is known from moral hazard literature (e.g. Arnott&Stiglitz 1988), this is not always the case. A way of circumventing this problem would be to assume that the effect of effort on the probability of accident is concave, and that the expected utility function V is linear in effort.

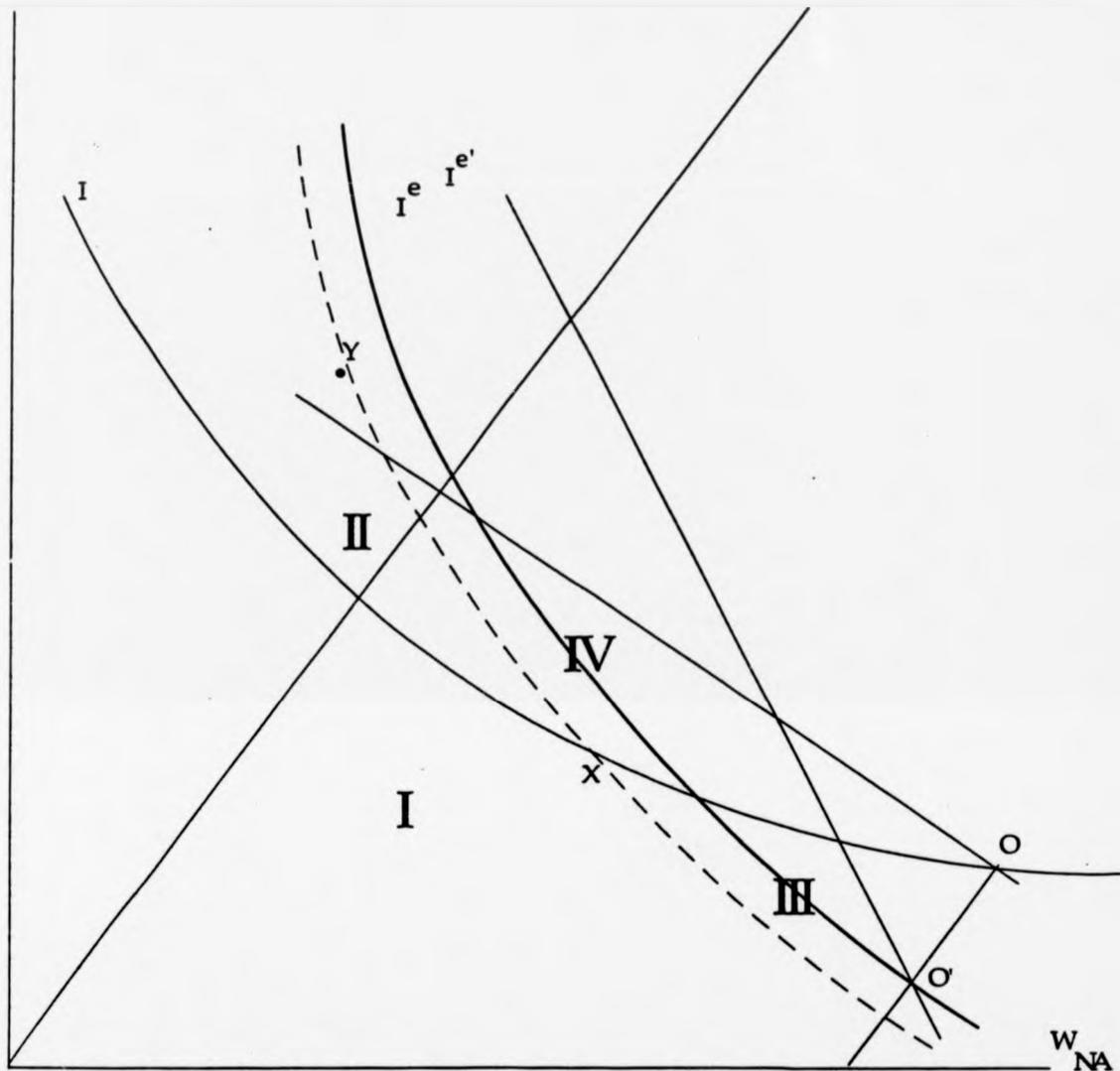


figure 1

to speak, choose between two different types: no effort and effort. Since the effort type has a lower probability of accident, and the same utility function otherwise, we know that its indifference curve is everywhere steeper than the indifference curve of the no effort type (this is the so-called single crossing property). Furthermore, if it is assumed that the effect of effort on the probability of accident is declining in effort, as is

usual in a moral hazard context, and that the customer is better off (if not insured) if she exerts at least some effort, then it follows that there is a level of effort which makes the customer indifferent between no effort and effort. This is the dotted indifference curve in figure 1. I will assume³ that the customer is better off exerting the assumed level of effort, e , than by not exerting it. Thus the dotted indifference curve of the effort type lies below the effort indifference curve going through point O' . In the following, I will call the two different indifference curves the "effort" and "no effort" indifference curves. If a effort and a no effort indifference curve give the same expected utility, they are labelled "equivalent"⁴. Thus in figure 1, the no effort indifference-curve going through point O and the dotted effort indifference curve are equivalent.

Let's concentrate for a moment on the no effort indifference curve on which the customer lies, and the equivalent effort indifference curve. Let's call these I and I' . The single crossing property ensures that they cross only once, at point X . The (W_{NA}, W_A) -space is divided into four subspaces, labelled I-IV. In space I, the customer is always worse off (or at most as well off as) than on the indifference curves I and I' . In space II, she can improve her expected utility by not exerting effort, compared to the indifference curve I' . Thus, if she would lie at point Y and exert effort, then

³ This assumption is not necessary for the analysis, but is made since some kind of assumption on this point has to be made to be able to draw figures. It is standard in moral hazard analysis.

⁴ Equivalent meaning $(1-p)U(W)+pU(W-L)=(1-p^e)U(W-e)+p^eU(W-L-e)$.

she could improve her expected utility by stopping to exert any effort. This would move her from the effort indifference curve to a no effort one that lies above the no effort indifference curve that is equivalent to the effort indifference curve that goes through point Y. Similarly, in space III, she can improve her expected utility by shifting from no effort to effort. Finally, any point in space IV would improve (or at least leave unchanged) her expected utility whether or not she exerts any effort, and depending on the point in that space she is better off, worse off, or indifferent between these two options. There is a boundary, going through point X and dividing spaces I and IV, on which the customer is indifferent between exerting effort and not exerting effort. That is, the boundary consists of points where equivalent effort and no effort indifference curves cross. On this boundary, the incentive compatibility constraint binds.

Although some effort indifference curves cross some no effort indifference curves above the 45°-line, the standard moral hazard assumptions mean that every effort indifference curve crosses the equivalent no effort indifference curve below the 45°-line. Otherwise the moral hazard problem would not exist, as the customer would be able to get full insurance (=a contract at the 45°-line), and would still exert effort.

What is the insurance firm's role in this setting? In this section, I assume that markets are competitive, ie. firms make zero expected profits. As in adverse selection models, iso-profit lines can be drawn. Especially, now

there are two types of iso-profit lines for each type of customer: one for no effort and another for effort. In figure 2, two of these, namely the ones giving zero expected profits, are shown. All the points on line π give the firms zero expected profits if the customer does not exert (any) effort, and similarly all the points on line π^e give zero expected profits if the customer exerts effort. They go through the initial points of the customer when no effort, respectively effort e , is exerted (line π goes through O , line π^e through O'). It is easy to show that the iso-profit lines for the effort-case are steeper than those for the no effort one. Competition drives the solution to the point (ie. to the contract) that gives the customer the highest possible expected utility, subject to the constraint that firms must make nonnegative profits. This point will lie on one or the other zero-profit line. The best possible contract that the firms can offer customers when no effort is exerted is depicted by point C in figure 2. It lies on the 45°- line and gives full insurance. The no effort indifference curve that goes through C is labelled I^0 , and the equivalent effort indifference curve I^{0e} .

Can the customer be better off than at C ? The point where the effort zero-profit isoprofit line intersects the I^0 -indifference curve is C^* . However, it lies in space IV of figure 1, which means that the incentive compatibility constraint does not bind and there thus is room for improvement of the customer's utility. The point we are looking for lies on the effort zero profit

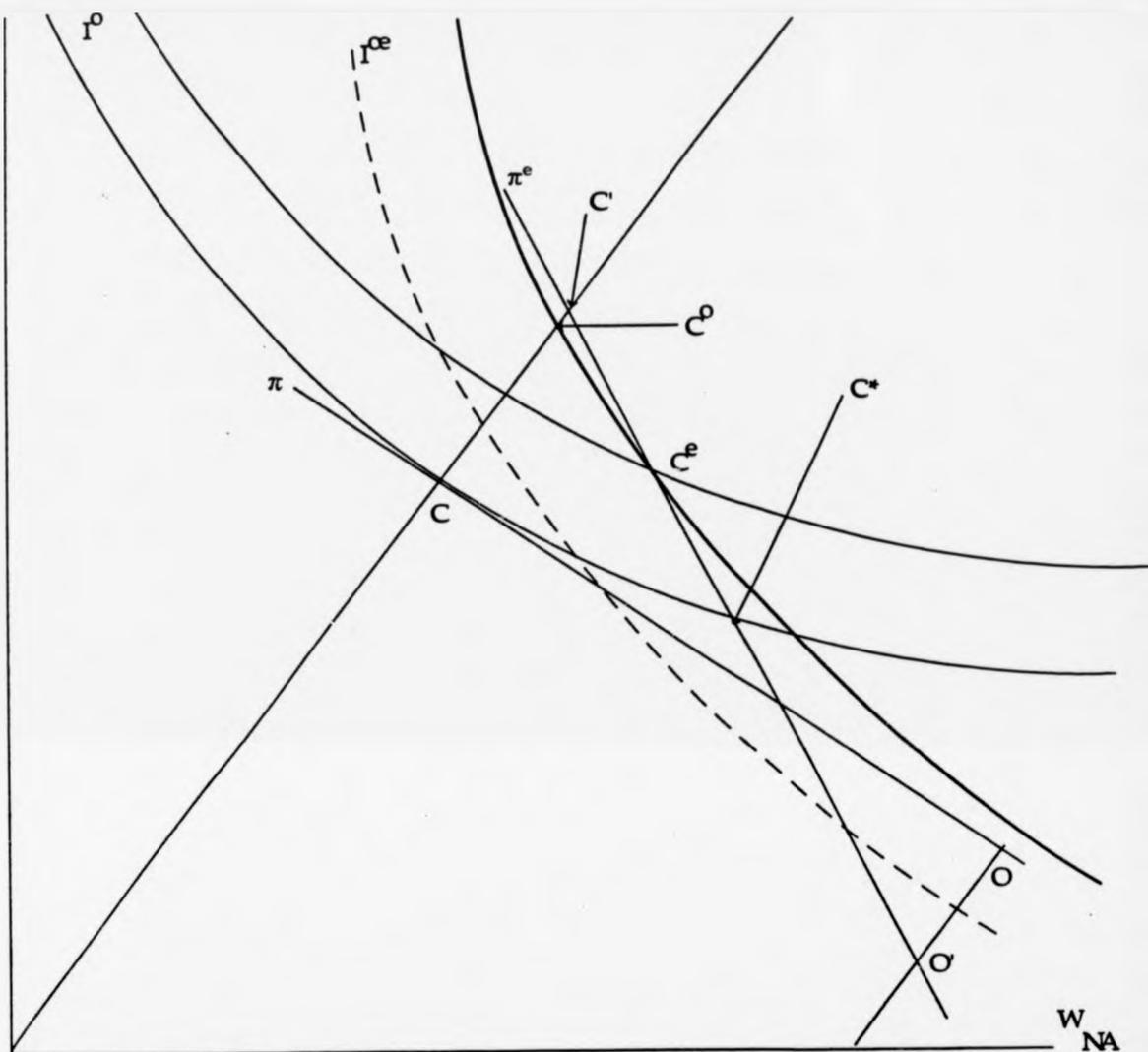


figure 2

line and on the boundary between spaces III and IV. This point is labelled C^* in figure 2. It lies on the intersection of spaces III and IV of figure 1, and on the effort zero-profit line. In space III, the customer could make herself better off by exerting effort, and all points in space IV were at least as good as those in other spaces. Thus here, the customer can, by exerting effort e , increase her expected utility compared to contract C , and the firms still make nonnegative (zero) profits. In a pure moral hazard world, with

the assumptions made, point C^e is the outcome in an insurance market with one type of customers. The reasons why C^e is the outcome are as follows: As it lies on the intersection of equivalent effort and no effort indifference curves, the incentive compatibility constraint binds. This means that it lies on the boundary between spaces I, II, III and IV of fig. 1. As it also lies on the effort zero-profit line, this means that the customer cannot be made better off without forcing the firm(s) to make expected losses. This is so since all points (=contracts) that would make the customer better off are above the respective zero profit lines: above the effort zero profit line in the case of spaces III and IV, and above the no effort zero profit line in the case of spaces II and IV.

It should be noted that point C^e does not lie on the indifference curve I^0 . If there were no moral hazard, but effort were observable and verifiable in courts, then the customers could get contract C' , which lies at the point of tangency of a effort indifference curve and the zero-profit line. The customers' loss of welfare due to moral hazard can be measured along the 45°-line as the distance between the effort indifference curves going through points (contracts) C^e and C' . This distance is $(C^e C')$.

Now that a machinery has been created to deal with moral hazard geometrically, this can be added to the traditional (=Rothschild&Stiglitz) analysis of adverse selection in competitive markets.

To add adverse selection, a second type of customer is introduced: the two types of customers differ only in respect to the probability of accident, and in this section I will maintain the standard assumption that only the customers know their type (low- or high-risk). Insurance firms know the expected proportion of high-respectively low-risk types in the population. In contrast to the traditional analysis, here the customers can influence their probabilities of accident (it is assumed that this happens in the same way and with same effectiveness for both types), as in the analysis above.

The adverse selection problem can be summarized in the so-called self-selection⁵ constraints:

$$(3) \quad V(p^x, \alpha^x, \beta^x, x) \geq V(p^y, \alpha^y, \beta^y, x) \quad i \neq j, \quad i, j \in \{H, L\}, \quad x, y \in \{0, e\}$$

They state that each type (ie. high- and low-risk) of customer has to prefer the contract that is designed for that type. The incentive compatibility constraints that result from the moral hazard problem can be written as follows:

$$(4) \quad V(p^e, \alpha^e, \beta^e, e) \geq V(p^i, \alpha^i, \beta^i, 0) \quad i \in \{H, L\}$$

As in the previous pure moral hazard analysis, the firms make zero

⁵ To avoid confusion, I will use the following terminology: self-selection constraints refer to the adverse selection problem, incentive compatibility constraints to the moral hazard problem.

expected profits. In line with the standard analysis, I will assume that the firms have to make zero expected profits per contract, ie. they cannot subsidise one type of customers with profits from contracts with the other type. Now there are two initial points, as before, but four zero-profit iso-profit lines: one initial point for no effort and another for effort, and two zero profit lines through both initial points, one for high-risk customers and the other for low-risk customers (see figure 3). The iso-profit lines of low-risk customers are steeper than those for high-risk customers⁶.

Let's assume (for notation) that the customers analyzed in the pure moral hazard case were high-risk. The best contract that they can get with no effort is C , and the best contract that they can get by exerting effort is C^e , as previously. This means that they choose C^e , and get partial insurance. What about the low-risk customers? In the pure adverse selection analysis, they are constrained by the existence (and nonseparability) of high-risk customers, and this is the case here, too. The fact that the firms cannot make the different types of customers apart means that they have to offer the low-risk types a contract that fulfils the self-selection constraint of high-risk customers. Thus the low-risk customers cannot be offered a contract that lies on a higher high-risk indifference curve than C^e , no matter whether this indifference curve is for effort (as the one on which the high-risk customers lie at point C^e) or the equivalent no effort indifference

⁶ The analysis can be carried out when the effort iso-profit lines of high-risk customers are steeper than no effort iso-profit lines of low-risk customers, and vice versa. Here, the first case prevails.

The contract that the low-risk customers would be offered in the absence of moral hazard is F that lies on the zero-profit line of low-risk customers that exert effort, and on the effort indifference curve of high-risk customers that is tangential with the effort high-risk zero-profit line at the 45° -line (the high-risk customers would get C'). In figure 3, low-risk customers are worse off because of moral hazard. This is easy to verify: Point F (which would entail effort that is assumed to be costlessly observable in a pure adverse selection model) is on an effort indifference curve that lies higher than the effort indifference curve that goes through point F' . The reason for regarding the effort indifference curves as the relevant ones is that pure adverse selection models can be thought of as models where effort is observable (and verifiable in courts), and thus can be brought into contracts. Even if there were no moral hazard on the part of low-risk customers, these would suffer from the fact that there is a moral hazard problem with the high-risk customers.

If there were no adverse selection, but only moral hazard, the customers could be offered contracts on the respective effort zero-profit lines, without taking notice of the self-selection constraints. These would be contract C' for the high-risk customers and contract D' for low-risk ones. At D' , the incentive compatibility constraint of low-risk customers binds. High-risk customers prefer D' to C' , too, but since the firms can tell the different customers apart, they cannot get it.



figure 5

welfare loss is:

$$(5) \quad N_H(C^e C') + (1-N_H)(F^e D')$$

If there is only adverse selection, the customers are offered C' and F . The high-risk customers get full insurance, and only the low-risk customers are

profit. But, as Rothschild and Stiglitz show in a pure adverse selection framework, no pooling equilibrium exists. Thus it is possible that there is no equilibrium in a competitive market. The effect of moral hazard is the same as in pure moral hazard models: the customers do not get full insurance, since the firms need to give the customers an incentive to take care, ie. to exert effort. Since both types are worse off with both adverse selection and moral hazard than with only one type of asymmetric information or at most as well off⁸, the society is also worse off (since the firms always make zero expected profits).

Again, welfare effects of asymmetric information can be studied. As the firms are assumed to make zero expected profits no matter what kind of information they have, the only welfare changes are for the customers. If there were perfect information, high-risk customers would get contract C' (see figure 5), and low-risk customers contract D' , giving both customer types full insurance. With both adverse selection and moral hazard, the high- and low-risk types get contracts C^e and F^e , respectively. The points that lie on the same respective effort indifference curves as these contracts, and on the 45°-line are C^o and F^o . As the proportion of high-risk customers in the population (the size of which is normalized to one) is N_H , the total

⁸ The high-risk customers get the same contract with adverse selection and moral hazard as with moral hazard only.

subjected to a welfare loss. This is of size $(F'D')$, as F' lies on the same low-risk effort indifference curve as the contract that low-risk customers are actually offered. If the adverse selection problem is solved, but the moral hazard one remains, then the high-risk welfare loss is the same as in the one customer type model presented earlier, but the low-risk customers' loss has to be added to get the society's loss, giving

$$(6) \quad N_H(C^{\circ}C') + (1-N_H)(D^{\circ}D')$$

Low-risk customers are better off with pure moral hazard than with pure adverse selection, but from the society's point of view pure adverse selection might be preferable. The reason for this is that with pure adverse selection, high-risk customers get their first best full insurance contract, contrary to the pure moral hazard case.

III.3 Vertical strategies and monopoly

The adverse selection model of Rothschild and Stiglitz (1976) has been extended to monopoly by Stiglitz (1977). In this section, I borrow his model's basic structure, but add moral hazard and change slightly some assumptions. The set-up of the model is otherwise identical to the previous section's model, but now there is just one firm in the market. To be able to identify a given customer, the monopoly has to hire an agent. I thus assume that a firm which is not vertically integrated sells its policies via

telephone or the like and that there is no possibility for a firm with no vertical structure to identify the customer type, but self-selection mechanisms. This "no vertical integration", or direct-selling (as I will subsequently call it) firm thus equals the firms in the previous papers. A vertically integrated firm, on the other hand, gets to know whether a customer belongs to the high- or low-risk group and, possibly, whether or not the customer has exerted effort. This makes it possible for a vertically integrated monopolist to discriminate between the customer types and to ensure that they exerted effort. The firm can thus reap all the surplus to itself, if it so wishes. Figures 6 and 7 clarify the situation.

Figure 6 presents the situation of a vertically integrated monopolist. Point O is the initial point where both customer types are without a contract, and exert no effort and point O' the initial point with effort. The firm wants to move to an equal-profit line as close to the origin as possible, and the customers want to get to as high an indifference curve as possible. The vertically integrated firm operates under two constraints: the policies it offers must give the customers at least the same utility as they have without a contract (these are the so-called participation or individual rationality constraints). Each contract can be described by the price paid if there is no accident, α , and the payment in case of an accident, β . A contract is completely described by the effect these two, the level of effort and the probability of an accident, p (since the original wealth is constant, it can be neglected here), have on the expected utility V of a customer. The

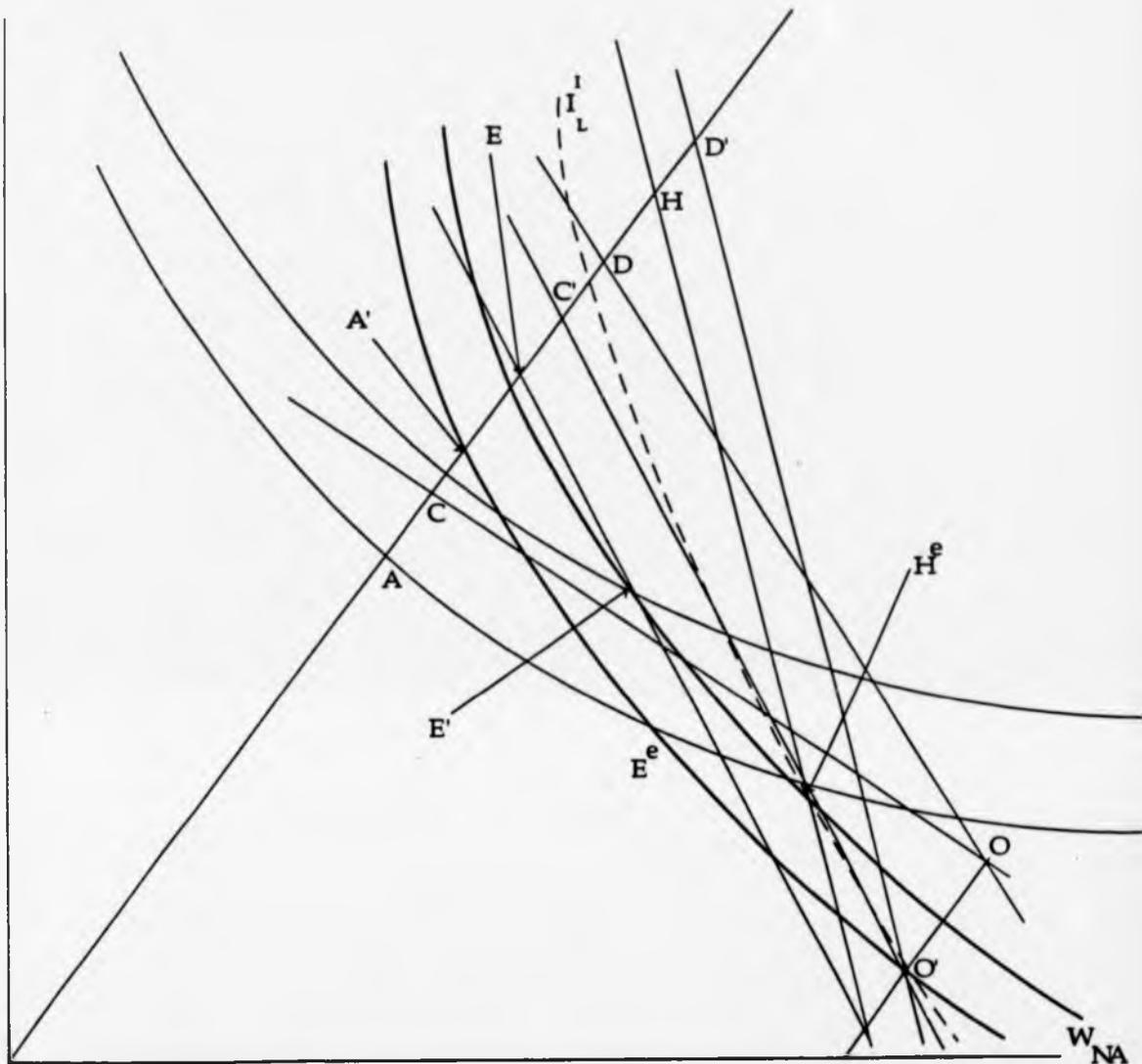


figure 7

facing the monopolist is now clear: it will balance the marginal loss from offering a more lucrative contract to high-risk customers to the marginal gain in offering a more profitable contract to low-risk customers. To get the optimal contract pair, it is not enough to compare marginal gains and losses per contract, but they have to be multiplied by the number of low- and high-risk customers, respectively, to get the true marginal effect.

insurance, effort (or the equivalent no effort) - level of utility, as is written on the right hand side in eq. (6). The binding of the participation constraint means that the customers' utility does not improve from buying an insurance, compared to no insurance cover. This is so because the vertically integrated monopolist is able to identify the customers and reap all surplus created by exchange. Since the monopolist maximizes profits and customers are risk-averse, the monopolist would like to sell both types of customers full insurance. With regard to the abilities of the agent, it is assumed that she can always tell the type of the customer. I will further assume that with an extra cost, it is possible to detect the effort that a customer has exerted. Of course, the firm only screens the effort levels of customers who had an accident.

The firm can then make a decision whether or not to screen the effort. When both customer type and effort can be detected, the monopoly can break the self-selection and incentive-compatibility constraints. This enables the monopoly to sell full insurance, and it only has to decide whether or not it wants the customers to exert effort. In both cases, however, it must take into consideration the fact that the customers can exert effort, and thus raise their utility level. This means that the firm cannot sell a contract that keeps the customers on the no effort indifference curve going through point O , but at most it can keep them on the effort indifference curve going through O' , or on the equivalent no effort indifference curve. In figure 6, the full insurance points that satisfy the participation constraint

for no effort and effort high-risk customers are points A and A' , respectively. The expected profits per contract can be measured along the 45°-line. The no effort zero-profit lines of the high- and low-risk customers pass the 45°-line through C and D , respectively, and the effort zero-profit lines through C' and D' . The profit per high-risk contract is (AC) for a no effort contract and $(A'C')$ for an effort contract. It should be noted that even if the customer in the end buys a no effort contract, the fact that she can exert effort in order to improve the contract offered to her. The decision on whether the contract includes or excludes effort is made by the monopolist, and this decision, given the assumptions, can be different for high- and low-risk customers. Everybody gets full insurance, however. The firm's expected profit is then

$$(8) \quad \Pi^{VIE} = N_H \max[(AC), (A'C')] + (1-N_H) \max[(BD), (B'D')] - (R + T)$$

where B and B' are the no effort and effort contracts of low-risk customers, R is the cost of establishing a vertical structure and screening all customers for their type, and T is the expected cost for screening the effort of those customers who had an accident. The value of T depends on the size of the population and on whether the monopolist sells effort contracts to both customer types. The superscript VIE stands for vertical integration and screening of effort. It is necessary to screen customers, since high-risk customers prefer contracts B and B' to contracts A and A' , and if customers are sold either contract A' (for high-risk customers) or B' (for low-risk

customers), they would prefer not to exert effort. If, given that no effort contracts are sold, a proportion δ would be left unscreened for type, then an expected number $N_H\delta$ high-risk customers would be able to buy contract B and thus incur an expected loss of $N_H\delta(CB)$ to the monopolist. I assume throughout this chapter that the marginal cost of screening the last customer has a smaller absolute value than the expected loss of not screening him. Another slight modification from the traditional set-up further necessitates the screening of all customers: the insurance firm knows only the expected number of high- and low-risk customers, not the exact numbers. If it knew exactly how many high-risk customers there are in the population, it could order its agent to stop screening after the agent had found all high-risk customers.

If the firm decides not to screen effort, it saves the effort screening cost T . Simultaneously, however, it cannot any more offer full insurance and expect the customers to exert effort. These would happily buy such a policy, and then not exert effort, thus increasing their utility. The firm can offer no effort full insurance policies (A for high-risk, B for low-risk customers). The no effort contracts' profits remain unchanged when the firm does not screen effort. The effort contracts have to obey the incentive compatibility constraints imposed by moral hazard, and this means in terms of figure 6 that the firm offers high-risk customers contract E^e . This contract keeps the customers on their effort indifference curve where the participation constraint binds, thus maximizing profits, and at the same

time the incentive compatibility constraint is binding. The firm's profit per high-risk contract is now measured as the distance between C' and the point where the 45°-line and the effort iso-profit line going through E^c cross, ie. point E in figure 6. Not screening effort thus results in a per contract decline in (gross) profits of $(A'E)$. As the firm still screens the customers for type, it can break the self-selection constraint and sell low-risk customers a policy that high-risk customers would prefer over the contract that they eventually get. The analysis of the low-risk customers follows that of the high-risk ones, resulting in either contract B (for no effort, and no screening of it) or G^c (effort, obeying the incentive-compatibility constraint). The firm's profit can be written as

$$(9) \quad \Pi^{VI} = N_H \max[(AC), (EC')] + (1 - N_H) \max[(BD), (GD')] - R$$

If effort for some reason cannot be screened, the vertically integrated monopolist's profit is (9). If effort can be screened, the firm chooses the larger of Π^{VI} and Π^{VIE} .

In addition to the two possibilities that involve some degree of screening, the firm has a third option, namely direct selling. If it chooses this organisational form, it has to take into account all three (the participation, self-selection and incentive compatibility) constraints of each customer type. Not all of these bind, however. As is known from earlier analyses of adverse selection (see eg. Fudenberg & Tirole 1991, ch. 7), only one

participation and one self-selection constraint bind: the self-selection constraint of the high-risk type and the participation constraint of the low-risk type. The incentive compatibility constraint binds for high-risk customers, but not necessarily for low-risk ones. The binding of a low-risk customer's participation constraint means that her utility does not improve from buying an insurance contract, compared to no insurance cover. The binding of the self-selection constraint of the high-risk customer means that the policy that is offered to low-risk customers lies on the same high-risk indifference curve as the policy that high-risk customers actually buy.

Let's assume that the direct-selling monopolist starts at contracts E^c (for high-risk customers) and O' (for low-risk customers, ie. no contract). It earns no money on the low-risk "contract" since it lies on the effort zero-profit line. On the other hand, this way it gets the maximum profits per contract out of high-risk contracts, since it is not possible to get to a higher high-risk equal-profits line than the one passing through E^c , because of the participation constraint of high-risk customers⁹. If the monopolist wants to earn money on the low-risk contracts, this means that it has to move up along I_L^l (remember, the participation constraint of low-risk customers always binds). At the same time, however, it has to offer a more lucrative contract than E^c to high-risk customers to prevent them from choosing the contract that is designed for low-risk customers (ie. H^c). The trade-off

⁹ I am assuming here that an effort contract provides larger profits than a no effort contract.

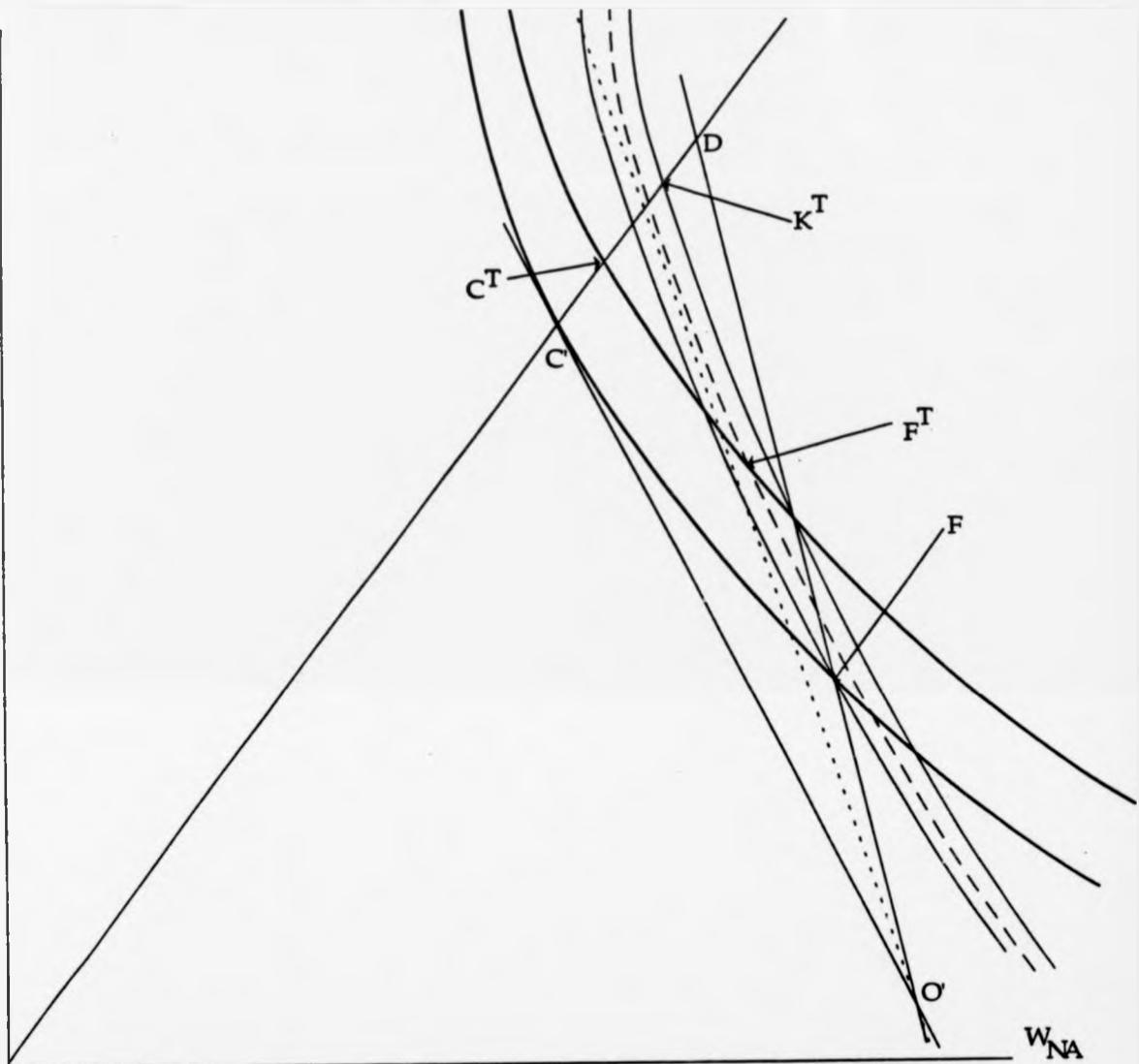


figure 8

Proof: The only reason for any firm to sell loss-making contracts is to increase total profits. By selling loss-making contracts to high-risk customers, a firm might be able to sell profitable contracts to low-risk customers and thus maximize total profits. As firm 1 is vertically integrated and can tell the customers apart, it does not need to sell loss-making contracts for this end. QED.

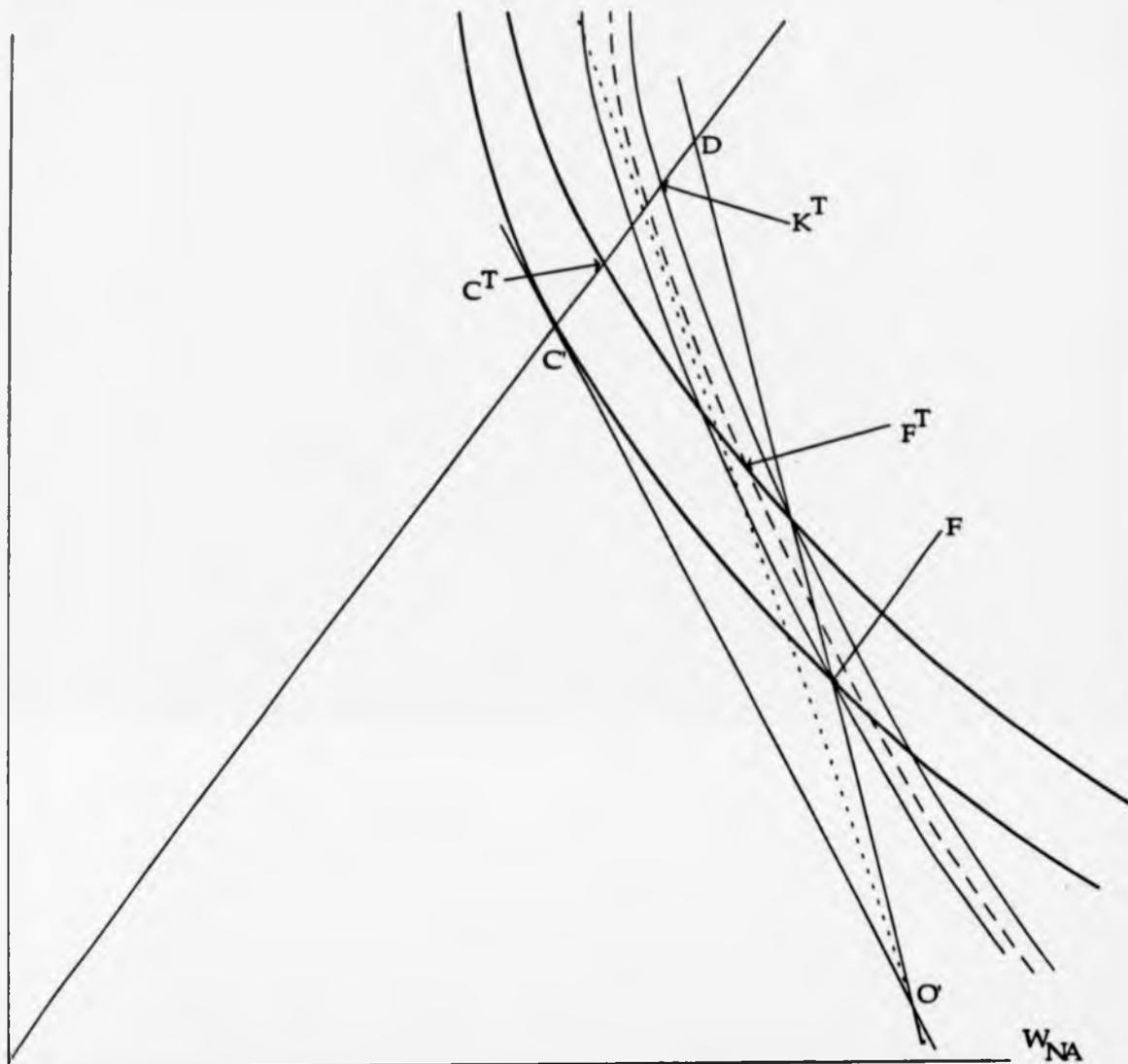


figure 8

Proof: The only reason for any firm to sell loss-making contracts is to increase total profits. By selling loss-making contracts to high-risk customers, a firm might be able to sell profitable contracts to low-risk customers and thus maximize total profits. As firm 1 is vertically integrated and can tell the customers apart, it does not need to sell loss-making contracts for this end. QED.

Equation (10) states the optimality condition:

$$(10) \quad -d(EC')N_H = d(HD')(1-N_H)$$

It is clear from (10) that the optimal contract pair depends on the proportions of high- and low-risk customers in the total population. If either of the two customer groups grows proportionally too big, then it does not pay to separate them any more, but the firm reverts to a pooling contract. That means that it offers the same contract to both customer groups. It also means that as it does not intend to separate the different customers, it will be a direct-seller. As N_H approaches 1, at some stage a point is reached where it does not pay to offer a contract to low-risk customers any more, since the loss from offering a more lucrative contract to the big high-risk contingent of customers is greater than the profit gain from the small number of low-risk contracts sold. In such a case the relation in (10) might hold with an inequality and a third constraint, namely the participation constraint of the high-risk customer, would bind in addition to the two constraints that always bind. Stiglitz (1977) analyzed this threshold for the case of a logistic utility function. The possibility of the market for low-risk contracts breaking down is thus real in the case of the direct-selling monopolist. At the other limit, with N_H approaching zero, pooling will eventually become more profitable than separation and the direct-selling insurance firm will offer the low-risk customers a contract with less than full insurance even if there is no moral hazard and the

participation constraint binding. The reason why the firm does not offer full insurance in the absence of moral hazard is that in the pooling case the relevant equal-profit line is the whole population equal profit line and this is not tangential to the low-risk indifference curves at the 45°-line unless there are no high-risk customers in the population. The few high-risk customers buy this contract, too. Notice that the high-risk customers thus never have a risk of not getting a contract.

When is it profitable to build a vertical structure? The direct-selling monopolist makes a total profit of

$$(11) \quad \Pi^{DS} = N_H(EC') + (1-N_H)(HD')$$

The first term is the total profit from high-risk contracts and the second the total profits from low-risk contracts. The monopolist compares this with the largest value of (8) and (9), and chooses the organizational form and the level of screening that maximizes its profits.

The welfare analysis is relatively simple: in the case of the vertically integrated monopolist, the firm is able to identify the customers and their effort and thus reaps all the surplus from exchange. The direct-selling monopolist, on the other hand, must rely on a "sweetener", to get the high-risk customers to reveal their type, and all customers to exert effort. Thus the high-risk types are better off with a direct-selling monopolist than with

a vertically integrated one. The low-risk customers make no gains in utility, whether we compare the two monopolies to each other or to the initial utility level. Low-risk customers only manage to exchange some wealth for a smaller risk (possibly a zero risk in the vertically integrated case), if even that. With a vertically integrated monopolist, the low-risk customers are always offered a contract, since it is profitable for the monopolist to become vertically integrated only if the proportion of low-risk customers is large enough. If the total welfare is expressed as the sum of customer utility and firm's profits, then the two solutions can be compared. The gain that high-risk customers make in the direct-selling case compared to the vertically integrated monopolist can be measured along the 45°-line as an increase in expected utility, measured in money (wealth). This increase is exactly as big as the decrease in the monopolist's expected profits, and these two effects thus offset each other. The low-risk customers get the same utility. This leaves as the only possible source of difference in total welfare the changes in profits that the monopolist gets from selling to the low-risk customers. The gross profits are bigger for the vertically integrated monopolist, but the screening costs have to be extracted from them¹⁰. This means then that as long as the costs of vertical integration are smaller than the extra profits from low-risk customers' contracts that the monopolist can make by being vertically integrated as opposed to

¹⁰ When taking the screening costs into account the (possible) gains in utility of the agent(s) should be taken into consideration. These can safely be assumed to be of a lot smaller magnitude than firm profits or the changes in all customers' welfare, and thus are neglected here.

being a direct-seller, there are welfare gains from vertical integration. There are two possible cases with respect to welfare: the low-risk customers get either full insurance (effort is monitored) or partial insurance (no monitoring of effort). I will analyse the first case here. The profit from full insurance low-risk contracts is (see figure 6) $(1-N_H)(B'D')$. Thus if equation (12) holds, there are welfare gains from vertical integration:

$$(12) \quad (1-N_H)(B'D') \geq R + T$$

As this is a stricter condition than (8), the profitability condition of vertical integration (when the monopolist sells full insurance to low-risk customers, and any insurance to high-risk customers), vertical integration can be privately profitable even when it is socially undesirable.

The model can be analyzed for different forms of asymmetric information, on the lines of the comparative analysis in the end of section 2. This is straightforward, but space consuming, and is thus left to the reader.

III.4 Vertical strategies and oligopoly

As I will show in this section, under certain conditions it is profitable for a firm to invest in a vertically integrated structure even when there is competition in the market. This will, however, only be profitable for one of the firms, and others choose the direct-selling strategy. This section

makes a slight deviation from the normal approach adopted in this chapter, since I will assume that there is no moral hazard. This is done in order to simplify the analysis. Moral hazard is introduced into the oligopoly model in the next section. I will assume, in line with Rothschild and Stiglitz, that competition is what they label price-quantity competition. This means that each firm offers the consumers pairs consisting of a price and a quantity (ie. level of cover), not just one or the other. This leads to competition being effectively Bertrand in nature, as noted by eg. Dasgupta and Maskin (1986b). The model that is analyzed here will be a duopoly model, but extending it to the n -firm case is trivial and does not affect the nature of equilibria, only their number. The equilibrium concept adopted in this section is that of Riley's (1979) reactive equilibrium:

Definition (Riley 1979, p. 350): A set of offers is a reactive equilibrium if, for any additional offer which generates an expected gain to the agent making the offer, there is another which yields a gain to the second agent and losses to the first. Moreover, no further addition to the set of offers generates losses to the second agent.

This section has a connection to the literature of vertical product differentiation. In that literature, it is assumed that customers' initial wealth varies. It can then be shown that these different customer groups buy different products and, especially, those customers with a higher wealth buy the better and more expensive good (Shaked and Sutton 1983).

Firms can choose different vertical strategies and this asymmetry also leads to vertical product differentiation, but only in one part of the market. Here, the motivation to vertical product differentiation does not spring from differences in initial wealth, but from the different preferences of customers that are due to different probabilities of accident. These differences (can) create an asymmetry in the vertical strategies of the firms. What is similar to the "pure" vertical product differentiation literature is that fixed investments (sunk costs) are a major element of the model. The vertically integrated firm can break the self-selection constraint of the high-risk customers because it can identify them. One result is that the two firms with asymmetric vertical strategies offer identical products for one customer group (high-risk customers) and (vertically) differentiated products to the other customer group (low-risk customers) and that the latter all prefer the product of the vertically integrated firm to that of the direct-selling firm. Here, the low-risk customers behave similarly to the wealthy customers of pure vertical differentiation models.

The game has three stages (and stage zero, where Nature chooses the exogenous variables, among them which firm gets to choose its vertical structure first):

- I firm k chooses its vertical structure
- II firm l observes k 's choice and chooses its vertical structure
 $k \neq l; k, l \in \{1, 2\}$
- III firm k observes firm l 's choice and both firms compete in price-

quantity contracts

The sequential structure of the game does not alter the nature of equilibria, but makes the game somewhat clearer. Actually, in the third stage the firms and customers play a game. The game between the firms is one of perfect information, but the game played between the firms and the customers has incomplete information. The third stage's game between the customers and the firms is essentially suppressed into the constraints of the firms' profit functions. As the firms are identical, the model will produce twice the number of equilibria found in the following treatment just by reversing the order of firms in stages I and II. In the following, I am assuming that it is firm 1 that gets to choose its vertical structure first. Firms are denoted by superscripts, customers by subscripts. Both firm 1 and firm 2 observe each others' choices of vertical structure before deciding what kind of contracts to offer to the customers. That they do simultaneously.

The maximization problem that the firms face is the following:

$$(13) \quad \max_{\alpha, \beta, R} \Pi^k = \sum_i x_i^k N_i [\alpha_i^k - p_i \beta_i^k] - R^k$$

where $x_i \in [0,1]$, $\sum x_i^k = 1$, $i \in \{H,L\}$, $R^k \in [0,\infty)$, $k \in \{1,2\}$

so that

$$(14a,b) \quad V(p_i, \alpha_i^k, \beta_i^k) \geq V(p_i, 0, 0) \quad \text{PC}$$

$$(15a,b) \quad V(p_i, \alpha_i^k, \beta_i^k) \geq V(p_i, \alpha_j^k, \beta_j^k), \text{ if } R^k < R; i \neq j; i, j \in \{H,L\} \quad \text{IC}$$

$$(16) \quad V(p_i, \alpha_i^k, \beta_i^k) > V(p_i, \alpha_i^l, \beta_i^l) \quad k \neq l$$

$$\Rightarrow x_i^k = 1, x_i^l = 0$$

$$(17) \quad V(p_i, \alpha_i^k, \beta_i^k) = V(p_i, \alpha_i^l, \beta_i^l) \quad k \neq l$$

$$\Rightarrow x_i^k = x_i^l = \frac{1}{2}$$

where

k = firms

i = customer groups

x_i^k = market share of firm k in customer group i

$R^k = S^k + V^k$ = total costs of screening

S^k = fixed costs of screening, to be entailed at stage I and II $S^k \geq 0$, where

$S^k = 0$ means direct-selling. That is, a vertically integrated firm will have

$S^k > 0$

V^k = variable costs of screening, to be entailed at stage III $V^k \geq 0$

Both firms can offer policies to all customers, and if their policies give the customers the same utility, then the customers of that group will be allocated evenly to the two firms. This is written down as equation (17) above. Note that it can well be that two different policies give the same expected utility to a given customer, and she is thus indifferent between them. The firm that offers a contract that gives a higher expected utility to a given customer-type captures all customers belonging to this group (equation (16) above). Competition is thus essentially Bertrand in nature. Actually, as we know (eg. see Fudenberg&Tirole 1991, ch.7), equation (14) only binds for low-risk customers and equation (15) for high-risk customers, if at all. That is, in this model, the vertically integrated firm can break the self-selection constraint(s), as is written in (15). Equation (15) also

in effect assumes that the minimum cost of an efficient vertically integrated structure (ie. one that accomplishes the task of identifying customer types) is R .

The equilibria of the game can be categorized according to the type, or "mix" of vertical strategies present in the equilibrium: there can be equilibria with symmetric or with asymmetric vertical strategies. This gives - theoretically - four possibilities in pure strategies:

- (i) both firms are vertically integrated
- (ii) firm 1 is vertically integrated, firm 2 is a direct-seller
- (iii) (ii) the other way round. Here, remember that firm 1 chooses its vertical structure first. That is, the order in which firms get to decide their vertical structures matters and makes (iii) different from (ii).
- (iv) both firms are direct-sellers

However, the following proposition is easy to prove:

Proposition 1: The following holds: Of the above categorization of equilibria,

- a) (i) cannot be an equilibrium
- b) (iii) cannot be an equilibrium in pure strategies¹¹

Proof of a: Competition is Bertrand and, if both firms are vertically

¹¹ It can be an equilibrium when firm 1 (and, possibly, firm 2) plays a mixed strategy. These cases are analyzed below.

integrated, there are sunk costs from stages I and II in establishing the vertical structure: This means that profits are at most $-S^k$ (when firms invest the sunk cost, but then realize that screening is unprofitable and revert to direct selling), thus negative, and firms can ensure zero profits by staying out.

Proof of b: Since firms are ex ante identical and firm 1 gets to choose its vertical structure first, if it finds it unprofitable to decide to be vertically integrated, then so must firm 2. QED.

From proposition 1 it follows that the pure strategy equilibria that we find are either in category (ii) or (iv). Those equilibria belonging to category (iv) have been analyzed by Dasgupta&Maskin (1986b), so they are left out of the current analysis. There are two questions to answer: under which conditions do we get an equilibrium that belongs to category (ii) and what kind of equilibria exist in category (ii)?

To answer the above two questions, we need to consider the different strategies open to the firms. Especially, we need to study out-of-equilibrium threats. The following lemma always holds:

Lemma 1: A vertically integrated firm (1) never offers a contract to high-risk customers that makes negative profits. This means, that firm 1 offers high-risk customers at most the contract that offers them full insurance at fair odds. In figure 8 this is contract C'.

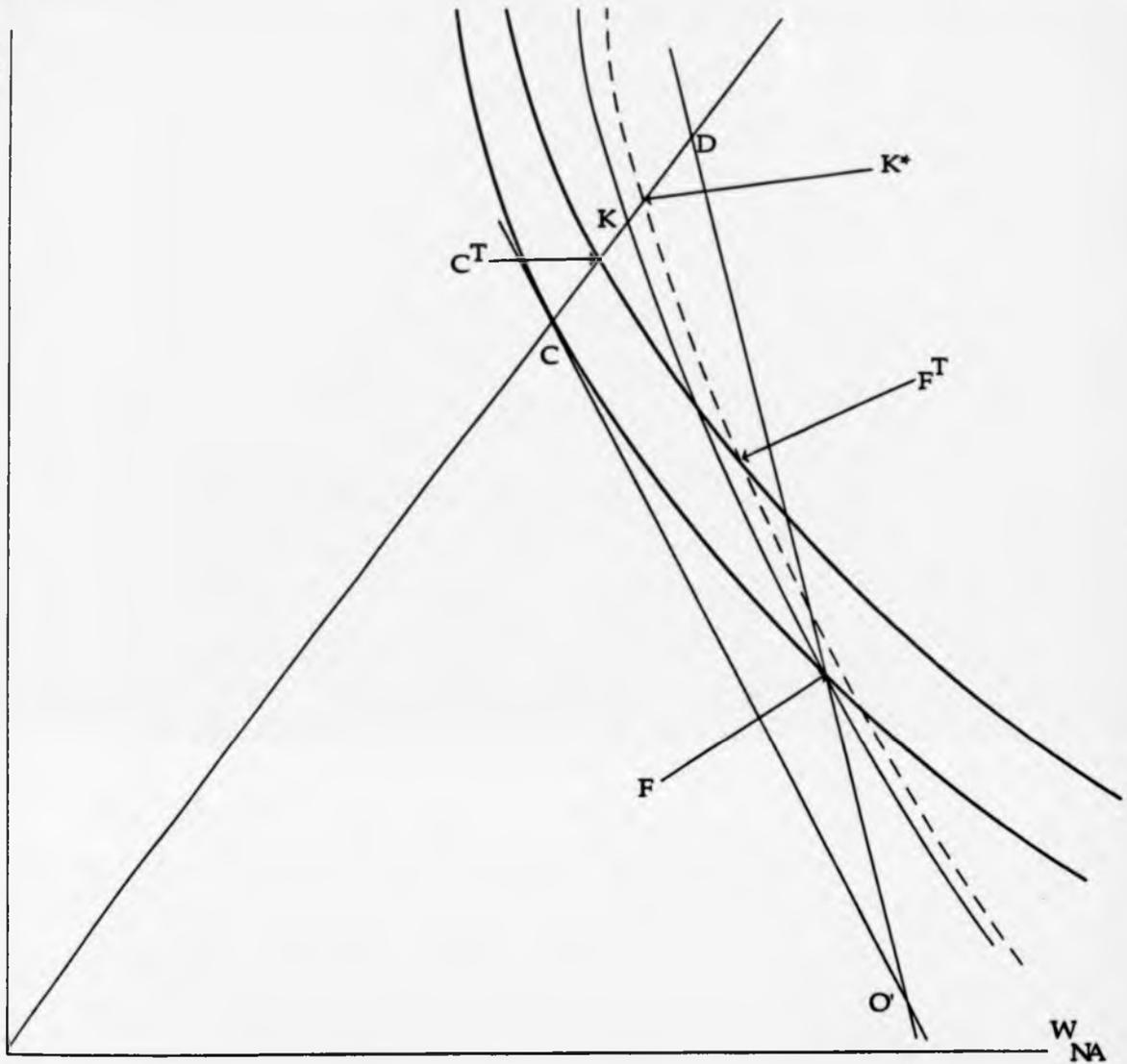


figure 9

full insurance is by a lower price, hence K and F are vertically differentiated. QED.

The equilibria with asymmetric vertical strategies can be divided further into two groups; one has multiple equilibria in mixed strategies, the other a unique pure strategy equilibrium. These are presented in propositions 6

The low-risk contract that has the self-selection constraint of high-risk customers binding when the high-risk contract is C' is F (again, see fig. 8). It might be, however, that the direct-selling firm 2 could make a better offer for both customer groups. These contracts can be found by solving the following programme:

$$(18) \quad \max_{\alpha, \beta} V(p_L, \alpha_L, \beta_L)$$

so that

$$(19) \quad V(p_H, \alpha_H, \beta_H) \geq V(p_H, \alpha_L, \beta_L) \quad \text{SS}$$

$$(20) \quad \sum_i N_i (\alpha_i - p_i \beta_i) = 0 \quad i \in \{H, L\}$$

$$(21) \quad V(p_H, \alpha_H, \beta_H) \geq \max V(p_H, \alpha_H, \beta_H)$$

For a discussion of this programme, see Spence (1978). The solution to (18) may or may not be (C', F) . Depending on the exogenous variables it is possible that the solution to (18) entails transfers from the low-risk group to the high-risk group and these contracts, call them (C^T, F^T) , are preferred by both groups to (C', F) . The solution to (18) is crucial to the equilibrium of the game since F^T is the biggest (possibly) credible separating threat that firm 2 can make. The important part is F^T . Note that equation (20) requires that total profits are zero, not the customer group-wise profits. Because competition is Bertrand in nature, this means that the high-risk customers get at least their first best, ie. contract C' . They might get an even better contract, depending on what the direct-selling firm can offer them while simultaneously breaking even.

The separating threat that is characterized by equations (18) to (21) is, however, not the only possible threat that firm 2 can make. It can make a threat where the customers are pooled, ie. offer the same contract to both high- and low-risk customers. Let's call the pooling contract that firm 2 might offer contract J^T . When we consider the threats that the two firms can make, there are two questions to ask: first, what is the biggest threat each firm can make and second, which firm's threat is credible? As for the first question, the solution for it is described in proposition 2.

Proposition 2: a) The vertically integrated firm's (firm 1) biggest threat is to offer contract pair (C', K^T) , whereby it makes a loss equal to the sunk cost of period one, S^1 .

b) The direct-selling firm can choose between a *separating threat* (C^T, F^T) and a *pooling threat* (J^T) . The separating threat is found by solving the programme (18)-(21). The pooling threat is found by drawing the *whole population zero-profits line* and choosing the contract on that line which maximizes the utility of low-risk customers. If the whole-population zero-profit line lies entirely below the low-risk indifference on which F^T lies, the direct-selling firm chooses the separating threat. If these two are tangential, the firm is indifferent and if they cross, the firm chooses the pooling threat.

Proof of a): Firm 1 has committed itself in stage I to a vertical structure. From proposition 1 we know that the vertically integrated firm will offer C' to the high-risk customers. Thus it chooses K^T so that the profits cover the variable costs of vertical integration. If it threatened to offer something

better, it would be an empty threat since firm 2 would know that firm 1 could save money by offering K^T . If it offered less, it could make the offer better for low-risk customers and still break even in stage III.

Proof of b): The programme (18)-(21) is designed to maximize the utility of low-risk customers and to guarantee zero overall profits. The biggest credible pooling threats are the ones that make zero profits and thus lie on the whole-population zero-profit line. Since the firms essentially compete for low-risk customers (because they are the ones that can provide positive profits to cover costs of either vertical integration or loss-making high-risk contracts), the biggest threat among these is the one that maximizes low-risk customers' utility. The choice between the two threats is clear. QED.

Proposition 2 can be clarified with help of figure 8. There I have drawn the indifference curves and contracts that break even individually for both groups. Let's assume that the contract pair (C^T, F^T) is the biggest separating threat firm 2 can make. The low-risk indifference curves going through F^T is dotted. To find out whether the direct selling firm chooses to make a separating or a pooling threat, draw into figure 8 the zero-profit line of the whole population (the dotted line in figure 8). As it is drawn, the separating threat is bigger, since the low-risk indifference curve going through F^T lies above the whole-population zero-profit line. The pooling contract J^T does not lie on the 45°-line since it maximizes the low-risk customers' utility and the slopes of low-risk customers' indifference curves are not equal to the whole-populations equal-profit lines at the 45°-line, but

to the low-risk equal-profit lines. As the proportion of low-risk customers in the whole population grows, the whole-population zero-profit line tends towards the low-risk zero-profit line and J^T moves closer and closer to full insurance, ie. the 45°-line. At some point the pooling contract comes a bigger threat than the separating contract.

The following proposition, which answers the second question relating to threats, namely which firm's threat is credible, is both important and easy to prove:

Proposition 3: The vertically integrated firm always captures the whole low-risk market and it does so if its threat K^T offers low-risk customers at least the same expected utility as the contract (either F^T or J^T) that the direct-selling firm offers them.

Proof: As long as the marginal costs of screening the customers are lower than the expected profit made out of a contract, it pays for the vertically integrated firm to screen. Imagine that firm 1 offers low-risk customers a contract that gives them the same expected utility as does the contract (either F^T or J^T) offered by firm 2. This would mean that the firms split the low-risk customers, according to equation (17). Now, if the marginal cost of screening a customer is just equal to the expected profit made out of an average contract, then firm 1 makes a loss of $-S^1$. But the direct selling firm makes a loss since it has designed its contract (-pair) so that it breaks even only if it captures the whole low-risk market. If the firms would split the

market, the direct-seller's threat is not credible. Then the vertically integrated firm will capture the whole market. Also, if the direct-selling firm can make the low-risk customers a better offer than K^T , then it does not pay for firm 1 to become vertically integrated since it can guarantee zero profits by being a direct-seller. Thus if vertical integration is profitable for firm 1, it will capture all low-risk customers. QED.

From proposition 3 it follows that we get two cases, one where the threat of the direct-selling firm is credible, and another where it is not. These are presented in proposition 4:

Proposition 4: Suppose there is a contract, offered by the direct-selling firm 2, (weakly) preferred to (C',F) by both customer types. Then there are two cases:

a) The threat is credible, ie. the costs of screening per contract are higher than the expected profit per contract. In this case the game always reverts to an equilibrium belonging to category (iv).

b) The threat is not credible, ie. it is profitable for firm 1 to undercut its rival in the low-risk customer market. This happens when the screening costs per contract are as high or lower than the expected (gross) profit per contract. This is a necessary, but not a sufficient condition for an equilibrium of category (ii) to exist.

Proof of a): Trivial.

Proof of b): At stage I of the game, firm 1 has committed itself to a vertically integrated structure. This means that at stage III it has to

maximize profits, taken as given the fact that it is vertically integrated. The part of investment R actually spent in stage I, S^1 , acts as a commitment. As long as firm 1 can do as well by screening the customers as by acting as if it were a direct-seller, it can choose to screen. From proposition 3 it follows that if firm 1 can offer a contract with same expected utility as F^T and J^T to the low-risk customers, it captures them all, leaving to firm 2 only the high-risk customers. The contract that firm 2 offers to high-risk customers, C^T or J^T is at least as good as contract C' , the contract firm 1 at most finds profitable to offer them. Especially, if the preference of high-risk customers of C^T or J^T over C' is strict, this means that firm 2 makes a loss because it only gets the loss-making high-risk contracts, not the profit-making low-risk contracts. This does not guarantee non-negative profits for firm 1, because the fixed costs of a vertical structure can be higher than the gross profit. Thus this is not a sufficient condition. QED.

An immediate lemma follows:

Lemma 2: If $(C', F) = (C^T, F^T)$, then firm 2 breaks even. If (C^T, F^T) is different from (C', F) and it and J^T are not credible threats, firm 2 offers (C', F) in order to maximize its profits. This means that as long as (C^T, F^T) or J^T is not a credible threat, the firms offer identical products in the high-risk customer market and share it in equal proportions. The direct-selling firm gets no low-risk contracts if it faces a vertically integrated rival. If $(C', F) = (C^T, F^T)$ this is equilibrium is a reactive equilibrium.

Proof: Follows from lemma 1, propositions 3 and 4, equation (17) and the

Bertrand nature of competition. QED.

Figure 8 can be used to clarify the calculation of which firm's threat is credible. Let's suppose that K^T is the low-risk contract that results in a loss of S^I to firm 1, ie. the vertically integrated firm covers its variable costs of screening in stage III of the game. Let's suppose then that the biggest threat of the direct selling firm is a pooling threat. If the whole-population zero-profit line goes below the low-risk indifference curve that goes through K^T (as is drawn in the figure), or at most is tangential to it, then the pooling threat of the direct-seller is not credible. If the biggest threat of the direct selling firm is a separating threat, then the reasoning goes as follows: If the low-risk indifference curve that goes through F^T (the direct-seller's low-risk contract) lies lower than the low-risk indifference curve going through K^T or at most is the same indifference curve (as drawn in figure 3), then the separating threat of the direct seller is not credible. If either the indifference curve going through F^T and K^T is the same or the indifference curve going through K^T is tangential to the whole-population zero-profit line, then the firms would share the low-risk market. This is, however, not possible according to Proposition 2. Sharing the market makes the direct-seller's threat empty and the vertically integrated firm captures the whole low-risk market. Note that in case of increasing screening costs firm 1 can readjust K^T to take this into account. Specifically, the vertically integrated firm will always have to screen all high-risk customers and, according to (17), half of the low-risk customers. It is thus

sufficient that firm 1 breaks even after screening that amount of customers. If the costs of screening increases when the number of customers screened increases, then the fact that the vertically integrated firm does not have to screen all customers (if the firms offer them contracts with the same expected utility) enables it to make a bigger threat.

To get the necessary and sufficient conditions for a type (ii) equilibrium (ie. with asymmetric vertical strategies) to exist, we obviously have to add a condition of non-negative profits to proposition 4b. The non-negativity constraint of the vertically integrated firm can be expressed, using notation from fig. 9, as

$$(22) \quad \Pi^{VI} = (1-N_H)(KD') - R \geq 0$$

where K is the profit-maximizing contract of the vertically integrated firm offered to low-risk customers. Now it is also clear that there is vertical product differentiation in a market with asymmetric vertical strategies. This is stated as proposition 5:

Proposition 5: The equilibrium with asymmetric vertical strategies produces vertical product differentiation in the low-risk customer market.

Proof: the vertically integrated firm offers low-risk customers K , which they all choose. The direct-selling firm offers them F , which lies on a marginally lower indifference curve than K . K gives full insurance and marginally higher expected utility. The only way partial insurance can compete with

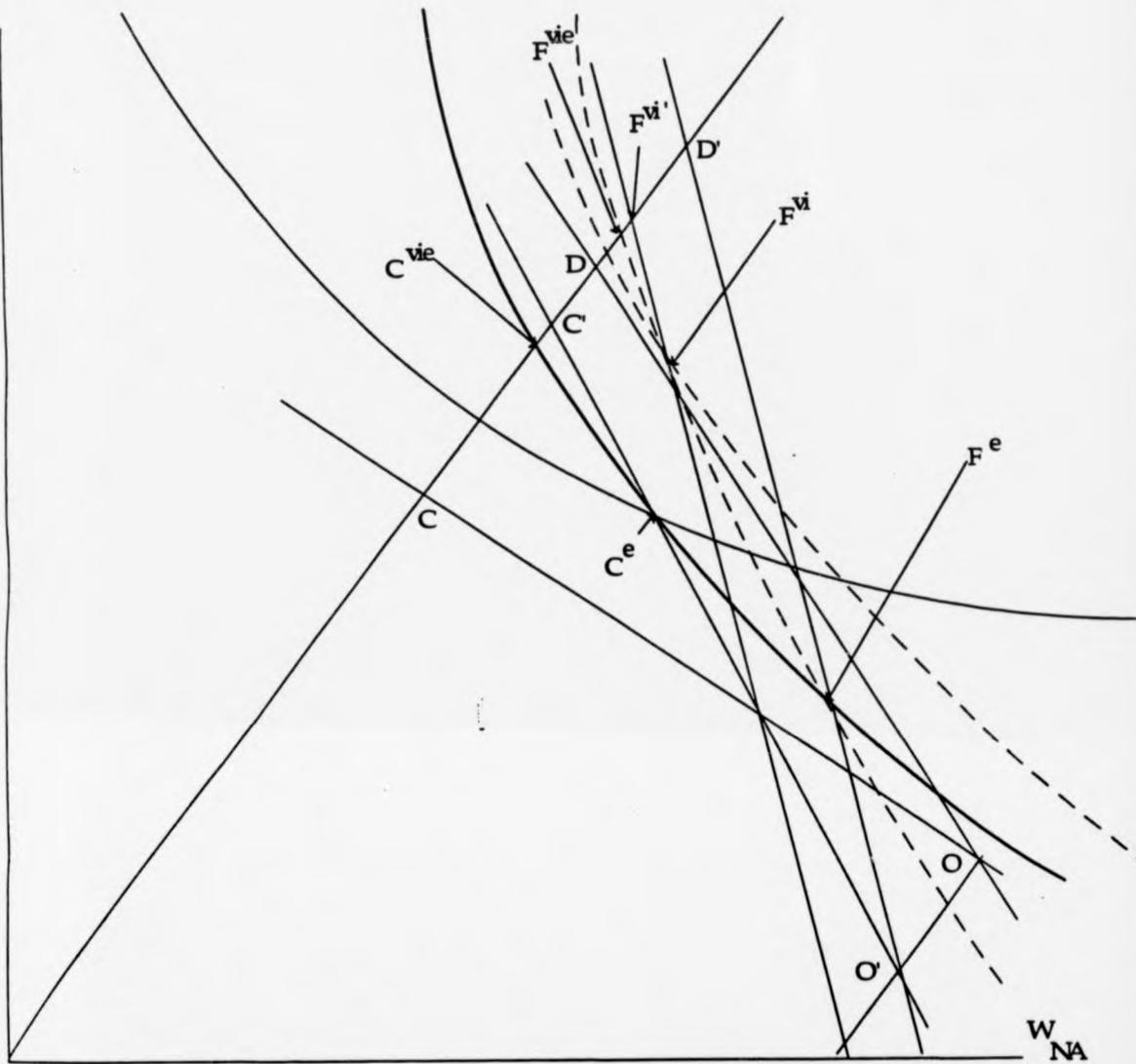


figure 10

same contracts as the firms in the competitive market of section 2. These are contracts C^e (for high-risk customers) and F^e (low-risk customers) in figure 10. As in the previous section, the vertically integrated firm captures the whole low-risk clientele, but is constrained by the contract that the direct seller offers them. The vertically integrated firm can still make

and 7.

Proposition 6: If the necessary and sufficient conditions are met and the vertically integrated firm makes zero profits (ie. (22) holds with equality), the game has an infinite number of subgame perfect reactive equilibria. These can belong to any of categories (ii)-(iv). Especially, the firm (2) that gets to choose its vertical structure after the other can under some conditions be the vertically integrated firm. The equilibria have the following characteristics:

- a) Both firms offer C' to high-risk customers
- b) The direct-selling firm offers F to low-risk customers
- c) The vertically integrated firm offers K to low-risk customers and captures them all
- d) Both firms make zero profits.
- e) Firms use the following mixed strategies: firm 1 plays "vertical integration" (VI) with probability ϕ and "direct selling" (DS) with probability $1-\phi$, $\phi \in [0,1]$. Then firm 2 always plays DS if firm 1 plays VI. If firm 1 plays DS, however, firm 2 can play VI with probability τ and DS with probability $1-\tau$, $\tau \in [0,1]$. Thus we have an infinite number of possible equilibria (and three pure strategy equilibria: (VI,DS);(DS,DS);(DS,VI)).

Proof of a: For the vertically integrated firm, this follows from lemma 1 and Bertrand nature of competition. For the direct-selling firm, this follows from lemma 2.

Proof of b: Follows from lemma 2.

Proof of c: The sufficient and necessary conditions of an equilibrium with asymmetric vertical strategies state that the vertically integrated firm captures the whole low-risk customer market. The profit-maximizing contract for this is K (see fig. 9) that is marginally preferred to F by low-risk customers.

Proof of d: Follows trivially from the non-negativity constraint (18) holding with equality and lemma 2.

Proof of e: From d) it follows that both firms make zero profits as long as both of them are not vertically integrated and as stated in proposition 1a, both make negative profits if they both are vertically integrated. That excludes equilibria belonging to category (i). Then firm 1 is indifferent between being a direct-seller and being vertically integrated and can thus randomize between these strategies. As firm 2 observes firm 1's choice of vertical structure between deciding its own, it, too, can randomize given that firm 1 is not vertically integrated. If firm 1 is vertically integrated, proposition 1a forces firm 2 to become a direct-seller. Since the choice of the weights (=probabilities) on different vertical strategies do not matter, firm 1 can choose ϕ freely, as can firm 2. These freedoms yield an infinite number of subgame perfect reactive equilibria. The three different equilibria in pure strategies are clear from the above discussion. QED.

Proposition 7: If the necessary and sufficient conditions hold and the vertically integrated firm earns strictly positive profits, ie. (22) holds with

inequality, the game has the following unique¹² pure strategy subgame perfect reactive equilibrium:

- a) Both firms offer C' to high-risk customers
- b) The direct-selling firm offers F to low-risk customers
- c) The vertically integrated firm offers K to low-risk customers and captures them all
- d) Firm 1 earns positive profits and firm 2 makes zero profits

proof of a-c: See proposition 6.

proof of d: Follows trivially from lemma 2 and (22) holding with inequality.

QED.

There are thus three possible equilibrium categories, when mixed strategies are included: categories (ii) and (iv) and the above analyzed mixed strategies. The mixed strategy equilibria case is "between" (ii) and (iv) in the sense that the outcome corresponds to either category (ii), (iii)¹³ or (iv), although the strategies are (can be) different.

Although the game is plagued with an infinite number of equilibria, it is noteworthy that this is a problem only with very specific values of the

¹² Remember that in this section it is assumed that firm 1 gets to choose its vertical structure before firm 2. Once this assumption is relaxed, the equilibrium is not unique, but there are two symmetric equilibria, both in pure strategies.

¹³ Categories (ii) and (iii) are similar in outcome, the difference being in which firm chooses the vertically integrated, which the direct-selling strategy.

exogenous variables that give zero expected profits for both the vertically integrated and the direct-selling firm. A small change in any of the exogenous variables will lead either to the traditional adverse selection model, or to the solution with asymmetric vertical (pure) strategies depicted in proposition 7, the main result of this section. The model also yields clear conditions under which the different equilibria exist. For the pure strategy equilibrium with asymmetric vertical structures, the model also yields some clear-cut predictions. Since the vertically integrated firm captures the whole low-risk customer market, and the firms split the high-risk customers, the vertically integrated firm is bigger. This of course follows partly from equation (17) which stated that the high-risk customers are split evenly between the firms. But even with other assumptions about how the high-risk customers are divided between the firms, it is highly likely that the vertically integrated firm is bigger than its competitor. This is so since in order for vertical integration to be profitable, a reasonably high proportion of low-risk customers is necessary. If $N_L > \frac{1}{2}$, the probability of firm 1 being bigger than firm 2 is obviously one. The operating costs of the vertically integrated firm are higher than those of the direct-seller, both in absolute and relative (ie. per contract) terms. Also, its claims expenditure is higher in absolute terms (at least under equation (17)), but lower per contract. The direct-seller's average claim per contract is $p_H d$, whereas the vertically integrated firm's is $[\frac{1}{2}N_H p_H d + (1 - N_H)p_L \beta_L] / [\frac{1}{2}N_H + (1 - N_H)]$. It is easy to verify that the latter is smaller than the former. And, by now self-evidently, the vertically integrated firm has

higher expected profits than the direct-selling firm.

Introducing competition into the model has a profound effect on the vertically integrated firm. The conditions under which vertical integration is profitable are much more stringent than in the monopoly case. Compared with a vertically integrated monopoly, a vertically integrated oligopolist loses most or all of its profits. A vertically integrated oligopolist does not make any gross profit on high-risk customers, and its profits per low-risk customer are down from the monopoly case. The costs of vertical integration, however, stay the same. If the consequences of competition are significant from the vertically integrated firm's point of view, they are significant from the customers' point of view, too. Both customer groups experience an increase in their utility, high-risk customers attaining (at least) their first-best level. The utility of low-risk customers does not increase as much, and ends up on a level equal to the standard model with only direct-selling firms. As in the case of monopoly, the welfare consequences of vertical integration are not universally good or bad. Using the same definition of welfare as in the monopoly part, the following can be concluded: compared to the direct-selling case, the high-risk customers lose $(C'C^T)$ each and the low-risk customers the distance (measured along the 45°-line) between the indifference curves on which contracts F and F^T respectively lie. Let's call the vertically integrated firm's contract that gives low-risk customers the same utility as F^T , K' (see figure 9). The losses (from vertical integration) of low-risk customers are cancelled out by the profit

gains of the firm, but the losses of the high-risk customers not necessarily. Thus the gains in profit (from vertical integration) net of the impact on welfare of low-risk customers have to be compared with the high-risk customers' loss of welfare. This is done in equation (23):

$$(23) \quad (1-N_H)(K^*D') - R - N_H(C^*C^T) \geq 0$$

The first term is the gross profit of the vertically integrated firm, after the welfare losses of low-risk customers (ie. $(1-N_H)(KK^*)$) have been subtracted. The second term represents the costs of vertical integration and the third term is the number of high-risk customers times the welfare loss that a high-risk customer experiences when the equilibrium shifts from a pure direct-selling equilibrium to one with asymmetric vertical strategies. If the pure direct-selling equilibrium tends towards the solution (C^*,F) where both groups' contracts individually break even, then the first term in (23) tends to $(1-N_H)(KD')$ and the last term to zero, ie. (23) tends to (22). In that case vertical integration always Pareto-dominates direct-selling¹⁴.

¹⁴ On the margin, where (22) holds with equality and (22)=(23), direct selling and the equilibrium with asymmetric vertical strategies are equal welfare-wise, given that the agents that the vertically integrated firm hired are kept at their reservation utility. If they experience a rise in utility, then, even in this case, vertical integration is welfare enhancing.

III.5 Oligopolistic vertical strategies with adverse selection and moral hazard

This section generalizes the results of the previous one to cases where moral hazard is present. However, I will introduce two simplifying assumptions; the first one is used by Rothschild and Stiglitz, and here in the section with competitive markets, namely that each type of contract must individually at least break even. This assumption does not alter the nature of the equilibrium outcomes, but enables me to use a simpler equilibrium concept than in section 4, namely the subgame perfect Nash equilibrium¹⁵. The second assumption is that the direct selling firm makes no pooling threats. This again simplifies the analysis without changing the nature of the equilibria. Thus here, the game is exactly the same as in the previous section, but now the incentive compatibility constraint of moral hazard is added to the maximization problem of the firms. In this section, I limit the analysis to the contracts offered in an equilibrium with asymmetric vertical strategies, and do not analyse the game as such¹⁶.

The effect of the above assumptions is essentially to reduce the scope of the threat that the direct seller can make: it can at most threaten to offer

¹⁵ This assumption makes the analysis somewhat simpler, and thus the figures more readable. The analysis goes through also when the reactive equilibrium is employed.

¹⁶ The analysis of the game follows that presented in the previous section.

(gross) profits per low-risk contract, since it can offer low-risk customers contracts (contract F^{VI} in figure 10, if it cannot monitor effort) with better coverage than the direct seller (since the vertically integrated firm can break the self-selection constraint). The vertically integrated firm's profit, if effort is not screened, is

$$(24) \quad \Pi^{VI} = (1-N_H)(F^{VI}D') - R$$

The firm makes no profit on high-risk customers, and its direct-selling rival makes zero expected profits. The only change compared to the pure adverse selection model is that neither customer type gets full insurance, since the incentive compatibility constraint of moral hazard binds for both organizational forms.

The nature of the equilibrium critically depends on whether or not the vertically integrated firm monitors effort. If effort is not or cannot be monitored, the high-risk contract that the firms offer is independent of organizational form, just as in the pure adverse selection model of the previous section. The firms thus divide the high-risk customers among themselves. The reason for this is that both firms find out the type of customers, the vertically integrated firm through screening and the direct seller by using contract design to make high-risk customers reveal their type. Both firms are constrained by the incentive compatibility constraint of moral hazard, however. Thus they cannot offer full insurance, but

competition drives the offered contract to the effort zero-profit line. If, however, the vertically integrated firm can monitor effort, the equilibrium changes quite dramatically: the capability of monitoring effort leads namely to a situation where the vertically integrated firm can offer a better contract to high-risk customers than the direct-seller. There is thus vertical product differentiation in both the high- and low-risk customer markets. The reason for this is that the incentive compatibility constraint does not any more bind the vertically integrated firm, and it can offer high-risk customers full insurance, and a contract that these strictly prefer to the contract that the direct-seller at most can offer. As the vertically integrated firm captures all low-risk customers even if it cannot monitor effort, this means that it is a monopolist that is constrained by the possibility of a direct-selling competitor. In terms of figure 10, the vertically integrated firm offers high-risk customers contract C^{vie} , which they strictly prefer to contract C^e , the contract that the direct-seller at most can offer them. The low-risk customers are offered a different contract, too, compared to the case where the vertically integrated firm cannot monitor effort. They get contract F^{vie} , which also provides full insurance. The vertically integrated firm makes a profit of size

$$(25) \quad \Pi^{VIE} = N_H(C^{vie}C') + (1-N_H)(F^{vie}D') - (R + T)$$

Compared to the case where effort is not screened, the firm makes a profit on high-risk contracts, and its profits per low-risk contract are higher.

These increases in gross profits have to be larger than the rise T in screening costs.

The welfare analysis follows closely that of the previous section for the case where effort cannot be monitored: it is possible that vertical integration is privately optimal and socially undesirable. When effort can be monitored, the situation is the same. But let's assume that the vertically integrated firm is present, and cannot monitor effort. Let's further assume that there is, say, a change in technology that makes screening of effort cheaper, so that it becomes optimal for the vertically integrated firm to screen effort, too. How does this affect social welfare? It turns out that this step from no screening of effort to screening of effort is always socially optimal when it is privately optimal. The reason for this is that the welfare of customers does not change since they stay on the same indifference curves. The only change is in the profits of the vertically integrated firm, because its rival makes zero expected profits whether or not it is present in the market, and has no sunk costs to lose by exiting. Thus, if the vertically integrated firm's profits increase, which of course is the condition for making the investment that allows screening of effort, then this is a Pareto improvement.

III.6 Other interpretations

As economists already are aware, the number of areas where asymmetric

information plays a role is huge. In the following, I list a few specific topics, where moral hazard and/or adverse selection are of importance, and the question studied in this chapter, namely the choice of organizational form, also plays a role. The close link between categorization and this chapter's analysis will also become apparent.

a. labour market, education and recruitment policy

The model can be interpreted in terms of the Spence (1974) labour-market model. Think of the monopoly¹⁷ as the firm (the employer) and high- and low-risk customers as potential employees with low, respectively high, productivity. Let the firm choose between using self-selection constraints as in the original model and hiring an agent that can find out the true type of any job seeker¹⁸. This would correspond to a version of the current model where moral hazard is absent. Further assume that the agent faces some competition, so that it cannot use monopoly pricing. The case where the monopolistic insurance firm decides to be vertically integrated corresponds to the firm hiring an agent, say a psychology consultancy, to screen applicants. This interpretation has alarming consequences to the value of education: if the employer can find out the true type of the job-applicant (and it is profitable for it to do so), then it does not pay to invest in education! The real consequences are probably less dramatic. Education

¹⁷ The earlier mentioned paper by Nalebuff&Scharfstein, as well as Guasch&Weiss (1980) study testing in a competitive market environment.

¹⁸ This section relies on Spence's assumption that education acts only as a signal.

would probably be diminished to serve the role of categorization: for example, big firms might only screen applicants that hold an MBA for their managerial positions, but it would not pay off to acquire a PhD to signal above average productivity (among those holding an MBA). The employer can thus make a decision whether to itself invest in alleviating the problem of asymmetric information, or to let the potential employees do the investments. Although there are papers studying testing in a labour market environment, they seem to emphasize different aspects of the matter. They do not discuss the effects testing has on the potential employee's decisions to invest in education, but use fines to replace educational choices of the employees. While the above point is not made here for the first time (at least Milgrom&Roberts (1992, p. 342) discuss it shortly), this is to my knowledge the first formalization (although in an insurance framework) of the problem. Here, as in the insurance model, the assumption that the agent can find out the true type of the customer/job applicant may seem restrictive. Adding uncertainty into the model would, however, only alter the calculation of the monopoly insurance firm/employer accordingly, as long as it is risk-neutral.

b. internal organization of creditors

Credit markets are a well-known example (see eg. Stiglitz&Weiss 1981) of markets where asymmetric information creates considerable damage, and also markets where many of the institutions (witness credit rating agencies, for example) and organizational forms reflect this fact. Again, the value of

extra information is central: does it pay, eg. for the potential creditor, to make an extra audit before granting a loan, and if so, then how would it change the terms of exchange? Does it pay to make credible the threat (by means of building an organization for this task) of an extra audit in case of bankruptcy, or is the ex ante screening enough? How does competition among creditors change the analysis? This chapter's analysis would suggest that a creditor that chooses not to monitor carefully would be wise to choose a niche where it can use observable characteristics of potential customers for categorization. If this is not possible to an extent large enough to ensure viable operation, then the creditor has to invest in an organization that is capable of screening the potential customers. This might not be profitable for all creditors, however, as the analyses of the oligopoly sections show.

c. internal organization and personnel policy

Another job market related interpretation is that of internal organization. The agent of the vertically integrated monopolist can be thought of as a supervisor, who either can or cannot screen the effort levels of employees. This would bring collusion into the analysis, as discussed by eg. Tirole (1986): the supervisor is payed to screen the employees' effort (and type), but these can bribe her to report (falsely) a high effort level. Collusion by its nature precipitates a dynamic analysis, which the current framework cannot provide with the tools used. But assuming collusion away, the question of whether or not to be vertically integrated can be reformulated

into whether or not to hire a supervisor. Here is an interesting link to the categorization literature mentioned in the introduction of this chapter: if the customers, or in this context, employees are homogenous enough, then categorization can work as a disciplining device. If employees (or their tasks) are (too) diverse, then a supervisor is a good investment. This brings the current analysis close to the idea of relative performance pay, and has implications to personnel policy. A firm might expect a (more) specialised workforce to have a higher theoretical productivity level than a homogenous workforce could at its best achieve. The price is that relative performance pay becomes impossible because of the heterogeneity of tasks and qualifications, and the theoretical edge in productivity might not be realised, if the firm does not build a supervision and incentive structure that unifies the workers' and the firm's interests. Although I have discussed the problem here strictly in an internal organization framework where the problem was how to pay lower level employees, the analysis extends itself easily to the question of how should the managers of a firm be rewarded.

d. organization of regulatory institutions

This is a timely topic in economics as the first analyses of 80's privatization programmes are appearing. As is well known in the literature, and on the field as well, there certainly is a moral hazard problem between the regulator and the firms, and there might even be an adverse selection problem, if entry is allowed. This chapter's monopoly model can be used

to analyse whether it pays for the regulator to invest in in-depth knowledge of the industry and firms so that the regulator can find out the true state of the industry, or whether it is better to lean back and rely on self-selection and incentive compatibility constraints in regulation. An example could be the regulation of insurance markets. The regulator can choose between hiring such expertise that it can screen the true costs of the firms, and the quality of their policies. If this is judged to be too expensive, the regulator has to settle with second best and formulate the regulations so that the policies firms offer are as close to first best as possible

III.7 Conclusions

Insurance markets display a plethora of different organizational forms: vertically integrated firms, partially vertically integrated firms, firms without a vertical structure, and firms that use a different kind of vertical strategy in different markets. One aim of this chapter was to shed light on these organizational issues. By no means do the models of this chapter provide a complete answer, but, I believe, they open up an interesting way to look at these questions. The assumption - as clearly is the case in practice - that a vertically integrated firm gets more information (here, more heroically, possibly perfect information) about prospective clients provides a plausible, and in terms of theoretical work on vertical structures, new motivation for vertical integration. One of the questions that was not tackled is the exact form of vertical integration that is optimal.

This issue has to some extent been discussed by Grossman and Hart (1986). What the models of this chapter do not analyze is the degree of vertical integration: here it is a zero-one decision.

The monopoly model is the "benchmark" model of this chapter. The model lends itself to other interpretations, too: eg. the decision of an insurance firm to be vertically integrated can be seen as the decision of an employer to hire an agent to screen job applicants, or as a decision of a firm to hire a supervisor to screen its production workers. The duopoly model provides clear results, yielding a unique subgame perfect reactive (and in some conditions, a subgame perfect Nash) equilibrium in asymmetric vertical strategies under given conditions. Not only do the firms choose different vertical structures, but they are (or, actually, one of them is) engaged in vertical product differentiation, too. A fixed investment can be used by one firm to relax price competition in the same way as in the vertical product differentiation models. Vertical product differentiation only arises in one part of the market. This model also yields a list of predictions concerning the differences in performance between the two (or more) firms with different vertical strategies. Most of the analysis was carried out geometrically by using the fact that the axes of the figures measure wealth of both customers and the firm(s) and the 45°-line can thus be used to measure expected utility and profits. To my knowledge, this is the first time that geometric analysis of adverse selection and moral hazard have been combined.

It should be noted that the models have been cast in the framework of only one insurance market, or more accurately, one group of customers in a given insurance market that cannot be told apart without direct screening. True world insurance markets will necessarily have several such groups, and as most insurance firms deal with several insurance markets, the number of such groups multiplies. Also, true world markets will probably display some degree of horizontal product differentiation, which was not dealt with in this chapter. Thus, when using these models to reflect real world observations, care should be taken to pay attention to these limitations of the models. Nothing prevents enlarging the models to entail several identifiable (by eg. categorization) customer groups, but as categorization is necessarily cheaper than screening, this would only amount to the duplication of the current models. These reasons have restricted the current analysis to theory only.

Some casual empirical observations can, however, be interpreted in light of the models' insights. In the mid-80's, the UK insurance industry was active in getting vertically integrated, buying up different agencies and making exclusive deals (eg. with building societies) to "get near" the customer. This trend has more or less been reversed in recent years, when new direct-selling firms have appeared and old firms have been busy establishing similar arrangements. This shift in vertical strategies may be due to a mistaken initial strategy, but also to a change in technology. If the current technology (whether that means better data, computers or actuaries

does not matter) gives the needed information without monitoring the client, then a branch network becomes obsolete. It might be, for example, that actuaries have become better in squeezing relevant information from data available through categorization, thus making monitoring unnecessary. It is enough that the accuracy is good enough to worsen the expected results (the amount of losses) less than are the certain savings of scrapping (or not establishing) a branch network. It is to be expected that the models are more suitable to situations where the information obtained by categorization is relatively inaccurate. Personal insurance business probably lends itself more easily to categorization than eg. industrial insurance.

Although the appearance of direct insurance might seem to be in conflict with the analysis, which after all shows that under certain circumstances vertical integration leads to higher profits than direct selling, this is not so. The analysis can be read the other way as well. It seems that in the UK insurance market, for some reason or the other, the conditions have changed such that vertical integration is not as profitable that is used to be. This change can be technical, as discussed above, but can also have its causes in changing tastes. Customers could, for example, be more ready to accept insurance policies as a commodity, and thus by them mainly on price. Whatever the reason, the model can be used to analyse it (with given restrictions, e.g. the preclusion of horizontal product differentiation). In the model, the higher profits of the vertically integrated firm stem from its

ability to identify low (and high)-risk customers. If the direct seller can achieve this with lower costs, then it should have higher profits, or drive the vertically integrated firms out of the market.

The adverse selection and moral hazard models, cast in an insurance framework, are cornerstones of economics of information. Most of the previous work on adverse selection has maintained the assumption that different customers cannot be told apart, and work on moral hazard the assumption that effort cannot be detected. By relaxing these assumptions, a surprisingly rich model arises. This model lends itself to study vertical integration, as I have done in this chapter. But, in more general terms, this chapter's models are about the value of extra information in different competitive situations, when information gathering is costly. In many, though not all, real world markets and exchange situations, the parties would benefit from having more information. Different constraints, time, money, etc. make the pursuit of perfect information often unprofitable. A utility-maximizing agent will conduct a calculation to decide whether the extra piece(s) of information is worth the disutility of getting it. Here, I have assumed that perfect information is obtainable, albeit with a cost. One obvious way to extend the model would be to assume that a vertically integrated firm does not obtain perfect, but only "better" information (say, in a first order stochastic dominance sense. For a discussion of the value of information along these lines, see Holmstrom 1979). To keep the model simple and to concentrate on the main questions, this was not done here.

IV OLIGOPOLISTIC SERVICES AND COST FUNCTION ESTIMATION

IV.1 Introduction

Retailing and services are two relatively neglected areas of industrial economics. Of services, banking is probably the one industry that has attracted the most interest (see Hannan 1991 and Mayer&Vives 1993). The Hotelling and Salop (see eg. Tirole 1988) models of product differentiation can be interpreted as models of retailing. One branch of the literature where retailing plays a prominent part is that of vertical relations (for a recent survey, see Waterson 1993), as these would not exist without at least two layers of organization, the other often being retailing. In recent years, one of the main criticisms of IO has been the wide(ning) gap between theoretical and empirical literature (see eg. Peltzmann 1991). This is apparent in the (empirical) literature on cost functions. It seems that here, neoclassical economics is still used as the framework of thought, and no lessons have been taken from the recent theoretical literature. The usual contents of this literature is a new data set, possibly combined with a new feature in the econometric function. As an example of the latter, see Braeutigam&Pauly (1986), who correct for quality bias in a regulated insurance environment.

This chapter rests on the observation that in services (and retailing, which from now on will be subsumed under the heading of services), the principal unit of production is the branch, and the production technology

used lies on this level. There can be important functions, such as logistics, that are not performed at branch level, but the actual product that the customer buys is produced there. Another observation (which is hardly new) is that the branch network is one of the strategic tools in the toolbox of services firms. Branches, through their location, are used to increase market power. As an example, think of a supermarket chain operating in the area of greater London, and another operating in Sweden. Let's assume that both use exactly the same technology. Both have a clientele of roughly 8 Mio., but these are located much more densely in London. To reach all potential customer the supermarket chain in London probably needs a number of branches that is far smaller than the number of branches needed in Sweden. If we assume that both chains sell the same amount of goods, then the chain in Sweden will have higher average costs because of the bigger number of branches. But even the London firm will probably not rely on just one store, even if this was technically feasible. If the only costs above the variable (=marginal) costs of purchasing the goods from manufacturers are the fixed costs of establishing a branch, then one store would be the cost minimizing solution for the supermarket chain. But the chain will have several in order to optimize the mix between market power and costs. There are two effects at work: *the captivation effect* of increasing market power through additional branches, and *the cost effect* of (possibly) increasing average costs, that the firm has to balance when deciding whether to increase output at the existing branch(es) through lower prices, or to add another branch to the network. A profit-maximizing firm thus

does not necessarily minimize firm level costs, as assumed in neoclassical production theory. This means that the firm-level cost function is not the dual of the production function, as traditional theory claims, but this relationship holds at the level of production unit, the branch. The above suggests that it is entirely possible, and if the fixed costs dominate variable/marginal costs, even probable, that we can observe situations where an industry exhibits increasing returns to scale at branch level, and diseconomies of scale at firm level. The effects of a multimarket environment can affect cost function estimations even in manufacturing. In an oligopolistic setting, it might be profitable for a firm to produce two products even if there are diseconomies of scope, if the strategic advantages override the cost-inefficiency. Profit-maximizing oligopolistic manufacturers take into account the cost and the strategic aspects and it is not at all certain that the cost-minimizing solution is the same as the profit-maximizing one.

For a multi-product firm, the characteristics that differentiate it from its rivals are the products it offers, and the location where these are offered. Thus the number (and location) of branches can be equalled to that of adding product lines. To use the above supermarket example, think of two rival supermarket chains. Let's assume that on the product level, there is no product differentiation, so that the beer *X* of chain *A* is the same as that of chain *B*'s (of course, both of them can have several beer labels on offer). The customer's decision which one's store to visit depend on the number

of products available, and the location of the store, if we take the view that a customer values variety as such as in Spence (1976) and Dixit&Stiglitz (1977), or if (s)he wants to buy several different types of beer. If *A* has a larger supply of different beers, but its nearest store is located further away than *B*'s, the customer will weigh these differences against each other when deciding which firm's store to visit. This means that for a multi-product firm, the branch network should be viewed as part of the output.

The above observations have major consequences for the estimation of cost functions of services industries. These are discussed in more detail in a later section. The empirical part of this chapter consists of constructing and estimating a cost function for a services industry that, especially relative to its importance for the functioning of a modern economy, has received little attention, namely insurance. The limitations of the data, a large number of products (39) and a small number of firms (21) requires some restrictions on the cost function so that results can be obtained. These are imposed so that general measures of firm- and branch-level economies of scale and scope can be obtained, the price being that product-level measures become unobtainable.

The rest of the chapter is organized as follows. In the next section, a theoretical model is constructed and solved in order to analyze the above discussed insights. In the third section, the data and the market under study, the Finnish non-life insurance market, are discussed. In that section,

the empirical model is constructed. I also compare it to those used in the existing literature. The fourth section contains the empirical results, and some comparisons to existing literature on cost function estimations in banking and insurance. The fifth section concludes.

IV.2 A model of branch networks

A branch network is strongly connected to the idea of clusterings of customers. Whether you think of a nation-wide branch network or a branch network in a single city, then each branch usually services a different set of customers. The situation is thus reminiscent of a multimarket environment. The study of oligopolies in a multimarket setting using game-theoretic tools is a fairly recent phenomenon: examples of such studies are Brander&Eaton (1984), Lal&Matutes (1989), Shaked&Sutton (1990) and Dobson and Waterson (1993). They specify the different markets in terms of different products, not different geographical markets, and as will be clear, this makes a difference. The customary modelling framework for branch networks is that of the Hotelling beach, but in its standard form, it does not allow for more than geographical product differentiation, or differentiation in tastes (ie. product differentiation only in one dimension). One answer to this is the hexagonal city literature (see eg. Nooteboom 1993), but it does not allow for discontinuously distributed customers. A more recent approach, that can be shown to nest several of the more traditional approaches is that of

discrete choice models (for a thorough discussion of the topic and its links to earlier models, see Anderson, de Palma and Thisse 1992). None of these models can easily deal with the nonconvexities arising in a locational setting. This old problem of urban economics (see Stahl 1987) is encountered in this study, too, and rather than tackling it head on, I will model around it.

The situation that I want to model is the following: think of two nearby cities, *A* and *B*, one (*A*) possibly larger than the other. Or think of a city centre and a suburb. There is a travelling cost to get from one city to the other. Furthermore, potential customers are heterogeneous when it comes to their tastes, and some are always willing to make the trip to get their preferred product. I am thus assuming that difference in tastes is greater than that in locations. This is an important assumption, since it will guarantee continuous, monotonic demands for the firms. Let the firms choose their location both with regard to tastes and geographical location. They can, however, only occupy one location on the taste axes. This means that they sell the (physically) same product in both cities. But they can have two branches, one in each city. The questions I want to study are:

- where do the firms locate themselves with regard to tastes?
- how many branches and where do they open?
- what are the prices they choose, and thus quantities and profits?
- specifically, do firms with different locations/branch networks price differently?

- is it possible that an oligopolistic firm does not minimize costs?

One possible way of modelling these problems would be to use representative consumers with quadratic utility functions, as eg. in Singh&Vives (1984). This approach has some inherent problems, however. Firstly, it is not very straightforward to endogenize the product differentiation decision of the firms, or the differences in the size of population. The model, as all the other models, runs into difficulty when the transport cost (distance) between the cities grows "too big": for a given price, a firm does not lure any demand from the other city, but lowering the price it gets at least some products sold. It also seems to me that in the case of geographically different customer populations, the representative consumer is not a very intuitive construct: it is easier to understand that some consumers from city *A* travel to *B*, and some don't, than to have a representative who does some of the shopping at *A*, some at *B*. Another possibility would be to use the discrete choice approach, but the solving of these models often rests on symmetry assumptions, and it is precisely the asymmetries that are of interest here. Furthermore, it would be difficult to make the transport costs not depend on the quantity bought, a feature of some models that has been criticized (see Stahl). For the products of interest here, eg. banking and insurance, household nondurables, it seems more plausible that the transport cost is a lump sum that does not depend on the amount bought.

There are extensions of the Hotelling framework into more than one dimension (see Eaton&Lipsey 1975, Ben-Akiva, de Palma&Thisse 1989). The latter assume that the geographical line is a Hotelling beach, and the consumers are distributed on a circle according to their tastes. This otherwise interesting approach does not easily allow non-homogenously located customers, either geographically or with regard to tastes, however.

The model used here is none of the above, but a simple extension of a Hotelling model. So let's assume that there are two cities, *A* and *B* (there could be more, but the assumption of two cities allows the study of the questions listed above, without undue difficulty), and the population of each city is uniformly distributed on a line of length one. The density of customers is 1 in the smaller city *B*, and $\alpha \geq 1$ in the larger city *A*. This line represents the tastes of the customers, and there is a taste cost, *d*, that decreases the utility derived from a product if the product's location on the taste line does not correspond to the location of the customer. I will assume that this disutility is quadratic in distance. The taste dimension of the model is thus a standard Hotelling model as in eg. Bonanno (1987), apart from the possibly different density of customers in city *A*. Further, it is assumed that the distance between the cities is *t*, reflecting the transport cost of travelling from one city to the other and back. This cost does not depend on a customer's location on the taste line, but is the same for all customers in the model. Modelling the transport (and taste) costs in this way corresponds to the intuition of them being independent of the

amount of goods bought (although I will assume that each consumer buys at most one good). To guarantee well behaving demand curves, I will assume that the taste cost of tastes is always greater than the transport cost, $t \leq d$. This assumption means that as long as the price differences are not too big (and the firms do not choose same location on the taste line), there are some customers that travel to the other city to get their most preferred product, and that a marginal change in price never results in a discrete jump in demand.

Call the two firms 1 and 2. To study the location decisions, I will assume that the firms make their location choices sequentially, and that it is firm 1 that acts first. This gives the following structure:

1. Firm 1 chooses its location on the taste line, and the geographical location of its first branch
2. Firm 2 ----- " -----
3. Firm 1 decides whether to open a branch in the city where it does not have a branch
4. Firm 2 ----- " -----
5. Firms compete in prices

There is a fixed cost that has to be incurred when a branch is opened. For simplicity, I will assume that this cost, denoted K , is constant (it could be made to vary with the number of branches, so that it would be either

lower or higher for the second branch than for the first one). I will also make the following assumptions:

- it is always profitable for both firms to establish at least one branch (they enter if expected profits are nonnegative)
- in case of ties, the firm prefers to locate in A
- the reservation utility, homogenous over all customers, is low enough to guarantee full market coverage even for a monopoly with one branch
- a firm has to set a uniform price (no intra-firm price discrimination)
- the production process is homogenous over firms at branch level (ie. product differentiation in tastes does not affect the production costs)
- marginal costs are zero

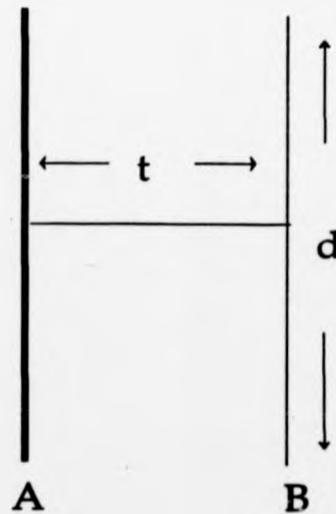


figure 1

The first of these is made largely for convenience, since it rids us of the monopoly cases, that are not of interest here. This is not to deny that they are interesting, and the model as such does allow for them. The second can be defended as follows: if there is some uncertainty of the characteristics of the rival (say, marginal costs), then the firm places itself on the safe side by tapping the larger clientele. This motivation will become clear later, when I discuss the location choices of firms, when one of them has two branches. The last one does not affect the results as long as we stick to the perfect information framework that I use here. Cremer and Thisse (1991) have shown that horizontal product differentiation, which I will concentrate on here, is a special case of vertical product differentiation.

If the transport cost t decreases, the model approaches the standard one-city Hotelling model, with a different customer density. If the taste disutility parameter d decreases, it means that t decreases, too (because of the assumption $t \leq d$), and the model approaches the standard Bertrand model of competition in prices.

The model takes the form of an H , where the vertical taste dimension is longer than the horizontal geographical dimension (see fig. 1). The customers are thus located on the vertical lines.

There are several possible combinations of geographical location. The firms could have just one branch each, and locate them in different cities (2

possibilities: 1 at *A* and 2 at *B* or vice versa), or in the same city (again two possibilities). One of the firms (with the assumption made, firm 1) could have two branches, whereas the rival has just one. The rival's branch can be in either city, giving again two possibilities. Lastly, both firms could opt for two branches. There are thus seven locational outcomes in theory.

Consider now a customer in city *A*, who is located at *x*. Her cost of acquiring firm 1's (when firm 1's location is z_1) product is either

$$(1) \quad p_1 + d(x-z_1)^2$$

or

$$(2) \quad p_1 + d(x-z_1)^2 + t$$

and similarly for firm 2, the only changes being the subscript of *p*, the price, and a change of distance measure to $((1-z_2)-x)$. It is easy to solve for *x*, thus getting the demands of the two firms in city *A*. The first proposition concerns the location of firms on the taste line.

proposition 1: the firms locate at the ends of the taste line.

A proof of this proposition can be found eg. in Tirole (1988). Basically, the derivatives of profits with respect to location are negative for firm 1 and

positive for firm 2, and they thus want to locate as far as possible from each other. This is a standard result of the Hotelling model with quadratic transportation costs, and the addition of another city to the model does not affect this result. The way distance is measured in the current model, firm 1 locates its product at point 0 and firm 2 at point 1.

If it is not profitable to have more than one branch per firm, the only thing that the firms have to decide is where to locate their branches. The following proposition states the result:

proposition 2: if it is not profitable for either firm to open more than one branch, then the firms will choose different locations, and firm 1 will open its branch in the bigger city A.

The proof of all the propositions are relegated to the appendix. What has to be shown is that, firstly, the profits are higher if the only branch is in A, and that given that firm 1 has already a branch in A, the profits of firm 2 are higher if it opens its only branch at B. There is thus only one possible equilibrium out of four possible ones, when the firms' optimal number of branches is 1.

It is possible to prove the following proposition regarding the one branch per firm equilibrium:

proposition 3: If the optimal number of branches per firm is 1, then firm 1 (located in A) sets a higher price and has a higher demand.

Intuitively, by positioning its branch in city A firm 1 captures the clientele there. This clientele is bigger than that in city B. The optimum for firm 1 turns out to be to exercise its market power through a higher price than its rival. This means that it surrenders some of its captive customers to firm 2, but squeezes a bigger slice of surplus out of the remaining ones. For this to be an equilibrium, it must not be profitable for either firm to open a new branch.

If the situation is changed so that (at least) one firm finds it profitable to open another branch despite an increase in fixed costs, the location decisions of the other firm is affected:

proposition 4: If it is profitable for just one firm (firm 1) to open two branches, then its rival will open its branch in city A. Firm 1 will set a higher price, and have a higher demand than its rival.

Firm 2's optimal location thus depend on whether or not it is profitable for firm 1 to open two branches. With the assumption of perfect information used here, this does not matter for firm 1, but if there were some uncertainty on, say the sizes of clientele in the two cities, then firm 2's choice of location would act as a signal to firm 1. To get this asymmetric

equilibrium in branches, it must be profitable for firm 1 to open a second branch, and unprofitable for firm 2 to counter that move by opening its second branch.

It is possible that both firms open two branches, but this is a prisoner's dilemma type of outcome: both firms would then prefer both of them having just one branch.

proposition 5: If firm 1's profits are higher with two branches than with one given that firm 2 has just one branch (and then the same applies to firm 2), and if firm 2's profits are higher with two branches than with one given the number of firm 1's branches, then the equilibrium number of branches is two for both firms. The profits are, however, lower compared to a situation where both firms would have just one branch.

Let's take the viewpoint of firm 1 on stage 3 of the game. The firm has to decide whether or not to open a second branch. Given the contents of proposition 5, the firm knows that if it does not open a second branch, firm 2 will. Since the firm's profits would be lower if it has just one branch and its rival two, compared to both having two, it decides to open a second branch. Then change to firm 2 in stage 4: since firm 1 already has two branches, and firm 2's profits are lower with one than with two branches, it opens a second branch. But since I have assumed that the market is covered, this means that there is duplication of fixed costs

without any increase in either prices or demand. Thus the firms lose the fixed costs of the second branch, compared to their first best. And actually, they lose even more, since their first best would also entail different geographical location according to proposition 1, and thus higher profits.

In the adopted duopoly setting, a dynamic game would allow for cooperation where both firms would open just one branch. This kind of cooperation, in turn, could be undermined if there are potential entrants to the market. Since the main aim of this section is to show the theoretical possibility of oligopolistic firms not minimizing costs, I do not study these questions in more depth.

The last proposition is probably the most important one, especially with regard to the empirical section:

proposition 6: An oligopolistic firm that maximizes profits does not necessarily minimize costs.

This is clear from the cases where either firm has more than one branch: the whole market is covered with just one branch per firm (with the assumptions made, even with a monopoly with one branch), but despite this firms can find it profit maximizing to open a second branch, thereby duplicating fixed costs and increasing average costs (since for neither firm does the demand more than double, which is a necessary condition for

average costs to decrease after the opening of the second branch).

To sum up the results that are relevant for the empirical part of this chapter: it is a possibility that an oligopolistic firm does not minimize costs, but is engaged in branch proliferation. Also, this means that firms forego potential economies of scale at branch level in order to gain market power. These equilibria thus display diseconomies of scale at firm level and economies of scale at branch level, as suggested in the introduction. Furthermore, a bigger firm (whether with one or two branches) sets higher prices to a good that has the same production costs. This means that turnover (or, in the case of insurance, premium income) is a biased measure of output, whether at product or firm level. A correct measure of output is the number of units sold, since this reflects truthfully the production volume, and hence the costs, of a firm.

IV.3 The data, the market and the empirical model

The data used in this study is from the Finnish non-life insurance market. This market provides an interesting test-bed for several reasons. All the firms rely on branches for distribution, and there are (almost) no brokers during the observation period 1989-1991. Thus it can be assumed that all firms rely on the same (or at least, closely related) technology. There are big differences in the sizes of the branch networks, the largest ones comprising of over a hundred branches, whereas the majority of firms rely

on just one branch. The number of products, as listed in Statistics Finland: Insurance¹, is large: 39 lines can be found for which both quantity and price (=premium income) data is available. The actual number of lines sold is even bigger, since for some categories, only the aggregate number of policies sold, or the aggregate premium income was available. The large number of products (=lines of insurance) together with the relatively small number of firms (21) makes it necessary to constrain the cost function in order to obtain any estimates. The large geographical size (over 300 000 sq. km.) relative to population (5 Mio.) suggests that in Finland, a branch network is an effective means to increase market power. This should mean, then, that the interplay between the captivation effect and the cost effect should figure prominently in the optimization problem that the firms face. This is indeed found to be the case, as will become clear when the empirical results are discussed.

The different lines of insurance are listed in table 1, together with the number of firms providing each line of insurance in any of the years of observation, 1989-1991, and the aggregate (1991) premium income of each line together with the four-firm concentration (1991 data) ratio per line².

¹ All other data but the number of branches is from Statistics Finland: Insurance. The number of branches per firm was kindly collected by the research dept. of the Association of Finnish Insurers, whose help is thankfully recognized.

² CR4 is calculated as the percentage of the number of policies sold that the four biggest firms in each line of insurance produce.

LINES OF INSURANCE IN THE FINNISH NON-LIFE INSURANCE MARKET (1991 DATA)

General category/subcategory	no. of firms	premium income (1000 FIM)	% of total	CR4
Statutory accident ins.				
general tariffs*	13	876 623	7,8	83,9
special tariffs*	13	1 156 707	10,3	88,1
other accidents	13	60 958	0,5	83,1
Other accident ins.				
continuous individual	14	338 930	3,0	68,4
cont. group ins.	13	108 581	0,9	91,2
traveller's ins. etc.	15	186 424	1,7	92,9
other ins.	12	13 387	0,0	87,6
Compulsory motor third party liability ins.				
	13	2 481 157	22,1	73,1
Motor vehicle ins.	16	1 676 160	14,9	80,0
Hull ins.				
ship hull, civil	9	141 645	1,3	89,8
ship hull, war	5	6 995	0,1	99,6
protection and indemnity liability	1	1 521	0,0	100,0
yacht ins.	14	71 931	0,6	75,5
aircraft hull ins.	8	9 177	0,1	97,9
aviation liability	6	4 034	0,0	99,5
Cargo ins.**	18	339 271	3,0	84,4
Fire and other				
combined property ins.				
households' fire ins.	13	57 071	0,1	63,3
households' compr.				
house etc. ins.	15	1 155 460	10,3	76,2
farm ins.	13	174 619	1,6	84,5
other ins.	4	165	0,0	100,0
industrial fire	15	287 783	2,6	86,8
trading fire	12	26 636	0,0	91,9
other fire ins.	11	69 689	0,1	93,4
real estate ins.	17	421 673	3,8	81,5
combined ins.				
(industrial)	11	188 278	1,7	70,8
combined ins. (trade)	12	111 411	1,0	93,3
combined ins. (other)	10	107 281	1,0	89,1
burglary and robbery ins.				
water damage ins.	11	8 989	0,0	95,1
glass and shield ins.	10	3 830	0,0	88,5
machinery				
breakdown ins.	9	64 175	0,6	82,2
other ins.	5	16 562	0,1	100,0
Loss of profit ins.	7	175 832	1,6	96,0
Forest ins.	13	32 360	0,3	85,8
Third party ins.	18	327 049	2,9	84,1
Credit ins.	18	317 577	2,8	71,0
Other ins.	19	160 231	1,4	92,9
Finnish reins.***	32	932 880		
Foreign ins.***	13	1 611 887		

no. of firms = no. of firms which actually sold policies in 1991, as opposed to being listed as (potentially) providing them

% of total = % of total direct domestic premium income. Thus not provided for the two last categories
CR4 is based on the output measure used, usually the no. of contracts

* measure of output working time in Mio. working hours, ** measure of output no. of accidents,

*** measure of output premium income. These lines can be viewed as part of the international market, where market power of Finnish firms is bound to be low

Table 1

Two biggest lines, the worker's compensation and traffic insurance, are regulated and compulsory. There is both price regulation (which has been gradually relaxed during the observation period, but not abolished) and entry regulation. A firm needs a licence to provide either of these lines of insurance, and not all firms have such a licence. It is not clear how many have opted not to apply for one voluntarily, and how many have not applied after judging that the likelihood of getting a licence is too small. The regulation of these lines can, for its part, strengthen the captivation effect in that firms engage in quality competition when price competition is banned or restricted. The total number of firms in the market is bigger than that in the sample. There are two main reasons for this: some firms, primarily industry captives that are really not active according to industry sources, have been excluded and the subsidiaries of firms have been merged into their parents. One firm was excluded because it was active

KEY INFORMATION ON FIRMS IN THE FINNISH NON-LIFE INSURANCE MARKET	
no. of firms in the market	21
highest number of branches	107
lowest number of branches	1
mean number of branches	32
no. of firms with one branch	10
average premium income (direct ins.)	509 000 000 FIM

table 2

only in one period and because its parent, a life insurance firm, went into receivership shortly after the observation period. Its market share was negligible, however. Key information regarding the firms is presented in table 2. Some of the firms are stock-owned, some mutuals. I assume that both organizational forms use the most efficient technology available.

The firms can be divided into two groups. The six biggest firms behave differently compared to the rest (see chapter II). All but one of these have a nation-wide branch network, the outlier being a firm specialising in big manufacturing customers. The other group consists of firms with heterogenous strategies: most of them have only one branch, and only one of them a national branch network. Some of them have a larger regional network, though.

The standard tool of cost function analysis currently is the translog cost function (Christensen, Jorgenson&Lau 1973). This, however, has the unfortunate feature that it does not allow for zero production of any one product. In the current sample, most firms do not produce all products, and thus a translog form cost function cannot be used. In one of the most comprehensive treatments of cost functions to date, Baumol, Panzar&Willig (1982, pp. 448-450) list five key requirements for a good cost function:

- firstly, it should be a proper cost function, one consistent with minimization of inputs. It should be nonnegative and nondecreasing, and linearly homogenous in input prices.

- secondly, it should be able to accommodate zero production of some products
- thirdly, it should not impose constraints on the data, but should be flexible
- fourthly, it should not require the estimation of an excessive number of parameters.
- fifthly, it should not impose any constraints on the values of the first and second partial derivatives

On these grounds, they end up recommending a quadratic cost function:

$$(3) \quad C = \alpha + \beta_i x_i + \sum \gamma_{ii} x_i^2 + \sum \sum_{i \neq j} \gamma_{ij} x_i x_j$$

where

x_i = output of product i

The quadratic cost function has two drawbacks: Firstly, it cannot be deduced theoretically. This is in my view a minor problem, since any cost function should be viewed as a statistical (or theoretical) approximation to the true, underlying cost function. Secondly, it does not allow an easy incorporation of inputs. With the particular data set used, this should not be a big problem either. Labour accounts for a major share of operating costs, and wages are centrally negotiated at the industry level. They can thus be assumed to be constant over firms.

The standard way of treating a cost function has been to estimate eq. (3) directly (for an insurance example, see Daly, Rao & Geehan 1985; for banking, eg. Lawrence 1989). Some studies have included a branch variable to "control for the number of branches" (eg. Murray & White 1983). The addition has usually been done simply by adding a linear branch-variable into equation (3). Other studies, eg. Kolari & Zardkoochi (1990) and Benston, Hanweck & Humphrey (1982), have treated the branch variable like any other variable without explicitly discussing whether branches are part of the out- or input. Thus the functional form becomes (for a translog cost function):

$$(4) \quad C = \alpha + \beta_i \ln x_i + \psi k + \frac{1}{2} \sum_{ii} \gamma_{ii} \ln x_i^2 + \frac{1}{2} \rho \ln k^2 + \frac{1}{2} \sum_{i \neq j} \gamma_{ij} \ln x_i \ln x_j + \frac{1}{2} \sum_{ij} \nu_{ij} \ln x_i \ln k + \frac{1}{2} \delta \sum_{pp} \rho_{pp} \ln x_i \ln x_j + \frac{1}{2} \nu \sum_p \rho_p \ln x_i \ln k \quad i \neq j$$

where

k = the no. of branches

The biggest problem empirically in this study is the large number of products. Although Baumol et. al. claim that the quadratic cost function does not "require the estimation of an excessive number of parameters", this is no longer true when the number of products is 39. A fully specified cost function would have 819 parameters. With the current data set, this would necessitate roughly 40 years of data to cover the parameters, and many more to produce reliable estimates. It is quite clear that there have

been major changes in the production technology over the latest 40 year period, most notably the introduction of computers, and thus even if such data were available, it would not solve the problem. The cost function has to be constrained so that a more limited set of data is adequate for estimation. This is achieved in two steps:

1) instead of using the output measure suggested by the theoretical model, I use a modified one. To be able to sum up different products, they are translated into a common measure, money. To avoid the possible market power effects that would result in different prices for differently sized firms, I use as price the per policy premium income of the biggest firm. Firm i 's premium income for any given line is obtained by multiplying the number of policies i sold, x_i by this "market-power free" price, p_i to get the measure of output, $p_i x_i$. This is somewhat similar to a Laspeyres quantity index.

2) I assume that the coefficients of the quadratic terms, β_i , and the coefficients of the cross-product terms, γ_i , are constant over products. To be able to achieve branch-level estimates, the branch-variables are separated.

The result of these procedures is equation (5):

$$(5) \quad C = \alpha + \beta \sum p_i x_i + \psi k + \gamma \sum (p_i x_i)^2 + \rho k^2 + \delta \sum \sum p_i p_j x_i x_j + \nu \sum p_i x_i k \quad i \neq j$$

The number of parameters has been thereby cut from 819 to 6. The difference between the approach adopted here and that of former studies that have included a branch variable is twofold: here, the branches are treated as part of the output, and, it is recognized that firms can expand production through adding branches even when that is more costly than increasing production at the existing branches would be. To quote Kolari&Zardkoochi (p. 441):"Bank managers may consider the existing bank facilities efficient, and, therefore, add branches to increase total bank output". The reasons for increasing the number of branches are thus very different.

The functional form adopted (5) nests the most commonly used alternative functional forms, namely the "pure" cost function with no branch variable (eq. (5a), which I subsequently call the traditional model I), and the cost function with a linear branch term (eq. (5b), traditional model II).

$$(5a) \quad C = \alpha + \beta \sum p_i x_i + \gamma \sum (p_i x_i)^2 + \delta \sum \sum p_i p_j x_i x_j \quad i \neq j$$

$$(5b) \quad C = \alpha + \beta \sum p_i x_i + \psi k + \gamma \sum (p_i x_i)^2 + \delta \sum \sum p_i p_j x_i x_j \quad i \neq j$$

The definitions of Baumol et. al. (pp. 50 and pp. 73) are used when estimates for economies of scale and scope are calculated. For economies of scale, two measures will be produced; for the firm-level and for the

branch-level. The former is achieved by letting all output variables, including the branch-variable, change, and the latter by holding the branch-variable constant. The overall economies of scale are defined as

$$(6) \quad S_N = C(x) / \sum y_i C_i$$

where

$C(x)$ = total cost of producing y products

x_i = the amount of product i produced

$C_i = \delta C / \delta x_i$, the partial derivative of C with respect to x_i

There are increasing, constant and decreasing economies of scale as S_N is bigger, equal or smaller than unity. The branch-level measure is obtained from (5) and (6) by holding all the branch variables constant. The measure of economies of scope is

$$(7) \quad S_E = [C(x_{N-i}) + C(x_i) - C(x_N)] / C(x_N)$$

The costs of producing the sets of products ($N-i$) and (i) separately are summed together, and the cost of producing the whole set (N) is subtracted from this. This is then divided by the cost of producing the whole set (N) of products. There are economies of scope if S_E is positive, and diseconomies of scope if S_E is negative.

In the insurance literature on cost functions, no agreement has been reached as to what measure of output to use. Skogh (1982) shows that premium income understates economies of scale, and suggests claims expenditure as a measure of output. Cho (1988) criticizes the use of claims as a measure of output, and suggests premium income. Subsequently, both have been used. The theoretical model of section 2 shows that in an oligopolistic environment, prices, hence premium income, vary systematically with firm size, and probably produce an upward bias. The criticism of section 2's model extends beyond insurance, however. In my opinion, there is a natural candidate for measuring output, namely the number of units (of a given product) produced. This requires the assumption that products in a given category are homogenous over firms, or that at least their production costs are homogenous. But these assumptions are already implicit in any cost function estimation, and thus this measure does not place any new restrictions on the empirical model.

IV.4 Empirical results and comparisons to earlier studies

The quadratic cost function (5) was estimated using the three-year (1989-1991) data set available. A one-way random effects specification was adopted³, as there was data available for the number of branches only for

³ The smallness of the data prevented a frontier analysis (see Bauer 1990 for a general discussion, Ferrier&Lovell 1990 and Berger&Humphrey 1991 for banking studies and Fecher, Perelman&Pestieau 1991 for an insurance study)

the year 1989. A one-way random effects model is of the general form

$$(8) \quad C = bX + e(i,t) + u(i)$$

(see Hsiao 1986) where C and X are matrices, there is an additional observation unit (in this study, firm) specific error term, $u(i)$. As there is no intra-firm variation in the branch variable, it would be collinear with the firm-dummies of a fixed-effects specification. The lack of branch data for 1990 and 1991 should not pose any problems with regard to the results, since changes in the number of branches after 1980 and before 1992 has been almost nonexistent.

In addition to the equation (5), its two constrained versions were estimated, too⁴: the traditional cost function without any branch variables (eq. (5a)), and the cost function with only a linear branch variable (eq. (5b)). Descriptive statistics of the estimating variables are presented in table 3, and the results of the estimations in table 4. There are three measures for economies of scale: one at the firm level and two at the branch level. The branch level measures differ in the number of branches: the first one is calculated at the level of just one branch, and the second at the level of the mean number of branches. The measures for economies of scope are calculated at two levels, similarly to the branch level economies of scale

⁴ All estimations were done using LIMDEP 6.0 (Greene 1991).

measures. Results for all three functional forms are presented in table 5⁵. It is interesting to compare the results from the different specifications, but before going into that it should be noted that an F-test (see table 4) clearly rejects both

DESCRIPTIVE STATISTICS OF ESTIMATION VARIABLES		
Variable	Mean	Standard error
operating costs	108 560,4	167 358,3
X	601 110,0	980 610,0
XX	166 720 000 000,0	363 170 000 000,0
XZ	370 820 000 000,0	877 350 000 000,0
K	22,6	31,6
KK	1492,6	2913,8
XK	103 950 000 000 000,0	481 690 000 000 000,0

operating costs = salaries + other social expenses + other operating expenses in 1991 money
 $X = \sum p_i x_i \quad i = 1, \dots, 39$
 $XX = \sum (p_i x_i)^2$
 $XZ = \sum p_i p_j x_i x_j \quad i \neq j$
 $K = \text{no. of branches in 1989}$
 $KK = K^2$
 $XK = \sum p_i x_i k$

table 3

⁵ In the calculations for the measures of economies of scale and scope, those coefficients that were insignificantly different from zero are assumed to be zero because of the low probability values of all these coefficients.

ESTIMATION RESULTS FOR THE QUADRATIC COST FUNCTION			
variable	pref. specification	traditional spec. I	traditional spec. II
const.	6199,2 (6935)	-10404 (8683)	-13407 (8337)
X	0,16685*** (0,03233)	0,36059*** (0,02389)	0,22407 (0,041)
XX	0,33E-07 (0,77E-07)	-0,40E-06*** (0,75E-07)	-0,22E-06*** (0,81E-07)
XZ	-0,15E-06*** (0,26E-07)	-0,53E-07*** (0,29E-07)	-0,51E-07** (0,27E-07)
K	-485,08 (620,5)		2423,2*** (606,0)
KK	49,327*** (8,07)		
XK	0,15E-11 (0,82E-11)		
R ²	0,977	0,950	0,956
F-test (d.f)		21,42*** (3,56)	48,97*** (1,56)
est. autocorr. coeff.	-0,196	-0,055	-0,038
F-test = F-test against the preferred specification; degrees of freedom in parentheses * = significant at the 10% level ** = significant at the 5% level *** = significant at the 1% level numbers in parentheses are standard errors			

table 4

traditional specifications. The proposed specification exhibits diseconomies of scale at firm level, and at the same time, a linear functional form (a positive fixed cost and constant marginal costs, see table 4), thus economies of scale at the branch level. The branch-level economies of scale are the greater, the more branches a firm has, since the fixed cost of the latest

added branch are increasing in the number of branches. This can be seen in table 5 when comparing the two branch level economies of scale measures of the preferred model. Thus, the bigger is a firm's branch network, the stronger is the cost effect, and (probably) the smaller is the captivation effect. Smaller firms forego smaller economies of scale if they decide to open yet another branch, but they, too, experience average costs that are higher with an added branch than they would by expanding output at the existing branches. There are overall economies of scope, which are produced by the negative coefficient of the crossproduct term XZ. Economies of scope seem to be stronger with a smaller number of branches. This effect is also due to the rise in fixed costs, when the branch network is expanded. When

ESTIMATES OF ECONOMIES OF SCALE AND SCOPE			
measure	pref. model	trad. model I	trad. model II
S_N firm level	0,84	2,95	1,72
$S_N (K=1)$	21,40	-	3,59
$S_N (K=22)$	31,32	-	5,84
$S_E (K=1)$	1,19	0,10	0,18
$S_E (K=22)$	1,12	-	0,43

S_N = economies of scale
 S_E = economies of scope
 $K = 1$: measured at level of firm having one branch
 $K = 22$: measured at level of firm having the mean no. of branches (=22)

table 5

comparing these results to those achieved with the traditional functional forms, it is easy to see the difference. The traditional specification with no branch variable exhibits large economies of scale, and small economies of scope compared with the preferred specification. It seems that the branch level economies of scale of the preferred model have been subsumed into the negative coefficient of the squared term XX , producing these results. Even the functional form with a linear branch term exhibits economies of scale at the firm level, as well as at the branch level. The scope estimate is closer to the traditional model than to the preferred model. With this specification, as with the preferred one, branch level economies of scale increase with the number of branches.

The theoretically and statistically preferred functional form, where branches are treated as part of the output, gives results that, as suggested in the introduction, are possible due to the captivation and the cost effect, if fixed costs dominate variable ones. There are decreasing returns to scale in the Finnish non-life market, but this is not due to the production technology, but to the market characteristics that the firms face. Large geographical area, and a widely scattered, but locally dense population, and the oligopolistic, partly regulated structure of the industry underline the effects of the branch network in increasing the market power of the firms. The clear difference in results, when compared to the two more traditional specifications, should also be noted. These both would suggest that the industry is a natural monopoly, indeed one very near to a

contestable market (in the case of the traditional model I, the results suggest a contestable market, and in the case of traditional model II, nearly so because the fixed costs of one branch, and more would be cost-inefficient, are close to zero) whereas the results of this study would point to the direction where a less concentrated structure would be more economical in cost terms.

The results of this study can be compared to those of Suret (1991) and Fecher et. al. (1992), which are the only non-life insurance studies using a flexible functional form (translog in both cases)⁶. Both studies report economies of scale. Suret, however, using Canadian data, only for a limited range of output. Suret reports no economies of scope, whereas Fecher et. al. use an aggregated output measure. Both studies use both premium income and claims as a measure of output. The results produced with premium income are biased, according to the theoretical model of section 2. In addition, Suret's study can have sample selection bias, as a few firms that do not produce all four lines of insurance are left out of the study. As the choice of product variety is the result of optimization, then there must be a rational reason for these firms not to produce all lines of insurance.

On the basis of these results, both traditional models are rejected. They also produced very different qualitative results compared to the preferred

⁶ For a summary of earlier insurance studies using other functional forms, see Suret.

model. It is possible that in a bigger market, say the U.S. banking market that has been the subject of several studies, the oligopolistic effects on cost functions are substantially weaker. This does not mean that they should not have been allowed for. The results of this study suggest that the traditional forms of cost function are likely to be misspecified, when used on data from a (oligopolistic) services industry.

The results of the econometric investigation allow for the study of the cost effects of the pursuit for market power through branch proliferation. As there are economies of scale at the branch level, the most cost effective way (abstracting from the transport costs of customers) would be to have just one branch per firm. Taking the geographic and populational facts into account, this would clearly be unfeasible. But if we assume that those firms (the five of the "big six") with an extensive branch network could let do with

- a) the mean number of branches (rounded to the nearest integer): 22 or
- b) the number of branches that the smallest of them has: 50 or
- c) the number of branches that the firm of the other group with the most extensive branch network has, namely 34,

then the extra costs can be calculated. The results of these calculations, that are intended more to give a feel for the level of extra costs rather than being exact estimates are presented in table 6, where both the market total, and the firm-wise extra costs are reported, as well as the percentage that

ESTIMATES OF THE COST OF BRANCH PROLIFERATION						
Firm ranking (size)	no. of branches	total operating costs (1991, 1000 FIM)	$\psi K + \rho KK$ with existing no. branches (1 000 FIM)	difference to $\psi K + \rho KK$ with mean no. of branches (22)	difference to $\psi K + \rho KK$ with min. of big firms' branches (50)	difference to $\psi K + \rho KK$ with max. of small firms' branches (34)
1st	107	705 034	564 745	540 871 (77%)	441 427 (63%)	507 723 (72%)
2nd	83	591 519	339 814	315 939 (53%)	216 496 (37%)	282 792 (48%)
3rd	67	302 548	221 429	197 555 (65%)	98 111 (32%)	164 407 (54%)
4th	63	268 733	195 779	171 905 (64%)	72 461 (27%)	138 757 (52%)
5th	50	360 184	123 318	99 443 (28%)	0 (0%)	66 295 (18%)
Σ				1 325 712 (60%)	828 496 (37%)	1 159 974 (52%)
numbers in parentheses in columns 3-6 are per cents of total operating expenses						

table 6

these extra costs represent from the total operating costs⁷ of the firms.

These estimates underline the importance of the captivation effect: the big firms seem to forego substantial economies of scale in order to gain market

⁷ The total costs used in table 6 are the true total costs of year 1991. The calculations were performed using estimated total costs, but the differences were minor: the average difference between two estimates was 0.1 percentage points.

power. Even the lowest average estimate indicates that 37% of the operating expenses are due to the captivation effect. Depending on which number of branches is taken as the yardstick, firm-wise estimates vary between 18 and 77%. Despite the crude character of these estimates, in my opinion it can be concluded that a nonnegligible part of the operating expenses of Finnish non-life insurance firms are due to the pursuit of market power.

IV.5 Conclusions

This study rests on the observation that the relevant unit of production in services production is the branch. In most industries, establishing a branch means that at least some fixed costs have to be incurred, and, as a result, there are economies of scale over at least some range of output at the branch level. Depending on the characteristics of the production technology, these might even span over the entire output range. Despite this we can observe fairly dense branch networks in several services. In this chapter, I argue that there are two off-setting forces that determine the size of the branch network. The captivation effect is the increase in market power that the firm acquires by expanding its branch network. The cost effect refers to the increase in (average) costs that this expansion results in. For a firm with a small branch network, the cost effect can be negative and the captivation effect is probably at its strongest, and thus that type of firm is likely to increase production through new branches. For big firms, the

situation is reversed, and thus they are more likely to expand production by increasing it at the existing branches. For a multiproduct firm, branches can be viewed as one of the products, as customers prefer a firm for a combination of the products that it offers, and the location where these are offered. A branch network can create an externality, too, in that people prefer the otherwise similar product of a firm that has a large branch network to support the product.

The market power that is achieved by a big branch network is often best exploited by setting prices at a different level than smaller rivals. This means that in any service industry, including retailing, turnover (or, for insurance, premium income) is a biased measure of output, and the correct measure is the number of products produced. The prominence of the branch network in the tool box of the firms means that it should not be left out of cost function estimations, but should be included in the same way as other products.

Theory suggests to us that an industry can simultaneously have increasing returns to scale at branch level and decreasing returns to scale at firm level. If the actual production technology exhibits economies of scale, this should be observed at the branch level. It might then be, however, that the market characteristics are such that the profit-maximizing strategy is to increase the number of branches, even at the cost of foregoing potential economies of scale that the production technology exhibits.

The above theory was used in formulating a cost function, used to estimate the economies of scale and scope in the Finnish non-life insurance industry. The insurance industry has been comparatively neglected, and there are only a few published papers on non-life insurance that use modern econometric techniques. A technique was developed that allowed the estimation of a cost function on a limited set of data.

The traditional functional forms were rejected. More importantly, they produced significantly different results. Whereas the preferred functional form confirmed the theoretical possibility of diseconomies of scale at firm level when there at the same time are economies of scale at the branch level, and economies of scope, the traditional functional forms gave results according to which there would be significant economies of scale both at firm-and branch level. The economies of scope-results were similar for all specifications, although the preferred model produced the biggest values of the scope measure. The importance of these differences can be best understood in the light of the kind of guidance they would give to regulators. If one of the traditional functional forms were used, there would be little reason from the efficiency point of view for the regulator to grant new entry licences. The preferred results, however, show that small firms are more cost-efficient, and thus these results would give ground to opposite policy advice.

On basis of the econometric results of the cost function estimation, the cost

effects of branch proliferation were calculated for those firms with a nationwide branch network. These proved to be of a very significant order, the lowest average measure that was obtained being 34%, when I assumed that the minimum number of branches needed is 50, corresponding to the size of the smallest nation-wide branch network of the "big six" firms.

Earlier studies on cost functions have, implicitly or explicitly, maintained the neoclassical assumption that the cost function is the dual of the production function. Even if a branch variable has been introduced in a similar manner to other variables, the assumption has been made that new branches are added only if economies of scale at the existing branches have been exhausted. The results of this study suggest that the more traditional functional forms are potentially misspecified, when applied to a service industry, and that branch proliferation might - and does - occur for other than efficiency reasons. Industrial economists have long known that an oligopoly does not minimize cost. This study shows, both theoretically and empirically, that oligopolistic firms do not necessarily do that either.

APPENDIX

Before going to the proofs of the propositions, let me state the profits and prices of firms 1 and 2 in the different possible geographical outcomes. The following notation is used:

α = density of customers in city A, $\alpha \geq 1$

p_i = price of firm i , $i = 1, 2$

d = taste cost parameter

t = transport cost between the cities

x = location in city A

y = location in city B

D_i = total demand of firm i

Π_i = profit of firm i

K = fixed cost of establishing a branch, $K > 0$

subscripts denote firms, superscripts the different locational outcomes

The following assumptions are made:

1. firm 1 is located at z_1 , firm 2 at $1-z_2$, $z_1 \leq 1-z_2$
2. $0 < t \leq d$
3. $\alpha \geq 1$
4. If the firms are indifferent with respect to geographical location, they locate in city A
5. In case of a tie, a firm chooses the smaller no. of branches
6. It is always profitable for both firms to establish at least one branch (they enter if expected profits are nonnegative)

7. in case of ties, the firm prefers to locate in *A*
8. The reservation utility, homogenous over all customers, is low enough to guarantee full market coverage even for a monopoly with one branch
9. A firm has to set a uniform price (no intra-firm price discrimination)
10. The production process is homogenous over firms at branch level (ie. product differentiation in tastes does not affect the production costs)
11. marginal costs are zero
12. length of the taste line is 1
13. travel costs between the cities are linear, and quadratic on the taste line
14. there are only two firms
15. the game has four stages, as explained in section IV.2
16. the fixed cost is positive

The different outcomes are labelled as follows:

- a) firm 1 in *A*, firm 2 in *B*
- b) both firms in the same city, each having one branch (by ass. 4 in city *A*)
- c) firm 1 in both cities, firm 2 in city *B*
- d) firm 1 in both cities, firm 2 in city *A*
- e) both firms in both cities

Since proposition 1 tells us that the firms always locate their branches in the ends of the taste line, the demand in city *A* in cases a) and c) is derived from

$$(1) \quad p_1 + dx^2 = p_2 + d(1-x)^2 + t$$

giving

$$(2) \quad x = (p_2 - p_1 + d + t)/2d$$

Demand in *B* is given by a similar equation to (1) in case a), by changing the subscripts and the location from *x* to *y*, giving

$$(3) \quad y = (p_1 - p_2 + d - t)/2d$$

In case c), the demand for firms is the same as in the standard Hotelling model with quadratic transport costs, namely

$$(4) \quad y = (p_1 - p_2 + d)/2d$$

The demands in case d) are derived similarly. In case b), the demands in the city where the firms are located is the normal Hotelling demand, similar to (4), and the demand from the city with no branches is for firm 1 is again similar to (4), since transport costs are the same to either firm and do thus not affect their respective demands (and, remember, it is assumed that the market is covered).

From the above, the demands, prices and profits can be derived. They are,

for each case, as follows:

- a) (1a) $D_1^a = (\alpha + 1)/2 + t(\alpha - 1)/6$
 (2a) $D_2^a = (\alpha + 1)/2 + t(\alpha - 1)/6$
 (3a) $p_1^a = d + t(\alpha - 1)/3(1 + \alpha)$
 (4a) $p_2^a = d - t(\alpha - 1)/3(1 + \alpha)$ $p_2 > 0$ by ass. 2 and 3
 (5a) $\Pi_1^a = d(\alpha + 1)/2 + t(\alpha - 1)/6 + t^2(\alpha - 1)^2/[18d(1 + \alpha)] - K$
 (6a) $\Pi_2^a = d(\alpha + 1)/2 - t(\alpha - 1)/6 + t^2(\alpha - 1)^2/[18d(1 + \alpha)] - K$
- b) (1b) $D_i^b = (\alpha + 1)/2$ $i = 1, 2$
 (2b) $p_i^b = d/2$ $i = 1, 2$
 (3b) $\Pi_i^b = d(\alpha + 1)/4 - K$ $i = 1, 2$
- c) (1c) $D_1^c = (\alpha + 1)/2 + \alpha t/6d$
 (2c) $D_2^c = (\alpha + 1)/2 - \alpha t/6d$
 (3c) $p_1^c = d + \alpha t/3(\alpha + 1)$
 (4c) $p_2^c = d - \alpha t/3(\alpha + 1)$
 (5c) $\Pi_1^c = d(\alpha + 1)/2 + \alpha t/3 + (\alpha t)^2/[18d(\alpha + 1)] - 2K$
 (6c) $\Pi_2^c = d(\alpha + 1)/2 - \alpha t/3 + (\alpha t)^2/[18d(\alpha + 1)] - K$
- d) (1d) $D_1^d = (\alpha + 1)/2 + t/6d$
 (2d) $D_2^d = (\alpha + 1)/2 - t/6d$
 (3d) $p_1^d = d + t/3(\alpha + 1)$
 (4d) $p_2^d = d - t/3(\alpha + 1)$
 (5d) $\Pi_1^d = d(\alpha + 1)/2 + t/3 + t^2/[18d(\alpha + 1)] - 2K$

$$(6d) \quad \Pi_2^d = d(\alpha + 1)/2 - t/3 + t^2/[18d(\alpha + 1)] - K$$

$$e) \quad (1e) \quad D_i^b = (\alpha + 1)/4 \quad i = 1,2$$

$$(2e) \quad p_i^b = d/2 \quad i = 1,2$$

$$(3e) \quad \Pi_i^b = d(\alpha + 1)/4 - 2K \quad i = 1,2$$

Here, superscripts denote cases (a,...,e) and subscripts denote firms. In the following, the propositions are restated and their proofs are given. I also state the conditions under which each branch network configuration is a subgame perfect Nash equilibrium.

proposition 2: if it is not profitable for either firm to open more than one branch, then the firms will choose different locations, and firm 1 will open its branch in the bigger city A.

proof: To prove this proposition, it has to be shown that if the branches of the firms are in different cities, the profits of the firm in city A are higher than those of its rival. This follows from eq. (1a) - (4a). Inspecting these it is easy to see that both the demand and price of firm 1 (which is assumed to have a branch in city A) are higher than its rival's, and thus also profits. The second part - that firm 2 locates in city B - can be proved by calculating the difference in profits that firm 2 gets in locating in either city

$$\Pi_2^a - \Pi_2^b = [3d + 2t + \alpha(3d-2t)]/12 + t^2(\alpha - 1)^2/[18d(\alpha + 1)] > 0$$

For this to be an equilibrium, it must be unprofitable for either firm to open a second branch. This is the case if the following relationship holds:

$$\Pi_1^a - \Pi_1^c = K - t(\alpha + 1)/6 + t^2(1 - 2\alpha)/[18d(\alpha + 1)] \geq 0$$

This holds for small enough t and α and large enough d and K . The relationship says that it must be unprofitable for firm 1 to open a new branch even if firm 2's first branch is located in city B, a situation that gives a two-branch firm 1 larger profits than a situation where firm 2's only branch is in city A. The weak inequality sign follows from assumption 5.

Q.E.D.

proposition 3: If the optimal number of branches per firm is 1, then firm 1 (located in A) sets a higher price and has a higher demand.

proof: by inspecting equations (1a) - (4a) Q.E.D.

proposition 4: If it is profitable for just one firm (firm 1) to open two branches, then its rival will open its branch in city A. Firm 1 will set a higher price, and have a higher demand than its rival.

proof: the firm with two branches will be firm 1. The profits of firm 2 if its only branch is in city A and B are, respectively, given by equations (6c) and (6d). It is easy to see that the latter is at least as large as the former. For this to be an equilibrium, firm 1's profits have to grow when opening

a second branch, and firm 2 must find it unprofitable to open a second branch. The following two relationships must hold:

$$\Pi_1^d - \Pi_1^a = t(3 - \alpha)/6 + \alpha(2 - \alpha)t^2/[18d(\alpha + 1)] - K \geq 0$$

$$\Pi_2^d - \Pi_2^e = (\alpha + 1)/4 - t/3 + t^2/[18d(\alpha + 1)] + K \geq 0$$

Let the first hold by equality, solve for K and insert into the second equation. It is easy to see that for small enough values of α, d and t , the latter holds.

Q.E.D.

proposition 5: If firm 1's profits are higher with two branches than with one given that firm 2 has just one branch (and then the same applies to firm 2), and if firm 2's profits are higher with two branches than with one given the number of firm 1's branches, then the equilibrium number of branches is two for both firms. The profits are, however, lower compared to a situation where both firms would have just one branch.

proof: For this to be an equilibrium, the following two relationships have to hold:

$$\Pi_1^d - \Pi_1^a = t(3 - \alpha)/6 + \alpha(2 - \alpha)t^2/[18d(\alpha + 1)] - K > 0$$

$$\Pi_2^e - \Pi_2^d = t/3 - t^2/[18d(\alpha + 1)] - (\alpha + 1)/4 - K > 0$$

Let the latter hold with equality, solve for K and insert into the former.

Then it is easy to see that the former holds for small enough α . Both inequalities are strong because of assumption 5. Q.E.D.

proposition 6: An oligopolistic firm that maximizes profits does not necessarily minimize costs.

proof: In the proofs for propositions 4 and 5 it was shown that for given parameter values, there can be either an asymmetric branch network equilibrium where firm 1 has two branches, or a symmetric one with both firms having two branches. As pointed out in the text, this branch proliferation leads to an increase in average costs through a duplication of fixed costs because the (possible) gain in market share is not large enough (the rise in market share is zero for both firms in case of the symmetric equilibrium and less than 100% for firm 1 in the asymmetric equilibrium) to compensate for the rise in costs. Q.E.D.

V SUMMARY

In the introductory chapter I stated as the objective of the thesis to apply industrial economics tools to the study of insurance markets in an intelligent way. One reflection of what I meant with this is apparent from the approach of the second chapter. Instead of applying a model deduced from game theory to the data at hand, I chose an indirect route and used the persistence of profits model. The reasons for this are stated in the chapter, but they are worth repeating here. The choice of the empirical model reflects my belief that the existing theoretical models do not adequately capture the special features of insurance markets in general (asymmetric information and all that follows from it) and of Finnish non-life insurance markets in particular (multiproduct firms, heterogeneous ownership forms, regulation). The chosen approach is a good first approximation that allows the analysis of the degree and type of competition in a market that escapes more formal modelling approaches. The results show that there are two distinct strategic groups, and that they behave differently from one another. The fringe firms do not compete so much with each other as with the leader firms. The leader firms either compete hard against each other or are engaged in some sort of (possibly tacit) collusion. The mergers of 1983 had clear effects on the behaviour of the market: the fringe became subject to tighter competition, and the punishment strategy used within the leader group became more pronounced.

The third chapter is very different in character compared with the second: the adopted approach is theoretical, and not linked to the institutional environment of any particular insurance market. The question asked is: does it pay for an insurance firm to hire an agent to screen the customers so as to find out their true type? Others than in most of vertical integration literature, here I study whether or not it pays to add another layer to the organization or not. The layer would not exist outside the organization (that is the implicit assumption of the model). It turns out that it can indeed be the case that a monopolist (or an oligopolist) wants to become vertically integrated. Furthermore, whether or not this is a socially optimal decision depends on the situation. The oligopoly model, especially its pure adverse selection version, yields some clear-cut predictions vis-à-vis direct selling firms and vertically integrated ones: the latter are more profitable, bigger, have lower per contract loss payments but higher in aggregate, and they are engaged in vertical product differentiation. As in the "pure" vertical product differentiation literature, also here a firm achieves this by making sunk investments. In the former literature these are in advertising or R&D, here they are directed into a branch network. The model - as is usual for models with asymmetric information - has other interpretations, too. It shows that if a firm is able to find out the true productivity of its job applicants, it does not pay for them to invest in education as in the standard models.

The fourth chapter in a way combines the approaches of the two previous

ones, containing both theoretical and empirical work. There it is argued that the standard cost function techniques cannot be straightforwardly applied to a services industry like insurance. The crucial difference to manufacturing is that in services (inclusive retailing) most firms are in effect multiproduct firms, each branch being the equivalent of a manufacturing plant. Others than in manufacturing, the location of branches has other effects than merely affecting transportation costs of the firm: the location decisions of branches and the decision of how many of them to have both affect the market power of the firm. It might well be (as turns out to be the case with the industry under study) that cost minimization is not equivalent to profit maximization. Oligopolistic services firms have to balance two effects: the increase in market power through an additional branch and the effect on costs that the additional branch has. These problems are studied in a modified Hotelling framework that has two cities. It turns out that it indeed is a theoretical possibility that there exist diseconomies of scale at the firm level and economies of scale at the branch level. This theoretical possibility is confirmed to be an empirical fact in the empirical part of the data, where a three year data set from the Finnish non-life insurance industry is used. Furthermore, the effects of this branch proliferation on firm costs is estimated: for those firms that have a nation-wide branch network, the part of costs that is due to the pursuit of market power is substantial, the lowest estimate being round 30% for any given firm. The results cast doubts on the results of earlier studies on banking and insurance, where the treatment of the

branch variable has seldom (and the interpretation of branch opening decisions never) been the same as in this study.

In both empirical chapters the smallness of the market, and thus the data, posed considerable difficulties, and the methods chosen reflect these. Apart from being an interesting question in itself, the decision to concentrate on strategic groups instead of firms was due to the fact that there are so few firms that are present for a long enough time. The adopted approach allowed also the inclusion of those firms that exit very early (after entry, or after start of the observation period). The solutions to this problem have varied in earlier studies, but the approach of this study precludes the possibility of sample selection bias. To be able to estimate a cost function, a way had to be found to restrict the excessive number of parameters. This was done by restricting the different products - apart from the branch variable, which was treated like any other product variable - to have same coefficients. Thus the linear terms, the power terms and the cross-product terms of the cost function had each the same coefficient for every product. To be exact, some variability was allowed, since the measure of output was the number of products (policies) sold, and these were converted into monetary units to allow a summing up of different products. Thus the true coefficient of any product is the regression coefficient times the price of the product. Prices that were used in this procedure were cleaned of market power influences by using the prices of the biggest firm in the sample.

As of future research, several possible avenues spring on mind. The obvious (but by no means easy) way to continue from chapter two is to build an explicit game-theoretical model that takes into account the institutional environment of the market. One possibility would be to treat the firms as if they were producing just one product, and use the breadth of products as a measure of (vertical) product differentiation. Another possibility is to concentrate on specific insurance policies and to try to measure the effects of asymmetric information. With regard to the theoretical chapter, testing the theory would, no doubt, be interesting. Before that would be possible, the model would have to be extended to allow for horizontal product differentiation, and categorization. A straightforward theoretical extension would be to study a market where brokers are allowed. Brokers should be modelled as agents of the customers, and thus the principal-agent relationship would be different from the one in the present model. In this extension, this relationship should be studied in more detail than has been done in chapter three. The last chapter is in this sense the easiest: both the theoretical model and the empirical part suggest some future research. The model of the theory section is - notwithstanding its length - only sketched, and warrants a full investigation. The empirical part suggests first of all that it would probably be worthwhile to re-estimate some of the cost function studies published in recent years in services, notably banking. It would also be interesting to include a frontier analysis, to test for the (only briefly outlined) predictions of the model

with respect to inefficiency measures¹. Another interesting option is to compare the cost-efficiency of banking and insurance on an inter-country set of data. The theory laid out in the fourth chapter, as well as the empirical results, suggest that the costs of these industries in any given country are to a great extent affected by the market environment, that is the geographical area and density of population. Also the base used in normal efficiency measures (in insurance always premium income) is subject to market power effects. To be really able to compare the effectiveness of firms in different countries (=environments), these effects should be accounted for.

¹ This was unfortunately not possible with the data set used in chapter four.

REFERENCES

- Akerlof, G. (1970) The market for lemons: Quality uncertainty and the market mechanism, *Quarterly Journal of Economics*, 89, 488-500
- Arnott, R. J. and Stiglitz, J. E (1988) Basic analytics of moral hazard, *Scandinavian Journal of Economics*, 90, 383-413
- Allen R. F. (1974) Cross-sectional estimates of cost economies in stock property liability companies *Review of Economics and Statistics*; notes, 100-103
- Anderson, S. P., de Palma, A. and Thisse, J.-F. (1992) *Discrete choice theory of product differentiation*, The MIT Press
- Ang J. S. and Lai T-Y. (1987) Insurance Premium Pricing and Ratemaking in Competitive Insurance and Capital Markets *Journal of Risk and Insurance* vol. LIV, no. 4 (Dec. 1987), 767-779
- Arrow, K. J. (1975) Vertical integration and communication, *Bell Journal of Economics*, Spring 1975, 173-183
- Barton, D. M. and Sherman, R. (1984) The price and profit effects of horizontal merger: a case study, *Journal of Industrial Economics*, vol. XXXIII, no. 2, 165-177
- Bauer P. W. (1990) Recent Developments in the Econometric Estimation of Frontiers *Journal of Econometrics*, 1990, 46, 39-56
- Baumol, W. J., Panzar, J. C. and Willig, R. D. (1982) *Contestable markets and the theory of industry structure*, Harcourt Brace Jovanovich, New York
- Ben-Akiva, M., de Palma, A. and Thisse, J.-F. (1989) Spatial competition with differentiated products, *Regional Science and Urban Economics*, 19, 5-19
- Benston, G. J., Hanweck, G. A. and Humphrey D. B. (1982) Scale economies in banking, *Journal of Money, Credit and Banking*, vol. 14, no. 4, 435-456
- Berger, A. N. and Humphrey, D. B. (1991) The dominance of inefficiencies over scale and product mix economies in banking, *Journal of Monetary Economics*, vol. 28, 117-148
- Bolton, P. and Bonanno, G. (1988) Vertical restraints in a model of vertical product differentiation, *Quarterly Journal of Economics*, 107, 555-570

- Bonanno, G. (1987) Location choice, product proliferation and entry deterrence, *Review of Economic Studies*, LIV, 37-45
- Bond E. W. and Crocker K. J., 1991, Smoking, Skydiving and Knitting; The Endogenous Categorization of Risks in Insurance Markets with Asymmetric Information *Journal of Political Economy* vol. 99, no. 1, 177-200
- Braeutigam, R. R. and Pauly, M. V. (1986) Cost function estimation and quality bias: the regulated automobile insurance industry, *Rand Journal of Economics*, vol. 17, no. 4, 606-617
- Brander, J. A. and Eaton, J. (1984) Product line rivalry, *American Economic Review*, vol. 74, 323-334
- Carlton, D. W. (1979) Vertical integration in competitive markets under uncertainty, *The Journal of Industrial Economics*, vol. XXVII, no. 3, 189-209
- Caves R. E and Porter M. E. (1977) From Entry Barriers to Mobility Barriers: Conjectural Decisions and Contrived Deterrence to New Competition *Quarterly Journal of Economics* 91 (May 1977), 421-441
- Caves R. E. and Ghemawat P., 1992, Identifying mobility barriers *Strategic Management Journal*, vol. 13, 1-12
- Christensen, L. R., Jorgenson, D. W. and Lau, L. J. (1973) Transcendental logarithmic production frontiers, *Review of Economics and Statistics*, vol. 55, no. 1, 28-45
- Cho, D. (1988) Some evidence of scale economies in worker's compensation insurance, *Journal of Risk and Insurance*, vol. 55, no. 2, 324-330
- Cremer, H. and Thisse, J.-F. (1991) Location models of horizontal differentiation: a special case of vertical differentiation models, *Journal of Industrial Economics*, vol. XXXIX, no. 4, 383-390
- Crocker, K. J. (1983) Vertical integration and the strategic use of private information, *Bell Journal of Economics*, Spring 1983, 236-248
- Crocker, K. J. and Snow A. (1986) The efficiency of categorical discrimination in the insurance industry, *Journal of Political Economy*, 94, 321-344, reprinted in Dionne, G. and Harrington S. E (ed.) 1992
- Cubbin, J. and Geroski, P., 1987, The convergence of profits in the long-run: inter-firm and inter-industry comparisons, *The Journal Industrial Economics*, vol. XXXV, no. 4, 427-442

- Cummins J. D. and Harrington S. E. (1987) The Impact of Rate-Regulation in U. S. Property-Liability Insurance Markets: A Cross-sectional Analysis of Individual Firm Loss Ratios *The Geneva Papers on Risk and Insurance* no. 42 (January 1987), 50-62
- Cummins J. D. and Vanderhei J. L. (1979) A note on the relative efficiency of property-liability insurance distribution systems *Bell Journal of Economics* 10, 709-720, reprinted in Dionne, G. and Harrington S. E. (ed.) 1992
- Daly, M. J., Rao P. Someshwar and Geehan, R. (1985) Productivity, scale economies and technical progress in the Canadian life insurance industry, *International Journal of Industrial Organization*, vol. 3, 345-361
- Dasgupta, P. and Maskin, E. (1986a) The existence of equilibrium in discontinuous economic games, I: Theory, *Review of Economic Studies*, LIII, 1-26
- Dasgupta, P. and Maskin, E. (1986b) The existence of equilibrium in discontinuous economic games, II: Applications, *Review of Economic Studies*, LIII, 27-41
- Dionne, G. and Harrington, S. E. (ed.) (1992) *The Foundations of insurance economics, readings in economics and finance*, Kluwer Academic Publishers, Boston
- Dixit, A. and Stiglitz, J. (1977) Monopolistic competition and optimum product diversity, *American Economic Review*, vol. 44, 297-308
- Dobson P., and Waterson, M. (1993) Product range and interfirm competition, *Warwick University working paper no. 9319*
- Eaton, B. C. and Lipsey R. G. (1975) The principle of minimum differentiation reconsidered: some new developments in the theory of spatial competition, *Review of Economic Studies*, vol. 42, 27-49
- Fecher, F., Perelman, S. and Pestieau, P. (1992) Scale economies and performance in the French insurance industry, *working paper, University of Liège*
- Ferrier G. D. and Lovell C. A. (1990) Measuring the cost efficiency in banking: econometric and linear programming evidence *Journal of Econometrics*, 1990, 46, 229-245
- Finsinger J. and Schmid F. A. (1991) Prices, Distribution Channels and Regulatory Intervention in European Insurance Markets *a paper presented at the EARIE conference 1991*

- Fudenberg, D. and Tirole, J. (1991) *Game theory*, MIT Press, Cambridge, Mass.
- Geroski P. A. (1990) Modelling persistent profitability in *The dynamics of company profits* ed. Mueller D. C. Cambridge University Press
- Geroski P. A. and Masson R. T., 1987, Dynamic market models in industrial organization *International Journal of Industrial Organization* vol. 5, no. 1 (March 1987), 1-14
- Geroski P. A., Masson R. T. and Shanaan J., 1987, The dynamics of market structure *International Journal of Industrial Organization* vol. 5, no. 1 (March 1987), 93-100
- Greene J., 1991 *LIMDEP Version 6.0 User's manual and reference guide* Econometric Software Inc. Bellport, NY.
- Grossman, S. and Hart, O. (1986) The costs and benefits of ownership: A theory of vertical and lateral integration, *Journal of Political Economy*, 94, 691-719
- Guasch, J. L. and Weiss, A. (1980) Wages as sorting mechanism in competitive markets with asymmetric information, *Review of Economic Studies*, 49, 653-664
- Hannan, T. H. (1991) Foundations of the structure-conduct-performance paradigm in banking, *Journal of Money, Credit and Banking*, vol. 23, no. 1, 68-84
- Hart, O. D. and Holmstrom, B. (1987) The theory of contracts, in T. Bewley, ed., *Advances in Economic Theory, Fifth World Congress*, Cambridge, Cambridge University Press
- Holmstrom, B. (1979) Moral hazard and observability, *Bell Journal of Economics*, 10, 74-91
- Hsiao, C., 1986, *Analysis of panel data*, Cambridge University Press
- Jaynes, G. D. (1978) Equilibria in monopolistically competitive insurance markets, *Journal of Economic Theory*, 19, 394-422
- Joskow P. L. (1973) Cartels, competition and regulation in the property-liability insurance industry *Bell Journal of Economics* 4, 327-427, reprinted in Dionne, G. and Harrington S. E (ed.) 1992
- Joskow P. L. and Rose N. L. (1989) The Effects of economic regulation in *Handbook of Industrial Organization*, vol. 1, ed. Schmalensee R. and Willig R.D. Elsevier Science Publishers

- Kim, E. H. and Singal, V (1993) Mergers and market power: evidence from the airline industry, *American Economic Review*, vol. 83, no. 3 549-569
- Kolari J. and Zardkoochi (1990) Economies of scale and scope in thrift institutions: the case of Finnish cooperative and savings banks *Scandinavian Journal of Economics* 92, (3), 437-451
- Lal, R. and Matutes, C. (1989) Price competition in multimarket oligopolies, *Rand Journal of Economics*, vol. 20, no. 4, 516-537
- Lambson V. E., 1992, Competitive profits in the long run *Review of Economic Studies*, 59, 125-142
- Lawrence, C. (1989) Banking costs, generalized functional forms, and estimation of economies of scale and scope, *Journal of Money, Credit and Banking*, vol. 21, no. 3, 368-379
- Levy D. (1987) The speed of the invisible hand *International Journal of Industrial Organization* vol. 5, no. 1 (March 1987), 79-92
- MacMinn R. D. and Witt R. C. (1987) A financial theory of the insurance firm under uncertainty and regulatory constraints *Geneva Papers on Risk and Insurance* 12, no. 42, 3-20
- Mayer, C. and Vives, X. (ed.) (1993) *Capital markets and financial intermediation*, Cambridge, CUP
- Mayers, D. and Smith, C. W. (1982) On the corporate demand for insurance, *Journal of Business*, vol. 54, 407-434, reprinted in Dionne, G. and Harrington S. E (ed.) 1992
- Milgrom, P. and Roberts, J. (1992) *Economics, organization and management*, Prentice-Hall, New Jersey
- Miyazaki, H. (1977) The rat race and internal labour market, *Bell Journal of Economics*, 8, 394-418
- Mueller D. C., 1977 The persistence of profits above the norm *Economica* 44, 369-380
- Mueller D. C., 1986, *Profits in the Long Run*, Cambridge University Press
- Mueller D. C., 1990, *The dynamics of company profits* ed. Mueller D. C. Cambridge University Press
- Murray, J. D. and White, R. W. (1983) Economies of scale and economies of scope in multiproduct financial institutions: a study of British Columbia credit unions, *Journal of Finance*, vol. XXXVIII, no. 3, 887-

- Nalebuff, B. and Scharfstein, D. (1987) Testing in models of asymmetric information, *Review of Economic Studies*, LIV, 265-277
- Newman H. H. (1978) Strategic Groups and the Structure-Performance Relationship *Review of Economics and Statistics* 60 (August 1978), 417-427
- Nooteboom, B. (1993) The hexagonal city and higher dimensions of product differentiation, *Rijksuniversiteit Groningen working paper*, July 1993
- OECD Report (1991) *Supervision of private insurance in Finland* (June 1991)
- Oster S. (1981) Intraindustry structure and the ease of strategic change *Review of Economics and Statistics* 63,
- Peltzmann, S. (1991) The handbook of industrial organization: a review article, *Journal of Political Economy*, vol. 99, no. 1, 201-217
- Pentikäinen T., Bonsdorff H., Pesonen M., Rantala J. and Ruohonen M. (1989) *Insurance solvency and financial strength* Finnish Insurance Training and Publishing Company Ltd Helsinki
- Porter M. E. (1979) The Structure within Industries and Companies Performance *Review of Economics and Statistics* 61 (May 1979), 214-228
- Prager, R. A. (1992) The effects of horizontal mergers on competition: the case of the Northern Securities Company, *Rand Journal of Economics*, vol. 23, no.1, 123-133
- Rasmusen, E. (1989) *Games and Information: An introduction to game theory*, Basil Blackwell, Oxford
- Raviv, A. (1979) The design of an optimal insurance policy, *American Economic Review*, vol. 69, no.1, 84-96, reprinted in Dionne, G. and Harrington S. E (ed.) 1992
- Rey, P. and Tirole, J. (1986) The logic of vertical restraints, *American Economic Review*, vol. 76, no. 5, 921-939
- Riley, J. G. (1979) Informational equilibrium, *Econometrica*, vol. 47, no. 2, 331-355
- Rothschild M. and Stiglitz J. (1976) Equilibrium in competitive insurance markets: An essay on the economics of imperfect information *Quarterly Journal of Economics* 90, 629-650, reprinted in Dionne, G.

and Harrington S. E (ed.) 1992

- Salant S. W., Schwitzer S. and Reynolds, R. J., 1983, Losses from horizontal merger: the effects of an exogenous change in industry structure on Cournot-Nash equilibrium *Quarterly Journal of Economics*, vol. XCVIII, no. 2, 185-199
- Salo S. (1980) *Vakuutuslaitosten lyhyen aikavälin ennustejärjestelmä (short-term forecasting system for insurance institutions)*, ETLA sarja B
- Shaked, A. and Sutton, J. (1983) Natural oligopolies, *Econometrica*, vol. 51, 1469-1484
- Shaked, A. and Sutton, J. (1990) Multiproduct firms and market structure, *Rand Journal of Economics*, vol.21, no.1, 45-62
- Shavell S. (1979) On Moral Hazard and Insurance *Quarterly Journal of Economics* 93, 541-562, reprinted in Dionne, G. and Harrington S. E (ed.) 1992
- Singh, N. and Vives, X. (1984) Price and quantity competition in a differentiated duopoly, *Rand Journal of Economics*, vol. 15, 546-554
- Skogh G. (1982) Returns to Scale in the Swedish Property-Liability Insurance Industry *The Journal of Risk and Insurance* 49, no. 2, 218-228
- Spence, M. (1974) *Market Signalling: Information Transfer in Hiring and Related Processes*, Cambridge, Mass.: Harvard University Press
- Spence, M. (1976) Product selection, fixed costs and monopolistic competition, *Review of Economic Studies*, vol. 43, 217-235
- Spence, M. (1978) Product differentiation and performance in insurance markets, *Journal of Public Economics*, 10, 427-447
- Stahl, K. (1987) Theories of urban business location, in E. S. Mills and P. Nijkamp (eds.), *Handbook of Regional and Urban Economics*, vol. 2: *Urban economics*. Amsterdam: North-Holland
- Stiglitz, J. (1977) Non-linear pricing and imperfect information: The insurance market, *Review of Economic Studies*, vol. 44, 407-430
- Suret J. M. (1991) Scale and Scope Economies in the Canadian Property and Casualty Insurance Industry *Geneva Papers on Risk and Insurance* 16, no. 59, 236-256
- Sutton J., 1991, *Sunk Costs and Market Structure: Price Competition, Advertising, and the Evolution of Concentration*, Cambridge, Mass., the

MIT Press 1991

- Tirole, J. (1988) *The theory of industrial organization*, the MIT Press, Cambridge, Mass.
- Toivanen O. (1992a) *Sääntely, kilpailutekijät ja kilpailu Suomen vakuutusmarkkinoilla: esitutkimus*, (Regulation, structure and competition in the Finnish insurance market; in Finnish only) Kilpailuviraston selvityksiä sarja 6/92 Publication series of the Office of Free Competition
- Toivanen O. (1992b) *An industrial organization study of the Finnish domestic non-life insurance market* unpublished Lic. Sc.-thesis, Turku School of Economics and Business Administration
- Valkonen, T. (1990) *Insurance company investment in Finland* (with English summary), ETLA (The Research Institute of The Finnish Economy) series C 56
- Waterson, M. (1989) Models of product differentiation, *Bulletin of Economic Research*, vol. 41, no. 1, 1-27
- Waterson, M. (1993) Vertical integration and vertical restraints, *Oxford Review of Economic Policy*, vol. 9, no. 2, 41-57
- Wilson, C. A. (1977) A model of insurance markets with incomplete information, *Journal of Economic Theory*, 16, 167-207
- Zweifel P. and Ghermi P. (1990) Exclusive vs. Independent Agencies: A Comparison of Performance, *Geneva Papers on Risk and Insurance*, vol. 15, no. 2, 171-192

THE BRITISH LIBRARY

BRITISH THESIS SERVICE

TITLE INDUSTRIAL ECONOMICS STUDIES IN
INSURANCE MARKETS.

AUTHOR Otto Iisakki
TOIVANEN

DEGREE Ph.D

AWARDING BODY Warwick University

DATE 1994

THESIS NUMBER DX186403

THIS THESIS HAS BEEN MICROFILMED EXACTLY AS RECEIVED

The quality of this reproduction is dependent upon the quality of the original thesis submitted for microfilming. Every effort has been made to ensure the highest quality of reproduction. Some pages may have indistinct print, especially if the original papers were poorly produced or if awarding body sent an inferior copy. If pages are missing, please contact the awarding body which granted the degree.

Previously copyrighted materials (journals articles, published texts etc.) are not filmed.

This copy of the thesis has been supplied on condition that anyone who consults it is understood to recognise that its copyright rests with its author and that no information derived from it may be published without the author's prior written consent.

Reproduction of this thesis, other than as permitted under the United Kingdom Copyright Designs and Patents Act 1988, or under specific agreement with the copyright holder, is prohibited.

DX

186403