Chapter 1

Introduction

The Economics of Insurance Markets

Part 1: Demand for Insurance

Ch.1: Expected utility based models of insurance demand;
extension to state-dependent utilities

Ch.2: Background risk. Multiple risks with constraints on insurability.
Insurance demand in the presence of complete and incomplete asset markets

Ch.3: Non-expected utility theories of insurance demand

Part 2: Supply of Insurance

Ch.4: Risk-pooling, risk spreading and the Law of Large Numbers.
Syndicates. Arrow-Lind Theorem

Ch.5: Models of the shareholder-owned insurance firm. Pricing risks.
Investment Policy. Solvency

Ch.6: Models of the mutual insurance firm. Agency Problems in mutuals

Ch.7: Reinsurance

Part 3: Structure of Insurance Contracts under Symmetric Information

Ch.8: Arrow model, Raviv model. Deductibles, coinsurance

Ch.9: Catastrophe insurance

Part 4: Asymmetric Information I: Adverse Selection

Ch.10: Monopoly and adverse selection. Categorical discrimination.

Single and multi-period contracts. Renegotiation.

Ch.11: Adverse selection in competitive insurance markets.

Rothschild/Stiglitz model and extensions. Categorical discrimination.

Multi-period contracts. Renegotiation.

Part 5: Asymmetric Information II: Moral Hazard

Ch.12: Monopoly/perfect competition and moral hazard. Single and multi-period contracts.

Renegotiation. Self-Insurance and self protection

Ch.13: Third party damage: liability and negligience rules

Ch.14: Insurance fraud

Part 6: Public Economics of Insurance Markets

Ch.15: Regulation of insurance markets.

Ch.16: Social insurance.

1.1 Detailed Outline

Many relevant aspects of the economics of insurance can be discussed with the help of one simple model. Consider an individual who has to decide how much insurance cover to buy. Formally, she maximizes her expected utility by choosing the optimal indemnity I:

$$E[U] = (1-\pi)U(\omega - P(I)) + \pi U(\omega - P(I) - L + I)$$

Here we assume that the individual has a von Neumann-Morgenstern utility function $U(x)$ which is increasing and concave. As we will see in section 1, concavity implies that the individual is risk averse. π is the probability that a loss occurs. The loss has the size L. For example, when you have a van Gogh which costs \$10 Mio., hanging in your room, the probability of having the painting stolen is say $\pi = 0.1\%$. In this case $L =$ \$10 Mio.. The individual can buy insurance cover I by paying the premium $P(I) = \rho I$. For every \$ which she wants to get paid in case of a loss, she has to pay ρ. ρ is called the premium rate, while I is the indemnity. Thus, if the premium rate is 0.2%, and you want to get all of your \$10 Mio. back in case someone steals your van Gogh, you have to pay \$20000 to the insurance company. Note that if the van Gogh is stolen, you paid the premium already, so after all you receive \$9,980,000 as the net payment, which is $I - P(I)$.

This model is the basis for all the discussions in the following chapters. In the first part of **chapter 1** we deal exclusively with this model and we investigate how the demand for insurance depends on the price of insurance and the degree of risk aversion.

However, there are limitations to the applicability of this model. Many real world features like deductibles, contracts with experience rating, regulation of insurance, etc. require more elaborate models.

We will now discuss these shortcomings and you will see that nearly every chapter in this book deals with one of the features which this simple model does not adequately describe. (We use arrows to indicate where the modified models differ from the basic model above.)

1. State Dependent Utility Function

$$E[U] = (1-\pi)\overbrace{U}^{\Downarrow}(\omega - P(I)) + \pi \overbrace{V}^{\Downarrow}(\omega - P(I) - L + I)$$

For some applications it is not sensible to assume that people have the same utility whether a loss has incurred or not, even if they are fully financially compensated for the loss. Assume that your gold bar is stolen, but fully covered by an insurance policy. In this case you probably will not mind the loss. You just go out and buy yourself another gold bar. In the case of your van Gogh being stolen this might be different. If you are very attached to this painting, you will feel worse off even if the insurance company pays out the full price you have paid for it. The reason is that a particular van Gogh is not a tradable good which can be rebought in the market. Another example is health insurance - if you break your leg, even with full insurance you will feel worse than when you are healthy. Life insurance also fits this category - surely your utility differs whether you are alive or dead. These aspects are discussed in detail in **chapter 1**, where we consider the demand for insurance for *irreplacable commodities*.

Even allowing for state dependent utility functions one might criticise the use of the framework of expected utility analysis. In **chapter 3** we discuss other decision theoretic models and their consequences for the demand of insurance.

2. Is there only one risk?

$$E[U] = E_{\tilde{\omega}}[(1-\pi)U(\overbrace{\tilde{\omega}}^{\Downarrow} - P) + \pi U(\overbrace{\tilde{\omega}}^{\Downarrow} - P - L + I)]$$

In the simple model above, the van Gogh is either stolen or not. However, in general individuals face more than one risk. Standard additional risks like car accidents, illness, fire, etc. can be covered by separate insurance contracts. But there are also uninsurable risks around - for example income risk, as the return on shares and bonds you own is uncertain, or because your job is not secure. You might not know for sure how much money you are going to inherit from your grandmother, whether you will marry into money or not, ... This feature is known as *background risk*. In **chapter 2** we analyse the situation where individuals face additional uninsurable risks (like the $\tilde{\omega}$ in the equation above). Now the demand for insurance will depend on whether those risks reinforce each other or whether they can be used as a hedging mechanism.

3. Where does $P(I)$ come from?

$$E[U] = (1-\pi)U(\omega - \overbrace{P(I)}^{\Downarrow}) + \pi U(\omega - \overbrace{P(I)}^{\Downarrow} - L + I)$$

In the simple model we have assumed that the individual faces some exogenous given premium function $P(I) = \rho I$. But who determines the premium? On which factors does it depend? In **part 2** of the book, where we discuss the supply of insurance, this will become clear. We consider both a monopoly insurer as well as an insurer on a competitive market.[1] We will discuss how shareholders react to risks by *diversifying* their risks (**chapter 5**), how mutuals enable the insured to *pool* their risks (**chapter 6**), and why reinsurance firms may be used to *spread* the risks (**chapter 7**). Furthermore, in those chapters we discuss the agency problems which result from the different institutional and legal forms of insurance firms.

4. Is there only one loss level possible?

$$E[U] = (1-\pi)U(\omega - P) + \pi \overbrace{\sum_i \pi_i}^{\Downarrow} U(\omega - P - \overbrace{L_i + I_i}^{\Downarrow})$$

In many situations a single loss level does not seem appropriate. Certainly, your van Gogh is either stolen or not, but in the case of fire, for example, it could be partly or completely damaged. If you have a car accident, the damage can vary between some hundred dollars and many hundreds of thousands. Similarly in aviation insurance: A claim could have the size of a few hundred dollars for a damaged suitcase, but can increase to many millions of dollars. As a matter of fact, one of the largest liability claims in the history of flight insurance was the ATTENTAT of the PanAm Boing 747 over Lockerbie. So far more than \$510 Mio. have been paid. More than one loss level is discussed with the help of the model of Raviv, which we present in **chapter 8**. This model provides a synthesis of the demand for and supply of insurance in the case of many loss levels. In this model we will see *deductibles* and *coinsurance* emerging. By deductibles it is meant that the first D dollars of the loss has to be paid by the insured. Coinsurance applies if an additional dollar of loss is only partially covered. This might be the case if for example the insurance covers a fixed percentage of the loss.

Another example, where losses can potentially take on many different values, are catastrophes, like earthquakes, floodings, vulcano eruptions, etc. One obsevres, however, that usually only a small part of the damage is covered by insurance. Sometimes those hazards are even referred to as *unsinsurable risks.* We will take a closer look on the issue of *catastrophe insurance* in **chapter 9**.

[1]Insurance companies compete by Bertrand price competiton, so two firms are enough to restore perfect competition. That is why we do not discuss oligopoly models.

5. Is π known?

$$E[U_i] = (1 - \overbrace{\pi_i}^{\Downarrow})U(\omega - P) + \overbrace{\pi_i}^{\Downarrow}U(\omega - P - L + I)$$

By determining the premium rate from the point of view of the insurer it is usually assumed that the probability of risk π is known. However, this must not necessarily be the case. You probably know much better whether you are a cautious or a wild driver, whether you have a healthy lifestyle or not. This is modelled by assuming that the insured knows her own π_i and the insurance company only has some expectation over the π_i. In those cases high risks types with a large π_i try to mimick low risk types and buy insurance which is not designed for them. This effect is known as *adverse selection.* In **chapters 10 and 11** we discuss the seminal paper by Rothschild and Stiglitz and other models which deal with this topic. The phenomenon of adverse selection allows us to understand why in some cases insurers offer several different contracts for the same risk. For your car insurance, for example, you might buy a contract with no deductible and a large premium or with a deductible and a lower premium. Offering different contracts is always a sign of a discriminating mechanism, which only makes sense if people differ in some unobservable characteristic. This analysis also allows us to discuss another feature which is commonly observed: *Categorical discrimination.* What are the pros and cons of conditioning a particular contract on gender, for example? Is it efficient to sell different contracts to male and female or to young and old drivers?

6. Is the loss probability exogenous or endogenous?

$$E[U] = (1 - \overbrace{\pi(e)}^{\Downarrow})U(\omega - P) + \overbrace{\pi(e)}^{\Downarrow}U(\omega - P - L + I) - \overbrace{c(e)}^{\Downarrow}$$

In many situations the loss probability can be influenced by the insured. The degree of attentiveness you devote to the road is something you have control of. By increasing your concentration the loss probability reduces $\pi'(e) < 0$. However, the more you concentrate the less time you have for phone calls with your mobile phone, listening to the radio, etc., so there are costs of concentrating ($c(e)$) which increase if one employs more effort: $c'(e) > 0$. If a person is completely insured, she might not employ any effort as she is not liable for any damage. This problem is known as *moral hazard* and is discussed in detail in **chapter 12**. Here we will find another reason why insurance companies may offer contracts with *partial insurance.*

The problem of not being fully accountable for one's own action is particularly relevant in the context of limited liability. Once a firm is bankrupt or once a person has no wealth, damages which are done to others cannot be reimbursed. The institutional response to this is *third party insurance*, the economics of which we consider in **chapter 13**.

7. Is the size of the loss observable?

$$E[U] = (1-\pi)U(\omega - P) + \pi U(\omega - P - \overbrace{L}^{\Downarrow} + I)$$

In some situations neither the occurence of a loss nor the size of the loss is easily observable by the insurance firm. In those situations the insured might be tempted to overstate the size of a loss or to claim a loss which has not occured. *Insurance fraud* is the topic of **chapter 14**. For obvious reasons the actual size of insurance fraud is difficult to measure. However estimates based on questionares suggest that for the personal liability insurance around 20% of all claims are fraudulent. We will discuss how contractual and institutional arrangements might cope with this problem.

8. Why only one period?

$$\begin{aligned} E[U] &= (1-\pi)U(\omega - P_0) + \pi U(\omega - P_0 - L + I_0) \\ &+ \overbrace{(1-\pi)[(1-\pi)U(\omega - P_N) + \pi U(\omega - P_N - L + I_N)]}^{\Downarrow} \\ &+ \overbrace{\pi[(1-\pi)U(\omega - P_L) + \pi U(\omega - P_L - L + I_L)]}^{\Downarrow} \end{aligned}$$

If an insurance is sold under perfect information, it does not make any difference whether many single period contracts or one many period contract are sold. In reality however we observe many contracts which have a dynamic component, like *experience rating* contracts in the car or health insurance industry. In those cases, individuals pay a different premium depending on whether a loss has occurred or not (P_L or P_N respectively). To understand this phenomenon one has to resort to asymmetric information, like the adverse selection or moral hazard models mentioned above. In **chapters 10 and 11** we consider this issue in the context of adverse selection and show how experience rating may appear endogenously. Also in **chapter 12**, as part of the discussion on moral hazard, *dynamic contracts* are considered. Another topic which is relevant when one discusses multi period contracts is the issue of *renegotiation and commitment*. The crucial point here is that even if ex-ante both the insurer and the insured agree to a longer lasting contract, ex-post it might be of advantage for both parties to change the terms of the contract in some circumstances.

9. Is there a need for governmental intervention?

In a competitive and complete market there does not exist a reason for the government to intervene. In chapters 2 and 10-14 incomplete markets are discussed. Here markets are incomplete either because some risks are not insurable or because there exists some form of asymmetric information. As we will discuss in those sections state intervention might be welfare enhancing. Those models do not explain, however, why there is *regulation* on the capital requirements of insurance companies in many countries. In **chapter 15** we discuss the arguments for and against such regulation, in a model where we allow insurance companies to go bankrupt.

The government intervenes into the insurance form in another, more direct way, namely by providing *social insurance*. Unemployment insurance, social benefits, and in some countries health insurance is provided by the state and not by private firms. The economic rationale for this is discussed in **chapter 16**.

Part I

Demand for Insurance

Chapter 2

Expected utility based models of insurance demand

Includes:

extension to state-dependent utilities

Chapter 3

Background risk. Multiple risks with constraints on insurability

Includes:

Insurance demand in the presence of complete and incomplete asset markets

Chapter 4

Non-expected utility theories of insurance demand

Part II

Supply of Insurance

Chapter 5

Risk-pooling, risk spreading and the Law of Large Numbers

Includes:

Syndicates. Arrow-Lind Theorem

Chapter 6

Models of the shareholder-owned insurance firm. Pricing risks

Includes:

Investment Policy. Solvency

Chapter 7

Models of the mutual insurance firm

Includes:

Agency Problems in mutuals

Chapter 8

Reinsurance

Part III

Structure of Insurance Contracts under Symmetric Information

Chapter 9

Arrow model, Raviv model, Deductibles, coinsurance

Chapter 10

Catastrophe insurance

Part IV

Asymmetric Information I: Adverse Selection

Suppose that you have won an offer from the 'endurance society' to join a professional group on a climb of the Mount Everest. Statistically, 3.3% of all climbers leaving the basis camp die, the ratio of people who reached the summit to those who died is 4:1. Still, considering the risk worth taking for the dream of your life, you decide to join the tour. At some point during the preparation for the tour, you recall your economics lecture where the usefullness of insurance was discussed. You find out that the life insurance company asks for a premium of $100 per year for an insurance cover of $100.000. Fortunately they do not ask whether you intend to climb the Mount Everest. This is an offer you cannot reject. With a chance of death of 3.3% and a premiun rate of 0.1% you decide to insure yourself with a sum of $2 Mio. Not that this is the money your family indeed needs in case of an accident, but as insurance is so cheap, why not buy more of it? Once you have started acquiring life insurance, you go on buying accident insurance, disability insurance, etc.

By doing so, you are just this kind of person the theory of adverse selection tries to deal with. Individuals, who know their own risk type better than the insurance company, use this knowledge when they buy an insurance contract. In the following we will investigate, how insurance companies might react to the phenomenon of adverse selection, and what role if any the state could play.

Adverse selection does not only arise in life insurance markets but pertains to all areas of the insurance industry: people know better whether they are reckless or careful drivers, whether they have healthy lifestyles or not, whether their property is well equipped against earthquakes or not. Although these examples contain some element of self-control, for example even a reckless driver might try to drive carefully, here we only consider that individuals differ a priori, i.e., before they acquire insurance. Influence over the risk probabilities will be discussed in the next part of this book, where we turn to moral hazard.

As with all of the literature on asymmetric information, insurance markets are only one example of their applicability. Banks handing out credits do not know the profitability of the projects, which however the creditor knows. Governments procurring defence equipment from the private sector have limited information on the costs of production which the firms themselves know much better.

Although we concentrate on the insurance sector, several of the ideas and approaches can be generalized to other areas in economics.

We will start with a perfect competitive market which is the basis for most of the discussion of adverse selection in the insurance literature. Then we turn to a monopoly insurer. That model is helpful as it provides the formal setup of most principal agent models with adverse selection in the literature. We then consider further issues like categorical discrimination, endogenous information acquisition, long term contracts and renegotiation.

Chapter 11

Adverse selection in competitive insurance markets

11.1 The basic model

Adverse selection is defined as the situation where the individual has better information about his risk type than the insurer. We then say that the individual risk is his private information. For simplicity, we concentrate on two types only: High risks and low risks, with risk probability π_h and $\pi_l < \pi_h$ of losing a sum L. The insurer only knows the ratio of high risks to low risks in society. This is given by $\gamma_h/(1-\gamma_h)$ so that the average risk in society is: $\gamma_h\pi_h + (1-\gamma_h)\pi_l$.

The expected utility of an individual of type i if he buys an insurance contract (pI, I) is:

$$EU_i(pI, I) = (1-\pi_i)U(W - pI) + \pi_i U(W - L + (1-p)I) \tag{11.1}$$

Here p is the premium rate, I is the amount of cover, W is the initial wealth of the individual. With probability $(1-\pi_i)$ no accident happens, and the insured has to pay the premium only. With probability π_i an accident occurs, the individual looses L, but is reimbursed the amount I. Note that even in this case the insured has to pay the premium.

As we know from the chapter on insurance demand under perfect information, if the insured chooses his optimal insurance cover I^* then I^* will be larger (smaller) than L if π is larger (smaller) than p. In Figure 11.1 this is shown.

On the two axes are the incomes in the two states (no-accident, accident), point $E = (W, W-L)$ describes the initial endowment without insurance. The two solid lines are the zero profit lines for each type, which have the slope $-(1-\pi_i)/\pi_i$ for $i = h, l$. If the contract line lies somewhere in between the two lines (the dotted line), then the high risks will overinsure (point H) while the low risks will optimally underinsure (point L).

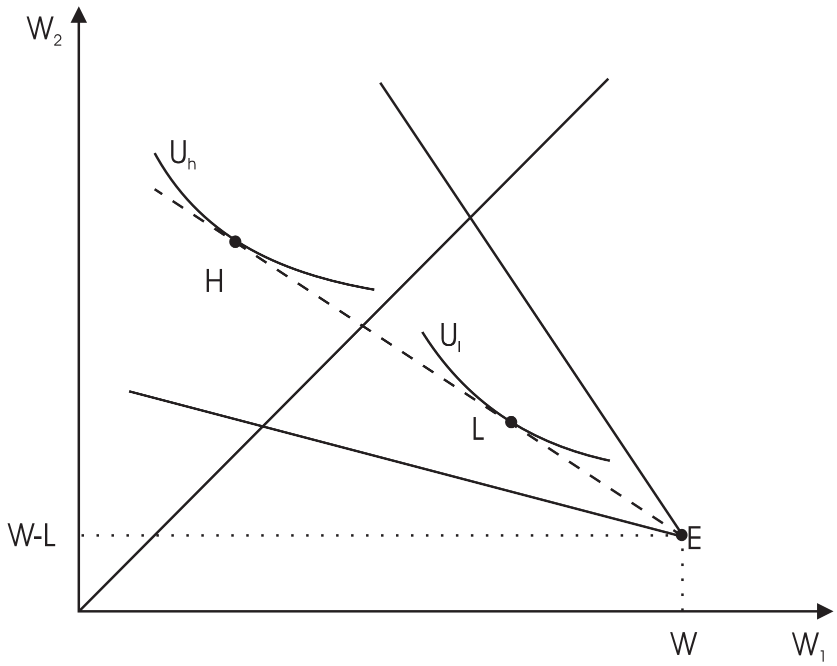

Figure 11.1: Optimal insurance cover.

Both types still face an income risk after the insurance purchase. U_h and U_l denote the indifference curves of the two types.[1] Note that due to this over- and underinsurance, a 'fair' premium rate $p = \gamma_h \pi_h + (1 - \gamma_h)\pi_l$ will lead to a loss for the insurance companies. Thus there are two inefficiencies arising: First, individuals do not buy the efficient amount of insurance, i.e., they either over- or underinsure. Second, prices have to be larger than the average risk probability to avoid loss. If there is a continuum of types, the insurance market may actually break down, as for any premium rate those lower risks who still buy a contract are not sufficient to subsidize the losses inflicted by the high risks. This argument works exactly as in Akerlof's (1970) famous 'lemons market' where the quality of the good is unknown. In such a case, governmental intervention might be useful, by, for example, obliging everyone to buy full insurance.

The analysis so far was restricted since it only concentrated on premium rates as the instrument available to the insurance companies. This may perhaps hold for life insurance, where individuals can buy several contracts from different companies,[2] but in other areas

[1] It is well known that along the certainty line, the 45 degree line, the slope of the indifference curves is $-(1 - \pi_i)/\pi_i$. Therefore the high (low) risk indifference curve is tangential to the dotted line to the left (right) of the certainty line.

[2] In Kifmann and Wambach (1999) we analyse in how far long term contracts and combined pension and life insurance contracts might be used by the life insurer to weaken the adverse selection problem (see exercise 11.1.).

of insurance markets this assumption does not necessarily hold. It was Rothschild and Stiglitz (1976) who first modelled firms which offer contracts specifying both premium and amount of indemnity as a reaction to adverse selection. For example in the case of car insurance, concepts like deductible or partial insurance cover are meaningful, and, as we will see, are very useful in dealing with the problems of asymmetric information.

In the following we allow insurance companies to set menus of price/indemnity contracts. An implicit assumption of the analysis is that individuals buy only one contract with only one insurance company. This may be achieved through a clause in the contract or through legal requirement, which e.g. forbids overinsurance. This point is discussed later on in more detail. The presentation is based on the work by Rothschild and Stiglitz, but differs from theirs in two respects: First, the modern terminology of a game, rather than a specific notion of equilibrium, is used. Second, firms are allowed to set a menu of contracts, while in Rothschild and Stiglitz' original paper firms which offer a single contract only were considered.

Consider the following game: As before, there are two types in society with N individuals. The risk probability of the types is π_i $(i = l, h)$ with $\pi_h > \pi_l$. The proportion of the h type is γ_h.[3] Both types have the same concave utility function $U(W)$. If uninsured, they obtain W if no loss occurs, and $W - L$ in case of an accident. There are $M \geq 2$ risk neutral firms in the market. The game proceeds as follows: At Stage 1, each firm i offers a menu of contracts $\{\omega_i^k = (P_i^k, I_i^k), k = 1, 2, ...\}$ which specify premium and indemnity.[4] At Stage 2, each customer chooses one of the contracts which are optimal for him, if any. If more than one firm offer this contract, individuals split between the firms equally. Then nature decides for each individual whether an accident occurs or not. Payments are made accordingly.

Before describing the equilibrium, one property has to be introduced:

Single Crossing Property: For every contract ω, the slope of the indifference curve of the low risks in a two-states-of-nature diagram is steeper than the slope of the high risks.

The single crossing property implies that indifference curves cross once only. This is usually assumed in all principal agent models, and it naturally holds in the present case, as $\frac{(1-\pi_h)U'(W_1)}{\pi_h U'(W_2)} < \frac{(1-\pi_l)U'(W_1)}{\pi_l U'(W_2)}$, where W_1 (W_2) is the income in the no-accident (accident) state. However, in exercise 11.3. it will be shown that, if individuals differ in their wealth

[3]In the present model, the insured do not act strategically, they just choose the best contract available. Therefore it suffices to assume that out of a population of N, $\gamma_h N$ types are high risks.

[4]It can easily be seen that contracts with random payments are never optimal, as they just confer some expected utility to the insured in case of an accident. This expected utility is the same for both types, so both types would be willing to pay the same amount of money in the accident state to avoid this uncertainty (see exercise 11.2.).

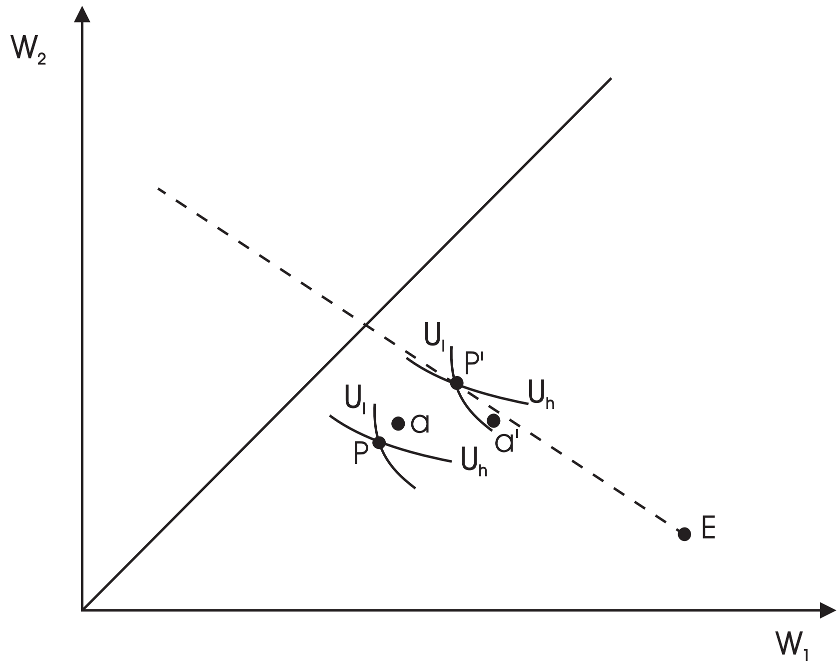

Figure 11.2: Non-existence of a pooling equilibrium.

or some other characteristic in addition to their risks, the single crossing property may be violated (see also Wambach, 2000). Then the results may well be different from those presented here.

Now we turn to the Nash-equilibrium where we concentrate on a symmetric equilibrium in pure strategies for the firms, where all customers of the same type choose the same contract. Mixed strategies, asymmetric equilibria and customers mixing between contracts will be discussed later.

The contracts offered in equilibrium define a set of contracts $\Omega = \{\omega_1, \omega_2, ..., \omega_n\}$. If all individuals of a certain type choose the same contract, then at most two active contracts are offered in equilibrium. Without loss of generality we do not consider that firms offer idle contracts as well.

Assuming that an equilibrium exists, five steps are required to derive the equilibrium contracts:

1. Non-existence of a pooling equilibrium:
Suppose $\Omega = \{\omega_p\}$, i.e., all firms offer the same contract in equilibrium. The corresponding final wealth of the insured is shown in Figure 11.2 (point P). For that to be a feasible outcome it has to lie below or on the pooling-zero-profit line (the dotted line), which is given by $W_2 = W_1 - L + I$ with $W_1 = W - P$ and $P = [\gamma_h \pi_h + (1 - \gamma_h)\pi_l]I$. Otherwise the firms would make a loss, which cannot be an equilibrium. Assume first, that this

outcome lies strictly below the pooling-zero-profit line. Then one firm might offer another contract with a slightly lower premium which leads to outcome a (see Figure 11.2) with which it attracts all customers. If a is close to P then this is surely better than also offering ω_p and only taking $1/M$ of the customers.[5] Therefore a point like P cannot be an equilibrium. Next assume that the outcome lies on the pooling zero profit line (P' in Figure 11.2). Due to the single crossing property, the indifference curves of the low and the high risk types cross each other at this point. From this it follows that there exist contracts which lead to outcomes like point a'. If a deviating firm offers such a contract while all the others still offer P' it will only attract the low risk types. As this contract would give approximately zero profit if taken by both risk types, it makes a profit if only the low risks buy it. Hence, we note as a first result that no pooling contract can be the equilibrium outcome.

Therefore suppose that two contracts are offered in equilibrium: $\Omega = \{\omega_l, \omega_h\}$, where the first is taken by the low risks, while the latter is taken by the high risks.

2. No contract makes a loss in equilibrium.

This is easy to see: if one of the two contracts makes a loss, then for every firm it is strictly better not to offer this contract, as long as the others still offer it. Note that this argument does not work in an asymmetric equilibrium: If only one firm offers a loss-making contract, then by withdrawing this contract, individuals would choose a different contract at Stage 2, which might change the profitability of the other contracts this firm offers. We return to this point later. As no contract can make a loss in a symmetric equilibrium, cross-subsidizing contracts, where for example the contract for the high risks makes a loss, while that for the low risks makes a profit, are ruled out. This is a very important point to keep in mind, because as we will see later on, cross-subsidizing contracts may be second best efficient.

3. No contract makes a profit in equilibrium.

Suppose that the contract for the low risks makes positive profits. There are two possible cases: Either the high risk types are indifferent between their contract and that for the low risks, or they are not. In the latter case, offering the low risks a slightly better deal, will attract all low risks and make approximately the same profit per insured. This is again the standard Bertrand argument. Therefore assume that the high risks are indifferent between the two contracts. Due to the single crossing property, there still exists a contract in the vicinity of ω_l such that only the low risks prefer this new contract. A firm offering this new contract would then again attract all low risks instead of only $1/M$ if it were to stick

[5]This argument is the same as that used for the proof that in the standard Bertrand oligopoly two firms are enough to restore perfect competition.

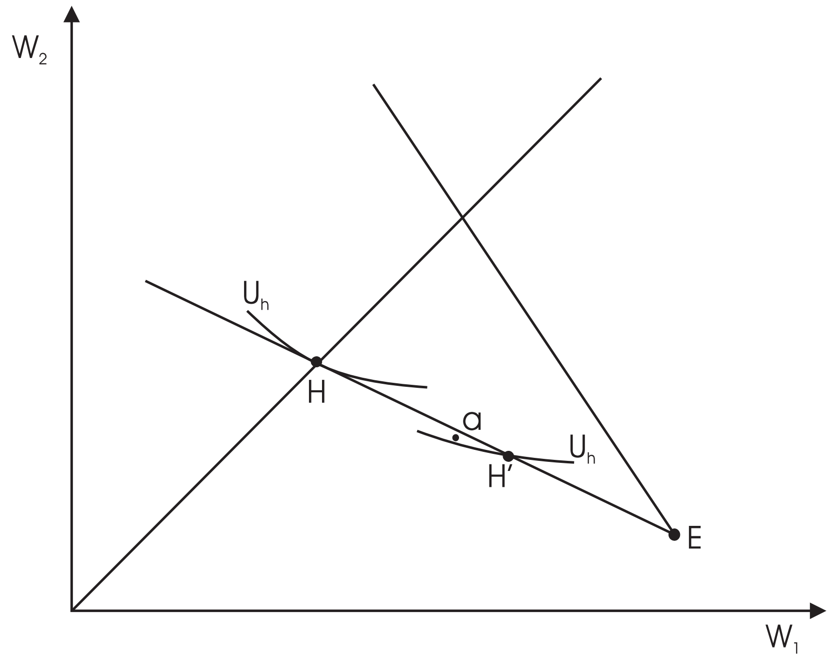

Figure 11.3: High risks obtain full insurance.

with the old contract and make a profit with those. Note that with a similar argument, profit making contracts for the high risks can be excluded.

4. The high risks obtain full insurance at the high risk fair premium.

As a result from steps 2 and 3 we know that the outcome for the high risks must lie on the high risks zero profit line. Suppose this point were at H' as drawn in Figure 11.3. As the indifference curve of the high risks cuts the zero profit line (remember that it is tangential only at the point of full insurance) offering a contract which leads to a would attract all high risks and lead to a profit for the company. Why are the low risks not a problem? We know that the contract for the low risks makes zero profit in equilibrium. Thus, even if the low risks now prefer a to their contract, so that they all switch to a once it is offered, that is good news for the deviating company, as it will make a profit with the low risks at this contract as well. The same argument works at a point of overinsurance, because also there, the high risks indifference curves cut the zero profit line. There is still scope for profitable deviations. The only possible outcome is at the point of full insurance where the indifference curve is tangential to the zero profit line (point H).

5. The low risks obtain partial insurance at their fair premium. The contract is such that the high risks are just indifferent between their contract and that for the low risks.

This can be seen in Figure 11.4. If the high risks are fully insured (outcome H), and the low risks receive a contract on their zero profit line, then outcome L is the best one can do for them. Any contract with less partial insurance (those outcomes which lie above

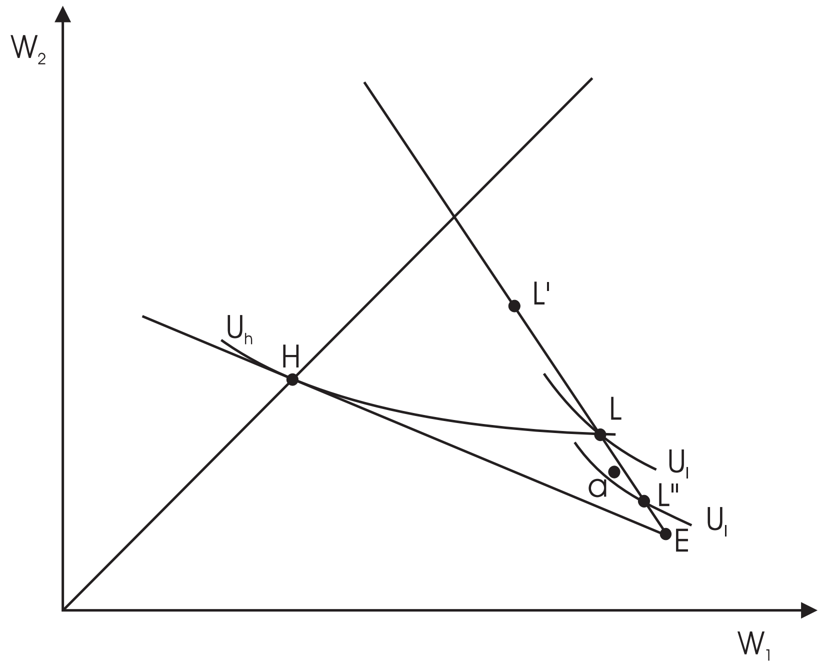

Figure 11.4: Low risks obtain partial insurance.

point L, e.g. L') are also preferred by the high risks. Therefore a firm offering such a contract would attract the high risks with whom it would make a loss. For any contract on the zero profit line below L (e.g. L''), there exists other contracts which all low risks and no high risk prefer and with which a firm can make a profit (e.g. a).

This finishes the proof: contracts $\{\omega_l^{RS}, \omega_h^{RS}\}$ which lead to outcomes $\{L, H\}$ are the famous Rothschild-Stiglitz (RS) contracts[6]: The high risks receive full insurance at their fair premium, while the low risks obtain partial insurance at their fair premium, and the high risks are indifferent between the two contracts.

This analysis is the basis for most of the work which has been done in the context of adverse selection in the insurance market. It displays two features we observe with many real world contracts: First, for a specific risk more than one contract is offered. In automobile insurance, for example, one can choose between different levels of deductible. This is usually also the case for health insurance contracts. To stress this point, offering different amount of cover for different premia is a typical sign of adverse selection. In a competitive industry, different contracts are only rationalizable if types differ in an unobservable characteristic which influences the profitability of a contract. Second, partial insurance is offered at a lower premium rate than the full insurance contract. For example, Puelz and Snow (1994) found that in a car insurance market where different deductibles

[6]These contracts are also known as least cost separating contracts.

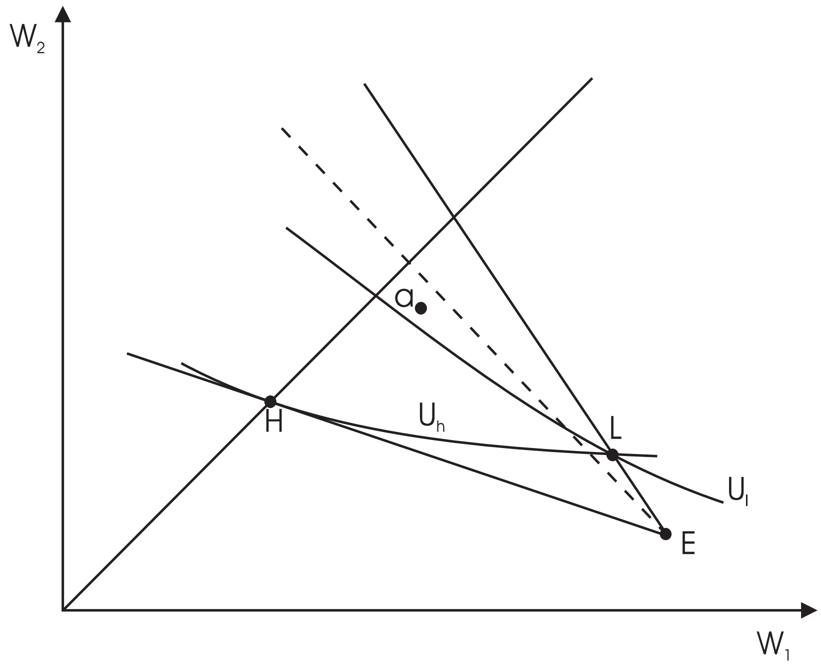

Figure 11.5: Pooling contracts dominate the RS contracts.

were offered, the average premium declined with the deductible chosen. Interestingly, they also obtained that there was only insignificant cross-subsidization between contracts. A result which one would expect from a Rothschild-Stiglitz equilibrium.

The structure of the RS contracts is very intuitive and gives a clear indication of possible market reactions to different risk types. Unfortunately, the discussion has one very serious limitation: The outcome may not be an equilibrium.

How could that be the case? The easiest way to see this is to show that in some cases a pooling contract might be such that it is preferred by both risk types to their RS contracts, while still making a profit. Such an outcome is drawn in Figure 11.5 (point a). This holds if there are sufficiently few high risks, which makes the pooling line (the dotted line) lie close to the low risk zero profit line. Then the low risks prefer a contract on the pooling line to their RS-contract. A firm offering a contract shortly below the pooling line would attract all customers and make a strictly positive profit. But, as shown in step 1 above, a pooling contract cannot be an equilibrium. Therefore in that case we have to conclude that there does not exist an equilibrium under the assumptions we made on the strategies. This is a serious problem, as for example in the health insurance market, one would expect that the ratio of high risks to low risks is relatively small.

Even if there is no profit making pooling contract which is better for the low risks than ω_l^{RS}, a pair of cross-subsidizing contracts (H', L') might be a profitable deviation,

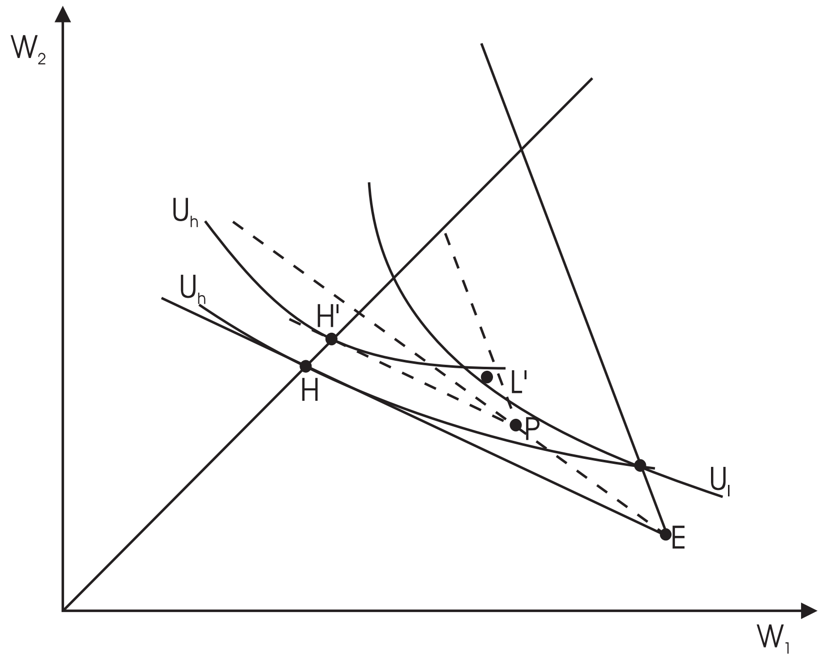

Figure 11.6: Cross-subsidzing contracts dominate the RS contracts.

as shown in Figure 11.6. If a firm offers such a pair of contracts, it would make a loss with the contract for the high risks, but a profit with the contract for the low risks. Then if there are (again) sufficiently many low risks, both risk types can be made better off, while the firm makes a positive profit overall. In Figure 11.6, P lies on the pooling zero profit line. Therefore any contract on the dotted line between P and H', which is the iso-cost curve for the high risks, gives the insurer the same loss with a high risk type. With outcome L' slightly below the iso-cost curve for the low risks and below the high risk indifference curve which goes through H', the combination (H', L') leads to a profit overall, if all consumers choose these two contracts. But, as shown in steps 2 and 3 above, profit making or loss making contracts cannot be equilibrium contracts. Thus also here we have to conclude that no equilibrium exists.

In Rothschild and Stiglitz original work, this cross-subsidization was not considered to be a problem, as only single-contract deviations were possible. But note that the pooling contract can also be seen as a degenerate pair of cross-subsidizing contracts. To summarize the results: If γ_h is sufficiently large, the RS contracts constitute a Nash equilibrium. If γ_h is smaller than some γ', a pair of cross-subsidizing contracts destabilizes the RS contract, so a Nash-equilibrium in pure strategies does not exist. For even lower $\gamma_h < \gamma'' < \gamma'$, a single pooling contract will be preferred by both types to the RS contracts.

11.2 The "equilibrium-non-existence" debate

To be precise, although the literature talks about the equilibrium-non-existence problem, so far we have only shown that an equilibrium where firms have symmetric and pure strategies, and all consumers of one type choose the same contract, does not exist. Let us therefore relax these assumptions:

Symmetric strategies on side of the firm: The only point in our proof of the RS equilibrium where this assumption played a role, was when we showed that loss-making contracts cannot exist in equilibrium. Now, if only one firm offers a loss-making contract which, say, is taken by the high risks, then by not offering such a contract, all high risks may choose another contract and could inflict a loss upon those firms offering this contract. Therefore the firm may perhaps rationally include the loss making contract in its menu of contracts. However, for that firm not to make a loss altogether, it must offer a contract to the other type, in this case the low risks, with which it makes a profit. But here, the same 'Bertrand-dynamic' as in step 3 above comes in: Due to the single crossing property, there always exists another contract for the low risks with slightly better terms which the high risks do not prefer. Another firm offering this contract will attract all low risks and make a profit with those. Therefore such a constellation breaks down. This point is worth stressing: Usually the argument brought forward against cross-subsidizing contracts in a competitive market is that firms would withdraw the loss making contract (as we did in step 2 in the previous section). In the case of asymmetric contract offers, however, cross-subsidizing does not work because the other firms do not allow the firm with the loss-making contract to recover its losses. It is the profit making contract which is not stable in a competitive environment.

All customers of one type choose the same contract: Suppose one type of customer mixes between contracts. This implies that all the contracts have to lie on the same indifference curve for this type. Then in general the firms make different profits per person with each of these contracts. But, with a similar proof as in step 2 and 3, neither loss making nor profit making contracts are possible in equilibrium. Therefore at most two contracts on the zero-profit line, one with underinsurance, the other with overinsurance are conceivable. However, if no incentive constraint binds at this point, full insurance is strictly better (as was shown in step 4). If, on the other hand, the incentive constraint binds such that full insurance is not feasible, then either the overinsurance contract (in case of the low risks) or the underinsurance contract (in case of the high risks) must have violated the incentive constraint (draw it in a diagram to see the argument). Thus customers of one type choosing different contracts cannot be an equilibrium.

Firms have mixed strategies: This is the formal game theoretic solution to the non-existence problem. In the case of two firms, which set two contracts each, Dasgupta and Maskin (1986) have shown that an equilibrium in mixed strategies exists if the RS contracts do not constitute an equilibrium. In this context, mixed strategies mean that each firm offers different sets of two contracts, each with some probability. The exact equilibrium is not known, however some information on the equilibrium can be obtained: First, firms make zero expected profit; second, with any contract pair offered, the high risks obtain full insurance at a fair or better premium and the low risks obtain partial insurance at an unfair premium. However, the economic interpretation of an equilibrium in mixed strategies is unclear: Are firms supposed to be randomizing over contracts each year/each day? In many contexts, mixed strategies are a sensible concept to use. As a description of the strategic interaction of an insurance markets, however, mixed strategies are more an indication for the limitation of our model. Perhaps it is too simplistic to assume that firms only offer contracts out of which customers choose the best one available. We have to look for more sophisticated games.

Formerly this was not done by extending the game structure, but by assuming different equilibrium concepts. The Rothschild and Stiglitz equilibrium definition is that *there is no contract outside the equilibrium set that, if offered, makes a profit.* (We already extended this to a menu of contracts.) In Wilson's equilibrium concept (1977), *every additional contract should stay profitable even if those contracts which make a loss after the introduction of the new contract, are withdrawn.* It is easy to see that in this case a pooling contract might survive in equilibrium: Consider step 1 from above again: Pooling was unstable because someone could offer a contract only to the low risks, i.e. to 'skim off' the good risks. However, in the Wilson concept, if someone tries to attract the low risks only, all others will withdraw their loss making pooling contract, because that contract would be bought by high risks only. Therefore also the high risks choose this newly offered contract, which makes it much less attractive to offer it in the first place. Wilson has shown that if the number of high risks is sufficiently large, his equilibrium coincides with the RS contracts. In exercise 11.4. you are asked to show that this so-called Wilson E2 equilibrium is a partial insurance contract on the pooling zero-profit line where the low risks indifference curve is tangential to that line,i.e. the best zero-profit pooling contract from the point of view of the low risks.

Extending Wilson, Miyazaki (1977) and Spence (1978) allow in addition that firms offer more than one contract. Therefore cross-subsidization between contracts becomes possible. This leads to the so-called WMS equilibrium, which is the solution to the following maximization problem:

$$
\begin{aligned}
&\max_{P_l,I_l,P_h,I_h} (1-\pi_l)U(w-P_l)+\pi_l U(w-P_l-L+I_l)\\
&\text{s.t.}\\
&(1-\pi_h)U(w-P_h)+\pi_h U(w-P_h-L+I_h)\geq(1-\pi_h)U(w-P_l)+\pi_h U(w-P_l-L+I_l)\\
&\gamma_h(P_h-\pi_h L)+(1-\gamma_h)(P_l-\pi_l I_l)\geq 0
\end{aligned}
\tag{11.2}
$$

The utility of the low risk type is maximized under the constraints that the high risks will not buy the contract designed for the low risks (incentive constraint) and that the firms make non-negative profit overall. The solution to this problem will be discussed in more detail in the section on categorical discrimination. Here it suffices to note that the high risks will always obtain full insurance, while the low risks obtain partial insurance. If γ_h is sufficiently large, then the WMS equilibrium corresponds to the Rothschild-Stiglitz outcome. If not, then the solution to the above maximization problem is a pair of cross-subsidizing contracts but never a pooling contract. As will be discussed in the section on categorical discrimination, the WMS contracts are second best efficient. There does not exist any other set of contracts which makes no-one worse off and someone better off, given the asymmetric information.

A different equilibrium concept was introduced by Riley (1979). In his *reactive equilibrium*, firms shy away from offering deviating contracts if another insurance company would react to such an offer by skimming off the desirable types. While in the Wilson concept firms anticipate that other firms will withdraw contracts as a result of their entry, here the deviating firms anticipate that at least one other firm will react by offering an additional contract. In that case, the Rothschild-Stiglitz outcome is stable. No-one deviates by offering a pooling contract or a pair of cross-subsidizing contracts as in both cases some other firm will profitably 'skim off' the low risk types.

The WMS equilibrium is attractive from an economic point of view, as the contracts are second best efficient. That is what a competitive market is expected to lead to: Pareto efficient outcomes. It is this feature of the WMS equilibrium which makes it quite popular in the insurance literature. On the other hand, the Riley concept rationalizes the Rothschild-Stiglitz outcome even if it does not constitute a Nash equilibrium. In both cases, however, by introducing new and to some degree arbitrary equilibrium conditions not much is achieved. We still have to look for a fully specified game to understand the economics of the insurance market: Why should a firm withdraw its contract in response to other firms' entry? How long does it take to withdraw? Why do only deviators fear responses, why do not those firms offering the equilibrium contracts fear deviators? What are other possible ways to interact strategically on the insurance market?

Both in the concepts of WMS and Riley, some form of dynamics, namely the possible reaction of firms after the contracts have been offered are considered. In the remainder of this chapter we will discuss a few selected models where this dynamic aspect is explicitly modelled as part of more elaborated games.

Hellwig (1987) introduced a third stage in the model described above. Again, firms offer contracts at Stage 1, then customers choose at Stage 2, but now firms can withdraw some (or all) of their contracts at Stage 3. This comes close to the Wilson concept, as firms now have the ability to withdraw some contracts, depending on what the other firms offered, and what the customers choose. The actual Nash equilibrium of this game is difficult to determine, as now customers by choosing contracts at Stage 2 reveal information which might be used in Stage 3. This is therefore a combination of a screening (by the firms) and signalling (by the customers) model. Those readers who are familiar with signalling models know that usually many equilibria are possible, depending on the out-of-equilibrium beliefs. This is not much different in this case, however, under specific belief refinements, only the Wilson pooling contract is robust, whenever the RS equilibrium does not exist. This approach is very useful as it explicitly attempts to model the equilibrium concept of Wilson. It shows, that an equilibrium in pure strategies always exists. However, whether the possibility of withdrawing an accepted contract is really the describing characteristic of an insurance market remains an open question.[7]

In Asheim and Nilssen (1996) the possibility of withdrawing contracts at the third stage is dismissed, but firms can now offer new contracts to their existing customers, who then either stick with their contract or choose the best new one on offer. Although this sounds a little bit like the Riley concept, it differs in so far as firms can only offer contracts to their own customers, and not to the whole market. The motivation for this model is the idea, that the insurance firm can renegotiate the contracts with its own customers. At Stage 3, any firm does not compete with the others for its customers anymore. Therefore it can offer cross-subsidizing contracts to its own customers as long as these contracts give them larger utility than the contract which made them sign in the first place. There is no danger of 'cream-skiming' by the other firms. It can be shown that, overall, the WMS outcome as the final contracts is the unique equilibrium of this game. This approach provides a justification for the use of the WMS equilibrium in insurance models. One

[7] An alternative model would be to let customers send some form of signal first, after which the firms make their contract offers, after which the customers choose the preferred contract (see Cho and Kreps, 1987). In that case, under appropriate belief refinements, the separating contracts of the Rothschild Stiglitz type are the equilibrium contracts. In the insurance market it is however not clear, what kind of signal the customers might give at Stage 1. In other models, like for example the credit market, this signal could be a collateral.

might criticise, however, that once renegotiation is explicitly introduced, it is not clear why the new contracts have to be offered to all customers. If customers have signed different contracts it is well conceivable that the new contracts differ, depending on which contract the individual has already signed. This renegotiation issue is a serious critique to all of the models presented here and we will devote a whole section to it later.

So far it is always assumed that some exclusivity condition can be enforced: Individuals only buy one insurance contract. But it is not obvious that this information is readily available to the insurance company: Who tells them whether there exists an additional policy or not? As suggested by Jaynes (1978) and later developed in game form by Hellwig (1988), the incentive insurance firms have to share or conceal information about their customers might be another strategic instrument available. This is modelled in a four stage game: In the first stage, firms offer contracts and decide whether or not an exclusivity requirement is attached to this contract. In the second stage, consumers choose a combination of contracts. Then firms at the third stage decide what information if any they want to divulge to which firms and at the fourth stage, they choose depending on the information they received whether or not to enforce the exclusivity condition. In equilibrium, customers buy two types of contract: The Wilson pooling contract is sold by firms who exchange information with each other. This policy is bought by everyone. The high risks amend this contract with a partial insurance contract at the high risk fair premium, such that they obtain full insurance. This latter additional contract is bought from firms who do not reveal information about their customers. Although this gives the impression that the endogenous treatment of inter-firm communication is the key point in this model, Hellwig stresses that the "sequential specification of firm behaviour which allows each firm to react to the other firms contract offers" solves the equilibrium problem. One interesting aspect of this result is that firms do not screen the market. The contract sold to the low risks is a pooling contract which all the high risks buy as well. Then the high risks obtain additional coverage from different insurers. An example for this could be seen in the German health insurance market: Everyone buys the same standard insurance package, while only some acquire additional coverage, the so-called 'add-on insurance' ("Zusatzversicherung").

A more direct solution to overcome the existence problem has been provided by Inderst and Wambach (1999). The authors assume that firms face capacity constraints, which might result from limited capital which is available to the insurer. In that case, by offering deviating contracts an insurer cannot be sure that it obtains the mix of risk types it desires, as not everyone will turn up at this firm. Under some assumptions on the severity of the capacity constraint and on the costs the customers face if they are rationed it is shown that indeed only the high risks will turn up if someone offers a deviating pooling contract or a pair of cross-subsidizing contracts. The reason is that due to the single crossing property the high risks gain much more from a deviating contract, so they are more willing to endure the rationing which will occur at the deviating insurer. Therefore no firm has an incentive to deviate, which implies that the RS contracts are always an equilibrium outcome of the game.

As a last example on how to overcome the non-existence problem we turn to evolutionary game theory. Ania et al. (1998) relax the assumptions that firms have perfect knowledge on the utility functions of the customers, their risk types, the number of different risk types, etc. Instead the authors assume that firms offer contract menus and imitate successful behaviour, i.e., in every period they observe the most profitable contracts on the market and copy those. In addition, once in a while they experiment with their own contracts (which is called 'mutation' in the literature). Experimentation and mutation both stand for different explanations of this dynamical feature: Either firms are supposed to experiment, trying to find ways to increase their profits or market shares by offering new contracts. Or they mutate, which means that they make mistakes in pricing their contracts, thus offering new ones by accident. Two results are shown: First, if no profit making pooling contract is better for the low risks than the RS contract, then the RS contracts are the long run outcome of this evolutionary game. If a pooling contract is preferred, then the RS contracts are still the long run outcome if experimentation takes place only locally, i.e. firms only add contracts close to the existing ones. The first result is interesting as it shows that even without detailed information firms can learn to offer screening contracts. Furthermore, in an evolutionary context, possible deviations via cross-subsidizing contracts are not a problem. Firms will quickly copy the profit making part of the cross-subsidzing pair of contracts while the first firm to offer this set of policies withdraws the loss making one. Then the system works itself back to the RS outcome. The second result points to the destabilizing force of pooling contracts which the RS contracts do not share: Pooling contracts can be destabilized by small changes in the contract structure. This evolutionary model is explicitly dynamic as it discusses the very long run outcome. It is limited as neither strategic contract settings from side of the firms, nor strategic choice of contracts from side of the customers is considered.

To summarize this section: The non-existence of an equilibrium in pure strategies of the Rothschild-Stiglitz model is still, after more than two decades, a puzzling problem in the insurance literature. To remedy it the simple two-stage game has to be extended. So far, there are good reasons to justify the RS outcome, the Wilson pooling outcome, the WMS cross-subsidizing contracts, and even a combination of Wilson and full insurance contracts.

Chapter 12

Further issues: Monopoly insurer, categorical discrimination, endogenous information acquisition

12.1 Monopoly insurer under adverse selection

A monopoly insurer is useful to consider as the formal setup is very similar to that used in most of the principal agent literature. In the latter case, a principal offers a set of contracts to one agent who has private information about his type while here, the set of contracts is offered to a population which consists of several types. In both cases, the formal expression for the (expected) profit of the principal is the same.

The monopoly insurer faces a population of two different risk types with proportion γ_h of high risks. The insurer is risk neutral while the individuals are risk averse.[1]

Before writing down a maximization problem we have to consider one further aspect, the so-called revelation principle. So far we assumed that the firms offer two contracts out of which each type chooses that one which is optimal for him, i.e. each type reveals himself. It is however not clear that this is indeed optimal: Why is it not better to have some individuals lying about their type? Are there no other 'mechanisms' which do not require truth telling, which are Pareto improving, or at least better for the principal? Fortunately, the answer is no. This is the result of the so-called *Revelation Principle*. This principle is discussed and proven elsewhere (e.g. Gravelle and Rees, 1992, pp.694-696), so we do not go into detail here. In short, this principle states that it suffices to consider truth-telling mechanisms only, as every other mechanism is equivalent to one where the agents reveal their type truthfully.

[1] The analysis on monopoly insurer under adverse selection was first done by Stiglitz (1977).

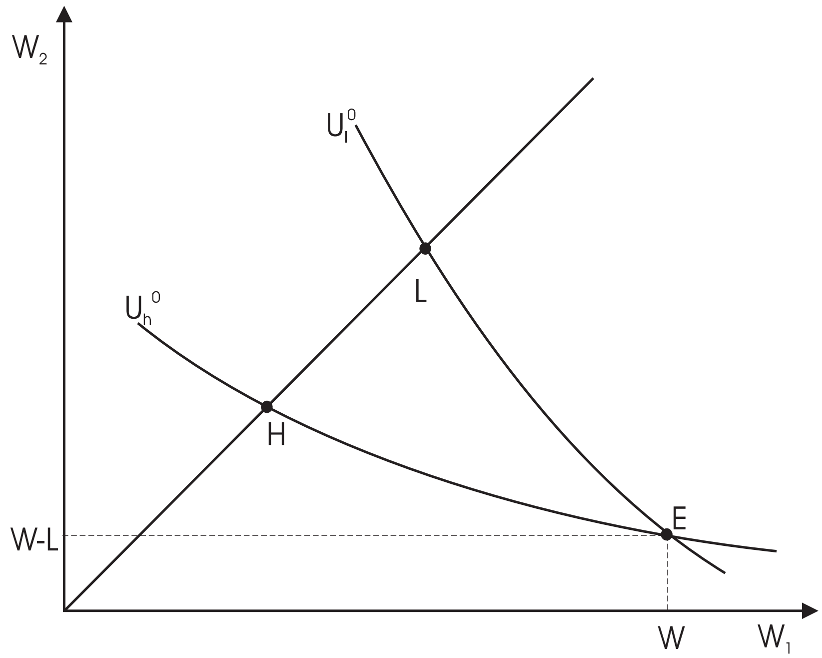

Figure 12.1: Insurance under symmetric information.

Let us first discuss the scenario under symmetric information. If the monopolist can observe the individual risk she would offer two contracts with full insurance as shown by points (H, L) in Figure 12.1. Both types are indifferent between insurance and no insurance. The indifference curves are denoted by U_i^0, $i \in \{l, h\}$. If such a pricing policy is possible, we would speak of price discrimination of degree one. It is quite clear that if information is private, everyone would claim to be a low risk type and choose outcome L if both L and H are offered. So let us now turn to asymmetric information.

We simplify notation by writing as $EU_i(P_i, I_i^n)$ the expected utility $(1-\pi_i)U(w-P_i)+\pi_i U(w-L+I_i^n)$. The expected utility if the individual has no insurance is written as $EU_i(0,0)$. Due to the 'Revelation Principle' we can write the maximization problem of the risk neutral monopolist as follows:

$$
\begin{array}{rcll}
\max_{(P_l, I_l^n, P_h, I_h^n)} & & \gamma_h N((1-\pi_h)P_h - \pi_h I_h^n) + & (1-\gamma_h)N((1-\pi_l)P_l - \pi_l I_l^n) \\
\text{s.t.} & & & \\
EU_l(P_l, I_l^n) & \geq & EU_l(0,0) & \text{PC (i)} \\
EU_h(P_h, I_h^n) & \geq & EU_h(0,0) & \text{PC (ii)} \\
EU_l(P_l, I_l^n) & \geq & EU_l(P_h, I_h^n) & \text{IC (i)} \\
EU_h(P_h, I_h^n) & \geq & EU_h(P_l, I_l^n) & \text{IC (ii)}
\end{array}
\tag{12.1}
$$

The monopolist maximizes her expected profit under four constraints: The first two are the *participation constraints (PC)*: Both agents must not be worse off, otherwise

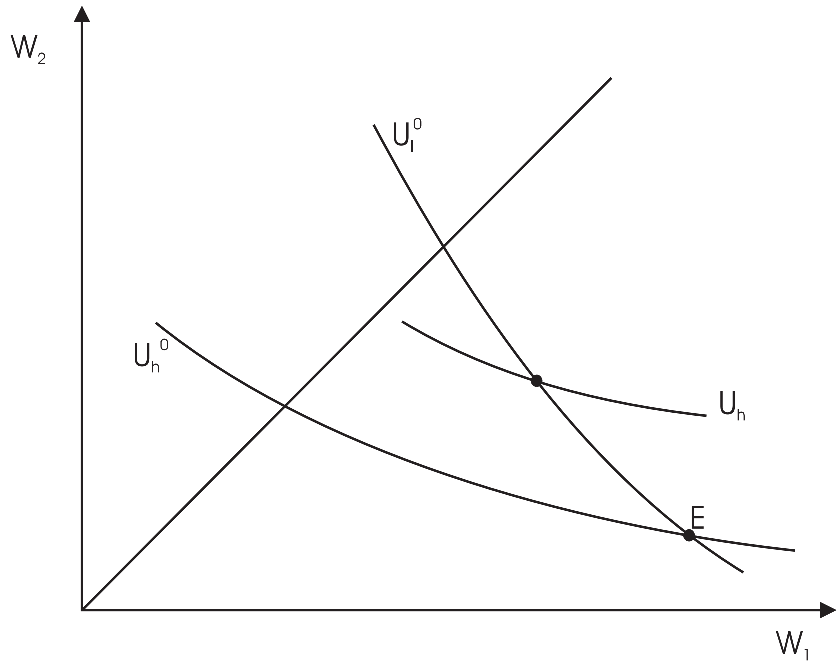

Figure 12.2: PC(ii) is not binding.

they would not buy insurance. The third and fourth constraints are the *self selection* or *incentive compatibility constraints (IC)*: Each agent must prefer to buy the policy which is intended for him to the policy which is for the other type.

We now go through a series of steps to work out the solution to this problem. First assume that two different contracts are offered in equilibrium.

1. Participation constraint (ii) is not binding.
See Figure 12.2. For all policies, where income is shifted from the state of no accident to the state of accident, the outside option utility indifference curve of the low risk type lies above that of the high risk. So any contract for the low risks gives the high risks a rent. IC(ii) then tells us that high risks must obtain a rent also with their contract.

2. Participation constraint (i) is binding.
See Figure 12.3. Suppose PC(i) is not binding (outcome L which lies on an indifference curve above the outside option utility level). Then the contract for the high risks must lie somewhere in the hatched region. There are two possibilities: Either the low risks are indifferent between their contract and that for the high risks, or they are not. In the latter case, by lowering the payment in the accident state the low risks obtain a contract which is still acceptable, and leads to a larger profit for the monopolist. In the former case, the indemnity for the high risks has to be lowered as well, such that the low risks still do not prefer the contract designed for the high risks. But as the high risks obtain a

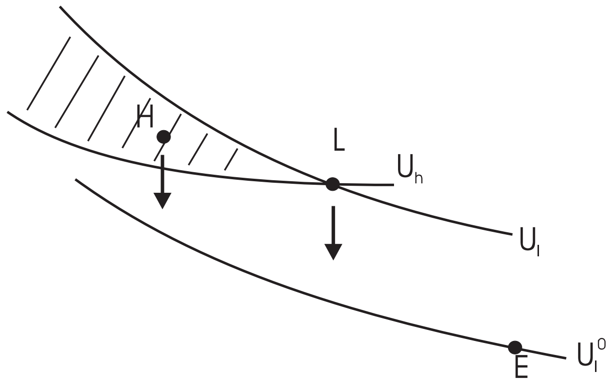

Figure 12.3: PC(i) is binding.

rent with their contract this does not violate any participation constraint.

3. Incentive constraint (i) is not binding.
We know that the low risks obtain a contract which does not give them a rent. If IC(i) binds, then the contract for the high risks would lie somewhere on the low risk indifference curve, as drawn in Figure 12.4 (point H). Take this contract as given: What would be the most profitable contract which could be offered to the low risks? One possibility is an outcome like L in Figure 12.4, where the indifference curve of the low risks is tangential to the low risk iso-profit line, which has the slope $-(1-\pi_l)/\pi_l$. But this is the point of full insurance. Alternatively, if that point lies to the left of H, a pooling contract will give the largest profit to the monopolist, i.e., $L = H$. Pooling contracts are discussed later on, so we ignore them for the moment. Given that L lies on the certainty line, the iso-profit line for the high risks (the dotted line) must be flatter at H than the indifference curve of the high risks. Therefore an outcome like a instead of H in Figure 12.4 would be bought by the high risks, does not violate any incentive constraint, and lead to a larger profit for the monopolist. Thus contract H could not have been optimal in the first place.

4. Incentive constraint (ii) is binding.
Suppose it were not binding. Then without this constraint, the maximization problem is the same as the one under full information, given that IC(i) does not bind. But in that case, we know that two full insurance contracts which do not satisfy IC (ii) are optimal. So this constraint has to be binding.

We have thus reduced the Kuhn-Tucker problem to a much simpler Lagrange problem

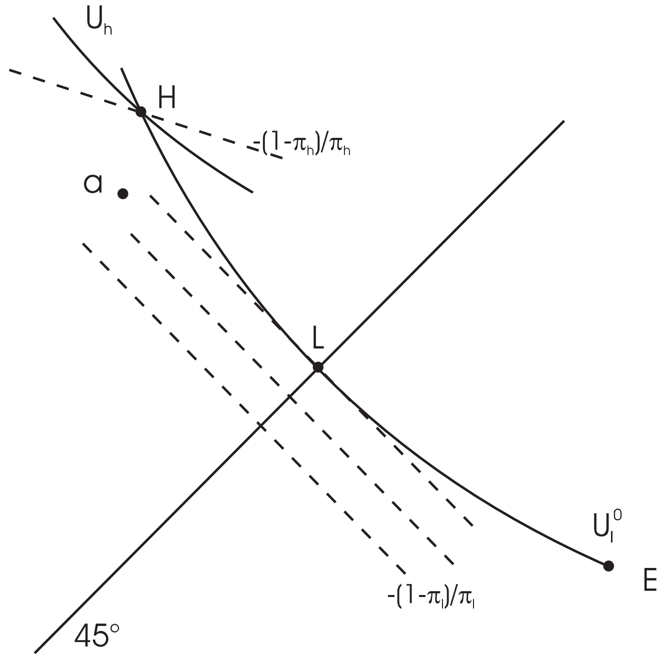

Figure 12.4: IC(i) is not binding.

with two binding constraints:

$$
\begin{array}{rcll}
\max_{(P_l, I_l^n, P_h, I_h^n)} & & \gamma_h N((1-\pi_h)P_h - \pi_h I_h^n) + & (1-\gamma_h)N((1-\pi_l)P_l - \pi_l I_l^n) \\
\text{s.t.} & & & \\
EU_l(P_l, I_l^n) & = & EU_l(0,0) & \text{PC} \\
EU_h(P_h, I_h^n) & = & EU_h(P_l, I_l^n) & \text{IC}
\end{array} \tag{12.2}
$$

Solving this problem leads to the first order conditions with respect to P_h, I_h^n:

$$
\begin{array}{l}
\gamma_h N(1-\pi_h) - \mu(1-\pi_h)U'(W-P_h) = 0 \\
-\gamma_h N\pi_h + \mu\pi_h U'(W-L+I_h^n) = 0
\end{array} \tag{12.3}
$$

where μ is the Lagrange parameter of the incentive constraint (IC). It directly follows that marginal utility of the high risks in both states of the world is the same: the high risks obtain full insurance. This can also be seen in Figure 12.5.

Take L, the contract for the low risks as given. As the high risks are indifferent between their policy and that of the low risks, the best the insurer can do is to offer that contract to the high risks where the indifference curve is tangential to the iso-profit curve. But this is the point of full insurance. This is the famous 'no-distortion-at-the-top' result. No low risk would like to claim to be a high risk and then buy the contract for the high risks. Therefore there is no need to distort the high risk contract to inefficient coverage levels. Such a result holds in all adverse selection problem. There is always one type, that one at

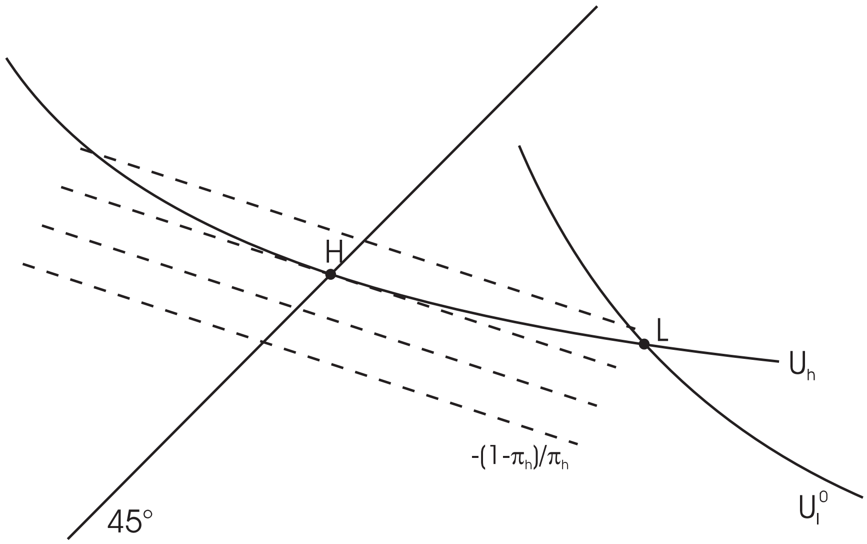

Figure 12.5: 'No distortion at the top'.

the top, whose contract specifies the efficient amount of whatever quantity is negotiated over.

Denoting by λ the Lagrange parameter of the participation constraint (PC), the first order conditions with respect to P_l, I_l^n become:

$$\begin{array}{l} (1-\gamma_h)N(1-\pi_l) - \lambda(1-\pi_l)U'(W-P_l) + \mu(1-\pi_h)U'(W-P_l) = 0 \\ -(1-\gamma_h)N\pi_l + \lambda\pi_l U'(W-L+I_l^n) - \mu\pi_h U'(W-L+I_l^n) = 0 \end{array} \tag{12.4}$$

Rearranging gives:

$$\begin{array}{r} U'(W-P_l) = \frac{(1-\gamma_h)N}{\lambda-\mu\frac{1-\pi_h}{1-\pi_l}} \\ U'(W-L+I_l^n) = \frac{(1-\gamma_h)N}{\lambda-\mu\frac{\pi_h}{\pi_l}} \end{array} \tag{12.5}$$

The only difference in the two expressions on the right hand side is the μ term in the denominator. As μ is positive and $\pi_h > \pi_l$, the marginal utility in state of no loss is smaller than the marginal utility in state of a loss. The low risks are underinsured.

This result also implies that pooling is never optimal. This can be seen in Figure 12.6. Suppose the principal would pool both types at a full insurance outcome P. If she instead offers the pair of outcomes (H, L) then the additional gain she obtains with the high risks is of the first order, while the reduction in profits with the low risks are of the second order. L is very close to the iso-profit line for the low risks which goes through P while the change in profit with the high risks (from P to H) is large. Thus it is always profitable not to pool.

To summarize the overall result: High risks obtain full insurance and receive a rent while the low risks obtain partial insurance without a rent. The degree of partial insurance

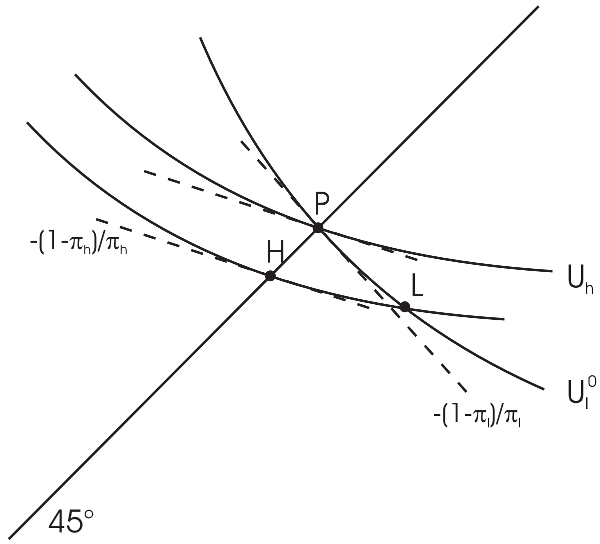

Figure 12.6: Pooling is never optimal.

depends on the ratio of high risks to low risks: If there are sufficiently many high risks the low risks might even not obtain any insurance at all. In that case, the monopolist only offers the full insurance contract to the high risks at which they do obtain their outside utility level.

As we are in a 'second best' world, the monopolist might try to use additional instruments to increase her profits. The same holds for the perfect competition case: If other means of screening are available, firms will use them to avoid the inefficiencies. One of these instruments will be discussed now: categorical discrimination.

12.2 Categorical discrimination

A common feature in the insurance markets is discrimination of customers: Young drivers pay a higher premium for automobile insurance. Furthermore, the premium depends on whether the car owner has a garage or not, what type of car he is driving, where he lives, etc. Health insurance premia are higher for older persons. They are higher for women, too.

On the other hand, usually one does not observe that the price of a good, like e.g. a TV set, depends on who buys it. In some cases, as for example railroad tickets, the price might depend on the age of the customer. But in general, this is perceived as a sign of market power. Usually it is optimal to charge price equal marginal costs. So what is different in the insurance sector if at all? Does discrimination, if it is not legally forbidden, develop endogenously in a market? Is there a need for governmental intervention? In this section we will tackle these questions.

Let us simplify the discussion by considering only two groups in society. For concreteness, call them males (with a proportion of λ_m) and females ($\lambda_f = 1 - \lambda_m$). As before, each individual faces a risk of losing L. If both parties have the same risk-probability, then in a competitive market without insurance costs the premium rate for both groups would be equal to their risk probability. In that case, discrimination does not change anything. So discrimination only makes sense if males and females differ in some relevant characteristic.

Assume that both groups have different risks, e.g. $\pi_m > \pi_f$. Again, in a competitive market, if discrimination is possible, both parties would receive full insurance at their fair premium, that is the males have to pay more than the females. If discrimination is forbidden, then the result depends on the equilibrium concept used. Here we concentrate only on the Rothschild-Stiglitz (RS) outcome and the Wilson-Miyazaki-Spence (WMS) equilibrium, as these are the concepts which are predominantly used in the literature. In a RS outcome, as we have seen in the previous chapter (Figure 11.4), the high risks (the males) receive full insurance at their fair premium, while the low risks (the females) obtain partial insurance at their fair premium. In that case it would be better for the females if categorical discrimination were allowed: they would receive full insurance at their fair premium. The males are indifferent between discrimination and no discrimination, in both cases they obtain the same contract. Discrimination would be Pareto improving. The intuition is that if discrimination is not possible, firms try to screen the market by other means. In this case by offering partial insurance contracts. As in general these means lead to inefficiencies, it might indeed be better to allow discrimination in the first place. This effect is particularly strong in a RS outcome.

If one considers a WMS equilibrium instead, this result does not hold. In that case the males are subsidized by the females. If discrimination is allowed, women fare better, while men are worse off. This is probably the standard result one would expect from a switch in regime from no to full discrimination: High risks are worse off while low risks are better off. But note that due to the inefficiencies which arise on an insurance market under adverse selection, a social planner would always prefer to discriminate: She could still offer the males the same policy even after discrimination while the females can be made strictly better off. This is shown in Figure 12.7. Starting from a WMS outcome (H, L), if discrimination is allowed, under perfect competition the new outcomes are (H', L'). However, a social planner could for example offer the policies (H, L'') which would be a Pareto improvement compared to (H, L). This latter feature distinguishes discrimination in the insurance market from price discrimination in other markets: It might help to

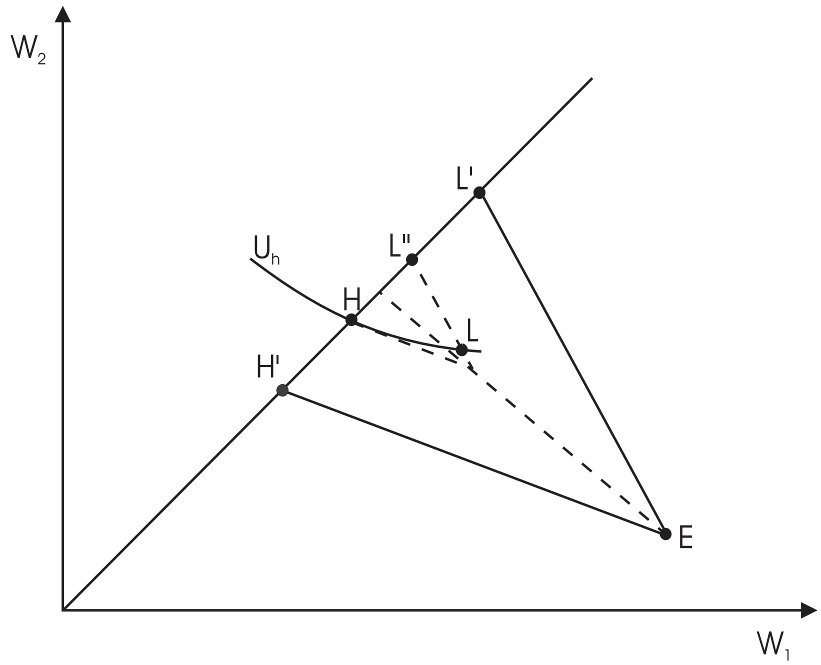

Figure 12.7: Discrimination can be welfare improving.

overcome some of the existing inefficiencies.

So far we have considered the case where the sex of the two types is a perfect signal of the riskiness. But in general one would expect that any category like sex or age does not reveal the risk type completely. There are probably still some cautious male and risky female drivers around. Denote by γ_m the proportion of high risks in the male population, and by $\gamma_f < \gamma_m$ the proportion of high risks in the female population. That is, the average male risk, which is $\gamma_m \pi_h + (1 - \gamma_m)\pi_l$ is larger than the average female risk.

To discuss the outcome in a competitive market, consider first the RS equilibrium, as shown in Figure 11.4. A crucial feature of the RS equilibrium is that the policies for the two risk types are independent of their actual proportion in society. High risks obtain full insurance at their fair premium while low risks obtain the best possible contract at fair terms, which the high risks do not prefer. In that case, discriminating with respect to sex would not change any contract offered.

This result does not hold in a WMS equilibrium. The cross-subsidizing pair of WMS contracts does depend on the ratio of the two types. To see this consider Figure 12.8.

The line $F - L$ denotes all feasible outcomes for the low risks. It is derived in the following way: Offer for the high risks a full insurance contract on the certainty line. Then draw the high risk indifference curve. That contract on this indifference curve, which has to be offered to the low risks such that the insurer makes zero profit altogether, is one

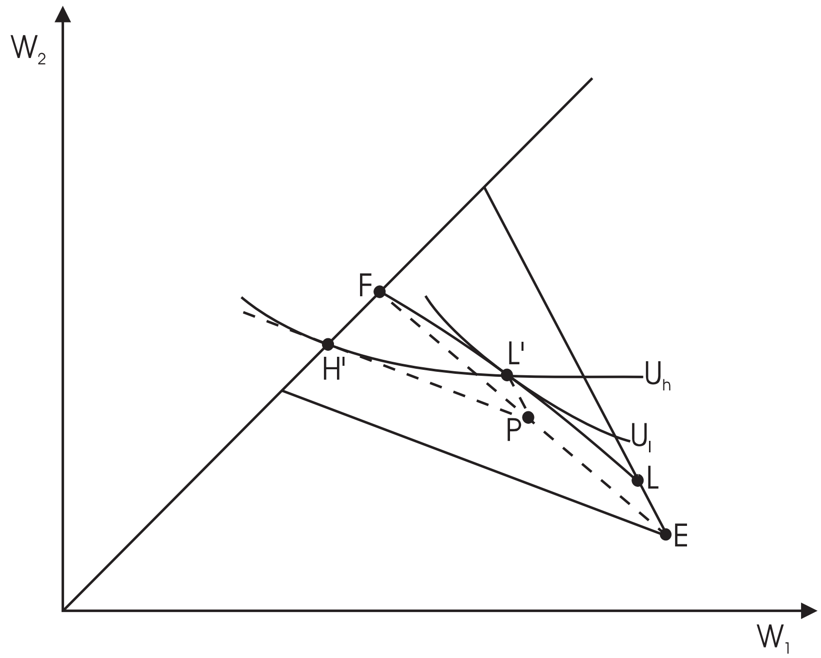

Figure 12.8: The efficient contract curve.

point on the line $F - L$. As an example consider the points (H', L'). H' provides full insurance for the high risks. L' is chosen such that the insurer makes zero profit if both risk types buy the two contracts in the proportion γ_h to $(1 - \gamma_h)$. To construct L', draw from H' the iso-profit line of the high risks, the dotted line. Where this line cuts the pooling-zero profit line (point P), draw the iso-profit line of the low risks. This line then cuts the indifference curve of the high risks at point L'. Thus the insurer is indifferent whether everyone buys contract P, or the low risks buy L' while the high risks acquire H'. Now by shifting H' along the certainty line, the contract curve $F - L$ can be obtained.

One endpoint of this curve, L is the RS contract of the low risks, while the other endpoint, F, must be the full insurance contract which lies on the pooling line. The WMS outcome is given by the best policy possible for the low risks along this line, here denoted by L'. All contracts above and including L', together with the corresponding contract for the high risks, denote the Pareto frontier in this case. Therefore the WMS contract pair is Pareto optimal as well. As a matter of fact, the WMS contract pair gives the low risk type the largest utility, given that asymmetric information is present, and the insurer make at least zero profit. The way we have drawn the curves in Figure 12.8, the RS pair of contracts is not Pareto optimal, but this only holds if the proportion of high risk types is small.

If γ_h changes, then contract F moves along the certainty line. For larger values of γ_h, F moves downwards. Furthermore, L' moves closer to contract L, whose position is

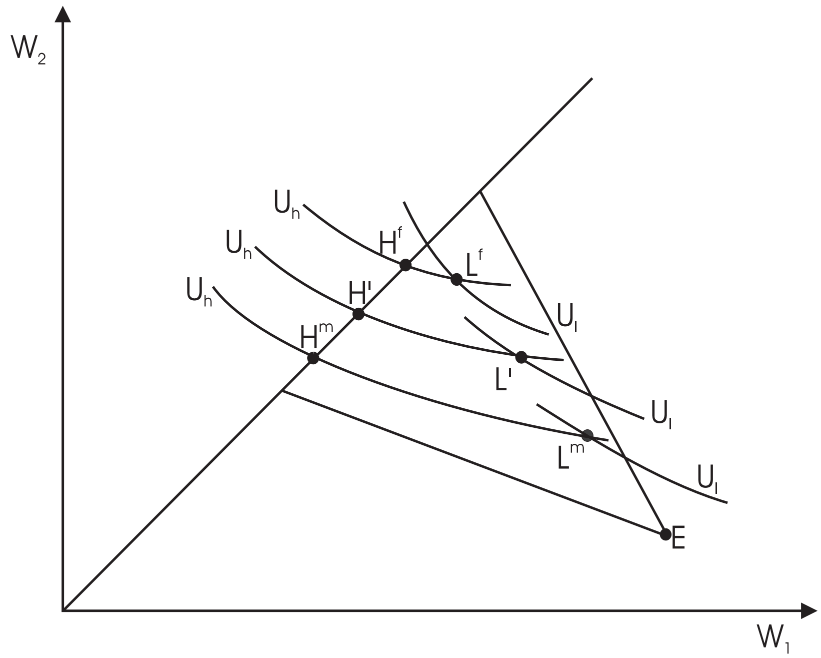

Figure 12.9: Unregulated discrimination is not Pareto improving.

independent of γ_h. For a large proportion of high risks, the RS contracts become Pareto efficient, which implies $L' = L$. If just a few high risks are in the population, then L' is shifted towards F, which itself moves close to the full insurance point at the low risk fair premium. With the help of this diagram we can make the following observations:

1. If it is possible to discriminate costlessly, then the insurance companies will do so. There will be winners and losers in the market. This can be seen in Figure 12.9. While initially at outcome pair (H', L'), the females move towards (H^f, L^f) while the males obtain (H^m, L^m). Note that in the case which we discussed in the beginning, if sex is a perfect indicator of the risk, $L^f = H^f$ at the low risk fair premium, while H^m has the fair premium of the high risks.

2. A social planner selling insurance policies could not do worse, and sometimes better if she discriminates. The reason is that by discriminating, the Pareto frontier shifts outwards. As we have seen above, if the signal is fully revealing, then this holds naturally (Figure 12.7). But also if the signal is only partially revealing, the efficiency gains could be used to make everyone better off.

Crocker and Snow (1986) have shown that with an appropriate tax system, where the different contracts are taxed differently, the state could implement any desired outcome, if firms behave according to the WMS equilibrium. In that case, but only in that case, everyone could be made better off by categorical discrimination. In general, however, without additional intervention by the government, there is usually someone who looses and someone who wins.

An interesting extension of the analysis here is the case of discriminating with respect to endogenous quantities, like the type of car, whether the car parks in a garage or not, whether someone smokes or not. In that case, discrimination might influence the behaviour of the insured. Consider the following situation: Half of the population likes to smoke and would be willing to pay \$20 per year for this privilege. The other half does not smoke. In addition, assume that smoking itself does not change the riskiness of a driver, but it happens to be the case that smokers are wilder drivers. So far, due to the legislation, everyone was fully insured at the same premium. Now discrimination is allowed. The insurance companies find out that smokers have on average more accidents than non-smokers, and they decide to charge them \$25 more per year, and the non-smokers \$25 less. What will happen? First, all smokers stop smoking and buy the contract of the non-smokers, as this is \$50 cheaper, so they gain \$30 alltogether. Then, the insurance companies make losses with the non-smokers policy (remember that smoking itself had no influence on the riskiness). Next year, they will raise the premium of the non-smoker contract again by \$25 to the old level while the contract of the smokers stays as it is. So after all, everyone pays the same as before, but the smokers are made worse off. Surely this was a simplified example[2], but the general point should be clear: If you can decide to which group you belong, the different policies sold to different groups will influence your behaviour.

To conclude this section: In contrast to most other markets, where the costs of production do not depend on the customer who buys the good, discrimination with respect to some characteristics may be welfare improving in the insurance sector. However, if discrimination is not accompanied by some other policy measures, there are usually some types who win and some others who lose.

12.3 Endogenous information acquisition

So far it was always assumed that individuals know which risk type they are, i.e. the information structure was exogenous to the model. In this chapter we investigate the situation where the individuals have no information ex-ante, but the possibility to acquire information about their risk type. In many real world situations, in particular in the health sector, this does seem to be the case: You decide for yourself whether to undergo a HIV test or not, or whether to take a genetic test or not.

In the simplest scenario there are three possible types in society. Those who have not

[2]A more elaborated model can be found in Polborn (1997), (see exercise 12.1).

taken the test (the 'uninformed'), those who were tested positive ('high risks') and those who where tested negative ('low risks').[3] Note that uninformed does only imply that the person has not undertaken the test. It does not imply that the person has no knowledge about his risk type at all. As individuals are assumed to be rational, an uninformed person knows that his risk is the average risk, which is defined below in more detail.

A point worth stressing is that we assume that ex-ante some people are already tested. This assumption influences the analysis, but it appears to be realistic, as those tests as mentioned above are usually done for diagnostic reasons with some individuals, independent of the insurance decision.[4]

We now discuss whether the uninformed have an incentive to undertake the test or not. There are two possible outcomes:

A) All individuals become tested, so in the end there are only positively tested individuals, the high risks, and negatively tested ones, the low risks.

B) Some or all of the uninformed do not undertake a test. Therefore in the end there are still three risk types in society.

Consider now the outcome in a competitive insurance market, where the insurer does not have access to the information whether the individual has undertaken a test or not, and in the former case, which test result was obtained. Following the discussion in Chapter 11 several equilibrium concepts could be considered. Here we restrict ourself to the Rothschild-Stiglitz outcome. This makes the analysis simpler, because the contracts do not depend on the distribution of types, only on the different number of risk types. However, some of the results have to be taken with care, as we will see later on. Furthermore, in particular with diagnostic tests in the health sector, the proportion of high risks is usually low. As we know from Chapter 11, this is exactly the situation where a Rothschild-Stiglitz outcome is not an equilibrium.

In scenario A, where all individuals become tested, the Rothschild-Stiglitz outcome is well known. These are the outcomes H, L as drawn in figure 12.3. (For the derivation recall chapter x.x). The high risks obtain full insurance at their fair premium, while the low risks obtain a partial insurance contract at their fair premium.

In scenario B, there are three risk types, where those who are uninformed are of 'average' risk. Formally, if the high risks have risk π_h, the low risks have risk π_l, and the proportion of high risks among the uninformed is γ_h, then the risk of the uninformed is $\pi_u = \gamma_h \pi_h + (1 - \gamma_h)\pi_l$. Then the Rothschild-Stiglitz outcome would be H, P, L' as drawn

[3] The analysis here is based on Doherty and Thistle, 1996. The case where the individuals have some prior knowledge about their risk type is discussed there as well.

[4] In exercise 12.2 you are asked to go through the analysis where ex-ante no one has been tested.

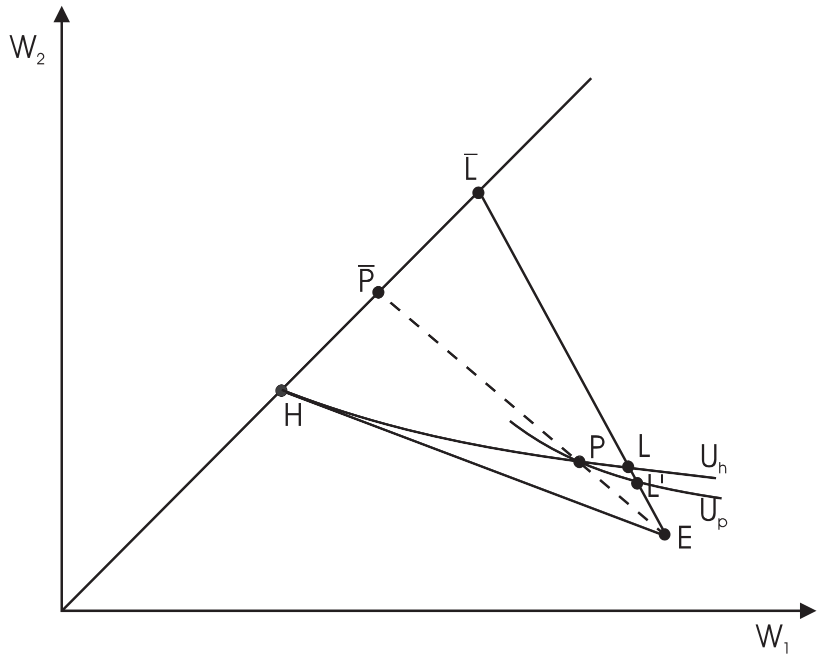

Figure 12.10: Possible final outcomes.

in figure 12.3. As before, the high risks obtain full insurance at their fair premium. The uninformed are now the low risks, compared to those positively tested. Therefore they obtain underinsurance with their fair premium (P), such that the high risks are exactly indifferent between H and P. To satisfy the incentive constraint for the uninformed, that they do not take the contract of the negatively tested, L' has to shift even further down the fair premium line of the low risks.

Assume first that individuals can undertake a test at zero cost. Then the following holds:

If the insurer can neither observe the test results, nor whether a test was undertaken or not, and if testing comes at zero cost, then everyone will become tested.

To prove this, we first need to show that in scenario B every uninformed person has an incentive to do the test. If someone is uninformed, he behaves as if his risk were π_u, and so he would buy contract P. If he undergoes the test, and he turns out to be a high risk, he buys H, otherwise L. Therefore the value of information (I) of the test is given by:

$$I = \gamma_h V(\pi_h, H) + (1 - \gamma_h) V(\pi_l, L') - V(\pi_u, P) \tag{12.6}$$

where $V(\pi, X)$ stands for the expected utility of a person with risk probability π having the state-dependent wealth X.

Due to the incentive constraint, $V(\pi_u, P) = V(\pi_u, L')$, one gets:

$$I = \gamma_h [V(\pi_h, H) - V(\pi_h, L')] + [\gamma_h V(\pi_h, L') + (1 - \gamma_h) V(\pi_l, L') - V(\pi_u, L')] \tag{12.7}$$

As $\pi_u = \gamma_h \pi_h + (1-\gamma_h)\pi_l$, the last term in brackets vanishes. Therefore:

$$I = \gamma_h [V(\pi_h, H) - V(\pi_h, L')] > 0 \tag{12.8}$$

The value of information is strictly positive, as high risks prefer outcome H to outcome L'. All the uninformed will undertake a test, therefore scenario B cannot be the outcome. The only possible outcome is therefore scenario A. If the Rothschild-Stiglitz contracts H, L are offered, and someone is uninformed, this person would choose outcome L. His incentive to test is given by:

$$I' = \gamma_h V(\pi_h, H) + (1-\gamma_h) V(\pi_l, L) - V(\pi_u, L) = 0 \tag{12.9}$$

where the last equality follows from $V(\pi_h, H) = V(\pi_h, L)$ and $\pi_u = \gamma_h \pi_h + (1-\gamma_h)\pi_l$. However, such a person, if he remains untested, would inflict a loss upon the insurer. Therefore (H, L) can only be the outcome if everyone is tested. That is the only possible equilibrium outcome.

Note that if H, L are the only outcomes, the incentive to test is zero. Now if tests are costly, no uninformed person would undertake a test in such a situation. This leads to a second result:

Assume that the insurer can neither observe the test results, nor whether a test was undertaken. If testing is sufficiently costly, then no uninformed person will undertake a test. If costs of testing are small, no equilibrium in the sense of Rothschild-Stiglitz exists.

The first point is obvious: large enough testing costs will distract every uninformed person to do the test. The second point is more delicate: As we have seen above, in scenario B the uninformed have an incentive to become tested. Now, if the costs of testing (for preciseness, these should be modelled in utility units) are small, scenario B cannot be the outcome. Therefore assume every insurer would offer the Rothschild-Stiglitz contracts leading to outcome H, L. But here the incentive to test is zero. So if there are still some people uninformed, they will not do the test. But then the low risk contract will be bought by the uninformed, which leads to a loss for the insurer offering this contract. Therefore this can also not be a stable outcome.

Note that this result depends on the assumption that the outcome is of the Rothschild-Stiglitz type. In a WMS world, the contracts offered depend on the ratios of types. Then it might well be conceivable that some uninformed do the test, until the cost of testing just becomes larger than the value of information. So far, a formal model for this case does not exist.

Let us now consider the situation where the insurer can observe whether a test was undertaken or not, but not the result.[5] Then, if someone undertakes the test, his final

[5]For a formal model see Crocker and Snow, 1992.

state-dependent wealth is either at H or L, depending on the test result. If he remains uninformed, his final wealth is $\bar{P}$, which is the full insurance contract at the uninformed fair premium (see figure 12.3). If someone is uninformed, his incentives to undertake the test are given by:

$$I'' = \gamma_h V(\pi_h, H) + (1 - \gamma_h) V(\pi_l, L) - V(\pi_u, \bar{P}) \tag{12.10}$$

Before discussing the result, let us consider as a final scenario the case where the insurer can observe everything, i.e. whether the test was taken or not, and the test result. Then the final outcomes are $H, \hat{L}$ for those tested, and $\bar{P}$ for the uninformed (see figure 12.3). Here, the incentive to test is:

$$I''' = \gamma_h V(\pi_h, H) + (1 - \gamma_h) V(\pi_l, \hat{L}) - V(\pi_u, \bar{P}) \tag{12.11}$$

As all contracts are on the certainty line, the individual faces a choice between obtaining outcome $\bar{P}$, or a lottery with outcome H with probability γ_h and outcome L with probability $(1 - \gamma_h)$. Due to $\pi_u = \gamma_h \pi_h + (1 - \gamma_h)\pi_l$ the expected value of both choices is the same, so the risk averse person would prefer to get the sure outcome $\bar{P}$, which implies that I''' is negative. This effect is called 'premium risk': If the uninformed undergoes a test, and this is observable, he faces the risk of being charged a larger premium. With $V(\pi_l, L) < V(\pi_l, \hat{L})$ it directly follows that also I'' is negative. To summarize:

If the insurer can observe whether a test was undertaken or not, independent of whether they can observe the test result, the private incentive to undertake a test is negative.

Note that if the insurer can only observe the test result, the scenario is a little bit like the one we discussed in the section on categorial discrimination. As we have seen in chapter x.x, the social value of information could then well be positive, depending on the equilibrium concept used, and whether an appropriate tax system is installed. Here we see that in such a case the private value of information is negative, so no one would be interested in doing the test.

A final interesting case to study is the so-called 'consent law' where individuals can show their test results to the insurer, but need not. Clearly, only those which are tested negative will do so. Depending on whether the insurer know if a test was undertaken or not, the final outcomes are either $H, \bar{P}, \hat{L}$ or $H, P, \hat{L}$. In the first case, the private incentive to undertake a test is negative, while it is positive in the second case. Therefore:

If the insurer cannot observe whether a test was undertaken, there is always a positive incentive to do the test. However, if the insurer have the information whether a test was done, this incentive vanishes, even if the insured can voluntarily reveal their test result.

As an application of these results, consider as an example the case of HIV tests.[6] Such a test is not only done for diagnostic reasons, but also to allow for preventive treatment. In this case, if the insurer are not allowed to ask whether the test was undertaken, individuals have an incentive to do the test. So forbidding insurer to ask for the test can be useful as it gives the individuals an incentive to do the test. In a Rothschild-Stiglitz world this would be optimal. However, in other equilibrium concepts, because of screening activities and efficiency losses due to adverse selection, the uninformed may be better off if all refrain from taking the test.

The other extreme would be to allow the insurer to ask for tests, undertake tests if they wish, etc., the so-called 'laissez-faire' regulation.[7] In that case, individuals have no incentive to take a test. This might well be welfare improving if costs of testing are large, and the behavioural responses to a test result, like preventive treatment, different savings behaviour, etc., are not significant. On the other hand, those people who for whatever reason undertake the test, face a potentially severe premium risk. A priori it is not clear which regime fares better.

[6]For a more elaborate discussion on the consequences in the health sector, see Chapter ?

[7]If there is no danger that individuals would undertake a test secretely, insurer would never require individuals to take a test before signing the contract. This only leads to premium risk, which the risk neutral insurer would avoid to burden on the insured, see exercise 12.3.

Chapter 13

Multi-period contracts and renegotiation

13.1 Multi-period contracts

In the real world, we observe many different forms of long term contracts: In automobile insurance this year's premium depends on whether or not the insured has had an accident last year. Life insurance contracts usually last for several years, and to exit from an existing contract is costly. The same holds for health insurance contracts. In unemployment insurance, the payment decreases with the time you are unemployed, while the premium usually does not depend on the duration of the employment. However, there also exist policies as in the legal or liability insurance market, which only last for one year. Renewal is possible, but the terms of the contract are independent of whether a policy was acquired last year. What are the reasons for these differences?

The models under symmetric information are not useful in understanding long term contracts. Suppose your risk of having an accident in any one year is π. In a competitive market you would receive insurance at the fair premium, and that in every year. No long term contract could do any better. As a matter of fact, any long term contract which makes zero profit in expectation is equivalent to a series of short term (one-period) insurance and saving contracts.[1] But if saving aspects are ignored, long term contracts do not improve efficiency under symmetric information.

We have to turn to asymmetric information to understand the economics of long term

[1] In some models, it is the saving motive which is the reason for long term contracts. If customers are credit constrained, for example, then a long term contract which specifies a very low premium in the beginning and a larger premium later on serves as a means of redistributing money across periods. Although in some cases this might be a relevant issue to discuss, we exclude this aspect in this section.

contracts. Asymmetric information in the form of adverse selection will be discussed in this section, while moral hazard will be dealt with in Chapter 16.

13.1.1 Finitely many periods

As before we simplify by considering two risk types only. One, the high risk type, has accident probability π_h, and the other has $\pi_l < \pi_h$. The proportion of the high risks in the population is γ_h. Now consider a two-period model, where the agents face the same risky environment in each period. In such a situation, a long term insurance contract for type i consists of 6 parameters:
P_i is the premium paid in period one. I_i^n is the net payment in case of an accident in period 1, paid out in period 1. $P_i(n)$ is the premium in period 2, if no accident has happened in period 1, and $P_i(a)$ is the corresponding premium if an $a$ccident has happened. $I_i^n(n)$ and $I_i^n(a)$ are the corresponding net indemnities.

We speak of a long term contract if the optimal contract is such that it differs from two short term contracts. The problem with this definition is that it is not clear what two short term contracts would look like. This is not a problem if one considers a Rothschild-Stiglitz equilibrium only: As long as the risk probabilities stay constant over time, and there are at least some high risks who buy on the spot market in the second period, the spot market contracts in the two periods are the same. It is more difficult in the WMS equilibrium concept, where the contracts depend on the ratio of high risks to low risks. In the optimum everyone will buy a long term contract, so no one buys on the spot market in period 2. This implies that the ratio of high to low risks is undetermined, which makes the WMS contracts difficult to determine.

Still, for our purpose we define a long term contract as a contract where the premium or indemnity in period 2 differs from that in period 1. As we will see, this interpretation is sufficient for the economics we want to bring across: The usefulness of experience rating.

Facing a long term insurance contract, the expected utility of an individual of type i is:

$$\begin{gathered}(1-\pi_i)U(W-P_i)+\pi_iU(W-L+I_i^n)+(1-\pi_i)[(1-\pi_i)U(W-P_i(n))+\\ \pi_iU(W-L+I_i^n(n))]+\pi_i[(1-\pi_i)U(W-P_i(a))+\pi_iU(W-L+I_i^n(a))]\end{gathered} \tag{13.1}$$

which we abbreviate by:

$$EU_i(P_i,I_i^n)+(1-\pi_i)EU_i(P_i(n),I_i^n(n))+\pi_iEU_i(P_i(a),I_i^n(a)) \tag{13.2}$$

By writing down this expression, we made a series of simplifying assumptions. First, we assumed that the income of the agent in both periods is the same. Different wealth in the different periods would make the formulation more messy. Furthermore, the spot market contracts might also differ between the two periods, as the insured are in general more or less risk averse if they are poorer or richer. But overall the economics stays roughly the same. Second, the agent does not save but consumes all the income he has. Allowing for savings would create different problems, some of which we will discuss below. But note that if the contract specifies full insurance at the same price in both periods, then there is no incentive to save anyway. A third assumption is that there is no discounting. This is just made for simplicity.

We discuss the result in a model of perfect competition by using the WMS equilibrium concept. The qualitative features of a long term contract do not differ when one uses the Rothschild-Stiglitz equilibrium instead, but formally the results are easier to see in the WMS framework. As the formalism for a monopoly insurer is dual to the WMS concept, also in this market form the results are qualitatively the same.

To derive the WMS equilibrium, the following optimisation problem has to be solved:

$$\begin{aligned}
&\max_{P_i, I_i^n, P_i(n), I_i^n(n), P_i(a), I_i^n(a)} EU_l(P_l, I_l^n) + (1-\pi_l)EU_l(P_l(n), I_l^n(n)) + \pi_l EU_l(P_l(a), I_l^n(a)) \\
&\text{s.t.} \\
&EU_h(P_h, I_h^n) + (1-\pi_h)EU_h(P_h(n), I_h^n(n)) + \pi_h EU_h(P_h(a), I_h^n(a)) \geq \\
&EU_h(P_l, I_l^n) + (1-\pi_h)EU_h(P_l(n), I_l^n(n)) + \pi_h EU_h(P_l(a), I_l^n(a)) \qquad \text{(IC)} \\
& \\
&(1-\gamma_h)\{(1-\pi_l)P_l - \pi_l I_l^n + (1-\pi_l)[(1-\pi_l)P_l(n) - \pi_l I_l^n(n)] + \\
&\pi_l[(1-\pi_l)P_l(a) - \pi_l I_l^n(a)]\} + \gamma_h\{(1-\pi_h)P_h - \pi_h I_h^n + \\
&(1-\pi_h)[(1-\pi_h)P_h(n) - \pi_h I_h^n(n)] + \pi_h[(1-\pi_h)P_h(a) - \pi_h I_h^n(a)]\} \geq 0 \qquad \text{(PC)}
\end{aligned} \tag{13.3}$$

This is the generalization of the maximization problem (11.2): Maximize the utility of the low risks such that the high risks prefer their contract to that of the low risks (the incentive constraint) and the insurance companies make no loss (the participation constraint).

First let us work through the first order conditions with respect to the high risk contract parameters, where λ (μ) is the Lagrange parameter of the incentive (participation) constraint:

$$\begin{array}{lr}
P_h & -\lambda(1-\pi_h)U'(W-P_h) + \mu\gamma_h(1-\pi_h) = 0 \\
I_h^n & \lambda\pi_h U'(W-L+I_h^n) - \mu\gamma_h\pi_h = 0 \\
P_h(n) & -\lambda(1-\pi_h)(1-\pi_h)U'(W-P_h(n)) + \mu(1-\pi_h)\gamma_h(1-\pi_h) = 0 \\
I_h^n(n) & \lambda(1-\pi_h)\pi_h U'(W-L+I_h^n(n)) - \mu(1-\pi_h)\gamma_h\pi_h = 0 \\
P_h(a) & -\lambda\pi_h(1-\pi_h)U'(W-P_h(a)) + \mu\pi_h\gamma_h(1-\pi_h) = 0 \\
I_h^n(a) & \lambda\pi_h\pi_h U'(W-L+I_h^n(a)) - \mu\pi_h\gamma_h\pi_h = 0
\end{array} \tag{13.4}$$

Note that the 'no-accident' ('accident') first order conditions are very much the same as those for period 1, the only difference is that they are multiplied by $(1-\pi_h)$ (or π_h respectively) which cancels out. Thus it follows that the high risks obtain full insurance, and their wealth in all states of the world is the same: $P_h = P_h(n) = P_h(a)$ and $I_h^n = I_h^n(n) = I_h^n(a) = L - P_h$. The 'no-distortion-at-the-top' result again. High risks obtain full insurance and no income variations across periods. The exact premium which is charged cannot be derived, as this depends on the degree of cross-subsidization which takes place. If there are many high risks around, so that the Rothschild-Stiglitz equilibrium becomes relevant, then $P_h = \pi_h L$. Note that for the high risks the optimal contract is not a long term contract as the terms of the contract do not differ between periods. The result as presented depends on the assumptions that wealth in both states are the same and utility functions do not differ. If these were relaxed, then the contract in period 2 would still not depend on whether an accident did occur or not in period one, but it would differ from the contract in period 1. Observe that marginal utility across states is equalized for the high risks, so this implies that wealth is equalized only if the utility functions are the same across states and time. And this implies that the contract is the same in the two periods, only if initial wealth in both periods is the same. Note also, that with such a contract the high risks have no incentive to save or dissave money.

The first order conditions for the low risks are slightly more elaborate:

$$\begin{array}{lrl}
P_l & -(1-\pi_l)U'(W-P_l) + \lambda(1-\pi_h)U'(W-P_l) + \mu(1-\gamma_h)(1-\pi_l) & = 0 \\
I_h^n & \pi_l U'(W-L+I_l^n) - \lambda\pi_h U'(W-L+I_l^n) - \mu(1-\gamma_h)\pi_l & = 0 \\
P_l(n) & -(1-\pi_l)^2 U'(W-P_l(n)) + \lambda(1-\pi_h)^2 U'(W-P_l(n)) & \\
 & +\mu(1-\pi_l)^2(1-\gamma_h) & = 0 \\
I_h^n(n) & (1-\pi_l)\pi_l U'(W-L+I_l^n(n)) - \lambda(1-\pi_h)\pi_h U'(W-L+I_l^n(n)) - & \\
 & \mu(1-\pi_l)\pi_l(1-\gamma_h) & = 0 \\
P_l(a) & -\pi_l(1-\pi_l)U'(W-P_l(a)) + \lambda\pi_h(1-\pi_h)U'(W-P_l(a)) & \\
 & +\mu\pi_l(1-\pi_l)(1-\gamma_h) & = 0 \\
I_h^n(a) & \pi_l^2 U'(W-L+I_l^n(a)) - \lambda\pi_h^2 U'(W-L+I_l^n(a)) - \mu\pi_l^2(1-\gamma_h) & = 0
\end{array} \tag{13.5}$$

Reformulating the expressions for the premia gives:

$$\begin{array}{lrl}
P_l & U'(W-P_l) & = [1-\lambda(1-\pi_h)/(1-\pi_l)]^{-1}\mu(1-\gamma_h) \\
P_l(n) & U'(W-P_l(n)) & = [1-\lambda(1-\pi_h)^2/(1-\pi_l)^2]^{-1}\mu(1-\gamma_h) \\
P_l(a) & U'(W-P_l(a)) & = [1-\lambda\pi_h/\pi_l(1-\pi_h)/(1-\pi_l)]^{-1}\mu(1-\gamma_h)
\end{array} \tag{13.6}$$

As

$$\frac{\pi_h}{\pi_l}\frac{(1-\pi_h)}{(1-\pi_l)} > \frac{(1-\pi_h)}{(1-\pi_l)} > \frac{(1-\pi_h)^2}{(1-\pi_l)^2} \tag{13.7}$$

it follows that

$$P_l(a) > P_l > P_l(n) \tag{13.8}$$

The agent is 'penalized' in period 2 if a loss occurred in period 1, but 'rewarded' for no loss. Note that there is nothing the agent can do about the accident, so that penalizing and rewarding are not meant in the sense that they give the insured an incentive to avoid losses. The contract structure is such that the high risks have no incentive to choose the contract designed for the low risks. And as they have a larger probability of having an accident, they are more afraid of the 'penalty' which might occur.

Reformulating the expressions for the indemnity, and comparing all equations it is easy to see that the low risks receive partial insurance in both periods, and that the indemnity can be ranked as well:

$$I_l^n(a) < I_l^n < I_l^n(n) < L - P_l(n) \tag{13.9}$$

In this case, the assumption we made before that the insured cannot save may become binding, as the agent anticipates that he has different outcomes in period 2.

If there are more than two periods, the high risks still obtain full insurance in every period, independently of whether an accident did occur or not, while the low risks obtain a long term policy with partial insurance in every period. The premium indemnity schedule is given by $(P_l(t,j), I_l^n(t,j))$ where t denotes time and j the number of accidents which occurred already. It can be shown that $P_l(t,j)$ is increasing in j while $I_l^n(t,j)$ decreases in j for constant t (Cooper and Hayes, 1987). Note that it does not play a role when exactly the accident happened, only how often an accident occurred. The reason is that the exact timing does not give more information on the risk type: The expected belief about the riskiness of two individuals, one of which had accidents in period 1,2,5, the other in 2,4,6, is the same.

This structure of the low risk contract is a feature we observe in many insurance markets, most notably automobile insurance. It is known as *experience rating* or *bonus-malus* system. To make the argument again: These contracts provide the low risks the optimal protection under the constraint that the high risks do not choose this contract. In the model presented here, experience rating is *not* a means to provide incentives to take more care about accident prevention.

Before finishing this section, one remark on general adverse selection models is in order. As we have stressed several times, the problem we face in the insurance market is formally quite similar to that faced by a monopolist selling a good, a government procuring weapons, and many other principal agent models. However, with respect to long term contracts, one difference between the standard applications and the insurance market occurs: In the insurance market, over time information of the type of agent is revealed, independent of which contract is signed. If someone has an accident, he is more likely a high risk than a low risk. On the other hand, if the monopolist does not sell any good or offers the same contract to all customers, then she is as smart in period 2 concerning the willingness to pay of her customers as she was in period 1. It actually turns out that in those standard principal agent models, the optimal long term contract is just the repetition of the one-period contract, which surely differs from the result obtained for the insurance market. You might ask yourself whether the monopolist, if she offers separating contracts, does not learn about the type of customer when he buys the good. That is true, but this information is endogenously determined, i.e., the customer reveals information on the basis of the contracts, and not exogenous as in the insurance market. However, the fact that the principal might potentially learn about the agent by the type of contract the agent chooses will appear soon, when we discuss renegotiation.

13.1.2 Infinitely many periods

Although usually insurance contracts do not last longer than a maximum of 30 years or so, considering infinitely many periods is useful as it clearly brings out the advantage long term contracts provide. Suppose you have a coin, which might be manipulated such that heads appears twice as often as tails. But you do not know for sure. What would you do to find out?

Yes, pay your younger sibling \$1 and ask him to throw the coin 1000 times, and to write down how often head appeared. One would expect that the manipulated coin comes up with head much more often. Once we have this number, we can calculate the probability of this happening under the two scenarios, which is given by $\left(\begin{array}{c} 1000 \\ N \end{array}\right) p^N (1-p)^{1000-N}$ where N is the number of heads. Thus, if head appeared 550 times, and the coin was expected to be manipulated with probability 1/2, then the revised belief that the coin is still manipulated is given by $2 \times 0.67^{550} 0.33^{450} / (0.5^{550} 0.5^{450} + 0.67^{550} 0.33^{450})$, which is approximately equal to 10^{-12}, quite a small number.

This effect can also be used in the insurance market. Actually, we have already done this before, just with 2 periods. If very many periods are possible, then one should obtain quite precise information about the riskiness of the type. In the limit, the information should be so good that the first best can be closely approximated. To achieve this, one would like to give both types of agents a contract which specifies full insurance in every period (because this is efficient), but some form of penalty if the observed number of accidents differs from that which is expected for this particular type. It is this latter point which makes the problem non-trivial: If both types receive full insurance at their fair premium, high risks would opt for the low risks contract. Therefore the low risk contract must provide some form of partial (or no) insurance if the number of accidents is too large to be expected from a low risk. This will deter the high risks from buying this contract. On the other hand, the low risks should obtain full insurance for themselves almost surely. This is slightly tricky to achieve. One way to do it is to set contracts in the following way: $(P_h, I_h^n) = (\pi_h L, (1-\pi_h)L)$ in each period, and

$$(P_l, I_l^n) = \begin{cases} (\pi_l L, (1-\pi_l)L) & \text{if } \frac{N}{T} < \pi_l + \delta(T) \\ (0,0) & \text{otherwise} \end{cases} \tag{13.10}$$

Here T is the period and N is the number of accidents which occurred so far.

The whole trick lies in finding the appropriate $\delta(T)$ function, which should be large enough that the low risks have only an infinitesimal risk of being penalized, and small enough that the high risks have a significant risk of being penalized if they choose this contract.

As is quite obvious from the remarks above, $\delta(T)$ will be a decreasing function in T, more periods allow one to get much better information about the true risk type of the agent.

One function, which satisfies the above mentioned requirements is

$$\delta(T) = \sqrt{2\gamma\pi_l(1-\pi_l)\log[\log[T]]/T} \tag{13.11}$$

where γ is some parameter larger than one.[2]

To summarize the results we obtained: Long term contracts allow for a weakening of the incentive compatibility constraint. Involuntarily the high risks reveal information about their type, because they have more accidents on average. So making the terms of the contract improve if no accident has occurred, but worsen if an accident did happen, the policy becomes more acceptable to the low risks than to the high risks. This is the structure known as bonus-malus systems or experience rating. In the limit of infinitely many periods, the first best can be achieved.

[2]The 'Law of the Iterated Logarithm' states that for any sequence of independent identically distributed random variables $\{x^t\}$, with finite mean $\bar{x}$ and finite variance σ^2, and for any $\gamma > 1$, almost surely

$$\lim_T \sup \frac{|\bar{x} - T^{-1}\sum_{i=1}^{T} x^i|}{\sqrt{2\gamma\sigma^2 log[log[T]]/T}} < 1.$$

With $\sigma^2 = \pi_l(1-\pi_l)$, the low risks average number of losses will almost surely for all but finitely many T be smaller than $\pi + \delta(T)$. It is slightly more demanding to show that the high risks indeed prefer their contract to that of the low risks. We refer the reader to Dionne (1983), who discusses the issue with a continuum of types.

13.2 Renegotiation

So far we have assumed that long term contracts are enforceable, which is to say that once the contract is signed, both parties will stick to it.

However, there are at least two situations where the enforcability of long term contracts is limited. One is for legal, the other for economic reasons. In some cases, laws prevent long term contracts. Most famous is the prohibition of slavery: you are not allowed to commit yourself to work with a company for 20 years, say. On the other hand, the company might well offer you a long term contract which it cannot breach, while you can. Similarly in insurance markets: In some sectors firms offer long term contracts, but the insured are allowed by law to opt out of the contract each year.

The other problem with long term contracts is economic: If the long term contract is signed, the insurer knows the type of the agent. But as usually the contract is inefficient (only partial insurance for the low risks), profitable renegotiation between the insurer and the insured could take place. And if both parties agree to renegotiate, then no court will forbid them to change the conditions of the contract. However, surely, if renegotiation could take place, this will be anticipated by the high risks, who then might choose the low risk contract in expectation of profitable renegotiations. So a priori it is not clear what will happen. Note that this problem of renegotiation already occurs before the first period starts, i.e., the initial contract could be immediately renegotiated. We will discuss all three issues in turn, first, when renegotiation may take place in the second period due to legal reasons, second, renegotiation in the second period due to efficiency reasons, and third, immediate renegotiation of contracts.

13.2.1 One-party commitment

First consider the possibility of renegotiation due to laws, which, although stated as if they allow the insured to change firms every year, in effect they prohibit that the insured commit themselves to a binding long term contract.

We work again in the two period model used before. If only the insurer can commit herself to long term contracts, but not the agent, the low risk might quit his contract if an accident occurred. Furthermore, a high risk type might perhaps choose the low risk type contract and, in case of an accident, change the insurer.

This introduces at least one further constraint in the optimisation problem:

$$EU_l(P_l(a), I_l^n(a)) \geq EU_l(P_l(s), I_l(s)) \tag{13.12}$$

where $(P_l(s), I_l(s))$ is the contract offered to the low risks on the spot market in period

2. As already discussed above, it is not quite clear what this contract is, as it is only offered out of equilibrium. In equilibrium the low risks do not opt out of their long term contract. However, as high risks could also buy single period contracts out of equilibrium, a good starting point for the analysis would be to assume that the spot contracts are the Rothschild Stiglitz contracts. But note that these policies only play a role in so far as they give the outside option of a low risk type on the spot market, they do not change the qualitative structure of the optimal contract. In addition there might be a further constraint for the high risks, as mentioned above.

Instead of going through all the equations again, we discuss the results:[3] High risks obtain full insurance as before. It also still holds that in case of no accident the low risks are rewarded, that is the premium is lower and the net indemnity is larger in the second period. The policy in case of an accident however is modified. It can be shown that:

$$EU_l(P_l(n), I_l^n(n)) > EU_l(P_l(a), I_l^n(a)) > EU_l(P_l, I_l^n) \tag{13.13}$$

The low risk type is better off in period 2 than in period 1, independent of whether he has had an accident or not, but he is even better off in case of no accident. This ranking of utilities is necessary such that the low risk type does not leave the insurer after period 1.

An interesting result which follows from this is that, if no cross-subsidization occurs, the insurance company makes an expected profit from the low risk type in period 1 and an expected loss in period 2. Even if the overall expected profit is not zero, insurer make lower profits per agent in later periods.

There is an interesting debate on whether insurer make a profit with their clients first and losses later on or whether it is the other way around. This model seems to suggest that it is the former case, insurer make losses later on such that customers do not prefer to change the company. As an example consider private health insurance markets, where in some cases the insured explicitly pays more in earlier periods to obtain lower premia in later periods.[4] Empirics however seem to suggest that insurer make losses first and profits later on (also known as 'lowballing', D'Arcy and Doherty, 1990). One possible explanation for this could be that the insurer first learn about the type of the agent and expropriate this knowledge at later stages. Competition then drives the market to zero expected profits, which implies losses first and profits later on (Kuhnreuter and Pauly,

[3]A detailed analysis can be found in Cooper and Hayes (1987).

[4]There are other reasons for this phenomena: First, such a contract has a saving element in it: save now for the higher future premia. Second, individuals learn more about their risks when they grow older. Locking the customer in might be a means to prevent him from going to another insurer if he turns out to be a low risk, and to stay with the insurer only if he is a high risk.

1985).

13.2.2 Renegotiation in period 2

In contrast to the last subsection, now both parties can again commit themselves to a long term contract. However, as discussed earlier, they cannot prevent themselves from renegotiating to an efficient contract once the type is known. This is actually what the Coase Theorem alludes two: *Bargaining will achieve an efficient allocation of resources whatever the allocation of property rights, if transaction costs are zero.* So if the insurer knows the type of agent for sure, such that there are no transactions costs due to asymmetric information, then an efficient outcome will be obtained.

If we assume that the insured are perfectly separated in period 1, this implies that both contracts for the low risk in period 2 have to specify full insurance, whether an accident has occurred or not. The only contracts which are fully efficient are those with full insurance. This then introduces two further constraints, the so-called renegotiation proofness constraints:

$$I_l^n(n) = L - P_l(n); \quad I_l^n(a) = L - P_l(a) \tag{13.14}$$

The contract can still be different in case of an accident and if no accident occurred, but it cannot specify partial insurance as before.

Note that due to renegotiation the low risks loose. The optimisation problem in a WMS equilibrium is the same as before, but now with two additional constraints. So the low risks cannot be made better off, although at first glance renegotiation seems to be a good thing to have. We can say even more in a Rothschild-Stiglitz equilibrium: If renegotiation is possible, then the high risks still obtain the same full insurance contract at their fair premium, the firms still make zero profit, and the low risks are *worse* off, a clear Pareto worsening. The inefficiencies in the contract were used to prevent the high risks from taking the low risk contract. If some mechanism forbids to use these inefficiencies, then the separation of types is much harder to achieve. Separation is still possible, but the contract in case of an accident has to be sufficiently bad, while that in case of no accident has to be very good. So in addition to the partial insurance he obtains in period one, the low risk type faces a larger future income risk than he would have had in the case of no renegotiation.

In this subsection we considered renegotiation for efficiency reasons which take place in period 2, once the contract is renewed. However, one might ask whether the parties would not like to renegotiate already earlier, once the original contract is signed, but before it is executed. This is one of the criticisms which apply to all the models we have presented so far, and to many of the models in the principal-agent literature. There it is always assumed that the principal makes a take-it-or-leave-it offer to the agent. However, if full screening were to take place by using inefficient contracts, immediate renegotiation would be profitable. We will devote a whole section to this problem.

13.3 Renegotiation before contract execution

Consider again a static insurance market where firms offer insurance contracts to a population of agents who are either high or low risks. As we have seen before, usually the low risks obtain partial insurance, while the high risks obtain full insurance. The population is fully screened, i.e. after signing the contract, the insurer knows the type of the insured perfectly. The contract for the low risks, however, is inefficient. So the insurer might call the agent in the evening, tell him that she now found out that he is of the low risk type, and whether he would not like to obtain full insurance instead of partial insurance. The premium could be such that both the low risk as well as the insurer are made better off.

This seems to be good news: First let both types sign their contracts and then renegotiate only with the low risks, so in the end all obtain full insurance and the adverse selection problem has nearly vanished. Has it? Surely not. The high risks will anticipate that they are called in the evening if they choose the low risk contract. So they would pick this contract as well, hope for renegotiation towards a full insurance contract, which is better than the policy they would have obtained otherwise.

One might argue, that this is perhaps the reason why the insurer does not start to renegotiate in the first place, as the high risks will mimic the low risks once again. But note that with the argument above, no renegotiation can also not be the outcome. If noone anticipates renegotiation, the high risks and low risks would separate into the two contracts. Then an insurer can profitably renegotiate with the low risks, and there is no fear that she picks up a high risk. So we have a real problem here: No renegotiation is not an equilibrium, as in that case renegotiation would be profitable. Renegotiation, starting from separating contracts is also not an equilibrium, because then the high risks would choose the low risks contract in the first place. More work is required to obtain a solution to this problem.[5]

We describe two attempts on how to solve this problem, which are to our knowledge the only models which deal with this issue.

The first one is based on Beaudry and Poitevin (1993). Consider an insurance market

[5]The commitment problem is well known in the literature. For example, Kreps (1990) refers to it on pp 677-679: 'The problem of credibility'.

where the agent who is either a high or a low risk type makes a contract proposal to the insurer each period. The contract specifies not just the terms of the insurance policy, but also at which period it will be executed.[6] The insurer can either accept or reject. If the insurer accepts, then the agent has the chance to make at least one additional offer before the contract is executed. If the agent makes another offer, and that is accepted, the former contract becomes void. Now the agent can make an additional offer, and so on. If the agent makes no further offer, or the insurer rejects the next offer, then the former contract stays valid. When the specified execution time arises, negotiation comes to an end, and the contract will be executed.

There are two features in this game worth noting: First, the agent makes all the offers. Thus in contrast to the models presented so far, this is a signalling rather than a screening model. The agent signals his type by offering specific contracts, rather than the insurer trying to screen the market by offering a menu of contracts. Nevertheless, letting the agent make all the offers seems to be a good starting point to model a competitive insurance market where the agent has all the 'bargaining-power'. On the other hand it is not clear why in the real world the insurer could not make the renegotiation offer herself. Once the first contract is signed, the notion of bargaining strength is not very useful anymore, because the outside option of both parties is exactly given by this signed contract. As we will see later on, the results depend quite critical on the assumption of agent-offers only. A second characteristic of this game is that no contract can be executed without having the chance to renegotiate the terms of the contract. The agent cannot make an offer to the insurer saying to execute immediately without negotiating further. Instead, there is always at least one round of further negotiating. Committing not to renegotiate is not possible.

Instead of going through the whole analysis, we will discuss one equilibrium path and motivate the result. As there is no discounting in the model, there is no unique equilibrium strategy. The final contracts however are under appropriate belief refinements unique. One possible equilibrium path is the following:

If the agent is of the low risk type, he offers the Wilson pooling contract in period 1. Recall that the Wilson pooling contract is that contract on the pooling zero profit line which the low risks find optimal. In period 2 he makes no additional offer.

If the agent is of the high risk type, he offers the Wilson pooling contract in period 1 as well. In period 2, however, the high risks make a new proposal, offering a full insurance contract which is such that the insurer makes the same profit with the high risk at this

[6] Note that 'period' does not refer to 'time periods' as in the previous sections, but rather to stages of the negotiation process. There is for sure no accident during the negotiation.

contract than at the Wilson pooling contract. In period 3 he makes no offer.
The insurer says yes to all the offers, so that the contracts are executed in period 3 for the low risks, and period 4 for the high risks.

This is a remarkable result: The equilibrium outcome is exactly the same as in the Jaynes-Hellwig equilibrium which we discussed previously. Low risks obtain the best pooling contract, while high risks obtain full insurance and impose the same loss to the insurer as if they were to buy the low risk contract. In the Jaynes-Hellwig model the reason for the final contracts were two groups of firms, one which is willing to exchange information about its customers, the others which would not do so. Here the outcome is the same, but at first glance the reason is very much different: High risks try to mimic the low risks as good as possible. This problem is so hard, that the low risks cannot make any contract offer which the high risks would not do as well. So the best they can achieve for themselves is the Wilson pooling contract. However, once they arrived at the Wilson pooling contract, the high risks cannot gain further from mimicking the low risks. So they amend their contract to full insurance, where the additional insurance is priced fairly. In effect, while in the Jaynes-Hellwig equilibrium other firms in the market negotiate further with the high risks and amend their contract, in this model the firms themselves cannot commit not to negotiate further, so they modify the contract of the high risks. In both cases, the high risks will gain as much as possible from mimicking the low risks before they reveal to be of the high risk type and buy additional insurance coverage.

This result clearly brings out one dynamic feature of renegotiation: High risks are much more difficult to separate from the low risks than in a model without renegotiation. Still, in the end both types separate, and the low risks obtain partial insurance as before. But this latter point is curious: We discussed in the beginning that partial insurance is inefficient, so that it would not occur if renegotiation is allowed and screening is perfect. However, here in the end low risks obtain different contracts than the high risks, and their contract is not efficient. Why do the low risks, once they are separated from the high risks, not negotiate further? For example, at stage 2, while the high risks demand their full insurance contract, the low risks could do as well with their fair premium.

In the equilibrium outlined above, this does not work due to the following out-of-equilibrium belief: If at some stage the agent offers any contract which is not the Wilson contract, the insurer believes that the agent is of the high risk type with probability one. So even if she had a previous belief of the agent to be of high risk with probability zero, once this agent makes an out-of-equilibrium contract offer, the insurer *switches* belief to the high risk type. At first glance this switch of belief appears strange: Although first the insurer was sure to have a low risk type with whom she negotiates, once this party reacts unexpected, she beliefs the agent to be of the high risk type for sure. However, for an equilibrium to exist, in this case a switch of belief is necessary: if the insurer's strategy in equilibrium were not to switch her belief, then the high risks would anticipate this and mimic the low risks even further.

The result is interesting, and the dynamics leading to it as well, because it shows very delicate problems of the renegotiation process. However, the assumption that only the agent makes all the offers is restrictive: This 'switch of belief' only works if the agent makes all the offers. If the insurer could make a renegotiation offer instead, just by making this offer no further information would be revealed, neither in nor out of equilibrium. Therefore, if the insurer has a belief along the equilibrium path that the agent is of the low risk type with probability one, she would always offer a full insurance contract, with which she makes more profit and the agent is better off. By making the offer herself, no belief will be changed or updated. However, as we have remarked above, the switch of belief is necessary to stabilize the equilibrium in the previous game, because otherwise the high risks would mimic the low risks even further. So it is not clear what happens if the insurer can make offers as well before the contract is executed.

There does not exist a non-cooperative model which allows for renegotiation offers to be made by the insurer, or by both parties. There is one approach, however, which tries to formalize the possibility of renegotiation in an axiomatic setup, which itself is independent of who makes the offer.[7] This approach specifies stability requirements the outcomes have to satisfy such that no further renegotiation will take place.

We will not go through the axioms here, but discuss one general result and its application to the insurance market. The general result is the following:

Given any candidate for a market outcome: If for any contract and type distribution choosing this contract, there exist at least one contract which makes everyone better off, then this outcome is not renegotiation proof.

This result describes what one would expect from renegotiation if the principal could make offers as well. Suppose negotiation has reached some stage, so that a contract is signed, and the principal has some belief of the type of the agent. Now, if she finds one contract which every agent in her belief set prefers and with which she increases her profit, she would offer it, and it would be accepted by everyone. Noone has to reveal his type by accepting the new contract, no change in the belief structure takes place. So there is no reason why this new arrangement should not go through. Consider on the other hand a situation where only with a menu of contracts, say two, everyone can be made better

[7] In the context of a monopoly insurer this was done by Asheim and Nilssen (1997). In Wambach (1999 a, 1999 b), the approach is extended to general principal agent models and to bargaining models.

off and only if the two risk types separate between the two contracts. Then it is not clear whether the principal will make such an offer. How does she know whether the high risks will indeed buy the contract which is designed for them? Maybe they anticipate further negotiations and choose the contract of the low risks? So any offer where types have to reveal some information has to be considered with care. But just offering a single contract, so that no further information has to be revealed, is a safe strategy to use.

Although this result is very intuitive (for a proof see Wambach, 1999 a), it has severe consequences for the insurance market. Assume that someone claims the outcome of the market process is the Rothschild Stiglitz outcome. Then there are two final contracts and two type distributions which belong to these contracts: Only high risks choose the full insurance contract, and low risks choose the partial insurance contract. Now the general result says that this is not renegotiation proof: For the partial insurance contract and the corresponding type distribution, namely low risks only, there indeed exists one contract which is better for both the insurer and the low risks. So the RS contracts cannot be final outcomes. With a similar observation, also the cross subsidizing contracts of the WMS type or the Jaynes-Hellwig type cannot be final outcomes. On the other hand, if someone claims that the Wilson pooling contract is the outcome, the general result has no bite. Sure, there exists a menu of contracts which could make everyone better off, but not a single contract.

This result already stands in contrast to the model discussed above, where only the agent could make the offers. There, both types are separated and the low risks obtain partial insurance. That is not possible anymore. If the principal knows that the agent is a low risk, she would offer him a full insurance contract. But then, what is the result?

One possible result is a pooling full insurance contract. This contract is efficient, and both types are pooled, so that no information is revealed. However, if the proportion of high risks is large, then even if the insurer makes zero profits, the low risks might not buy the full insurance pooling contract, as they are better off without any contract at all. But this cannot be an equilibrium, because it is separating and inefficient. While the high risks obtain full insurance the low risks obtain no insurance at all. So there must be other equilibria apart from the pooling, full insurance contract.

For the case where the utility function of the agent is of constant relative risk aversion, Asheim and Nilssen (1997) show that a combination of two contracts can be the equilibrium outcome: One is full insurance, which only the high risks, but not all of them choose. The other is a partial insurance contract which some of the high risks and all low risks choose. So the high risks are indifferent between the two contracts. The proportion of high risks buying this partial insurance contract must be such that for the group of buyers there does not exist a single contract which is better for everyone. Technically, the indifference curve of the low risk type in a two-states of the world diagram has to be steeper than or tangential to the (pooling) iso-profit line at this point. The derivation of this result is rather messy, as many stability criteria have to be checked. But note that with the arguments given above, if an equilibrium exists, that was to be expected: First, full insurance for everyone is only possible, if both types are pooled. The high risks would never accept a full insurance contract at a worse deal than the low risks obtain. Second, if full insurance at the pooling line is not acceptable for the low risks, then somehow the low risks must obtain partial insurance. Third, if only the low risks obtain partial insurance, that cannot be an equilibrium due to the general result given above. Therefore the partial insurance contract must be (partially) pooling. Fourth, the general result also tells us that the pooling ratio must be such that there does not exist a single contract which makes everyone better off, i.e., sufficiently many high risks have to buy this contract as well.[8]

With this we end the chapter on adverse selection. We have discussed one possible reason for many aspects which can be observed in the real world: Partial insurance contracts, categorical discrimination, experience rating or bonus-malus systems. Although the phenomena of adverse selection has been known for a long time, and the formalism was established more than than 20 years ago, markets under adverse selection are still a very active research area both in the insurance as well as in the general economics literature. On the applied side challenging questions exist, where a good grasp of adverse selection seems necessary to understand the economics: Two examples: Should an insurer be allowed to use information coming from genetic tests or not; why does the market for crop insurance do so badly in many countries?

On the theory side, many open questions concerning the foundations of adverse selection models remain: What is the appropriate way to model renegotiation? Does this perhaps solve the equilibrium non existence problem, or are there other characteristic features of the insurance market which so far have been overlooked?

Also on the empirical side much more work needs to be done to understand the severeness of adverse selection in different circumstances, and the applicability of the models presented here to the real world. Do insurer really screen the market or do they mainly pool risks? Which sectors of the insurance market suffer most under adverse selection?

[8]One clarifying remark: The result by Asheim and Nilssen does not depend on whether the low risks would choose the full insurance contract at the pooling premium or not. As a matter of fact, they discuss a monopolist insurer who would always choose a combination of the two contracts, i.e., some screening, than to offer full insurance for everyone.

We now turn to another problem of asymmetric information, which similar to adverse selection model has turned out to be one of the mostly discussed topics over the last twenty years: Moral hazard.

Part V

Asymmetric Information II: Moral Hazard

The introduction of the saftey belt was celebrated as a major step forward to reduce the number of fatal accidents. This seems obvious, as the risk of having major injuries is largely reduced. In many countries, saftey belts are now legally required to be worn.

However, although the number of injuries per accident decreased, the number of accidents actually increased. This latter effect was so dominant that the overall number of fatalities stayed roughly constant. And the incidence of injuries changed: Major injuries shifted away from the drivers of cars to pedestrians and cyclists. So altogether probably only the repair industry profitted from the introduction of safety belts.[9]

You might wonder what this has to do with insurance? Safety belts are like an insurance device. In case of an accident, you can be sure that less damage will occur to you. This is quite similar to an insurance contract, which in case of an accident pays out, making the damage less severe. Now people, who installed the 'safety belt insurance' felt less inclined to drive safe, as they would have if no saftey belt existed.[10] This is a common phenomena in the insurance market, known as moral hazard. Other examples are the following: Health insurance may induce people to be less carefully when doing dangerous sports; having a property insurance will make you think whether it is really necessary to take care of your premises; with a crop insurance, a peasant may work less hard to cultivate his fields.

We have been quite careful only to cite examples where individuals provide less effort in case of full insurance. In the literature, it is often assumed that individuals invest less financially in case of full insurance. For example, you acquire fire insurance and do not install fire sprinklers or extinguishers. You have earthquake insurance which leads you to build your house less earthquake-proof. The problem with the latter examples is that if the moral hazard problem consists of underinvestment, it is perhaps possible to write the efficient investment into the contract. Surely, a clause 'indemnity is only paid if Sprinkler system A is installed' would lead the homeowner to install that system. Even if monitoring of the investment is costly, it is possible a priori, and that should be included in the model. Therefore, in the following we concentrate on effort costs and not financial costs.

Unfortunately, the term 'moral hazard' has a second meaning in the insurance literature. It is also used if people who are health insured consume more health services than would be optimal. This effect arises because insurance companies pay for treatment only and do not indemnify the patient. To distinguish this effect from the underprovision of

[9]For a lucid account of this phenomena see Chapter 1 of Landsburg, 1993.

[10]People were asked whether they drove more dangerous with their saftely belts on - they declined. However, when they were asked whether they would drive more carefully if their car had no safety belts, they agreed.

effort as discussed here, will denote it as ex-post moral hazard, because the behaviour occurs *after* the accident has happened. Ex-post moral hazard will be discussed in chapter X.

Apart from the insurance context, models with (ex-ante) moral hazard cover a wide range of economical phenomena. Some examples: A manager, who is 'insured' by receiving a fixed wage, has no incentive to work hard. Banks, who invest into foreign countries, and which are 'insured' through bailouts by the IMF if that country collapses, have less incentives to screen the projects they finance carefully. Students, who have passed their midterm exam with a good grade, and which are thus 'insured' against failing the whole term, are less inclined to work hard for the final exam.

The last example is quite instructive as it gives us a hint of how to overcome some of the problems moral hazard creates. In many cases students do not just receive a pass/fail mark, they can also acquire different degrees like distinction, a price for the best exam, etc.. Thus, even if they are 'insured' against failing, there is still the incentive to work hard to obtain a good grade. In the models we present, a similar effect will hold: Individuals must somehow profit from the effort they put in. If they don't, they become lazy.

In the following we will unravel the literature step by step: We start with a simple example, only two outcomes and two effort levels ('lazy' and 'hard working'). This model serves two purposes: First, it discusses how the costs of effort can be modelled, and second, we see partial insurance appearing as a second best contract. Then the model is extended to more than two effort levels. In this context, the famous problem of the 'first-order approach' will be discussed. In a third step, continuous outcomes are considered. The most general case is instructive as it teaches us that not many general results can be obtained. However, one result which emerges is that moral hazard, although it creates inefficiencies, does not lead to a breakdown of the market.

Having derived the most general form in the static model, in the following chapter we turn to dynamic moral hazard problems, i.e. with more than one period. This is done in two steps: First, two periods are considered and second, infinitely many periods. The focus in this part lies on the question whether long-term contracts like e.g. experience rating can be useful to deal with moral hazard. Allowing for more than one period, renegotiation may become an issue again. This is discussed in the final section of chapter 15. In chapter 16 we turn to limited liability, which is another reason why people behave less carefully. Here we deal with the question whether mandatory insurance might be welfare improving or not. Finally, in chapter 17, insurance fraud is discussed.

Chapter 14

Single period contracts under moral hazard

14.1 The simplest model

As in the master model which reoccurs so often in this book, there are only two states of the world: One with no loss, the other where the loss occurs. In contrast to the standard model, in this chapter the probability of the damage is not exogenous, but can be influenced by the insured. Formally:

$$E[U] = (1 - \pi(e))U(W - P) + \pi(e)U(W - L + I^n) - c(e) \tag{14.1}$$

where we have introduced $I^n = I - P$ as the net payment in case of a loss. Here, $\pi(e)$ is the probability that a loss occurs, which satisfies $\pi'(e) < 0$, i.e. more effort (e) leads to a lower probability of an accident.

We have written $c(e)$ with the assumption that $c'(e) > 0$ as the 'cost of effort' in utility units ('utils'). It is obvious that if someone puts in effort to prevent an accident from happening, then this must be costly in some form. However, how to model this simple insight is far less trivial. Unfortunately, there is no axiomatic approach which can tell us how to do it optimally (like e.g. the axioms of expected utility lead to von Neumann-Morgenstern utility functions). Several possibilities exist, we discuss only two: First, costs could be monetary, i.e. $U(w, e) = U(w - c^m(e))$. The advantage of this way of modelling is that it is easily interpreted. However, as we have argued above, if costs are indeed monetary, in many cases contracts could condition on these costs. Then, the moral hazard problem would disappear.[1] An alternative, which we use in the following, is: $U(w, e) = U(w) - c(e)$, i.e. the utility function is additively separable in income and

[1] For an analysis with monetary effort costs, see Arnott and Stiglitz (1988 a).

effort. In this case the reduction in utility is independent of the state of the world, i.e. whether a loss has occured or not, 'costs' of effort are $c(e)$. Furthermore, the preferences over lotteries do not depend on the amount of effort taken.[2]

One might wonder why $U(w, e)$, the most general formulation would not be appropriate. The main reason for this and for the choice we make is practicability: We will see by going through the chapter, that it is not easy to always find a solution to the moral hazard problem. For many specifications, the mathematical problem is not well defined. Although we learn and have learned a lot from the moral hazard literature, this last point shows one weakness of it: Many results only hold for specific formulations.

This should not distract us from going through the models, as the few general results we obtain are quite powerful. Furthermore, also from specific models, there is much to learn on how the different effects interact.

Let us now go into the model. As mentioned above, there are only two states of the world. In addition, the agent has the choice between two effort levels: e_1 ('lazy') and $e_2 > e_1$ ('hard working').

First consider the 'first best', where effort is observable and contractable. We know from Chapter X that in that case full insurance will be optimal. The premium will be $P = \pi(e)L$, depending on the effort level. Therefore either effort level e_1 or effort level e_2 is optimal, depending on which of the two expressions is larger:

$$U(w - \pi_1 L) - c(e_1) >< U(w - \pi_2 L) - c(e_2)$$

where $\pi_i = \pi(e_i)$, $i = 1, 2$.

To make the problem interesting, let us assume that effort level e_2 is the first best effort level. If e_1 is preferred, then even in the case of non-observability of effort the moral hazard problem ceases to exist, as the agent can receive his full insurance contract as before and just stay lazy.

Now turn to the second best:
Suppose, that even under asymmetric information the higher effort is desired. The contract (premium/indemnity) must be designed such that the agent will indeed work hard.

[2]Another advantage of an additively separable utility function is that random contracts, where the indemnity is paid out with some probability smaller than one, are never optimal. In case of monetary effort costs, this results does not hold (see Arnott and Stiglitz, 1988 b CHECK!!).

The optimization problem is then the following:

$$\begin{array}{ll}
\max_{P,I^n} & (1-\pi_2)U(W-P)+\pi_2 U(W-L+I^n)-c(e_2) \\
& \text{s.t.} \\
\text{P.C.}: & (1-\pi_2)P-\pi_2 I^n \geq \Pi \\
\\
\text{I.C.}: & (1-\pi_2)U(W-P)+\pi_2 U(W-L+I^n)-c(e_2) \\
& \geq (1-\pi_1)U(W-P)+\pi_1 U(W-L+I^n)-c(e_1)
\end{array} \tag{14.2}$$

As in the last chapter, P.C. stands for participation constraint, i.e., the insurance company must obtain at least profit Π to agree to trade with the agent. Note, that by varying Π the whole efficiency boundary of this problem can be reached. Thus if $\Pi = 0$, we are in the competitive market situation. If Π is large enough, the solution to the monopoly problem will be obtained. This holds for moral hazard problems, because ex-ante both parties have the same information. The asymmetric information issue arises after the contract is signed, when the insured decides on which effort to choose. This is in contrast to adverse selection models, where one party has an informational advantage. There the structure of the result depends on how the bargaining power is distributed.[3]

I.C. is the incentive compatibility constraint. As effort is not contractable, the contract must be such that it is better for the agent to put in effort e_2 instead of e_1. This looks quite similar to the adverse selction problem discussed in the previous part ofthis book. There, however, the incentive compatibility constraint was such that it prevented one type of agent to choose the contract of the other type. Here, there is only one type of agent. But this agent must have an incentive to put in the desired effort level, which makes the I.C. necessary.

This is a Kuhn-Tucker problem, but fortunately we know that both constraints have to be binding. P.C. binds as otherwise through a decrease of P by $\epsilon/U'(W-P)$ and an increase of I^n by $\epsilon/U'(W-L+I^n)$ with $\epsilon > 0$ and small, the incentive constraint does not change, the participation constraint only changes marginally, which is all right if it was slack before, while the utility of the insured increases (check it!). Note that this follows from the assumption of additively separable utility functions. In the case of monetary costs, for example, it might happen that the P.C. does not bind at the optimum. The I.C. must be binding, because without it, we know that full insurance would be optimal. But that would lead the agent to put in effort e_1, which is a contradiction.

[3]In the insurance market under adverse selection, both in the monopoly problem as well as under perfect competition, the high risks obtain full insurance while the low risks are underinsured. However, for other forms of principal agent models the party whose contract is distorted may well depend on which party has the bargainig power.

The Lagrange function is then given by:

$$
\begin{aligned}
L = \; & (1-\pi_2)U(W-P) + \pi_2 U(W-L+I^n) - c(e_2) + \lambda[(1-\pi_2)P - \pi_2 I^n - \Pi] \\
& +\mu[(1-\pi_2)U(W-P) + \pi_2 U(W-L+I^n) - c(e_2) \\
& -(1-\pi_1)U(W-P) - \pi_1 U(W-L+I^n) + c(e_1)]
\end{aligned}
\tag{14.3}
$$

The first order conditions with respect to P and I^n are:

$$
\begin{aligned}
& -(1-\pi_2)U'(W-P) + \lambda(1-\pi_2) - \mu[(1-\pi_2)U'(W-P) - (1-\pi_1)U'(W-P)] = 0 \\
& \pi_2 U'(W-L+I^n) - \lambda\pi_2 + \mu[\pi_2 U'(W-L+I^n) - \pi_1 U'(W-L+I^n)] = 0
\end{aligned}
\tag{14.4}
$$

If we denote $U'(W-P) = U_1'$ and $U'(W-L+I^n) = U_2'$, then the first order conditions can be written as:

$$
\begin{aligned}
\tfrac{1}{U_1'} &= \lambda^{-1} + \tfrac{\mu}{\lambda}\tfrac{(1-\pi_2)-(1-\pi_1)}{(1-\pi_2)} \\
\tfrac{1}{U_2'} &= \lambda^{-1} + \tfrac{\mu}{\lambda}\tfrac{\pi_2-\pi_1}{\pi_2}
\end{aligned}
\tag{14.5}
$$

An expression of this form will reappear over and over in the literature on moral hazard. One over the marginal utility is equal to a constant plus another constant times an expression, which depends positively on the change in probability for different effort levels for that state and negatively on the probability of that state. Note that the last factor can also be written as $1-(1-\pi_1)/(1-\pi_2)$ (and $1-\pi_1/\pi_2$ respectively). The ratio of the probabilities is also known as the likelihood ratio. If the likelihood ratio becomes smaller, then U' decreases which implies that the wealth in that state increases. For a very small likelihood ratio one can be relatively sure that this state arises under the desired effort level. I.e. if $(1-\pi_1)/(1-\pi_2) = 0.1$, then it is ten times more likely to obtain the state no-accident if effort level e_2 is used instead of e_1. Then consumption is large in those states to give an incentive to work hard. In this example, if $\pi_2 = 0.1$, π_1/π_2 will be equal to 9.1. So, in the accident state consumption will be much lower to give a strong incentive to prevent the accident, which is quite unlikely under the larger effort level.

In the present case note that as $\pi_2 < \pi_1$, we have $U_1' < U_2'$ which implies that $P < L - I^n$, i.e., there is less than full insurance. Although that was to be expected, it is useful to denote this as the first general insight:

To implement higher effort levels, the agent must not obtain full insurance.

This is one example of what was mentioned in the introduction, the agent has to profit somehow from putting in more effort. Here, this is achieved by giving him a larger utility if no accident occurs than in case of an accident. A common feature observed in insurance contracts, namely deductibles and/or partial insurance, can be explained by this.

The problem is not solved yet. We have calculated how the contract would look like if the higher effort level is implemented. It is not clear, however, whether this is optimal. Although in a first best world, a higher effort level may be preferred, in a second best world it might be better to implement the lower effort level (see Exercise 14.1). So to find the overall solution, one has to check whether the high effort level with partial insurance or the low effort level with full insurance will give the insured a larger utility.

14.2 Many effort levels

In this section we extend the model in one direction, namely to allow for more than two effort levels. The formal implementation of many effort levels gave rise to a lengthy debate in the literature which goes under the heading of the 'First-Order Approach'. We will say more on this later.

Suppose the possible effort levels are $e \in E$, where E is some discrete or continuous set. The problem, which has to be solved, is the following:

$$\begin{array}{ll} \max_{e\in E,P,I^n} & (1-\pi(e))U(W-P)+\pi(e)U(W-L+I^n)-c(e) \\ & \text{s.t.} \\ \text{P.C.}: & (1-\pi(e))P-\pi(e)I^n \geq \Pi \\ & \\ \text{I.C.}: & e = \text{argmax}_{\tilde{e}\in E} \quad [(1-\pi(\tilde{e}))U(W-P)+\pi(\tilde{e})U(W-L+I^n)-c(\tilde{e})] \end{array} \tag{14.6}$$

The optimal contract has to be such that the agent prefers to choose effort e, i.e. e must maximize his utility given the contract. The way the I.C. is written captures this problem, but is unfortunately not very helpful for finding a solution. How do we deal with an *argmax* function? There are two possibilities on how to treat the I.C. in such a way that standard analysis can be used. One is to have discrete effort levels, i.e. $E = \{e_1, e_2, ...\}$ and write the incentive compatibility constraint for each effort level:

$$\begin{array}{ll} \forall e_i \in E \;\; \text{with} \;\; e_i \neq e & (1-\pi(e))U(W-P)+\pi(e)U(W-L+I^n)-c(e) \geq \\ & (1-\pi(e_i))U(W-P)+\pi(e_i)U(W-L+I^n)-c(e_i) \end{array}$$

This direction was pursued by Grossman and Hart (1986). The big advantage of this approach is that, together with a finite number of outcomes, it is possible to show that the maximisation problem given in (14.6) is well-defined, i.e., after reformulation the Kuhn-Tucker Ansatz satisfies the conditions of a concave programming problem, for which we know that a solution exists.

The other possibility, and that is how we will proceed, is to use continuous effort levels and replace the incentive compatibility constraint by the first order condition for the agent. This is the so-called *First Order Approach*, among others used by Holmström (1979). It was Mirrlees who detected a potential flaw in this approach - it might not be well defined!

The problem is that it is not clear whether the first order condition for the agent does describe the unique maximum: it could well describe a minimum, a saddle point, or a local, but not global maximum. If one wants to use the first order approach, one therefore always has to check whether the problem is well defined. Fortunately, in the present case it is, if we make a further assumption on the second derivative of the cost and probability function. Let us check this. For an interior solution, the I.C. can be replaced by:

$$\text{I.C.}: \quad -\pi'(e)[U(w-P) - U(w-L+I^n)] - c'(e) = 0 \tag{14.7}$$

With the assumptions above, namely $\pi'(e) < 0$, $c'(e) > 0$ it already follows that $U(w - P) > U(w - L + I^n)$, i.e. partial insurance. Now we have to check the second order condition:

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

This holds for any partial insurance contract if $c''(e) > 0$ and $\pi''(e) > 0$, i.e. costs of effort are convex, and probability is a convex function of effort. Larger effort becomes more and more costly, and less and less productive.

The I.C. above already tells us that to implement any effort level larger than e_{min}, partial insurance is necessary. This is a very neat way of proving this result.

Allowing for more than two effort levels, another issue arises. Does the agent work harder or less hard in a situation under moral hazard compared to the first best, i.e. where effort is observable. As a first guess one would expect that due to the unobservability of effort the agent will work less hard. However, as you are asked to show in a specific example in exercise 14.1., the following holds:

> *In a second best world, the agent may either work less or harder than in the first best world.*

The trade-off in moral hazard problem is: Should I provide more extensive insurance, which is good as the agent is risk averse, or is less insurance better, as this gives the agent an incentive to avoid the accident. In a second best world, it is not at all clear which of these effects is more dominant. Therefore nothing can be said about the effort level compared to the first best case.

So far only two outcomes were possible. A partial insurance contract could therefore be either a contract with a deductible, or with coinsurance, or with a combination of these two. We now turn to continuous outcomes, to shed more light on the optimal contract structure.

14.3 Continuous losses

This section discusses the most general case, with a continuous loss distribution. Here we distinguish two cases: Loss-prevention and loss-reduction. The former is easy to define and refers to the case, where the agent can influence the probability of a loss. The definition of the latter is slightly more complicated, and we will confer the discussion to one after the next subsection.[4]

14.3.1 Loss-prevention

Losses are random with a distribution function $F(L)$ and density $f(L)$, defined on $L \in [\underline{L}, \bar{L}]$. As before, the agent only controls the probability that a loss occurs, and not the distribution of losses.

With the help of the 'first order approach', the optimisation problem becomes the following:

$$\begin{array}{lll} \max_{e,P,I^n(L)} & (1-\pi(e))U(W-P) + \pi(e)\int_{\underline{L}}^{\bar{L}} U(W-L+I^n(L))f(L)\mathrm{dL} - c(e) & \\ & \text{s.t.} & \\ \text{P.C.}: & (1-\pi(e))P - \pi(e)\int_{\underline{L}}^{\bar{L}} I^n(L)f(L)\mathrm{dL} \geq \Pi & (14.8) \\ & & \\ \text{I.C.}: & \pi'(e)[-U(W-P) + \int_{\underline{L}}^{\bar{L}} U(W-L+I^n(L))f(L)\mathrm{dL}] - c'(e) = 0 & \end{array}$$

Let us quickly check whether the second order condition for the agent has the correct sign:

$$\pi''(e)[-U(W-P) + \int_{\underline{L}}^{\bar{L}} U(W-L+I^n(L))f(L)\mathrm{dL}] - c''(e) < 0$$

which is satisfied if, as before, $\pi''(e) > 0$ and $c''(e) > 0$. An additional assumption, which is often used in insurance economics is that the indemnity should not be negative. The motivation for this is that if the insured faces a negative indemnity he would not report a loss. So this introduces an additional constraint: $I(L) \geq 0$ (or $I^n(L) \geq -P$). Then, at the optimum, it holds:

$$\pi(e)U'(W-L+I^n(L))f(L) - \lambda\pi(e)f(L) + \mu\pi'(e)U'(W-L+I^n(L))f(L) \leq 0 \quad (14.9)$$

and $I^n(L) = -P$ if expression (14.9) is strictly lower than zero.

[4] In the literature, loss-prevention is also known as self-protection, while loss-reduction is referred to as self-insurance, (see e.g. Ehrlich and Becker, 1972). These wordings are ambiguous. It is not clear, for example, whether the common notion of protecting oneself does not also refer to loss reduction. For example, a bullet proof veste as a means of self protection does not prevent an attempt of murder, but it lowers the severity of the attack.

If the equality sign holds, reformulating this expression yields:

$$\frac{1}{U'(W - L + I^n(L))} = \lambda^{-1} + \mu/\lambda \frac{\pi'(e)}{\pi(e)} \tag{14.10}$$

This equation looks quite similar to equation (14.5): one over the marginal utility is equal to some constant plus a term which depends on the change in probability divided by the probability. This latter expression is also known as the differential form of the likelihood ratio. Note that the right hand side does not depend on L. This implies, that $I^n(L) - L$ has to be constant. By inspection of the incentive compatibility constraint it it clear that this constant has to be lower than P. With the additional constraint that $I^n(L) + P$ cannot be negative, this leads to the following result:

> *If the agent can only influence the probability of an accident, then the optimal insurance contract has a deductible:* $I(L) = \max[L - D, 0]$.

This result is very instructive as it shows how incentive contracts operate. In this case, the agent can only influence the probability of loss, not the loss distribution. So to make him work hard, he needs to be punished in case a loss occurs, but rewarded for no loss. This is achieved by giving the agent an income of $W - P$ in the no-loss state, but $W - P - D$ in the loss state, if loss exceeds D. Apart from small losses, where we assumed that the agent cannot pay money back to the insurance company, the agent has the same income independent of the size of the loss. That is, he is fully insured against variations in the size of the loss. There is no reason to distort the agent, e.g. to provide full insurance for low loss levels, and partial insurance for high loss levels, as the agent cannot influence the distribution of losses. However, the agent is only partially insured against the occurence of a loss, as he has to pay the deductible D himself. This is done to give him an incentive to reduce the loss probability.

Such an intuition also holds in a more general context: When designing an incentive contract, one has to be aware which quantities are influenced by the agent, and which are not, and condition the contract only on the former. For example, it used to be (and sometimes still is) quite common to pay asset managers an 'incentive' contract, which conditions on the return they make on a portfolio. On first glance, this makes sense, as the return is what counts in the end. However, if say the portfolio consists entirely of stocks, the job of the asset manager is to do better than some stocks index like the Dow Jones. Therefore the wage should condition on the relative performance of the portfolio with respect to that index.

Turning the argument around, another result can be shown to hold: If something observable and contractable is influenced by the agent's effort, then the contract should condition on this (Holmström, 1982, see also exercise 14.2). This so called 'sufficient statistic result' is quite strong, as it implies that optimal contracts should condition on possibly very many quantities. For example, in the case of a car accident, the indemnity should condition on the speed of the car, whether the radio was turned on or not, whether the driver was in a phone coversation, etc. as long as these quantities are correlated with the preventive effort the agent has taken, and if they can be observed ex-post. Partially, this is achieved by the negligience clause, i.e. the insurance company pays less if someone behaved negligient.

The preventive effort a car drivers exerts does not only influence the probability of an accident, but also the severity. This is where we turn to now.

14.3.2 Loss-reduction

Loss reduction refers to a situation where by putting in effort the agent can influence the size of the loss. The straightforward way to formalize loss reduction would be to let loss be a function of effort, i.e., $L = L(e)$ with $L'(e) < 0$. However, if this is a deterministic function, even if e is not observable, the first best can be achieved. How?

Determine the first best effort level e^{FB} and premium P^{FB} in a contract with full insurance, i.e. $I^{FB} = L(e^{FB})$. In the second best world, where effort is not observable, consider the following contract: In case of a loss the insurance company pays the agent the size of the loss if the realized loss is smaller than $L(e^{FB})$. If the realized loss is larger than $L(e^{FB})$ the agent receives nothing. For this service, the agent has to pay P^{FB}. What will the agent do? He will surely not work harder than e^{FB}, as this leads to the same full insurance outcome, but with larger effort costs. If he works less than e^{FB}, in case of a loss he will receive nothing. But this cannot be better than working e^{FB}. So the first best is obtained.

There are two criticisms one might have with the derivation of the 'second best' contract above: First, one might argue that the insurance company usually does not know the loss function. This is a fair comment, but to model this, asymmetric information has to be imposed, which together with moral hazard is an even harder problem to solve. Second, the insurance company might not observe the size of the loss with a precision necessary for such a contract. However, if the size of the loss is not observable in general, the agent could always claim a much larger loss than has actually occured. This phenomenon will be discussed later in the chapter on insurance fraud. So to conclude, although very intuitive, a deterministic loss function is unfortunately not useful at all to model a moral hazard problem of unobservable loss reduction activities.

The only way out of this problem is to assume that the loss function is stochastic, where

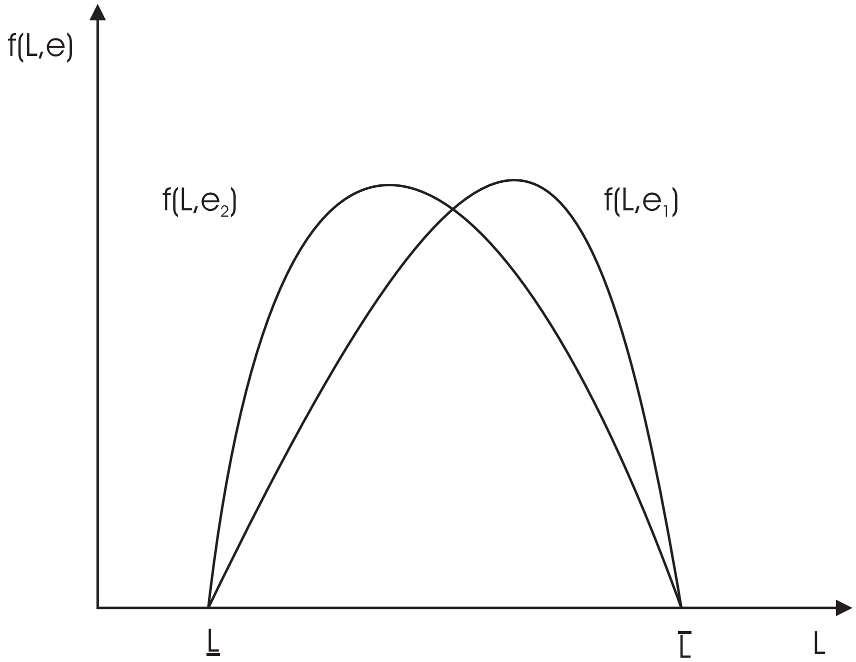

Figure 14.1: Accident propability density for two effort levels.

each effort level determines another distribution function $F(L,e)$ with density $f(L,e)$. Larger effort than implies a reduction in expected loss.

Larger effort shifts the loss curve such that the mean moves to the left. Also here we have to be very careful. Larger effort must shift the curve in such a way that the support of the loss function stays the same for all effort levels (see Figure 14.1).

If for different loss levels the supports were different, then to prevent the agent from engaging in a particular effort one could penalize him very hard if losses occur which are impossible under the other effort levels. But then, the first best would again be possible.[5]

By using the first order approach, the optimisation problem becomes:

$$\max_{e,P,I^n(L)} \quad (1-\pi)U(W-P) + \pi \int_{\underline{L}}^{\bar{L}} U(W-L+I^n(L))f(L,e)\mathrm{dL} - c(e)$$

s.t.

$$\text{P.C.}: \quad (1-\pi)P - \pi \int_{\underline{L}}^{\bar{L}} I^n(L)f(L,e)\mathrm{dL} \geq \Pi \tag{14.11}$$

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

Check the second order condition for the agent:

$$\pi \int_{\underline{L}}^{\bar{L}} U(W-L+I^n(L))f_{ee}(L,e)\mathrm{dL} - c''(e) < 0 \qquad ?$$

[5]Economists like to call this punishment: 'burn the agent in oil'. This however only works, if the agent can be punished sufficiently hard. In the case of limited liability, which is discussed in Chapter 16, this is not possible.

Here is where the problem alluded to earlier arises. It is not clear at all, that this expession is smaller than zero for *any* net indemnity function $I^n(L)$. f_{ee} is the second derivative of the distribution function with respect to effort, which can well be positive for some effort levels. As mentioned above, there is a large literature around the first order approach, and only under restrictive assumptions can this approach be justified.[6]

We do, as many have done in the literature before us, close both eyes and assume that the first order condition describes the global maximum. But note, that if you want to use this approach for a given specific problem, once you have found an optimal premium-indemnity schedule, you have to check whether the second order condition indeed holds.

Given the above maximization problem, the first order condition with respect to $I^n(L)$ is:

$$\pi f(L,e)U'(W-L+I^n(L)) - \lambda\pi f(L,e) + \mu\pi f_e(L,e)U'(W-L+I^n(L)) = 0 \quad (14.12)$$

or, after reformulating:

$$\frac{1}{U'(W-L+I^n(L))} = \lambda^{-1} + \mu/\lambda\frac{f_e(L,e)}{f(L,e)} \quad (14.13)$$

Again, an equation quite similar to equations (14.5) and (14.10) above.

What can be said about the form of the optimal contract? Actually not much at all. There exist distribution functions, for which the indemnity is not even increasing in the size of the loss, or where it increases faster than the loss. However, in exercise 14.4. we show that if the distribution function satisfies the monotone likelihood ratio property (MLRP), which says that the function $f(L,e_2)/f(L,e_1)$ is decreasing in L if e_2 is larger than e_1, than at least we can deduce that $-L+I^n(L)$ is a decreasing function, i.e. the agent has lower utility for larger losses. MLRP can be interpreted as implying that a larger loss is more probable to have been occured under the lower effort level. Although MLRP is quite often assumed, it must not hold necessarily. Furthermore, it can still be true that $I^n(L)$ decreases for larger losses.

These are really bad news. The result can be quite arbitrary, and may not even look like an insurance contract at all. So the predictive power for real world insurance contracts is quite weak. One way to remedy this problem is to make further assumptions on the set of potential contracts. In the last section we have already assumed that insurance contracts cannot specify a negative indemnity. Furthermore, quite often it is also assumed that the indemnity cannot be larger than the size of the loss $I \leq L$, to prevent overinsurance, as this would give an incentive to enforce the accident. If the agent can manipulate the size of the loss as well, then $0 \leq I'(L) \leq 1$ has to hold. All these assumptions, in addition with

[6]See e.g. the papers by Rogerson (1985) and Jewitt (1988), and Exercise 14.3.

MLRP yield an optimal contract with possibly full insurance for low loss levels, partial insurance thereafter and a non-decreasing indemnity schedule. But note that most of these features are driven by the additional assumptions we made, and not by the moral hazard problem.

One might argue that the generality we use for both the utility function and the distribution function are really not necessary. Firms do not know much about $f(L,e)$. They might know $f(L)$ from past experience, however how this distribution function changes for different effort levels is sureley not well-known. Also, the insured has very little knowledge on how his effort influences the loss-statistic exactly. It might perhaps be reasonable to assume a simple form for the distribution function. A first attempt would be to model the loss distribution function as a normal distribution, where the agent controls the expected value. That is, $f(L,e) = \sqrt{2\pi\sigma^2}^{-1} \exp -\frac{(L-(\bar{L}-e))^2}{2\pi\sigma^2}$. In that case $f_e(L,e)/f(L,e) = a + bL$, where a and b are some constant parameters (see Exercise 14.5). This is a very simple expression. If in addition the individual is assumed to have a logarithmic utility function $u(w) = ln(w)$, and using the result from equation (14.13), linear contracts would seem to be the outcome. I.e. insurance contracts specify a percentage rate of losses they cover. Unfortunately this conclusion is wrong: If, as we assumed, losses are normally distributed, L can take on very large values. With a linear indemnity function, wealth can become negative. However, the logarithm of a negative number is not defined. Somehow it seems that we must have done something wrong with the mathematics. A well defined model, a correct calculation, but now this caveat. As a matter of fact, this is an example of a problem where the first order approach fails. Mirrlees showed (and we do this in exercise 14.5) that if the agent can be punished sufficiently hard, the first best can be closely approximated with the help of a trigger contract similar to the one we have discussed above for deterministic loss functions: full insurance for losses below some value L^* and punishment (no insurance) for loss levels above this value. So we should keep in mind that the results presented above only hold if the first order approach is valid.[7]

On the positive side, one result which is worth mentioning, is that the insurance market will not break down completely, i.e. it is never optimal to sell no insurance contract at all. This can be seen in the following way: $(P, I^n(L)) = (0,0)$ would imply a corner solution to the maximization problem (14.11), which yields for the first order condition with respect to $I^n(L)$:

$$\pi f(L,e)U'(W-L) - \lambda\pi f(L,e) + \mu\pi f_e(L,e)U'(W-L) \leq 0 \tag{14.14}$$

[7] Also, as Grossmann and Hart (1983) have shown in an example with discrete effort levels, MLRP is not sufficient to guarantee that the agent is worse off for larger loss levels.

On the other hand, taking the derivative of the Lagrange function with respect to P at $P = 0$ gives:

$$-(1-\pi)U'(W) + \lambda(1-\pi) \leq 0 \tag{14.15}$$

As $\int_{\underline{L}}^{\bar{L}} f(L,e)dL = 1$ for all effort levels, it follows that $\int_{\underline{L}}^{\bar{L}} f_e(L,e)dL = 0$. So there exists values of L at which $f_e(L,e) > 0$. For those loss levels the inequality (14.14) implies that $U'(W-L) \leq \lambda$, while from the inequality (14.15) it follows that $U'(W) \geq \lambda$ which is a contradiction. So moral hazard creates inefficiencies, but does not lead to a complete market breakdown.[8]

To summarize the results we have obtained in the static model:

- If the agent should provide preventive effort, then partial insurance is necessary.
- The second best effort level can be lower or larger than the first best effort level.
- Moral hazard leads to inefficiencies, but the market does not break down completely.
- 'Sufficient Statistic Result': The optimal contract should not condition on quantities which do not reveal any additional information on the choice of effort by the agent. On the other hand, every signal whose occurence does provide information about the effort chosen by the agent, should be part of the optimal incentive scheme.
- In case of loss-prevention, a simple deductible contract is optimal.
- In case of loss-reduction, the optimal contract depends on the exact characteristics of the environment. The structure of the contract is closer to a coinsuance contract than to an insurance policy with a deductible.

  We now ask the question whether, like in the case of adverse selection, a dynamic contract which lasts over several periods can give better incentives to the agent and will avoid some of the inefficiencies.

[8] This is different under uniform pricing, where insurance firms can only set premium rates (see Exercise 14.6.).

Chapter 15

Multi-period contracts and renegotiation

15.1 Many periods

To understand the economics of long-term contracts, we start with a simple two period model, where the moral hazard problem only occurs in the first period. In both periods, the individual can either work hard or be lazy. However, for the moment we assume that in period 2 he will work hard for sure. In addition, we assume that there are only two states of the world in each period.

Define the following quantities:
e_i: effort of the insured: $i = 1$: lazy, $i = 2$: hard working.
c_i: utility costs of effort if the insured exercises e_i.
π_i: probability of an accident, given effort e_i.
C_k^1: consumption of the agent in period 1, if no accident ($k = n$) or if an accident happened ($k = a$).
C_{kj}^2: consumption of the insured in period 2, if in period 1 no (a) accident happend ($k = n$ or $k = a$ respectively) and if an accident did not or did happen in period 2 ($j = n$ or $j = a$ respectively).

If the individual works hard in period 2, and exercises effort e_i in period 1, his expected utility is given by:

$$(1-\pi_i)U(C_n^1) + \pi_i U(C_a^1) - c_i + (1-\pi_i)[(1-\pi_2)U(C_{nn}^2) + \pi_2 U(C_{na}^2) - c_2] + \pi_i[(1-\pi_2)U(C_{an}^2) + \pi_2 U(C_{aa}^2) - c_2] \tag{15.1}$$

It is assumed that the insured does not discount. Discounting would not change the result.

Assume that the insurer would like the agent to work hard in period 1. Otherwise, full insurance in period 1 would be optimal. Under the assumption of a competitive market, where the insurer makes zero profit, the optimal contract is then given by the solution to the following maximization problem:
Maximize (15.1) with $i = 2$, subject to

$$\begin{gathered}(1-\pi_2)[U(C_n^1)+(1-\pi_2)U(C_{nn}^2)+\pi_2 U(C_{na}^2)]+\pi_2[U(C_a^1)+(1-\pi_2)U(C_{an}^2)+\pi_2 U(C_{aa}^2)]-2c_2 \geq \\ (1-\pi_1)[U(C_n^1)+(1-\pi_2)U(C_{nn}^2)+\pi_2 U(C_{na}^2)]+\pi_1[U(C_a^1)+(1-\pi_2)U(C_{an}^2)+\pi_2 U(C_{aa}^2)]-c_1-c_2\end{gathered}$$

$$(1-\pi_2)[C_n^1+(1-\pi_2)C_{nn}^2+\pi_2 C_{na}^2]+\pi_2[C_a^1+(1-\pi_2)C_{an}^2+\pi_2 C_{aa}^2] \leq \tilde{W}^1+\tilde{W}^2 \tag{15.2}$$

The first constraint is the incentive compatibility constraint: The insured must be better off by working hard in period 1, then by staying lazy. The second constraint is the participation constraint of the insurer. The consumption of the agent must not be larger than his overall wealth in the two periods, which is given by $\tilde{W}^1 = W^1 - \pi_2 L$ plus $\tilde{W}^2 = W^2 - \pi_2 L$. W^i denotes the wealth of the insured in period i, if no accident occurs. $\pi_2 L$ is the fair insurance premium for each period. Note that it is equivalent whether we specify the premium and indemnity in every period or the consumption of the agent, as long as the agent cannot transfer money between periods. As it turns out, discussing consumption levels simplifies the analysis.

Denoting the two Lagrange parameters by λ and μ, and taking the first order conditions with respect to $C_n^1, C_{nn}^2, C_{na}^2$ yields:

$$\begin{array}{ll} C_n^1 & U'(C_n^1)[(1-\pi_2)-\lambda((1-\pi_2)-(1-\pi_1))]-\mu(1-\pi_2)=0 \\ C_{nn}^2 & U'(C_{nn}^2)[(1-\pi_2)^2-\lambda((1-\pi_2)^2-(1-\pi_1)(1-\pi_2))]-\mu(1-\pi_2)^2=0 \\ C_{na}^2 & U'(C_{na}^2)[(1-\pi_2)\pi_2-\lambda((1-\pi_2)\pi_2-(1-\pi_1)\pi_2)]-\mu(1-\pi_2)\pi_2=0 \end{array} \tag{15.3}$$

And therefore:

$$U'(C_n^1) = U'(C_{nn}^2) = U'(C_{na}^2) \tag{15.4}$$

The agent is fully insured in period 2 if no accident occured in period 1 ($C_{nn}^2 = C_{na}^2$), and in addition, he consumes as much in period 2 as he does in period 1. By taking the other derivatives it follows that also $C_a^1 = C_{an}^2 = C_{aa}^2$, but with $C_a^1 < C_n^1$.

The intuition for this result is the following: The agent receives less in state of an accident than in case of no accident. This gives him an incentive to work hard in the first period. In the second period, the agent has no influence on the accident probability, so he is fully insured. That he consumes as much in the second period as he does in the first, is the *income smoothing* effect. The agent prefers a constant income stream to a variable one. With this construction, the incentive is distributed on the two periods: If an accident happens, the agent is not only worse off in this period, but also in the next period. In other words, the agent still feels the consequences of todays accident tomorrow. On the other hand, this also implies that the deductible the agent has to pay in period 1 is less severe than the according deductible in a one period model. Actually, if the number of periods is increased, and the incentive problem is still only in the first period, the first best will be closer and closer approximated. The difference between the consumption levels following an accident or no accident becomes smaller and smaller.

This is a noteworthy result: Many periods allow that incentives for todays work are distributed over time. A similar result holds for example for a manager: She might work hard today not just because of this years bonus payment, but also to increase her chances to become CEO in 10 years time.

Observe that the agent obtains full insurance in period 2. This is also what he would get if he were to buy a short term insurance contract in period 2, as there is no incentive problem at this stage. So one might wonder whether this consumption pattern could not also be achieved with a series of short term contracts. Actually, it can. Write $C_n^1 = W^1 - P^1 - S_n$, $C_{nn}^2 = C_{na}^2 = W^2 - \pi_2 L + S_n$ where P^1 is the premium the agent has to pay in period 1, S_n is the amount of money he saves in period 1, and $\pi_2 L$ is the fair premium in period 2. (Similarly, $C_a^1 = W^1 - P^1 - L + I^1 - S_a$ and so on.) Working backwards we see that two short term contracts will achieve the optimal outcome. In period 2, the insured buys a full insurance contract at the fair premium $\pi_2 L$. If he has no accident in period 1, he has wealth of $W^1 - P^1$. Depending on the size of W^2, the agent will either save or lend money, in order to smooth his income across the two periods. In any case, he will choose S_n such that $W^1 - P^1 - S_n = W^2 - \pi_2 L + S_n$. In the first period this behaviour will be anticipated from the insurer, so she sets (P^1, I^1) such that the agent obtains the same final consumption stream as with the optimal long term contract, which also induces hom to work hard in period 1.

This is another remarkable result: Although, as we have seen above, the incentives are distributed across periods, this can be achieved with single period contracts. Thus there is no need to write a long-term contract.

We now turn to the more complicated case, where the agent can influence his effort also in period 2. A similar result as above will hold, but now we have to assume that the insurer can control the saving of the agent, which was not necessary before. But as before we will see that in period 2 the agent is better off if he had no accident in period 1, and this can again be achieved with short term contracts.[1]

If the agent influences the risk probability in period 2 as well, and the technology stays the same across periods, then two further incentive constraints have to be fullfilled:

$$\begin{aligned}(1-\pi_2)U(C^2_{nn})+\pi_2 U(C^2_{na})-c_2 &\geq (1-\pi_1)U(C^2_{nn})+\pi_1 U(C^2_{na})-c_1\\ (1-\pi_2)U(C^2_{an})+\pi_2 U(C^2_{aa})-c_2 &\geq (1-\pi_1)U(C^2_{an})+\pi_1 U(C^2_{aa})-c_1\end{aligned} \tag{15.5}$$

Whether an accident occured or not in period 1, the agent must prefer to work hard in period 2 than to stay lazy.

Denoting the Lagrange parameters for these two additional constraints as λ_1 and λ_2, taking the first order conditions of the maximization problem (15.1) with respect to $C^1_n, C^2_{nn}, C^2_{na}$, and rearranging yields:

$$\begin{aligned}U'(C^1_n) &= [1+\lambda\tfrac{\pi_1-\pi_2}{1-\pi_2}]^{-1}\mu\\ U'(C^2_{nn}) &= [1+\lambda\tfrac{\pi_1-\pi_2}{1-\pi_2}+\lambda_1\tfrac{\pi_1-\pi_2}{(1-\pi_2)^2}]^{-1}\mu\\ U'(C^2_{na}) &= [1+\lambda\tfrac{\pi_1-\pi_2}{1-\pi_2}+\lambda_1\tfrac{\pi_2-\pi_1}{(1-\pi_2)\pi_2}]^{-1}\mu\end{aligned} \tag{15.6}$$

and therefore $C^2_{nn} > C^1_n > C^2_{na}$. A similar result can be obtained for the accident case (Do it!).

The agent is better off in period 2 if he has no accident than if he has an accident. This is necessary to give the agent the right incentive in period 2. Interestingly he consumes more in period 2 if he has no accident than in period 1. This is the income smoothing effect. The agent 'saves' money in period 1 to cover for the bad days in period 2, namely the accident state. If however no accident happens, then he can consume more. This is a structure we know from the real world, where it comes in the form of bonus-malus contracts.

It is slightly tricky to show that this outcome can also be obtained with short term contracts only. Short term contracts differ from long term contracts in two respects. First, a short term contract maximizes the utility of the agent in period 2 conditional on the I.C. and a zero profit condition for the insurer. Thus, after the agent obtains his final wealth at the end of period 1, savings and insurance for period

[1]See also Fudenberg, Homström and Milgrom (1990) and Malcomson and Spinnewyn (1988).

2 must be chosen such that his utility is maximized. So far, the overall utility of the agent was maximized, but not necessarily his second period utility. The second difference is that the insurer has to make zero profit in both periods, while with a long term contract she could have shifted profits across periods. Both differences can be summarized in the following additional constraint, which we write for simplicity only for the case where no accident happened in period 1. The consumptions $C_n^1, C_{nn}^2, C_{na}^2$ are the solution to a short term contract in period 2, if they maximize the following expression

$$U(C_n^1) + (1-\pi_2)U(C_{nn}^2) + \pi_2 U(C_{na}^2) - c_2 \tag{15.7}$$

subject to the first constraint in (15.5) and

$$C_n^1 + (1-\pi_2)C_{nn}^2 + \pi_2 C_{na}^2 \leq W_n^1 + \tilde{W}^2 \tag{15.8}$$

where $W_n^1 = W^1 - P^1$ is the wealth the agent has at the end of period one, before he decides how much to consume at period one, if no accident has happened. The maximization problem guarantees that the untility of the agent is maximized. The additional condition ensures that the insurer makes zero profits.

We now show that both additional features of a short term contract will not change the optimal contract structure. Take the second point first. Assume, a long term contract is such that it makes a loss in the second period, if no accident happened in the first period. If the first period payment is given by P_1 and S_n^1 is the amount the agent saves, such that $C_n^1 = W^1 - P^1 - S_n^1$, then a second period loss by the insurer implies:

$$-S_n^1 + (1-\pi_2)C_{nn}^2 + \pi_2 C_{na}^2 = \tilde{W}^2 + \Pi$$

where Π is size of the loss. Consider the following long term contract instead: Decrease both P^1 and S_n^1 by Π. Keep all the other consumptions constant. Note that with this change, the agent still consumes in every state the same amount, as C_n^1 depends only on the difference between P^1 and S_n^1. But for the insurer, a profit which so far accrued in period 1, is now shifted to period 2. And his overall profit in period 2 is zero. The same reasoning holds if the insurer makes a profit instead, i.e. $\Pi < 0$.

This comes about because the insurer fulfills two jobs: She insures the agent, but she also shifts money across periods. As long as she controls the savings of the agent, and with that the consumption of the agent, it is mere accounting whether the profit occurs earlier or later.

Having satisfied ourself that (15.8) is no real constraint, we now turn to the utility of the agent. The proof that the agent's utility in period 2 is also maximized with an optimal long term contract, works by contradiction. Suppose the long term contract is such that the utility is not maximized in period 2. I.e. there exists a different set of consumptions $C_n^1(*), C_{nn}^2(*), C_{na}^2(*)$ which still satisfies (15.8) and the first constraint of (15.5) but which makes the agent better off in period 2. But then consider the following long term contract instead: In period 1, if no accident happens do the following: With probability θ, let the agent pay P^1 as before, but now choose the consumption pattern $C_n^1(*), C_{nn}^2(*), C_{na}^2(*)$. With probability $(1-\theta)$ choose $P_-^1 > P^1$, and the optimal consumption in period 2 for this (lower) wealth, which we denote by $C_n^1(-), C_{nn}^2(-), C_{na}^2(-)$. P_-^1 must be so large, that the agent is worse off with his contract in period 2 than he was with the original long term contract. So instead of the former long term contract, the agent obtains if no accident occurs either a contract which specifies the same premium and makes him better off, or a contract which gives him a really bad time. The probability of these two contracts (θ) is chosen such that

$$U(C_n^1) + (1-\pi_2)U(C_{nn}^2) + \pi_2 U(C_{na}^2) - c_2 = \theta[U(C_n^1(*)) + (1-\pi_2)U(C_{nn}^2(*))+ \\ \pi_2 U(C_{na}^2(*)) - c_2] + (1-\theta)[U(C_n^1(-)) + (1-\pi_2)U(C_{nn}^2(-)) + \pi_2 U(C_{na}^2(-)) - c_2]$$

Note that with such a contract, the agent is not worse off overall than he was with the original long term contract. Also, his incentive to work hard in period 1 has not changed. The insurer on the other hand, makes zero profit in period 2 with all contracts. But she is now better off in period 1 with this new contract, because $P^1 < \theta P^1 + (1-\theta)P_-^1$. Thus the long term contract could not have been optimal.

The only reason why the long term contract should not maximize the agent's utility in period 2 is to give the agent a better incentive to work hard in period 1. What this proof shows is that one can give this incentive already in period 1. By the way, the same argument holds if we had chosen the accident state instead of the no-accident state.

So overall we have shown that the additional constraints due to the assumption of short term contracts are not binding. The economic rationale for this result is the following: Remember the sufficient statistic theorem we discussed in the previous chapter. Only quantities which are influenced by the agent's effort should be relevant for his payment. As nothing in period 2 depends on the effort of the agent he exercises in period 1, two separate contracts for the two periods are sufficient. Still, as in the example above, the optimal contract for period 1 is designed such that an optimal contract for period 2 is anticipated.

In the derivation of this result the requirement that the insurer can control the saving of the insured was crucial. If she cannot, new problems arise. The problem here is that if savings are not observable, and the insured has a mixed strategy for his saving behaviour, then in the second period we have a combination of a moral hazard and an adverse selection problem. Adverse selection in so far, as the insurer does not know the wealth of the agent and with that the risk aversion of the agent. Unfortunately, as shown by Chiappori et al. (1994), the optimal contract which implements any other than the minimum effort level, involves randomized saving. The structure of the optimal contract under these circumstances has not yet been derived. There is however one exception where the savings of the insured do not create a problem. That is the case if the agent's utility function has constant absolute risk aversion, and his effort costs are monetary, i.e. $U(W,e) = -\exp[-r(W - c(e))]$. In that case in the incentive constraints in the second period (15.5), wealth $[(W - P^1)$ and $(W - L - P^1 + I^1)]$ cancels out, so that the amount of savings is irrelevant for the incentive structure. Then again, short term contracts are optimal.

To summarize the result: Many periods allow a shifting of the penalty across periods. Someone is worse off in later periods if he has an accident today, i.e. his consumption tomorrow differs whether he had an accident today or not. This is desirable due to the income smoothing effect. With an appropriate combination of savings and insurance contracts, the optimal outcome can be achieved via short term contracts. In light of these results, bonus-malus systems can be seen as an approximation to the optimal contract structure: Incentives are shifted across periods, but as savings are not controllable by the insurer, long term contracts are used. However, the optimal structure of these contracts is still unknown.

15.1.1 Infinitely many periods

As in the chapter on adverse selection, with infinitely many periods the first best can be obtained. There are several ways to see this. One is to use a contract similar to the one we used in the adverse selection case: Pay full insurance as long as the average risk probability is close enough to the one expected under the first best effort level. If not, penalize the agent. This is the procedure used in Rubinstein and Yaari (1983) and Radner (1981). As the argumentation is quite similar to the one outlined in chapter 13, we will not go into detail here.

Instead we discuss the solution along the lines of Fudenberg, Holmström and Milgrom (1990). They make the assumption that the agent is not allowed to borrow any money, but he can save. In addition it is assumed that the technology is the same in every period, as well as the wealth of the agent. I.e., without insurance the agent obtains W in every period if no accident occurs, and $W - L$ in case of an accident. The first best utility of the agent would be to consume $W - \pi_2 L$ in every period, which gives him a utility of: $U^* = U(W - \pi_2 L) - c_2$. Where, as before, e_2 is the larger effort level, which would be implemented in the first best.

Fudenberg et al. show the following: There exists a series of short term contracts such that for every $\epsilon > 0$, there exists a discount rate $\delta(\epsilon) < 1$, such that the agent can ensure himself an average utility level $U^* - \epsilon$ for all $\delta > \delta(\epsilon)$.

The interesting thing about this result is the specific series of short term contracts used: Namely no insurance at all. The agent is fully responsible for what he does. The crux of the proof is to find a consumption strategy so that the agent can ensure himself a utility level close to the first best. This is achieved if the agent consumes slightly less than $W - \pi_2 L$ if his savings are large enough, say larger than some $\bar{W}$. If savings fall below this critical level, he consumes $W - L$ in every period and restarts saving. With such a strategy, wealth follows a stochastic process (a submartingale) with a positive drift rate. From the theory of martingales it is known that eventually wealth will not fall below the critical level with a probability arbitrarily close to one. So if the discount rate is sufficiently small, the 'bad years' in the beginning, where the agent accumulates savings, do not count for the overall average utility.

So infinitely many periods are a nice thing to consider theoretically, but are not very relevant for the insurance market.

15.2 Renegotiation

Having discussed long term contracts we now turn to renegotiation which has given us a really hard time in the context of adverse selection. Fortunately, there is much less to worry if we have a moral hazard problem instead.

What we have seen in the last section is that under specific assumptions, long term contracts are just repetitions of short term contracts. But each short term contract is immune against renegotiation. Renegotiation is only relevant if based on the information the two parties have, there exists a contract which is better for both. But the short term contract was exactly designed in such a way that is was optimal.

And just by signing, no further information is revealed as it was the case in the adverse selection model. In moral hazard problems, the insured are all considered to be the same. As long as the effort is not choosen, no further information is relevant.

This latter point, however, can give rise to a renegotiation problem. Assume the following scenario: A ship owner wants to insure his ship against accidents and drowning. The ship still needs to be built, and a crew has to be found. You would like to give the owner an insurance contract with a deductible, so that the owner has an incentive to build the ship solidly, and find a good crew to sail with it. So you agree on a ten year contract, specifying a premium for each year, and a deductible. The owner builds his ship, recruits a crew and sets it off into the sea. Then he knocks at your door and has the following request: "I know that we arranged an insurance contract with a deductible. I understand the logic of this contract (he also attended this course once), so I build my ship properly and found a really good crew. But now, as the ship is on sea, there is nothing I can do about the quality anymore. So why don't we change the contract to a full insurance contract. I do not need any incentives anymore." Your are sceptical: that he claims to have build his ship properly and that he found a good crew is cheap talk - the owner would have said so in any case. But still, the argument for a full insurance contract is convincing: Now that the ship is on sea, there is nothing which can be done to improve the ship and it's crew. However, you realize that the owner probably anticipated that you will find his logic compelling. So if he anticipated a full insurance contract, he probably did not put any effort in building a solid ship and finding a proper crew. This time you loose. If you signed an incentive contract in the beginning under the assumption that it will not be renegotiated, you indeed, as a rational person, will modify the terms of the contract now that the incentive problem ceases to exist. But as a rational person, would you have signed the contract in the first place?

This renegotiation problem is discussed by several authors. The basis model used is always the same: The insurer offers a contract to the insured, which the agent can sign. At stage 2 the insured exercises his effort, and then the contract is renegotiated. Finally, at stage 3, the outcome occurs and payments are made.

Fudenberg and Tirole (1990) assume that when it comes to renegotiating, the insurer makes a new take-it-or-leave-it offer to the insured. Their main result is that under these assumptions the agent will always use a mixed strategy over his effort levels, if any other than the lowest effort level is chosen. Why is that? Assume that the agent exercises a specific effort level with probability one. If that is the strategy of the agent, at the renegotiation stage the insurer knows the effort of the agent for sure. There is no further incentive problem at the renegotiation stage, so indeed the optimal thing to do is to give the agent full insurance. However, this is anticipated by the agent, so he would not have worked this effort level before, at least if that was not the lowest level anyway. This is exactly the story given above. Anticipating full insurance the agent does not work hard. Therefore no equilibrium exists where the agent works hard with probability one. Assume instead that the agent chooses a mixed strategy. Then at the renegotiation stage, the principal does not know which effort precisely the agent has exercised. This is therefore a situation under adverse selection, so the principal will offer contracts as discussed in the chapter on adverse selection. The equilibrium is such that the menu of contracts offered by the insurer is the best response to the mixed strategy of the agent, while the agent's mixed strategy is the best response to the anticipated contract menu. In this case, renegotiation really does not make life easier.

Ma (1994) changes a 'slight' detail in the setup of the game. He considers the case where the agent makes the renegotiation proposal instead of the principal. Therefore at this stage we have no longer a screening model, but a signalling one, where the agent can with his contract proposal signal his type. Remember that in the chapter on renegotiation with adverse selection it turned out that who makes the contract offer can make a lot of difference. This is similar here: Ma shows that in his setup the standard second best contract can be obtained (under specific belief refinements). Therefore renegotiation is not a problem in his model. The reason for this result is that the principal can again have all different beliefs outside the equilibrium. So the agent would not propose a full insurance contract, as that makes the principal belief that he has put in the lowest effort level. If the contract in the beginning is designed in the right way, exercising a larger effort level can then be optimal for the agent.

Hermalin and Katz (1991) finally change another detail of the setup. They allow the principal to observe the effort put in by the agent. Effort is still not verifiable in the sense that it cannot be written into the contract, but the principal can see how much effort was put in. Take the ship owner example from above: Although it might be difficult to write into a contract that the crew has to be excellent, go on together well, not consume too much beverages, etc., it might well be possible that you see and jugde for yourself how good the crew is. In that case, the first best can be achieved: Renegotiation is good news. This works the following way:

The insurer offers an incentive contract which implements the first best effort level. That is the contract is such that the insured receives the maximum utility if he exercises this first best effort level. This is in general *not* the second best contract. Then, when it comes to renegotiation, the principal makes a take-it-or-leave-it offer to the insured. For any effort level he observes, he will make a full insurance offer which leaves no additional rent to the insured and this offer will be accepted by the insured. Anticipating this behaviour, the agent will put in the optimal effort. Any other effort level gives the agent a lower expected utility level under the original contract. When it comes to renegotiation the agent would then also obtain lower utility, because renegotiation does not increase the utility of the agent.

The difference in this model to the models above is that information is no longer asymmetrically distributed. Both the agent and the insurer can observe the effort, although it can still not be written into a contract. If effort were verifiable, i.e. could be written into a contract, then the first best is easily achieved. If it is only observable, then with an appropriate mechanism a first best can also be obtained (see also Moore and Repullo (1988)). In this case, a simple renegotiation procedure where the insurer makes a take-it-or-leave-it offer is just such a mechanism.

This finishes the first part on moral hazard. The main result we obtained was that incentives of all possible forms have to be given to alleviate the problem of asymmetric information. This can be in the form of contracts with deductibles, self insurance, but also bonus-malus system can serve this purpose. Still, there remains a lot to do in the context of moral hazard. What is the form of optimal contracts? Why do we mainly observe insurance contracts with deductible and/or proportional insurance coverage, but not with more complicated contract structures? What are optimal long run contracts if saving is not observable? Similar as with the theory of adverse selection, also on the applied side more work could be done: Which sectors in the insurance industry suffer most under moral hazard? How severe are these effects?

We now turn to limited liability, which, even if the agent is risk neutral, might give rise to moral hazard behaviour.

Chapter 16

Third party damage: liability and negligience rules

Limited liability refers to the notion that one party (the injurer), which causes damage to a second party (the victim), has not the financial means to compensate the victim fully.[1]

Examples are abound: The producer of a children's toy might go bankrupt if he is sued for product liability as a consequence of a child's death. Firms are liable for the safety of their worker. In case of a severe accident the costs might exceed the financial resources of the firm. Car drivers might not have enough wealth to cover very expensive accidents. Doctors might be sued for medical malpractice for sums which they are not able to pay for.

In the above cases, both the injurer and the victim are well defined people or institutions. This is not always the case. As an example consider a nuclear power station, where the owner does not have the financial means to cover all potential damages.[2] Here damages are done to the environment, possibly to many countries and people, so that the notion of the victim is more difficult to define. Also the following example falls in this category with a well defined injurer, but not so well defined victim: A person who undertakes a risky activity might in case of a bad outcome become dependent on the support of the state, for which he is not liable. In that case the 'victim' are the tax-payer, who have to cover the costs of the social transfer this person obtains.

[1]In the literature the problem of limited liability is also referred to as wealth constraints or the judgement proof problem (see e.g. Shavell, ?).

[2]For example, the Tchernobyl desaster lead to losses larger than ???.

Insurance is offered for automobile accidents, comprehensive general liability, medical malpractice, product liability, and so on. Pension scheme, health insurance and unemployment insurance are institutions which among other issues have to deal with the problems caused by wealth constraints. Some of these insurances are compulsory, like e.g. automobile insurance (at least in most western countries), others are voluntarily. This is a topic we will discuss later on: When should insurance be obligatory? Further issues, which we consider, are: Will the injurer provide enough care to avoid the losses? Might prohibition of insurance be a sensible policy?

Before starting with the analysis, one remark on liability rules is in order. In the literature, much effort has been devoted to discuss the different incentives of the injurer under different liability rules. Most commonly discussed are strict liability, which implies that the injurer has to compensate the victim always, and negligence, which leads to compensation only if not sufficient care has been undertaken. We will comment on these rules in the final section of this chapter. For the remainder of this chapter, however, we concentrate on strict liability. The reason is that if negligence is possible, this implies that a court can observe and determine the level of care the injurer has undertaken. In that case the insurer could condition the contract on this information, which leads to a first-best result.[3]

In this chapter we consider risk neutral agents only. There are two reasons for this assumption: First, problems with limited liability also arise for firms, where risk-neutral behaviour is a good approximation to reality. Second, risk neutrality allows us to focus on the problems caused by limited liability. Adding risk aversion will surely modify the optimal risk sharing arrangements, but will not change substantially the results we obtain.

In the following section we define the problem. We then first discuss why, if the agent is risk neutral and does not face wealth constraints, unobservability of effort is not a problem. Then we turn to the case with limited liability, where we calculate the optimal insurance contract. In Section X.X. policy proposals will be discussed.

[3] In the literature it is sometimes assumed that although the court can determine the care of the injurer, the insurer will not use this information when designing the contract. This, however, would be suboptimal.

16.1 Setting up the problem

Consider an agent who contemplates undertaking a risky project. His personal gain is given by $G(\theta)$, where θ denotes the state of the world. A second party incurs a loss if the project is undertaken of size $L(\theta)$. θ lies in between $[\underline{\theta}, \bar{\theta}]$ and is ordered such that $G(\theta) - L(\theta)$ is a decreasing function. That is, the loss to society is larger the larger θ is. θ is distributed according to the distribution function $F(\theta, e)$ where e denotes the effort of the agent devoted to care.[4] Providing effort is costly to the agent. These costs are denoted by e.[5]

So, given that the injurer exercises effort e, his expected utility, if he is not liable for any damage he incurs on society, is given by:

$$U_I = \int_{\underline{\theta}}^{\bar{\theta}} G(\theta) dF(\theta, e) - e$$

The expected loss to the victim is equal to:

$$U_V = \int_{\underline{\theta}}^{\bar{\theta}} L(\theta) dF(\theta, e)$$

Let us start by calculating the first best effort level, which is given by the maximization of the following expression:

$$\int_{\underline{\theta}}^{\bar{\theta}} [G(\theta) - L(\theta)] dF(\theta, e) - e \tag{16.1}$$

The optimal effort is determined by:

$$\int_{\underline{\theta}}^{\bar{\theta}} [G(\theta) - L(\theta)] dF_e(\theta, e) = 1 \tag{16.2}$$

Denote this effort by e^{FB}.

If the agent faces no liability at all for the harm he is doing to a third party, he maximizes utility according to:

$$\int_{\underline{\theta}}^{\bar{\theta}} G(\theta) dF_e(\theta, e) = 1$$

[4] The model is general enough to encompass the previous models used in Chapter X. Note that for $\theta = \underline{\theta}$, the overall 'loss' $G(\theta) - L(\theta)$ is minimized, which can be interpreted as the no-loss situation. In a situation where accident occurs with probability π and the agent influences only the distribution of the loss (see Chapter X.X), take $F(\underline{\theta}, e) = 1 - \pi$. This implies that the probability of no accident is $1 - \pi$, independent of the effort chosen.

[5] As long as no further assumption is made on $F(\theta, e)$, modelling the costs by e (and not $c(e)$) is completely general, as one could always make the transformation: $\hat{e} = c(e)$, $\hat{F}(\theta, \hat{e}) = F(\theta, c^{-1}(\hat{e}))$.

It is not clear whether the optimal effort under no-liability is larger or smaller than the first best effort level. It can be shown (Exercise 16.1), however, that if the injurer only controls the probability of an accident, effort under no liability will be smaller than e^{FB}.

As a side remark, note that the situation without liability is similar to the one discussed in microeconomics textbooks under the heading of externalities. The typical example being a firm pollutes a river which poses negative externalities on the laundry downstream.

We will not turn to strict liability, where the injurer is fully liable to the harm he imposes on someone else. Let us first discuss this situation where the injurer faces no wealth constraints.

16.2 Strict liability and no wealth constraints

As we have seen in the previous chapters, if effort by the agent is unobservable *and* the agent is risk averse, moral hazard occurs, which leads to efficiency losses. Here we face a similar situation, the injurer exercises unobservable effort. However, in this case we assume that he is risk-neutral. This changes the results completely. As is shown below, the first-best will be obtained. However, the proof relies on the fact that the agent can be punished sufficiently hard, which is not feasible if the agent is wealth constrained.

It is very easy to see that a first-best can be obtained. If the agent is fully liable for any harm he causes to the second party, he maximizes the utility

$$\int_{\underline{\theta}}^{\bar{\theta}} [G(\theta) - L(\theta)] dF(\theta, e) - e \tag{16.3}$$

which leads to the same first-order condition as equation (16.2).

This seems obvious: If the agent faces full responsibility for his actions, he will devote the optimal care. Note that if the injurer is risk averse, this result will not longer hold. In that case we are back to the models of the previous chapters. In those models it is the trade-off between incentives and insurance which drives the moral hazard problem. If there is no need to insure anyone, the party who exercises the effort will, by being fully responsible for the outcome, choose the optimal effort. Incentives without insurance are not a problem.

There is an interesting interpretation of this result: Consider the case where $G(\theta) = 0$ for all θ and $-L(\theta) > 0$. That is the agent has no direct benefit from the project, while the second party gains from the project. This is a standard principal-agent situation where the agent is e.g. the manager of a firm, while the principal is the owner, who needs the manager to run the business. The result from above says that optimally the manager faces full responsibility for his actions. This is done by paying him $-L(\theta) - F$ in state θ, where F is some fixed sum. But such a payment scheme implies nothing else than that the owner is selling the business, i.e. the payment stream, for a sum F to the agent.

This is a more general result: If the agent is risk neutral, and effort is not observable, the optimal thing to do is to sell all payment streams, which are influenced by the behaviour of the agent to the agent. He must be made the residual claimant for all his tasks. Note that this result only holds if the agent does not face wealth constraints and if no one else has to provide effort. This latter point will be discussed in Exercise 16.2., where the effort of both parties influences the outcome.

16.3 Limited liability

Why does such a payment scheme as the one discussed above not work in general? If $\theta = \bar{\theta}$, $G(\theta) - L(\theta)$ can be so negative, that the agent has not enough wealth to cover these costs. If, for example, the wealth of the agent W is equal to some $G(\hat{\theta}) - L(\hat{\theta})$ with $\hat{\theta} < \bar{\theta}$ he knows that he will not pay for the losses if the outcome is really bad. Then, instead of maximizing equation (16.3), he will maximize:

$$\int_{\underline{\theta}}^{\hat{\theta}} [G(\theta) - L(\theta)]dF(\theta, e) + (1 - F(\hat{\theta}, e))[G(\hat{\theta}) - L(\hat{\theta})] - e \tag{16.4}$$

If θ becomes larger than $\hat{\theta}$, which occurs with probability $1 - F(\hat{\theta}, e)$, the injurer only pays $[G(\hat{\theta}) - L(\hat{\theta})]$ and not the full loss which he caused to the other party. There is nothing left which can be taken away from him to compensate the victim. As before, it is not clear whether this leads to less or more effort by the injurer. Exercise 16.1 shows, however, that if the injurer only controls the probability of a loss, limited wealth leads to less effort exercised.

We now turn to the insurance decision. As before, the injurer faces limited wealth, however he is strictly liable for what he does. Suppose first, that the insurance decision is compulsory. The question we now turn to is how does the optimal insurance contract look like? Probably it will not provide full insurance. Results from the previous chapters have shown that full insurance leads to insufficient care undertaken. To derive the optimal contract, let us formalize the maximization problem:

$$
\begin{array}{lcr}
\max_{e,I(\theta)} & \int_{\underline{\theta}}^{\bar{\theta}}[W+G(\theta)-L(\theta)+I(\theta)]dF(\theta,e)-e & \\
 & \text{s.t.} & \\
\text{P.C. :} & \int_{\underline{\theta}}^{\bar{\theta}} -I(\theta)dF(\theta,e)\geq 0 & (16.5)\\
\text{I.C.} & e=\arg\max_{\tilde{e}}\int_{\underline{\theta}}^{\bar{\theta}}(W+G(\theta)-L(\theta)+I(\theta))dF(\theta,e)-e & \\
\text{L.L. :} & W+G(\theta)-L(\theta)+I(\theta)\geq 0 &
\end{array}
$$

The utility of the injurer, who faces strict liability, is maximized under the constraints that the insurer makes no loss, effort is chosen optimally by the agent, and the limited liability constraint: wealth of the agent cannot be smaller than zero. Note that because insurance is compulsory, there is no participation constraint for the agent.[6]

If the wealth constraint becomes binding in the optimum, the solution to the problem takes the following form:[7]

$$
I(\theta)=\begin{cases} -P & \text{if } W+G(\theta)-L(\theta)\geq P \\ -W-G(\theta)+L(\theta) & \text{if } W+G(\theta)-L(\theta)<P \end{cases} \tag{16.6}
$$

If the state of the world is such that the agent is sufficiently wealthy, he will pay a premium P independent of the exact realization of the state θ. For those states where $G(\theta)-L(\theta)$ is small (negative), the agent obtains marginally full insurance, i.e. the insurer pays all losses which exceed the agent's wealth. P is set such that the insurer makes zero profits overall. As before, it is not clear whether in general the agent will work harder or less hard in case he buys compulsory insurance. However, as you are asked to show in Exercise 16.1., if the agent only controls the probability of the loss, he will work less hard under insurance.

The result can be easily understood if we recall the solution to the problem without limited liability. There the optimal contract took the form: $I(\theta)=0$, i.e. the injurer was fully responsible for his actions. If the outcome is such that the agent can pay for it, we would still like to make him the residual claimant for his actions. This is done by making the (negative) transfer independent of the outcome. This transfer works fine as long as $W+G(\theta)-L(\theta)\geq P$. In that case, the limited liability constraint does not bind, so the solution has the same structure as the one for the problem

[6] For the standard insurance problem, one can interpret $I(\underline{\theta})$ as $-P$, where P is the premium the agent pays to the insurer.

[7] For a derivation see Innes (1990).

without this constraint. However, in the region where the wealth constraint binds, i.e. $W+G(\theta)-L(\theta)\geq P$, this constraint forces us to set $I(\theta)=-W-G(\theta)+L(\theta)$.

This contract takes the form of an excess loss reinsurance contract. The insurer only pays out if income minus loss, i.e. $G-L$, exceed a threshold.[8]

If the flow of payments is such that the insurer pays the victim, as it is the case in automobile insurance, the optimal contract looks different. In that case, the utility of the agent is given by

$$\int_{\underline{\theta}}^{\bar{\theta}}[W+G(\theta)+\tilde{I}(\theta)]dF(\theta,e)-e \tag{16.7}$$

The injurer only obtains his private benefit $G(\theta)$ (which might well be constant) and the transfer form the insurer, while the injured is paid by the insurer directly. Following the results from above, the optimal insurance contract is easily derived:

$$\tilde{I}(\theta)=\begin{cases} -P-L(\theta) & \text{if } W+G(\theta)\geq P+L(\theta) \\ -W-G(\theta) & \text{if } W+G(\theta)<P+L(\theta) \end{cases} \tag{16.8}$$

For small losses, the insurer passes on the costs to the agent. If the losses are large, the agent is pushed to the lowest wealth possible.

This modelling of risk neutral agents with wealth constraints is very popular in the literature for several reasons. First the optimal contract structure is relatively simple, so it facilitates modelling more complicated problems. Second, even if the insurer has a monopoly (or equivalent the principal has all the bargaining power) the agent will receive a rent in general, which does not hold in the model presented in the previous chapters. For some applications (e.g. farm workers, procurement firms, etc.) such a feature may be desired.

So far we have assumed that insurance is compulsory. If it is not, the injurer will only buy insurance if this increases his utility. However, for a risk neutral injurer the expected gain is all which counts. If by buying insurance he has to pay for losses which accrue to the victim, which he otherwise would not have to pay for, his expected profit will decrease.

Thus a risk neutral injurer will never buy insurance voluntarily. This general result does not hold if the injurer is risk averse. In that case by buying insurance the agent might take the risk out of the risky activity even in states where he would be able to

[8] As an interesting application of this result consider again the shareholder-manager situation. In that case the contract is an option contract. For bad outcomes, the manager obtains zero (or some fixed sum), while for good outcomes, he receives the whole profit plus this fixed amount, which may well be negative.

pay for the victim. This insurance aspect might be sufficient to buy insurance: But note that also from his point of view the insurance premium is unfair - the insurer has to pay for losses the agent otherwise would not have considered. (See Exercise 16.3).

16.4 Policy proposals

In the analysis of this chapter is was assumed that both the injurer and the victim were risk neutral. In that case, risk allocation is no welfare issue. So the only welfare loss due to limited liability arises through inefficient effort level taken by the injurer. We have seen that in general it is not clear under which regime (no-liability, strict liability and no insurance, strict liability and insurance) the injurer will exercise care optimally. However, Exercise 16.1 shows, that there are good reasons to believe that under strict liability and no insurance the effort by the injurer is the largest, but still not as large as the first-best effort level. So there seems to be the case for forbidding insurance. However, with this argument we have ignored two other aspects, which a policy maker has to take into account: First, many injured are risk averse, and second, there might be a concern for distributional issues. The first point is clear. Concerning the second point: Under a regime of no insurance, much of the harm the injurer causes to the victim has to be covered by the victim. For risk neutral parties this is just a transfer, and not a welfare loss, however, it might well be considered unfair that the victim has to pay for her losses herself. Although this fairness argument seems quite acceptable - if someone crashes into my garden with his car, he should be liable for all costs - there are many situations where the victim does not obtain compensation: If someone builds a house close to yours, you are not compensated for the noise and dirt the builders inflict upon you.

In the following we will leave the distributional issue beside, and discuss the case for and against insurance from a mere efficiency point of view.

Forbidding insurance.
From the argument above there seems to be a case to forbid insurance, as no insurance might (but need not) lead to larger effort provided by the injurer to prevent losses. However, as we have also seen in this chapter, a risk neutral injurer will not buy insurance anyway - so we have to turn to risk averse injurers to see whether forbidding insurance makes any sense.

As a matter of fact, it does not: If insurance is voluntarily, and the injurer buys

insurance, then:
- the injurer must be better off than without insurance, otherwise he would not have bought it.
- the insurer makes non negative profits, otherwise he would not have sold the insurance, so he is made better off as well.
- also the victim is made better off if the injurer is insured: she will be paid by the insurance if a loss occurs, otherwise she would have obtained nothing if the loss exceeds the wealth of the injurer.

Note that this argument holds as long as insurance covers the loss of the victim, i.e. liability insurance.[9] If, on the other hand, the injurer could buy insurance for the risk which accrues to him only, (i.e. against variations in $G(\theta)$), then the victim does not benefit at all from the insurance. But she might well suffer, as the insurance could lead the injurer to take less care. So in this case forbidding insurance might make sense. Examples for this kind of insurance are:
- ???

Mandatory Insurance.
As we have seen above, it is not clear whether individuals will take more or less care if they acquire insurance. We have also seen above, that a risk neutral injurer will not buy insurance voluntarily. A risk averse insurer might do so, but he is less willing to buy insurance, if his wealth is low anyway and/or if the potential loss is very large. The expected size of the loss is also relevant as the insurance has to pay for it, so it enters the utility of the injurer via the insurance premium. If the expected loss is large, the insurance becomes very unfair from point of view of the insurer. But if the expected loss is very large, it is quite likely that from a welfare point of view the project should not be undertaken at all. In that case, making insurance mandatory might be a means to prevent activities which impose a large burden on third parties.

Mandatory insurance can also be welfare improving if the victim cannot insure himself against the damage the injurer imposes upon him. In that case, forcing the injurer to buy liability insurance might be the only way to reduce the risk to the injurer.

But apart from these efficiency arguments, the main reason for mandatory insurance is fairness: If damage accrues to me, I expect the person who does the damage to pay for it. (Experiments?)

[9] For a more detailed discussion see Polborn (1998).

Negligence vs. strict liability.

As mentioned in the introduction to the chapter, the legal regime is also one topic discussed together with the problem of limited liability. For a contract theorist, the main distinction between these forms is the difference in information which is available to the parties. If the level of care is unobservable, what we have assumed in this chapter, the regime can only condition on the outcome, that is $G(\theta)$ and $L(\theta)$ in the terminology from above. As for most people and situations the functions will be different, the only possible regimes are no liability and strict liability. It is to be expected that with strict liability the incentives of the injurer to provide care are closer to the optimal level than with no liability. However, it is possible to create examples were this does not hold.

The other possibility is that the level of care is observable. In that case a legal structure like negligence rule makes sense. However, if effort is observable and can be verified by a court, it is quite obvious that the first-best can be obtained: Make insurance mandatory. The insurance company will condition its payment on the care taken and only pay if optimal (first-best) effort has been undertaken. Then the injurer exercises optimal effort and the victim will be fully compensated. This is not an unrealistic scenario - automobile insurance for example conditions on the degree of negligence which was undertaken by the driver: Drunk drivers do not obtain anything from the insurer if they caused an accident.

As we have seen, limited liability rises several issues where economists have to be concerned with. From a theory point of view, the neat way to model moral hazard with risk neutral agents who face wealth constraint is worth noting. From a policy perspective, the results depend on the economic environment: It is not possible to derive a clear-cut predictions concerning insurance and legal structure. This has to be decided for every situation in turn. The only thing we could do is to make the economic forces at hand clear: induce the injurer to provide sufficient care and insure the victim, if she is risk averse, for the losses she has to suffer. In some situations, this might be optimally done with mandatory insurance, sometimes voluntarily insurance might fare better.

Chapter 17

Insurance fraud

still to be written

Part VI

Public Economics of Insurance Markets

Chapter 18

Regulation of insurance markets

Chapter 19

Social insurance

Bibliography

1. ADVERSE SELECTION

Akerlof, G.A. (1970): The market for lemons: Quality uncertainty in the market mechanism, *Quaterly Journal of Economics* 84, 488-500

Ania, A., T. Tröger, A. Wambach (1998): An evolutionary analysis of an insurance market under adverse selection, *University of Vienna working paper 9808*

Asheim, G.B. and T. Nilssen (1996): Non-discriminating renegotiation in a competitive insurance market, *European Economic Review* 10, 1717-1736

Asheim, G.B. and T. Nilssen (1997): Insurance monopoly and renegotiation, *Economic Theory* 9, 341-354

Beaudry, P. and M. Poitevin (1993): Signalling and renegotiation in contractual relationships, *Econometrica* 61, 745-782

Cho, I. and D. Kreps (1987): Signalling games and stable equilibria, *Quaterly Journal of Economics* 52, 179-222

Cooper, R. and B. Hayes (1987): Multi-period insurance contracts, *International Journal of Industrial Organization* 5, 211-231

Crocker, K.J. and A. Snow (1986): The efficiency effects of categorical discrimination in the insurance industry, *Journal of Political Economy* 321-344

Crocker, K.J. and A. Snow (1992): The social value of hidden information in adverse selection economies, *Journal of Public Economics* 48, 317-347

D'Arcy, S.P. and N. Doherty (1990): Adverse selection, private information and lowballing in insurance markets, *Journal of Business* 63, 145-164

Dasgupta, P. and E. Maskin (1986): The existence of equilibrium in discontinuous economic games, II: Applications, *Review of Economic Studies* 53, 27-41

Dionne, G. (1983): Adverse selection and repeated insurance contracts, *Geneva Papers on Risk and Insurance* 8, 316-332

Dionne, G. and N.A. Doherty (1992): Adverse selection in insurance markets: A selective survey, in: Contributions to insurance economics, George Dionne (ed.), Kluwer Academic Publishers, 97-140

Doherty, N.A. and P.D. Thistle (1996): Adverse selection with endogenous information in insurance markets, *Journal of Public Economics* 63, 83-102

Gravelle, H. and R. Rees (1992): Microeconomics, Longman Group UK, 2nd ed.

Hellwig, M.F. (1987): Some recent developments in the theory of competition in markets with adverse selection, *European Economic Review* 31, 319-325

Hellwig, M.F. (1988): A note on the specification of interfirm communication in insurance markets with adverse selection, *Journal of Economic Theory* 46, 154-163

Hoy, M. (1982): Categorizing risks in the insurance industry, *Quaterly Journal of Economics*, 321-336

Inderst, R. and A. Wambach (1999): Competitive insurance markets under adverse selection and capacity constraints, CEPR Discussion Paper No.2269

Jaynes, G.D. (1978): Equilibria in monopolistically competitive insurance markets, *Journal of Economic Theory* 19, 394-422

Kifmann, M. and A. Wambach (1999): Adverse selection in life insurance markets, mimeo

Kreps, D.M. (1990): A course in microeconomic theory, Harvester Wheatsheaf

Mas-Colell, A., Whinston, M. D., Green, J.R. (1995): Microeconomic Theory, Oxford University Press.

Miyazaki, H. (1977): The rat race and internal labour markets, *Bell Journal of Economics* 8, 394-418

Polborn, M. (1997): Endogenous discrimination in insurance markets, mimeo

Puelz, R. and A. Snow (1994): Evidence on adverse selection: Equilibrium signalling and cross-subsidization in the insurance market, *Journal of Political Economy* 102, 236-257

Riley, J.G. (1979): Informational equilibrium, *Econometrica* 47, 331-359

Riley, J.G. (1983): Adverse selection and statistical discrimination, further remarks, *Journal of Public Economics* 20, 131-137

Rothschild, M. and J. Stiglitz (1976): Equilibrium in competitive insurance markets: An essay on the economics of imperfect information, *Quaterly Journal of Economics*, 629-649

Stiglitz, J. (1977): Monopoly, nonlinear pricing, and imperfect information: The insurance market, *Review of Economic Studies* 44, 407-430

Strohmenger, R. and A. Wambach (1999): Adverse selection and categorical discrimination in health insurance markets: The effects of genetic tests, to appear in *Journal of Health Economics*

Wambach, A. (1999 a): Renegotiation before contract execution, CEPR Discussion Paper 2189

Wambach, A. (1999 b): Bargaining and renegotiation, Münchener Wirtschaftswissenschaftliche Beiträge 99-03

Wambach, A. (2000): Introducing heterogeneity in the Rothschild-Stiglitz model, to appear in *The Journal of Risk and Insurance*

Wilson, C. (1977): A model of insurance markets with incomplete information, *Journal of Economic Theory* 16, 167-207

2. MORAL HAZARD

Arnott, R. and J.E. Stiglitz (1988 A): The basic analytics of moral hazard, *Scandinavian Journal of Economics* 90, 383-413

Arnott, R. and J.E. Stiglitz (1988 B): Randomization with asymmetric information, *Rand Journal of Economics* 19, 344-362

Chiappori, P.A., I. Macho, P. Rey and B. Salanie (1994): Repeated Moral Hazrad: The Role of Memory, Commitment and the Access to Credit Markets, *European Economic Review* 38, 1527-1553

Ehrlich, I. and G. Becker (1972): Market insurance, self insurance and self protection, *Journal of Political Economy* 80, 623-648

Fudenberg, D., B. Holmström and P. Milgrom (1990): Short Term Contracts and Long Term Agency Relationships, *Journal of Economic Theory* 51, 1-31

Fudenberg, D. and J. Tirole (1990): Moral Hazard and Renegotiation in Agency Contracts, *Econometrica* 58, 1279-1319

Grossman, S. and O. Hart (1983): An Analysis of the Principal-Agent Problem, *Econometrica* 51, 7-45

Hermalin, B. and M. Katz (1991): Moral Hazard and Verifiability, *Econometrica* 59, 1735-1754

Homström, B (1979): Moral Hazard and Observability, *Bell Journal of Economics* 10, 74-91

Homström, B (1982): Moral Hazard in Teams, *Bell Journal of Economics* 13, 324-340

Jewitt, I. (1988): Justifying the First-Order Approach to Principal-Agent Problems, *Econometrica* 57, 1177-1190

Langsberg, S.E. (1993): The Armchair Economist, Economics and Everyday Life, *The Free Press*

Ma, C.T.A. (1994): Renegotiation and optimality in agency contracts, *Review of Economic Studies* 61, 109-129

Malcomson, J. and F. Spinnewyn (1988): The Multiperiod Principal-Agent Problem, *Review of Economic Studies* 55, 391-408

Mirrlees

Moore, J. and R. Repullo (1988): Subgame Perfect Implementation, *Econometrica* 56, 1191-1220

Pauly, M.V. (1974): Overinsurance and public provision of insurance: The roles of moral hazard and adverse selection, *Quarterly Journal of Economics*, 44-62

Polborn, M. (1998): 'Mandatory insurance and the judgment proof problem, *International Review of Law and Economics* 18, 141-146

Radner, (1981):

Rogerson, W. (1985): The First-Order Approach to Principal-Agent Problems, *Econometrica* 53, 1357-1368

Rogerson, W. (1985): Repeated Moral Hazard, *Econometrica* 53, 69-76

Rubinstein, A. and M.E. Yaari (1983): Repeated Insurance Contracts and Moral Hazard, *Journal of Economic Theory* 30, 74-97

Shavell, S. (1979): On moral hazard and insurance, *Quaterly Journal of Economics* 11, 541-562

Shavell, S. (...) Judgement proof problem, ...

3. Others

Besley, T. (1989): Publicly provided disaster insurance for health and the control of moral hazard, *Journal of Public Economics* 39, 141-156

Bond, E.W. and K.J. Crocker (1997): Hardball and the soft touch: The economics of optimal insurance contracts with costly state verification and endogenous monitoring costs, *Journal of Public Economics* 63, 239-264

Cass, D., G. Chichilnisky and H.-M. Wu (1996): Indivudual risk and mutual insurance, *Econometrica* 64, 333-341

Coate, S. (1995): Altruism, the Samaritan's dilemma and government transfer policy, *American Economic Review* 85, 46-57

Cook, P.J. and D.A. Graham (1977): The Demand for insurance and protection: The case of irreplacable commodities, *Quaterly Journal of Economics* 91, 143-156

Coulson, N.E., J.V. Terza, C.A. Neslusan and B.C. Stuart (1995): Estimating the moral-hazard effect of supplemental medical insurance in the demand for prescription drugs by the elderly, *American Economic Review* 85, 122-126

Cummins, J.D. and M.A. Weiss (1993): Future developments in financial services: Insurance, *Journal of Banking and Finance* 566-568

Doherty, N.A. and H. Schlesinger (1983): Optimal insurance in incomplete markets, *Journal of Political Economy* 91, 1045-1054

Fields, J.A. and D. Tirtiroglu (1991): Agency-theoretic implications for the insurance industry: A review of theoretical and empirical research, *Quaterly Journal of Business and Economics*, 40-61

Geehan, R. (1977): Returns to scale in the life insurance industry, it Bell Journal of Economics, 497-514

Grossman, R.S. (1992): Deposit insurance, regulation and moral hazard in the thrift industry: Evidence from the 1930's, *American Economic Review* 82, 800-821

Haslinger F. and P.-J. Horgby (1997): High-cost protection in health insurance, *Finanz Archiv* 54, 1-25

Hill, R.D. (1979): Profit regulation in property-liability insurance, *Bell Journal of Economics*, 172-191

Keeley (1990): Deposit insurance, risk, and market power in banking, *American Economic Review* 80, 1183-1200

Lewis, F.D. (1989): Dependents and the demand for life insurance, *American Economic Review* 79, 452-467

Machnes, Y. (1995): Insurance and decision-taking, *Economics Letters* 48, 139-144

Schlesinger, H. (1984): Optimal insurance for irreplacable commodities, *Journal of Risk and Insurance* 51, 131-137

Selden, T.M. (1993): Should the government provide catastrophic insurance?, *Journal of Public Economics* 51, 241-247

Shioshansi, F.P. (1982): Insurance for irreplacable commodities, *Journal of Risk and Insurance* 49, 309-320

Sinn, H.-W. (1978): The Efficiency of insurance markets, *European Economic Review* 11, 321-341

Sloan, F.A. (1990): Experience rating: Does it make sense for medical malpractice insurance, *American Economic Review* 80, 128-133

Smith, E. and R. Wright (1992): Why is automobile insurance in Philadelphia so damn expensive?, *American Economic Review* 82, 756-772

Spence M. (1978): Product differntiation and performance in insurance markets, *Journal of Public Economics* 10, 427-447

Sung, J. (1997): Corporate insurance and managerial incentives, *Journal of Economic Theory* 74, 297-332

Toivanen, O. (1997): Economies of scale and scope in the Finnish non-life insurance industry, *Journal of Banking and Finance* 21, 759-779

Viscusi, W.K. (1979): Insurance and individual incentives in adaptive contexts, *Econometrica* 47, 1195-1207