Four Essays in Insurance

Inaugural-Dissertation
zur Erlangung des akademischen Grades
Doktor der Wirtschafts- und Sozialwissenschaften
(Dr. rer. pol.)
an der Friedrich-Alexander-Universität Erlangen-Nürnberg

vorgelegt von
Michael Sonnenholzner
2005

Referent: Prof. Achim Wambach, Ph.D.
Korreferent: Prof. Dr. Berthold U. Wigger
Promotionstermin: 12.07.2005
Acknowledgements

This doctoral dissertation was written at the University of Erlangen-Nürnberg under the supervision of Professor Achim Wambach Ph.D. from 2002 to 2005. I would like to thank Professor Wambach for the encouragement and scientific advice he provided during my studies in Nuremberg. Working at his chair was a very pleasant experience due to an excellent and cooperative atmosphere, challenging and inspiring scientific work and a low non-scientific workload. Furthermore, I want to thank Professor Dr. Berthold U. Wigger for agreeing to be the second reviewer of this thesis.

I wish to thank my colleagues Andreas R. Engel, Jesko Herre, Kristina Kilian, Alexander Rasch, Rüdiger Reissaus and Kerstin Windhövel. They supported my work with helpful comments and were good fellows also in the non-academic part of life. Furthermore, I would like to thank Ursula Briceño La Rosa for handling most of the chair’s administrative duties and thus increasing the available time for research significantly. I am also grateful to our student helpers, especially Sarah Borgloh, Martin Götz, Wanda Mimra and Oliver Poletti.

Financial support by the Staedtler Stiftung is gratefully acknowledged.

The third paper (chapter 4) of this thesis was partly written during a research visit to the University of Toulouse. Financial support was provided by a Marie Curie grant from the European Union. I am very grateful to both institutions. During my stay in Toulouse, Professor Christian Gollier and Professor Bertrand Villeneuve provided very valuable advice on my research project.

Last but not least, I want to thank my friends and family who supported me during the recent years and thus made a great contribution to this work.

July 2005

Michael Sonnenholzner
## Contents

1 Introduction .................................................. 1

2 Moral Hazard .................................................. 9
   2.1 General remarks ........................................... 9
   2.2 The model with finite efforts ............................ 10
      2.2.1 The general program ............................... 11
      2.2.2 The program with two efforts .................... 12
   2.3 Continuous effort levels ................................. 14
   2.4 Continuous losses ....................................... 14
      2.4.1 Loss prevention .................................. 15
      2.4.2 Loss reduction .................................. 16
   2.5 Exclusive Contracts ..................................... 18
   2.6 Many periods ............................................ 18
   2.7 Renegotiation ............................................ 19

3 Oligopoly in Insurance Markets ............................ 21
   3.1 General remarks .......................................... 21
   3.2 Factors favoring oligopolies: Standard industrial organization models 22
      3.2.1 Capacity constraints ............................... 22
      3.2.2 Product differentiation ............................ 23
      3.2.3 Search and switching costs ....................... 25
      3.2.4 Barriers to entry and regulation .................. 26
   3.3 Factors favoring oligopolies: Insurance specific models ............... 27
      3.3.1 Adverse selection when consumers differ in risk and wealth 27
      3.3.2 Competition under moral hazard ................... 28
### 4 The Role of Patience in an Insurance Market

4.1 Introduction ........................................... 30
4.2 The Model .............................................. 33
   4.2.1 Insurers ........................................ 33
   4.2.2 Consumers ..................................... 34
4.3 Effort border ......................................... 35
4.4 Equilibria ............................................. 36
   4.4.1 Definition of Equilibria ....................... 36
   4.4.2 Analysis of Equilibria ...................... 37
4.5 Conclusion ........................................... 55

### 5 Insurance Brokers and Advice Quality

5.1 Introduction ........................................... 58
5.2 Related Literature .................................... 60
   5.2.1 Explaining coexistence: Principal agent models .. 60
   5.2.2 Explaining coexistence: Broker provides service .. 66
5.3 The Model ............................................. 72
   5.3.1 Consumers ..................................... 72
   5.3.2 Insurance agents ............................... 73
5.4 Monopoly broker without threat of new entry ............ 74
   5.4.1 Monopoly broker’s price and signal quality .. 74
   5.4.2 Socially optimal price and signal quality .... 77
5.5 Incumbent broker with threat of new entry ............... 79
   5.5.1 Full flexibility of $p_I$ and $s$ ............... 81
5.5.2 Full flexibility of $p_I$ but $s$ cannot be changed . . . . . . . 82
5.5.3 Neither $p_I$ nor $s$ can be changed . . . . . . . . . . . . . . . 84
5.6 Policy implications . . . . . . . . . . . . . . . . . . . . . . . . . . 85
5.6.1 Advantageousness of regulation . . . . . . . . . . . . . . . . 85
5.6.2 Regulation using $s$ and $p_I$ . . . . . . . . . . . . . . . . . . 86
5.6.3 Regulation using $s$ only . . . . . . . . . . . . . . . . . . . . 86
5.7 Conclusion . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 89
1 Introduction

This thesis contains four essays which are all from the broader field of insurance economics. The chapters can be read independently, although there is a certain connection between the articles. Chapter 2, which deals with moral hazard, and chapter 3, which is on oligopolistic models in the insurance market, are paving the way for chapter 4, which contains an article on the role of patience in an insurance market. This is because chapter 4’s model includes moral hazard and adds to the literature cited in chapter 2. Finally, the last section of chapter 4 raises the issue of insurance brokers, whose role is explored in more detail in chapter 5.

Chapter 2 presents an overview on moral hazard in an insurance context. Moral hazard arises because the insurer cannot observe or verify the insuree’s actions after the policy is signed. As a consequence, the insuree has an incentive to change his behavior to the detriment of the insurer. In the case of ex-ante moral hazard, the consumer employs suboptimal precautionary effort in order to avoid the loss before the occurrence of the event insured. Therefore, the optimal contract under moral hazard is usually characterized by partial insurance in order to provide incentives for the insuree to employ precautionary effort.

An overview on oligopoly in insurance is given in chapter 3. The key characteristic of an oligopoly is that firms do not take price as given like under perfect competition. In contrast, there are few enough firms in the market in order that their price or quantity decision influences the overall market outcome. As a consequence, a situation of strategic competition among firms arises which may lead to above normal profits. There is some empirical evidence that insurers enjoy such above normal profits, at least in some markets (e.g. Murat et al. (2002)). However, the standard model of price competition (Bertrand (1833)), which is usually used for the insurance market, yields a zero profit for insurers. Therefore, the aim of chapter 3 is to give an overview of explanations of how an above normal profit can arise in spite of price competition. We break the analysis apart into two sections: standard industrial organization models and insurance specific models.
The articles of chapter 2 and 3 were published in the Encyclopedia of Actuarial Science (Wiley, 2004). I would like to thank Achim Wambach for very useful advice and for being a co-author.

The Role of Patience in an Insurance Market is explored in chapter 4. The article is motivated by various empirical studies which find that personal discount rates seem to vary substantially among individuals. Among others, evidence for this has been found by Warner and Pleeter (2001) or Frederick et al. (2002). The typical personal discount rate is reported to be anywhere in the range between 0 and 30%. As an insurance contract usually requires to prepay the premium for several periods (e.g. 12 months), there exists a role for the personal time preference in the ex-ante valuation of the contract. We analyze the effects of differences in the personal discount rates among individuals in the context of a traditional competitive insurance market. For this purpose we employ a two-period model with the simultaneous presence of moral hazard and adverse selection which draws on de Meza and Webb (2001). However, in contrast to de Meza and Webb, adverse selection does not occur regarding risk aversion but the individual’s personal discount rate, which can either be high (impatient) or low (patient). Moral hazard is modelled along the traditional lines: A high or a low level of precautionary effort in order to avoid the loss can be employed, leading to a low or high loss probability. The downside of high effort is that it is more expensive for the consumer than low effort.

The model is set up as a game with 4 stages throughout 2 periods. In stage 1, insurers make irrevocable offers of contracts which specify premium and indemnity. In stage 2, clients buy at most one contract from one insurance company. If they buy they have to pay the premium. In stage 3, the consumer decides whether to employ a high or low level of unobservable effort in order to avoid the loss. The costs of effort have to be paid for in this stage as well. Finally, in stage 4, the loss occurs or not and the indemnity is paid out in case of a loss. Stages 1 to 3 take place in period 1, whereas stage 4 occurs at period 2. In this sense, the consumers need to pay for the premium and the effort up-front. Consumers are identical up to their personal discount factor, which can either be high (patient) or low (impatient).
There are two or more risk neutral insurers in the market who compete in contracts. The insurers are assumed to know the distribution of patience types in society but they cannot observe the patience type of an individual client. Furthermore, they know the specification of the model, especially the utility function of the consumers, the two available effort levels and the resulting loss probabilities. This allows insurers to infer the effort level employed by patient and impatient consumers for any contract they offer.

The equilibria of the model are solved for with the help of the consumers’ indifference curves in the premium-indemnity space (contract space). Under identical effort, the indifference curves of patient types turn out to be steeper than of the impatient types. We establish an assumption which ensures a single crossing property also for diverging efforts in order to obtain stable equilibria. For each patience type there may exist an effort border, which separates the regions in the premium-indemnity space where the respective patience type employs high and low effort. If there is no effort border in the contract space, the accordant patience type will employ low effort under each possible contract. Patient types are more inclined to employ high effort.

First we look at a situation in which the impatient types never employ high effort but there is a high effort region for the patient types in the premium-indemnity space. We find a unique separating equilibrium in which the impatient types employ low effort and buy little or even no cover. In contrast, patient types employ high effort and buy more cover than the impatient types. As low effort results in a high loss probability, impatient types may be regarded as high risks, whereas patient types are low risks. In that sense, we obtain a similar result as de Meza and Webb (2001), which they refer to as the opposite of adverse selection: High risks buy little or no cover, whereas low risks are more fond of buying insurance. This result seems to be in line with some empirical findings. For example, Cawley and Philipson (1999) analyze whether data from life insurance is consistent with the adverse selection hypothesis as established by Rothschild and Stiglitz (1976). Their finding is that the data exhibits the opposite of the expected pattern, as there is a negative covariance between risk and quantity. Therefore, it seems as if it were actually the low risks who are more interested in
buying insurance. The separating equilibrium has a second interesting property: the contract for the patient types can be profit-making. The intuition behind this result is that undercutting the patient types’ contract would attract all impatient types as well, what results in an expected loss for the insurer. More information on profit-making equilibria which follow the same line of argument and empirical evidence for profit-making insurance contracts is provided in chapter 3 of this thesis.

If the assumption ensuring the single crossing property does not hold, the equilibrium outcome may be that insurers do not provide insurance at all. In this regard, accounting for differences in personal time preference among consumers can help to explain why insurance is not offered for some risks. Moreover, depending on the shape of indifference curves, we also find a unique zero-profit pooling equilibrium.

Regarding the profit-making separating equilibrium, we analyze if providing a randomized insurance contract could help to separate agents better and thus allows insurers to compete down profits. Following Arnott and Stiglitz (1988), we distinguish between ex-ante and ex-post randomization, which relates to whether the result of the lottery is made known to the consumer before or after his choice of effort. Under ex-ante randomization the consumer commits to buy any contract the sales agent proposes on the basis of a lottery carried out before the consumer chooses his effort level. Although unrealistic, such a lottery is indeed capable of achieving a perfect separation of types and a complete elimination of profits. Ex-post randomization can be achieved by linking the indemnity payment to the precise circumstances of the loss which are random themselves (Villeneuve (2003)). We show that a compensated ex-post randomization cannot improve separation of types. This is because both consumer types have the same utility function. Therefore, the compensation will encompass the same certainty equivalent for patient as well as for impatient types.

Finally, profits could also be competed away as a consequence of product differentiation. By introducing a discount policy in addition to existing full service policies, an insurer could improve screening for patient types. We assume, that the difference between full service and discount insurance is the time it takes until
the indemnity is paid out in case of a loss. Patient types do not mind such a
delayed payment as much as impatient types. We show that there exist cases in
which the profits can be competed away completely, depending on the shape of
the incentive compatibility constraint. Otherwise, at least a part of the profit
will persist in equilibrium.

This article was partly written during a research visit to the University of
Toulouse. I am grateful to Christian Gollier, Harris Schlesinger and Bertrand
Villeneuve, who provided very valuable advice. Furthermore, I wish to thank
Achim Wambach, who is a co-author of this article, for many fruitful discussions.
Earlier versions have been presented in seminars at the University of Erlangen-
Nürnberg and at the 2004 seminar of the European Group of Risk and Insurance
Economists (EGRIE) in Marseille.

Chapter 5 deals with Insurance Brokers and Advice Quality. Looking at this issue was motivated by the announcement of the EU’s directive
2002/92/EC regarding the mediation of insurance on January 15, 2003. Among
other arrangements, the directive sets up minimum requirements for the training
of insurance agents. In Germany, for example, they need to attend 230 hours of
professional training before they are allowed to sell policies to clients. The model
we create is intended to isolate the effect of an insurance agent’s advice quality
on social welfare. This allows to analyze whether regulation of advice quality is
useful. In this sense, the model is related to Gravelle (1994). Furthermore, the
model also contributes to the literature explaining the coexistence of brokers and
exclusive agents (e.g. Grossman and Hart (1986), Posey and Yavaş (1995)).

In order to isolate the effect of advice quality, we abstract from price dispersion
in the insurance market. Price search is of course an important service provided
by insurance brokers (e.g. Posey and Tennyson (1998)). However, the role of
the broker as analyzed in chapter 5 is to find the insurance policy which fits the
consumer’s needs best rather than looking for the cheapest policy. In our view,
this way of modelling the problem comes closest to the EU’s intention when
passing the directive mentioned above. In the end, one can think of the various
models as building blocks of the real insurance market.

We assume that there are two insurance policies from two different insurers
in the market. The policies are similar and cost the same, but they differ in some details which are not obvious to the consumer. The reason for the difference is exogenous. It may stem from insurers’ heterogeneity regarding marketing perceptions of what consumers want, comparative cost advantages or diverging regulation. For each consumer there exists an ex-ante probability that one particular of the two policies maximizes his expected utility. Consumers are assumed to be equally distributed on a line between zero and one regarding this ex-ante probability. Consumers who feel sufficiently sure can buy directly from the respective insurer (or an exclusive agent). However, those consumers who feel uncertain can contact an insurance broker. The broker will give them advice and sell the policy he recommends, however for a markup. He gives the correct recommendation with a certain probability, which we refer to as advice quality. Before entering the market, the broker decides on how much he wants to invest in advice quality. Higher advice quality is assumed to be associated with increasing and convex costs.

First, we compute the equilibrium price and advice quality of a monopoly broker in the case where no new entry is possible. Then we compare the result to the socially optimal price and advice quality. It turns out that the monopoly broker will charge a price which is too high and offer an advice quality which is too low. The reason for this is that the monopoly broker ignores the positive external effect of higher advice quality on the consumers’ surplus. The socially optimal price is zero as we abstract from any costs apart from sunk entry costs paid for building up advice quality. Therefore, the monopoly broker would suffer a loss in the social optimum which would have to be accounted for, possibly with non-distorting lump sum taxes raised elsewhere in the economy.

Second, we look at a situation in which the incumbent broker is subject to potential entry. In this context we assume that there is no stage of the game in which the incumbent is alone in the market before the threat of entry. The market outcome depends on the flexibility of price and advice quality once these values have been chosen by the incumbent broker for the first time. As a framework for the analysis, a 4-stage game is used. In stage 1, the incumbent broker chooses his price and advice quality. In stage 2, a potential entrant decides whether to
enter or not. If there is no entry, the game continues with stage 4. In stage 3, the incumbent broker and the new entrant simultaneously decide over their price and advice quality in Bertrand style competition. In this stage, it is relevant if the incumbent can change the values of price and advice quality which were chosen in stage 1. Finally, in stage 4, consumers buy and payoffs are realized. Different equilibria of the game arise depending on the flexibility of price and advice quality once the incumbent has chosen them in stage 1.

For computing equilibria, we distinguish 3 subcases: (1) full flexibility of both variables, (2) flexibility of price only and (3) inflexibility of both variables. In subcase (1), the equilibrium is that there will be no entry. The incumbent will offer the monopoly price-advice quality combination. The reason for this is that due to the flexibility of both variables, full fledged Bertrand competition is triggered by entry. Together with sunk entry costs this is sufficient for the incumbent to establish credible entry deterrence. In subcase (2) the equilibrium will be characterized by no entry as well. The incumbent may offer a better price-advice quality combination for consumers as the monopoly combination. However, this is only the case if the incumbent could not deter entry even by offering a price of zero under the monopoly advice quality. Then it is unavoidable for him to offer a better advice quality from the outset as he cannot change it anymore if entry should occur. Finally, in subcase (3) the equilibrium is no entry and the incumbent offering a zero-profit contract which maximizes the consumers’ expected utility. This is because of the inflexibility of both variables Bertrand competition must be fully anticipated by the incumbent in order to deter entry.

As to welfare the following result is obtained: In the case without threat of entry and in all of the 3 subcases with threat of entry welfare in the resulting equilibria is always below the socially optimal welfare level. Therefore the question of regulation arises. In the following we use the term “regulation of advice quality” in the sense of prescribing a binding minimum advice quality. If regulation can use both price and advice quality as an instrument, the first best welfare level can be achieved. Then, the broker’s loss will have to be accounted for somehow, possibly with non-distorting lump sum taxes. If price is not available as a policy instrument but regulation is restricted to influencing advice quality, 3 cases
need to be distinguished: (i) If the equilibrium is the monopoly outcome without threat of entry or subcase (1) (with threat of entry), regulation of advice quality only is always increasing social welfare. This is because a marginal increase in advice quality will keep profits constant as they were maximized with respect to advice quality, while consumer surplus is increased. (ii) If the equilibrium is subcase (2), the universal benefit of regulation cannot be guaranteed. However, simulation results suggest that there exists a benefit for a broad range of parameter specifications. (iii) If the equilibrium is subcase (3), regulation of advice quality only is never advantageous. This is because the incumbent is maximizing the consumers’ surplus anyway, subject to a zero profit. A further increase of advice quality will keep the broker on his zero-profit locus while the consumers’ surplus will be decreased.

In the real world, the brokers’ markup may not be available as a regulation instrument. On the one hand, it is very tedious and expensive to analyze the markup of thousands of similar insurance policies and a large bureaucracy would be required. On the other hand, introducing regulation for brokers’ commissions might be difficult to achieve politically as it contradicts the spirit of the EU’s 1994 insurance market deregulation. All in all, the results within the framework of the model suggest that regulating advice quality but not price may be reasonable in two out of three market situations.

This article has been accepted for the World Risk and Insurance Economics Congress 2005 in Salt Lake City.
2 Moral Hazard

In an insurance context, moral hazard comprises the phenomenon that having insurance gives the insured an incentive to alter his behavior to the detriment of the insurer. This manifests itself by the insuree employing suboptimal precautionary effort to avoid the loss. We discuss the insurance specific moral hazard model and several extensions with regard to many effort levels, continuous loss distribution, multi-period contracts and renegotiation.

2.1 General remarks

Moral hazard in an insurance context comprises the phenomenon that having insurance gives the insured an incentive to alter his behavior to the detriment of the insurer. In contrast to adverse selection, moral hazard is characterized by symmetric information between the insured and the insurer at the time the contract is signed. The information asymmetry arises afterwards due to unverifiable actions of the insured (hidden action).\footnote{Generally, a problem of moral hazard can arise due to hidden action or hidden information (Arrow (1991)). In the latter case the agent learns private information about the state of the world after the contract has been signed. As most of the literature on moral hazard in insurance, this article will focus on hidden action.}

Moral hazard comes in two forms: ex-ante moral hazard and ex-post moral hazard. If the change of the insured’s behavior takes place before the occurrence of the event insured this is called ex-ante moral hazard. The insured has an incentive to use less effort to prevent the loss or accident as soon as he signed the insurance policy. For example, if the insured has bought a motor insurance he might drive less carefully. Ex-post moral hazard is mostly relevant for health insurance. Here the insured changes his behavior after the event insured occurred by demanding too much treatment.\footnote{For further information on ex-post moral hazard refer to Folland et al. (1997).} Usually the unqualified term “moral hazard” refers to ex-ante moral hazard on which we will concentrate in the following.

The basic cause of moral hazard is due to the fact that the insurer cannot write a complete contract which would specify the exact level of precaution effort the insuree has to employ in every contingency. This is not possible if the
insuree’s effort is not observable or observable but not verifiable. The issue of non-observability is especially relevant if the effort is of a non-material nature and does not cause financial costs like the endeavor of driving carefully all the time. In contrast, if effort is equivalent to a financial investment like the installation of fire extinguishers, moral hazard is less of an issue. The insurer could simply include a clause in the insurance policy that in case of a fire the indemnification will only be paid if fire extinguishers were installed. Given that the insurer can verify ex-post whether or not the required investment was made the moral hazard problem disappears.

The fundamental trade-off in the context of moral hazard is as follows: More insurance has a positive effect on the insured’s expected utility since the consumer is assumed to be risk averse. But at the same time more insurance has a negative impact on the insured’s expected utility due to a more severe moral hazard problem which has to be accounted for with a higher premium. This trade-off describes what the literature often refers to as the right mix between risk-sharing and incentives.

The relevance of moral hazard models is not limited to insurance but the problem arises in a broad range of so-called principal-agent relationships. For example, an employee whose payment is independent from his performance may have less incentives to work hard.

2.2 The model with finite efforts

In this section we want to present the insurance specific moral hazard model when the insured can choose his effort $e$ from a finite set $E = \{e_1, e_2, ..., e_n\}$ only. It is assumed that there are two states of the world: “loss” and “no loss”. Moral hazard arises because the loss probability $\pi(e)$ depends on the level of effort the insured employs: $\pi(e_i) < \pi(e_j)$ if $e_i > e_j$, which means that a higher effort level leads to a lower loss probability. The cost of effort in utility units for the insured is $c(e_i)$ and a higher effort level causes more costs than a lower effort level. The expected utility of the insured is given by

$$E[U] = (1 - \pi(e)) U(W - P) + \pi(e) U(W - L + I^n) - c(e)$$
where $U$ is a utility function implying risk aversion ($U' > 0, U'' < 0$), $W$ is the initial wealth of the insured, $P$ is the insurance premium, $L$ is the amount of the loss and $I^n$ is the net indemnification in case of a loss (indemnification $I$ minus premium $P$).

A noteworthy feature of the above equation is the formulation of the effort term which implies that the insured’s utility function $U(W, e)$ is additively separable in wealth $W$ and effort $e$. The reduction in utility due to the effort does not depend on the state of the world but is always the same, no matter whether a loss occurred or not. The elegance of this approach is that the insured’s preferences over lotteries do not depend on the level of effort employed which facilitates mathematical tractability. Different ways of modelling effort may make the problem hard to solve or be less plausible. For example, a possible alternative would be to model monetary effort costs $U(W, e) = U(W - c(e))$ to represent effort as a financial investment like the installation of fire extinguishers. By doing so the effort is easily interpreted. However, as mentioned above, insurance policies can often condition on these precautionary investments and then the moral hazard problem ceases to exist.

If there is full, symmetric information, i.e. the insurer can perfectly verify and therefore condition the contract on the effort level, one calls this a first best world. In the second best however, there is asymmetric information between the insurer and the insuree: effort is not verifiable and contractible anymore.

### 2.2.1 The general program

In the second best world, it is possible that the insured prefers to employ the minimum effort. In this case he needs no incentives and can obtain full insurance for the corresponding fair premium. However, in order to make the problem interesting we assume in the following that the insured is better off with a contract implementing a higher effort level via an incentive scheme. Following Grossman and Hart (1983), the optimal second best contract can be found by solving the
following program:

\[
\max_{e \in E, P, I^n} (1 - \pi(e))U(W - P) + \pi(e)U(W - L + I^n) - c(e)
\]

s.t. P.C.: \((1 - \pi(e))P - \pi(e)I^n \geq 0\)

I.C.: \((1 - \pi(e))U(W - P) + \pi(e)U(W - L + I^n) - c(e) \geq (1 - \pi(e_i))U(W - P) + \pi(e_i)U(W - L + I^n) - c(e_i) \quad \forall e_i \in E\)

The optimal contract specifies a premium \(P\) and a net indemnification \(I^n\) such that the insured’s expected utility is maximized given the participation constraint (P.C.) and the incentive constraints (I.C.). The participation constraint states that the insurance company has to make non-negative profits because otherwise it would not be interested to write an insurance contract.\(^3\) The incentive constraints reflect the condition in order to make the insured employ the desired effort level \(e\): the insured’s expected utility using effort \(e\) must be (weakly) higher than under any other effort \(e_i\).

2.2.2 The program with two efforts

Solving the general program above is a tedious process involving the use of Kuhn-Tucker conditions. However, in the case with just two possible effort levels the optimal second best contract can be found much easier. The incentive scheme to implement the high effort results from the following optimization problem:

\[
\max_{P, I^n} (1 - \pi_h)U(W - P) + \pi_h U(W - L + I^n) - c_h
\]

s.t. P.C.: \((1 - \pi_h)P - \pi_h I^n \geq 0\)

I.C.: \((1 - \pi_h)U(W - P) + \pi_h U(W - L + I^n) - c_h \geq (1 - \pi_l)U(W - P) + \pi_l U(W - L + I^n) - c_l\)

where \(\pi_h = \pi(e_h), \pi_l = \pi(e_l), c_h = c(e_h), \text{ and } c_l = c(e_l)\).

In this case we can take advantage of the fact that both constraints have to be binding: The I.C. has to be binding because otherwise moral hazard would not be a problem and full insurance would be optimal. But this would make the insured

\(^3\)This reflects the implicit assumption that insurers are in a market of perfect competition. Other assumptions on the market structure, e.g. a monopoly insurer, do not change the insight provided.
use the low effort, which is a contradiction to the assumption that the insured’s expected utility is at a maximum with the high effort. The P.C. must also be binding because otherwise one could increase the expected utility of the insured by marginally lowering the premium and increasing the net indemnification in a suitable way which does not violate the I.C.\textsuperscript{4} This works out if an additively separable utility function is used. Under different assumptions on the relationship between utility and effort the result of a non binding P.C. may prevail as any change in indemnity or premium would modify the incentive structure. Such a situation is analyzed by Bennardo and Chiappori (2003), who show that even under price competition insurers may sustain positive profits in the presence of moral hazard.

In a fully specified model one can explicitly solve for the optimal second best contract \((P^{SB}, I^{SB})\) from the two binding constraints. But also in the general case we can obtain a conclusion about the optimal second best contract. Rearranging the incentive constraint while using the fact that it must be binding yields:

\[
(\pi_l - \pi_h) \left[ U(W - P) - U(W - L + I^n) \right] = c_h - c_l
\]

The left hand side of the equation describes the advantage of the insured using the high effort: Because the loss probability is smaller under the high effort he can enjoy the higher utility of the no loss state more often. In equilibrium, this advantage is balanced with the additional costs of the high effort, which are written on the right hand side.

Since the right hand side of the equation is positive \((c_h > c_l)\) and \(\pi_l > \pi_h\), the utility difference also needs to be positive. However, this can only be achieved if the indemnification \(I = I^n + P\) is smaller than the loss \(L\). This leads to the first core insight about the design of the optimal second best contract: In order to implement the high effort level in the presence of moral hazard, the insured must not obtain full insurance. The optimal contract balances the trade-off between the insured’s benefit of having greater insurance and his benefit of having less insurance which provides a better incentive to avoid the loss. Thus this result can explain that in practice we observe partial insurance and deductibles.

\textsuperscript{4}This can be done by lowering \(P\) by \(\epsilon/U'(W - P)\) and increasing \(I^n\) by \(\epsilon/U'(W - L + I^n)\) with \(\epsilon\) being small and positive.
2.3 Continuous effort levels

When allowing for continuous effort levels the procedure of writing the incentive compatibility constraint for every single effort is not possible anymore. A workaround is to replace the incentive compatibility constraint by the first order condition for the insured. This method is called first order approach (Holmström (1979)). When using the first order approach the incentive constraints of the general program above are replaced with the following expression:

\[ \text{I.C.} : - \pi'(e)(U(W - P) - U(W - L + I^n)) - c'(e) = 0. \]

As in the case of just two possible efforts, one can derive from the incentive constraint that in order to implement any effort level higher than the minimal effort partial insurance will be necessary.

The second order condition is:

\[ -\pi''(e)[U(w - P) - U(w - L + I^n)] - c''(e) < 0. \]

This is valid if costs of effort are convex \((c''(e) > 0)\) and the loss probability is a convex function of effort \((\pi''(e) > 0)\).

When allowing for more than two effort levels, the question arises as to how the insured will behave when effort is not contractible as compared to the first best. Unfortunately there is no clear answer because the result depends on the exact form of the trade-off between risk sharing and incentives. Thus the insured may use either more or less effort in the second best world.

2.4 Continuous losses

So far we have considered the case with only two possible outcomes or states of the world: either a certain loss occurs or not. This section extends the analysis of the moral hazard problem to situations with a continuous loss distribution. To structure the analysis the loss is modelled to be dependent on two components: the probability \(\pi\) that a loss occurs and a continuous distribution \(F(L)\) of the size of the loss \(L\) with support \([L, \bar{L}]\), given that the loss occurred. Then the insured’s expected utility is

\[
E[U] = (1 - \pi)U(W - P) + \pi \int_{\underline{L}}^{\bar{L}} U(W - L + I^n(L))dF(L) - c(e).^5
\]
In this context one can distinguish between loss prevention and loss reduction (Winter (2000)). Loss prevention describes the case where the loss probability $\pi(e)$ depends on the insured’s effort but not the distribution $F(L)$ of the loss. For example, the installation of high quality door locks reduces the probability of a theft. However, if a burglar manages to overcome the locks, the value of the property stolen is not influenced by the quality of the locks anymore.

Loss reduction refers to a situation where the insured’s effort influences the distribution of the loss $F(L, e)$ given that the loss occurred but not the loss probability $\pi$. An example might be that precautionary measures against natural disasters like floods or tornados are loss-reducing since currently it is impossible to have an influence on their probability of occurrence. In the literature, among others in Ehrlich and Becker (1972), loss prevention is sometimes referred to as self-protection, while loss reduction is also known as self-insurance.

2.4.1 Loss prevention

In the case of loss prevention the optimal second best insurance contract has a very simple structure: it features a deductible. The insuree receives an indemnification which is smaller than the loss by the amount of the deductible. It is not necessary to provide coinsurance beyond the deductible because it is not in the hands of the insured to influence the distribution of the size of the loss. However, he needs an incentive to employ high effort to avoid the loss which is accomplished by the deductible.

The generalization of this intuition is that the optimal incentive contract will condition only on variables which convey information about the agent’s effort, even if this information is very small. In this context such a variable is said to be a sufficient statistic for the agent’s effort. On the other hand, the optimal contract does not condition on variables which have no informational value about the agent’s effort but are just noise. This result is known as sufficient statistic result (Holmström (1979), Shavell (1979), Holmström (1982)). To some degree real world insurance contracts incorporate this feature via the negligence clause.

\footnote{Technically, the dichotomy of having $\pi$ and $F(L)$ could also be condensed into a single loss distribution function.}
2.4.2 Loss reduction

The case of loss reduction is technically more difficult to handle and the result is less clear cut. The first thing to note is that the straightforward way of modelling the loss as being a deterministic function of effort is not helpful. By doing so the insurer could find out the effort indirectly by observing the loss which would make effort contractible again. Then the first best effort level can be implemented with a contract which pays no indemnification if the sum of the loss is such that it is clear that the insured employed less than the first best effort. Therefore the loss function needs to establish a stochastic relationship between effort and loss: A larger effort level lowers the expected loss but the support $[L, \overline{L}]$ must remain the same. If there were loss levels which occur only under the low effort the insured could be punished very hard if the insurer observes such a loss. Then again the first best would be attainable.

Using the first order approach the optimization problem for the loss reduction case becomes:

$$\max_{e, P, I} E[U] = (1 - \pi)U(W - P) + \pi \int_{L}^{\overline{L}} U(W - L + I^n(L))f(L, e)dL - c(e)$$

s.t. P.C.: $(1 - \pi)P - \pi \int_{L}^{\overline{L}} I^n(L)f(L, e)dL \geq 0$

I.C.: $\pi \int_{L}^{\overline{L}} U(W - L + I^n(L))f_e(L, e)dL - c'(e) = 0$.

The first order condition with respect to $I^n(L)$ is

$$\frac{1}{U'(W + T(L))} = \lambda^{-1} + \frac{\mu f_e(L, e)}{\lambda f(L, e)} ,$$

where $T(L) = -L + I^n(L)$ can be seen as a transfer representing the effective monetary impact of the loss on the insured’s wealth. The fraction $\frac{f_e(L, e)}{f(L, e)}$ is the differential form of the so-called likelihood ratio. If the distribution function satisfies the monotone likelihood ratio property (MLRP), the likelihood ratio is a decreasing function in the absolute size of the loss $L$. That is, a higher loss is more likely to have occurred under a lower effort. This can be seen by noting

---

6However, if the agent is subject to limited liability it might be the case that he cannot be punished hard enough. Further implications of limited liability are analyzed by Innes (1990).
that MLRP implies that $\frac{f(L,e_2)}{f(L,e_1)}$ is a decreasing function in $L$ for $e_2 > e_1$. When MLRP holds, the first order condition above states that the transfer $T(L)$ will be smaller the larger the loss: $T'(L) < 0$.

For the first order approach to be valid it is necessary to make sure that the problem is concave in effort. In order to verify this we restate the objective function by integrating it by parts and subsequently differentiate twice with respect to effort, what yields the following condition:

$$-\pi \int_L^T U'(W + T(L)) T'(L) F_{ee}(L, e) \, dL - c'' < 0.$$ 

This shows that MLRP together with $F_{ee}(L, e) < 0$, which is called concavity of the distribution function condition (CDFC), is sufficient for the first order approach to be valid. There has been a long debate in the literature under which conditions the first order approach is well defined (see e.g. Mirrlees (1975), Rogerson (1985) and Jewitt (1988)).

Now we turn to the contract resulting from the optimization problem: It may have the unattractive feature that $I''(L)$ is a decreasing function in $L$. The nature of insurance business, in particular the incentive to report losses and to inflict damages impose further restrictions on the contract: the indemnity should not exceed the loss ($I \leq L$) and should also be increasing ($I'(L) \geq 0$). Under these additional assumptions the optimal contract with moral hazard on loss reduction exhibits possibly full coverage for low loss levels and partial insurance for higher losses with a non-decreasing indemnity schedule.

Generically, optimal second best contracts implement an incentive scheme which is not necessarily linear in payoff, although linear contracts are often observed in practice (e.g. share cropping). Holmström and Milgrom (1987) develop a model in which the agent controls the drift rate of a Brownian motion in a continuous time model. Under the assumption that the agent’s utility function exhibits constant absolute risk aversion the optimal contract is indeed linear.

\footnote{In the non insurance specific but standard principal-agent model the condition is convexity of the distribution function, see e.g. Rogerson (1985).}
2.5 Exclusive Contracts

As we have seen above the solution to the moral hazard problem is to provide partial insurance only. However, insurees might buy insurance contracts from several insurers in order to increase their level of cover. Such non-exclusive contracts are a problem because they undermine the insurees’ incentive to employ a high effort and inflict losses upon insurers who relied on a low loss probability. In practice insurers might try to enforce exclusive contracts by requiring the presentation of original documents or by exchanging data. An analysis of the problems due to non-exclusive contracts is by Bisin and Guaitoli (1998).

2.6 Many periods

In the context of multi-period contracts the question whether a long-term contract between the insurer and the insured can mitigate the moral hazard problem has been extensively discussed in the literature. A first intuition would be that under multi-period contracts moral hazard is less of a problem. If a loss occurs the insurer can punish the careless insured in later periods by demanding a higher premium (“experience rating”). Thereby bonus-malus contracts which can be observed in practice could be explained. By the argument of income smoothing, the insured would be better off as the incentive to employ high effort is spread over several periods. However, as pointed out by Winter (2000), such a long-term contract is not necessarily better than a single-period contract. To establish this result, the following assumptions are necessary: a finite number of periods, a constant risk aversion (no income effects in the demand for insurance) and the absence of informative signals for the insurer. Then punishing the insured in subsequent periods for a current loss is equivalent to reducing his coverage under the current single-period contract by the corresponding present value. Therefore, there is no difference for the insured between the multi-period and the single-period contract, if the same interest rate is applicable for him as for the insurer. In this case there is no change in the incentive structure which would potentially reduce the moral hazard problem. Repeated or long-term contracts only have an influence on moral hazard under relative stringent assumptions like an infi-
nite number of periods or because the insuree can save money only through the insurer.\textsuperscript{8}

A model in which experience rating actually does eliminate the inefficiency of moral hazard is by Rubinstein and Yaari (1983). However, their result is driven by the assumption that there are infinite periods. The authors show that the social optimal level of care can be elicited by a suitable “no-claims discounts” (NCD) strategy. Under such a strategy the insurer will charge a low premium as long as the average size of the claims the insuree filed so far is consistent with the average expected size of claims under the desired effort level. Otherwise the insured will be punished with a high premium.

### 2.7 Renegotiation

As we have seen above, insurance contracts involve partial insurance or deductibles in order to provide an incentive for the insured to exert a high effort level. However, if there exists a stage when the insured has chosen his effort irrevocably, he does not need incentives anymore and might propose to the insurer to change his contract to full insurance. This raises the issue of renegotiation. For example, imagine the builder of a space-shuttle whose insurance contract features a deductible to make him take care about the quality of the shuttle to be built and the crew to be hired. When the shuttle is finished and on its mission in space the builder wants to renegotiate the insurance contract to full insurance. Should the insurer agree? The argument for full insurance seems convincing because from this stage onwards there is no moral hazard problem anymore which would justify the deductible. However, the problem is that the insuree may have anticipated to eventually get full insurance which would make him use less than the desired effort from the outset.

The renegotiation problem is considered by several authors. Fudenberg and Tirole (1990) discuss a model in which the insurer makes a new take-it-or-leave-it offer to the insuree when it comes to renegotiating. Their result is that the insured chooses a mixed strategy over his effort if he employs any other than

\textsuperscript{8}Malcomson and Spinnewyn (1988) as well as Fudenberg et al. (1990) analyze under which circumstances long-term contracts provide an efficiency advantage over repeated short-term contracts.
the lowest effort level. This is because if the insuree would use an effort level above the minimum effort with certainty the insurer indeed should provide full insurance for a premium corresponding to that certain effort level. But then the insuree would anticipate getting full insurance and employ the minimum effort. Further aspects of renegotiation are analyzed by Ma (1994) and Hermalin and Katz (1991).
3 Oligopoly in Insurance Markets

Although empirical evidence on oligopolistic behavior in some sectors of the insurance industry exists, relatively few theoretical models have been developed. We discuss the cause for this discrepancy, elaborate on the classical oligopoly models and their applicability to the insurance market, and introduce the few models which are specific to the insurance industry. The underlying mechanism in these models is that undercutting a rival makes the pool of risks unattractive and thus oligopolistic market power can be sustained.

3.1 General remarks

An oligopoly is a situation in which relatively few firms compete in a market for a given product. The key characteristic of an oligopoly is that firms are not passively taking the market price as given like under perfect competition. Instead their own actions influence the overall market outcome which leads to strategic behavior in competition. In such a situation it is possible that firms enjoy positive (above normal)\(^9\) profits. The focus of this article is on how to explain positive profit equilibria in insurance markets.

Empirical evidence seems to suggest that at least some insurance sectors can be characterized as having an oligopolistic market structure. With respect to the US market, Nissan and Caveny (2001) find that some lines of property and liability insurance are significantly more concentrated than a comparable collection of other industries (data from 1996 and 1998). In the UK the largest ten property insurers have a market share of 85\(^{10}\). An empirical analysis of the Australian general insurance industry by Murat et al. (2002) suggests that competition is less than perfect and thus insurers do command some extent of market power. Their reasoning is that under perfect competition firms should be forced to completely pass on any increase in production costs (e.g. wage increases) to the consumers as firms make zero profits and thus have no leeway. In their study, however, they

---

\(^9\)We use the term “profit” not in an accounting but in a microeconomic sense, which includes the opportunity costs of equity capital. Therefore a microeconomic “zero profit” can still be consistent with a positive accounting profit. In this context positive profits as opposed to a zero profit are called “above normal”.

\(^{10}\)Source: The Association of British Insurers, www.abi.org.uk
find that insurance companies do not pass on the whole cost increase. A recent study by Chiappori et al. (2005) suggests, that market power is probably present in the French car insurance market.

Interestingly, theoretical work on positive profit oligopoly models in insurance markets is very rare (with a few exceptions discussed below). The traditional theoretical literature assumes that either a monopoly insurer is present or that there is a situation equivalent to perfect competition. So the empirical and theoretical view do not seem to fit well together but rather create a kind of a puzzle. In the following we will explore the reasons for this discrepancy and will outline possible avenues where economic theory in the context of oligopolies and strategic competition can be applied to understand oligopolistic behavior in the insurance sector.

3.2 Factors favoring oligopolies: Standard industrial organization models

In an insurance context Bertrand\textsuperscript{11} competition is the most plausible mechanism because competition takes place in prices (i.e. premiums) and insurers cannot produce coverage in advance. Bertrand competition yields the same outcome as perfect competition, namely zero profits, even if there are only two firms. This is sometimes referred to as Bertrand paradox. As a result, there is no space left between the two extremes of of a monopoly (one insurer) and perfect competition (two or more insurers). Therefore there are virtually no theoretical models which consider oligopoly insurers making above normal profits and most of the classical insurance articles use Bertrand competition.\textsuperscript{12} However, the standard literature on industrial organization discusses several ways to get around the Bertrand conclusion, which will be introduced below.

3.2.1 Capacity constraints

One reason why the Bertrand paradox might not hold true is the presence of capacity constraints, which has been modelled by Edgeworth (1897). If a firm

\textsuperscript{11}For a detailed analysis of Bertrand competition, see Tirole (1988).

\textsuperscript{12}See, for example, the seminal paper on adverse selection by Rothschild and Stiglitz (1976).
cannot serve all the customers it could attract by undercutting its rival, the incentive to do so attenuates. The outcome is that in Nash equilibrium firms charge a price above marginal costs. A similar result obtains in the presence of convex marginal costs. When producing further units drives up costs more than proportional this creates a kind of “soft” capacity constraint the firm runs into. This might play a role in an insurance context as Cummins and Zi (1998) find that more than 30% of US insurers - mostly large companies - operate in an area of decreasing returns to scale.

Capacity constraints seem to be a relevant factor in the insurance industry as well (Cummins et al. (2002)). As explained in Inderst and Wambach (2001), these constraints might arise because the sales force is in the short run limited, capacity to proceed claims and orders is limited, and, above all, the necessary equity to back up insurance contracts cannot be increased easily.

In Edgeworth’s model capacity constraints have been exogenous. In a fundamental article Kreps and Scheinkman (1983) describe a situation in which capacities are determined endogenously. They model a two stage game in which firms simultaneously choose their capacity in the first stage and in the second stage set their price in a Bertrand game. Under efficient rationing\textsuperscript{13} the game exactly yields the Cournot (1838) outcome\textsuperscript{14} which is characterized by positive equilibrium profits.\textsuperscript{15}

\subsection*{3.2.2 Product differentiation}

One means for a firm to get away from perfect competition is to have a different product than its rivals. Thus undercutting the price of a competitor will not attract all the customers due to their different tastes. In the case of horizontal product differentiation different characteristics of the product have an ambiguous link to the consumers’ preferences: For example, some people like blue cars better

\textsuperscript{13}“Efficient rationing” describes a way how customers partition to different insurers if there is overdemand for the policies of one insurer. In this context overdemand means that an insurer has more potential customers than he can actually serve.

\textsuperscript{14}For a detailed contemporary description of Cournot competition see, for example, Shy (1996) or Tirole (1988).

\textsuperscript{15}As for the general result, care has to be taken. Davidson and Deneckere (1986) show that for virtually any other rationing rule this result does not hold true.
than red cars and vice versa. With respect to insurance an example might be life insurance policies with a fair premium but different maturities.

In the case of vertical product differentiation different product characteristics have a clear cut link to the consumers’ preferences: All consumers would rather prefer to have insurance with a lower deductible if the premium was the same. However, not all of them are willing to actually pay the associated higher premium.

**Horizontal product differentiation**

The classic model of horizontal product differentiation is by Hotelling (1929). Consider a situation where two shops which are placed some distance from each other compete for customers. It is assumed that both vendors sell the same product and that customers are distributed on a straight line between the two shops. Now the two shops compete in prices. The product differentiation stems from the fact that customers have transportation costs which depend on the distance they have to walk to the respective shop. It turns out that a higher degree of product differentiation (higher transportation costs) leads to higher prices.\(^{16}\) In a more elaborate game, the shop owners decide at an initial stage where to locate their shops. In this two stage game d’Aspremont et al. (1979) show that there will be maximum differentiation. If the customers are approximately uniformly distributed, the shop owners will build their stores as far apart as possible and both will make positive profits.\(^{17}\) With the assumption that all insurance policies are the same the argument of horizontal product differentiation easily extends to the location of insurance branch offices.

**Vertical product differentiation**

A model of vertical product differentiation is given by Shaked and Sutton (1982). A high quality firm and a low quality firm compete in prices for customers whose

---

\(^{16}\)One can also think of the distance between the shops being an interval representing a sorted array of consumers with different tastes. Then transportation costs correspond to disutility consumers suffer when they do not exactly get the product matching their taste i.e. their position on the interval.

\(^{17}\)Salop (1979) considers in contrast to Hotelling a circular arrangement of customers. This facilitates the extension of the model to a situation with more than two firms. It turns out that in a situation where market entry is free but connected with fixed costs in equilibrium only a limited number of firms are in the market which charge markup prices.
willingness to pay more for the high quality product differs. The outcome of the model is that under certain assumptions both firms make positive profits with the high quality firm charging a higher price. An interesting extension is that even if entry costs converge to zero only a finite number of firms will enter the market. That is because price competition of high quality firms will crowd out low quality firms.

For an insurance company there are quite a number of ways of how to make their products better. Schlesinger and von der Schulenburg (1991) mention the role of perceived service quality in this context. Examples are quick and easy payment of indemnities, the availability of a local agent or bonus/malus systems to adjust the premium. Product differentiation is not limited to the very product itself but may also involve its distribution and marketing approach. Here the range spans from high price full service insurers to discount direct insurers.

### 3.2.3 Search and switching costs

In a well-known article Diamond (1971) proposes a model with consumer search costs with a surprising result: Already minimal search costs yield an equilibrium outcome in which no consumer searches and firms charge the monopoly price. The idea behind this model is that if every consumer anticipates that all firms charge the monopoly price, it does not make sense for the consumer to start a costly search for lower prices. If the firms anticipate that there will be no search for lower prices anyway, it does not make sense for them to reduce their prices in order to attract more customers. Thus the monopoly price can be a stable equilibrium outcome.

Nowadays one can easily surf the internet for insurance comparison portals which claim to find the best offer for the consumer. With respect to life insurance, Brown and Goolsbee (2002) find empirical evidence that such insurance portals increased competition significantly and reduced prices by eight to fifteen percent.

A similar impact is created by switching costs which represent any inconvenience an insuree may have in order to cancel a police and switch to a different insurer. Here one can think of the opportunity costs of time and the disutility of tedious work like having to learn terms and procedures of cancelling, writing
letters or filling out forms.\textsuperscript{18}

Schlesinger and von der Schulenburg (1991) create a combined model which considers Bertrand competition of insurers in the presence of search and switching costs in addition to product differentiation. Incumbent insurers are located on a Salop-like circular street which represents consumer tastes regarding product differentiation. Now new insurance firms enter the market. It is shown that search and switching costs provide market power to the existing insurers and reduce the market share of new entrants.

3.2.4 Barriers to entry and regulation

When a market is split up among members of a profit-making oligopoly one should expect that new firms are interested in entering this market to gain a share in these profits. However, as Bain (1956) pointed out, there might be a number of barriers which hamper market entry. Examples are technological barriers, threats by the incumbent firms (e.g. price wars) and government created barriers like license requirements. Barriers to entry have been analyzed in detail by Encaoua et al. (1986) and von Weizsäcker (1980).

In the insurance sector government regulation plays an important role and can be considered as a barrier to entry. For example before the EU-wide insurance deregulation in 1994 German insurers needed to get their insurance products and premiums approved by a regulating authority. Rees and Kessner (1999) argue that the downside of such a high degree of regulation is that insurers with inefficient high costs are protected from competition instead of being driven out of the market which leads to a loss of insurance buyers’ welfare. This downside of a government protected quasi cartel was taken into account for the sake of consumer protection against insurance insolvency.\textsuperscript{19}

The analysis of barriers to entry provides a prominent role for sunk costs. Usually a potential new entrant is required to invest certain entry costs (e.g. specific equipment) which can be regarded to be sunk (at least partly) and need to be recovered by future profits. As Stiglitz (1987) pointed out, already small

\textsuperscript{18}For the influence of switching costs on market structure see also Klemperer (1987) and Farrel et. al (1988).

\textsuperscript{19}For a critical evaluation, see e.g. Rees et al. (1999).
entry costs may establish a barrier to entry: As long as there is no entrant in the market the incumbent earns the monopoly profit minus his own sunk entry costs. However, should a rival come into the market both firms will end up with a loss due to Bertrand price competition and sunk entry costs, which is not attractive for a potential new entrant. An example of sunk entry cost in the insurance sector might be expenditures on risk research and actuarial expertise. Following Sutton (1991), the analysis of sunk costs may also provide valuable insights on how to explain an industry’s market concentration level.

3.3 Factors favoring oligopolies: Insurance specific models

While the models presented so far are standard in the industrial organization literature and apply potentially to many industries, this section deals with models which employ features which are specific to the insurance industry. The main idea behind all models is that charging a lower price (and thus undercutting a rival) might be undesirable, because this will change the risk portfolio in a detrimental way. Thus prices higher than the competitive price can be sustained. Although the main effect is similar among the models we present, the reason why the risk portfolio becomes unattractive differs.

3.3.1 Adverse selection when consumers differ in risk and wealth

In the classic Rothschild-Stiglitz model (1976) risk is the only dimension of asymmetric information between the insurers and the consumers. This is represented by two types of consumers: a high risk type and a low risk type. In the case of an existing equilibrium insurers offer two separating contracts, one for the high risk type and one for the low risk type. High risk types have no incentive to mimic low risk types because the equilibrium separating contracts are designed in a way that high risks are indifferent between their own and the low risks’ contract. Furthermore the equilibrium contracts are making zero profits due to Bertrand-style price competition among insurers. Thus the original Rothschild-Stiglitz model

---

20 In the model by Rothschild and Stiglitz (1976) an equilibrium may not exist for reasons which are beyond the scope of this overview article.
cannot explain profit making insurance oligopolies.

Several authors (Smart (2000), Villeneuve (2002), Wambach (2000)) extend this model and introduce a second dimension of asymmetric information by considering the case in which consumers can either have low or high wealth (or high or low risk aversion respectively) in addition to having different risks. As risk aversion differs with the wealth a person has, his demand for insurance is not only influenced by his risk type, but also by his wealth level. Now equilibrium situations are possible in which insurers make profits even when there is Bertrand competition and free market entry. The mechanics behind this result is that undercutting the premium of the profit making equilibrium contract will lead to an expected loss. This is because lowering the premium will not exclusively attract the consumer types which generate the expected profit away from the competitors, but mainly high risk types which generate an expected loss.

3.3.2 Competition under moral hazard

Bennardo and Chiappori (2003) argue that moral hazard might be one reason why even under price competition the insurers are able to sustain positive profits. The argument goes as follows: In a situation with moral hazard, insurance contracts must specify a deductible (or a coinsurance rate) as otherwise the insurees will not exert effort to prevent the damage (or to lower the size of the damage). Thus insurers are unable to increase the payments to the insurees in case of a loss, as this will lower effort. The other way insurers can compete away their profits is by reducing the premium charged for the contract. Bennardo and Chiappori show that if marginal costs of effort increase in the wealth of a person, this reduction in premium might also lead to a reduction in effort undertaken by the insuree, thus increasing the expected loss. In such a situation an insurer might make positive profits as any change in the contract structure induces the agent to exert less effort.

3.3.3 Competition between a mutual and standard insurers

Fagart et al. (2003) deal with market interactions between standard insurance companies and a mutual. In a mutual the insurees are also the owners of the
company. Their premia are adjusted ex post to balance premium revenue and indemnity expenses which leaves the mutual with a zero profit. The authors show that for a mutual it is always in the interest of the insurees to have more members. The authors refer to this fact as a positive network effect. This, however, does not hold true for a standard insurer, which is owned by its shareholders. Here the network effect may be positive or negative, depending on the amount of capital funds. The reason is that in case of a bankruptcy the remaining equity will have to be split up among more insurees and thus everybody gets less if more people are insured.

Profit making oligopolies can now be sustained. If one standard insurer undercuts his rivals, he might not attract additional customers. Although the insurer is now cheaper than his competitors, switching to this insurer might be unattractive for the insurees as this leads to an increase in the number of insured which, as explained above, makes the insurance contract less valuable.

### 3.3.4 Risk averse insurers

Polborn (1998) develops an insurance model along the lines of Wambach (1999). He considers two risk averse insurers which are engaged in Bertrand competition. In equilibrium both insurers charge a price above marginal costs and enjoy positive profits. This is because undercutting the rival would imply taking over the risks of the whole market. Thus the insurers face a trade off between profits and risk which is exactly balanced in equilibrium. The critical assumption is whether insurers are risk averse or not. Usually it is argued that shareholder value maximizing companies should not be concerned with their risk because risk averse shareholders can diversify their portfolio themselves. However, several arguments have been brought forward why companies might behave as if they were risk averse (see e.g. Cummins et al. (1998)).
4 The Role of Patience in an Insurance Market

4.1 Introduction

Empirically, personal discount rates vary to a significant degree among people. For example, Warner and Pleeter (2001) use data from the US military downsizing program of the early 1990’s to estimate the discount rates of separatees who could choose between an annuity and a lump-sum payment. Their estimates of discount rates range from 0 to over 30%. Frederick et al. (2002) survey articles that try to estimate the annual discount rate of individuals. Across and within the various studies there is a tremendous variance in results which take values from zero to infinity, with some results even being negative. These findings seem to underline the relevance of accounting for different time preferences among consumers. However, this issue has to our knowledge not received attention so far in the context of insurance markets.

Since the typical insurance contract requires insurees to pay the premium up-front for several periods (e.g. months), there exists a role for time preference in the consumers’ ex-ante valuation of the contract. The aim of this article is to analyze the effects of differences in the personal discount rate of individuals in a competitive insurance market. For this purpose we employ a two-period model with both moral hazard and adverse selection in the spirit of de Meza and Webb (2001). However, it is assumed that the informational asymmetry is not with regard to risk aversion but the individuals’ personal discount rate, which can either be high (impatient) or low (patient). This corresponds to a low (impatient) or high (patient) discount factor. The discount factor will be used for modelling purposes throughout this chapter.

We show that there exists a separating equilibrium in which patient consumers use high effort and buy a profit-making insurance contract. In contrast, impatient consumers use low effort and buy a contract with lower cover than the patient consumers or even prefer to remain uninsured. In this sense, accounting for differences in the personal discount rate helps to explain both positive profits and the opposite of adverse selection (“advantageous” selection as it is called by
de Meza and Webb).

This finding contrasts with traditional models of adverse selection in insurance, for example as represented by Rothschild and Stiglitz (1976), which usually suggest that there should be zero profits for insurers in a competitive insurance market. This result is based on the assumption that there is Bertrand competition in prices, i.e. premiums. Furthermore, Rothschild and Stiglitz predict that it is the high risks who are more keen on buying insurance. However, there is some evidence which does not seem to fit these predictions. In an empirical study of the Australian general insurance industry, Murat et al. (2002) find that insurers are able to sustain a considerable amount of market power. This result is reached since insurers do not completely pass on increases in their production costs to consumers. However, this would be expected from a perfectly competitive industry, as it has no leeway to bear part of the cost increases itself. A further hint that market power may exist at least in some insurance sectors comes from Nissan and Caveny (2001) who find that some lines of property and liability insurance in the US are significantly more concentrated than a comparable collection of other industries.

Furthermore, the effect that high risks want to buy more insurance cannot be readily verified in practice either. Sometimes even the opposite seems to be true. For example, Dionne et al. (2001) criticize an empirical study by Puelz and Snow (1994) which finds that adverse selection is a relevant problem for automobile insurance. They show that under refined estimation methods the result cannot be confirmed. Cawley and Philipson (1999) analyze whether data from life insurance is consistent with the adverse selection hypothesis. They report that in several regards the data exhibits the opposite of the expected pattern. For example, there is a negative covariance between risk and quantity. This suggests that it is actually the low risks who are inclined to buy more insurance.

We also explore to what extent random contracts and differentiated contracts may allow insurers to compete further for profitable patient customers such that profits are reduced or even eliminated. A compensated randomized contract potentially increases the expected utility of the patient types while leaving the
utility of the impatient types constant. If this worked out, a better separation of types could be achieved. We show that ex-ante randomization, where the result of the lottery is revealed before the choice of effort, can indeed achieve a perfect separation of types. Therefore, premiums for patient customers are driven down by competition until profits disappear. The downside of ex-ante randomization is that in this model it is equivalent to the quite unrealistic idea of throwing the dice in the insurance agent’s office in order to determine which contract the insuree will be offered.

A more realistic way of thinking about randomization is, when the result of the lottery is revealed only after the choice of effort. Such randomness might be achieved in practice by linking the indemnity payment to criteria which are out of the insuree’s control, e.g. the precise circumstances of an accident. We show that ex-post randomization can never make the patient types better off in comparison to the deterministic contract. Therefore, profits can persist even if the possibility of ex-post randomization exists.

A different way of how insurers can further compete for profitable patient insurees is to make use of the consumers’ different time preferences. This can be done by differentiating the insurance product into a relatively expensive full service policy and a cheaper discount policy. We assume that the different service level is defined by the time it takes until the indemnity is paid out in case the customer reports a loss. The impatient types are inclined to buy full service insurance whereas the patient types do not mind the delayed payment of indemnities under discount insurance that much. Competition in discount policies may cause profits earned with patient types to disappear completely. However, profits can still persist at least to some extent, depending on the shape of the incentive compatibility constraint.

This article draws on de Meza and Webb (2001) who develop a model which can explain the opposite of adverse selection in a competitive insurance market. This follows from adverse selection on risk aversion in the simultaneous presence of moral hazard. Consumers are split into timid and bold people, whereby the bold ones are less risk averse. They can choose between a high or low level of unobservable precautionary effort in order to avoid the loss. De Meza and Webb
show that for certain parameter values a unique separating equilibrium exists in which timid individuals buy insurance and employ high effort whereas bold ones remain uninsured and employ low effort. In contrast to de Meza and Webb, adverse selection in the present model occurs regarding the consumers’ personal time preferences.

Apart from standard industrial organization explanations, there is already some related insurance specific literature which can explain positive profits even in a competitive insurance market (see chapter 3 of this thesis).

4.2 The Model

The model we employ in this chapter consists of a game with 4 stages throughout 2 periods:

1) Insurance companies make irrevocable offers of contracts that specify both premium $P$ and indemnity $I$.

2) Clients buy at most one contract from one insurance company. If two or more insurers offer the same optimal contract, clients randomize with equal probability. When buying insurance, a client has to pay the premium $P$ up-front.

3) The consumer decides which unobservable effort level $e$ to choose in order to avoid the loss.

4) The loss occurs or not and the indemnity is paid out in case of a loss.

Stages 1 to 3 take place in period 1, whereas stage 4 takes place in period 2.

4.2.1 Insurers

There are two or more risk neutral insurers in the market who compete in contracts. There is informational asymmetry regarding the patience type of consumers: Insurers know the distribution of patience types in society, but they cannot identify the patience type of an individual who wants to buy a contract. Furthermore, the consumers’ utility function, the two available effort levels and the resulting loss probabilities are also known by the insurers. This enables them
to conjecture the effort level employed by insurees of each type under any contract correctly, even though effort is unobservable.

4.2.2 Consumers

It is assumed that the consumers’ ex-ante expected utility $EU$ can be described by

$$EU = U(w - P) + \delta \left[ (1 - s(e))U(w) + s(e)U(w - L + I) \right] - c(e)$$

with $U$ being a concave, time-additive utility function with exogenously given risk aversion. Premium is denoted by $P$ and indemnity by $I$. Furthermore, there is period income $w$, personal discount factor $\delta$, effort $e$ with $e \in \{e_l, e_h\}$, loss probability $s(e)$ with $s(e_h) < s(e_l)$, costs of effort $c(e)$ with $c(e_l) < c(e_h)$ and loss $L$. To simplify notation we define $s_n := s(e_n)$ and $c_n := c(e_n)$ for $n \in \{h, l\}$.

The level of effort is chosen and paid for in the first period, whereas the benefit of the effort is in effect in the future. This way of modelling effort is especially adequate when it is a technical necessity to make precautionary provisions before or at the very beginning of the insurance contract. For example, when building a house the agent can decide whether or not to use fire-retardant materials. This investment will carry on its beneficial effect over the policy period of a fire insurance. Another example is travel health insurance. Here the insuree can decide whether or not to get vaccinated before departure. Sometimes an audit conducted by the insurer in case a loss is reported allows to infer the effort level employed to a certain extent. However, a perfect inference seems to be unlikely in many cases. Then the two effort levels in this model can be interpreted as the residual consumer’s discretion with regard to effort which cannot be detected in an audit anymore.

The population shall be split up in a fraction $\gamma$ of patient people with a high discount factor $\delta_p$ and a fraction $1 - \gamma$ of impatient people with a low discount factor $\delta_i$.

Via the implicit function theorem, the slope of the consumers’ indifference curves $IC$ in the premium-indemnity space (figure 4.1, page 39) can be verified
to be
\[ S(\delta, e, P, I) := \frac{dP}{dI} \bigg|_{e} = \frac{\delta s(e)U'(w-L+I)}{U'(w-P)} > 0. \] (4.1)

For extremely low patience with \( \delta = 0 \), the indifference curve is a flat horizontal line. With increasing \( \delta \), the indifference curve becomes steeper at every given point in the premium-indemnity space (figure 4.1).

Given the same effort level \( e \), the indifference curve of a patient type \( IC_p \) is steeper than the indifference curve of an impatient type \( IC_i \) for any contract \( \{P, I\} \). This can be seen by computing the derivative of (4.1) with respect to \( \delta \):
\[ \frac{\partial S(\delta, e, P, I)}{\partial \delta} \bigg|_{e} = \frac{s(e)U'(w-L+I)}{U'(w-P)} > 0. \]

Furthermore, consumers’ indifference curves are concave:
\[ \frac{dS(\delta, e, P, I)}{dI} = \frac{\delta s(e)U''(w-L+I)}{U'(w-P)} < 0. \]

### 4.3 Effort border

Depending on parameter values, there may be a border line in the premium-indemnity space which describes when a certain patience type is exactly indifferent between high and low effort. To determine this effort border \( EB \) we compute a consumer’s advantage in expected utility from employing high effort:
\[ A(\delta) := EU_h - EU_l = \delta[s_l - s_h] \left[ U(w) - U(w-L+I) \right] - (c_h - c_l) \]

If \( A(\delta) \) is positive, it pays to employ high effort, whereas otherwise the individual is better off employing low effort.

In order to obtain a condition under which the individual is indifferent between high and low effort, we set \( A(\delta) \) equal to zero and obtain:
\[ U(w-L+I) = U(w) - \frac{c_h - c_l}{\delta(s_l - s_h)} \] (4.2)

When the individual switches from high effort \( e_h \) to low effort \( e_l \) (e.g. while \( P \) is constant and \( I \) increases marginally), it becomes apparent from (4.1) that the indifference curve exhibits a kink at the effort border, as its slope becomes steeper. The indifference curve of the patient types \( IC_p \) in figure 4.1 gives an example.
According to (4.2), the position of the kink is determined by the following parameter values:

- a larger $\delta$, a larger difference in loss probabilities for low and high effort $(s_l - s_h)$ and a larger loss $L$ move the kink to the right in the premium-indemnity space

- a larger cost difference between high and low effort $(c_h - c_l)$ moves the kink to the left. This is also the case for a larger $w$ if $I < L$, because $U$ is concave.

As the position of the kink does not depend on $P$, the curve of kinks (effort border) is a linear vertical line in the premium-indemnity space. If the circumstances are such that the effort border of a patience type is located at an indemnity level of zero or below, the individuals of this patience type never employ high effort.

### 4.4 Equilibria

In what follows, we restrict our analysis to pure strategy subgame-perfect Nash equilibria.

#### 4.4.1 Definition of Equilibria

Before analyzing equilibria of the model, we outline the definition of separating and pooling equilibria, following de Meza and Webb (2001). The outside option contract (no insurance) is denoted by $O$.

**Separating Equilibrium**

A separating equilibrium is characterized by the following four properties, whereby $C_n^*$ with $n \in \{i, p\}$ stands for the contract chosen by type $n$ in equilibrium.

**Incentive compatibility**

There must be no incentive for an impatient type to buy the contract for a patient type and vice versa:

$$EU_i(C_i^*) \geq EU_i(C_p^*)$$

$$EU_p(C_p^*) \geq EU_p(C_i^*)$$
**Effort incentives**

Since effort is unobservable, the individual employs high effort only if this is advantageous in terms of a higher expected utility:

\[
e = \begin{cases} 
  e_h, & \text{if } A(\delta) \geq 0 \\
  e_l, & \text{if } A(\delta) < 0 
\end{cases}
\]

**Participation**

Consumers cannot be forced to buy insurance, but they insure themselves voluntarily, if this results in a higher expected utility than the outside option of remaining uninsured:

\[
EU_n(C^*_n) \geq EU_n(O) \text{ for } n \in \{i, p\}
\]

**Profit maximization**

By offering \(C^*_i\) and \(C^*_p\) each insurer maximizes his profit, given that \(C^*_i\) and \(C^*_p\) are offered by his competitors. No insurer can earn more by offering a different contract or by offering no contract at all.

**Pooling Equilibrium**

A pooling equilibrium \(C^*\) is characterized by the following three properties:

**Effort incentives**

\[
e = \begin{cases} 
  e_h, & \text{if } A(\delta) \geq 0 \\
  e_l, & \text{if } A(\delta) < 0 
\end{cases}
\]

**Participation**

\[
EU_n(C^*) \geq EU_n(O) \text{ for } n \in \{i, p\}
\]

**Profit maximization**

Given that \(C^*\) is offered by his competitors, each insurer maximizes his profit by offering \(C^*\).

4.4.2 Analysis of Equilibria

The analysis of equilibria is based on the following assumptions:
Assumption 4.1 For simplicity it is assumed that the market interest rate is zero.

This assumption does not change the main insights of this article but merely helps to facilitate the analysis.

Assumption 4.2 Even under high effort \(e_h\), a patient type \((\delta_p)\) has a steeper indifference curve than an impatient type \((\delta_i)\) under low effort \(e_l\) for any contract \(\{P, I\}\).

According to (4.1), this requires that \(\delta_p \, s_h > \delta_i \, s_l\). This condition allows for stable equilibria in the sense that deviations are unattractive which will be pointed out in more detail later on. The typical loss probability in most lines of insurance can be considered relatively small in comparison to the typical discount rate as suggested by the literature cited above. Therefore, the condition does not seem overly restrictive as long as a low level of precautionary effort does not cause \(s_l\) to skyrocket too much. Furthermore, the possibility of an audit may impose a lower bound on \(e_l\). Later on we discuss the situation where this assumption is not satisfied.

The situations which arise from the various parameter constellations can be grouped according to which types employ high effort at least in some area of the contract space. Possible answers are none, only the patient types or both types. Since the case in which no one ever employs high effort is not very interesting, we concentrate on the latter two cases which we denote by A and B respectively.

A) Impatient types do not employ high effort under any contract

In this section we analyze case A, which shall be characterized by the following property:

A.1 The impatient types do never employ high effort but there is a region for the patient types in the contract space where they employ high effort.

This is to say that the effort border of the impatient types \(EB_i\) is not visible in the premium-indemnity space, as it is located at a weakly negative indemnity.
However, the effort border of the patient types $EB_p$ is located at a strictly positive indemnity, as shown in figure 4.1. This property imposes restrictions on the permissible parameter values according to (4.2).

Figure 4.1: Contract space in a two-period competitive insurance market where consumers differ in patience. There is a unique separating equilibrium in which the impatient types employ low effort and buy zero-profit contract $D$ with low indemnity, whereas the patient types employ high effort and buy profit-making contract $E$ with high indemnity.

**Separating Equilibrium**

The resulting equilibria can be seen most easily by proceeding diagrammatically. Figure 4.1 depicts the contract space (premium-indemnity space). The concave lines are indifference curves of the patient types ($IC_p$) and the impatient types ($IC_i$), respectively. At some critical indemnity level, there is the patient types’ effort border $EB_p$. To the left of it they voluntarily employ high effort, as their expected utility is higher than under low effort. To the right of it they employ low effort. Switching to the low effort level makes the loss probability jump from
to \(s_l\). Therefore, the patient types’ indifference curves have a kink at their effort border, because insurance is now more valuable again.

The insurance company’s fair premium lines for high and low effort are straight lines with slope \(s_h\) and \(s_l\), respectively, because it is assumed that the insurer is risk neutral and that the market interest rate is zero. The fair pooling line describes all contracts which yield zero profit for the insurer, if they are bought by the whole population of patient and impatient people.

Regarding equilibrium, we distinguish between two cases. This is because the relevant indifference curve of the impatient types can cut the patient types’ effort border below (figure 4.1) or above (figure 4.3, page 43) the fair pooling line (point \(B\)).

**Proposition 4.1** Consider case A with the impatient types’ discount factor \(\delta_i\) being such that \(IC_i\) through \(D\) (or \(O\), if impatient types prefer being uninsured) (figure 4.1) cuts \(EB_p\) above \(A\) but below \(B\) at some point \(E\), and \(IC_p\) through \(E\) is always below the fair pooling line. If assumptions 4.1 and 4.2 hold, then there exists the following unique separating equilibrium: The patient types employ high effort and buy the profitable contract \(E\). If \(s_l < \frac{\delta_i s_l U'(w-L)}{U'(w)}\), the impatient types employ low effort and buy zero-profit contract \(D\) which has lower cover than contract \(E\). Otherwise, the impatient types remain uninsured.

**Proof** The patient types are strictly better off with contract \(E\) in comparison to being uninsured. The impatient types are indifferent between contract \(E\) and contract \(D\) (or not being insured at all) and thus, as in Rothschild and Stiglitz (1976), it is assumed that they buy \(D\).

Now we challenge contract \(E\) by considering deviations to the left of \(EB_p\). Offering contracts above \(IC_p\) does not make sense. This is because below \(IC_i\) only impatient types are attracted which results in losses. Contracts above \(IC_i\) cannot attract any customers. By offering a contract below \(IC_p\) an insurer could attract all patient types. However, due to assumption 4.2, such a contract would also attract all impatient types. Since the fair pooling line is above the deviating contract, it necessarily leads to losses.

Deviations to the right of \(EB_p\) are also loss-making. This is because in this
region both types employ low effort and the patient types cannot even be attracted away from contract $E$ by offers on the fair premium line under low effort.

The patient types prefer contract $E$ to any other contract which lies on $IC_i$ through $D$ (or $O$, if impatient types prefer to be uninsured) and is on or above the applicable fair premium line. Therefore, there can be no other separating equilibrium. A pooling equilibrium would have to be on or above the fair pooling line. Since $IC_p$ through $E$ is always below the fair pooling line by assumption, there cannot be a pooling equilibrium that would not be destroyed by $D$ (or $O$) and $E$. Uniqueness follows. □

**Situation with assumption 4.2 not being satisfied**

An interesting question to analyze at this point is what happens when assumption 4.2 does not hold. This means that the indifference curve of the impatient type under low effort is steeper than the indifference curve of the patient type under high effort. Suppose that the impatient types do not want to buy insurance even for their fair premium ($IC_i$ does not touch the fair premium line under low effort). Then there are two possibilities with regard to the fair pooling line: It can either cross the indifference curve of the patient types through $O$ ($IC_p$ in figure 4.2, page 42) or not. In the first case, a contract on the fair pooling line dominates $O$, but it can always be destroyed by a profit-making deviating offer which only attracts patient types. In the end, no equilibrium exists. In the second case $O$ is an equilibrium. This shows that consumer heterogeneity with regard to time preference and simultaneous moral hazard may contribute to the fact that some risks are uninsurable. A similar result is obtained by Chiu and Karni (1998) in the context of private unemployment insurance. The authors demonstrate that adverse selection on the employees' preference for leisure together with moral hazard regarding the employees' effort of working hard can explain the absence of private unemployment insurance as an equilibrium outcome.

**Pooling Equilibrium**

**Proposition 4.2** Consider case $A$ with the impatient types' discount factor $\delta_i$ being such that $IC_i$ through $D$ (or $O$, if impatient types prefer being uninsured)
If assumption 4.2 is not satisfied, $IC_i$ through $O$ does not cross the fair premium under low effort and the fair pooling line does not cross $IC_p$, then the equilibrium contract is $O$.

Figure 4.2: If assumption 4.2 is not satisfied, $IC_i$ through $O$ does not cross the fair premium under low effort and the fair pooling line does not cross $IC_p$, then the equilibrium contract is $O$.

(figure 4.3, page 43) cuts $EB_p$ above $B$ but below $C$ and $IC_p$ through $B$ is always below the fair premium line under low effort. Under assumptions 4.1 and 4.2, there exists a unique zero-profit pooling equilibrium in which the impatient types employ low effort, the patient types employ high effort and both types buy contract $B$, if $IC_p$ at point $B$ is steeper than the fair pooling line to the left of $EB_p$ (condition $C1$). Otherwise, there exists no equilibrium in pure strategies, if new insurers can enter the market.

**Proof** Suppose first that condition $C1$ holds. Both types are strictly better off buying contract $B$ instead of remaining uninsured. Now we consider deviating contracts to the left of $EB_p$: Offering a deviating contract above $IC_p$ through $B$ does not make sense as no patient types can be attracted. Offering a profitable deviating contract below $IC_p$ through $B$ is loss-making since, by condition $C1$, such contracts are below the fair pooling line and would also be bought by all
Figure 4.3: Unique zero-profit pooling equilibrium in which both types buy contract $B$. The impatient types employ low effort, whereas the patient types employ high effort.

impatient types (assumption 4.2).

Deviating contracts to the right of $EB_p$ are also loss-making. This is because in this region both types employ low effort and cannot even be attracted away from $B$ by contracts on the fair premium line under low effort.

The patient types prefer contract $B$ to all other contracts on or above the fair pooling line. Thus, there cannot be another pooling equilibrium. It must be that in a candidate for a separating equilibrium the impatient types get contract $D$ (or $O$, if impatient types prefer being uninsured) and are indifferent between their own contract and the contract of the patient types which would be $S$. Due to incentive compatibility, the separating contract for the patient types must be on or above $IC_i$ through $D$ (or $O$). Therefore, both types will prefer contract $B$ to the candidate for a separating contract and $B$ is a unique pooling equilibrium.
Now we consider the situation in which condition \( C1 \) does not hold which is depicted in the box in figure 4.3. In this case, there is a point \( T \) to the left of \( EB_p \) where \( IC_p \) is tangent to the fair pooling line. Due to assumption 4.2, \( T \) is no stable equilibrium, since a profitable deviating contract like \( G \) can be offered. Moreover, \( B \) can also be destabilized by offering a contract like \( F \) which attracts both types but still is above the fair pooling line. This reasoning holds true for any contract apart from \( T \) on and above the fair pooling line. Since \( IC_p \) is below the fair premium line under low effort there cannot be a pooling equilibrium to the right of \( EB_p \) either.

Suppose \( S \) is part of a candidate for a separating equilibrium. All insurers offering \( S \) earn a profit per contract which is represented by the distance \( \overline{AS} \). If there are too many insurers offering \( S \), a deviation slightly to the south-west of \( S \) can be profitable for an individual insurer. This is the case, if the pooling profit earned by serving all consumers is greater than the share in profit by serving only patient consumers. The consequence is that no equilibrium is possible due to the single crossing property (assumption 2) in the area to the left of \( EB_p \). Even if there are not too many insurers offering \( S \), new entrants in the insurance market are attracted and \( S \) is destabilized. Thus, a separating equilibrium cannot exist. In the end, neither a pooling nor a separating equilibrium exists, if condition \( C1 \) does not hold. \( \square \)

**Random contracts**

In this section we analyze in how far random contracts can be used by insurers in order to compete for profitable customers.\(^1\) Along the lines of Arnott and Stiglitz (1988), we can distinguish two cases with regard to timing: The realization of the lottery can be made known to the consumer either before (ex ante) or after (ex post) his choice of effort.

Under ex-ante randomization we can think of the consumer committing to accept a contract which is the outcome of a lottery \( L \) played subsequently at the insurance agent’s office. The resulting contract is immediately made known

---

\(^1\)We thank Bertrand Villeneuve for suggesting this issue. Random contracts within an insurance framework are also discussed by Arnott and Stiglitz (1988), Prescott and Townsend (1984) and Villeneuve (2003).
to the consumer who can then decide which effort level to employ. Consider contract $E$ in figure 4.1 (page 39). As pointed out above, there is no profitable way for an insurer to undercut $E$ with a deterministic contract. However, an insurer could offer a lottery $L = \{(E, X); (\mu, 1 - \mu)\}$ instead of contract $E$. This lottery would imply to offer contract $E$ with probability $\mu$ and contract $X$ with probability $(1 - \mu)$, whereby $X$ is somewhere on $IC_i$ through $E$ to the right of $EB_p$. Contract $E$ is profitable, whereas $X$ is loss-making. Therefore, there is a certain mixing probability $\mu$ for each possible $X$ for which the lottery yields zero profit for insurers. Furthermore, such a lottery $L$ is strictly preferred by a patient consumer in comparison to the deterministic contract $E$, because there is a probability $1 - \mu$ of being better off. An impatient consumer has nothing to gain from such a lottery, since both contracts $E$ and $X$ offer exactly the same utility to him. In this situation, the introduction of random insurance contracts indeed enables insurers to compete further for profitable clients even at point $E$ in figure 4.1, until zero profits are reached.

Under ex-post randomization, the consumer is presented a lottery and commits to buying whatever contract will be the outcome. However, the uncertainty is resolved only after the choice of effort. In the present setting, ex-post randomization is more realistic than ex-ante randomization which is not observed in practice. As pointed out by Villeneuve (2003), ex-post randomization can be achieved by linking the insuree’s reimbursement to criteria which are random themselves. For example, insurers can evaluate the precise circumstances of the loss. If the insuree is found to have been negligent, this allows for reducing the coverage payment accordingly. However, the legal definition of negligence most often will offer quite some room for interpretation. Generally, there are many aspects with regard to the circumstances under which the loss occurred which are out of the insuree’s control, e.g. time or weather.

The question is now whether an ex-post randomized indemnity payment, which is compensated by a higher expected value of the indemnity, can achieve a better separation of types. This may allow to make the patient types better off while leaving the expected utility of the impatient types unchanged. Again, we use contract $D$ for the impatient types and contract $E$ for the patient types in
figure 4.1 as a starting point. Since the impatient types are indifferent between contract $D$ and contract $E$, we can write their expected utility under contract $D$ as

$$EU_i = U(w - P_E) + \delta_i[(1 - s_l)U(w) + s_l U(w - L + I^E + \tilde{\epsilon})] - c_i,$$

where $P_E$ and $I^E$ denote the premium and the indemnity under contract $E$ while the random variable $\tilde{\epsilon}$ with $E(\tilde{\epsilon}) = 0$ represents the randomness of the indemnity. The expected utility of the patient types under contract $E$ is

$$EU_p = U(w - P_E) + \delta_p[(1 - s_h)U(w) + s_h U(w - L + I^E + \tilde{\epsilon})] - c_h$$

$$= U(w - P_E) + \delta_p[(1 - s_l)U(w) + s_l U(w - L + I^E + \tilde{\epsilon})] - c_l.$$

Since $\delta_p > \delta_i$, the patient types suffer even more from the introduction of $\tilde{\epsilon}$. However, if the impatient types are compensated by an additional certain indemnity payment for $\tilde{\epsilon}$ until they have the same expected utility as without the randomization, the expected utility of the patient types will remain unchanged as well. This is because both types have the same utility function and hence the same risk aversion. As in Arnott and Stiglitz (1988), the lottery over indemnities can be replaced by the same certainty equivalent indemnity. In the end, a randomized indemnity payment cannot achieve a better separation of types which is in line with Arnott and Stiglitz’ (1988) proposition 10.

**B) Impatient types employ high effort under some contracts**

**Basic Analysis**

Case B shall be characterized by the following properties B.1 and B.2 which ensure that both types want to buy insurance. This gives rise to the possibility that both patience types can be separated by making use of their different patience and offering differentiated insurance policies to them. We leave this issue to the next section and start with the simple case of no product differentiation.

**B.1 Both the impatient as well as the patient types employ high effort in some region of the contract space.**

The parameters in (4.2) take values for which the indifference curves of both types feature a kink.
B.2 The impatient types are interested in buying insurance at their fair premium under high effort.

This property requires the slope of the impatient types’ indifference curve without insurance \((P = 0 \text{ and } I = 0)\) not to be flatter than the fair premium line under high effort. This is captured by the following condition:

\[
s_h < \frac{\delta_i s_h U''(w - L)}{U'(w)}
\]

Proposition 4.3 Consider case B with the impatient types’ discount factor \(\delta_i\) being such that \(IC_i\) through \(D\) (figure 4.4, page 48) cuts \(EB_p\) above \(A\) but below \(B\) at some point \(E\), and \(IC_p\) through \(E\) is always below the fair pooling line. If assumptions 4.1 and 4.2 hold true, then there exists a unique separating equilibrium in which both types employ high effort. The patient types buy profit-making contract \(E\) and the impatient types buy zero profit contract \(D\).

Proof Both types are strictly better off than without insurance. The impatient types are indifferent between their own contract \(D\) and the patient types’ contract \(E\). Therefore, it is assumed that the impatient buy their own contract \(D\). Deviations from \(D\) are not an issue, as \(D\) maximizes the impatient types’ expected utility subject to a zero profit by construction.

Deviations from \(E\) to the left of \(EB_p\): It does not pay to offer a contract above \(IC_p\) through \(E\). This is because above \(IC_i\) through \(D\) no customer can be attracted. Below it, only impatient types who employ low effort can be attracted, what leads to losses. This also holds true for all points on \(IC_i\) through \(D\), with the only exemption being \(D_i\) since \(IC_i\) does not cross the fair pooling line to the right of \(EB_i\) as we look at the case where \(E\) is below \(B\). However, offering \(D\) instead of \(E\) is not attractive either, because it involves zero profits. Below \(IC_p\) through \(E\) both customer types are attracted due to assumption 4.2. However, since this area is below the fair pooling line, deviating offers below \(IC_p\) through \(E\) cause losses as well.

To the right of \(EB_p\), profitable offers are not possible either. In this region, not even offers on the fair premium line can attract any patient customers away from \(E\).
Figure 4.4: Both types feature a kink in their indifference curves. There is a separating equilibrium in which both types employ high effort. The patient types buy contract $E$ (profit-making) whereas the impatient types buy contract $D$ (zero profit).

The patient types prefer contract $E$ to any other contract which lies on $IC_i$ through $D$ and is on or above the applicable fair premium line. Therefore, another separating equilibrium cannot exist.

A pooling equilibrium would have to be on or above the fair pooling line. Since $IC_p$ through $E$ is always below the fair pooling line, there cannot exist a pooling equilibrium which would not be destroyed by $E$. Uniqueness follows. □

**Full service versus discount insurers**

We use the separating equilibrium of the previous section (figure 4.4) as a starting point and assume now that the insurers can be separated into full service insurers as well as discount insurers with a less attractive service. Service quality shall be represented by the following stylized fact: Full service insurers are characterized by a quick and hassle-free payout of indemnities in case the customer reports
a loss. Discount insurers, in contrast, pay indemnities only with a delay. This might be because they want to save costs by having fewer staff who process claims or by earning interest on withheld indemnities.

An alternative interpretation of this specification could be in terms of exclusive agents versus independent agents. Kim and Smith (1996) suggest that exclusive agents are more reluctant when it comes to paying out indemnities while independent agents provide a better service regarding claims settlement. This is because they can threaten to move their customers to a different insurer if claims are not settled fairly and promptly.

The expected utility of an individual buying full service insurance is unchanged with regard to the situation before and given by

\[ EU^F = U(w - P^F) + \delta \left[ (1 - s(e))U(w) + s(e)U(w - L + I^F) \right] - c(e), \]

where superscript \( F \) stands for full service insurance.

The expected utility of a consumer buying discount insurance is

\[ EU^D = U(w - P^D) + \delta \left[ (1 - s(e))U(w) + s(e)U(w - L + t(\delta)I^D) \right] - c(e), \]

where superscript \( D \) stands for discount insurance. We assume that a discount insurer pays out indemnities with a delay of one period, so the indemnity is paid out in the third period only. In order to facilitate the analysis, we assume that there exists a function \( t(\cdot) \), with \( t(\delta) < \delta \) for \( \delta \in (0, 1) \) and \( t'(\delta) > 0 \), which transforms \( \delta \) in such a way that a consumer is indifferent between receiving \( t(\delta)I^D \) in period 2 and \( I^D \) in period 3.

By comparing \( EU^F \) and \( EU^D \), information about the relative position of indifference curves under full service (\( IC^F \)) and under discount insurance (\( IC^D \)) can be obtained. Subtracting \( EU^D \) from \( EU^F \) while assuming that \( EU^D = EU^F \) yields:

\[ U(w - P^F) - U(w - P^D) + \delta s(e)[U(w - L + I^F) - U(w - L + t(\delta)I^D)] = 0 \]

From this equation it can be seen easily that two indifference curves \( IC^F \) and \( IC^D \) of a particular patience type \( \delta \), representing the same expected utility and the same effort level, intersect at the same \( P \), if \( I^F = I^D = 0 \). If \( I^F = I^D \)
is positive and $\delta < 1$, $IC^D$ is below $IC^F$ for the same effort level. As will be pointed out in the next paragraph, $IC^F$ becomes steeper due to the switch to low effort at a lower indemnity than $IC^D$. Therefore, $IC^D$ is also below $IC^F$ for indemnities for which the effort level is not the same under the two different forms of insurance. A graphical illustration is given in figure 4.5 on page 51. Indifference curves $IC^F_p(E)$ and $IC^D_p(I)$ represent the same expected utility level for patient types, since they intersect at point $H$ where the corresponding indemnities are zero. In the same line of argument, $IC^D_i(J)$ represents the same expected utility level than $IC^F_i(D)$ for impatient types, since both indifference curves intersect at point $G$.

In analogy to (4.2), the effort border under discount insurance is described by

$$U(w - L + t(\delta)I) = U(w) - \frac{c_h - c_l}{\delta(s_l - s_h)}.$$ 

It is apparent that the effort border under discount insurance is located further to the right in the contract space, if $t(\delta) < 1$.

An example of a situation with product differentiation is given in figure 4.5 on page 51. There are four different effort borders $EB^n_m$ and indifference curves $IC^n_m$ where $n \in \{i, p\}$ denotes the patience type and $m \in \{F, D\}$ denotes the insurance type ($F$=full service, $D$=discount). Point $I$ marks the intersection of $IC^D_p(I)$ and $EB^D_p$. Point $J$ denotes the intersection of $IC^D_i(J)$ and $EB^D_i$.

Before proceeding to the analysis of equilibria, we introduce a modification of assumption 4.2:

**Assumption 4.3** Even under high effort $e_h$, a patient type ($\delta_p$) has a steeper indifference curve than an impatient type ($\delta_i$) under low effort $e_l$ for any discount insurance contract $\{P^D, I^D\}$.

In order not to inflate the scope of this article, we want to focus on one case only: It is not possible to make the impatient types better off in comparison to full service contract $D$ with a discount contract that is not loss-making. This requires that $IC^D_i(J)$ (figures 4.5 and 4.6) is always below the applicable fair premium line under discount insurance, which is based on $s_h$ to the left of $EB^D_i$ and $s_l$ to the right of it.
Figure 4.5: Situation with product differentiation in which $I$ is above $F$ and $J$ is below $F$. This allows for a separating zero-profit equilibrium in which both types employ high effort. Impatient types buy full service insurance contract $D$, whereas patient types buy discount insurance contract $F$.

**Definition 4.1** The area between $IC^D_p(I)$ and $IC^D_p(J)$ which is above $s_h$ is defined as the “incentive compatible improvement area for the patient types” (ICIAP).

In this area, the patient types can be made better off in comparison to $E$ with a discount insurance contract without attracting impatient types away from their full service contract $D$. Furthermore, an insurer who offers a contract in ICIAP achieves at least a zero profit, if the patient types employ high effort.

Regarding the following proposition, point $F$ (figures 4.5 and 4.6) is defined to be the intersection of $EB^D_p$ and the fair premium line under high effort.
Figure 4.6: Situation with product differentiation in which I and J are above F. This results in a separating positive-profit equilibrium in which both types employ high effort. Impatient types buy full service insurance contract D (zero profit), whereas patient types buy discount insurance contract J (profitable).

Proposition 4.4 Consider case B in the presence of product differentiation, an \( IC_p^D(I) \) which is always below the fair premium line under low effort, an \( IC_p^D(F) \) which is steeper than \( s_h \) to the left of \( EB_p^D \), an \( IC_i^D(J) \) which is always below the applicable fair premium line and the fair pooling line under discount insurance, and an \( IC_F^F(E) \) which is always below the fair pooling line under full service insurance (figures 4.5 and 4.6). If there is an ICIAP to the left of \( EB_p^D \), discount insurance can attract patient customers away from the profitable full service contract E, and one of the following equilibria arises:

a) If ICIAP contains F, then a unique zero-profit separating equilibrium arises in which both types employ high effort. Impatient types buy full service contract
whereas patient types buy discount insurance contract $F$ (figure 4.5).

b) If ICIAP does not contain $F$, but a point like $J$ below $I$, then a unique separating equilibrium arises in which both types employ high effort. Impatient types buy full service contract $D$, which yields a zero profit, whereas patient types buy discount insurance contract $J$, which yields a positive profit (figure 4.6).

**Proof** If there is an ICIAP area to the left of $EB_{D}^{p}$, then insurers can make the patient types better off in comparison to full service contract $E$. In addition, insurers will achieve at least a zero profit by doing so, since only patient high effort types will be attracted by such discount policies and the ICIAP area is defined to be above the fair premium line under high effort, which is based on $s_h$.

Now we consider part a) of the proposition when ICIAP contains $F$. The consumers’ participation constraint is met, since both contracts $D$ and $F$ make the respective consumer types better off than having no insurance.

As to stability of $D$, it is impossible to attract impatient types away from $D$ with a discount policy which would be profitable. This is because $IC_i^D(J)$ is assumed to be always below the applicable fair premium line under discount insurance and to intersect $EB_{D}^{p}$ at some point $J$ which is below $F$. Therefore, competition between insurers offering a discount policy to the patient types is not limited by the incentive compatibility constraint, as insurers run into zero profits first. Moreover, it is impossible to attract impatient types away from $D$ by offering any other full service contract which would not be loss-making. This is because $D$ is maximizing the impatient types’ expected utility subject to a zero profit by construction.

Also contract $F$ is stable. It is assumed that $IC_p^D(F)$ is steeper than $s_h$ to the left of $EB_{D}^{p}$. Moreover, it follows from the assumptions that it is below the fair premium line under low effort to the right of $EB_{D}^{p}$. Therefore, the best discount contract for the patient types which does not inflict losses upon insurers is the corner solution at point $F$. This contract makes the patient types employ high effort and involves a zero profit for insurers. Deviations from $F$ offering discount contracts are not an issue. In order to attract patient types away from $F$ they would have to be loss-making.

As to uniqueness, the impatient types prefer full service contract $D$ to any
other full service contract which lies on or above the applicable fair premium line. Offering discount insurance to impatient customers which would provide the same level of expected utility to them as \( D \) is impossible, since \( IC_i^D(J) \) is assumed to be always below the applicable fair premium line. The patient types prefer discount insurance contract \( F \) to any other full or discount insurance contract which is on or above the applicable fair premium line. Therefore, there can be no other separating equilibrium than \( \{ D, F \} \).

A discount pooling equilibrium would have to be on or above the fair pooling line under discount insurance. However, \( IC_i^D(J) \) is assumed to be always below the fair pooling line under discount insurance. Therefore, the impatient types would not want to buy such a pooling contract. Since \( IC_p^F(E) \) is assumed to be always below the fair pooling line under full service insurance, there can be no full service pooling equilibrium either, because it would be destroyed by \( E \). Uniqueness follows.

With regard to part b) of the proposition, the participation constraint of the consumers is met as well, because both patience types are better off than without insurance.

Showing the stability of \( D \) follows the same line of reasoning as under part a). However, this time competition between insurers offering discount insurance to patient types is limited by the incentive compatibility constraint, which is binding at point \( J \).

Now we challenge the stability of \( J \) by considering deviating discount contracts. As \( J \) is below \( IC_p^D(I) \), which represents indifference to \( IC_p^F(E) \), deviations involving full service contracts are irrelevant, since even the best non-loss-making full service contract \( E \) makes the patient types worse off than \( J \).

To the left of \( EB_p^D \), offering a discount contract above \( IC_p^D(J) \) does not make sense: Above \( IC_p^D(J) \) and below \( IC_i^D(J) \) only impatient types can be attracted. To the right of \( EB_i^D \), they employ low effort, which results in a loss. To the left of \( EB_i^P \), they employ high effort, but since \( IC_i^D(J) \) is assumed to be always below the fair premium line under discount insurance, this results in a loss as well. Above \( IC_i^D(J) \) no customers can be attracted. Due to assumption 4.3, deviating offers below \( IC_p^D(J) \) would attract both types of customers. However, since this
region is below the fair pooling line under discount insurance, such deviations are loss-making. To the right of $EB^D_p$, there are no profitable deviations either. This is because the patient types employ low effort and can not even be attracted away from $J$ with contracts on the fair premium line under discount insurance and low effort, resulting in a loss probability of $s_l$.

As to uniqueness, the impatient types prefer full service contract $D$ to any other full service contract which lies on or above the applicable fair premium line. Offering discount insurance which would provide the same level of expected utility to impatient customers as $D$ is impossible without incurring a loss. This is because $IC^D_i(J)$ is assumed to be always below the applicable fair premium line. The patient types prefer discount insurance contract $J$ to any other full service or discount insurance contract which is on or above the applicable fair premium line and meets the incentive compatibility condition. As a result, there can be no other separating equilibrium than $\{D, J\}$.

A discount pooling equilibrium would have to be on or above the fair pooling line under discount insurance. However, $IC^D_i(J)$ is assumed to be always below the fair pooling line under discount insurance. Therefore, the impatient types would not want to buy such a pooling contract. Since $IC^F_p(E)$ is always below the fair pooling line under full service insurance, there can be no full service pooling equilibrium either because it would be destroyed by $E$. Uniqueness follows. □

4.5 Conclusion

In this article we employ a 2-period competitive insurance model, which is characterized by the simultaneous presence of moral hazard and adverse selection. Moral hazard is modelled along the traditional lines and is assumed to occur with regard to unobservable precautionary effort which can either be high or low. However, adverse selection occurs with regard to the personal discount rate of consumers which can be high or low as well. It is assumed that consumers decide whether or not to buy insurance in the first period. If so, they have to pay the premium up-front and decide about the precautionary effort level they wish to employ. In the second period, the consumer faces the risk of a loss. This setup is meant to capture the fact that real world insurance contracts usually require the
consumer to pay the premium up-front for several periods (e.g. 12 months) which creates a role for the consumer’s personal discount rate in the ex-ante valuation of contracts.

The different time preference among consumers opens the door for separating equilibria in which the patient types employ a high effort level and buy high cover, while impatient types employ low effort and buy little or even no cover. Since high effort implies low risk and vice versa, this result is equivalent to saying that the low risks are more fond of buying insurance. Thus, the prevailing outcome is the opposite of the traditional adverse selection theory where adverse selection takes place with regard to consumer risk types. Furthermore, it is possible that the equilibrium contract for the patient types is profitable. In this case, undercutting the premium would attract all the impatient types who employ a low effort. This would result in a loss for the insurer.

If profits exist, they turn out to be quite robust. Even though ex-ante randomization enables insurers to compete away any profits, the validity of this concept for the real world insurance market is doubtful since it is not observed in practice. Ex post randomization, however, is not capable to increase expected utility of the patient types in comparison to the deterministic equilibrium contract. Thus there is no better separation of types and profits persist.

A further way of how profits with the patient types can be competed away is by introducing differentiated insurance products in the form of full service and discount insurance. We think of discount insurers as being cheaper but having a service which is not as good. We assume that the service level is defined by the delay with which an insurer pays out indemnities in case of a loss. Since patient consumers do not mind a later payment that much, this kind of product differentiation can potentially achieve a better separation of agents which allows for competing down remaining profits. However, it may be that there are still some profits left in equilibrium, depending on the shape of the incentive compatibility constraint.

Theoretical models of asymmetric information in the insurance market can be grouped according to whether they employ pure adverse selection (e.g. Rothschild and Stiglitz (1976)), pure moral hazard (e.g. Arnott and Stiglitz (1988)), or a
combination of both (e.g. de Meza and Webb (2001) or the present model). Both pure theories on their own suggest that in equilibrium a higher cover implies a higher risk, thus leading to a positive correlation between cover and loss probability. In a recent article, Chiappori et al. (2005) establish sets of conditions under which this positive correlation property follows also from models with simultaneous moral hazard and adverse selection. The present model satisfies one set of these conditions, which require that the consumers know their loss probabilities and their risk aversion be identical and publicly known. Interestingly, there can still be a negative correlation between loss probability and cover, as can be seen from the separating equilibrium depicted in figure 4.1 on page 39. The reason for this discrepancy is that different personal discount rates translate into different curvatures of patient and impatient types’ present value indifference curves. This is equivalent to different risk aversions in a one-period model.

In this model we abstract from the existence of a capital market. If consumers can borrow and transfer money between periods, this will surely have a substantial impact on the power of using consumers’ time preference as a screening device. However, capital market imperfections like borrowing constraints and deviations between the consumers’ personal discount rate and the market interest rate might still be interesting issues to be analyzed in the context of this model. We leave this for future research.
5 Insurance Brokers and Advice Quality

5.1 Introduction

On January 15, 2003, the European Union announced directive 2002/92/EC regarding the mediation of insurance. The directive’s aim is to harmonize the distribution of insurance within the EU’s member countries and to protect insurance customers. Among other arrangements, the directive requires insurance and reinsurance intermediaries to “possess appropriate knowledge and ability, as determined by the home Member State of the intermediary.”

In Germany, for example, the requirement is to attend 230 hours of professional training. This directive gives reason to have a closer look at the role of advice quality in the insurance intermediation business, which is the aim of this paper.

The model we employ is intended to isolate the effect of an insurance broker’s advice quality on social welfare and on the division of the insurance market between brokers and dependent agents. Therefore, we abstract from price differentials for similar insurance policies. This is in contrast to existing literature in this field which associates the role of an insurance broker with a price search service for consumers. This is certainly an important aspect. However, the focus of this chapter shall be on the broker’s advice quality not with regard to finding the cheapest policy but to recommending the right policy which covers the consumer’s needs best. In our view, this way of modelling the problem comes closest to the intention the EU had in mind when passing the insurance mediation directive mentioned above.

In our model there are two similar insurance policies from two different insurers in the market. Both policies cost the same. However, they differ in some details which are not obvious to the consumers. The reason for the difference is treated as exogenous. We assume that consumers are equally distributed with regard to their ex-ante probabilities that a particular one of the available insurance policies is best for them. Those consumers who feel sufficiently sure as to which policy fits their needs best will buy directly from an exclusive agent. The

\(^2\text{See directive 2002/92/EC of the European Parliament and of the Council, article 4(1).}\)

\(^3\text{In this article we will use the terms “exclusive agent” and “dependent agent” interchangeably.}\)
others consult an insurance broker who will sell them the policy he recommends, however for a markup. The broker’s advice will be correct with a certain probability which we henceforth refer to as advice quality. The broker can invest in his advice quality which is assumed to cause increasing and convex costs.

A monopoly broker who is not subject to potential market entry by competitors will always offer an advice quality which is lower than in the social optimum. This is because the monopoly broker ignores the positive external effect of higher advice quality on the consumers’ surplus. Furthermore, he will also charge a price\(^4\) which is higher than in the social optimum. The reason for this is that the investment in advice quality is a fixed payment which is sunk at the time when the broker sells his policies. All in all, welfare in the situation of an uncontested monopoly broker is strictly lower than in the social optimum. The first best welfare level can only be achieved by regulation of both the broker’s price and his advice quality. However, even if regulation is limited to advice quality as a policy instrument, welfare can be increased in many situations.

We also look at situations where the incumbent broker faces the threat of market entry by competitors. Even in such a situation the incumbent broker remains the only broker in the market in equilibrium. The intuition behind this result is that once a competitor enters the market Bertrand price competition will be triggered. As investments in advice quality are sunk at this stage, the incumbent broker will undercut the new entrant by offering better advice quality or a lower price until he reaches a zero profit. This mechanism of fierce undercutting allows him to establish credible entry deterrence. The equilibrium price-advice quality combination offered by the contested incumbent broker will depend on the flexibility of price and advice quality once they have been chosen for the first time. We analyze the situations of full flexibility of both price and advice quality, flexibility of price but inflexibility of advice quality as well as inflexibility of both variables. Again, the resulting welfare level is always below the socially optimal welfare level. Only regulation of both price and advice quality can achieve the optimal welfare. However, in the case of full flexibility of both variables, even reg-

\(^4\)Price in the context of the broker shall denote the broker’s markup only, not the total price of a policy bought from a broker.
ulation of advice quality only is always welfare enhancing. This result also holds in the case of flexibility of price but inflexibility of advice quality for a broad range of parameter specifications. Only if both variables are inflexible, calling for a binding minimum advice quality is detrimental to social welfare.

Altogether, analyzing the role of brokers’ advice on finding the policy which fits a consumer’s needs best provides two main insights: Firstly, the broker market will in general be characterized by a price which is too high and an advice quality which is too low in comparison with the social optimum. Secondly, even regulation which is restricted to improving advice quality can increase social welfare in many situations. This seems important as regulation of price may not be feasible and also not preferable in practice. Reasons for this might be high costs of analyzing the broker’s markup of thousands of similar insurance policies and the large bureaucracy which would be associated with it. However, regulating advice quality via a minimum training level as arranged by the EU’s directive may be a reasonable procedure in two out of three market situations which are looked at in the theoretical framework of this paper.

5.2 Related Literature

There are already some papers that deal with insurance intermediaries and that mainly focus on the coexistence of insurance brokers and exclusive (dependent) agents. The basic question is why brokers can survive in the market even though they are apparently more expensive (e.g. Joskow (1973), Cummins and VanDerhei (1979) and Barrese and Nelson (1992)) and thus seem to be inefficient.

5.2.1 Explaining coexistence: Principal agent models

Two main branches can be identified in the literature. The first one tries to explain the brokers’ existence with agency conflicts. The second branch focuses on value adding services the broker provides to his clients.

Agency conflicts between insurers and agents are analyzed by Marvel (1982), Grossman and Hart (1986) as well as Sass and Gisser (1989). In their article, Mayers and Smith Jr. (1981) confine themselves to a brief statement on the expected intensity of agency problems between the insurance company and the
insurance agent, as well as the insurance agent and the policyholder. The authors express the view that with an independent agent the insurer-agent conflict is relatively greater whereas the agent-policyholder conflict is relatively smaller. Therefore, insurers who offer high-price, high-service policies would rather use independent agents whereas exclusive agents tend to represent low-service, low-price policies.

The basic argument of Marvel (1982) is that exclusive dealing is a means to enforce supplier property rights. For example, an insurance company may invest in substantial advertising and pass on some of the costs to the distribution system via lower commissions. Then brokers have an incentive to sell policies from rival insurers who do not advertise and thus pay higher commissions. The author points out that using exclusive agents prevents such free riding and protects the insurer’s investment in advertising. In this sense, coexistence of brokers and exclusive agents can be explained. This is because higher efficiency of company-level promotion favors exclusive agents. However, if salesmen-initiated marketing effort is more efficient this creates a tendency towards independent agents. As pointed out before, a lower commission for exclusive agents can be explained in the same line of argument. One implication resulting from this theory is that exclusive agent insurers (direct writers) should be more successful in lines of business where advertising is more important than services provided by the agent and thus no high commissions need to be paid. Moreover, direct writers should have higher expenditures for advertising as a fraction of their premiums than insurers using brokers. Marvel (1982) also tests his theory empirically and finds that brokers are very dominant in selling insurance to business clients whereas regarding private clients the share of exclusive agents is significantly higher. The author concludes that advertising is probably more important in the mass business with private clients whereas personal contacts can be used more successfully by agents in the business lines. Moreover, it is confirmed empirically that a higher requirement for extensive agent services, as represented by high commission expenditures, gives brokers a larger market share. Finally, the data suggest that insurers who use brokers spend a smaller proportion of their premiums for advertising what is also in line with the theory.
The article by Grossman and Hart (1986) is concerned with optimal ownership of the residual rights of a firm’s assets when only incomplete contracts can be written. For example, in the contractual relationship between two firms (e.g. a supplier and a manufacturer) it could be very costly to write a contract including all the specific rights firm 1 requires regarding firm 2’s assets. Then it might be efficient to let firm 1 purchase firm 2 and thus give firm 1 control of the residual rights over firm 2’s assets. The authors set up a 2-period-model. In period 1 both firms make relation-specific investments. Both firms’ managers can observe each others investment, there is neither asymmetric information nor uncertainty. In period 2, both managers choose a further strategy variable, which represents residual rights of control, for example measures which influence product quality. Both, the startup investment in period 1 and the choice of the strategy variable in period 2 determine the managers’ private benefit (e.g. perquisites) in period 2. It is assumed that in period 1 the investment, the strategy variable and the benefit are noncontractible. However, the strategy variable shall become contractible in period 2 and costless renegotiation with a 50% sharing of gains is possible. Therefore, renegotiation will always take place in period 2 and both managers receive half of the gain from cooperation which is realized in addition to the benefit of the noncooperative Nash equilibrium regarding the strategy variable. The noncooperative Nash equilibrium for the investment in period 1 involves full anticipation of renegotiation in period 2. According to the sharing rule the first order condition of period 1’s Nash equilibrium puts 50% weight on the noncooperative solution in period 2. However, in the first best with full contractibility 100% weight would be put on the cooperative solution since it maximizes the sum of both managers’ benefits. This explains why the firms’ investment decision in period 1 is distorted in the case of nonintegration. In the case of integration, firm 2 for example is taken over by firm 1 in period 2 and firm 1’s manager chooses firm 2’s strategy variable. An analog line of reasoning as in the nonintegration case shows that period 1 investments are distorted in the case of integration as well. The intuition behind this is as follows: After buying the assets of firm 2, firm 1’s residual rights give it the power to appropriate firm 2’s surplus in contingencies which are not contractually specified. Therefore, firm 2’s
incentive to invest beforehand is decreased which leads to underinvestment while firm 1 tends to overinvest. As a result, nonintegration, firm 1 control or firm 2 control should be chosen depending on where the least amount of surplus is lost due to investment distortions.

Grossman and Hart (1986) apply their line of reasoning to the case of insurance distribution. Firm 1 is the insurance company, firm 2 is the agent and the accordant asset is the client list. The authors point out that ownership of the client list determines whether an agent is an independent agent (broker) or an exclusive agent. A broker owns the client list what allows him to switch clients to whatever insurer he wants. Moreover, an insurance company is not allowed to contact a broker’s clients in order to sell their policies, not even if the broker terminates the business relationship with the insurer. In order to create a situation analogous to the model above, the agent is required to make a startup investment in the beginning and to receive his payoff later. According to Grossman and Hart, this can be achieved with a commission structure which is “back-loaded”. This means that for newly written business the agent gets a commission that does not quite cover his initial acquisition costs, what represents the agent’s startup investment. However, if the client renews his policy the agent receives a payoff via a commission which is higher than his renewal costs. The authors argue that such a commission structure is quite realistic when the insurance company wants to motivate the agent to spend effort (e.g. fast claims handling) in order to achieve renewals and thus establish profitable long lasting client relationships. The insurance company’s startup investment may consist of, among others, advertising, training of agents or product development. The investments of both parties can be distorted by noncontractibles (1) to the detriment of the insurer if the agent owns the client list and (2) to the detriment of the agent if the insurer owns the client list. With regard to (1) it is argued that the insurer could develop a new product but cannot solicit business with the broker’s clients without his permission. Moreover, the broker might switch clients to rival insurers. Examples for case (2) are that the insurer may decrease his advertising expenditures, offer lower quality products or raise his prices to an uncompetitive level. All in all, the decision about whether to use brokers or exclusive agents should be made
according to where the least amount of surplus is lost due to distortions of the startup investments. Grossman and Hart also look at insurance branches in which the agent’s effort in order to achieve renewals is less important, e.g. for whole life insurance policies. Then the agent’s startup investment is less important and increased list ownership of the insurance companies should be observed. Indeed, for whole life insurance the percentage of brokers as part of the distribution system is significantly lower (19.4%) than, for example, for substandard insurance (55.9%).

A further theory explaining the apportionment of insurance distribution between brokers and exclusive agents is put forward by Sass and Gisser (1989). In contrast to Grossman and Hart (1986) as well as Marvel (1982) their argument is independent of incontractibles and supplier property rights and relates to reducing the agency costs inherent in commission contracts. The authors’ starting point is the general share contracting problem: A broker’s sales effort determines the payoff of the insurers he represents. As the broker is only entitled to a share of the payoff (commission), he employs less effort than the insurers would. It is argued that switching the broker to be an exclusive agent for one particular insurer will induce him to provide a higher level of effort for this insurer than before. This is because taking away the possibility from the broker to market rival products lowers his opportunity costs of selling the policies from his new single principal. However, the broker will only agree to be an exclusive agent if he can earn at least as much money as before. Therefore the authors argue that it is easier for large insurers who operate in a large area with a high population density to rely on exclusive agents. Smaller insurers may not be able to generate enough business for an exclusive agent and thus use brokers. Furthermore, it is demonstrated that the insurer’s profit maximizing commission for an exclusive agent is lower than for a broker. Sass and Gisser (1989) test their model also empirically with data from listed property-liability insurers. It turns out that insurers’ size and the population density do indeed play a positive and significant role for the use of exclusive agents. However, in contrast to Marvel (1982), advertising has an insignificant effect only. The prediction that the optimal com-

---

mission rate will be mainly driven by the choice of the distribution system is also found to be supported by the data.

Finally, the article by Kim, Mayers, and Smith Jr. (1996) contains an empirical study which makes a systematic approach in order to identify the most important determinants of an insurer’s choice between brokers and dependent agents. The analysis is tailored around contracting problems between insurers and agents, as well as insurers and clients. The design of the study and the data used are capable of providing evidence regarding the validity of the insurer-agent related articles described above. Moreover, the study also looks at a conflict between the insurance company and the policyholders which was pointed out by Mayers and Smith Jr. (1981): The insurance company may have an incentive to expropriate its policyholders. This is because premiums are prepaid and later on the insurer could deny to deliver the promised payments or services. However, the extent to which such exploitation is possible depends on the discretionary leeway of the insurance company. In some lines of business high insurer discretion is necessary due to a higher frequency of disputes over loss amounts, limited information on loss distributions or less stability of the legal environment. Moreover, it is argued that also the insurance company’s ownership structure plays a role in this regard. For example, Lloyds or closely held stock companies may have better possibilities to exploit clients than mutuals. It is pointed out that independent agents can be used by insurers to bond themselves against expropriative behavior. This is because independent agents protect their policyholders since they often handle small claims themselves and could also threaten to switch their clients to a different insurer. Therefore the authors expect that independent agents should play a more significant role for particular lines of insurance and ownership structures.

The empirical results of Kim, Mayers, and Smith Jr. (1996) are obtained by investigating the association between the type of distribution system (independent or exclusive agents) and the following variables: insurer’s ownership structure, lines of insurance, advertising expenditures, firm size, geographic concentration and cost structure. The lines of insurance are identified to be the main driving force, whereby higher-price, higher-service policies appear to be
sold predominantly by independent agents. Ownership structure is also a significant influence. In this regard, support for the hypothesis of Mayers and Smith Jr. (1981) is found. Furthermore, the results are consistent with the argument of Sass and Gisser (1989) whereupon higher geographic concentration argues for exclusive agents. Furthermore, advertising is also a significant factor in the sense of Marvel (1982) and Grossman and Hart (1986). However, Marvel’s finding that insurers who specialize in private lines use exclusive agents could only be verified for large insurers because small insurers tend to rely on independent agents.

5.2.2 Explaining coexistence: Broker provides service

One limitation of the agency-theoretic models introduced in the previous section is that they do not take account of the consumers’ choice between the two distribution systems. Therefore, a different branch of theoretical literature has evolved which mainly looks at value added services the broker provides to his clients. However, mixed empirical results are reported as to whether or not independent agents really do provide better service quality to their customers. An empirical analysis by Etgar (1976) finds that brokers do not necessarily provide better services in all dimensions (e.g. assistance in risk analysis, placing insurance applications or claim handling), but may even be worse than exclusive agents in some regards. Also Cummins and Weisbart (1977) basically get the same result. On the other hand, empirical support for the hypothesis that brokers actually do provide a higher quality of service also exists. For example, Barrese, Doerpinghaus, and Nelson (1995) obtain this result on the basis of using complaint data in the context of private passenger automobile insurance. In the same line, Berger, Cummins, and Weiss (1997) find that the cost differential between independent and exclusive agents is more due to different service qualities than market imperfections.

Posey and Yavaş (1995) develop a search model where insurers and potential clients search for each other to form a match. Both insurers and clients differ in their costs of doing the search on their own and can either be a low or high search cost type. By using a broker, an insurance company can avoid incurring search costs but has to pay commission instead. Consumers and insurers decide
independently which distribution system to use. Prices are exogenous in the sense that it is assumed that each of the two insurer types charges the accordant zero profit price. Therefore, the nature of the model is not really one of price search but rather of finding a good match. A consumer wants to find an insurer who offers the kind of policy he is interested in. Not all insurers offer the same products, for example due to comparative cost advantages. Therefore, only a subset of the insurers offers an appropriate policy for a particular consumer, whereas the rest of the insurers does not. The commission charged by the brokers is also assumed to be exogenous. The authors show that there exist equilibria in which brokers and exclusive agents coexist even if a policy bought from a broker is more expensive than from a low cost direct writer. This model is related to the one in this chapter in the sense that it focuses on finding a good match between the insurance policy and a consumer’s needs as well. However, in Posey and Yavaş (1995) the brokers’ commission is exogenous and their advice quality is not modelled.

Posey and Tennyson (1998) is an extension of Posey and Yavaş (1995) as price is now determined endogenously. Therefore, the model has now the character of price search. In contrast to Posey and Yavaş (1995), all insurers offer only one homogeneous insurance product. There are high and low production cost firms as well as high and low search cost consumers. Each insurer decides whether to be a direct writer or use independent agents. Two methods are available for the consumers to locate price information: They can decide to search for the lowest price on their own by contacting exclusive agents, one after another, until their reservation price is met (“sequential search”). Due to their search costs consumers will only do a finite maximum number of searches, what gives firms some market power. Alternatively, it is assumed that they contact only one independent agent and buy the cheapest policy available (“nonsequential search”). However, independent agents do not represent all insurers who are not direct writers. Instead, it is assumed that an independent agent represents a particular number of insurers who are randomly assigned. Again, Bertrand competition is prevented because the event characterized by charging a price higher than marginal costs and still being the cheapest insurer in the sample of a particular independent agent has a strictly positive probability. Insurers’ profit
maximizing price depends on their production costs. Due to market power, even insurers with higher costs can survive and price dispersion arises in the coexistence equilibrium, both in the direct writer as well as in the independent agent market. The dispersion of prices relative to the dispersion of production costs is found to be smaller in the direct writer sector than in the independent agent sector. A further result of the model is that the mean level of price (cost) for direct writers is smaller than the mean price (cost) for independent agent insurers. The authors find these properties of the equilibrium price distribution to be consistent with empirical data regarding automobile insurance.

A dynamic extension of the price search approach is by Seog (1999). The author analyzes when coexistence is also a long run equilibrium. In his model, price dispersion in the direct writer and broker insurance sector is exogenously given. Initially, consumers, who differ in search costs, are totally ignorant about the price dispersion. Each consumer’s search costs correspond to an exogenously given search number which represents how often the consumer will search. According to the search number, a consumer samples consumers of the previous generation who have already bought insurance in the past. This sample is assumed to be split into former broker clients and former direct writer clients in a way which is proportional to the market share of the two distribution systems. Each period a new generation replaces the old one. The result of the model is that coexistence is only a long run equilibrium if the fraction of the population which searches many times and the fraction which searches few times is balanced.

Another interesting aspect is highlighted by Venezia, Galai, and Shapira (1999). In their model the value added service of brokers is not price search but a better claims service. This is because in contrast to dependent agents they can threaten to switch their clients to a different insurer if a claim is not settled promptly and fairly. As a result the authors obtain a separating equilibrium in which high risk clients choose a broker whereas low risk clients buy directly. Therefore they argue that the observed cost differential may not be due to broker’s inefficiency but rather due to self-selection of consumers.

In Gravelle (1994) the service provided by a broker is giving advice about the suitability of an insurance product. The article mainly deals with the relative
efficiency of alternative broker remuneration systems. However, there is also a
section on brokers’ advice quality which puts the article in the closer neighborhood
of this chapter’s model. Two remuneration systems are compared: commission
versus fee-for-advice. Under the latter system, the broker is paid a fee for advice
directly from the consumer irrespective of any sales rather than a sales dependent
commission from the insurance company. The article analyzes the incentives of
the two remuneration systems on brokers and the consequences for welfare.

In the model there is one identical insurance product in the market which
is offered by several insurers who incur all the same constant marginal costs.
Consumers are uninformed about the utility they could derive from buying the
insurance product. This is because the utility is assumed to consist of a gross
benefit net of the degree of mismatch between the insurance product’s benefits
and the consumer’s needs. Consumers know the distribution of the mismatch in
society and the expected mismatch but they do not know their own realization
of the mismatch. If a consumer is contacted by an honest broker he learns about
his mismatch and can make an informed buying decision. However, if a consumer
is not contacted by a broker he remains uninformed. In this case he uses the ex-
pected mismatch to make a buying decision at the end of the period. A consumer
is contacted by at most one broker. Regarding brokers it is assumed that there
is unimpeded entry into the broking market. Each broker incurs a contact cost
and an opportunity cost relative to his reservation wage. A contact cost function
which is increasing in the number of brokers represents a negative marketing ex-
ternality (congestion): the more brokers are already in the market and compete
for clients the harder it becomes to achieve a contact with a consumer. The
opportunity cost function can be thought of representing an increasing ordering
of brokers according to their reservation wage. Due to free entry the number of
brokers in the market is determined by the marginal broker’s zero profit condi-
tion. Welfare consists of the unweighted sum of the expected surpluses of the
informed consumers, the uninformed consumers, the brokers, the insurance com-
panies and the taxpayers. Solving the maximization problem of the first best
allocation yields that the insurance policy should be priced at marginal costs.
The first best number of brokers is found by equating its marginal social benefit
to its marginal social costs. The marginal social benefit is the utility gain of one additional consumer who becomes informed by one more broker. The marginal social costs of one more broker are his own contact and opportunity costs as well as his detrimental external congestion effect on existing brokers.

In a market equilibrium, insurers engage in Bertrand competition via the use of brokers and therefore end up with a zero profit. This is because brokers are the only means to sell insurance policies. The author obtains that the market equilibrium with commission is not first best efficient. The product price is too high because both brokers and insurance companies ignore the congestion externality. The equilibrium number of brokers is inefficient as well. The reason for this is that it will be determined by the marginal broker’s zero profit condition instead of the according first order condition of the first best maximization problem. The marginal broker ignores his positive external effect on an additional informed customer and his negative congestion externality on existing brokers. Depending on the relative size of these two externalities, the equilibrium number of brokers may be too high or too low. Furthermore, the market equilibrium is not even second best efficient. This is because a regulator could increase social welfare either by establishing a suitable tax system or by controlling the commission.

Under a fee-for-advice system the broker is not paid a commission from the insurer. Instead, the consumer can pay a fee in order to get the broker’s advice regarding his mismatch. However, this fee has to be paid even if the consumer decides not to buy the policy afterwards. As insurers are not burdened with the commission anymore, Bertrand competition will drive down the policy price to marginal costs. However, the crucial factor for social welfare is the price of advice. If all consumers have the same gross benefit, their willingness to pay for the broker’s advice will also be the same. As consumers are passive and each consumer is contacted by one broker at most, the brokers are essentially in a monopoly position. Therefore, the brokers maximize their revenue and extract the whole rent the consumers gain from information. As a consequence, all consumers buy but their surplus is not larger than if they had remained uninformed. Again, brokers will ignore congestion externalities and enter the market until the marginal broker’s profit is zero. All in all, the author shows that for the equilibrium un-
der the fee-for-advice system the same result as for the commission system holds true: it is neither first- nor second-best efficient. Too many brokers will enter the market because, in contrast to the commission system, brokers internalize all the consumers’ rent from information.

Gravelle (1994) concludes with a section on brokers’ advice quality. The resulting advice quality under the two different remuneration systems is analyzed as well as the resulting welfare implications. Advice quality is modelled as the degree of a broker’s understating the mismatch vis-à-vis a consumer in order to increase the probability of a sale. In other words, poor advice quality is equivalent to a broker’s blunt lying and talking a consumer into buying the policy even though he would be better off without insurance. As lying cannot be detected by consumers all brokers lie under a commission system in order to maximize their expected revenue. Under a fee-for-advice system, however, brokers give accurate advice. This is because lying cannot increase the consumers’ willingness to pay for advice. Even though advice quality is clearly worse under a commission system it is demonstrated that the implications for social welfare are ambiguous. This is because lying under a commission system may generate a gain in the brokers’ rents which is higher than the consumers’ loss of surplus. The intuition behind this result is that not all consumers are worse off when the brokers lie because some of them would have bought with accurate advice anyway. In addition, lying leads to more brokers in the market. So more consumers are contacted which can also generate a positive effect for welfare if they had not bought if uncontacted.

The perspective employed by Gravelle (1994) regarding advice quality is different to the one in the present paper in several regards. Poor advice quality comes from deliberate lying of a broker rather than from insufficient training as in the present paper. Furthermore, consumers are assumed to be totally passive. In the present model consumers are thought of acquiring some information themselves before they decide actively whether to contact a broker or a dependent agent. Moreover, the focus of Gravelle (1994) is on whether it is advantageous for a consumer to buy a single undifferentiated insurance product. The situation we have in mind in this paper is that it is pretty clear that the consumer is better off buying a particular type of insurance. However, it is unclear which policy out
of a bundle of similar ones from different insurers the consumer should choose. In this regard we think of a more sovereign consumer who can use the internet or other information sources for a rough evaluation but who may lack the sophistication and experience to master the situation when it comes to the details. Finally, in Gravelle (1994) the broker’s decision which level of advice quality to employ is driven by revenue considerations only. In contrast, the costs of building up advice quality play a crucial role in the present paper.

Gravelle (1994) draws on two similar papers which are Gravelle (1991) and Gravelle (1993). Gravelle (1991) analyzes the welfare implications of the use of brokers in a market for life insurance. It is motivated by the sometimes advocated opinion that insurers’ competition in the commissions they pay to brokers leaded to excessive commissions. This resulted in higher prices of insurance and was detrimental to consumers. Therefore the question arises whether regulating brokers’ commissions or allowing insurers to collude can actually improve social welfare. The author finds that this may indeed be the case due to marketing externalities. The basic line of reasoning is the same as in Gravelle (1994). However, there is neither an analysis of fee-for-advice nor of advice quality. Gravelle (1993) is in some respect an advancement of Gravelle (1991) because it also takes brokers’ advice quality into account.

5.3 The Model

We assume that there are two insurance policies $A$ and $B$ of two different insurers in the market. These policies basically cover the same risk but differ in details. For example, both could provide fire insurance for private homes. However, one policy includes the family car, whereas the other one includes business equipment which is stored at home.

5.3.1 Consumers

Consumers are split equally into two types: For the first group policy $A$ provides a higher expected utility. The opposite is true for the second group. However, there is heterogenous information among consumers: Some have a good assessment regarding what policy to buy, whereas others only have a vague or no idea
whether to choose \( A \) or \( B \). We assume that consumers arrive at these different probabilities, because they are heterogenous with regard to their search costs or dislike for doing research on insurance policies. In order to model this situation, each consumer is assigned an ex-ante probability \( \pi_A \) with which policy \( A \) is better for him. In the spirit of Hotelling (1929), we assume that consumers are equally distributed with regard to \( \pi_A \) on a line between zero and one. Thus, a person who is located at \( \pi_A = 1 \) is completely sure that policy \( A \) is better for him than policy \( B \). Similarly, at \( \pi_A = 0 \) buying \( B \) is surely the best choice. However, all other consumers have a lower level of confidence as to which policy they should buy and somebody with \( \pi_A = \frac{1}{2} \) has absolutely no clue.

5.3.2 Insurance agents

Consumers have the possibility to contact a dependent insurance agent and directly buy policy \( A \) or \( B \) for a price of \( p_A \) or \( p_B \), respectively. Alternatively, they can seek the advice of an independent agent (insurance broker). The broker will provide a recommendation and sell the according policy, however for a markup of \( p_I \) in addition to the original price of the policy. The central issue in this paper is the probability \( s \) with which the broker gives the correct recommendation: \( \text{Prob}(R_i|i) = s \), where \( R_i \) denotes the event “broker recommends policy \( i \)” given that the consumer is of type \( i \) with \( i \in \{A, B\} \). For simplicity, we assume that the broker will recommend the wrong policy with probability \( (1 - s) \): \( \text{Prob}(R_i|\neg i) = 1 - s \). Once costs rise from zero for the first time, the broker incurs strictly increasing and convex costs \( C(s) \) for improving his signal quality \( s \), which will be pointed out in more detail later on.

In the remainder of the article we use the following notation: \( V_{ij} \) describes the valuation of consumer type \( i \) for policy \( j \). For example, \( V_{AA} \) characterizes the expected utility a consumer of type \( A \) derives from getting the correct policy \( A \), whereas \( V_{AB} \) represents the expected utility of consumer type \( A \) from getting the wrong policy \( B \), with \( V_{AA} > V_{AB} \). Furthermore, we assume symmetry, i.e. \( V_{AA} = V_{BB} \), \( V_{AB} = V_{BA} \) and \( p_A = p_B \). For notational convenience we define \( a := V_{AA} - V_{BA} \).
5.4 Monopoly broker without threat of new entry

In this section we compute the price and the signal quality of a monopoly broker when no new entry in the broker market is possible. Then we compare the results to the socially optimal price and signal quality (which would be chosen by a benevolent social planner).

![Figure 5.1: Independent agent (broker) customers’ and dependent agent customers’ expected utility, represented by $EU^i(s)$ and $EU^d(\pi_A)$, respectively.]

5.4.1 Monopoly broker’s price and signal quality

A rational consumer with $\pi_A > \frac{1}{2}$ who buys directly from a dependent agent will buy policy $A$. Then his expected utility is

$$EU^d = \pi_A(V_{AA} - p_A) + (1 - \pi_A)(V_{BA} - p_A)$$

$$= \pi_A a + V_{BA} - p_A$$

When a consumer buys from an independent agent, his expected utility is

$$EU^i = \pi_A \left[ s(V_{AA} - p_A) + (1 - s)(V_{AB} - p_B) \right] +$$

$$\left( 1 - \pi_A \right) \left[ s(V_{BB} - p_B) + (1 - s)(V_{BA} - p_A) \right] - p_I$$

$$= sa + V_{BA} - p_A - p_I$$ (due to symmetry assumption) \hspace{1cm} (5.1)$$

Equating $EU^d$ and $EU^i$ and solving for $\pi_A$ yields the probability $\hat{\pi}_A$ of the marginal consumer who is indifferent between the broker and buying directly
from the dependent agent:

\[ \hat{\pi}_A = s - \frac{p_I}{a} \]

This equation shows that the more clients buy from the broker the higher his signal quality \( s \), the lower his price \( p_I \) and the higher the consumers’ advantage \( a \) from buying the correct policy. A graphical representation is given in figure 5.1. Consumers whose ex-ante probability \( \pi_A \) is between \( 1 - \hat{\pi}_A \) and \( \hat{\pi}_A \) buy from the broker, since their expected utility \( EU^i(s) \) will be higher than if they buy from a dependent agent. This would give them an expected utility of only \( EU^d(\pi_A) \). It is noteworthy that expected utility when buying from the broker \( EU^i(s) \) is no function of \( \pi_A \) but of \( s \). Therefore, \( EU^i(s) \) is a flat line in figure 5.1. A higher \( s \) would shift the line upwards, thus leading to higher broker client surplus and therefore also raising aggregate consumer surplus.

In the model we assume that the client always buys the policy the broker recommends. Therefore, the question arises under what circumstances a consumer whose ex-ante probability favors policy \( A (\pi_A > 0.5) \) is willing to accept a broker’s recommendation to buy policy \( B \). Applying Bayes’ rule yields that this is only the case if \( s > \pi_A \). As \( \hat{\pi}_A = s - \frac{p_I}{a} < s \), this condition is always satisfied.

Due to symmetry, the overall demand for the broker’s service is \( 2(\hat{\pi}_A - \frac{1}{2}) \). The demand and the costs of the signal quality enter the profit function \( G \) of the broker which is

\[ G = 2 \left[ s - \frac{p_I}{a} - \frac{1}{2} \right] p_I - C(s) \quad (5.2) \]

To make the problem interesting, we assume that there is a region in the \((s, p_I)\)-space where the monopoly broker makes a strictly positive profit. This implies that costs \( C(s) \) will rise slowly enough as advice quality is increased from \( s = \frac{1}{2} \). Otherwise, no broker would be willing to enter the market. The monopoly broker will maximize \( G \) via the choice of \( s \) and \( p_I \) under the constraints of \( p_I \geq 0 \) and \( \frac{1}{2} \leq s \leq 1 \).

Solving the first order conditions of the unconstrained problem yields the
following equations:

\[ C'(s) = a(s - \frac{1}{2}) \]  \hspace{1cm} (5.3)

\[ p_I = \frac{1}{2}a(s - \frac{1}{2}) \]  \hspace{1cm} (5.4)

In order to find a unique absolute maximum under the unconstrained problem, \( G \) must be strictly concave. The corresponding Hessian of \( G \) is

\[ H = \begin{bmatrix} -C''(s) & 2 \\ 2 & -\frac{4}{a} \end{bmatrix} \]

For strict concavity the Hessian must be negative definite everywhere. This is satisfied, if \(-C''(s) < 0\) and \(\frac{4}{a}C''(s) - 4 > 0\) which is equivalent to \(C''(s) > a\). It can be seen that \( C(s) \) has to be sufficiently convex, as determined by \(C''(s) > a\), in order to guarantee strict global concavity of \( G \). However, assuming convex costs seems to be quite plausible, because improving the signal from mere guessing is relatively cheap, while achieving a perfect signal can be thought of being infinitely expensive, with increasing marginal costs in between.

As a next step we assume global strict concavity of \( G \) and try to replace the constrained maximization problem by an unconstrained one. Inspection of (5.4) shows that \( p_I \geq 0 \) will always be satisfied as \( a \geq 0 \) and \( s \geq \frac{1}{2} \). From (5.3) follows that \( s \geq \frac{1}{2} \) will always be satisfied as well, because we assume that \( C'(s) \geq 0 \) and \( a > 0 \). Therefore, the constraints of the maximization problem can be reduced to the condition \( s \leq 1 \). Then we plug (5.4) into (5.2) and obtain

\[ \hat{G} = \frac{1}{2}a\left(s - \frac{1}{2}\right)^2 - C(s) \]

which gives a more condensed version of the unconstrained problem in only one variable \( s \). It becomes clear that an optimal \( s < 1 \) will occur, if \( C(s) \) is convex enough as \( s \) converges to one. In the real world, it may be considered plausible that \( \lim[C(s)]_{s\rightarrow1} = \infty \), as described above. Therefore, the analysis to follow in this paper will be based on the following assumption:

**Assumption 5.1** Regarding the cost function \( C(s) \) it holds that \( C''(s) > a \) and \( \lim[C(s)]_{s\rightarrow1} = \infty \). Furthermore, \( a > 0 \), \( C'(s) \geq 0 \), \( C(\frac{1}{2}) = 0 \), \( C'(\frac{1}{2}) = 0 \) and \( C(s) < \frac{1}{2}a(s - \frac{1}{2})^2 \) on some interval \( s \in (\frac{1}{2}; \frac{1}{2} + \epsilon) \), with \( \epsilon > 0 \).
Altogether, assumption 5.1 guarantees global concavity of the monopoly broker’s profit function (5.2) and an interior solution for the profit maximizing \( s \). In order to make the problem interesting, the last condition ensures that \( C(s) \) rises slowly enough as \( s \) increases from \( s = \frac{1}{2} \) for the monopoly broker’s profit function to be strictly positive in some region of its domain.

After having motivated assumption 5.1, we are ready to interpret the solutions to the unconstrained maximization problem as being identical with the solutions to the relevant constrained maximization problem: The intuition behind (5.3) is that the optimal advice quality of the monopoly broker, which shall be denoted \( s^M \), can be found by equating the marginal costs of a quality increase \( C'(s) \) to the marginal increase in turnover which is the term \( a(s - \frac{1}{2}) \) on the right hand side. To see this, let turnover be defined by \( T(p_I, s) = 2(s - \frac{p_I}{a} - \frac{1}{2})p_I \). The total differential is \( dT = \frac{\partial T}{\partial p_I} dp_I + \frac{\partial T}{\partial s} ds \). However, \( \frac{\partial T}{\partial p_I} = \frac{\partial C}{\partial p_I} = 0 \) due to the first order condition. This yields \( \frac{dT}{ds} = 2p_I = a(s - \frac{1}{2}) \), if evaluated at the optimal price \( p_I \). It becomes clear that the monopoly broker ignores the positive externality of increased advice quality on the consumers’ surplus. Furthermore, it follows from (5.4) that the monopoly broker charges a strictly positive price \( p_I \) if \( a > 0 \) and \( s > \frac{1}{2} \).

### 5.4.2 Socially optimal price and signal quality

The socially optimal signal quality maximizes welfare \( W(s, p_I) \) subject to \( p_I \geq 0 \) and \( \frac{1}{2} \leq s \leq 1 \). Social welfare is the sum of the surplus of the clients who buy directly, the broker’s clients and the broker:

\[
W(s, p_I) = 2 \int_{\hat{\pi}_A(s, p_I)}^{1} \left( \pi_A a + V_{BA} - p_A \right) d\pi_A \\
+ 2(\hat{\pi}_A(s, p_I) - \frac{1}{2})(sa + V_{BA} - p_A - p_I) \\
+ 2(\hat{\pi}_A(s, p_I) - \frac{1}{2})p_I - C(s) \\
= s^2 a - sa + a - \frac{p_I^2}{a} + V_{BA} - p_A - C(s) .
\]

(5.5)

As before regarding the broker’s profit function, we solve the unconstrained maximization problem first. Taking the derivatives of \( W \) with respect to \( s \) and \( p_I \) yields:
\[
\frac{dW}{ds} = 2as - a - C'(s) = 0
\]
\[
\frac{dW}{dp_I} = -\frac{2p_I}{a} = 0.
\]
It follows that
\[
C'(s) = 2a(s - \frac{1}{2}) \quad \text{and} \quad p_I = 0. \tag{5.6}
\]
This shows that the constraint regarding \(p_I\) is not binding. Since we assume \(C'(s) \geq 0\) and \(a > 0\), the constraint \(s \geq \frac{1}{2}\) is not binding either. Therefore, the only constraint which cannot be guaranteed at this point is \(s \leq 1\). Let \(s^*\) denote the socially optimal signal quality and \(\hat{s}\) the solution of the unconstrained maximization problem as represented by (5.6). Then we have
\[
s^* = \begin{cases} 
\hat{s}, & \text{if } \hat{s} \leq 1 \\
1, & \text{if } \hat{s} > 1 
\end{cases} \tag{5.7}
\]
As we are interested in finding a unique absolute maximum of \(W\) we compute the accordant Hessian, which is
\[
H = \begin{bmatrix} 
2a - C''(s) & 0 \\
0 & -\frac{2}{a} 
\end{bmatrix}.
\]
It turns out that \(W(s, p_I)\) is globally strictly concave, if \(C''(s) > 2a\) for each \(s\). Thus we make the following assumption which will also be implied throughout this paper:

**Assumption 5.2** The costs of investing in advice quality \(C(s)\) are convex enough to guarantee global concavity of the social welfare function (5.5), what requires \(C''(s) > 2a\).

Then the following result can be obtained:

**Proposition 5.1** The monopoly broker will provide a socially suboptimal signal quality, \(s^M < s^*\), and charge a price which is higher than in the social optimum.
Proof The result follows directly from comparing (5.3), (5.4) and (5.6). □

It holds that \( p_I = 0 \) in the social optimum, because in (5.2) we abstract from any marginal variable costs associated with providing the broker service.

5.5 Incumbent broker with threat of new entry

In practice there are no substantial barriers to entry in the insurance broker market. Therefore, this section looks at the case when the incumbent broker is not shielded from potential competition by outsiders who may enter the market. The objective is to analyze whether the threat of new entry can induce the incumbent broker to offer a better price and signal quality than the suboptimal levels which prevailed in the previous section. In order to do that we need to define how competition works once a newcomer enters the broker market. The following game structure is used as a framework for the analysis:

1. The incumbent broker chooses a price-advice quality combination \((p_I^{I1}, s^{I1})\) and pays \(C(s^{I1})\).

2. A potential entrant decides whether or not to enter the broker market. If there is no entry, the game continues with stage 4.

3. If entry occurred, the incumbent broker and the new entrant engage in Bertrand style competition by choosing their price \( p_I \) and their advice quality \( s \) simultaneously:

   (a) entrant: \( p_I^E, s^E \), payment of \( C(s^E) \)

   (b) incumbent: \( p_I^{I2}, s^{I2} \), payment of \( C(s^{I2}) - C(s^{I1}) \), if \( s^{I2} > s^{I1} \)

4. Consumers buy and payoffs are realized. If the incumbent and the entrant offer the same price and advice quality, the market is assumed to split evenly.

Bertrand competition as opposed to Cournot competition seems to be appropriate in this context. This is because insurance brokers provide an intangible service upon request rather than producing countable quantities of some sort of product.
Different equilibria of the game arise depending on whether or not the circumstances in the insurance broker market allow changes of \( p_I \) and \( s \) after they are chosen by a broker for the first time. This is relevant for the incumbent only, because he makes a choice concerning his price and advice quality in stage 1 and may want to change his choice in stage 3 depending on whether or not entry has occurred. We look at three cases: 1) Full flexibility of both \( p_I \) and \( s \). In this case, the incumbent can change both variables in stage 3 of the game at will. 2) Flexibility of \( p_I \) but inflexibility of \( s \). In this case, the incumbent is committed to the advice quality chosen in stage 1 of the game: \( s_I^2 = s_I^1 \). 3) Inflexibility of both \( p_I \) and \( s \). Then, the incumbent is fully committed to both variables as chosen in stage 1: \( s_I^2 = s_I^1 \) and \( p_I^2 = p_I^1 \).

The equilibria will be outlined below with the help of figure 5.2 (page 81) which shows an example of a contour plot of a monopoly broker’s profit function according to (5.2). The horizontal axis represents signal quality \( s \) and the vertical axis price \( p_I \). Point \( M \) marks the peak of the profit function, whereas the outer-most contour depicts the zero-profit locus whereby the costs of the according advice quality \( C(s) \) have already been subtracted. The sloped line which is tangent to the zero-profit locus at \( T \) and its parallels show broker clients’ indifference curves according to (5.1). The socially optimal contract is represented by point \( W \).

Before proceeding with the analysis of the equilibria of the game, we prove the following lemma which will turn out to be useful later on:

**Lemma 5.1** The monopoly broker’s profit-maximizing price \( p_I^m \) under the constraint of a binding minimum advice quality \( \bar{s} \) is described by \( p_I^m(\bar{s}) = \frac{1}{2}a\bar{s} - \frac{1}{2}a \). In figure 5.2, this relationship is represented by the grey line through \( M \), \( C \) and \( R \). Starting from the monopoly contract \( M \), broker clients’ surplus and therefore total consumer surplus is increasing in \( \bar{s} \) when moving along \( p_I^m(s) \).

**Proof** The first part of the lemma is obtained immediately by maximizing (5.2) over \( p_I \) for any given \( \bar{s} \). Regarding the second part of the lemma, it holds that \( \frac{dp_I^m}{ds} = \frac{1}{2}a \), while the steepness of the broker clients’ indifference curves according to (5.1) is \( a \). Thus, the broker clients’ surplus is increasing when moving
Figure 5.2: Contour plot of a monopoly broker’s profit function after taking account of the costs of the according advice quality.

along $p_I^m(s)$ to the right of $M$. As a consequence, total consumer surplus will be increasing as well, because direct clients will only become broker clients if they are better off. □

5.5.1 Full flexibility of $p_I$ and $s$

The equilibrium of the game is solved for via backward induction and leads to the following result:

**Proposition 5.2** If there is full flexibility of $p_I$ and $s$, the equilibrium of the game is as follows: The incumbent broker will charge the monopoly price and provide the monopoly advice quality $s^M$ which is represented by point $M$ in figure 5.2. The potential competitor will not enter the market.
Proof In stage 2, the entrant will only enter, if he ends up with a profit. If both \( p_I \) and \( s \) can be changed freely, Bertrand competition in \( s \) and \( p_I \) is triggered by the entry. The costs of advice quality \( C(s_I) \) the incumbent paid in stage 1 are sunk in stage 3. However, the costs of the entrant’s advice quality and the incumbent’s incremental advice quality are not sunk yet and thus still relevant for their decision. Therefore, the process of anticipated undercutting in stage 3 will end at point \( T \). At this point the entrant can no longer attract consumers with a contract they prefer to \( T \) without sustaining a loss. The incumbent would even be willing to undercut his rival beyond \( T \) as long as he can achieve a positive turnover in order to recover at least part of his sunk entry costs \( C(s_I) \). Under these circumstances, entering in stage 2 is never profitable. Therefore, the incumbent broker will not change his behavior in stage 1 due to a potential entrant but charge the monopoly price-advice quality combination \((p_M^I, s_M^I)\) at point \( M \). □

The result shows that the power of unleashed Bertrand competition together with sunk entry costs is strong enough to deter entry. This holds regardless of the incumbent broker’s offer in stage 1, even if it is the monopoly \((p_I, s)\)-combination at \( M \). Suboptimal advice quality still prevails just as if there was no potential competitor who could enter the market.

5.5.2 Full flexibility of \( p_I \) but \( s \) cannot be changed

If \( s \) can no longer be changed after the incumbent has chosen it, the power of Bertrand competition as a weapon to deter entry is weakened somewhat. An inflexible advice quality is quite plausible, because further training of a broker takes time. On the other hand, employing a lower advice quality on purpose may be possible, but under Bertrand competition only better quality is relevant.

Before looking at the resulting equilibrium, we establish the following lemma, which will be useful later on:

**Lemma 5.2** There cannot be two brokers in the market who are vertically differentiated with regard to advice quality.

**Proof** Broker customers achieve the same expected utility level regardless of their ex-ante probability \( \pi_A \) (see figure 5.1 on page 74) and thus no differences in the willingness to pay for advice quality arise. □
Lemma 5.2 rules out that two brokers can be in the market in equilibrium. Therefore, a situation of vertical product differentiation in the sense of Shaked and Sutton (1982) cannot arise.

**Proposition 5.3** If there is full flexibility of $p_I$ but $s$ cannot be changed after it has been chosen by the incumbent in stage 1, the equilibrium of the game is as follows:

1. If the consumers’ indifference curve through $s^M$ and $p_I = 0$ (figure 5.2) does not cut through the zero-profit locus of the incumbent broker’s profit function, he will charge the monopoly price and provide the monopoly advice quality $s^M$ which is represented by point $M$ in figure 5.2.

2. If the consumers’ indifference curve through $s^M$ and $p_I = 0$ (figure 5.2) does cut through the zero-profit locus of the incumbent broker’s profit function, he will offer a price and advice quality $s^C$ as represented by point $C$. If $a > 0$, contract $C$ is characterized by a higher advice quality, a higher price and a higher consumer surplus than contract $M$.

In both cases the potential competitor will not enter the broker market.

**Proof** Suppose the incumbent broker has chosen $M$ in stage 1 and is now committed to the corresponding advice quality which shall be denoted $s^M$. However, he can still engage in Bertrand competition using $p_I$ to undercut a new entrant.

Entering the market with a worse advice quality than $s^M$ is never attractive for a newcomer. Since the incumbent’s entry costs in the form of investments in advice quality $C(s^I)$ are sunk in stage 3, the limit of the incumbent’s undercutting would be offering $s^{I2} = s^M$ and a price $p^{I2}$ of zero which is equivalent to zero turnover. The corresponding indifference curve of the consumers is the one in the middle in figure 5.2 which goes through $p_I = 0$ and $s = s^M$. The consumers’ indifference curves will always have a positive slope as long as $a > 0$. Attracting customers away from the incumbent broker would therefore require the entrant to charge a negative price. This will provoke a loss for the entrant.
On the other hand, entering with a better advice quality than the incumbent broker might be profitable. Suppose again that the incumbent broker offers $M$. As before, maximum undercutting under Bertrand competition will lead the incumbent to end up offering $s^{I2} = s_M$ and $p^{I2} = 0$. However, if the corresponding indifference curve cuts through the zero-profit locus (as in figure 5.2), the entrant can offer a contract which is just marginally below this indifference curve but above the zero-profit locus. Doing so will allow the entrant to attract all consumers and make a profit, whereas the incumbent would end up with a loss of $C(s^{I1})$. Therefore, the incumbent broker will anticipate this outcome when choosing his offer in stage 1 and try to deter entry in stage 2. In order to do that the incumbent has to choose an advice quality $s^{I1}$ which is just high enough to prevent an entrant to make profits, if Bertrand competition forces the incumbent to set $p^{I2} = 0$ in stage 3. As depicted in figure 5.2, the lowest possible and therefore profit-maximizing advice quality $s^C$ for which this condition holds is determined by the intersection of the horizontal axis at $p_I = 0$ and the indifference curve which is just tangent to the zero-profit locus at point $T$. Then, the incumbent broker maximizes his profit at the resulting advice quality $s^C$ which leads to point $C$. Moreover, lemma 5.2 shows that coexistence of two brokers who are vertically differentiated with respect to advice quality cannot occur in equilibrium.

As a result, the equilibrium is characterized by the incumbent broker offering $C$ in stage 1 and the entrant deciding not to enter in stage 2, if the relevant indifference curve cuts through the zero-profit locus as described above. If not, the incumbent broker will offer contract $M$ in stage 1 and entry will not occur either. It follows directly from lemma 5.1 that $C$ entails a higher price, a higher advice quality and higher consumer welfare than $M$, if $a > 0$. □

5.5.3 Neither $p_I$ nor $s$ can be changed

**Proposition 5.4** If there is no flexibility regarding both $p_I$ and $s$ after the variables have been chosen by the incumbent in stage 1, the equilibrium of the game is as follows: The incumbent broker will offer a zero-profit contract as represented by $T$ in figure 5.2. The potential competitor will not enter the broker market.
**Proof** Suppose that the incumbent broker in stage 1 chooses any other contract within or on the zero-profit locus apart from $T$. Then, a new entrant could attract all clients by offering a contract which makes them better off but is still profitable. There would be no possibility for the incumbent broker to react. As a consequence, he has to fully anticipate this situation already in stage 1. Since the indifference curves of the broker clients through $T$ represent their maximal surplus subject to zero profit, $T$ is the only contract the incumbent broker can offer in order to deter entry. □

The intuition behind this result is that the power of Bertrand competition as a threat to deter entry is completely lost. In the example of figure 5.2 advice quality $s^T$ is again closer to but still below the social planner’s level $s^*$.

### 5.6 Policy implications

As shown in the previous section, in equilibrium there will always be just one insurance broker in the market. However, the contract offered by this incumbent broker depends on the flexibility of price $p_I$ and advice quality $s$ and points $M$, $C$ or $T$ may prevail as a result. This section deals with the question, if and how the incumbent broker should be regulated under these circumstances.

#### 5.6.1 Advantageousness of regulation

Regulation is reasonable if it increases social welfare. The first best level of welfare shall be defined as being the maximum attainable sum consisting of the broker clients' surplus, the dependent agent clients’ surplus and the broker’s profit. Point $W$ in figure 5.3 (page 87) represents the first best level of welfare on the vertical axis and the associated advice quality on the horizontal axis. Points $W$, $T$, $M$ and $C$ in figure 5.3 correspond to the respective points in figure 5.2. The same is the case with $R$ whose role will be explained below.

Social welfare in any market outcome, with or without threat of entry, is always below the first best level of social welfare. In terms of the graphical example given in figures 5.2 and 5.3, the welfare associated with $M$, $C$ and $T$ is always smaller than the first best level of welfare as represented by $W$. This is because due to assumption 5.2, $W(p_I, s)$ is strictly concave with a unique absolute
maximum at point $W$. Due to assumption 5.1, it holds that $s > \frac{1}{2}$ for each point $M$, $C$ and $T$ which requires the incumbent broker to charge a price $p_I > 0$ in order to achieve at least zero profit. Therefore, $W$ never coincides with either $M$, $C$ or $T$, since each of these points is associated with a $p_I > 0$, while $W$ is always associated with $p_I = 0$.

It follows that regulating the incumbent broker always offers the possibility to achieve a higher social welfare. This raises the question of how the incumbent broker should be regulated. Basically there are two variables which can be influenced by the regulator in order to increase social welfare: price $p_I$ and advice quality $s$.

### 5.6.2 Regulation using $s$ and $p_I$

As shown before, in a market outcome the monopoly broker will always charge a strictly positive price, while in the social optimum the broker’s price is zero. Thus, the first best level of social welfare can only be achieved by regulating both the broker’s price $p_I$ and advice quality $s$. However, the downside of doing so is that a regulated price $p_I = 0$ will inflict losses upon the monopoly broker which need to be covered somehow. As suggested by Gravelle (1991), this could be done by a lump sum tax which is just high enough to cover the monopoly broker’s expenses on advice quality.

Another possibility is to choose $T$ instead of $W$ as a regulation goal. In this case the monopoly broker makes zero profit. When there is total inflexibility of both $p_I$ and $s$ this result will prevail as a market outcome anyway and no further regulatory action needs to be taken.

### 5.6.3 Regulation using $s$ only

Even though only regulating both $p_I$ and $s$ is capable of securing the first best level of social welfare, using $p_I$ as a policy instrument for regulation might not be possible. Prescribing $p_I$ is difficult in practice, because the broker’s fee is hidden in the total premium charged from the client. Apart from that, the regulator would have to compare this hidden commission in thousands of similar policy pairs which creates substantial costs of regulation. Furthermore, as the broker’s
commission is part of the policy’s total premium, regulating $p_I$ would require to regulate again the price of insurance policies, at least with regard to broker’s commissions. This might be difficult to achieve politically, as it contradicts the spirit of the EU’s 1994 insurance market deregulation.

If there are reasons which prevent $p_I$ from being used as a regulation instrument, it follows that regulating $s$ only can never achieve the first best level of social welfare. The question arises how at least a second best regulation could be implemented. When the regulator prescribes a binding $s$, the monopoly broker will charge the corresponding profit-maximizing price $p^M_I(s)$ as described by lemma 5.1. The concave line in figure 5.3 (page 87) shows the social welfare associated with such a regulated $s$, when the monopoly broker charges the corresponding price $p^M_I(s)$. Point $R$ marks the maximum attainable social welfare and the corresponding advice quality $s^R$ in the case where only $s$ is regulated. Therefore, we call the associated welfare level second best optimal.

![Figure 5.3: Welfare under different competition situations.](image)

**Proposition 5.5** The advice quality associated with $R$, denoted by $s^R$, is always higher than the advice quality associated with $M$, denoted by $s^M$: $s^R > s^M$. 

87
Proof Starting from point $M$, it follows from lemma 5.1 (page 80) that the monopolist will move along $p_I^m(s) = \frac{1}{2}as - \frac{1}{4}a$ when prescribing a binding $s$ with $s = s^M + \epsilon$, with $\epsilon > 0$. However, $\frac{\partial G}{\partial s^M} = 0$ and $\frac{\partial G}{\partial p^M} = 0$ ((5.3) and (5.4), page 76). Therefore, it holds that $\frac{dG}{ds} = 0$ for a marginally small $\epsilon$. Let $Q$ denote total consumer surplus, being the sum of broker and dependent agent clients’ surpluses. Then, total welfare is represented by $W(s, p_I^m(s)) = G(s, p_I^m(s)) + Q(s, p_I^m(s))$. Lemma 5.1 shows that $\frac{dQ}{ds} > 0$ at point $M$. Thus, $\frac{dW}{ds} > 0$. □

The proposition shows that in the situation where $M$ is the market outcome, either with or without threat of entry, regulation is welfare enhancing even if it is limited to $s$ as a regulation instrument. This is because a higher $s$ is always advantageous for the consumers, taking into account that the broker sets his price along the profit maximizing path $p_I^m(s)$ as described by lemma 5.1.

Regulation of $s$ only may also improve welfare in a situation where $C$ is the market outcome, as shown in figure 5.3. However, it is difficult to tell analytically if $C$ could also be to the right of $R$. This is because $C$ is determined via the tangency condition at $T$ and $C$’s relative position with regard to $R$ depends critically on the shape of the zero-profit contour as depicted in figure (5.2). This zero-profit contour is driven by the shape of $C(s)$ which may take on unlimited different forms. Therefore, we have no universal analytical rule at hand under what circumstances $s^C$ could be equal to or larger than $s^R$, thus rendering the regulation of a binding minimum $s$ pointless. However, extensive simulations with a family of cost functions of the form $C(s) = k(s-0.6)^z$ yield that $s^C < s^R$ generally holds true. Only when $k$ grows very large or $z$ grows small we obtain $s^C = s^R$ in the limit. Details can be found in the appendix. This shows that regulation of $s$ in this context can be seen to be welfare increasing at least in a broad range of specifications.

However, if $T$ is the market outcome, increasing $s$ via regulation is never useful. This is because the monopoly broker will move away from the point of tangency $T$ on the zero profit contour. Therefore, social welfare will be smaller than at $T$ because the broker’s profit is still zero whereas consumer surplus decreases.
5.7 Conclusion

This paper analyzes the coexistence of insurance brokers and dependent agents with a focus on advice quality. Due to heterogeneity regarding search costs, consumers are assumed to arrive at different ex-ante probabilities that one particular out of a range of similar insurance policies fits their needs best. Consumers can buy directly from a dependent agent or consult an insurance broker who will give them advice which policy from the available ones is best for them. However, this service comes at a cost, since the broker will sell the client the recommended policy at a markup. His advice is of a particular quality which is represented by the probability that he recommends the correct insurance product. Before offering his service, the broker invests in his advice quality which is assumed to cause increasing and convex costs.

We break down the analysis into two parts according to whether or not a broker who is already in the market is subject to competition from potential new entrants. We obtain the result that an uncontested monopoly broker will offer less advice quality and charge a higher price than a benevolent social planner would choose in order to maximize social welfare. The outcome in the case with potential new entrants depends on the flexibility of the broker’s advice quality and price once entry has occurred. However, the first best welfare level is never achieved as a market outcome in this case either.

The fact that the first best welfare level is never realized, irrespective of whether new entry may occur or not, raises the issue of regulation. The social optimum can only be achieved by regulating both advice quality and price at the same time. However, price may not be available as a policy instrument for regulation. In this case, even regulation of advice quality alone increases social welfare in the case of full flexibility of price and advice quality. The same may hold true if advice quality is inflexible, depending on the shape of the broker’s cost function of investing in advice quality. However, if both variables are inflexible, regulation of advice quality alone no longer improves social welfare.

One characteristic of our model is that perfect observability of brokers’ advice quality by consumers is implied. We have in mind, that the results serve as a kind of benchmark: Advice quality in the unregulated market is found to be
suboptimal even in the limit of perfect observability. Therefore, unobservable advice quality will tend to be even worse. However, the EU’s directive also contains increased requirements regarding the broker’s documentation obligations and liability for wrongdoing. In particular, the broker has to create a written report explaining why selling a particular insurance product was an appropriate solution for the consumer’s needs. These measures are meant to be an incentive for brokers to be honest and give the best advice possible within their capabilities.

Of course, this theoretical model abstracts from many other aspects which are relevant characteristics of the insurance distribution system. As pointed out before, agency problems or a broker’s value added services, like price search or superior claims service, are subjects which have been discussed by various authors before. However, it is nonetheless interesting that focusing on brokers’ advice quality, as one building block of reality which has not received much attention before, can provide a rationale behind the EU’s directive on insurance mediation in two out of three market situations.
Appendix

Relative position of $s^R$ and $s^C$

The following graphs show simulation outputs of $s^R$ and $s^C$ for varying parameters of a family of cost functions of the form

$$C(s) = \begin{cases} 
  k(s - 0.6)^z & \text{for } s \geq 0.6 \\
  0, & \text{otherwise.}
\end{cases}$$

It becomes apparent that $s^R$ is always larger than $s^C$, although the difference $(s^R - s^C)$ converges against zero if $k$ grows large or $z$ becomes small. This shows that social welfare can always be increased at least slightly by regulating a minimum advice quality $s$ in the cases which have been simulated. However, the possible welfare gain decreases sharply in the limit of a large $k$ or a small $z$.

Further parameter specifications for the simulations were $V_{AA} = 100$, $V_{BA} = 95$ and $p_A = 70$.

Figure 5.4: $s^R$ and $s^C$ for various $k$ with $z = 1.5$.

Figure 5.5: $(s^R - s^C)$ for various $k$ with $z = 1.5$. 

91
Figure 5.6: $s^R$ and $s^C$ for various $z$ with $k = 6$. Figure 5.7: $(s^R - s^C)$ for various $z$ with $k = 6$. 
References


LEBENSLAUF

Persönliche Daten

Name Michael Sonnenholzner
geboren am 25.04.1975
Geburtsort Wasserburg am Inn

Ausbildung

1981 -1985 Grundschule Eiselfing
1985 - 1994 Luitpold-Gymnasium Wasserburg am Inn
07/1994 - 06/1995 Wehrdienst
09/1995 - 06/1997 Ausbildung zum Bankkaufmann bei der Bayerischen
Vereinsbank AG, Rosenheim
11/1997 - 05/2002 Studium der Volkswirtschaftslehre an der Ludwig-Maximilians-
Universität München, Abschluss als Diplom-Volkswirt;
Alumni-Preis für junge Volkswirte
10/2002 - 03/2005 Wissenschaftlicher Mitarbeiter am Lehrstuhl für Volkswirtschaftslehre, insbesondere Wirtschaftstheorie, an der
Friedrich-Alexander-Universität Erlangen-Nürnberg
10/2003 - 03/2004 Forschungsaufenthalt an der Universität Toulouse 1
20.06. - 29.06.2004 Teilnahme an der Summer School ”The Economics of Risk Bearing”, Hebrew University, Jerusalem