ESSAYS IN THE ECONOMICS OF INSURANCE MARKETS

GEORGE W. BLAZENKO

B.A., Simon Fraser University, 1978

M.A., University of Western Ontario, 1979

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF

THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY
in

THE FACULTY OF GRADUATE STUDIES
(Faculty of Commerce and Business Administration)

We accept this thesis as conforming
to the required standard

THE UNIVERSITY OF BRITISH COLUMBIA
February, 1985
© George W. Blazenko, 1985

In presenting this thesis in partial fulfilment of the requirements for an advanced degree at the University of British Columbia, I agree that the Library shall make it freely available for reference and study. I further agree that permission for extensive copying of this thesis for scholarly purposes may be granted by the head of my department or by his or her representatives. It is understood that copying or publication of this thesis for financial gain shall not be allowed without my written permission.

Department of Commerce and Business Administration

The University of British Columbia
1956 Main Mall
Vancouver, Canada
V6T 1Y3

Date April 19, 1985

ABSTRACT

This dissertation investigates several aspects of the economics of insurance markets.

First, conditions under which policies with deductible provisions, coinsurance provisions, or premium rebates are optimal are given. Results of previous papers on indemnity costs are considered as special cases. For both personal and commercial lines of insurance, further applications consider income taxes, interest income and acquisition costs.

Second, the effect of individuals' characteristics on the trade-off between risk-sharing and incentives in a competitive insurance market affected by moral hazard is studied. An increase in the utility cost of effort decreases both prevention and coverage, while an increase in productivity of effort decreases loss frequencies and increases coverage. Decreases of utility in the loss state increase both prevention and coverage. Additional results establish the effect of wealth and risk aversion changes.

Third, features of insurance markets that affect the use of reinsurance are examined. An active reinsurance market exists when the direct market is imperfectly competitive. The manager of an insurance firm with monopoly power takes reinsurance in preference to holding this on own account. Market power in the reinsurance market also restricts reinsurance. The manager of a monopsonistic insurer takes reinsurance when risk aversion is greater than that of clients; this is consistent with the interests of owners. The use of reinsurance is then decreasing with the ratio of policy-holder to manager risk aversion coefficients. Costs incurred by either insurers or reinsurers in the reinsurance market reduce the use of reinsurance, while costs incurred by insurers in the original transaction leave coverage provided by insurers themselves unchanged.

Table of Contents

List of Figures

ACKNOWLEDGEMENTS

I am indebted to the members of my dissertation committee for their assistance. I would also like to thank Professor W.T. Ziemba and Professor P.P. Boyle for their encouragement and helpful suggestions.

Financial support from the Social Sciences and Humanities Council of Canada is gratefully acknowledged, as well as some support received from National Research Council Grant 67-7147.

CHAPTER ONE

INTRODUCTION AND SUMMARY OF RESULTS

If markets for assets were perfect and access to information costless, individuals could diversify risk, at no cost, by offering for general sale, claims to possible outcomes. Some organized exchanges, such as security markets, closely approximate this ideal, and have developed to facilitate such transactions. However, for other types of risks, transaction and information costs are significant relative to the benefits of risk sharing. In some cases these costs are sufficiently great that no exchanges are possible. However, in many other cases, organizations emerge who assume risks that individuals, because of costs, cannot trade amongst themselves; these are insurance companies. Economies of scale in the sale of contracts and the benefits of specialization, in the tasks of risk evaluation and claim settlement, serve to reduce average transaction costs. Insurance markets provide a means to diversify certain types of risks not tradeable in other markets.

Transactions and information costs cause insurance markets to develop separately from other financial markets. Their characteristics are also greatly affected by these costs. The purpose of this dissertation is to investigate the effects of several forms of transaction and information costs on insurance markets.

Chapter two considers the effect of the economic environment on the type of policies that are exchanged. The main result is a proposition that outlines in a general manner, the cases for which deductible, coinsurance, or policies with premium returns are optimal. Transactions costs, incurred by either the insurer or policy-holder, that depend upon either the premium or indemnity, result in deductible policies with coinsurance above a given level of loss. Results of Raviv (1979), on insurer settlement costs, and Townsend (1979), on policy-holder claim administration costs, are special cases of the general proposition presented. Because of the tax environment, insurance policies with premium returns in the form of dividends, rebates or retrospective rate returns may be exchanged. In this case, like in life insurance, the non-life insurance contract has a pure insurance element and a savings element. This can develop because of the asymmetric taxation of payments associated with insurance transactions between policy-holders and insurers in personal lines of insurance. For insurers, premiums are income and claim payments and premium returns are deductible expenses, while for policy-holders, premiums are not deductible expenses and claim payments and policy returns are not income. Policies with premium returns may also be exchanged, in either personal or commercial lines of insurance, because insurers' reserves for retrospective rate credits are tax deductible before they are actually paid. In either of these cases, policy-holders earn an after tax rate of return on premium returns greater than they receive from other sources of risk free investment.

In chapter three, insurers' informational costs of observing individuals' efforts to reduce the probability of loss are taken to be infinite. As is well known, the result of this is that even in the absence of other costs, an individual's risk cannot be entirely diversified through the insurance market. Individuals must hold some of the risk as an incentive to avoid the loss. This chapter focuses on how the characteristics of individuals affect the trade-off between risk-sharing and the required incentive effects of partial coverage. Economic agents buy insurance and undertake self-protection (activities which reduce the probability of loss) for the same reason - to mitigate the adverse effect of a loss of wealth accompanying damage to or destruction of an insurable asset. Since they have the same purpose, it might be suspected that insurance and prevention are substitutes. However, an increase in utility cost of prevention decreases both prevention and coverage. In this case incentives dominate substitution effects. Similar logic underlies the result that if prevention and productivity are homogeneous of degree zero in their effect on the probability of loss, then an increase in productivity (this may be the result of regulatory policy) decreases loss frequencies and increases coverage. If utility is state dependent, a decrease in utility in the loss state which does not affect marginal utility of income, increases both prevention and coverage. In this case, the increased incentive for self-protection is strong relative to the increased demand for coverage. As a consequence, increased insurance coverage does not reduce self-protection. Additional results establish the effect of wealth or risk aversion changes. In the special case of the HARA family of utility functions, if risk aversion is decreasing, increasing wealth reduces coverage sold, while if risk aversion is increasing, increasing wealth decreases prevention. If risk cautiousness is in a neighborhood of zero, (risk aversion may be increasing or decreasing) increases in wealth decrease both prevention and coverage. On the other hand, if risk tolerance is shifted upwards for a utility function with decreasing risk aversion, coverage sold decreases, while if risk aversion is increasing, prevention is increased. For risk tolerance in the neighborhood of zero, upward shifts have unambiguous effects on both coverage and prevention.

Chapter four is concerned with the effectiveness of insurance markets in the diversification of risks. Some characteristics of insurance markets enhance while others hinder this diversification. If insurers are price-takers, a perfectly competitive reinsurance market leads to complete diversification of insurable risk, even though this is tradeable only amongst insurers and reinsurers. This result is in contrast to that of Doherty and Tinic (1981) who argue that in capital market equilibrium reinsurance is redundant. They assume that the shareholders of insurance companies have well diversified portfolios, in which case, they are essentially risk neutral with respect to insurable risks. This means that such risks are completely diversified once an individual transacts with an insurer, and reinsurance is not required. This chapter allows the possibility that insurable risks may not be completely diversifiable in the immediate transaction with an insurer. This is the case if the direct market is less than perfectly competitive. As a result, transactions costs and the extent of market power on the part of insurers or reinsurers hinder risk spreading. Transactions costs incurred by insurers or reinsurers in the reinsurance exchange reduce the use of reinsurance. Transactions costs incurred by insurers in the original exchange reduce the use of reinsurance, but leave unchanged the coverage provided by insurers themselves. A monopolistic insurer always has a demand for reinsurance, however, this is decreased relative to a competitive market. To maximize expected profits, a monopolistic reinsurer restricts the use of reinsurance. The manager of a monopsonistic insurer takes reinsurance only if risk aversion is greater than that of clients. The use of reinsurance is then decreasing with the ratio of policy-holder to manager risk aversion coefficients. Moreover, this increases the value of the insurance firm.

CHAPTER TWO

OPTIMAL INDEMNITY CONTRACTS

2.1 Preliminaries

The economic environment in which risks occur can be expected to have significant effects on characteristics of insurance contracts. Many of the factors that affect the contract such as investment opportunities, transactions costs, government regulations, and taxes, are largely outside the control of participants. Moreover, these factors affect the ability of insurers to market and profitably sell policies. From a competitive standpoint policies must be designed to reflect the economic environment and adjusted for its changes. The purpose of this chapter is to examine the design of insurance contracts under differing economic situations. The main result is a proposition that outlines, in a general manner, the cases for which policies with deductible provisions, coinsurance provisions or policies with rebates or dividends are Pareto optimal. Previously obtained results on indemnity costs are considered as special cases. Further special cases treat interest income, premium taxes, income taxes, brokerage fees and both acquisition and commission expenses. The advantage of a general approach is that the determination of Pareto optimal policy design is greatly simplified. This is done with a straight-forward comparison of marginal effects of indemnity and premium on policy-holder and insurer final wealths. The need for separate proofs of the Pareto optimality of contract design for each and every economic environment investigated is obviated. At the same time, a general approach highlights the situations which do not induce either the rebating of premiums or policies with deductibles. For example, neither tax deductibility of loss, without corresponding taxation of indemnity, nor transactions costs that depend upon the extent of loss lead to either rebating of premiums or policies with deductible provisions. The reason is that neither of these factors affect the relationship of premiums to indemnity between insured's and insurer's final wealths. In addition, applying the general proposition yields insights and testable propositions on the operations of insurance markets.

Most analyses of the demand of insurance assume particular contractual forms, usually deductible or coinsurance types. For example, Gould (1969), Pashigian, Schkade and Menefee (1966), Mossin (1968), and Schlesinger (1981) have considered the choice of a deductible level. On the other hand, Smith (1968), Mossin (1968) and Mayers and Smith (1983) have treated the demand for coinsurance. Another area of the insurance literature considers the endogenous determination of the contract form. This research can be subdivided into articles that consider the "moral hazard" problem and those that do not. Examples of the former are Spence and Zeckhauser (1970), Harris and Raviv (1978), and Holmström (1979). They conclude that insurance contracts contain both risk sharing and incentive features. The incentive features serve to mitigate the adverse effects that unmonitored actions of insureds have on insurers. Mayers and Smith (1981) focus on monitoring and control mechanisms that arise in insurance contracts to lessen conflicts of interest between contract parties. Articles that examine endogenous determination of contract form but not the moral hazard problem are Borch (1960), Arrow (1963), Raviv (1979), and Townsend (1979). Borch was the first to show that if insurer and insured are both risk averse the optimal form of contract is of the coinsurance type. Arrow showed that if the premium was a function of the actuarial value of non-negative indemnity the optimal insurance policy provides full coverage above a deductible. Raviv and Townsend considered "dead weight" costs relative to both insured and insurer. Raviv examined insurer administrative and settlement costs, while Townsend considered insureds' costs of presenting and verifying claims. In both cases, a policy with a deductible is optimal. Townsend, however, did not consider the conditions under which the optimal deductible is non-zero. Huberman, Mayers and Smith (1983) have examined optimal insurance policies with and without moral hazard. For purposes here, the most interesting result is that if insurers are risk neutral, expenses are concave with respect to indemnity and the actuarial value of loss and expenses is charged as a premium, then a policy with a disappearing deductible is optimal. Such a policy has the property that loss minus indemnity decreases to zero as a function of the loss. With the loss as an upper bound on indemnity this result indicates that eventually full coverage is provided, thus the name disappearing deductible.

The application section of the present chapter focuses on three aspects of the economic environment and their effect on Pareto optimal policies. These aspects are riskless investment, taxation and transactions costs. Since interest income is of fundamental importance in insurance, it plays a prominent role in this chapter. Insurers typically receive premium income before claims are paid, interest income is earned in the interim. This ability to earn interest means that premiums are lower the higher is the interest rate. Reduced premiums reflect an implicit interest payment by insurers to individuals for the use of their cash between receipt of premiums and payment of claims[1]. In addition, differing income tax treatments of personal and commercial lines of insurance has important implications for optimal insurance contracts in the presence of interest income. In personal lines of insurance, if transactions costs are sufficiently small, a policy with a dividend is optimal. In commercial lines, full coinsurance policies are generally optimal, however, if the deductibility of reserves for unpaid claims and for retrospective premium credits is considered, a rebate policy is also optimal. In either of

[1] In the framework of the Capital Asset Pricing Model, Fairley (1979), and Hill (1979) have analyzed property-liability insurance rates with the inclusion of investment income. Fairley also examines equilibrium profit rates for insurers in the presence of taxes. Mayers and Smith (1982) have considered tax effects on the corporate demand for insurance.

these cases policies with provisions for return of part of the premium are optimal because individuals can earn an after tax rate of return greater than achievable from other sources of risk free investment. This increased rate of return is possible because of the nature of the tax environment. Like some forms of life insurance, policies contain both a pure insurance element and a savings element. The savings element is induced by favorable tax treatment. Transaction costs incurred by either the policy-holder or insurer that depend upon either the total premium or indemnity lead to policies with deductible provisions. Such policies are a compromise between risk sharing and economizing on transaction costs. Abandoning risk sharing at low loss levels where it is relatively unimportant saves on costs and allows risk sharing to be maintained to a greater extent at high loss levels where it is more significant. The study of transactions costs is particularly appropriate for insurance policies since they are generally more prominent than in other financial contracts. Moreover, costs incurred at the inception of the policy are on average more significant than claims adjustment expenses.

The remainder of this chapter proceeds as follows: the next section develops the setting for the problem, the notation to be used and the main result; Section 3 applies this result to important cases; Section 4 contains a brief summary.

2.2 Optimal Insurance Policies

The insured faces a random loss[2] x, $0 \leq x \leq T$, with density $\phi(x)>0$. Insurance indemnity is given by the schedule[3]

[2] Unlike a number of recent papers in the economics of insurance (i.e., Mayers and Smith (1983), Doherty and Schlesinger (1984), and Turnbull (1983)), it is assumed that the insurable risk is the individual's only source of random wealth. Alternatively, results in this chapter remain unchanged if random non-insurable wealth is additive in wealth and independent of the insurable loss.

[3] In both Raviv (1979) and Huberman, Mayers, and Smith (1983) it is assumed that indemnity is less than the loss (Raviv also assumes convex indemnity costs which means

$$0 \leq I(x). \qquad (2.1)$$

The effects on the insured's and the insurer's final wealths with respect to indemnity I, are given by twice differentiable functions $g_1(I)$ and $g_2(I)$, with $g_1(0)=g_2(0)=0$, and g_1', $g_2'>0$. These wealth effects may not equal I, because of taxes or transaction costs. The premium exchanged for the indemnity is P. The effect on insured and insurer final wealths are given by differentiable functions $f_1(P)$ and $f_2(P)$, with $f_1(0)=f_2(0)=0$, f_1', $f_2'>0$. With interest income the effects on insured's and insurer's final wealths are not equal to the premium, P. The function f_1 contains an opportunity cost component, whereas the function f_2 recognizes insurers' ability to earn interest on premiums. Finally, the effect of the loss x on the insured's final wealth is given by the differentiable function h(x), with $h(0)=0$, $h'>0$. This wealth effect of loss may not equal x because of tax deductibility.

The insured's final wealth is,

$$r_t\omega - f_1(P) - h(x) + g_1[I(x)]$$

where ω is initial wealth and r_t is one plus the after tax rate of return on riskless investment[4]. The insurer's final wealth is,

$$r_\tau W + f_2(P) - g_2[I(x)]$$

where W is initial wealth and r_τ is one plus the after tax rate of return on riskless investment for corporations. Utility functions for insured and insurer are U and V respectively; U is strictly concave while V is concave. The latter assumption allows the

that the upper bound constraint will never be violated). There are two important cases when indemnity can exceed the loss. The first is when the insured incurs costs of claim settlement. In this case indemnity is a compensation for both loss and expenses and, therefore, may be greater than the loss. The second case is when policies with rebates are optimal, in this case $I(0)>0$.

[4] As will be seen later in the application section of this chapter, the definition of f_1 incorporates the fact that the premium P is not available for riskless investment.

possibility that the insurer is risk neutral. To find Pareto optimal contracts the insured's expected utility is maximized subject to the insurer reaching a required utility level.

$$\underset{P,I(\cdot)}{\text{maximize}} \int_0^T U\left\{r_t\omega - f_1(P) - h(x) + g_1[I(x)]\right\}\phi(x)dx \tag{2.2}$$

subject to (2.1) and

$$\int_0^T V\left\{r_T W + f_2(P) - g_2[I(x)]\right\}\phi(x)dx \geq K \geq V(r_T W). \tag{2.3}$$

Equations (2.1), (2.2), and (2.3) describe a problem in the calculus of variations with an isoperimetric constraint and a non-negativity constraint on the state variable (for more details, see for example Intrilligator 1971, pp 318-320). Holding the premium constant, a Pareto optimal indemnity schedule, I^*, can be characterized by the following first order conditions[5],

$$I^* = 0 \text{ if } U'\left\{r_t\omega - f_1(P) - h(x)\right\}g_1'[0] - \lambda V'\left\{r_T W + f_2(P)\right\}g_2'[0] \leq 0 \tag{2.4}$$

$$U'\left\{r_t\omega - f_1(P) - h(x) + g_1[I^*]\right\}g_1'[I^*] - \lambda V'\left\{r_T W + f_2(P) - g_2[I^*]\right\}g_2'[I^*] = 0 \text{ for } 0 < I^* \tag{2.5}$$

where λ is the Lagrangean multiplier on the isoperimetric constraint. Necessary conditions are sufficient for a maximum when the intermediate function

[5] Second order necessary conditions, the Legendre condition, the Weierstrass condition, and the Weierstrass-Erdmann corner conditions are trivially satisfied in this problem.

$$U\left\{r_t\omega - f_1(P) - h(x) + g_1[I]\right\} + \lambda V\left\{r_T W + f_2(P) - g_2[I]\right\} \tag{2.6}$$

is concave in the state variable I. This assumption is made for the remainder of this chapter.

If a policy with a deductible is optimal the relation between the premium and the deductible $\bar{x}>0$ is

$$U'\left\{r_t\omega - f_1(P) - h(\bar{x})\right\}g_1'[0] - \lambda V'\left\{r_T W + f_2(P)\right\}g_2'[0] = 0. \tag{2.7}$$

Differentiating yields,

$$\frac{\partial P}{\partial \bar{x}} = \frac{-R_u\left\{r_t\omega - f_1(P) - h(\bar{x})\right\}h'(\bar{x})}{R_u\left\{r_t\omega - f_1(P) - h(\bar{x})\right\}f_1' + R_v\left\{r_T W + f_2(P)\right\}f_2'} < 0,$$

where $R_u\{\cdot\}$, $R_v\{\cdot\}$ are the insured's and insurer's Arrow-Pratt risk aversion indices, respectively. In all examples considered, $\frac{h'(\bar{x})}{f_1'(P)} \leq 1$, which means $\frac{\partial P}{\partial \bar{x}} \geq -1$. For a dollar increase in the optimal deductible the premium will fall by less than a dollar. Equation (2.5) implicitly defines the optimal indemnity schedule I^* as a function of x and P, but since P is an implicit function of $\bar{x}$ through (2.7), I^* can be considered a function of x and $\bar{x}$.

If a policy with a rebate or a dividend is optimal, the relation between the premium and the rebate "a" is

$$U'\left\{r_t\omega - f_1(P) + g_1(a)\right\}g_1'(a) - \lambda V'\left\{r_\tau W + f_2(P) - g_2(a)\right\}g_2'(a) = 0 \tag{2.8}$$

Differentiating, and assuming the intermediate function is concave in I, the premium is increasing with respect to the rebate. For this policy, I^* may be considered a function of x and "a".

Proposition 2.1: A Pareto Optimal insurance contract has a deductible, provides full coinsurance or provides for the rebating of premiums respectively, depending upon whether $\frac{f_1'}{g_1'}$ is greater than, equal, or less than $\frac{f_2'}{g_2'}$ for all I,P.

Proof: See the appendix.

The first ratio in the proposition is the insured's marginal rate of wealth substitution of premiums for effective loss. In other words, the increase in indemnity, for an increase in premium, required by the insured if final wealth is to remain constant when loss sustained is x. The second ratio is the insurer's marginal rate of substitution of current for future wealth. This rate of substitution is the maximum increase in indemnity, for an increase in premium, which leaves the insurer's final wealth unchanged when the individual's loss is x. If $\frac{f_1'}{g_1'} < \frac{f_2'}{g_2'}$, insurance coverage should be expanded. Final wealths of both insurer and insured can be increased with an increase in both premium and indemnity. If this relation holds for all I and P it is optimal for insurer and insured to make the rebate, and thus the premium as large as possible. On the other hand, if $\frac{f_1'}{g_1'} > \frac{f_2'}{g_2'}$, for a particular loss x, any perturbation in the policy which leaves unchanged the final wealth of one party reduces that of the other. The reason for this loss is dead weight costs in the insurance transaction which impede the flow of funds between the contract parties. The proposition indicates that the best way to economize on these costs is a positive deductible. In many cases this is further accomplished with reduced coinsurance above the deductible. If $\frac{f_1'}{g_1'}=\frac{f_2'}{g_2'}$, a zero deductible, or in other words, a full coinsurance policy is optimal.

Policy-holder dividends are in essence premium rebates. Generally, however, unlike retrospective rate credits, dividends to policy-holders are unrelated to claiming behavior of the policy-holder. They typically depend upon the insurance company's overall experience, and the discretion of directors[6]. If a policy-holder expects or is lead to expect policy-holder dividends or rebates he/she faces the possibility of losing these should the insurer go bankrupt. The higher the premium the more effort the insured expends to choose an insurer who is certain to be able to meet these payments. This search entails costs associated with determining the financial position of the insurer, comparing the refund policy with competing financial instruments and extra negotiation and transactions costs. On the other hand, insurers need increased sales efforts to promote high dividend or retrospectively rated policies. These notions can be reflected in the model by assuming that $f_1(P)$ is increasing convex and that $f_2(P)$ is increasing concave. This possibility means that the expression $\frac{f_1'}{g_1'}$ can be less than the expression $\frac{f_2'}{g_2'}$ for low values of P, but greater than for large values of P. Equation (2.18) in the appendix illustrates that the optimal dividend is finite.

[6] Of course, there are other reasons why insurers pay dividends than those examined in this chapter. The dividend is determined ex post; if there are costs of financial distress for an insurance company, charging a premium which exceeds expected losses and rebating the difference is an effective bond against these. Of course policy-holders must be compensated for such a service. It is argued in the next section that they can expect to receive greater after tax return from rebated premiums than from other sources of risk free investment.

2.3 Applications

2.3.1 *Costly Claim Settlement, Raviv (1979)*

The insurer's costs of claims settlement are assumed to depend upon the extent of indemnity. In the notation of section 2.2,

$$r_t=1,\ r_\tau=1,\ h(x)=x,\ f_1(P)=P,$$

$$f_2(P)=P,\ g_1[I]=I,\ g_2[I]=I+c[I].$$

The last term, c[I], gives the insurer's cost upon settling a claim I. Assuming $c'\geq 0$, $\frac{f_1'}{g_1'}=1\geq\frac{f_2'}{g_2'}=\frac{1}{1+c'}$. This result establishes the main theorem in Raviv's paper, that the optimal deductible is non-zero if and only if $c'>0$.

Using the appropriate functions above and differentiating (2.5) yields the marginal coverage

$$\frac{\partial I^*(x,\bar{x})}{\partial x}=\frac{R_u(A)}{R_u(A)+R_v(B)(1+c')+c''/(1+c')},\quad \bar{x}\leq x\leq T \tag{2.9}$$

where $A=\omega-P(\bar{x})-x+I^*(x,\bar{x})$, $B=W+P(\bar{x})-I^*(x,\bar{x})-c\left[I^*(x,\bar{x})\right]$.

Since concavity of the intermediate function insures that the denominator of this expression is positive, marginal coverage is always positive. In the special case of a risk neutral insurer and a concave cost function, marginal coverage is greater than one. If loss minus indemnity decreases sufficiently fast (i.e., to zero), and if the loss is considered an upper bound on indemnity, this result illustrates a disappearing deductible. See Smith and Bickelhaupt (1981) for a discussion of this result. In a later paper this same result is obtained by Huberman, Mayers and Smith (1983) (HMS). A policy with a disappearing deductible has the disadvantage that it may induce claimants to increase the extent of loss once it has occurred. The control of this expost moral hazard problem is considered by HMS.

2.3.2 *Costly Claim Verification, Townsend (1979)*

Individuals incur costs to verify and administer claims made against insurers. Following Townsend, policy-holders' settlement costs are assumed to depend upon indemnity. Therefore,

$$r_t=1,\ r_T=1,\ h(x)=x,\ f_1(P)=P,$$

$$f_2(P)=P,\ g_1(I)=I-\psi(I),\ g_2(I)=I.$$

The term $\psi(I)$ gives the insured's costs of verification[7], $\psi(0)=0$, $\psi'\geq 0$. For an optimal schedule I^*, $\psi'(I^*)<1$, otherwise both insured and insurer may be made better off by reducing indemnity. Thus, $\frac{f_1'}{g_1'}=\frac{1}{1-\psi'}\geq\frac{f_2'}{g_2'}=1$. Hence an optimal policy has a deductible if and only if $\psi'>0$. This result extends the analysis in Townsend (1979) by establishing the conditions under which a non-zero deductible is optimal.

2.3.3 *Administrative, Acquisition, and Underwriting Costs, Premium Taxes, Commissions*

In Canada and the United States, premium taxes are levied by provincial and state governments as a percentage of premiums charged. The rate is generally between two and four percent. Agent's commissions for selling policies are also charged as a

[7] The assumption $\psi(0)=0$ means that there are no fixed verification costs. Townsend, also examines a problem with a risk neutral insurer, fixed verification but no variable costs. In this case, a deductible policy is optimal.

percentage of premiums. It is further assumed that administrative expenses incurred by insurers when writing a policy are increasing with respect to premiums. Then

$$r_t=1,\ r_\tau=1,\ h(x)=x,\ f_1(P)=P,$$

$$g_1[I]=I,\ g_2[I]=I,\ f_2'\leq 1.$$

Thus, $\frac{f_1'}{g_1'}=1\geq\frac{f_2'}{g_2'}=f_2'$, and the optimal policy has a deductible provision.

2.3.4 *Brokerage Costs*

Insurance brokers are often employed by individuals or firms to act on their behalf in the purchase of insurance. The fee for this service is invariably increasing with respect to premiums paid. Thus,

$$r_t=1,\ r_\tau=1,\ h(x)=x,\ f_2(P)=P,$$

$$g_1[I]=I,\ g_2[I]=I,\ \text{and}\ f_1'(P)\geq 1.$$

Thus, $\frac{f_1'}{g_1'}=f_1'\geq 1=\frac{f_2'}{g_2'}$, and the optimal policy has a deductible provision.

2.3.5 *Combinations of Applications 3.1, 3.2, 3.3, 3.4*

In the above applications, $f_1'\geq 1$, $f_2'\leq 1$, $g_1'\leq 1$, $g_2'\geq 1$. Applying the proposition, any combination of these possibilities leads to a policy with a deductible.

The deductible in both this and the above sub-sections results from transactions costs. To economize on these costs, risk sharing is reduced. The proposition indicates that the best way to economize is with a deductible provision. The deductible is a compromise between risk sharing and economizing on transactions costs. Abandoning risk sharing at low loss levels where it is relatively unimportant to the individual saves on costs and allows risk sharing to be maintained to a greater extent than would otherwise be possible at high loss levels.

2.3.6 *Income Taxes and Personal Lines of Insurance*

In personal lines of insurance, premiums paid by individuals, are not tax deductible. In Canada, losses are not deductible and indemnity is not taxed as income. In the United States, however, uninsured casualty losses of more than one hundred dollars are tax deductible[8]. In this section, both these situations are considered.

Corporations are generally allowed for tax purposes to carry forward or backward operating losses against taxable income. The effect of this opportunity is to average actual tax payments in any one year. This provision is particularly important in insurance because it means government shares in insurers' gains and losses. The government is an implicit partner in an insurer's risk enterprise. This partnership affects insurers' attitudes to risk. Insurers who have incurred many years of losses or new insurers are not eligible for immediate tax refunds. Instead losses must be carried forward; the effect depends upon future positive taxable income and opportunity costs associated with receiving a tax refund in future years. For simplicity it is assumed that insurers have previously paid taxes and are eligible for immediate tax refunds.

The effect of carry back provisions of the tax code is that insurers incur only $1-\tau$ of losses, where τ is the corporate tax rate. Thus the insurer's final wealth is

---

[8] The loss must also be greater than 10 per-cent of income. This provision of the U.S. tax code is ignored.

$$W+(1-\tau)\Big[rW+RP-I(x)\Big],$$

where R=1+r. The term in the square brackets is the insurer's taxable income. This income is composed of both premiums and interest. The insurer pays taxes if taxable income is positive and receives a refund on previously paid taxes if it is negative. This expression implicitly assumes that a policy-holder dividend or rebate (i.e. $I(0)>0$) is deductible by an insurer for calculation of taxable income. Under section 140 of the Canadian Income Tax Act, mutual and joint stock non-life insurers are allowed, in computing income, to deduct amounts paid or credited to policy-holders within the policy year or one year thereafter by way of dividends, refunds of premiums and refunds of premium deposits. A similar deduction is available in the United States through Section 832(c)(2.11) of the tax code. Tax authorities have viewed policy-holder dividends of non-life insurance policies as premium returns, as such they have not been taxed (even though investment earnings are involved). The reason for this is the relatively short term nature of non-life insurance policies which makes them ill suited as significant savings devices.

In Canada, the insured's final wealth after the occurrence of an insured casualty loss is

$$\Big[1+r(1-t)\Big](\omega-P)-x+I(x), \tag{2.10}$$

where t is the personal tax rate, and it is assumed that all interest income is taxable[9]. In the notation of section 2.2,

[9] In the United States all interest income is taxable, in Canada the first one thousand dollars of interest or investment income is exempt from tax.

$$r_t=1+(1-t)r,\ r_\tau=1+(1-\tau)r,\ h(x)=x,\ f_1(P)=\Big[1+r(1-t)\Big]P,$$

$$f_2(P)=(1-\tau)RP,\ g_1[I]=I,\ \text{and}\ g_2[I]=(1-\tau)I.$$

Thus, $\frac{f_1'}{g_1'}=1+r(1-t)\leq\frac{f_2'}{g_2'}=R$. If both the interest rate and the personal tax rates are positive a dividend policy is optimal.

To see how this result arises, first observe that the insured's opportunity cost of premiums paid is $r_t=1+r(1-t)$. For each dollar of dividend, the insurer in essence borrows an amount between $\frac{1}{r_t}$ and $\frac{1}{(1+r)}$, invests this at the rate $1+r$ and makes a profit of between $(1-\tau)\left[\frac{(1+r)}{r_t}-1\right]$ and 0. Both insurer and insured can, as a result, be made better off. Presumably, the insured cannot borrow at the after tax rate $1+r(1-t)$ and, therefore, (2.10) is appropriate only when $\omega>P$. As a result final wealths cannot be made infinitely large.

If the interest portion of the policy-holder dividend were to be taxed, its advantage dissipates. In this case the optimal policy is of the coinsurance type, with marginal coverage

$$\frac{\partial I^*}{\partial x}=\frac{R_u}{R_u+R_v\cdot(1-\tau)}.$$

Because the government shares in the insurer's gains and losses, the insurer has an implicit partner. This reduces the effective risk aversion of the insurer and results in increased insurance coverage. In insurance markets where the supply of coverage is influenced by insurers' risk aversion, equilibrium insurance rates decrease with increases in the tax rate[10]. On the other hand, if insurers are risk-neutral corporate taxes have

[10] This decrease results because of a reduction in the "risk premium" included in the pricing of insurance. However, the major impact of an increase in corporate tax rates is likely to be felt in the value of the insurance firm as a whole rather than on the insurer's

no effect on insurance premiums.

If underwriting costs, premium taxes and agents' commissions are introduced into this sub-section, then

$$r_t=1+(1-t)r,\ r_\tau=1+(1-\tau)r,\ h(x)=x,\ f_1(P)=[1+r(1-t)]P,$$

$$g_1[I]=I,\ f_2(P)=RP(1-\tau)(1-\mu),\ \text{and}\ g_2[I]=(1-\tau)I.$$

The expression for $f_2(P)$ assumes that underwriting costs are proportional to premiums and equal μP. In this case, $\frac{f_1'}{g_1'}=1+r(1-t)$ and $\frac{f_2'}{g_2'}=R(1-\mu)$. If insurers' acquisition costs are large per premium dollar, the advantage of premium rebates disappears. For example, if the interest rate is 10% and the personal tax rate is 30%, then when initial contract costs are greater than approximately 2.7% per premium dollar a deductible rather than a dividend policy is optimal. Unless transactions costs are small, dividends are not induced by the tax environment. However, for individual cases, particular lines of insurance, or with technological change that reduce the costs of underwriting, acquisition costs may be sufficiently low to induce the rebating of premiums. Rebating is most likely for renewal policies, policies that have multiple year terms, or for insurers who use direct marketing rather than agents. Moreover, the marginal cost of obtaining premium dollars for investment purposes may be substantially less than that for pure insurance purposes. Lower marginal costs enhance the possibility that insurers gain financially from rebating premiums. Of course, even if a deductible policy is optimal, differing tax treatments of individuals and insurers makes the optimal deductible less than would otherwise be the case. The above expressions for $\frac{f_1'}{g_1'}$ and $\frac{f_2'}{g_2'}$ indicate that

---

pricing decision. With an increase in corporate taxes investors are likely to be attracted to bonds and fixed income securities and away from equity securities.

even if transactions costs dominate, the level of the deductible is influenced by both the riskless interest rate and the personal tax rate. This result implies that there are tax clientele effects in the sale of personal insurance. In other words, persons with greater tax rates purchase policies with lower deductibles. The reason for this relationship is that with a higher tax rate the opportunity cost of giving up a dollar in insurance premium is lessened. A lesser opportunity cost is equivalent to a reduced premium charge for insurance. As a result, individuals with greater tax rates are attracted to policies with greater indemnity, these policies have a lower deductible.

Finally in this sub-section, a casualty loss in the United States is considered. The insured's final wealth is

$$\left[1+r(1-t)\right](\omega-P)-(1-t)\left[x-I(x)\right].$$

This expression reflects the fact that premiums are not tax deductible, but that the uninsured portion of the loss is. Thus, $\frac{f_1'}{g_1'}=\frac{\left[1+r(1-t)\right]}{(1-t)}\geq\frac{f_2'}{g_2'}=R$, and a policy with a deductible provision is optimal. The reason for this result is that the implicit taxation of indemnity reduces its desirability. Moreover, this reduced desirability occurs even though the loss itself is tax deductible. This result illustrates that the determining factor for deductible or rebate policies is the relationship of premiums to indemnity between an insured's and an insurer's final wealths. Frictions associated with the loss itself have no bearing on the Pareto optimality of insurance contracts.

2.3.7 *Income Taxes and Commercial Insurance*

Premiums paid for commercial insurance are deductible business expenses. The principle determining this deductibility is found in normal commercial accounting practice rather than in tax laws. Repairs for damaged property are deductible expenses, associated indemnity is taxable as income (see for example section 12(2.1)(f) in the Canadian income tax act). In the case of loss or destruction of property, indemnity may be viewed as a replacement of fixed assets, and is therefore, not taxable. The difference between the loss and indemnity is, however, a deductible expense[11]. A loss of trading assets, such as stock in trade or cash, through theft, holdup, robbery or embezzlement is allowed as a deduction. The amount is net of any insurance recovery. Tax treatment of liability premiums and net recovery is similar.

In all these cases, the insured's final wealth is,

$$\omega+(1-\tau)\Big[r(\omega-P)-P-x+I(x)\Big]$$

This expression encompasses the assumption that corporations can carry back losses for tax refunds. It also assumes that any premium refund is taxed as income to the insured. This taxation is simply a recapture of the premium deduction.

In the notation of section 2.2

$$r_t=1+(1-\tau)r,\ r_\tau=1+(1-\tau)r,\ h(x)=(1-\tau)x,\ f_1(P)=(1-\tau)RP,$$

$$g_1[I]=(1-\tau)I,\ f_2(P)=(1-\tau)RP,\ \text{and}\ g_2[I]=(1-\tau)I.$$

Since $\frac{f_1'}{g_1'}=R=\frac{f_2'}{g_2'}$, a full coinsurance contract is optimal. Notice that this result holds

[11] Compensation for loss or destruction of capital property is deemed to be proceeds of disposition of the property and thus may result in recapture of depreciation, (capital cost allowance in Canada) and/or a capital gain. These aspects of tax law are not considered in this chapter. In Canada, x-I(x) is the terminal loss associated with loss of the asset, in the United States, it is the allowable deduction for loss. See Mayers and Smith (1982,pp289) for further details.

even if the insurer and the firm purchasing insurance have differing tax rates.

Compensation for loss or destruction of inventory or for loss of profits is considered business income. This income is, therefore, taxed at normal rates. Correspondingly, losses are not tax deductible. Since the loss is not tax deductible, $h(x)=x$, but f_1, g_1, f_2, and g_2 remain as above, and therefore, a full coinsurance policy remains optimal.

The taxation of non-life insurers is essentially the same as the taxation of a standard corporation. One of the major exceptions, however, is the treatment of contingent liabilities. For standard corporations, with few exceptions, no allowance is made for contingencies in computing taxable income. Such liabilities are deducted from net income only when paid or accrued. Invariably at fiscal year end an insurer will be unaware of its exact liabilities. Many claims will be in the process of investigation, adjustment, or litigation, other potential claims are likely to be unreported. Insurers, in a departure from strict accrual accounting, deduct estimates of exposure on the occurrence of an insured event even though liability is not fixed or determinable. Moreover, these amounts need not be discounted to account for the fact that actual payments need not be made for considerable periods of time. This deductibility is especially important for accident, liability, and sickness insurance, where the delay between claim reporting and actual payment, can be lengthy. In this chapter, it is assumed not only that insurers give truthful estimates of future liabilities, (this may be assumed to be a result of penalties imposed by taxing authorities for reserves that subsequently turn out to be unreasonable), but are also able to predict with certainty what these liabilities will be. It is further assumed that although loss occurs at the end of the policy period, indemnity may not be paid until afterwards. Because of this delay, indemnity must be discounted to be comparable to other amounts either paid or received at the end of the policy term. This discounting should have a random component since the actual date of indemnity will not be known (see Fairley (1979) for the distribution of times to settlement for automobile property and bodily injury claims). For simplicity the random nature of discounting is ignored; the effect on insurer and insured final wealths of indemnity I to be paid at a date not necessarily the same as the end of the policy period is assumed to be hI, $0<h<1$. The parameter h is related to the concept of a "funds generating factor" used by Fairley (1979) and Cummins and Nye (1981). Cummins and Nye measure this as the sum of the unearned premium reserve plus the reserves for claim settlement divided by earned premiums. A value greater than one indicates that a dollar of premium income can, on average, be invested by the insurer for a greater length of time than the policy period. Estimates indicate that liability coverages have the greatest funds generating factors. A value of h less than one indicates a funds generating factor greater than one. The insured's final wealth is[12]

$$\omega+(1-\tau)\Big[r(\omega-P)-P-x\Big]+h(1-\tau)I(x). \tag{2.11}$$

The insurer's final wealth is

$$W+\Big[rW+RP-hI(x)\Big]-\tau\Big[rW+RP-I(x)\Big]. \tag{2.12}$$

The second term gives non-tax related aspects of the insurer's income. The expression in the square brackets of the last term gives the insurer's taxable income. This expression encompasses the assumptions that future liability can be perfectly predicted and can be deducted for current tax purposes. This expression also implies that if a

[12] If the loss is sustained only when indemnity is paid, for example as is the case in liability insurance, then the discounted loss is hx. Major results remain unchanged with this modification of the problem.

premium rebate is to be paid, this is done at the time the claim is settled, and that the amount is currently deductible for taxes. This deduction is not allowable if the rebate is paid in the form of a policy-holder dividend since these are deductible only when paid (see for example Lenrow et al. pp. 195-196). On the other hand, the deduction is allowable if the rebate is paid as a retrospective premium credit. The reason the deduction is possible is that reserves set aside for the payment of these premium returns are included in the insurer's unearned premium reserve which is deductible in the same way that the insurer's loss reserve is deductible (Lenrow et al. pp. 174-175). This deductibility means that I(x) in (2.12) may be interpreted as the sum of indemnity and experience dependent premium returns.

In the notation of section 2.2

$$r_t=1+(1-\tau)r,\ r_\tau=1+(1-\tau)r,\ h(x)=(1-\tau)x,\ f_1(P)=(1-\tau)RP,$$

$$g_1[I]=h(1-\tau)I,\ f_2(P)=(1-\tau)RP,\ \text{and}\ \ g_2[I]=(h-\tau)I.$$

The assumption $g'_2>0$ implies that $h>t$. This restriction means that the tax advantage of indemnity's deductibility never offsets the disadvantage of its discounted payment. The above expressions indicate that, $\frac{f'_1}{g'_1}=\frac{R}{h}\leq\frac{f'_2}{g'_2}=\frac{R(1-\tau)}{h-\tau}$, and therefore the optimal policy is of the rebate type.

To see why the rebate policy arises, let the implicit interest rate the insurer pays for a rebated premium dollar be r^*. The insurer invests this at $(1+r)$ and pays tax on the entire amount. A discounted equivalent rebate of h is paid, but the insurer receives a reduction of current taxes or a tax refund of τ. For a rebate of h, the insurer makes

$$\frac{(1-\tau)R}{(1+r^{*})}-(h-\tau). \tag{2.13}$$

The insured's opportunity cost for a dollar of premium is $(1-\tau)R$. The rebate received is equal to h, which after tax is equal to $(1-\tau)h$. The effect on the insured's final wealth is therefore

$$\frac{-(1-\tau)R}{(1+r^{*})}+(1-\tau)h. \tag{2.14}$$

If $\frac{R}{h}\leq(1+r^{*})\leq\frac{(1-\tau)R}{(h-\tau)}$, it is possible to make both equations (2.13) and (2.14) positive. This opportunity arises because although the insurer pays $h<1$ for a rebate, a full dollar is deductible from current net income for tax purposes. Both parties may be made better off by this arrangement as long as the insurer has current tax liabilities to be reduced or can carry back losses for tax refunds.

There have been a number of papers, for example, Anderson (1971), Balcarek (1966) and Anderson and Thompson (1971) that have examined the question of whether insurers overestimate loss reserves in order to reduce current tax liabilities. Taken together, however, these papers are inconclusive. Cummins and Nye (1981) argue that effective tax rates of property-liability insurers are significantly less than those quoted in tax codes because of insurers' ability to manipulate reserves. The importance of results just presented, is to point out that one way insurers can reduce taxes is to use retrospectively rated policies in lines of insurance with large funds generating factors. Moreover, this tax reduction can be achieved with no fear of penalty from taxing authorities. This result implies that retrospectively rated policies are most likely to be found in lines of insurance with large funds generating properties. In Cummins and Nye (1981) Workman's Compensation is estimated to have a large funds generating factor of 1.48. At the same time, the bulk of retrospectively rated policies are written in this line of insurance. The funds generating factors for Auto Liability and Auto Physical Damage are estimated to be 1.174 and 0.594 respectively. The fact that deductibles are prevalent in Auto Physical Damage rather than Auto Liability is consistent with the results of this chapter.

2.4 Summary

A generalized theory of single period Pareto optimal insurance contracts is presented. The main view is that many aspects of the economic environment in which an insurable risk occurs, will ultimately be impounded into the design of optimal insurance policies. Results of previous papers on indemnity costs are considered as special cases. Further applications consider income taxes, interest income and various acquisition costs. With transactions cost that depend upon either the total premium or indemnity, incurred by either the insured or insurer, a policy with a deductible provision is optimal. In personal lines of insurance, because of differing tax treatments of insurers and individuals, and if transactions costs are sufficiently low, a policy with dividends is optimal. However, a deductible policy is optimal for casualty losses in the United States. In commercial lines of insurance full coinsurance policies are generally optimal. If, however, the deductibility of particular reserves is considered, a policy with retrospective rate credits is optimal.

2.5 Appendix to Chapter Two

Proof of Proposition 2.1: A Pareto optimal indemnity schedule is obtained by maximizing the insured's expected utility subject to the constraint that the insurer achieve a minimum utility. Section 2 gives the conditions for this maximization with respect to the indemnity schedule for a fixed premium. This appendix completes the problem by considering the maximization with respect to the premium as well. Because the premium is related to the deductible through the implicit relation (2.7), this problem may alternatively be cast in the form of maximizing with respect to the deductible. The insured's expected utility with indemnity schedule $I^*(x,\bar{x})$ derived from equations (2.4), (2.5), and (2.7) is

$$U^*(\bar{x})=\int_0^{\bar{x}} U\left\{r_t\omega - h(x) - f_1(P)\right\}\phi(x)dx$$

$$+\int_{\bar{x}}^{T} U\left\{r_t\omega - h(x) - f_1(P) + g_1[I^*]\right\}\phi(x)dx.$$

The derivative with respect to $\bar{x}$ is

$$-P'f_1'\int_0^{\bar{x}} U'\left\{r_t\omega - h(x) - f_1(P)\right\}\phi(x)dx$$

$$+\int_{\bar{x}}^{T} U'\left\{r_t\omega - h(x) - f_1(P) + g_1[I^*]\right\}\left[-f_1'P' + g_1'\frac{\partial I^*}{\partial \bar{x}}\right]\phi(x)dx.$$

The insurer's expected utility is

$$V^*(\bar{x})=\int_0^{\bar{x}} V\left\{r_T W+f_2(P)\right\}\phi(x)dx+\int_{\bar{x}}^{T} V\left\{r_T W+f_2(P)-g_2[I^*]\right\}\phi(x)dx.$$

The derivative with respect to the deductible is

$$V'\left\{r_T W+f_2(P)\right\}f_2'P'\int_0^{\bar{x}}\phi(x)dx$$

$$+\int_{\bar{x}}^{T} V'\left\{r_T W+f_2(P)-g_2[I^*]\right\}\left[f_2'P'-g_2'\frac{\partial I^*}{\partial \bar{x}}\right]\phi(x)dx.$$

If the derivative of

$$U^*(\bar{x})+\lambda V^*(\bar{x}) \tag{2.15}$$

with respect to $\bar{x}$ evaluated at $\bar{x}=0$ is positive, then a policy with a deductible provision is optimal. Constraint (2.3) serves to determine λ.

The derivative of expression (2.15) with respect to $\bar{x}$ is

$$-P'\int_0^{\bar{x}}\left[U'\left\{r_t\omega-h(x)-f_1(P)\right\}f_1'-\lambda V'\left\{r_T W+f_2(P)\right\}f_2'\right]\phi(x)dx$$

$$-P'\int_{\bar{x}}^{T}\left[U'\left\{r_t\omega-h(x)-f_1(P)+g_1[I^*]\right\}f_1'-\lambda V'\left\{r_T W+f_2(P)-g_2[I^*]\right\}f_2'\right]\phi(x)dx. \tag{2.16}$$

If $\frac{f_1'}{g_1'}>\frac{f_2'}{g_2'}$, then

$$\frac{U'f_1'}{U'g_1'} > \frac{\lambda V'f_2'}{\lambda V'g_2'}.$$

Rearranging, considering the case $I^* > 0$, and using equation (2.5) implies,

$$\frac{U'f_1'}{\lambda V'f_2'} > \frac{U'g_1'}{\lambda V'g_2'} = 1. \tag{2.17}$$

Using the result (2.17), expression (2.16) evaluated at $\bar{x}=0$ is positive, and therefore a positive deductible is optimal. The value will depend upon the characteristics of the insurer and insured, the economic environment, (i.e. interest rate, taxes and transactions costs) and the loss distribution. If transactions cost are sufficiently great it may be the case that $\bar{x}=T$, in which case no insurance transaction is possible.

If $\frac{f_1'}{g_1'} = \frac{f_2'}{g_2'}$, expression (2.16) is zero when evaluated at $\bar{x}=0$, and negative (using equation 2.4) when evaluated at positive values of $\bar{x}$. A full coinsurance policy is, therefore, optimal.

If $\frac{f_1'}{g_1'} < \frac{f_2'}{g_2'}$, then (with the results (2.4) and (2.5))

$$\frac{U'f_1'}{\lambda V'f_2'} < \frac{U'g_1'}{\lambda V'g_2'} \leq 1.$$

The second term equals 1 when $I^* > 0$ and is less than 1 if $I^* = 0$. Hence expression (2.16) is negative for all $\bar{x} \geq 0$. Since negative deductibles are not possible rebate policies are considered.

For rebate policies the equivalent to expression (2.16) is

$$-P'\int_0^T \left[U'\left\{r_t\omega - h(x) - f_1(P) + g_1[I^*]\right\}f_1' - \lambda V'\left\{r_T W + f_2(P) - g_2[I^*]\right\}f_2'\right]\phi(x)dx$$

$$+\int_0^T \left[U'\left\{r_t\omega - h(x) - f_1(P) + g_1[I^*]\right\}g_1' - \lambda V'\left\{r_t W + f_2(P) - g_2[I^*]\right\}g_2'\right]\frac{\partial I^*}{\partial a}\cdot\phi(x)dx. \quad (2.18)$$

The last term is zero, and since $P'(a)>0$ the entire expression is positive for all "a". Together these results mean that in the case under consideration, rebating of premiums increases expected utility of both insurer and insured.

CHAPTER THREE

MORAL HAZARD AND RISK CLASSIFICATION

3.1 Preliminaries

In situations where an individual faces risk, it is typically advantageous for this risk to be shared with others. Contracts that are written for this purpose depend to a large extent on the ability of the individual to influence the probability or magnitude of events, and the ability of contracting parties to observe these activities. Risk-sharing generally reduces an individual's incentive to avoid or mitigate unfavorable events. With unobservability of loss prevention such contracts reduce, but do not eliminate an individual's risk; some must be held to maintain incentives. This relationship between coverage and incentives is called the moral hazard problem. Seminal contributions include Arrow (1963), Spence and Zeckhauser (1971), Pauly (1974) and Mirrlees (1975). More recent contributions include Shavell (1979a,1979b), Holmström (1979) and Grossman and Hart (1983).

The effectiveness of partial coverage as a means to maintain incentives depends upon the characteristics of individuals. The purpose of this chapter is to analyze how these characteristics affect the trade-off between risk-sharing and incentives[13]. One of the fundamental premises of this chapter is that there is a close connection between moral hazard in the presense of asymmetric information and risk classification as

[13] Such comparative static analysis is not appropriate in principal/agent problems unless it is specified how these changes affect the agent's reservation wage (for an example, see Grossman and Hart (1983)). However, in the insurance model used in this chapter, comparative static analysis raises no difficulty because an assumption of actuarial pricing replaces the notion of a reservation wage.

practiced by the insurance industry. If an insurer could observe an individual's loss prevention, the only risk classification required would be with respect to this prevention and associated productivity, in other words, only variables that directly affect the probability of loss. However, if an insurer cannot observe loss prevention it risk classifies using attributes not directly related to prevention or productivity. An insurer does so if these attributes influence the individual's choice of prevention. The analysis that follows is, therefore, important to insurers in the design of risk classification schemes and the policies that are offered or promoted within risk classes. It is also important as a means to determine how exogenous economic shocks or regulatory policy affect risk-sharing and incentives and, as a result, the profitability of contracts that insurers sell or contemplate selling.

Economic agents buy insurance and undertake self-protection (activities which reduce the probability of loss) for the same reason - to mitigate the adverse effect of a loss of wealth accompanying damage to or destruction of an insurable asset. Since they have the same purpose, it might be suspected that insurance and prevention are substitutes. However, an increase in the utility cost of prevention decreases both prevention and coverage. In this case incentives dominate substitution effects. Similar logic underlies the result that if prevention and productivity are homogeneous of degree zero in their effect on the probability of loss, then a decrease in productivity increases loss frequencies and decreases coverage. If utility is state dependent, a decrease in utility in the loss state which does not affect marginal utility of income, increases both prevention and coverage. In this case, the increased incentive for self-protection is strong relative to the increased demand for coverage. As a consequence, increased insurance coverage does not reduce self-protection. Additional results establish the effect of increases in wealth or risk aversion. If risk aversion is decreasing, increasing wealth reduces coverage sold, while if risk aversion is increasing, increasing wealth decreases prevention. If risk cautiousness is in a neighborhood of zero, (risk aversion may be increasing or decreasing) increases in wealth decrease both prevention and coverage. On the other hand, if risk tolerance is increased for a utility function with decreasing risk aversion, coverage sold decreases, while if risk aversion is increasing, prevention is increased. For risk tolerance in the neighborhood of zero, increases have unambiguous effects on both coverage and prevention.

The present chapter is most closely related to that of Grossman and Hart (1983), Shavell (1979a,1979b) and Pauly (1974). Grossman and Hart (G-H) show that increases in an agent's marginal utility cost of self-protection or risk aversion (when the agent's utility of wealth is exponential and wealth and actions are multiplicatively separable) decrease a principal's welfare. This result differs from those decribed above in that it does not consider the associated effect on either the agent's self-protection or the incentive schedule (coverage). Moreover, the G-H result is obtained from a problem in which the principal's welfare is maximized with an action which is implementable at minimum cost. In an insurance context, this is equivalent to a monopolistic insurer. In contrast, the analysis of the present chapter assumes a competitive insurance market and actuarial pricing. Shavell (1979a,1979b) shows that the moral hazard problem disappears when the marginal productivity of prevention is either zero or infinitely large. This result on productivity is extended in the present chapter by considering the effects of changes in the marginal productivity of prevention for both intermediate as well as extreme values.

The model with which the above issues are examined is essentially that of Pauly (1974). The important difference is in the characterization of optimal prevention and coverage. That used in this chapter facilitates comparative static analysis. The model and the characterization of the solution are presented in the following section. Section 3.3 carries out the comparative static analysis.

3.2 The Insurance Model

3.2.1 *Expected Utility*

Let U(W,X) and V(W,X) be the consumer's concave utility functions of wealth, W, and self-protection, X, in the no-loss and loss states respectively, with $U(\cdot,X) \geq V(\cdot,X)$, $0 < \frac{\partial}{\partial W}U(W,X) \leq \frac{\partial}{\partial W}V(W,X)$, $\frac{\partial}{\partial X}U(W,X) < 0$, $\frac{\partial}{\partial X}V(W,X) < 0$. The fact that self-protection appears directly in the utility function indicates that it is not a monetary variable. Decreasing utility with self-protection indicates that self-protection is an activity which individuals avoid (other things being equal, in particular the probability of loss being constant). In automobile insurance, self-protection might relect care in driving. Concavity implies decreasing marginal utility of wealth and increasing marginal disutility of self-protection. The second set of inequalities above indicates increased marginal utility of wealth in the loss state. The probability of incurring a loss is $\pi(X)$. Self-protection is unobservable to the insurer. Assume that $\pi' < 0$, $\pi'' > 0$, and that $\pi(\cdot)$ is thrice differentiable.

Preferences of the individual depend upon wealth and self protection. It is assumed they are consistent with the expected utility function

$$\left[1 - \pi(X)\right]U(S-P,X) + \pi(X)V(S-P+q-L,X),$$

with $X \geq 0$, $0 \leq q \leq L$, and where

S = initial wealth,

L = monetary value of the risky asset,

q = insurance coverage,

P = premium.

If wealth and self-protection are additively separable in the utility function, and a suitable transformation of the variable X is made in the expected utility function so that $U(W,X)=U(W)-cX$, $V(W,X)=V(W)-cX$ then this becomes[14]

$$\left[1-\pi\left(X\right)\right]U\left(S-P\right)+\pi\left(X\right)V\left(S-P+q-L\right)-cX. \tag{3.1}$$

Additive separability of utility between monetary and non-monetary factors is important[15]. Because of this, any transformation of that section of the utility function dependent upon wealth changes the marginal rates of substitution between wealth and self-protection. However, additive separability is plausible because self-protection is assumed non-monetary. The implication of this is that an individual's self-protection to self-protect does not affect preferences for consumption goods. The parameter c is called the utility cost of self-protection. Although this characteristic may not be

[14] Pauly (1974), Shavell (1979) and Arnott and Stiglitz (1983) examine models where the probability of loss depends upon monetary expenditures rather than non-monetary loss prevention as in (3.1). The assumption made by these papers that monetary loss prevention expenditures are unobservable to the insurer is not always suitable. For example, an insurer may often determine, even after destruction by fire, whether a home or business had sprinkler systems, smoke detectors or fire extinguishers. Even if this were not the case the insured may have receipts to verify such expenditures were made. Since loss prevention expenditures are economic transactions they often have a history in themselves which insurers can trace. Loss prevention that effects the insured's utility directly, (self-protection, carefulness etc.) will generally entail no economic transaction and no physical record. In this case the assumption of unobservability is more plausible. Moreover, risk classification in non-life insurance is for the most part unrelated to monetary expenditures, but rather attributes associated with loss-prevention. Modelling self-protection as non-monetary rather than monetary is therefore appropriate for this chapter.

[15] The linearity of self-protection in the utility function is not restrictive. Suppose instead that the utility cost of self-protection, $C(Y)$, is increasing convex with respect to self-protection Y. Make the change of variable $cX=C(Y)$. The inverse transformation $Y=C^{-1}[cX]$ is increasing concave. The probability function is decreasing convex, which means that after making the change of variable from Y to X, the probability function is convex with respect to the new self-protection variable X (see Mangasarian (1970) for details).

directly unobservable, it is assumed that classifying variables used by insurers, (such as marital status, parenthood, age, and gender in automobile insurance) determine it uniquely.

3.2.2 *The Premium Schedule*

The relation between the premium the insurer charges and the coverage the individual chooses depends upon the insurer's observations (subsequent to the sale of insurance) of loss frequencies associated with coverage levels for different risk types. To analyze how these observations are impounded into the premium schedule, the individual's choice problem after the purchase of coverage is considered. When choosing prevention, the individual takes the characteristics of the insurance policy (P,q) as given, this is because self-protection is unobservable to the insurer. The first order condition for a maximum with respect to prevention is[16]

$$\pi'\left[V-U\right]-c=0, \tag{3.2}$$

with $V\equiv V\left(S-P+q-L\right)$, $U\equiv U\left(S-P\right)$. This expression may be interpreted to mean that the insured chooses self-protection only after the purchase of insurance. The fundamental moral hazard problem is that the chosen level of self-protection is influenced by coverage. The first term in (3.2) is the marginal benefit of increased self-protection - the amount by which the probability of loss is decreased times the gain in utility between the loss and no loss states. The second term is the marginal cost of increased self-protection. The insured increases self-protection until the marginal benefit is no longer greater than the marginal cost. Since the insurer is risk neutral in a competitive market the premium charged for coverage q, P(q), is[17]

[16] Later results illustrate that $V-U$ (and $V'-U'$ as well), is non-zero. This insures that second order sufficiency conditions in the individual's maximization with respect to self-protection are satisfied. Later in this chapter, non-negativity of $V-U$ and $V'-U'$ is also important to avoid possible division by zero.

[17] Because $\partial X/\partial P$ is non-negative, it can be shown that there exists a unique solution

$$P=\pi\left[X\left(P,q,c\right)\right]\cdot q, \tag{3.3}$$

where X(P,q,c), found by solving equation (3.2), is the insured's self-protection if a premium P is charged for coverage q. This is actuarially fair regardless of the fact that the insurer cannot observe self-protection. Because the insurer is assumed to be large, loss frequencies for particular risk types are the same as an individual's probability of loss.

The marginal price of coverage is given by

$$P'(q)=\frac{\left[\frac{\partial X}{\partial q}\cdot\pi'\cdot q+\pi\right]}{\left[1-\frac{\partial X}{\partial P}\cdot\pi'\cdot q\right]}. \tag{3.4}$$

In order to insure that the sign of this derivative is positive, the partial derivatives $\frac{\partial X}{\partial q}$ and $\frac{\partial X}{\partial P}$ are found. From equation (3.2),

$$\frac{\partial X}{\partial q}=\frac{-\pi'\,V'}{\pi''\left[V-U\right]}\leq 0, \tag{3.5}$$

$$\frac{\partial X}{\partial P}=\frac{\pi'\left[V'-U'\right]}{\pi''\left[V-U\right]}\geq 0. \tag{3.6}$$

The fact that self-protection increases with the premium charged is the result of decreasing marginal utility which from equation (3.2) increases the marginal benefit of

---

of (3.3) for some value of P. The left hand side is strictly increasing while the right is non-increasing with P. Because the right is greater than or equal the left at P=0, and because of continuity of X(P,q,c), (continuity is insured by the implicit function theorem and the fact that the denominator of the expression in (3.6) is negative) there is a unique P value at which they are equal.

self-protection. The fact that self-protection decreases with coverage indicates that, holding premium effects constant, the insured's incentive to take care is reduced by insurance coverage. Substituting expressions (3.5) and (3.6) into (3.4) yields

$$P'(q)=\frac{-\pi'^{2}V'q+\pi''[V-U]\pi}{\pi''[V-U]-\pi'^{2}q[V'-U']}\geq 0. \tag{3.7}$$

This expression illustrates that the premium schedule P(q) is increasing.

Besides being increasing, the premium schedule is typically convex. This raises the question of whether there might be an advantage to an insured in transacting with more than one insurer. In general, in the courts, property, liability and health insurance policies have been interpreted as contracts of indemnity. That is, an insurer pays only for actual loss to the policy-holder and then only to the face value of the contract. Should the insured receive payments from other sources, for example from another insurer, the first need only contribute the difference up to the face value of its own policy. Transactions costs arguments aside, this generally eliminates any incentive the insured might have to contract with several insurers. This is assumed in the remainder of this chapter. Pauly (1974) and Jaynes (1978) study insurance markets where insurers cannot observe an individual's total purchase of insurance.

3.2.3 *Optimal Self-Protection and Coverage*

Optimal coverage and self-protection can be found by substituting $P=\pi(X)\cdot q$ into equations (3.1) and (3.2) and maximizing the first (insured's expected utility) subject to the second (first order condition for a maximum of self-protection)[18] and the conditions $X\geq 0$, $0\leq q\leq L$. This is equivalent to the insured choosing coverage and self protection to maximize utility subject to a premium schedule that recognizes the incentive for decreased self-protection with coverage and the condition that the premium be actuarially fair.

The Lagrangean for this problem is

$$\left[1-\pi(X)\right]U(S-\pi(X)\cdot q)+\pi(X)V(S-\pi(X)\cdot q+q-L)-cX$$

$$+\lambda\left\{\pi'\left[V(S-\pi(X)\cdot q+q-L)-U(S-\pi(X)\cdot q)\right]-c\right\},$$

where λ is the Lagrangean multiplier. First order conditions for a maximum, interior to the constraints $X\geq 0$, $0\leq q\leq L$, are

$$-\pi'\cdot q\cdot\left[(1-\pi)U'+\pi V'\right]+\lambda\left\{\pi''[V-U]-\pi'^{2}q[V'-U']\right\}=0, \tag{3.8}$$

[18] Mirrlees (1975) shows that it is not generally appropriate in a principal-agent setting to maximize a principal's utility subject to a first order condition obtained from the agent's problem of maximizing utility with respect to unobservable action unless this problem has a unique solution. Since this is the case in the current chapter, no difficulty arises from using the first order condition approach. Mirrlees also shows that there is a class of problems for which there is no significant loss of efficiency as a result of self-interested unobservable behavior. This requires a sequence of contracts for which the agent is increasingly penalized for losses exceeding greater and greater amounts. In this chapter, such contracts are clearly not appropriate. An insurance contract which applied a penalty on top of the loss sustained by the individual would evaporate the demand for insurance.

$$\pi(1-\pi)[V'-U']+\lambda\pi'\Big[(1-\pi)V'+\pi U'\Big]=0, \tag{3.9}$$

$$\pi'[V-U]-c=0, \tag{3.10}$$

with $V\equiv V\Big(S-\pi(X)\cdot q+q-L\Big)$, $U\equiv U\Big(S-\pi(X)\cdot q\Big)$. Eliminating λ from (3.8) and (3.9), the set of first order conditions reduce to

$$F_1\equiv\pi'^{\,2}V'U'q+\pi''[V-U][V'-U']\pi(1-\pi)=0, \tag{3.11}$$

$$F_2\equiv\pi'[V-U]-c=0. \tag{3.12}$$

Equations (3.11) and (3.12) define two implicit relations between X and q. The derivative of the second is

$$\frac{\partial X}{\partial q}=-\frac{\pi'\Big[(1-\pi)V'+\pi U'\Big]}{\pi''[V-U]-\pi'^{\,2}q[V'-U']}<0. \tag{3.13}$$

The derivative of the first is

$$\frac{\partial X}{\partial q}=-\frac{\dfrac{\partial F_1}{\partial q}}{\dfrac{\partial F_1}{\partial X}},$$

where

$$\frac{\partial F_1}{\partial q}=\pi'^{\,2}V''U'\cdot q\cdot(1-\pi)-\pi'^{\,2}V'U''\cdot q\cdot\pi+\pi'^{\,2}V'U'$$

$$+\pi''\Big[V'(1-\pi)+U'\pi\Big][V'-U']\pi(1-\pi)+\pi''[V-U]\Big[V''(1-\pi)+U''\pi\Big]\pi(1-\pi). \tag{3.14}$$

In order to sign this derivative, note that all terms are positive except the first. However, the last term can be separated into two positive terms

$$\pi'' [V-U]V'' (1-\pi)^2\pi + \pi'' [V-U]U'' \pi^2(1-\pi). \tag{3.15}$$

When equation (3.11) is satisfied, the first term of (3.15) and the first term of (3.14) equal

$$\frac{-V'' \pi'' [V-U][V'-U']\pi(1-\pi)^2}{V'} + \pi'' [V-U](1-\pi)^2 V'' \pi.$$

Multiplying by $V' \cdot U'$, rearranging terms and simplifying, this reduces to

$$V'' U'^2 \pi'' [V-U]\pi(1-\pi)^2 \geq 0.$$

The positivity of this term implies that $\frac{\partial F_1}{\partial q} > 0$ when (3.11) is satisfied. This together with continuity of F_1 and the fact that F_1 is negative when $q=0$ and positive when $q=L$ (this requires the assumption that $V'(\cdot,X)=U'(\cdot,X)$), means that for fixed X, the solution to equation (3.11) is unique (moreover, for every X there is a q that satisfies this equation). This means that if there is a solution to equations (3.11) and (3.12), with self-protection and coverage both positive, it is also unique. If $\pi(0)$ is sufficiently large, for example if $\pi(0)=1$, the expression in (3.11) is positive when $X=0$, but approaches zero when X is large. This, together with the uniqueness of the solution to (3.11) means that if for a fixed q there is a solution to equation (3.11), the expression F_1 first decreases with X in the neighborhood $X=0$, becomes negative and then increases to approach zero. This means $\frac{\partial F_1}{\partial X}$ is negative when (3.11) is satisfied. Together, the facts that $\frac{\partial F_1}{\partial q}$ is positive and $\frac{\partial F_1}{\partial X}$ is negative (when equation (3.11) holds) mean that the implicit relation defined by (3.11) is positive. Since second order sufficient conditions for a maximum are satisfied[19], the intersection between the positive relation

[19] Second order necessary conditions for a maximum require $\partial F_1/\partial X \cdot \partial F_2/\partial q - \partial F_1/\partial q \cdot \partial F_2/\partial X \geq 0$, this is the same as the condition that the determinant of the bordered Hessian be positive. The results $\partial F_2/\partial q \leq 0$, $\partial F_2/\partial X \leq 0$ and $\partial F_1/\partial X \leq 0$, $\partial F_1/\partial q \geq 0$ (when equation (3.11) holds) in-

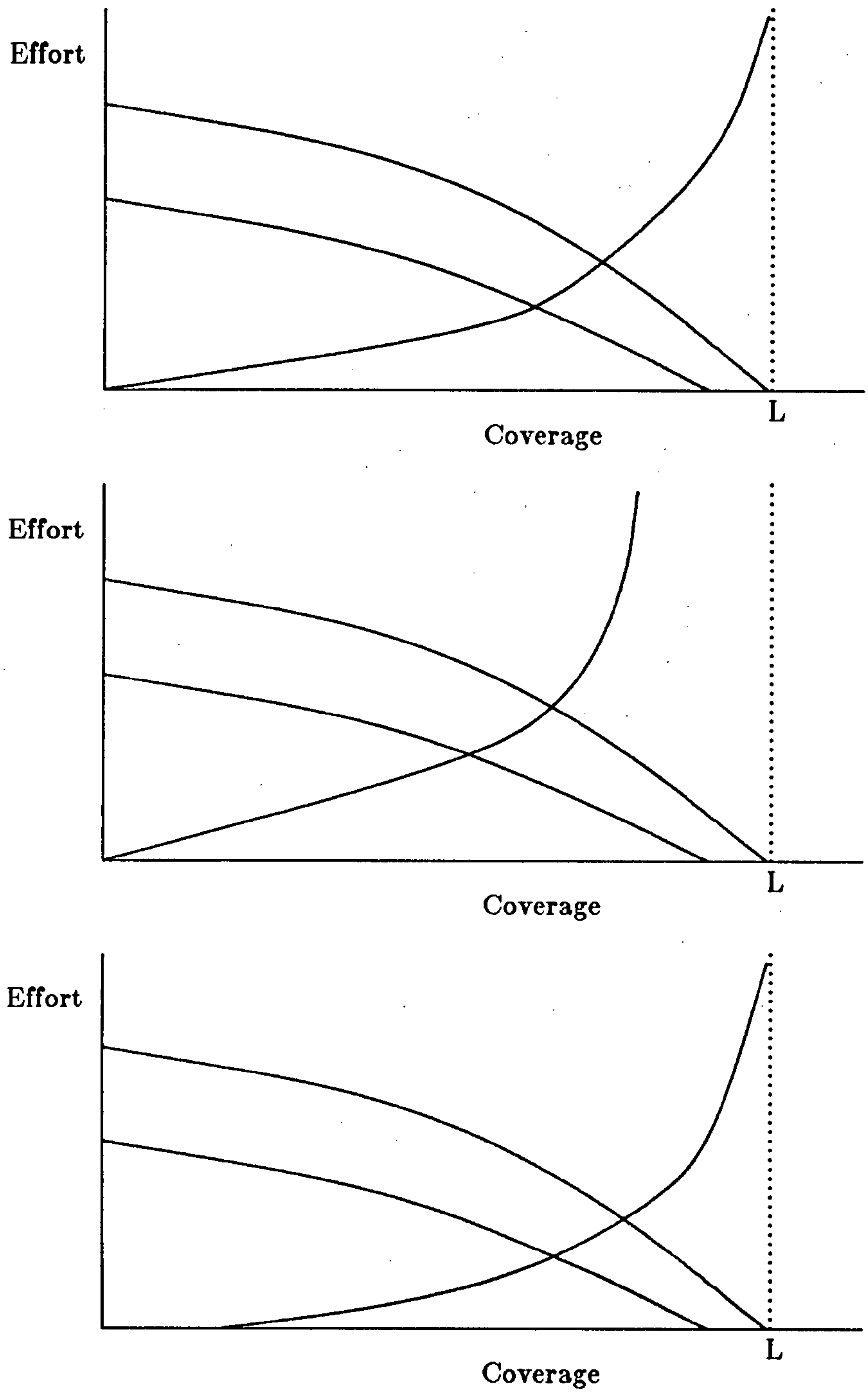
defined by (3.11) and the negative one defined by (3.12) give locally optimal insurance coverage q^* and self-protection X^*. The relations defined by equations (3.11) and (3.12) are plotted in Figure 1.

The downward sloping curves give prevention/coverage pairs that are consistent with actuarial pricing on the part of an insurer. In this chapter, they are called incentive curves because they out-line the coverages required to induce individuals to choose given protection levels. In other words, for points along this curve, if an insurer expects an individual's prevention to be X^0 when coverage is q^0, and if the individual in fact chooses coverage q^0 then he/she also chooses prevention X^0, fulfilling the insurer's expectations. Points above this line yield negative expected profit to the insurer, while points below yield positive expected profit. The upward sloping curves in these figures give points at which for a fixed prevention level the individual maximizes utility with respect to coverage. As such, they are called demand for coverage curves. The expression F_1 may be given the following interpretation. At any prevention, coverage pair say X^0, q^0, the insured is charged the actuarial premium $\pi(X^0) \cdot q^0$ for coverage q^0, but must decide whether to buy an incremental unit of coverage. The individual makes this decision on a presumption that either he/she will actually adhere to this level of prevention or that the insurer does not charge extra for reduced prevention associated with increased coverage (either of these presumptions are justified only in equilibrium). The insurer on the other hand charges a premium rate for this marginal unit that reflects the fact that the extra unit induces the individual to decrease prevention. Because the individual makes the purchase decision on the extra unit of coverage on the assumption that his/her prevention will not change, he/she views the price of the marginal unit as actuarially unfair. Moreover, with higher levels of coverage, the marginal utility of increased coverage is smaller. For some level of coverage, the marginal unit is not

sure that it is satisfied.

purchased. These are the points that make up the upward sloping curves in Figure 1. This argument may also be made mathematically as follows. The expression in (3.13) evaluated at an arbitrary prevention, coverage pair X^0, q^0 is the amount an insurer expects prevention to fall for an increase in coverage. Multiplying by $\pi'(X^0)$ to reflect the increased actuarial cost of coverage and by q to reflect the fact that this increased cost is paid on intramarginal as well as marginal units of coverage and adding $\pi(X^0)$ gives the amount the insurer must charge to break even on an incremental unit. If the individual is charged a premium rate $\pi(X^0)$ for units up to q^0, but the calculated marginal coverage cost for the extra unit, the marginal change in utility (except for a negative dividing term) is given by expression F_1 in (3.11). The individual bears the premium cost of any reduction in prevention induced by the marginal unit of coverage. To the left of the upward sloping curves in Figure 1 the utility value of incremental coverage to the individual is greater than the cost of the marginal unit; to the right, the opposite is true. The fact that this curve is increasing indicates that the portion of the marginal premium that accounts for decreased individual prevention with coverage is decreasing with anticipated prevention by the individual. The intuitive reason for this is the convexity of the probability function.

Figure 1: Optimal Coverage and Self-Protection



3.3 Risk Classification

3.3.1 *Utility Cost of Effort*

A fundamental aspect of the problem at hand is that partial rather than full coverage is required to maintain an insured's incentive to take care. In fact, with full coverage, the individual always undertakes the minimum self-protection $X=0$. An increase in the utility cost of self-protection creates an incentive for the insured to undertake less self-protection subsequent to the purchase of insurance, than expected by the insurer. This erosion in the incentive to take care can only be abated by decreased insurance coverage, this decrease is induced by an upward rotation in the premium schedule. In contrast, without such premium effects, individuals with higher utility costs of self-protection substitute coverage for more costly self-protection. Proposition 3.1 indicates that both prevention and coverage decrease with the utility cost of self-protection. In this case, the role of insurance coverage as a provider of incentives is stronger than its role as a substitute for self-protection.

Proposition 3.1: Increases in the utility cost of self-protection increase accident frequencies and decrease coverage sold.

Proof: Note that the demand for coverage curve, equation (3.11), is unaffected by the utility cost of self-protection parameter **c**. The incentive curve (3.12) is, however, shifted downward by an increase in **c** (this shift is equivalent to an upward rotation in the premium schedule). This downward shift implies that both optimal coverage and self-protection decrease with the utility cost of self-protection. □

This result may also be interpreted to mean that classes of insureds with the highest utility costs of self-protection have the highest accident frequencies but the lowest coverages.

Even though the demand for coverage schedule is increasing and the incentive schedule is shifted upward by decreases in the utility cost of self-protection, it is not always the case that when $c \to 0$ the optimal solution approaches full coverage. One indication of this possibility, is that in (3.12) when $c \to 0$ it need only be the case that $X \to \infty$ or $q \to L$ (if $U(\cdot)=V(\cdot)$), but not necessarily both. In other words the optimal solution may have infinite self-protection but less than full coverage. In this case, the demand for coverage curve (3.11), approaches infinity before full coverage is reached. This possibility is illustrated in the second diagram of Figure 1. When, however, the greatest lower bound of π, denoted by $\bar{\pi}$ is greater than zero and $U'(\cdot)=V'(\cdot)$, the optimal solution is $(X \to \infty\ , q \to L)$. To see this result, first note that $(X^* < \infty\ , q \to L)$ cannot be an optimal solution. Although (3.12) is satisfied, the expression F_1 in (3.11) is positive rather than zero. This possibility means that when $c \to 0$, $X^* \to \infty$ is part of the optimal solution. In this case the individual's utility approaches $(1-\bar{\pi})U(S-\bar{\pi}q)+\bar{\pi}V(S-\bar{\pi}q-L)$. Since this expression is increasing in q, the optimal solution must have full coverage[20]. This result is illustrated in the first and third diagrams of Figure 1.

Using the envelope theorem, the change in the individual's maximized utility with respect to the utility cost of self-protection, is given by

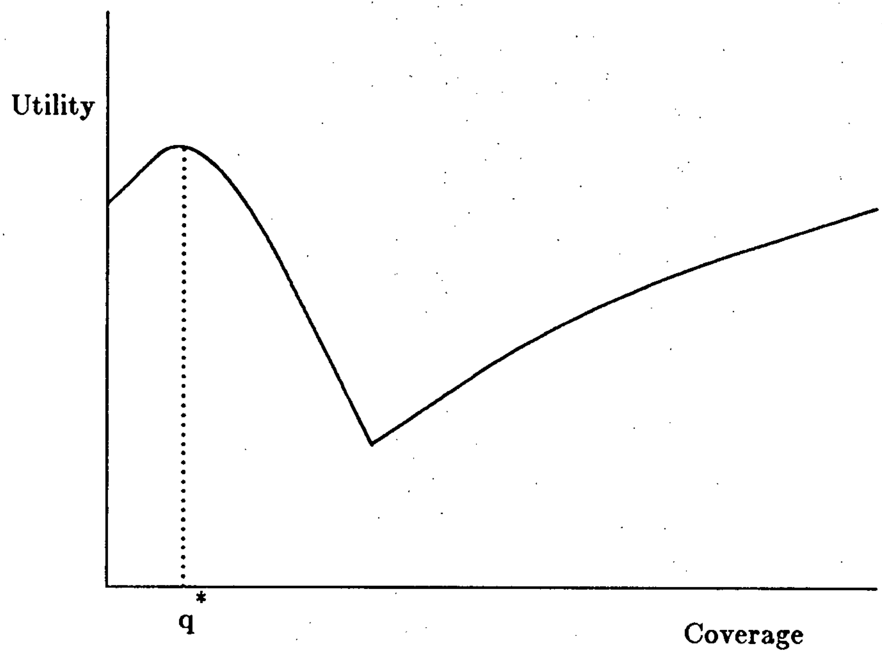
$$-\left[\pi V'+(1-\pi)U'\right]\frac{\pi' q^*}{\pi''[V-U]-\pi'^{\,2}q^*[V'-U']}-X^*.$$

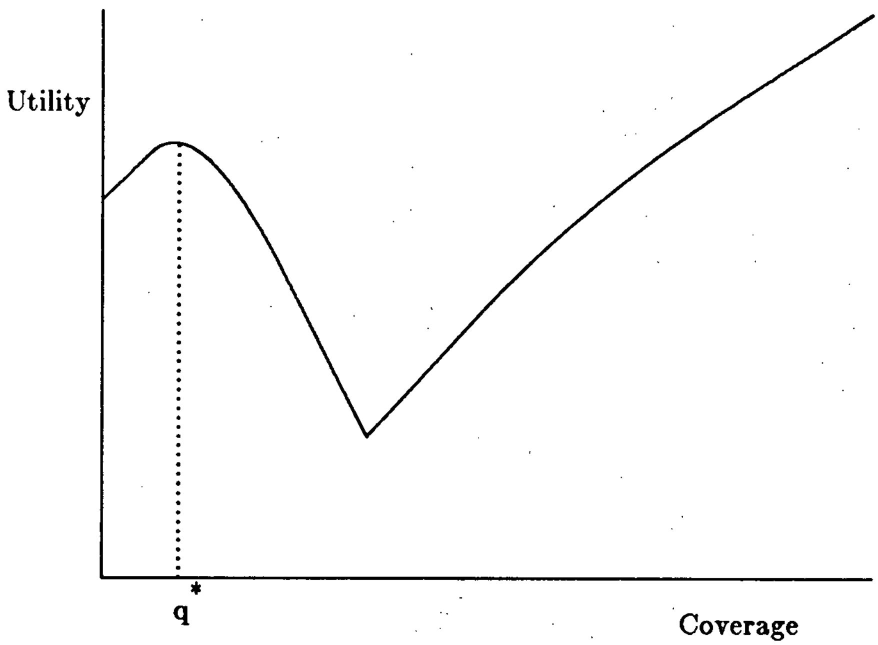
This expression is negative for all positive q^* and X^* and approaches zero only when q^* and X^* approach zero. If $\pi(0)=1$, then $(X=0, q=0)$ is a solution of (3.11) (see the first and second diagrams of Figure 1). Together, these two results imply that as the utility cost of self-protection increases, maximized utility approaches $V(S-L)$. This utility level is obtained well before the utility cost of self-protection goes to infinity. This

[20] The last part of this proof has been adapted from Shavell (1979a,pp459).

result is illustrated by the fact that the relation between X and q that satisfies equation (3.12) collapses to the single point $X=0, q=0$, when $c=\pi'(0)\left[V(S-L)-U(S)\right]$. Of course for any greater values of the utility cost of self-protection, $V(S-L)$ is also the individual's utility. In contrast to the case $\pi(0)=1$, when $\pi(0)<1$, as the utility cost of self-protection increases utility associated with the solution to equations (3.11) and (3.12) approaches $[1-\pi(0)]U(S-\pi(0)\bar{q})+\pi(0)V(S-\pi(0)\bar{q}+\bar{q}-L)$, where $\bar{q}$ is the smallest level of coverage for which (3.11) and (3.12) have a solution (see the third diagram of Figure 1). This utility is never greater than the utility the individual obtains by undertaking no self protection, but taking full coverage $V(S-\pi(0)L)$. The two expressions are equal only when $\pi(0)=1$ because in this case $\bar{q}=0$. This result indicates that for high values of the utility cost of self-protection, the solution of (3.11) and (3.12), $(X^*, \; q^*)$, is a local maximum but not a global maximum. The relation between utility and coverage when the non-linear premium schedule (3.3) is used and the individual chooses self-protection $X(P(q),q,c)$ is shown in Figure 2. When coverage is large the constraint $X\geq 0$ becomes binding. For this level of self-protection the premium $\pi(0)q$ is charged, and at this point the individual's utility is increasing with coverage. This relationship is illustrated by the kink in Figure 2. The individual's maximized utility is obtained either with the solution of (3.11) and (3.12), $(X^*, \; q^*)$, or $X=0, q=L$. For low values of the utility cost of self-protection the first solution yields greater utility, for larger values, the converse is true.

Fiqure 2: Utility vs. Coverage





3.3.2 *Productivity of Prevention*

The probability of loss is assumed to depend upon prevention and a productivity parameter "a," i.e., $\pi = \pi(X,a)$. The productivity parameter is observable to the insurer and may be given several interpretations. It may represent characteristics of risk types associated with the the ability to reduce the probability of loss and thus may be used by insurers for classification. It may represent loss-prevention that is undertaken by insurers on behalf of their policy-holders. The extent to which insurers pursue these activities depends upon their effects on both loss frequencies and coverages sold. On the other hand, the productivity variable may summarize the available technology with which individuals are able to affect the probability of loss. For example, in automobile insurance this is influenced by highway design, traffic control and types of vehicles. Finally, the productivity variable is affected by regulatory policy. Examples include reductions in speed limits, vehicle safety standards and the enforcement of seat-belt laws. The last two interpretations of the productivity of prevention are studied in Peltzman (1975), Viscusi (1984) and Crandall and Graham (1984).

In this subsection is is assumed that productivity and prevention are homogeneous of degree zero in their effect on the probability of loss, i.e., $\pi(X,a) = \pi(\frac{X}{a})$. This form of the relation between prevention and productivity is used in Shavell (1979a). The term X/a is called effective self-protection because it is this that determines the probability of loss. An increase in "a" increases the probability of loss so that this represents a decrease in productivity.

Proposition 3.2: If prevention and productivity are homogeneous of degree zero in their effect on the probability of loss, an increase in productivity decreases loss frequencies and increases coverage sold.

Proof: Equations equivalent to (3.11) and (3.12) are

$$\frac{1}{a^2}(\pi'(\frac{X}{a}))^2 V' U' q+\frac{1}{a^2}\pi'' [V-U][V'-U']\pi(1-\pi)=0 \tag{3.16}$$

$$(\frac{1}{a})\ \pi'(\frac{X}{a})[V-U]-c=0. \tag{3.17}$$

Making the change of variable $Y=\frac{X}{a}$, and multiplying both sides of (3.16) by a^2 yields

$$(\pi'(Y))^2 V' U'q+\pi'' (Y)[V-U][V'-U']\pi(Y)(1-\pi(Y))=0, \tag{3.18}$$

$$(\frac{1}{a})\pi' (Y)[V-U]-c=0. \tag{3.19}$$

This set of equations is the same as (3.11) and (3.12) except for the term $(\frac{1}{a})$ in the second equation. Since Y and q are negatively related in (3.19), but positively related in (3.18), a decrease in the productivity parameter "a" (increase in productivity) increases effective self protection $\frac{X^*}{a}$ and optimal coverage q^*. □

As was the case in Proposition 3.1, Proposition 3.2 implies that classes of insureds with the highest accident frequencies also take the lowest coverages. The reason for this is the same as with the utility cost of self-protection, the role of partial coverage as a provider of incentives outweighs its role as a substitute for self-protection.

Proposition 3.2 has the interesting interpretation that as the ability to affect the probability of loss is decreased, loss frequencies increase while coverages offered by insurers decrease. With transactions costs, for example fixed costs in the provision of insurance, this result indicates that individuals with the least ability to prevent losses are the most likely to be unable to obtain insurance coverage from insurers. This is consistent with the observation that elderly drivers often have difficulty obtaining insurance in competitive markets[21].

[21] A similar interpretation is appropriate when considering the utility cost of self-protection. That is, individuals with the greatest utility costs of self-protection will have the greatest difficulty obtaining insurance coverage. This result is consistent with the ob-

If equation (3.19) is multiplied by "a", it is clear that the comparative static results obtained above relating the utility cost of self-protection c and optimal effort and coverage X^*, q^* also hold for the relation between the productivity of self-protection "a" and optimal effective self-protection and coverage $(\frac{X^*}{a}, q^*)$. In particular if "a" approaches zero $\frac{X^*}{a} \to \infty$ and a sufficient condition for full coverage is $\pi(\infty)>0$; as "a" becomes large, and if $\pi(0)=1$, $\frac{X^*}{a}$ and q^* both approach zero and utility approaches its minimum $V(S-L)$, if however, $\pi(0)<1$, the individual switches to zero self-protection but full coverage when utility falls below[22] $V(S-\pi(0)L)$.

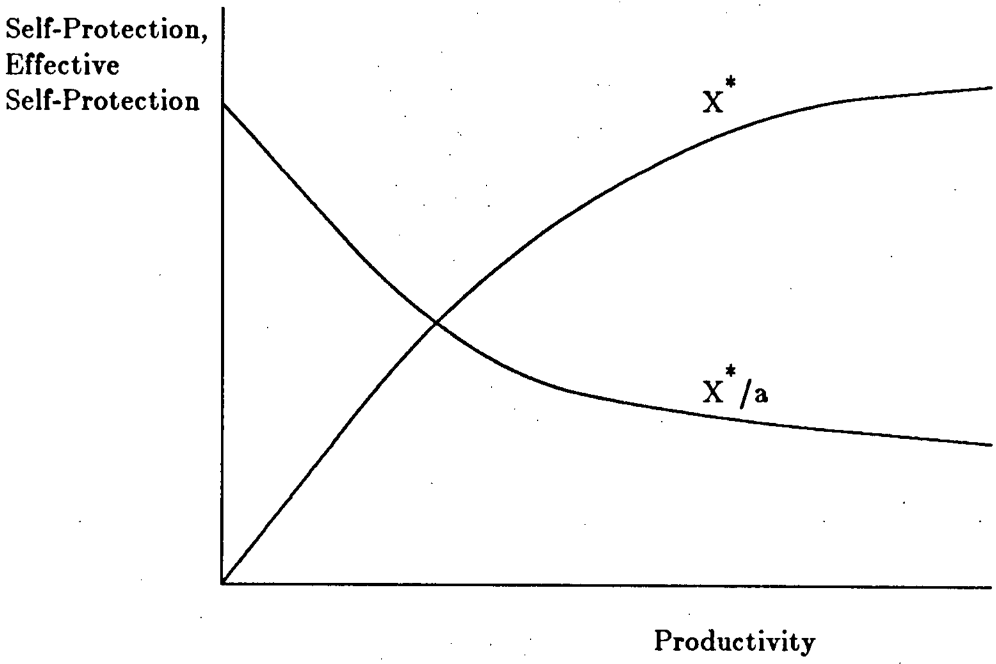
Proposition 3.2 shows that loss frequencies decrease with the productivity of prevention. This is because effective prevention $\frac{X^*}{a}$ increases. However, it will often be the case that prevention itself, X, decreases. This relationship is illustrated in Figure 3. If the productivity parameter "a" is viewed as the stringency of safety regulation then this Figure illustrates an offsetting behavioral response. This forms the basis of Peltzman's (1975) study of automobile safety regulation. Peltzman argues that safety regulations may reduce risk to drivers and their passengers, but this may be more than offset by an increased risk to pedestrians through an increase in driving intensity. While the model in this chapter is not sufficiently general to consider this question directly, the above Figure may be interpreted, in a broad manner, to be consistent with this conclusion. Driving intensity, $1/X$, is increased in response to safety regulation. If the probability of accidents that involve pedestrians depends upon X, while the probability of

servation that youthful drivers often have difficulty obtaining coverage in competitive insurance markets.

[22] Shavell (1979) has shown that when self-protection is monetary, $a \to 0$ and $\pi(\infty)>0$ implies full coverage, and under the same conditions, if "a" is large coverage is also full. The results in the present chapter differ in that they apply to non-monetary self-protection and are more complete in that they also consider the relation between productivity and coverage for intermediate as well as extreme values for the productivity parameter "a".

all accidents depends upon X/a, then safety regulations reduce total accident frequencies, but increase the frequency of accidents involving pedestrians. Viscusi (1984) suggests the possibility that adverse behavioral response to safety regulation may be sufficiently great as to actually increase the probability of loss. This suggestion is inconsistent with the result in Proposition 3.2.

Fiqure 3: Productivity and Self-Protection



3.3.3 *State Dependent Utility*

While the example in the above sub-section supports the view that safety regulations induce an adverse behavioral response, in general, this conclusion depends upon the form of the relationship between prevention and productivity. For example, if productivity and prevention are additively separable in their effects on the probability of loss, it is easy to find cases where increases in productivity decrease, but eventually increase prevention. On the other hand, safety regulations may affect utility in the loss state rather than the probability of loss. Examples in automobile regulation include seat-belts, energy-absorbing steering columns, and passive restraint systems. Proposition 3.3 shows that safety regulation that increases utility in the loss state always reduces prevention. Another interpretation of an increase in utility in the loss state is the non-financial service component of insurer's claim settlement. Examples include the speed of claim settlement, loaner vehicles, and presettlement payments.

Proposition 3.3: An increase in utility in the loss state which does not affect marginal utility of wealth reduces both prevention and coverage sold.

Proof: Define a utility parameter K such that V(W,X)=V(W)+K−cX. An increase in K reduces the difference between utilities in the loss and no-loss states. This reduces the incentive for prevention, the incentive curve is shifted downwards. On the other hand, an increase in K reduces the need for insurance coverage, the demand for coverage curve is shifted to the left. This can be seen by noting that F_1 in (3.11) is increased by an increase in K. These two shifts mean that the effect on coverage is unambiguously negative. To find the effect on prevention, note that, $\dfrac{\partial X^*}{\partial K}=\dfrac{\dfrac{-\partial F_1}{\partial K}\cdot\dfrac{\partial F_2}{\partial q}+\dfrac{\partial F_2}{\partial K}\cdot\dfrac{\partial F_1}{\partial q}}{D}$, $D=\left\{\dfrac{\partial F_1}{\partial X}\cdot\dfrac{\partial F_2}{\partial q}-\dfrac{\partial F_1}{\partial q}\cdot\dfrac{\partial F_2}{\partial X}\right\}\geq 0.$ Since $\dfrac{\partial F_1}{\partial K}=\pi''[V'-U']\pi(1-\pi)$, $\dfrac{\partial F_2}{\partial K}=\pi'$ and $\dfrac{\partial F_2}{\partial q}=\pi'\Big[(1-\pi)V'+\pi U'\Big]$, $\dfrac{\partial F_1}{\partial K}\cdot\dfrac{\partial F_2}{\partial q}$ is equal to the negative of $\dfrac{\partial F_2}{\partial K}$ times the fourth term of $\dfrac{\partial F_1}{\partial q}$ given in equation (3.14), and these two terms, therefore, cancel. The first and fifth terms of $\frac{\partial F_1}{\partial q}$ have been shown to be positive at (X^*,q^*) so that all remaining terms are negative. The effect on prevention is, therefore, negative. □

This proposition illustrates that an increase in utility in the loss state has a strong adverse effect on the incentive for self-protection relative to the decreased demand for coverage. While the service component of an insurer's claim settlement policy may increase the number of policies sold, this has an adverse effect on loss frequencies and the level of coverage that can be offered in individual policies. At the same time, safety regulations that affect utility in the loss state may be expected to increase loss frequencies.

3.3.4 *Initial Wealth*

This subsection examines the effect of changes in wealth on prevention and coverage. Attention is restricted to the HARA family of utility functions[23], $V(W)=U(W)=(\frac{1}{B-1})(BW+g)^{1-1/B}$. This family includes the most common utility functions as special cases. The parameter B is called risk cautiousness. The Arrow-Pratt measure of absolute risk aversion is $(BW+g)^{-1}$. The inverse of this expression is called risk tolerance.

Proposition 3.4: If utility of wealth is HARA, (i) with decreasing risk aversion increases in wealth reduce coverage sold, (ii) with increasing risk aversion prevention is decreased with wealth, (iii) for utility functions in the neighborhood of constant risk

---

[23] The letters HARA stand for "hyperbolic absolute risk aversion." The utility function is so called because absolute risk aversion is given by $(BW+g)^{-1}$, the inverse called risk tolerance is linear in wealth. The parameter B is called risk cautiousness. It is easily shown that when $B>0$, risk aversion decreases with wealth, when $B<0$, risk aversion increases with wealth. When $B=-1$ the utility function is quadratic, when $B>1$ it is a power utility function, when $B\to 0$, a utility function with constant risk aversion is approached, when $B\to 1$, a logarithmic utility function is approached. Risk causitiousness, B, cannot equal zero or one. Note that $g\geq 0$, and it is required that $BW+g>0$.

aversion increasing wealth decreases both coverage and prevention.

Proof: The essence of the proof is to show that the incentive curve defined by (3.12) is shifted downwards by an increase in wealth, while the demand for coverage curve defined by (3.11) is shifted to the left if risk aversion is decreasing with wealth, to the right if risk aversion is increasing with wealth and is essentially unchanged if risk aversion is essentially unchanged by wealth. These shifts establish (i), (ii) and (iii).

Risk tolerance is given by BW+g, so that risk aversion is increasing or decreasing with wealth depending upon whether $B<0$ or $B>0$. Define $W_1=S-\pi(X)\cdot q$, $W_2=S-\pi(X)\cdot q+q-L$, then the first step is to show that $\pi'[V-U]-c$ decreases with S. The derivative of this expression is

$$\pi'\left[(BW_2+g)^{-1/B}-(BW_1+g)^{-1/B}\right].$$

This expression is non-positive regardless of the sign of B.

Next it is shown that $\frac{[V-U][V'-U']}{V'\,U'}$ increases with S when $B>0$, decreases with S when $B<0$, and is unchanged with S when B approaches zero. Increases or decreases in this expression correspond to left or rightward shifts in the demand for coverage curve. This expression equals

$$\frac{1}{B-1}\left[(BW_2+g)^{1-1/B}(BW_1+g)^{1/B}-(BW_1+g)-(BW_2+g)+(BW_1+g)^{1-1/B}(BW_2+g)^{1/B}\right].$$

The derivative with respect to S equals

$$Z=\frac{1}{B-1}\left[(B-1)(a+1/a)+(b+1/b)-2B\right],$$

where $a=\left[\frac{BW_1+g}{BW_2+g}\right]^{1/B}$, $b=\left[\frac{BW_1+g}{BW_2+g}\right]^{1/B-1}$. Since $W_2\leq W_1$ and $(B-1)(a+1/a)+(b+1/b)$ equals 2B when $W_2=W_1$, the sign of

$(B-1)(a+1/a)+(b+1/b)-2B$

can be determined by the sign of the derivative with respect to W_2. This is

$$(B-1)\Big[-(BW_1+g)^{1/B}(BW_2+g)^{-1/B-1}+(BW_1+g)^{1/B-1}(BW_2+g)^{-1/B}$$
$$+(BW_2+g)^{1/B-1}(BW_1+g)^{-1/B}-(BW_2+g)^{1/B-2}(BW_1+g)^{-1/B+1}\Big].$$

This expression equals

$$(B-1)(a+1/b)\Big[(BW_1+g)^{-1}-(BW_2+g)^{-1}\Big].$$

This expression is negative when $B<0$, or $B>1$, and positive when $0<B<1$. This result means $(B-1)(a+1/a)+(b+1/b)-2$ is positive for $B<0, B>1$ and negative for $0<B<1$. As $B \to 1$, $a \to \frac{(W_1+g)}{(W_2+g)}$ and $b \to 1$. As $B \to 0$, both "a" and b approach $e^{(W_1-W_2)/g}$. This result means that $(B-1)(a+1/a)+(b+1/b)-2B \to 0$ as $B \to 0$ or $B \to 1$. These results mean that Z is negative for $B<0$, positive for $B>0$ and $Z \to 0$ as $B \to 0$. Using L'Hospital's rule, it can also be shown that Z is finite for $B \to 1$. □

Increasing wealth decreases the incentive for prevention, while it increases or decreases the demand for coverage depending upon whether risk aversion is increasing or decreasing. The combination of these two effects gives the results of Proposition 3.4.

3.3.5 *Risk Aversion*

This sub-section examines the effect of Risk Aversion on prevention and coverage. The restriction to Hara utility function is maintained.

Proposition 3.5: When utility of wealth is HARA,

(i) if risk aversion decreases with wealth, an increase in risk tolerance decreases coverage sold,

(ii) if risk aversion increases with wealth, an increase in risk tolerance increases prevention,

(iii) if risk cautiousness is in the neighborhood of zero, and:

(a) risk aversion is less than one ($g>1$): when risk cautiousness is in a positive neighborhood of zero, an increase in risk tolerance increases prevention and decreases coverage; when risk cautiousness is in a negative neighborhood of zero, an increase in risk tolerance increases both prevention and coverage.

(b) risk aversion is greater than or equal to one ($g\leq 1$): when risk cautiousness is in a positive neighborhood of zero, an increase in risk tolerance decreases both prevention and coverage; when risk aversion is in a negative neighborhood of zero, an increase in risk tolerance increases prevention and decreases coverage.

Proof Since risk aversion is $(BW+g)^{-1}$ a decrease in risk aversion is viewed the result of an increase in the parameter g. There are three parts to this proof. First it is shown that a decrease in risk aversion decreases the demand for coverage. In other words, the demand for coverage curve is shifted to the left. Secondly, it is shown that the incentive to self-protect is decreased or increased with decreased risk aversion depending upon whether risk aversion is decreasing or increasing with wealth. In other words, incentive curve is is shifted downwards or upwards with decreased risk aversion depending upon whether risk aversion decreases or increases with wealth. The last part of the proof shows that the effect of risk aversion on the incentive for self-protection is discontinuous in the neighborhood of zero risk cautiousness.

Part 1: It is shown that $\frac{[V-U][V'-U']}{V'U'}$ is increasing with g. The derivative with respect to g is Z/B where Z is defined in the proof of proposition 3.4. Recall that Z has been shown to be positive for B>0, negative for B<0 and approaches zero as $B \to 0$. This result means that Z/B is always non-negative. Appendix A shows that $0 \le \frac{Z}{B} < \infty$ as $B \to 0$.

Part 2: It is shown that $\pi'[V-U]-c$ is decreasing or increasing with g depending upon whether B is positive or negative. The derivative is

$$ZZ = \frac{\pi'\left[(BW_2+g)^{-1/B}-(BW_1+g)^{-1/B}\right]}{B}.$$

The numerator is always negative, so that $\pi'[V-U]-c$ increases with g when B<0, and decreases with g when B>0.

Part 3: It is shown that when B approaches zero from the right, ZZ approaches zero or minus infinity depending upon whether g>1 or $g \le 1$; when B approaches zero from the left, ZZ approaches positive infinity or zero depending upon whether g>1 or $g \le 1$. Note that

$$ZZ = \pi' \frac{g^{-1/B}}{B}\left[(BW_2/g+1)^{-1/B}-(BW_1/g+1)^{-1/B}\right].$$

As B approaches zero from the left or right the term in the square brackets approaches

$$\left\{\exp[-W_2/g]-\exp[-W_1/g]\right\} \ge 0.$$

This result means that when

$$g>1, \quad \begin{cases} B \to 0_+ & ZZ \to 0 \\ B \to 0_- & ZZ \to \infty, \end{cases}$$

$$g \leq 1, \quad \begin{cases} B \to 0_+ & ZZ \to -\infty \\ B \to 0_- & ZZ \to 0. \end{cases}$$

Parts 1 and 2 of this proof establish parts (i) and (ii) of the proposition. When ($g>1$ and $B \to 0_+$) or ($g \leq 1$ and $B \to 0_-$), the incentive effect with respect to a decrease in risk aversion is slight. In other words, the incentive curve is essentially unaffected. The resulting leftward shift in the demand for coverage curve means that prevention is increased while coverage is decreased. When $g>1$, and B approaches zero from the left, the fact that $ZZ \to \infty$, means that the resulting upward shift in the incentive curve is significant. This means that both coverage and prevention increase. When $g \leq 1$, and B approaches zero from the right, the fact that $ZZ \to -\infty$, means that the resulting downward shift in the incentive schedule is significant. As a result of this shift both coverage and prevention decrease. This result completes part (iii) of the proposition[24]. □

Perhaps the most interesting result in Proposition 3.5 is that it is possible with a decrease in risk aversion for both prevention and coverage to increase. In this case the increased incentive for protection resulting from an increase in the difference between utilities in the loss and no-loss states, dominates the reduced need for coverage. This result indicates that when incentive effects are present, the role of the Arrow-Pratt calculation as a measure of risk aversion is obscured. The reason for this is that the Arrow-Pratt measure affects not only the demand for coverage but also the incentive for self-protection.

[24] The magnitude, but not the direction of shifts in both the demand for coverage and the incentive curves are affected if that section of the utility function dependent upon wealth is multiplied by a positive multiplying factor. The magnitude of shifts is affected because the marginal rates of substitution between wealth and prevention are changed. Results in Proposition 3.4 part iii and Proposition 3.5 part iii can, as a result, change if the positive multiplying factor goes to zero or infinity as risk cautiousness goes to zero.

3.4 Summary

This chapter has studied the effect that characteristics of individuals have on the trade-off between risk sharing and incentives in insurance markets affected by moral hazard. In particular, major results identify the effect of utility cost of self-protection, productivity of effort, utility in the loss state, wealth and risk aversion.

3.5 Appendix to Chapter 3

In this appendix it is shown that $0 \leq \frac{Z}{B} < \infty$ as $B \to 0$. All limits are taken as $B \to 0$. Note that

$$\text{lm}\{\frac{Z}{B}\} = \text{lm}\{\frac{1}{B-1}\} \cdot \text{lm}\left\{ \left[(B-1)(a+\frac{1}{a}) + (b+\frac{1}{b}) - 2B \right] / B \right\}$$

$$= -\text{lm}\left\{ \left[(B-1)(a+\frac{1}{a}) + (b+\frac{1}{b}) - 2B \right] / B \right\}.$$

Since both the numerator and denominator of the term in curly brackets approach zero, L'Hospital's rule is used. This requires finding

(A) $\text{lm}\{a+\frac{1}{a}\}$,

(B) $\text{lm}\left\{\frac{d}{dB}(a+\frac{1}{a})\right\}$,

(C) $\text{lm}\left\{\frac{d}{dB}(b+\frac{1}{b})\right\}$.

(A)

$$\operatorname{lm}\{a\}=\exp\left[\frac{W_1-W_2}{g}\right],$$

so

$$\operatorname{lm}\{a+\frac{1}{a}\}=\exp\left[\frac{W_1-W_2}{g}\right]+\exp\left[-\frac{W_1-W_2}{g}\right].$$

(B)

$$\operatorname{lm}\{\frac{d}{dB}(a+\frac{1}{a})\}=\operatorname{lm}\{\frac{da}{dB}\}\cdot\left\{1-\exp[-2\frac{W_1-W_2}{g}]\right\}.$$

And,

$$\operatorname{lm}\{\frac{da}{dB}\}=\operatorname{lm}\left[\frac{g(W_1-W_2)/[(BW_1+g)(BW_2+g)]-\left\{\ln[(BW_1+g)/(BW_2+g)]\right\}/B}{B}\right]\cdot\operatorname{lm}\{a\}.$$

Both numerator and denominator of the term in the large square brackets approach zero, the numerator approaches zero because both terms approach $\frac{(W_1-W_2)}{g}$. Thus,

$$\operatorname{lm}\{\frac{da}{dB}\}=-\left[\frac{(W_1-W_2)(W_1+W_2)}{g^2}\right]\cdot\operatorname{lm}\{a\}-\operatorname{lm}\{\frac{da}{dB}\}.$$

Therefore,

$$\operatorname{lm}\{\frac{da}{dB}\}=-(\frac{.5}{g^2})(W_1-W_2)(W_1+W_2)\exp\left[\frac{(W_1-W_2)}{g}\right].$$

and,

$$\operatorname{lm}\left\{\frac{d}{dB}(a+\frac{1}{a})\right\}$$

$$=-(.5/g^2)(W_1-W_2)(W_1+W_2)\exp\left[(W_1-W_2)/g\right]\left\{1-\exp\left[-2(W_1-W_2)/g\right]\right\}.$$

$$\text{(C)}\quad \frac{d}{dB}(b+1/b)=db/dB\cdot[1-1/b^2].$$

Since $b=a\left[(BW_2+g)/(BW_1+g)\right]$,

$$db/dB=(da/dB)\left[(BW_2+g)/(BW_1+g)\right]-a\cdot\left[g(W_1-W_2)/(BW_1+g)^2\right].$$

Thus,

$$\text{Im}\{db/dB\}=\text{Im}\{da/dB\}-\left[(W_1-W_2)/g\right]\cdot\exp\left[(W_1-W_2)/g\right],$$

and,

$$\text{Im}\left\{\frac{d}{dB}(b+1/b)\right\}=$$

$$\left\{-.5\left[(W_1-W_2)(W_1+W_2)/g^2\right]\cdot\exp\left[(W_1-W_2)/g\right]\right.$$

$$\left.-\left[(W_1-W_2)/g\right]\cdot\exp\left[(W_1-W_2)/g\right]\right\}\cdot\left[1-\exp[-2(W_1-W_2)/g]\right].$$

Using (A), (B) and (C),

$$\text{Im}\left\{\frac{\left[(B-1)(a+1/a)+(b+1/b)-2B\right]}{B}\right\}=$$

$$\exp[\frac{(W_1-W_2)}{g}]+\exp[-\frac{(W_1-W_2)}{g}]-\frac{(W_1-W_2)}{g}\cdot\exp[\frac{(W_1-W_2)}{g}]\left\{1-\exp[\frac{-2(W_1-W_2)}{g}]\right\}-2$$

$$=e^{(z)}+e^{(-z)}-z\left[e^{(z)}-e^{(-z)}\right]-2,$$

where $z=\frac{(W_1-W_2)}{g}\geq 0$. This expression approaches zero as $z\to 0$ and the derivative $-z\left[e^{(z)}+e^{(-z)}\right]$ is negative. This result means that $\mathrm{lm}\{\frac{Z}{B}\}$ is both finite and non-negative as $B\to 0$. Moreover, $q<L$, means $W_2<W_1$, which implies $z>0$, and this means Z/B is strictly positive.

CHAPTER FOUR

THE ECONOMICS OF REINSURANCE

4.1 Preliminaries

A reinsurance contract is a risk exchange between insurers. The insurer accepting the risk, in exchange for premium payment is called the reinsurer. The insurer transferring the risk, which it has obtained from its customers in exchange for premium payments, is called the ceding firm, direct insurer, or original underwriter. The following passage summarizes a common view of reinsurance as a mechanism to reduce the probability of excessively large losses.

> When an insurance company reinsures a part of its portfolio, it buys security and pays for it. The company will forego a part of its expected profits in order to reduce the possibility of inconvenient losses. (Borch [1961, pp. 35])

The implication of this statement, that insurers have a natural demand for reinsurance which supports the reinsurance market, is based on the implicit assumption that the relationship between an insurer and reinsurer is analogous to the relationship between an insured and insurer. This being the case, insurers are willing to pay more than the expected value of losses in order to diversify. This view, however, ignores the fact that insurers are not endowed with risk. If a unique market price exists for risk, under conditions of capital market equilibrium the reinsurance exchange will not enhance the market value of the insurer. In such a case, reinsurance is redundant because shareholders of an insurer hold diversified portfolios. This point has been made by Doherty and Tinic (1981). The assumption of shareholders with diversified portfolios is essentially equivalent to a risk neutral insurer, in which case, of course, no need exists for reinsurance. The present chapter allows the possibility that insurable risks are not completely diversifiable in the immediate transaction with an insurer. This possibility exists if the insurance market is imperfectly competitive. This market is modeled by assuming that managers of insurers underwrite in a risk averse manner. Owners have no reason to force risk neutral underwriting; in fact, this action serves to decrease the value of the firm.

The purpose of this chapter is to examine the features of insurance markets that enhance or restrict the use of reinsurance. If insurers are price takers, but have some market power (possibly the result of barriers to entry), this power is eroded and risks are completely diversifiable when the reinsurance market is perfectly competitive. In other words, expected profits attract reinsurers. Costs incurred by either insurers or reinsurers in the reinsurance transaction reduce the use of reinsurance, while costs incurred by insurers in the original market leave coverage provided by insurers themselves unchanged. The manager of an insurer with monopoly power takes reinsurance in preference to maintaining this coverage. At the same time, this action is not contrary to owners' interests. Market power in the reinsurance market restricts its use, while the manager of a monopsonistic insurer takes reinsurance when his or her risk aversion is greater than that of clients. Doing so increases the value of the firm. The use of reinsurance is then decreasing with the ratio of policy-holder to manager risk aversion coefficients.

The remainder of this chapter proceeds as follows: Section 4.2 considers the competitive reinsurance market; Section 4.3 considers transactions costs; and Section 4.4 considers imperfect competition.

4.2 Competitive Reinsurance Market

Individuals possess assets that are susceptible to partial loss. They have the option of complete or partial protection against loss of wealth by paying a fee, the premium, to an insurer. The insurer may also transfer a portion of the risk to a reinsurer by paying a premium. It is assumed that individuals, insurers and reinsurers maximize utility functions of the mean and variance of final wealth. The justification for the assumption of risk averse insurers is given in the following paragraph. The marginal rates of substitution between mean and variance are given by $-c/2, -d/2, -r/2$, where the parameters c,d,r refer to consumer, direct insurer, and reinsurer respectively.[25]

Let, W_c, W_d, W_r represent initial wealths of insured, insurer and reinsurer. Individuals choose the fraction Q of loss x they receive in compensation when such a loss occurs; actual compensation is Qx. Letting α be the premium per unit coverage (premium rate) and since Q is the partial coverage purchased, the premium[26] is αQ. Policyholder final wealth is[27]

$$W_c - x + Qx - \alpha Q.$$

Under the assumptions, the expected utility of final wealth is

$$W_c - \mu + Q\mu - \alpha Q - .5c\sigma^2(1-Q)^2, \tag{4.1}$$

where μ and σ^2 are the mean and the variance of the loss distribution. Maximizing yields the demand for partial coverage

---

[25] The only utility of wealth functions and probability distributions consistent with the assumptions are an exponential utility function with a normal loss distribution or a power utility function with a lognormal distribution. Analysis remains essentially unchanged with any loss distribution and utility functions of the HARA class. In particular, sharing rules remain linear (see for example Mossin (1973, pp. 113-117)), and the figures presented in this chapter are unchanged except that they become non-linear. An important advantage to the assumptions used in this chapter is that they lead to closed form solutions.

[26] The assumption that the premium is proportional to coverage is innocuous. Any monetary transfer that does not depend upon the loss does not affect Pareto Optimality. This formulation is similar to that used by Mayers and Smith (1983).

[27] To keep the analysis and the algebra of this chapter as simple as possible the fact that both insurers and policy-holders invest in both risk-free and risky assets is ignored.

$$Q_c(\alpha)=1-\frac{\alpha-\mu}{c\sigma^2}. \tag{4.2}$$

As expected, full coverage is taken if $\alpha=\mu$, the actuarially fair premium rate.

In the absence of opportunities for reinsurance, the final wealth of the insurer is

$$W_d-Qx+\alpha Q.$$

Assuming that the insurer is operated by a manager, the expected utility of final wealth is

$$W_d-Q\mu+\alpha Q-.5dQ^2\sigma^2.$$

Maximizing with respect to Q, yields the amount of coverage the manager would like to supply at premium rate α,

$$Q_d(\alpha)=\frac{\alpha-\mu}{d\sigma^2}.$$

Equating the demand for coverage $Q_c(\alpha)$, with the amount the insurer is willing to accept yields the equilibrium premium rate

$$\mu+\sigma^2(1/c+1/d)^{-1}.$$

Correspondingly, the partial coverage exchanged is $(1/d)(1/c+1/d)^{-1}\times 100$ per cent of the loss x. Substituting these results into the insurer's expected final wealth yields

$$W_d+\sigma^2\frac{(1/d)}{(1/c+1/d)^2},$$

which is the market value of the insurer. Even if the owners of the insurer hold diversified portfolios, they will not choose a manager who underwrites in a risk neutral manner. If they did so, the value of the firm would be W_d which is less than the above expression. In fact, without competitive restrictions owners will choose managers who underwrite as if they had the same risk aversion coefficient as policy-holders. In other words, with risk aversion coefficient d=c. Moreover, this action is equivalent to choosing a risk neutral manager who acts as a monopolist in the setting of premium rates. To see this, note that a risk neutral monopolistic insurer maximizes $W_d+(\alpha-\mu)Q_c(\alpha)$ with respect to the premium rate. Managers modelled as risk averse, with $0\leq d\leq c$, may be interpreted to act with market power to the extent permitted by competition in the insurance market. Because competition forces premium rates downwards, it is not in owners' interests for managers to underwrite as if they were more risk averse than policy-holders.

It is assumed that insurers cede a proportion M of under-written risks to a reinsurer in exchange for a premium of λM, where λ is the premium rate. A manager's expected utility is,[28]

$$W_d+(\alpha-\mu)Q-(\lambda-\mu)M-.5d(Q-M)^2\sigma^2. \tag{4.3}$$

The manager's problem is to maximize expected utility with respect to Q and M subject to the conditions

$$0\leq Q\leq 1,\ 0\leq M\leq 1.$$

Derivatives of (4.3) are

---

[28] It is assumed that reinsurers can transact with insurers only and cannot operate in the direct market. One of the reasons that the insurance and reinsurance markets may be considered separate is that the relation between an insurer and its reinsurer is very close. The insurer informs the reinsurer of underwriting details, policy-holder characteristics, and accounting data. An insurer is unlikely to be willing to supply such information to a competitor in the direct market. According to Bickelhaupt (1983, pp. 824) seventy five percent of reinsurance in the United States is with professional reinsurance companies. The international character of reinsurance also separates the direct and reinsurance markets geographically.

$$(\alpha-\mu)-d(Q-M)\sigma^2, \text{ and} \tag{4.4}$$

$$-(\lambda-\mu)+d(Q-M)\sigma^2. \tag{4.5}$$

These expressions reveal that (4.3) does not have an interior maximum if $\alpha \neq \lambda$. If an equilibrium is to exist, insurance and reinsurance transactions must take place on original terms, i.e., $\alpha=\lambda$. To see why this is the case, let $\alpha>\lambda$; an insurer will pass on as many and as much of contracts of the above form as possible. In so doing, expected final wealth can be made infinitely large. Of course, other insurers notice the possibility for profits and offer lower rates in the direct market to attract business from the first insurer. These profits are available until $\alpha=\lambda$. On the other hand, if $\alpha<\lambda$, insurers do not use the reinsurance market. For every unit of coverage they pass to reinsurers they lose money. Insurers are better off by reducing their reinsurance purchases and the coverage they offer in the direct market. Reinsurers must reduce their rates to that charged in the direct market if they are to attract any business. Together, these results imply $\alpha=\lambda$. In other words, if the insurer is a price taker in both the direct and reinsurance markets, reinsurance must take place on equal terms.

With the result that $\alpha=\lambda$, (4.4) and (4.5) collapse to one expression. Setting this expression equal to zero indicates that the net coverage the manager offers is

$$Q-M=\frac{\alpha-\mu}{d\sigma^2}. \tag{4.6}$$

This coverage is also that which is offered in the absence of reinsurance, Q_d. This result implies that for any premium rate, price taking insurers have no innate demand for reinsurance, but only a willingness to supply a given level of net coverage. Insurers are indifferent to the gross amount of coverage provided in the direct market or supplied from the reinsurance market. The implication is that insurers are not attracted to the reinsurance market as a supplier of security. With equal premium rates, an insurer does not gain or lose by taking reinsurance as long as it compensates this position with increased coverage in the direct market. This coverage is taken and given on equal terms, and therefore, does not affect either final wealth or utility. An insurer's indifference to reinsurance is partly based on the view that transactions in the reinsurance market have no effect on the equilibrium premium rate. However, if all insurers take reinsurance, they <u>will</u> affect this rate. Through this process reinsurers indirectly affect the competitiveness of the original market.

The above discussion indicates that even if managers of an insurer are risk averse, price taking behavior implies that they have no inherent demand for reinsurance. However, reinsurance transactions emerge because reinsurers can increase their expected wealth by inducing insurers to take reinsurance through the use of lesser rates than those that exist in the direct market. Consider managers of reinsurers with expected utilities of final wealth

$$W_r - M\mu + \lambda M - .5rM^2\sigma^2.$$

The reinsurance offered is then,

$$M_r = \frac{\lambda - \mu}{r\sigma^2}. \tag{4.7}$$

In a price taking market, an insurance transaction takes place when the demand for coverage equals the amount insurers are willing to supply and when the demand for reinsurance equals the amount reinsurers are willing to supply. With the introduction of reinsurers, coverage offered in the direct market is the net coverage of insurers plus the coverage of reinsurers. This sum is

$$(1/d+1/r)\frac{\alpha - \mu}{\sigma^2}. \tag{4.8}$$

Equating this expression with the demand for coverage yields the equilibrium premium rates[29]

[29] It is a straight-forward matter to introduce many policies into the insurer's portfolio. For example, if all individuals are identical and have identical uncorrelated risks,

$$\alpha=\lambda=\mu+\sigma^2(1/c+1/d+1/r)^{-1}. \tag{4.9}$$

Notice that these rates are increasing with respect to the reinsurer's risk aversion, and as the reinsurer's risk aversion becomes large, the premium rate approaches the level that would exist if no reinsurance were available. The implication is that reinsurers induce insurers to take coverage by offering reduced rates. The result is a reduction in the market price of insurable risks. The above discussion may be summarized as

Proposition 4.1: When managers of insurers are price takers, but perfect competition does not exist to force risk neutral underwriting, reinsurers will induce an active reinsurance market through their willingness to accept risks with positive expected profits.

In Figure 4, the curve 0a is the manager's marginal utility cost of coverage. In the absence of reinsurance α^0 is the equilibrium premium rate. The curve 0b is the horizontal sum of insurer and reinsurer's marginal utility costs of coverage. With reinsurance, the equilibrium premium rate is α^1.

The introduction of reinsurers is equivalent to the addition of competitors in the direct market. Since $\alpha^1>\mu$ expected profits can be made by reinsurers who wish to enter the market. The arguments can be repeated and insurers eventually transact with many reinsurers.[30] All transactions are on original terms and equilibrium premium rates are decreasing with the number of reinsurers. These rates are

---

and the insurer incurs a cost of zN^2 when transacting with N policy-holders, the optimal number of policies is $N^*=[(\alpha-\mu)Q-.5dQ^2\sigma^2]/(2z)$.

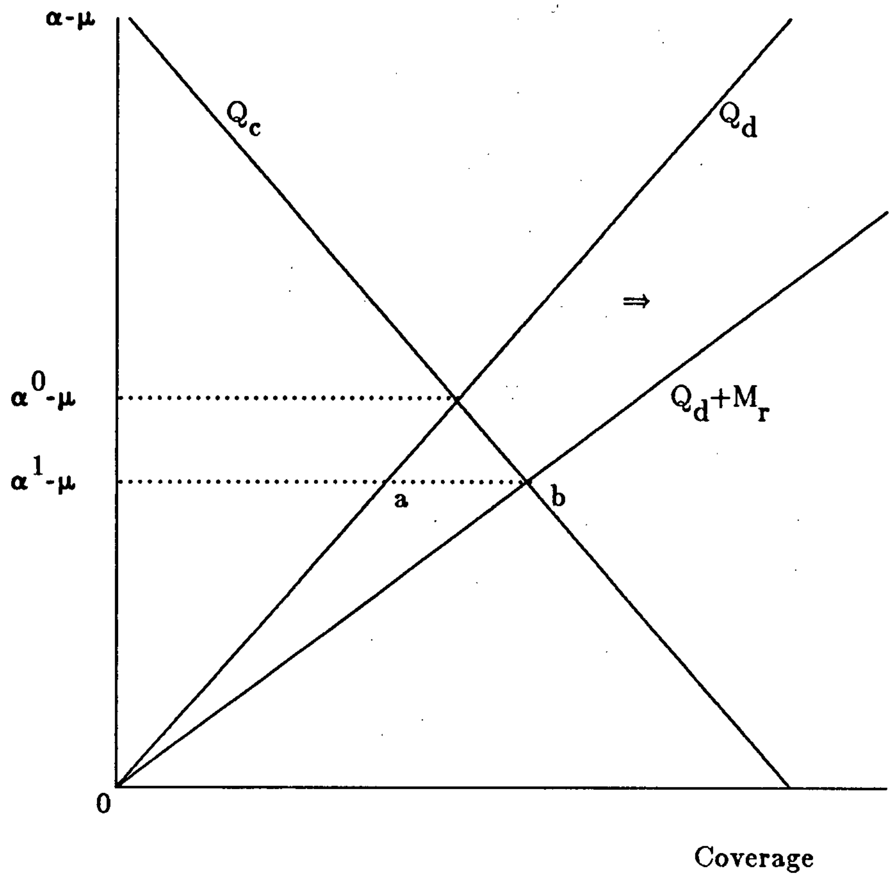
[30] This statement assumes that reinsurers do not transact with one another. However, exactly the same results are obtained if the first reinsurer passes risk to subsequent reinsurers. In fact, the same results are obtained if all reinsurers act as insurers in the direct market and consumers transact with more than one insurer. This more general statement indicates that in the absence of transactions costs, institutional structure has no effect on pure risk sharing.

$$\alpha=\lambda_i=\mu+\sigma^2[1/c+1/d+\sum(1/r_i)]^{-1},\ i=1,\ldots n,$$

where n is the number of reinsurers. Graphically this result means that the supply curve in Figure 4 becomes a horizontal line and intersects the demand for coverage curve at the point $Q=1$, $\alpha=\mu$. This finding can be interpreted to mean that reinsurers themselves provide the impetus for the reinsurance market. The result is increased competitiveness of insurance markets in general. The entire risk is spread among an insurer and reinsurers, each taking an infinitesimal share. This market may be thought of as perfectly competitive because of the infinitely elastic supply curve.

The potential demand for reinsurance is measured by the distance between the curves Q_c and Q_d below the excess (above actuarial) premium rate $\alpha^0-\mu$. This demand means that in the absence of perfect competition in the reinsurance market, managers of reinsurers are expected by owners to underwrite with risk aversion $r\leq(1/c+1/d)^{-1}$. Other factors being equal, managers of reinsurers underwrite in a less risk averse fashion than managers of a direct insurer.

Figure 4: A Price Taking Insurance Market



4.3 Administration Costs

This section examines the manner in which transactions costs restrict the use of reinsurance. Assume that both insurers and reinsurers incur administrative costs in the reinsurance market, while insurers also incur costs in the direct market. Let insurers' costs be proportional to Q and M and the ith reinsurer's costs be proportional to coverage supplied M_i, $(M=\sum M_i)$[31]. The expected utility of the manager of a direct insurer is

$$W_d - Q\mu + \sum M_i\mu + (\alpha - p)Q - \sum(\lambda_i + k)M_i - .5d\sigma^2(Q - \sum M_i)^2.$$

where p and k are costs per unit of direct and reinsurance coverage. The expected utility of the manager of the ith reinsurer is,[32]

$$W_i - M_i\mu + (\lambda_i - s)M_i - .5r_i(M_i)^2\sigma^2$$

where s is the cost per unit of coverage.

From the perspective of insurers, $\alpha - p$ and the terms $\lambda_i + k$ play the same role as α and λ in the previous section. If an equilibrium is to exist

$$\alpha - p = \lambda_i + k, \; i=1,...n, \tag{4.10}$$

where n is the number of reinsurers. This expression yields a fundamental relation

---

[31] These assumptions are chosen, in large part, for analytic tractability. If costs depend upon the total premium or indemnity, it is shown in Chapter Two of this thesis that indemnity schedules with deductible properties are appropriate.

[32] The same results are obtained if a first reinsurer passes some of the risk to a subsequent reinsurer who again passes some of this on, etc., and if a receiving reinsurer pays to a ceding reinsurer an "over-riding" commission. That is, if a ceding reinsurer's costs per unit coverage are s, the receiving reinsurer absorbs some of these costs by reducing its premium rate by s per unit coverage.

between the direct and reinsurance premium rates. The direct premium rate, net of transactions costs, equals the gross premium rate of reinsurers. This equality implies that all transactions costs reinsurers pass on to insurers are passed on to the direct market. In addition to these costs, insurers pass on portions of their own transaction costs.

Equilibrium premium rates are[33]

$$\alpha=\mu+\sigma^2 z^{-1}+s\sum(1/r_i)z^{-1}+p(1/d+\sum 1/r_i)z^{-1}+k\sum(1/r_i)z^{-1}, \tag{4.11}$$

$$\lambda_i=\mu+\sigma^2 z^{-1}+s\sum(1/r_i)z^{-1}-(p/c)z^{-1}-k(1/c+1/d)z^{-1},\ i=1,2...n, \tag{4.12}$$

where $z=[1/c+1/d+\sum 1/r_i]$. These equations indicate that if administrative costs are borne solely by reinsurers, reinsurance takes place on the original terms. Premium rates increase with reinsurers' transactions costs because they decrease reinsurers' willingness to offer coverage. Reinsurers require an increase in premium for every unit of coverage held. This increase is transferred dollar for dollar to the direct market; this transfer is implied by relation (4.10). On the other hand, if administrative costs are incurred by insurers in the original or the reinsurance transaction, the direct premium rate increases, while reinsurance rates decrease. Rates increase in the direct market because insurers pass on portions of costs to those insured; they decrease in the

[33] The introduction of administrative costs raises the possibility that these costs might be so large that reinsurance and even insurance transactions might not be possible. The minimum premium rate that is acceptable to reinsurers is $\mu+s$. Therefore, reinsurers are induced into the market only when the equilibrium reinsurance rate (4.12) is greater than this amount, i.e., $\mu+\sigma^2 z^{-1}+q(\sum 1/r_i)z^{-1}-p/cz^{-1}-k(1/c+1/d)z^{-1}\geq\mu+s$. Rearranging, $s+p+k\leq(\sigma^2+p/d)[1/c+1/d]^{-1}$. Since this inequality is independent of the number of reinsurers, price competition does not force insurers into the reinsurance market if it does not hold. This condition insures a positive demand for reinsurance and a positive transaction in the direct market, since if $k=s=0$, it reduces to the condition for the demand for coverage to be positive at an insurer's reservation premium rate $\mu+p$.

reinsurance market because otherwise the direct premium rate, net of costs, is less than the premium rate in the reinsurance market. This situation reduces the attractiveness of reinsurance to insurers. Reinsurers must, as a result, adjust rates downwards to maintain competitiveness. This adjustment implies that portions of insurer costs are assumed in both the direct and reinsurance market.

Proposition 4.2: Transactions costs incurred by insurers or reinsurers in the reinsurance exchange reduce the use of reinsurance. This reduction is also true for transaction costs incurred by insurers in the original exchange, but coverage provided by insurers themselves is unchanged.

Proof: If the reinsurance market is perfectly competitive, the equilibrium direct premium rate is $\mu+p+k+s$, and the reinsurance rate is $\mu+s$. Coverage in the direct market is $1-(p+k+s)/(\sigma^2 c)$. The portion of this coverage supplied from the reinsurance market is $1-(k+s)(1/c+1/d)/\sigma^2-p/(\sigma^2 c)$, while the portion supplied from the direct market is $(k+s)/(\sigma^2 d)$. The fact that coverage in the reinsurance market decreases with all transactions costs while coverage held by the insurer is invariant to direct market costs, establishes the result. □

A perfectly competitive reinsurance market is equivalent to risk neutrality. This neutrality means that with no reinsurance transactions costs, all coverage offered in the direct market is passed on to the reinsurance market. This transfer means that the effect on direct market coverage of insurers' direct market costs is totally absorbed in the reinsurance market. This absorption is illustrated by the insensitivity of insurers' net coverage to per unit costs in the direct market, p. When costs in the reinsurance market increase, reinsurance becomes expensive relative to direct insurance. Therefore, the reinsurance market loses coverage not only because of reaction to price increases from consumers in the direct market, but also because of substitution of direct for reinsurance coverage. This statement explains why the net amount of coverage held by insurers is increasing in the insurer and reinsurer per unit costs of reinsurance coverage k and s.

4.4 Imperfect Competition

This section examines the manner in which imperfectly competitive insurance markets[34] restrict the use of reinsurance. It is assumed that four levels of competitive behavior exist. An insurance market is perfectly competitive if all participants are price takers and no barriers to entry exist. In this case managers of insurers underwrite in a risk neutral manner. A market is competitive as opposed to perfectly competitive if all participants are price-takers, but barriers to entry exist. These barriers might take the form of capitalization requirements by regulatory authorities or large fixed costs associated with developing a clientele and distribution network. In such a market insurers make positive expected profits because insurers choose managers who underwrite in a risk averse manner. Finally, an insurer has market power if it can make pricing decisions on the basis of knowledge of a contracting party's demand for coverage or willingness to supply reinsurance coverage or, in combination with this, can also choose reinsurers on the basis of risk preference.

4.4.1 *Competitive Insurance Market, Monopolistic Reinsurer*

Assume that a reinsurer has market power on the basis of knowledge of the demand for reinsurance that is forthcoming from the direct market at different premium rates, but insurers remain price takers in both the direct and reinsurance markets.[35] In this case, the principle of reinsurance on original terms remains valid. As in

[34] The extent of competition in the insurance industry is examined by Quirin and Waters (1974). For purposes of this chapter it is sufficient to note that many features appear competitive while others do not.

[35] The terms monopolist and monopsonist are meant to include the notions of monop-

Section 2, insurers have no inherent demand for coverage. To find the equilibrium premium rate in this market, substitute the individual's demand for coverage Q_c into (4.6) and solve to obtain the excess demand for coverage from the direct market

$$M_d = 1 - (1/d + 1/c)\frac{\alpha - \mu}{\sigma^2}. \tag{4.13}$$

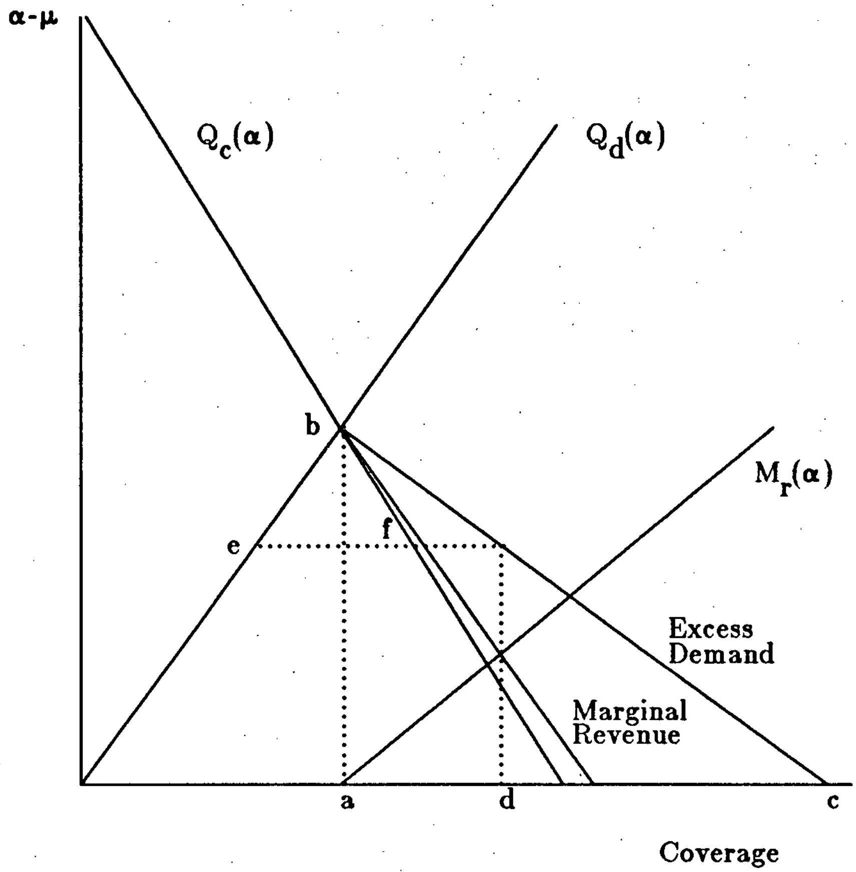
This expression is the demand for coverage in the direct market above that which the insurer wishes to hold on its own account at the premium rate α. It may be interpreted to be the demand for reinsurance. Graphically, this demand is the distance between the curves $Q_d(\alpha)$ and $Q_c(\alpha)$ in Figure 4. It is interesting to note that the elasticity of this excess demand for coverage curve is always greater than that of the original direct demand curve. Reinsurers face two responses to an increase in price; the direct market reacts to prices passed on by insurers, and insurers themselves react by increasing coverage held on account. The excess demand for coverage is represented in Figure 5 by the line bc and is measured from the point a. A line that divides this triangle in half gives the monopolistic reinsurer's marginal revenue curve. The intersection with the reinsurer's marginal cost curve gives both the price that will be charged, and measuring from point "a", the amount of reinsurance. This reinsurance coverage is the distance ad measured on the excess demand curve and the distance ef measured between the curves $Q_d(\alpha)$ and $Q_c(\alpha)$. The premium rate in both the direct and reinsurance market is less than the equilibrium premium rate without reinsurance. The introduction of a reinsurer, even a monopolistic reinsurer, increases effective competition in the direct market.

---

olistic competition and monopsonistic competition respectively. Expected profits do not attract entry to the industry because of financial or other barriers. Spellman, Witt and Rentz (1975) and McCabe and Witt (1977) consider insurance markets under conditions of monopolistic competition.

In contrast to the perfectly competitive market the risk is not completely spread amongst insurers and reinsurers. In order to maximize profits, it is in the interest of the reinsurer to restrict the use of reinsurance.

Figure 5: Monopolistic Reinsurer



4.4.2 *Monopolistic Insurer*

Assume that an insurer has market power on the basis of knowledge of policy holders' demand for coverage (4.2). In the absence of reinsurance, the manager will maximize expected utility

$$W_d+(\alpha-\mu)Q_c(\alpha)-.5dQ_c(\alpha)^2\sigma^2.$$

with respect to the premium rate. The result is the premium rate $\mu+\sigma^2(c+d)/(2+d/c)$, and the market value of the firm is

$$W_d+\sigma^2\frac{c+d}{(2+d/c)^2}.$$

This value is strictly decreasing with the risk aversion of the manager. If the manager can apply monopoly power on the basis of knowledge of policy-holders' demands for coverage, the value of the firm is decreased if the manager also underwrites risks in the risk averse fashion.

The remainder of this sub-section considers the reinsurance market when the manager has the opportunity to underwrite risks in the risk averse manner, in other words, without absolute parity to owner's interests. Later it is seen that if the reinsurance market is sufficiently competitive, the manager can act with absolute parity to owners' interests and will at the same time take reinsurance.

If the manager of the direct insurer is a price taker in the reinsurance market, first order conditions for maximization of expected utility with respect to α and M are

$$Q(\alpha)+(\alpha-\mu)Q'(\alpha)-d[Q(\alpha)-M]Q'(\alpha)\sigma^2=0, \tag{4.14}$$

$$(\lambda-\mu)-d[Q(\alpha)-M]\sigma^2=0. \tag{4.15}$$

Notice that equation (4.15) implies that the monopolistic insurer holds net coverage $(\lambda-\mu)/(d\sigma^2)$. Insurers hold units of coverage on their own account as long as the marginal cost of doing so (the curve $Q_d(\alpha)$ may be interpreted to be the manager's marginal cost curve), is less than the reinsurance premium rate. This coverage is up to point "a" in Figure 6. Any units of coverage offered in the consumer market beyond this amount are passed on as reinsurance. The monopolist charges a higher rate in the direct market than λ in order to take advantage of the fixed price in the reinsurance market. It acts as a risk neutral monopolist with respect to the setting of this price since it receives the profit $(\alpha-\lambda)Q_c(\alpha)$ with certainty. Maximizing with respect to α indicates that the relation between premium rates is

$$(\alpha-\mu)=.5\sigma^2c+.5(\lambda-\mu). \tag{4.16}$$

This relation can also be obtained by substituting (4.15) into (4.14). The relationship between α and λ is displayed graphically in Figure 6. Point b gives the gross amount of coverage; point a is the monopolist's net coverage. The difference is the demand for reinsurance coverage when the premium rate charged is λ. This coverage is

$$M_d(\lambda)=.5-(1/d+.5/c)\frac{\lambda-\mu}{\sigma^2}. \tag{4.17}$$

This expression is found by substituting (4.16) into (4.15) or by appropriately measuring the distance between points a and b in the figure.

The reinsurance supply curve is drawn from point a in Figure 6. Demand for and supply of reinsurance are equated at $\lambda^1-\mu$. Using distances on the figure and appropriate equations or by simply using equations (4.17), (4.7) and subsequently (4.16), equilibrium premiums rates are,

$$\alpha=\mu+\frac{\sigma^2[1+c/r+c/d]}{[2/r+2/d+1/c]}, \tag{4.18}$$

$$\lambda=\mu+\frac{\sigma^2}{[2/r+2/d+1/c]}. \tag{4.19}$$

The direct rate is greater than the reinsurance rate because of the insurer's monopoly power. The sum of the areas ocg and gdef give the manager's expected utility. The triangle acd gives the reinsurer's expected utility.

Proposition 4.3: A manager of an insurer who exerts monopoly power takes reinsurance coverage.

Proof: The manager's expected utility upon transacting with a reinsurer is

$$W_d+\sigma^2\frac{[c/r^2+2c/(dr)+.5/d+c/d^2]}{[2/r+2/d+1/c]^2}.$$

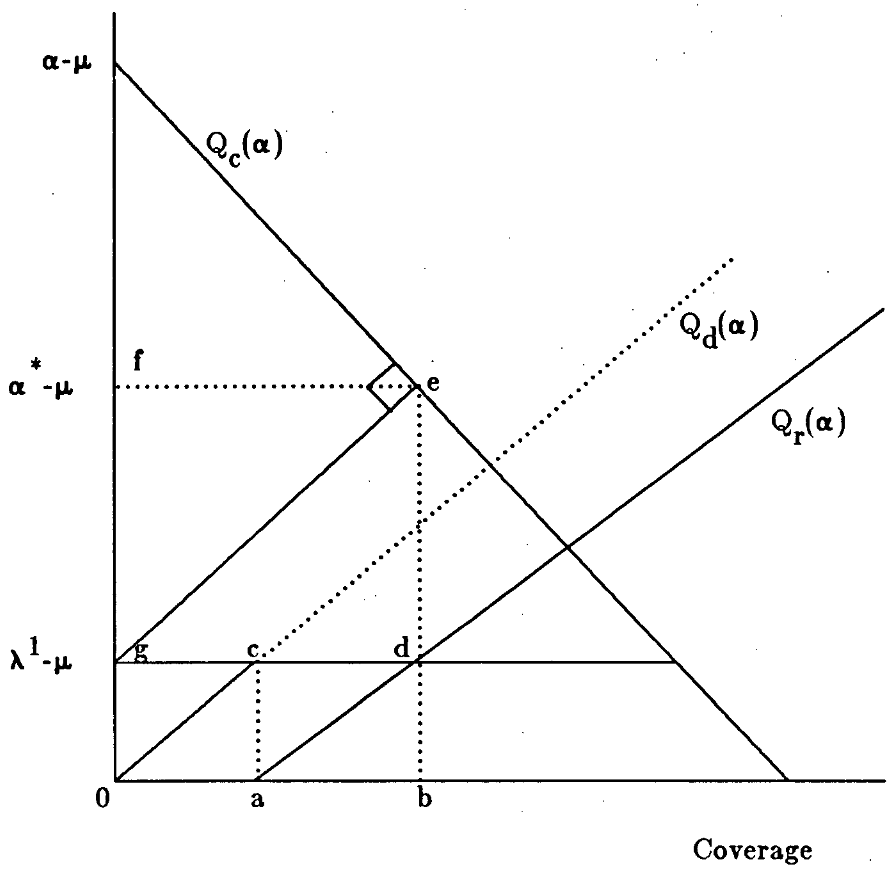
The limit of this expected utility as r becomes large is the same as the utility the manager would obtain with no reinsurance. Moreover, this expression is decreasing in r. □

The fact that the expected utility of the direct market manager is decreasing in the reinsurer's risk aversion illustrates that it is in the manager's interest to promote competition in the reinsurance market. Utility is greatest when the reinsurance market is perfectly competitive. When the risk aversion of reinsurers is large, the rates (4.18) and (4.19) reduce to the direct rate that a monopolist charges without reinsurance and the reinsurance rate that induces zero reinsurance demand, (see equation 4.17). If reinsurers have a zero risk aversion coefficient, the above rates reduce to $\mu+.5\sigma^2c$ and μ respectively; the direct rate is the same as that charged by a risk neutral monopolist, which is the rate that maximizes expected profit, $(\alpha-\mu)Q_c(\alpha)$. In this case the insurer takes and passes on the coverage 0.5. Measuring "price" as $\alpha-\mu$ it charges a rate for which the demand curve has unit elasticity. This position is extremely attractive for the monopolist. It receives the largest expected profit possible and does not participate in any claim payment; in other words, the insurer simply acts as a broker. This situation is also attractive for the manager who can act with absolute parity to the owners' interests and at the same time avoid loss distributions to which he or she is risk averse. The risk aversion of a monopolistic manager is irrelevant to owners when the reinsurance market is perfectly competitive.

In Section 2, insurers were assumed to be price-takers, but also possessed market power that was measured by the extent to which manager's risk aversion was greater than zero. In the current subsection, market power is more pronounced as evidenced by the fact that a competitive reinsurance market does not dissipate the competitive imperfections in the direct market. In this case the direct market insurer substitutes reinsurance for coverage held on its own account and makes a profit on the divergence between the direct market rate and the reinsurance premium rate.

Because of costs involved with setting up distribution networks and becoming familiar with local market conditions and government regulations, reinsurers rely heavily on direct insurers and brokers for the development of their underwriting portfolios. The above analysis illustrates the nature of the market power that can emerge as a result.

Figure 6: Monopolistic Insurer



4.4.3 *Monopsonistic Purchase of Reinsurance*

Suppose that the insurer is monopsonistic[36] on the basis of knowledge of reinsurers' willingness to hold coverage. If insurers remain price takers in the direct market, a manager maximizes utility with respect to Q and λ. The first order conditions are

$$\alpha-\mu-d(Q-M)\sigma^2=0, \tag{4.20}$$

$$-M-(\lambda-\mu)M'+d(Q-M)\sigma^2M'=0, \tag{4.21}$$

where M' is the derivative of the reinsurance supply function (4.7) with respect to λ. The first order condition with respect to Q implies that insurers' net supply of coverage in the direct market is the same as that without reinsurance.

Using the reinsurance supply function (4.7) and the first order conditions (4.20) and (4.21) yields

$$(\alpha-\mu)=2(\lambda-\mu). \tag{4.22}$$

This equation reflects the fact that insurers are attracted to the reinsurance market because they can force the purchase price of reinsurance coverage below the price at which they sell coverage in the direct market. Using this relationship and equations (4.7) and (4.20), the gross supply of coverage is

[36] The North American agency system of marketing insurance gives agents a strong bargaining position relative to insurers. Property rights associated with renewals have generally been granted to agents by court decisions. This right means that an insurer cannot direct bill its policy-holders to secure renewal business. In this chapter insurers market policies directly, but possess the same bargaining position relative to reinsurers. This position may be taken as one of the sources of market power in the present section.

$$\frac{(1/d+.5/r)(\alpha-\mu)}{\sigma^2}. \tag{4.23}$$

This supply never exceeds (4.8), the gross supply when all parties are price takers. This result means that in equilibrium, monopsonistic insurers take less reinsurance than insurers who are price takers in the reinsurance market. Notice also that (4.23) implies that a monopsonistic insurer acts as if it were a competitive insurer with risk aversion $(1/d+.5/r)^{-1}$.

Equilibrium occurs in the direct market when the demand for coverage equals the gross supply. Equating (4.2) with (4.23) and subsequently using (4.22) yields

$$\alpha=\mu+\frac{2\sigma^2}{2/d+2/c+1/r}, \tag{4.24}$$

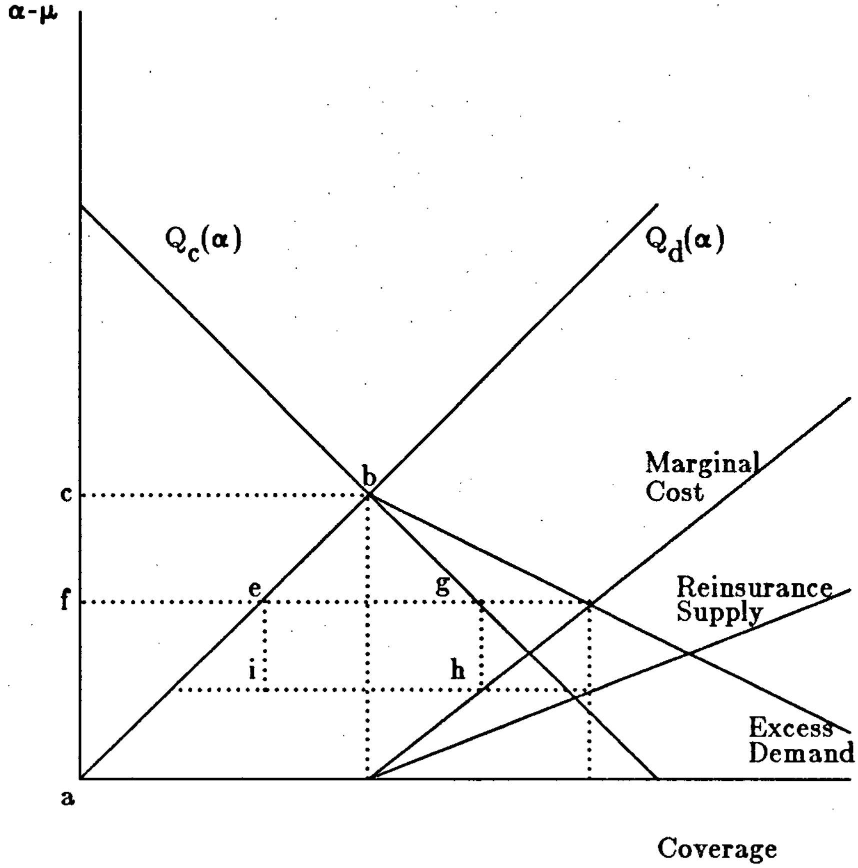
$$\lambda=\mu+\frac{\sigma^2}{2/d+2/c+1/r}. \tag{4.25}$$

The direct rate is greater than the reinsurance rate because of market power. Equilibria with and without reinsurance are illustrated in Figure 7. The manager's surplus without reinsurance is the triangle abc, whereas the surplus with reinsurance is the triangle aef. Insurers also receive a welfare gain in the reinsurance market because they supply coverage at a greater premium rate than they receive it from reinsurers. This gain is represented by the rectangle eghi. The manager's expected utility is higher with reinsurance when these two areas exceed the former.

Notice that both premium rates (4.24) and (4.25) approach the actuarial rate when reinsurer risk aversion is low. In other words, if the reinsurance market is perfectly competitive, expected profits in the direct market are completely eroded. In this case, direct market insurers' attempts to take advantage of monopsony power in the reinsurance market enhance the competitiveness of the direct market. An important condition for this result to occur is that insurers be price takers in the direct market.

In the remainder of this section, an expanded form of monopsony power is considered. It is assumed that managers of insurers have the ability to choose reinsurers with desirable risk aversion properties. They do so in order to insulate the domestic market from the effects of the type of indirect competition just considered. In other words, managers might jointly wish to restrict reinsurers' access to the domestic market. The extent to which access is restricted can be measured by reinsurer risk aversion. This restriction is equivalent to managers setting a minimum rate at which reinsurance transactions must take place. A reinsurer with a zero risk aversion coefficient is equivalent to a perfectly competitive reinsurance market and unobstructed access to transactions with domestic insurers. In this case reinsurers charge an actuarially fair rate. Insurers expand sales of coverage in the direct market because the direct rate is greater than the reinsurance rate. This expansion continues until the direct market rate is also actuarially fair; policy-holders obtain full coverage, all of which is passed on as reinsurance. This process completely erodes the expected profits of insurers. At the opposite extreme a reinsurer with an infinitely large risk aversion is equivalent to a ban by managers on reinsurance transactions. Proposition 4.4 indicates that managers prohibit reinsurance when their risk aversion is less than that of clients. Any access by reinsurers reduces managers' expected utility. On the other hand, when managers' risk aversion is greater than that of clients, by choosing a reinsurer with risk aversion sufficiently great, or equivalently by setting a minimum reinsurance premium rate sufficiently high, it is always possible for managers of insurers to increase expected utility. Moreover, with complete discretion over choice of reinsurers they effectively act as risk neutral monopolists in the direct market. Of course, the extent to which managers could maintain this position depends up on their joint ability to enforce reinsurance restrictions.

Figure 7: Monopsonistic Insurer



Proposition 4.4: Managers with market power in the reinsurance market purchase reinsurance when their risk aversion is greater than that of policy-holders. When managers jointly permit reinsurance they do so to maximize market power in the direct market.

Proof: The manager's expected utility upon transacting with a reinsurer with risk aversion r is,

$$W_d + \frac{\sigma^2(2/d+1/r)}{(2/d+2/c+1/r)^2}.$$

The limit of this expression as r becomes large equals the manager's expected utility in the absence of the reinsurance transaction. It is strictly increasing in r when $d \leq c$. The manager does not take reinsurance if his or her risk aversion is less than that of policy-holders. On the other hand, if the manager is more risk averse than consumers, $d > c$, it is always possible to transact with a reinsurer who increases expected utility. A manager's expected utility with reinsurance is greater than his or her expected utility without reinsurance when $r \geq (2d/c^2 - 2/d)^{-1} > 0$. Expected utility is maximized when $r = (2/c - 2/d)^{-1} > 0$. For this value of r, the manager's expected utility is $W_d + c\sigma^2/8$ and the price charged in the direct market is $\mu + .5\sigma^2 c$. This price is the same as that charged by a risk neutral monopolist in the direct market. □

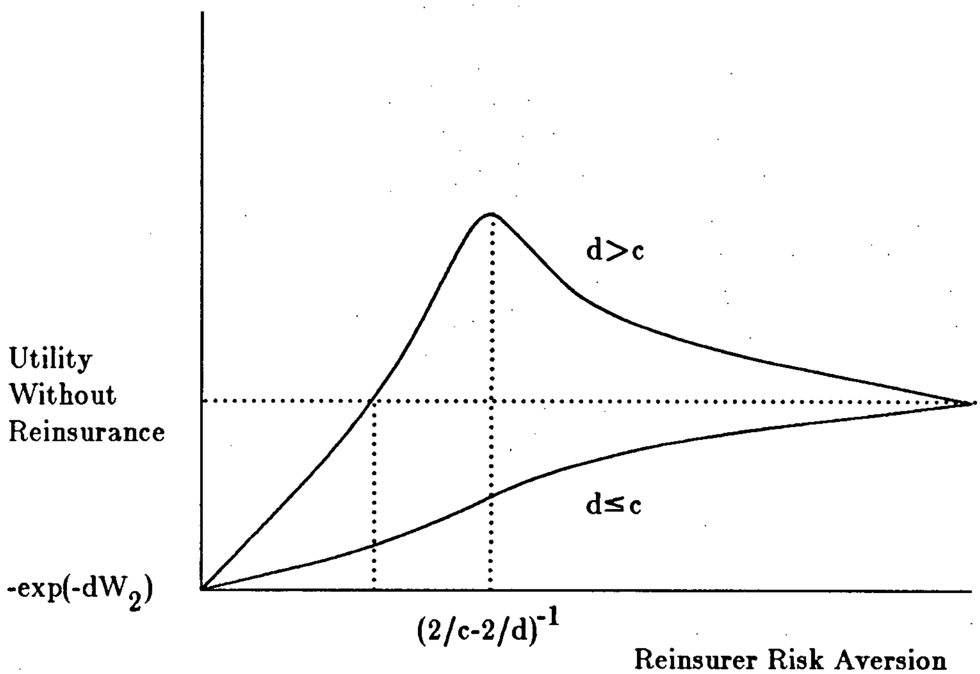
When managers jointly choose reinsurance to maximize utility, coverage provided in the direct market is 50 percent of the loss. Also, $.5c/d \times 100$ percent of the total risk is held by the insurer while $.5(1-c/d)$ is passed on as reinsurance. Figure 8 illustrates the manager's expected utility as a function of the risk aversion of the reinsurer.

Since one would expect business firms to be less risk averse than individuals, one may interpret Proposition 4, to imply that, ceteris paribus, reinsurance should be more prevalent in commercial rather than personal lines of insurance. Proposition 4.4 indicates that when managers are more risk averse than policy-holders, managers' expected utilities are increased by taking reinsurance, and in particular are maximized by taking reinsurance from a reinsurer with risk aversion coefficient $r=(2/c-2/d)^{-1}$. Moreover the market value of the insurer is also increased. To see this, note that the value of the firm is

$$W_d+\frac{\sigma^2(4/d+1/r)}{(2/d+2/c+1/r)^2},$$

and that when a manager maximizes utility by choosing a reinsurer with risk aversion $r=(2/c-2/d)^{-1}$, this expression is greater than $W_d+(.25)c\sigma^2$, the market value of a risk neutral monopolistic insurer. In this case monopsony expected profits exceed the loss in expected profits from decreased rates in the direct market.

Figure 8: Monopsonistic Expected Utility



4.5 Summary

The purpose of this chapter is to investigate the features of insurance markets that restrict or enhance the use of reinsurance. Contrary to the view of Doherty and Tinic (1981) that capital market equilibrium negates the need for reinsurance, it is shown that active reinsurance markets emerge when the direct market is imperfectly competitive. In the model studied, imperfect competition is a sufficient condition for reinsurance. The extent to which managers of direct insurers utilize reinsurance is examined. Two important factors determining this utilization are transactions costs and the degree of market power in both the direct market and the reinsurance market exerted by either the direct insurer or the reinsurer.

BIBLIOGRAPHY

Dan R. Anderson, "Effects of Under and Over_evaluations in Loss Reserves," *Journal of Risk and Insurance,* 1971, 38, 585-600.

John Anderson and Howard Thompson, "Financial Implications of Over-Reserving in Non-Life Insurance Companies," *Journal of Risk and Insurance,* 1971, 38, 333-342.

R.J. Arnott, and J.E. Stiglitz, "Equilibrium in Competitive Insurance Markets: The Welfare Economics of Moral Hazard," Discussiion Paper, Institute for Economic Research, Queen's University, 1982.

-----. "Moral Hazard and Optimal Commodity Taxation," Discussion Paper, Institute for Economic Research, Queen's University, 1983.

Kenneth J. Arrow, "Uncertainty and the Welfare Economics of Medical Care," *American Economic Review,* 1963, 53, 941-973.

Raphel J. Balcarek, "Effect of Loss Reserve Margins in Calendar Year Results," *Proceedings of the Casualty Actuarial Society,* 1966, 53, 1-16.

David L. Bickelhaupt, *General Insurance,* Homewood,Illinois: Richard D. Irwin Inc., 1983.

Karl Borch, "The Safety Loading of Reinsurance Premiums," *Skandinavisk Aktuaraietidskift,* 1960, 163-184.

-----. "Some Elements of a Theory of Reinsurance," *The Journal of Insurance,* September 1961, 35-43.

-----. *The Mathematical Theory of Insurance,* London: Lexington Books, 1974.

M. Boyer and G. Dionne, "Variations in the Probability and Magnitude of Loss: Their Impact on Risk," *Canadian Journal of Economics,* August 1983, 16, 411-19.

Michael J. Brennan and Ray Solanki, "Optimal Portfolio Insurance," *Journal of Financial and Quantitative Analysis*, 1981, 16, 279-290.

Robert L. Carter, *Reinsurance*, London: Kluwer Publishing, 1979.

Robert W. Crandall and John D. Graham, "Automobile Safety Regulation and Offsetting Behavior: Some New Empirical Estimates," *The American Economic Review Proceedings,* 1984, 74, 328-331.

J. David Cummins and David J. Nye, "Portfolio Optimization Models For Property-Liability Insurance Companies: An Analysis and Some Extensions," *Management Science,* 1981, 27, 414-430.

G. Dionne, "Moral Hazard and Search Activity," *Journal of Risk and Insurance,* September 1981, 48, 422-434.

-----. "Moral Hazard and State Dependent Utilities," *Journal of Risk and Insurance,* September 1982, 49, 405-22.

N.A. Doherty and S.M. Tinic, "A Note on Reinsurance Under Conditions of Capital Market Equilibrium," *Journal of Finance,* September 1981, 949-53.

N.A. Doherty and H. Schlesinger, "Optimal Insurance in Incomplete Markets," *Journal of Political Economy,* March 1983, 1045-54.

I. Erhlich and G. Becker, "Market Insurance, Self-Insurance, And Self-Protection," *Journal of Political Ecomomy,* 1972, 80, 623-48.

William B. Fairley, "Investment Income and Profit Margins in Property Liability Insurance: Theory and Empirical Results," *Bell Journal of Economics,* 1979, 10, 192-210.

John P. Gould, "The Expected Utility Hypothesis and the Selection of Optimal Deductibles for a Given Insurance Policy," *Journal of Business,* 1969, 42, 143-151.

Sanford Grossman and Oliver Hart, "An Analysis of the Principal-Agent Problem," *Econometrica,* 1983, 51, 7-46

Milton Harris and Arthur Raviv, "Some Results on Incentive Contracts with Application to Education and Employment, Health Insurance, and Law Enforcement," *American Economic Review,* 1978, 68, 20-30.

George H. Head and Michael L. Smith, "Guidelines for Insurers in Pricing Deductibles," *Journal of Risk and Insurance* , 1978, 45, 217-238.

Raymond D. Hill, "Profit Regulation in Property-Liability Insurance," *Bell Journal of Economics,* 1979, 10, 172-191.

Bengt Holmstrom, "Moral Hazard and Observability," *Bell Journal of Economics,* 1979, 10, 74-91.

Gur Huberman, David Mayers, and Clifford W. Smith, "Optimal Insurance Policy Indemnity Schedules," *Bell Journal of Economics,* 1983, 14, 415-426.

Michael D. Intrilligator, *Mathematical Optimization and Economic Theory* , Englewood Cliffs, N.J.: Prentice Hall, 1971.

G.D. Jaynes, "Equilibria in Monopolistically Competitive Insurance Markets," *Journal of Economic Theory,* December 1978, 17, 394-422.

Yehuda Kahane, "Ratemaking and Regulation in Property-Liability Insurance," *Quarterly Review of Economics and Business,* 1977, 17, 97-112.

-----, "The Theory of Insurance Risk Premiums - A Re-Examination in the Light of Recent Developments in Capital Market Theory," *The ASTIN Bulletin,* 1978, 10, 223-239.

R. E. Kilstrom, D.R. Romer and S. Williams, "Risk Aversion with Random Initial Wealth," *Econometrica,* July 1981, 49, 911-20.

Alan Kraus, and Steven Ross, "The Determination of Fair Profits for the Property Liability Insurance Firm," *Journal of Finance,* 1982, 37, 1015-1028.

J.J. Laffont, *Essays in the Economics of Uncertainty,* Cambridge, Massachusetts: Harvard University Press, 1980.

Gerald I. Lenrow, Ralph Milo, and Anthony P. Rua, *Federal Income Taxation of Insurance Companies,* New York: John Wiley, 1979.

Brian G.M. Main, "Business Insurance and Large Widely-Held Corporations," *The Geneva Papers on Risk and Insurance* , July 1982,7, 248-57.

O.L. Mangasarian, "Convexity, Pseudo-Convexity and Quasi-Convexity of Composite Functions," *Cahier du Centre d'E'tudes de Recherche Ope'rationelle* , 1970,12, 114-122.

David Mayers, and Clifford W. Smith, "Contractual Provisions, Organizational Structure, and Conflict Control in Insurance Markets," *Journal of Business* , 1981,54, 407-434.

-----. "On the Corporate Demand for Insurance," *Journal of Business,* 1982, 53, 281-296.

-----. "The Interdependence of Individual Portfolio Decisions and the Demand for Insurance," *Journal of Political Economy,* 1983, 91, 304-311.

George M. McCabe, and Robert C. Witt, "Insurer Optimizing Behaviour and Capital Market Equilibrium," *Journal of Risk and Insurance,* September 1977, 44, 447-467.

Dennis Moffet, "Optimal Deductibles and Consumption Theory," *Journal of Risk and Insurance,* 1977, 44, 669-682.

Jan Mossin, "Aspects of Rational Insurance Purchasing," *Journal of Political Economy* , 1968, 76, 533-568.

David J. Nye, "An Evaluation of Risk Classification Systems in Automobile Insurance." Report Prepared for the Florida Department of Insurance, 1979.

Peter B. Pashigian, Lawrence Schkade and George Menefee, "The Selection of an Optimal Deductible for a Given Insurance Policy," *Journal of Business*, 1966, 39, 35-44.

M. Pauly, "Overinsurance and Public Provision of Insurance," *Quarterly Journal of Economics,* February 1974, 88, 44-62.

Sam Peltzman, "The Effects of Automobile Regulation," *The Journal of Political Economy,* 1975, 83, 677-725.

J.W. Pratt, "Risk Aversion in the Small and in the Large," *Econometrica,* January 1964, 32, 122-36.

D.G. Quirin, P.J. Halpern, P.J. Kalymon, B.A. Mathewson, W.R. Waters, "Competition, Economic Efficiency and Profitability in the Canadian P-L Industry," Toronto: Insurance Bureau of Canada, 1974.

Arthur Raviv, "The Design of an Optimal Insurance Policy," *American Economic Review,* 1979, 69, 84-96.

M. Rothschild and J. Stiglitz, "Equilibrium in Competitive Insurance Markets: An Essay in the Economics of Imperfect Imformation," *Quarterly Journal of Economics,* 1976, 90, 629-49.

Harris Schlesinger, "The Optimal Level of Deductible in Insurance Contracts," *Journal of Risk and Insurance,* 1981, 48, 465-481.

S. Shavell, "On Moral Hazard and Insurance," *Quarterly Journal of Economics,* November 1979, 93, 514-62.

-----. "Risk Sharing and Incentives in the Principal Agent Relationship," *Bell Journal of Economics,* 1979, 10, 55-73.

Michael S. Smith and David L. Bickelhaupt, "Is Coinsurance Becoming Obsolete?" *Journal of Risk and Insurance,* 1981, 48, 95-110.

Vernon Smith, "Optimal Insurance Coverage," *Journal of Political Economy,* 1968, 76, 68-77.

Lewis J. Spellman, Robert C. Witt, and William Rentz, "Investment Income and Non-Life Insurance Pricing," *Journal of Risk and Insurance,* December 1975, 42, 567-577.

Michael Spence and Robert Zeckhauser, "Insurance Information and Individual Action," *American Economic Review,* 1971, 61, 380-387.

Joseph E. Stiglitz., "Monopoly, Non-Linear Pricing and Imperfect Information: The Insurance Market," *Review of Economic Studies,* 1977, 44, 407-30.

Robert M. Townsend, "Optimal Contracts and Competitive Markets with Costly State Verification," *Journal of Economic Theory,* 1979, 21, 265-293.

Stuart M. Turnbull, "Additional Aspects of Rational Insurance Purchasing," *Journal of Business,* 1983, 56, 217-229.

Kip Viscusi, The Lulling Effect: The Impact of Child-Resistant Packaging on Aspirin and Analgesic Ingestions," *The American Economic Review Proceeding,* 1984, 74, 324-327.

Bests Aggregates and Averages of the Property Casualty Industry, Morristown, N.J.: A.M. Best Company, 1982.