The Design of an Optimal Insurance Policy

By ARTUR RAVIV*

Almost every phase of economic behavior is affected by uncertainty. The economic system has adapted to uncertainty by developing methods that facilitate the reallocation of risk among individuals and firms. The most apparent and familiar form for shifting risks is the ordinary insurance policy. Previous insurance decision analyses can be divided into those in which the insurance policy was exogenously specified (see John Gould, Jan Mossin, and Vernon Smith), and those in which it was not (see Karl Borch, 1960, and Kenneth Arrow, 1971, 1973). In this paper, the pioneering work of Borch and Arrow—the derivation of the optimal insurance contract form from the model—is synthesized and extended.

The incentive to insure and insurance decisions have been treated extensively by Gould, Mossin, and Smith. They analyzed the problem of rational insurance purchasing from the point of view of an individual facing a specific risk, given his wealth level and preference structure. In their analysis the individual is offered an insurance policy specifying the payment to be received from the insurance company if a particular loss occurs. The individual may choose the level of the deductible, the level of the maximum limit of coverage, or the fraction of the total risk which is to be insured. Since the premium paid by the individual is directly related to the features chosen, the optimal insurance coverage involves balancing the effects of additional premium against the effects of additional coverage. In this approach the terms of the policy are assumed to be exogenously specified and are imposed on the insurance purchaser.

Borch (1960) was the first to take the more general approach of deriving the optimal insurance policy form endogenously. He sought to characterize a Pareto optimal risk-sharing arrangement in a situation where several risk averters were to bear a stochastic loss. This framework was then used by Arrow (1971) to obtain Pareto optimal policies in two distinct cases: 1) if the insurance seller is risk averse, the insured prefers a policy that involves some element of coinsurance; (i.e., the coverage will be some fraction (less than 1) of the loss); and 2) if the premium is based on the actuarial value of the policy plus a proportional loading (i.e., the insurer is risk neutral) and the insurance reimbursement is restricted to be nonnegative, the insurance policy will extend full coverage of losses above a deductible. Arrow (1973) extended this result to the case of state dependent utility functions. In this case, the optimality of a deductible which depends upon the state was proved. Robert Wilson also dealt with the endogenous determination of optimal risk-sharing arrangements, focusing on the incentive problem and the existence of surrogate functions. Consequently, constraints on the contract or costs associated with contracting were not included.

The purpose of this paper is to explain the prevalence of several different insurance contracts observable in the real world. The previous studies addressed this issue with a diversity of underlying assumptions and, therefore, the essential ingredients of the model that give rise to the insurance policy's various characteristics are not clear. For example, in Arrow's 1971 paper it is unclear whether an insurance policy with a deductible is the consequence of risk neutrality of the insurer, nonnegativity of the insurance coverage, or loading on the premium. Could a deductible be obtained when the insurer is risk averse? Could the loading be interpreted as risk premium? Could we explain the prevalence of deduct-

*Associate professor of economics, Carnegie-Mellon University and Tel-Aviv University. I would like to thank E. Green, M. Harris, R. Townsend, T. Romer, D. Epple, and G. Constantinides for several helpful discussions.

ibles and coinsurance in insurance policies? These and other questions can be answered only from a general formulation of the insurance problem, a formulation in which the previous models are imbedded.

In this paper I undertake the development of such a model. Using the same basic framework as Borch (1960), and Arrow (1971, 1973) the choice of contracts subject to various restrictions on the class of feasible contracts is considered. The analysis generalizes and extends the previous results in several directions: 1) The form of the Pareto optimal insurance contract is identified under general assumptions regarding the risk preferences of both the insurer and insured. The necessary and sufficient conditions leading to deductibles and coinsurance are investigated. 2) The cost of insurance is explicitly recognized and shown to be the driving force behind the deductible results. This clarifies the results obtained by Arrow (1971). 3) I show the conditions under which an insurance policy with an upper limit on coverage is adopted. 4) All results are extended to the case where more than one loss can occur during the period of insurance protection. A thorough understanding of the above issues not only contributes to our understanding of insurance policies, but provides a foundation for the analysis of optimal contracts in more general situations.

In this paper, the insurance policy is characterized by the premium paid by the insured and by a coverage function specifying the transfer from the insurer to the insured for each possible loss. The admissible coverage functions are restricted to be nonnegative and less than the size of the loss. Provision of the insurance is costly, with the cost consisting of fixed and variable (depending on the size of the insurance payment) components. The premium depends on the insurance policy and the insurance cost through a constraint on the insurer's expected utility of final wealth.

It is shown that the Pareto optimal insurance policy involves a deductible and coinsurance of losses above the deductible. The deductible is strictly positive if and only if the cost of insurance is a function of the insurance coverage. In other words, if the cost of providing insurance is independent of the insurance coverage then the Pareto optimal contract does not have a deductible. I conclude that the deductible clause in insurance policies exists due to two sources: the nonnegativity constraint on the transfer from the insurer to the insured and the variable insurance cost. The coinsurance arrangement is due to either the risk aversion of the insurer or the nonlinearity of the insurance costs. The exact functional relationship for the coinsurance is derived. These results are generalizations of Arrow's 1971 work and point out the crucial assumptions underlying his results. (Arrow's results are included as special cases.) Contrary to what might be inferred from Arrow it is shown that 1) insurer's risk neutrality is neither a necessary nor sufficient condition for a policy to have a deductible; 2) insurer's risk aversion is not a necessary condition for coinsurance; 3) an optimal policy may involve a deductible and a coinsurance. A policy with an upper limit on coverage, a feature common in major medical, liability, and disability insurance, is shown not to be Pareto optimal. To explain the prevalence of upper limits, a model is provided in which a risk-averse insurer is restricted (by regulation) in determining his premium by an actuarial constraint.

These results are initially derived under the assumption that at most one loss can occur during the period of insurance protection. When several losses are allowed to occur during the insurance period (from a single or several perils), the optimal insurance policy should be based on the aggregate loss and possess the same characteristics as previously discussed. If a policy with a deductible was optimal in the single loss case, then the policy should stipulate a deductible from the total claims when two or more losses occur. Similarly, in the case of insurance with upper limits, the upper limit should apply to the aggregate loss of the insured and not to each loss separately.

The assumptions and the model are specified in the next section, followed by a characterization of optimal insurance coverage when the premium is exogenously fixed. Section III completes the determination of Pareto optimal insurance policies. The behavior of a risk-averse insurer is investigated in Section IV while Section V extends all previous results to the multiple loss case. Conclusions and a critique are contained in the last section.

I. Assumptions and the Model

The insurance buyer faces a risk of loss of x, where x is a random variable with probability density function $f(x)$. Assume that $f(x) > 0$ for $0 \leq x \leq T$.[1]

The insurance policy is characterized by the payment, denoted by $I(x)$, transferred from the insurer to the insured if loss x obtains. Let us refer to $I(x)$ as the *insurance policy* or as the *coverage function*. Any admissible coverage function satisfies

$$(1) \qquad 0 \leq I(x) \leq x \qquad \text{for all } x$$

This constraint reflects the assumption that an insurance reimbursement is necessarily nonnegative and cannot exceed the size of the loss. The latter implies that the insured cannot gamble on his risk. This constraint also implies $I(0) = 0$; there is no reimbursement if there is no loss. The price paid by the insured, *the premium*, is denoted by P. Provision of insurance is costly due to administrative or other expenses and this cost is a deadweight loss relative to the insurer and the insured. It is assumed that the cost consists of fixed and variable (depending on the size of the insurance payment) components; $c(I)$ denotes the *cost* when the insurance payment is I with

$$(2) \qquad c(0) = a \geq 0, \qquad c'(\cdot) \geq 0, \qquad c''(\cdot) \geq 0$$

The insurer is assumed to maximize the expected value of his utility, which is a concave function of wealth; $V(W)$ denotes the utility function of the insurer with $V'(W) > 0$ and $V''(W) \leq 0$ for all W. Thus, the insurer is assumed to be risk averse (but not necessarily strictly risk averse). The special case of risk-neutral insurer, $V''(W) = 0$, is of particular interest.

If W_0 denotes the initial wealth of the insurer, then after selling the insurance policy and receiving the premium P, his final wealth is $W_0 + P - I(x) - c(I(x))$ if the loss x obtains. In other words, the insurer exchanges his initial certain utility $V(W_0)$ for the expected utility $E\{V[W_0 + P - I(x) - c(I(x))]\}$. A necessary condition for the insurer to offer such a policy is

$$(3) \qquad E\{V[W_0 + P - I(x) - c(I(x))]\} \geq V(W_0)$$

In the special case of a risk-neutral insurer, the risk premium equals zero and equation (3) takes the form: $P \geq E[I(x) + c(I(x))]$. Here, the policy is evaluated by the insurer according to the actuarial value of the coverage and cost. Often, in the insurance literature, it is assumed that

$$(2') \qquad a = 0,\ c'(I) = l \qquad \text{for all } I$$

That is, the costs are proportional to the insurance payment (fixed percentage loading l). In this case, the constraint on the policies offered is:

$$(3') \qquad P \geq (1 + l)E[I(x)]$$

On the insurance demand side, the insured is assumed to maximize the expected value of his utility of wealth. The insured's utility function of wealth is denoted by $U(w)$ with

$$(4) \qquad U'(w) > 0,\ U''(w) < 0 \qquad \text{for all } w$$

If w is the initial level of wealth, x the loss (a random variable), $I(x)$ the payment received from the insurer when loss x occurs, and P the premium paid for the insurance coverage, then the insured's final wealth is $w - P - x + I(x)$. Without purchasing insurance, his final wealth is $w - x$ when the loss x occurs. Thus, a necessary condition for purchasing the coverage $I(x)$ for a premium P is

$$(5) \qquad E\{U[w - P - x + I(w)]\} \geq E\{U[w - x]\}$$

[1]We could assume $f(x) \geq 0$. However, this would complicate the exposition without adding any content.

Necessary conditions for a *given* insurance contract to be acceptable to each party were given above. In order for a contract to be acceptable to both sides, both (3) and (5) have to be satisfied. In what follows, we will assume that the set of acceptable insurance contracts which satisfy these necessary conditions is nonempty. From this set, a Pareto optimal insurance policy will be chosen.

To find the form of the Pareto optimal insurance contract, we find the premium P and the function $I(\cdot)$ that maximize the insured's expected utility of final wealth subject to the constraint that the insurer's expected utility is constant. The problem is stated as follows:

$$(6) \quad \underset{P,I(x)}{\mathrm{Max}}\ \overline{U}(P,I) \equiv \int_0^T U[w - P - x + I(x)]\, f(x)dx$$

subject to (1) and

$$(7) \quad \overline{V}(P,I) \equiv \int_0^T V[W_0 + P - I(x) - c(I(x))]\, f(x)dx \geq k$$

where k is a constant and $k \geq V(W_0)$.

The above problem is solved in two steps. First, in Section II, the premium P is assumed fixed and the form of the optimal insurance coverage is found as a function of P. Second, in Section III, the optimal P is chosen, thus completing the solution to our problem.

II. Optimal Insurance Coverage for a Fixed Premium

The next theorem characterizes the solution to equation (6) subject to constraints (1) and (7) when P is fixed. The theorem states that optimal insurance policies have one of two possible forms: there is either a deductible provision coupled with coinsurance of losses above the deductible, or there is full coverage of losses up to a limit and coinsurance of losses above that limit. Coverage functions satisfying (8) below are referred to as *policies with a deductible*. The *deductible* $\overline{x}_1$ is the largest loss not covered by the insurance policy. Policies satisfying (9) are referred to as policies with an upper limit on full coverage or *policies with upper limit*. The *upper limit* $\overline{x}_2$ is the largest loss which is fully covered by insurance. Usually, the coinsurance level is the proportion of the loss covered by insurance. In our analysis this proportion varies with the size of the loss. Consequently, the *coinsurance* is defined as the marginal coverage, $I^{*\prime}(x)$. From (10) it is seen that the coinsurance depends on the risk preferences of the insurer and the insured as well as on the cost function $c(\cdot)$.[2]

THEOREM 1: *The solution* $I^*(x)$ *to equation* (6), *subject to constraints* (1) and (7) *when P is fixed, takes one of the two forms* (8) *or* (9) *where*

$$(8) \quad \begin{aligned} I^*(x) &= 0 && \text{for } x \leq \overline{x}_1 \\ 0 < I^*(x) &< x && \text{for } x > \overline{x}_1 \end{aligned}$$

$$(9) \quad \begin{aligned} I^*(x) &= x && \text{for } x \leq \overline{x}_2 \\ 0 < I^*(x) &< x && \text{for } x > \overline{x}_2 \end{aligned}$$

In both cases, in the range where $0 < I^*(x) < x$, the marginal coverage satisfies

$$(10) \quad I^{*\prime}(x) = \frac{R_U(A)}{R_U(A) + R_V(B)(1 + c') + c''/(1 + c')}$$

where

$$A = w - P - x + I^*(x)$$
$$B = W_0 + P - I^*(x) - c(I^*(x))$$

$R(\cdot)$ denotes the index of absolute risk aversion, and c', c'' are evaluated at $I^*(x)$.

PROOF:

Constraint (7) is binding at the optimum. Since the specified problem is solved via optimal control theory we rewrite (7) as

$$(11) \quad \begin{aligned} \dot{z}(x) &= V[W_0 + P - I(x) - c(I(x))]\, f(x) \\ z(0) &= 0 \\ z(T) &= k \end{aligned}$$

Using $I(x)$ as the control variable and

[2]When we have no contraints on the insurance function and when $c(\cdot) \equiv 0$ then (10) is identical to the sharing rule given by Wilson.

$z(x)$ as the state variable, the Hamiltonian for this problem is

$$H = \{U[w - P - x + I(x)] + \lambda V[W_0 + P - I(x) - c(I(x))]\} f(x)$$

Since the Hamiltonian does not depend on the state variable it is clear that the auxilary function λ is constant with respect to x.

The necessary conditions for the optimal coverage function to maximize the Hamiltonian subject to constraint (1) are

$$(12) \quad I^*(x) = 0 \quad \text{if } \mathbf{J} \equiv U'(w - P - x) - \lambda V'(W_0 + P - a)(1 + c'(0)) \leq 0$$

$$(13) \quad I^*(x) = x \quad \text{if } \mathbf{K} \equiv U'(w - P) - \lambda V'(W_0 + P - x - c(x))(1 + c'(x)) \geq 0$$

$$(14) \quad U'[w - P - x + I^*(x)] - \lambda V'[W_0 + P - I^*(x) - c(I^*(x))][1 + c'(I^*(x))] = 0 \quad \text{for } 0 < I^*(x) < x$$

First, note that either (12) or (13) has to occur for some x, although both cannot be satisfied simultaneously. This follows directly from the fact that $\mathbf{J}$ (as defined in (12)) is continuous and increasing in x while $\mathbf{K}$ is continuous and decreasing in x. If

$$\mathbf{L} \equiv U'(w - P) - \lambda V'(W_0 + P - a)(1 + c'(0)) \geq 0$$

then (12) cannot obtain for $x > 0$. If $\mathbf{L} \leq 0$ then (13) cannot obtain for $x > 0$. Hence the optimal solution satisfies either (12) and (14), or (13) and (14). Define $\bar{x}_i$, $i = 1,2$ from

$$(12') \quad U'(w - P - \bar{x}_1) - \lambda V'(W_0 + P - a)(1 + c'(0)) = 0$$

$$(13') \quad U'(w - P) - \lambda V'(W_0 + P - x_2 - c(\bar{x}_2))(1 + c'(\bar{x}_2)) = 0$$

Clearly, $\bar{x}_i$ is uniquely defined by (12′) and (13′), respectively. (The special case $\bar{x}_1 = \bar{x}_2 = 0$ occurs if $U'(w - P) - \lambda V'(W_0 + P - a)(1 + c'(0)) = 0$.) As a result, the optimal coverage function takes one of the two forms (8) or (9). In both cases, for $x > \bar{x}_i$ (14) is satisfied. Differentiating with respect to x and using the earlier definitions of A and B we obtain for $x > \bar{x}_i$:

$$(15) \quad U''(A)[I^{*\prime}(x) - 1] + \lambda V''(B)[1 + c'(I^*(x))]^2 I^{*\prime}(x) - \lambda V'(B) c''(I^*(c)) I^{*\prime}(x) = 0$$

Substituting λ from (14) and solving for $I^{*\prime}(x)$, (10) is obtained.

Notice also that, since the Hamiltonian does not depend on the state variable, the sufficiency theorem concavity requirement as shown by Morton Kamien and Nancy Schwartz is satisfied trivially. Hence, $I^*(x)$ satisfies the necessary and sufficient conditions of optimality.

Theorem 1 is easily interpreted. In the absence of constraint (1), risk aversion of both parties implies that Pareto optimal coverage involves sharing the risk according to the sharing rule (10). Equation (10) is a differential equation which, together with a boundary condition, results in a coverage function. Whether the optimal policy has a deductible or an upper limit depends on the appropriate boundary condition. For example, if the boundary condition is $I(\bar{x}_1) = 0$, then the policy has a deductible. The appropriate boundary condition depends on the fixed premium. When the premium is P, let $I_P(x)$ denote the function which solves the differential equation (10) with the boundary condition $I_P(0) = 0$. Since $0 < I'_P < 1$, this function also satisfies constraint (1). To verify whether this function is the solution to the problem the insurer's expected utility $\overline{V}(P, I_P)$ must be evaluated. Three cases could occur: 1) If $\overline{V}(P, I_P) = k$, then (7) is satisfied, $I_P(0) = 0$ is the appropriate boundary condition, and I_P is the optimal coverage function. 2) If $\overline{V}(P, I_P) < k$, then I_P is not the solution since (7) is violated. To increase the insurer's expected utility to the required level, the payment to the insured has to be reduced for some losses. This could be achieved by a boundary condition specifying that $I^*(0)$ is negative. However, the constraint $I^*(x) \geq 0$ becomes binding and the appropriate boundary condition is $I^*(x) = 0$ for $x \leq \bar{x}_1$. In this case, the optimality of the deductible policy is obtained. 3) If $\bar{V}(P, I_P) > k$, the coverage can be increased. This increases the insured's expected utility, while the constraint (7) on insured's utility is not violated. The appropriate boundary condition is $I^*(0) > 0$, which together with the constraint $I(x) \leq x$, results in $I^*(x) = x$ for $x \leq \bar{x}_2$. In this case, the policy with an upper limit is obtained.

Let P_0 be the fixed premium corresponding to the first case above and let $I_0(\cdot)$ be the function solving (10). Thus, $\bar{V}(P_0, I_0) = k$ and the coverage function has the property $\bar{x}_1 = \bar{x}_2 = 0$ (i.e., (P_0, I_0) is the policy with no deductible or upper limit provision). Denote

$$(16) \qquad S_1 = \{P \mid \bar{V}(P, I_P) \leq k\}$$

$$\text{and} \qquad S_2 = \{P \mid \bar{V}(P, I_P) \geq k\}$$

By definition, $P_0 \in S_i$ for $i = 1$ and 2. Lemma 1 summarizes the discussion above and states that the optimal coverage involves a nontrivial deductible if $P \in S_1$ and a nontrivial upper limit if $P \in S_2$.

LEMMA 1: *For* $P \in S_i$, $i = 1$ *or* 2 *and* $P \neq P_0$, $I^*(x)$ *is specified by* (8) *and* (10) *or* (9) *and* (10), *respectively, with* $\bar{x}_i > 0$, $i = 1$ *or* 2.

Lemma 2 determines the effect of a change in $\bar{x}_i$ on $I^*(x)$ for $P \in S_i$. It is stated that for policies with a deductible, the coverage function decreases with the deductible level. Similarly, for policies with upper limit, the coverage function increases with the upper limit. If the sharing proportion $I^{*\prime}$ was constant, then these results clearly follow from the changes in the initial conditions of the differential equation. The proof that it is correct in the present, more general, case is given in Appendix A.

LEMMA 2: *a) If* $P \in S_1$ *then* $\partial I^*/\partial \bar{x}_1 < 0$ *for* $x > \bar{x}_1$ *and b) If* $P \in S_2$ *then* $\partial I^*/\partial \bar{x}_2 > 0$ *for* $x > \bar{x}_2$.

III. Pareto Optimal Insurance Policy

In the previous section the premium was assumed fixed. Therefore, in Theorem 1, $\bar{x}_1$, $\bar{x}_2$ and $I^*(x)$ are functions of P. We proceed to determine P^*, thus completing the determination of the Pareto optimal insurance policy. Theorem 2 proves that the search for the optimal premium can be restricted to the subset S_1 of premiums which generate policies with a deductible. Within this subset, Theorem 3 characterizes the necessary and sufficient conditions for the deductible to be nontrivial. These two results together complete the derivation of Pareto optimal policies and allow us to clearly distinguish the cases under which we would expect to observe deductibles and coinsurance clauses in insurance contracts. The results obtained by Arrow (1971) are treated as special cases thus allowing us to focus on the specific assumptions which generate these results.

The next theorem states that any insurance policy with an upper limit is dominated by the policy (P_0, I_0) with zero upper limit. In other words, the pure sharing arrangment dominates any policy with an upper limit. Intuitively, starting with the (P_0, I_0) policy, any increase in $\bar{x}_2$ (from $\bar{x}_2 = 0$) has the effect of increasing insurance coverage for all losses which, in turn, increases the dead-weight loss due to increased insurance costs and therefore is suboptimal.

THEOREM 2: $\bar{U}(P_0, I_0) \geq \bar{U}(P, I^*)$ *for all* $P \in S_2$.

The proof consists of comparing the slopes of the indifference curves for the insured and the insurer in $P, \bar{x}_2$ space. It is shown that for an incremental increase in $\bar{x}_2$ the insured is willing to increase P less than is required for the insurer to remain indifferent. Because of the limited space and since the proof is similar to the proof of Theorem 3 the details are omitted. The interested reader can receive the proof from the author upon request.

After showing that the Pareto optimal

insurance policy will not be of the upper-limit type, we now investigate the conditions under which the Pareto optimal policy will or will not include a deductible clause. The next theorem specifies that a (nontrivial) deductible is obtained if and only if the insurance cost depends on the insurance payment.

THEOREM 3: *A necessary and sufficient condition for the Pareto optimal deductible to be equal to zero is* $c'(\cdot) \equiv 0$ (*i.e.*, $c(I) = a$ *for all* I).

The proof is given in Appendix B. We compare the insurer's and the insured's tradeoff between x_1 and P. If $c'(\cdot) = 0$, it is shown that for a marginal increase in the deductible the amount the insured is willing to pay in premium is less than that required by the insurer. On the other hand, if $c'(\cdot) > 0$, the insured is willing to pay more than what is required by the insurer and, therefore, the deductible is greater than zero.

Theorems 2 and 3 characterize the Pareto optimal insurance policy. What are the implications of these results regarding the contract form that we would expect to observe? The persistence of deductibles is explained by Theorem 3. If the cost of insurance depends on the coverage, then a nontrival deductible is obtained. This result does not depend on the risk preferences of the insured or the insurer. I stress this fact to point out that Arrow's (1971, Theorem 1) deductible result was not a consequence of the risk-neutrality assumption. Rather, it was obtained because of the assumption that insurance cost is proportional to coverage. For completeness Arrow's result is reproduced as a special case of my treatment.[3]

COROLLARY 1 (Arrow 1971): *If* $c(I) = lI$ *and the insurer is risk neutral, the Pareto optimal policy is given by*

$$I^*(x) = \begin{cases} 0 & \text{for } x \leq \bar{x}_1 \\ x - \bar{x}_1 & \text{for } x > \bar{x}_1 \end{cases} \tag{17}$$

where $\bar{x}_1 > 0$ *if and only if* $l > 0$.

PROOF:

By Theorem 3, the Pareto optimal policy involves $\bar{x}_1 > 0$ iff $c' = l > 0$. The form of the coverage function is specified by (10). Since $R_V \equiv 0$ and $c'' = 0$, we have $I^{*\prime} = 1$ for $x > \bar{x}_1$. Thus $I^*(x)$ is given by (17).

Even if risk neutrality is not assumed, we still obtain a nontrivial deductible. The coverage involves, however, a coinsurance arrangement for losses above the deductible. The coinsurance level is given by (10) and in this special case is

$$I^{*\prime}(x) = \frac{R_U(A)}{R_U(A) + R_V(B)(1 + l)} < 1$$

This is the generalization of Arrow's (1971, Theorem 1) result to the risk-averse insurer case. Risk aversion of the insurer causes the coinsurance of losses above the deductible. Because of the insurance costs, the deductible is strictly positive.

What are the conditions that lead to a coinsurance arrangement? As already pointed out, risk aversion on the part of the insurer could be the cause for coinsurance. With no insurance costs this was proved by Arrow (1971, Theorem 2). In this case, there is no deductible. Our results prove, however, that a Pareto optimal policy may include a deductible and coinsurance of losses above the deductible.

COROLLARY 2: *If the insurer is risk averse, then the Pareto optimal insurance policy involves coinsurance of losses above the deductible.*

PROOF:

From (10) it is clear that $I^{*\prime} < 1$ for $x > \bar{x}_1 \geq 0$.

[3]Strictly speaking, Arrow (1971) did not fully discuss the Pareto optimal policy. His theorem characterizes only the optimal coverage function for a given premium.

Risk aversion, however, is not the only explanation for coinsurance. Even if the insurer is risk neutral, coinsurance might be observed, provided the insurance costs are a strictly convex function of the coverage. The intuitive reason for this result is that the cost function nonlinearity substitutes for the utility function nonlinearity.

COROLLARY 3: *If the insurer is risk neutral and $c'' > 0$, the Pareto optimal policy involves a deductible, $\bar{x}_1 > 0$, and coinsurance of losses above the deductible.*

PROOF:

Since $c' > 0$, we know $\bar{x}_1 > 0$ by Theorem 3. From (10) we have that for $x > \bar{x}_1$,

$$I^{*\prime}(x) = \frac{R_U(A)}{R_U(A) + c''/(1 + c')} < 1$$

In this section, the Pareto optimal insurance policy which was shown to specify a deductible and coinsurance of losses above the deductible was characterized. The deductible was shown to be strictly positive if and only if the insurance cost depended on the insurance payment. Coinsurance results from either insurer risk aversion or the cost function nonlinearity.

IV. Policies with Upper Limit on Coverage

The previous section explained deductibles and coinsurance arrangements in insurance contracts, as well as why Pareto optimal insurance policy does not involve an upper limit on coverage. In this section, I attempt to explain the prevalence of upper limits on coverage which are frequently incorporated in major medical, liability, and property insurance. The explanation rests on the fact that insurance companies are frequently regulated and therefore operate subject to a regulatory constraint. In what follows, it is argued that the upper limit on insurance coverage is desired by the insurance seller restricted in his policy offering by an actuarial constraint. Intuitively, the insurer is required to sell a policy with a prescribed actuarial value, for any given premium. This actuarial value might be smaller than the expected monetary loss, and then the policy cannot fully cover all potential losses. Being risk averse, the insurer prefers to allocate this given policy actuarial value to full coverage of small losses and limited coverage of large losses, rather than any other feasible form of the coverage. Heavy losses above that limit will not be insured under this contract.

To make the above statements precise, the results of the previous section are specialized to characterize the insurance policy desired by the insurance seller. Assume that the insurer devises contracts so as to maximize his expected utility of final wealth:

$$(18) \quad \bar{V}(P, I) = \underset{P, I(x)}{\mathrm{Max}} \int_0^T V[W_0 + P - I(x) - c(I(x))]\, f(x)dx$$

The class of feasible insurance contracts is restricted by (1) and by the assumption that the premium received is required (by regulation) to be a function of the policy's actuarial value.[4] Denoting this function by R, assume

$$(19) \qquad P = R\left\{ \int_0^T I(x)\, f(x)dx \right\}$$

Equation (19) specifies a general relationship between the premium charged and the actual value of the policy. This specification is consistent with the procedure used by regulatory agencies under the prior approval laws which are the predominant form of regulation of the property-liability insurance industry. (See Paul Joskow for the description of the pricing behavior in this industry.) In general, rates are established so as to yield a particular rate of return on sales (premiums). As Joskow states: "A standard rate of return on sales figure

[4]The result of this section would not change if we assume that the premium depends on the actuarial value of the coverage and insurance cost.

of 5 percent is employed in most states as a result of a recommendation by the National Association of Insurance Commissioners in 1921" (p. 394). Under this procedure, the pricing formula is

$$P(1 - .05) = \text{Expected Losses} + \text{Operating Expenses}$$

which is a special case of our formulation in (19).

Theorem 4 characterizes the solution to the insurer's problem specified above, stating that if a risk-averse insurer selects an insurance policy to maximize his expected utility, then the policy offered fully covers losses up to certain upper limit, and covers no losses above the limit.

THEOREM 4: *The solution to problem* (18) *subject to constraints* (1) *and* (19) *is* P^* *and* $I^*(x)$ *such that*

$$(20) \qquad I^*(x) = \begin{cases} x \text{ for } x \leq \bar{x} \\ \bar{x} \text{ for } x > \bar{x} \\ 0 \leq \bar{x} \leq T \end{cases}$$

and $\bar{x} = \bar{x}(P^*)$.

PROOF:

Starting with a fixed P, the optimal solution is shown to have the form of equation (20). The determination of the optimal P^* is then discussed.

The policies obtained in Theorem 1 can be viewed as the solution to the following problem: Maximize the insurer's expected utility of final wealth subject to constraint (1) and a restriction on the insured's expected utility of final wealth: $EU[w - P - x + I(x)] = c_1$, where c_1 is a constant. If the utility function U is linear and P is given, this restriction can be rewritten as $EI(x) = c_2$, which is equivalent to constraint (19). Thus, the solution to the present problem is obtained directly from Theorem 1 by specifying $R_U(\cdot) = 0$. This yields $I^{*\prime}(x) = 0$, for $x > \bar{x}_i$. In the first case, $I^*(x) = 0$ for $x \leq \bar{x}_1$, and, therefore, $I^*(x) = 0$ for all x. In the second case, $I^*(x) = x$ for $x \leq \bar{x}_2$, and, therefore, $I^*(x) = \bar{x}_2$ for $x > \bar{x}_2$ thus proving that the optimal form of the contract is given by (20). The constant $\bar{x}_2$ is determined by the optimal premium P^*, which depends on the function R and the insurance cost. If loading is sufficiently high, full coverage of all losses could be obtained (i.e., $\bar{x}_2 = T$). On the other hand, loading could be low enough so that no insurance is offered (i.e., $\bar{x}_2 = 0$). In general, therefore, $0 \leq \bar{x}_2 \leq T$.

V. Optimal Insurance Policies when Multiple Losses Can Occur

In the previous sections the insurance policy contracted between the insurance buyer and the insurance seller was analyzed. My model, however, incorporated the simplifying assumption that only a single loss can occur during the period of insurance protection. This assumption appears to be too restrictive; business firms and individuals may be faced with risks that could result in more than one loss during the period of insurance protection. Furthermore, the insurance buyer will typically purchase several different policies to cover different perils that he faces. The present analysis will extend the results of the previous sections to derive the properties of an optimal policy when several potential losses are faced by the insured. To facilitate notation, the proofs have been restricted to the case of two potential losses. The analysis carries over to more general cases.

The insurance buyer is assumed to face two potential losses during the period of insurance coverage. His total monetary loss is $x_1 + x_2$ where x_i, $i = 1, 2$, are assumed to be random variables defined on $[0, T_i]$ with a joint probability density function $f(x_1, x_2)$.

The insurance policy is characterized by the payment $I(x_1, x_2)$ transferred from the insurer to the insured if losses x_1, x_2 obtain. As before, $I(x_1, x_2)$ is referred to as the insurance policy or coverage function. Similar to condition (1) a restriction is imposed on the insurance function:

$$(21) \quad 0 \leq I(x_1, x_2) \leq x_1 + x_2 \qquad \text{for all } x_1, x_2$$

The Pareto optimal coverage function $I(x_1, x_2)$ is obtained by maximizing the insured's expected utility of final wealth subject to the constraint that the insurer's expected utility exceeds a given constant. The problem is then stated as follows:

$$(22)\quad \underset{P,I}{\text{Max}} \int_0^{T_2}\int_0^{T_1} U[w - P - x_1 - x_2 + I(x_1, x_2)] f(x_1, x_2) dx_1 dx_2$$

subject to

$$(23)\quad \int_0^{T_2}\int_0^{T_1} V[W_o + P - I(x_1, x_2) - c(I(x_1, x_2))] f(x_1, x_2) dx_1 dx_2 \geq k$$

and

$$(24)\qquad 0 \leq I(x_1, x_2) \leq x_1 + x_2$$

The above problem has the form of an iso-parametric problem in the calculus of variations with the additional constraint (24). Since the unknown function $I(x_1, x_2)$ depends on two variables, the extension of the simple Euler equation can be used to include two dimensions and constraints in order to derive the optimal insurance policy. Rather than proceeding along those lines, we first prove that the optimal function depends only on the sum $(x_1 + x_2)$.[5] Thus, the coverage function depends on one variable, the aggregate loss, and all previous results apply to this aggregate loss.

THEOREM 5: *Let $I^*(x_1, x_2)$ be the solution to the problem* (22) *subject to constraints* (23) *and* (24). *Then, $I^*(x_1, x_2)$ depends on the sum $(x_1 + x_2)$ only, i.e., $I^*(x_1, x_2) = \bar{I}^*(x_1 + x_2)$.*

The proof consists of showing that for any function $I(x_1, x_2)$ which does not depend on the sum only, there exists another coverage function $I^*(x_1, x_2)$ which increases the objective (22), is feasible, and depends on the sum only. The detailed proof is tedious and can be obtained from the author upon request. In what follows I provide the intuition behind the result and its proof.[6]

For simplicity, suppose that with probability $p(p')$ the losses are x_1, $x_2(x_1', x_2')$. Thus, the total loss is $x_1 + x_2$ or $x_1' + x_2'$ with probabilities p and p', respectively. Assume that $x_1 + x_2 = x_1' + x_2' = y$. Consider a coverage function $I(\cdot, \cdot)$ which does not depend only on the sum; $I(x_1, x_2) \neq I(x_1', x_2')$. A risk averter prefers to exchange any uncertainty for a certain outcome. In particular, the insured's expected utility for these two states can be increased by providing a coverage function which depends on y only:

$$pU[w - P - y + I(x_1, x_2)] + p'U[w - P - y + I(x_1', x_2')] \leq U[w - P - y + I^*(y)]$$

where $I^*(y) = pI(x_1, x_2) + p'I(x_1', x_2')$. Thus, the function I is dominated by the function I^*. I^* is the "weighted average" or the expected value of the coverages of equal total losses. By a similar argument, it can be shown that the insurer also prefers the coverage I^*. The above intuitive argument can be generalized. The driving force for the proof is, as above, the concavity of the utility functions; the dominance of I^* is established via Jensen's inequality.

Recall that our objective is to find optimal insurance policies when the insured is facing two potential losses. In Theorem 5 it was proved that any Pareto optimal coverage function depends only on the aggregate loss. The aggregate loss is denoted by $y = x_1 + x_2$ with probability density function $g(y)$. Using Theorem 5, we can rewrite problem (22)–(24) as: Find a coverage function $I(y)$ to maximize

$$\int_y U[w - P - y + I(y)] g(y) dy$$

subject to the constraints

[5] Borch (1962) proved that "any Pareto optimal set of treaties is equivalent to a pool arrangement" (p. 428). In his analysis, however, insurance was costless, and there were no constraints imposed on the feasible insurance policy. Therefore, he did not obtain the deductible or upper-limit results and could not generalize these results to the multiple loss case.

[6] I am indebted to Arie Tamir for suggesting this intuitive approach.

$$\int_y V[W_0 + P - I(y) - c(I(y))]g(y)dy \geq k$$
$$0 \leq I(y) \leq y$$

The above problem has the same structure as the problem considered in Sections II–IV with y replacing x. Thus, all the results regarding optimal insurance policies hold unchanged when the insured faces more than one risk, when the loss considered is the aggregate loss from all those risks. For example, if a risk-neutral insurer offers insurance policies and incurs linear cost, then it was proved that the Pareto optimal policy involves full coverage of losses beyond the deductible. Hence, we can now state:

COROLLARY 1′: *If $c(I) = lI$ and the insurer is risk neutral, the Pareto optimal policy is given by:*

$$I^*(x_1, x_2) = \begin{cases} 0 & \text{for } x_1 + x_2 \leq \bar{x} \\ x_1 + x_2 - \bar{x} & \text{for } x_1 + x_2 > \bar{x} \end{cases}$$

where $\bar{x} > 0$ if and only if $l > 0$.

Similarly, all the theorems of the previous sections can be now restated with the only difference being that the loss considered is interpreted as the aggregate loss during the period of insurance protection.

VI. Conclusions

In this paper the prevalence of different insurance contracts was explained. It was shown that the Pareto optimal insurance contract involves a deductible and coinsurance of losses above the deductible. The deductible feature was shown to depend on the insurance costs. The coinsurance is due to either risk or cost sharing between the two parties. The upper limits on insurance were shown to be Pareto suboptimal. Their prevalence was shown to be in the interest of the regulated insurer. All results were obtained for single as well as multiple losses.

Two shortcomings of the above analysis should be noted. First, adverse selection problems were not analyzed; both the insurer and the insured were assumed to know the probability distribution function of the losses. Second, moral hazard problems were ignored; the monetary loss was assumed exogenous and not under the insured's control. A detailed analysis of the optimal contracts in these cases is much more difficult and was not attempted here.

Appendix A

PROOF of Lemma 2:

a) If $P \epsilon S_1$, $I^*(x) = 0$ for $x \leq \bar{x}_1$. For $x > \bar{x}_1$, $I^*(x) = \int_{\bar{x}_1}^{x} I^{*\prime}(t)dt$, where $I^{*\prime}$ is given by (10). Differentiating with respect to $\bar{x}_1$,

$$\frac{\partial I^*(x)}{\partial \bar{x}_1} = -I^{*\prime}(\bar{x}_1) + \int_{\bar{x}_1}^{x} \frac{\partial I^{*\prime}(t)}{\partial I^*} \cdot \frac{\partial I^*(t)}{\partial \bar{x}_1} dt$$

Solving this equation yields

$$\frac{\partial I^*(x)}{\partial \bar{x}_1} = -I^*(\bar{x}_1) \exp \left\{ \int_{\bar{x}_1}^{x} \frac{\partial I^{*\prime}(t)}{\partial I^*} dt \right\} < 0$$

Part (b) is proved similarly.

Appendix B

PROOF of Theorem 3:

From (7) we have that for all $P \epsilon S_1$

$$\bar{V}(P, I^*) = \int_0^{\bar{x}_1} V[W_0 + P - a] f(x)dx + \int_{\bar{x}_1}^{T} V(B) f(x)dx = k$$

By differentiating we obtain expressions for $d\bar{P}/d\bar{x}_1$ when $\bar{V}$ and $\bar{U}$ are held constant. These are shown on page 95. From (12′) and (14) we have that for $x \geq \bar{x}_1$

$$\text{(A1)} \quad \frac{U'(A)}{U'(w - P - \bar{x}_1)} = \frac{V'(B)(1 + c')}{V'(W_0 + P - a)[1 + c'(0)]}$$

Therefore,

$$\text{(A2)} \quad \frac{1}{U'(w - P - \bar{x}_1)} \int_{x_1}^{T} U'(A) \frac{\partial I^*}{\partial \bar{x}_1} f(x)dx =$$

$$\frac{dP}{d\bar{x}_1}\Big|_{\bar{V}=const} = \frac{\int_{\bar{x}_1}^{T} V'(B)(1+c')\frac{\partial I^*}{\partial \bar{x}_1} f(x)dx}{\int_0^{\bar{x}_1} V'[W_0+P-a]f(x)dx + \int_{\bar{x}_1}^{T} V'(B)[1-(1+c')\frac{\partial I^*}{\partial P}]f(x)dx}$$

$$\frac{dP}{d\bar{x}_1}\Big|_{\bar{U}=const} = \frac{\int_{\bar{x}_1}^{T} U'(A)\frac{\partial I^*}{\partial \bar{x}_1} f(x)dx}{\int_0^{\bar{x}_1} U'(w-P-x)f(x)dx + \int_{\bar{x}_1}^{T} U'(A)\left(1-\frac{\partial I^*}{\partial P}\right)f(x)dx}$$

$$\frac{1}{V'(W_0+P-a)[1+c'(0)]} \int_{x_1}^{T} V'(B)(1+c')\frac{\partial I^*}{\partial \bar{x}_1} f(x)dx$$

To prove sufficiency, assume $c'(\cdot) = 0$. From (A1) and since x in the first integral is smaller than $\bar{x}_1$ we have

$$\text{(A3)}\quad \frac{1}{U'(w-P-\bar{x}_1)} \left\{\int_0^{\bar{x}_1} U'(w-P-x)f(x)dx + \int_{\bar{x}_1}^{T} U'(A)\left(1-\frac{\partial I^*}{\partial P}\right)f(x)dx\right\}$$
$$\leq \frac{1}{V'(W_0+P-a)} \left\{\int_0^{\bar{x}_1} V'(W_0+P-a)f(x)dx + \int_{\bar{x}_1}^{T} V'(B)\left(1-\frac{\partial I^*}{\partial P}\right)f(x)dx\right\}$$

Dividing (A2) by (A3) and recalling that, by Lemma 2, $\partial I^*/\partial \bar{x}_1 < 0$

$$\frac{dP}{d\bar{x}_1}\Big|_{\bar{U}=const} \leq \frac{dP}{d\bar{x}_1}\Big|_{\bar{V}=const}$$

with the equality holding only if $\bar{x}_1 = 0$. Thus the optimal policy is $\bar{x}_1 = 0$ and the premium is P_0 as was claimed.

To prove necessity, assume $c'(0) > 0$. There exists $y > 0$ such that

$$\frac{U'(w-P)}{U'(w-P-y)} = \frac{1}{1+c'(0)}$$

Using (A1) for $\bar{x}_1 > y$ we obtain

$$\text{(A4)}\quad \frac{1}{U'(w-P-\bar{x}_1)} \left\{\int_0^{\bar{x}_1} U'(w-P-x)f(x)dx + \int_{\bar{x}_1}^{T} U'(A)\left(1-\frac{\partial I^*}{\partial P}\right)f(x)dx\right\}$$
$$\geq \frac{1}{V'(W_0+P-a)[1+c'(0)]} \left\{\int_0^{\bar{x}_1} V'(W_0+P-a)f(x)dx + \int_{\bar{x}_1}^{T} V'(B)(1+c')\left(1-\frac{\partial I^*}{\partial P}\right)f(x)dx\right\}$$
$$> \frac{1}{V'(W_0+P-a)[1+c'(0)]} \left\{\int_0^{\bar{x}_1} V'(W_0+P-a)f(x)dx + \int_{\bar{x}_1}^{T} V'(B)\left[1-(1+c')\frac{\partial I^*}{\partial P}\right]f(x)dx\right\}$$

The last inequality is obtained since $c' > 0$. Dividing (A2) by (A4) and since $\partial I^*/\partial \bar{x}_1 < 0$ we have that for $\bar{x}_1 < y$

$$\frac{dP}{d\bar{x}_1}\Big|_{\bar{U}=const} > \frac{dP}{d\bar{x}_1}\Big|_{\bar{V}=const}$$

Therefore, the optimal deductible level in this case is different from zero, as we argued.

REFERENCES

Kenneth J. Arrow, *Essays in the Theory of Risk Bearing*, Chicago 1971.

———, "Optimal Insurance and Generalized Deductibles," Rand Corp., R-1108-OEO, Feb. 1973.

K. Borch, "The Safety Loading of Reinsurance Premiums," *Skand. Aktuarietidskrift*, 1960, 162–84.

———, "Equilibrium in a Reinsurance Market," *Econometrica*, July 1962, *30*, 424–44.

J. P. Gould, "The Expected Utility Hypothesis and the Selection of Optimal Deductibles for a Given Insurance Policy," *J. Bus., Univ. Chicago*, Apr. 1969, *42*, 143–51.

P. L. Joskow, "Cartels, Competition and Regulation in the Property-Liability Insurance Industry," *Bell J. Econ.*, Autumn 1973, *4*, 375–427.

M. I. Kamien and N. L. Schwartz, "Sufficient Conditions in Optimal Control Theory," *J. Econ. Theory*, June 1971, *3*, 207–14.

J. Mossin, "Aspects of Rational Insurance Purchasing," *J. Polit. Econ.*, July/Aug. 1968, *76*, 533–68.

V. L. Smith, "Optimal Insurance Coverage," *J. Polit. Econ.*, Jan./Feb. 1968, *76*, 68–77.

R. B. Wilson, "The Theory of Syndicates," *Econometrica*, Jan. 1968, *36*, 119–32.