R-1108-OEO

February 1973

Optimal Insurance and Generalized Deductibles

Kenneth J. Arrow

A Report prepared for

OFFICE OF ECONOMIC OPPORTUNITY

25th Year

Rand

SANTA MONICA, CA. 90406

The research reported herein was performed pursuant to a grant from the Office of Economic Opportunity, Washington, D.C. The opinions and conclusions expressed herein are solely those of the author, and should not be construed as representing the opinions or policy of any agency of the United States Government.

R-1108-OEO
February 1973

Optimal Insurance and Generalized Deductibles

Kenneth J. Arrow

A Report prepared for

OFFICE OF ECONOMIC OPPORTUNITY

Rand
SANTA MONICA, CA. 90406

Published by The Rand Corporation

PREFACE

This report is a study in the economics of optimal insurance. It was prepared for the Office of Economic Opportunity as part of a project to study the financing of medical care services for the poor and near-poor.

In assessing policy toward health care financing, it is important to determine what features characterize an optimal medical insurance policy. The author had earlier considered this problem and had shown that under certain assumptions, including a positive administrative fee or loading, the optimal policy was characterized by complete insurance above a deductible amount. One of the assumptions was that the utility function for income be independent of the state of the individual. This appears excessively restrictive, especially for medical insurance, where the marginal utility of income may vary considerably depending upon the individual's state of health.

Relaxation of this assumption indicates that the basic result cited above still holds but that the deductible amount depends upon the state of the individual. More precisely, insurance is such that the marginal utility of income never exceeds a certain critical level.

One of the assumptions made here is that medical insurance can be analyzed as a lump-sum transfer that does not affect consumption except through an income effect. This is known not to be the case. (See, for example, Charles Phelps and Joseph Newhouse, *Coinsurance and the Demand for Medical Services*, R-964-OEO/NC, The Rand Corporation (forthcoming). The author is currently working on the problem of how the results of this report change when that assumption is relaxed.

SUMMARY

This report is intended as a contribution to the theory of demand for insurance. In many circumstances, it appears that, given a range of alternative possible insurance policies, the insured would prefer a policy offering complete coverage beyond a deductible. In an earlier paper (Arrow, 1963; reprinted in Arrow, 1971, pp. 212-216), this argument was developed for the case where the risk being insured against was, effectively, loss of income. Recently, Ehrlich and Becker (1972) have extended these results considerably, as well as analyzing other responses of the insured to the price of insurance, responses beyond the scope of this study. For some other related work, see Pashigian, Schkade, and Menefee (1965), Smith (1968), and Gould (1969). However, income is not the only uncertainty, especially in the context of health insurance, and only under special and unrealistic circumstances can it be held that the other uncertainties have income equivalents. Put loosely, the marginal utility of income will in general depend not only on the amount of income but also on the state of the individual or, more generally, on the state of the world.

The insurance policies considered here specify a cash payment for each possible state of the world. Therefore, they do not include exactly the typical health insurance policy, which provides for reimbursement of actual expenses in whole or in part. There is a two-fold difference between the cash payment contingent on a state of the world and the reimbursement policy: (1) the payment in the latter case depends on a decision of the patient and the physician and not only on the objective state of affairs; in particular, it ignores the price elasticity of demand for medical services[1]; and (2) the reimbursement policy

[1]It is well known that contingent cash payment policies are more efficient means of resource allocation than alternatives, such as reimbursement policies. I believe the latter exist because of an informational inequality between insurer and insured. The insured (and his agents, such as his physicians) knows the extent of his illness in much more detail than the insurer. Indeed, the insurer usually has no independent source of information and could create one only at considerable

is a payment in kind, not in fungible cash, so that the satisfaction derived by the insured from an insurance payment may be lower than that from an equal cash payment.[1]

The choice of insurance policies is constrained in two ways: (1) the insurer is assumed to be risk-neutral, so that premiums depend only on the expected insurance payment, but because of administrative or other expenses, the premium may exceed the actuarial value; and (2) insurance payments are nonnegative. The latter assumption not only conforms to everyday observation but is virtually necessitated by moral hazard; in general, it may be difficult to collect from an insured even in a state whose outcome is favorable to him.

More explicitly, it is assumed that the insured is given the probabilities of different states, his initial (pre-insurance) income in each, a utility function for income in each state, which may vary from state to state (see Section I), and a given ratio of expected benefits to premiums (not exceeding 1). Then he can freely choose a premium and a system of nonnegative insurance payment for all the states so as to maximize his expected utility.

In the case where the utility function for income is the same for all states (Section II) the optimal policy has a simple form; a critical

---

cost both to himself and to the insured. Hence, the policy cannot be written as a schedule of cash payments for the different states because the insurer and the insured do not have equal access to this information. The payments themselves, or some related magnitudes, such as quantities of different kinds of medical care, on the other hand, are statements about the world that are verifiable by both parties to the transaction. I intend to develop this theme in subsequent work.

[1]Fire and other forms of casualty insurance typically pay money for the occurrence of certain states, without restricting the subsequent expenditures of the individual; if my house burns down, the insurance payment is not contingent on my rebuilding it. Medical insurance, however, invariably is only a payment for medical expenses; I cannot choose to make a claim on the grounds of illness and then spend the money on a vacation. The difference is important only if the constraint is in fact binding. That is, if, given the illness and the payment of insurance, I would in fact choose to use the entire payment and perhaps even more for medical expenses even if I were free to divert the payment to other uses, then the medical insurance is simply an alteration of income. The effectiveness of the expenditure constraint in practice is worthy of some investigation.

income level is selected, and the insurance payment is equal to the extent to which actual income falls short of the critical level. The distribution of a given volume of insurance among states is independent of the individual's utility function, though the total amount demanded does depend on it.

When utility of income is state-dependent (see Section III), the policy is characterized by a critical marginal utility of income, with payments being made so as to bring the marginal utility down to the critical level in those states where this is needed. An algorithm is found for determining the optimal policy (see also Section VI).

If the expected benefits or the premium shift in any way, the optimal post-insurance incomes in those states with positive insurance all shift in the same direction, the changes being proportional to a magnitude known as the risk tolerance (Section IV). It can also be shown that if we hold expected benefits constant, then the critical marginal utility increases with the premium (and therefore post-insurance incomes in all insured states decrease); on the other hand, for constant premium, the critical marginal utility falls as expected benefits rise.

These partial results can be combined to give a full characterization of the optimal policy, to determine both the optimal premium and the optimal critical marginal utility. With some modifications, the policy is characterized by two relations: a budget constraint, that the ratio of expected benefits to premium be the specified benefit-premium ratio; and the condition that the ratio of the expected to the maximum marginal utility of post-insurance income equal this benefit-premium ratio (see Sections V and VII). In the actuarially fair case, when the benefit-premium ratio is 1, all states are insured; otherwise, the probability that an insured state occurs is less than the benefit-premium ratio.

The comparative statics of the optimal policy are examined with respect to shifts in the various parameters. As the benefit-premium ratio increases, the critical marginal utility falls; the premium rises in general, but there may be exceptions. The effects of changes in the probabilities of different states are rather complicated to describe (see Theorem 7 in Section IX). Finally (Section X), it is shown that

an increase in pre-insurance income in any state decreases the critical marginal utility and therefore increases the post-insurance income in any insured state. The effect on the premium paid depends on whether the state for which pre-insurance income rose was insured or not; in the first case, the premium decreases; in the second, it increases.

ACKNOWLEDGMENTS

The author thanks Robert Shishko, of the Rand Corporation, for his comments, which were useful in improving the exposition.

CONTENTS

I. THE DEPENDENCE OF UTILITY ON THE STATE OF THE WORLD

In the now-customary formulation of decisionmaking under uncertainty, there is assumed to be a fixed set of possible states of the world. For present purposes, it suffices to assume that the set is finite. The states will be indexed by $s = 1,\ldots,S$. The probability of state s will be denoted by p_s. In the usual formulation of the expected-utility hypothesis, the decisionmaker is assumed to be confronted with a set of alternative specifications of consequences in each state of nature, say x_s, where x_s is a consequence if state s occurs. Then it is hypothesized that among alternative specifications of sets of consequences, the individual will choose the one that maximizes

$$\sum_s p_s U(x_s), \tag{1}$$

for some *utility function* U(x). In this formulation, the utility function does not depend on s. The hypothesis is that originally formulated by Daniel Bernoulli in 1783, with particular stress on the case where x_s is simply income in state s; the point of view is that of an individual whom the alternative states affect only through his income. In more modern discussions, from Ramsey (1931) on down, it is shown that hypothesis (1) can be deduced from some apparently convincing assumptions on rational behavior in the presence of uncertainty.

Nevertheless, in many applications, it seems more reasonable to have the utility function depend on s. This is likely to be the case where health is involved. The health of an individual clearly affects the utility of other consequences (for example, money income). If recovery from an illness depends on the availability of medical care, then the marginal utility of income in a state of illness will typically be higher than in a state of health. This is true in a perfectly operational and ordinalist sense, with no illegitimate comparisons of marginal utility; starting from a situation in which income is the same in the states of health and illness, the individual would choose to buy an actuarially fair insurance policy to be paid in event of illness, so

that at his preferred point, disposable income would be less in the healthy state than in the ill state.[1]

In my earlier paper (Arrow, 1963) I took this partially into account by supposing that an illness is equivalent to a reduction of income. That is, for each state s, there is a cash equivalent, $\bar{x}_s$, say, so that if x_s is the income in state resulting from some of the possible decisions, the individual maximizes

$$\sum_s p_s U(x_s - \bar{x}_s).$$

However, this now appears to be inadequately general although sometimes appropriate. It is indeed true that the marginal utility of income will be raised in the presence of illness. But there are two limitations: (1) There is, after all, no reason why the *shape* of the utility function under illness should be the same as that in good health at a lower level of income. More specifically, the risk aversion will in general be different in the two cases; it might be expected to be greater under illness if a minimum degree of medical care is essential but more has rapidly diminishing usefulness. (2) The simple cash-equivalent model implies that states of the world that are utility-reducing are also

[1] The theory of risk-bearing under uncertainty uses as a basic concept the *state of the world,* a complete description of the economically relevant aspects of the entire world. The illustrations of the last paragraph, on the other hand, refer only to the *state of the individual;* and indeed insurance, as we ordinarily understand it, is a schedule of payments contingent on events relating to the individual alone. As a general proposition, insurance policies contingent on the state of the individual alone are not optimal and may not even be feasible. Even though the insured cares only about his personal state, the insurer has to be concerned with the prospect of payments to other insured, which depend on their states; thus, it would clearly be possible to introduce a life insurance policy more favorable *ex ante* to both parties if the payment made for any one death were lower when a great many other individuals died. Nevertheless, in the common situation where the risks to different insured individuals are independent, and there are many such, it is possible to write policies depending only on the state of the insured without loss of efficiency, for then the state of the world, except insofar as it concerns the insured, is a statistical certainty. Hence, in the rest of this study, the "states" can be interpreted indifferently as states of the world and states of the insured.

marginal-utility-increasing. But in the case of illness, this is clearly not necessarily so. There are many states of invalidism in which medical care is of little use and the possibility of deriving satisfaction from consumption is small. Thus the marginal utility of income will be low rather than high in these states, even though they are very unsatisfactory and yield low utilities.

For these reasons, it seems to be more satisfactory to state the criterion for choice as the maximization of

$$\sum_s p_s U_s(x_s). \tag{2}$$

As far as I know, such an expression was first used by Eisner and Strotz (1961) in the context of life insurance. Clearly, being dead is rather different from being alive. An individual who takes out life insurance has some utility for money for his family after his death, but the shape of the utility function in that case might well be different from its shape when he is alive. The expression (2) has also been defended by Hirshleifer (1970, pp. 218, 220-221).

For those accustomed to the axiomatic treatment of Ramsey and his followers, there may be a bit of a puzzle here. Those axioms, which seem compelling enough, lead to (1), not to (2); that is, to a utility function that does not depend on the state of nature. Properly understood, there is no contradiction. The simplest way to look at it is to consider that (1) holds, provided that x_s is interpreted to include *all* the effects on the individual of being in state s. From the viewpoint of economic analysis, however, only *some* of these effects are relevant. More specifically, insurance can relate only to some but not to all components of the consequence vector. Thus, the fact of having fallen ill is not in itself something the economic system can affect; it can only alter the medical services and other purchasable commodities a person receives. If we assume that prices are unaffected by the state of the world, then all economic effects can be subsumed into changes in income. The illness itself must be regarded as unalterable by economic activity. Hence, the full consequences x_s can be written as a pair, (y_s, z_s), where y_s is income in state s, and z_s is an unalterable characteristic

of the state. Then (1) asserts that the individual strives to maximize

$$\sum_s p_s\, U(y_s,\ z_s),$$

but the only variables under his control are the y_s's. We can therefore write

$$U(x_s) = U(y_s,\ z_s) = U_s(y_s),$$

and then (1) is equivalent to the maximization of

$$\sum_s p_s\, U_s(y_s), \tag{3}$$

which is an operational equivalent of (2), with only the economic variables displayed.[1]

---

[1]Instead of reinterpreting the results of the usual axiom system, it would be possible to challenge it and derive (2) or (3) from an altered system. The guilty axiom is one that asserts a consequence possible under any one state of the world is possible under any other. Although the word "possible" here does not mean "feasible," it might be argued that some consequences are not only infeasible but inconceivable under some states of the world; if one is ill, one cannot be well. If this axiom is dropped, then only the weaker form (2) emerges.

It might be remarked here that the more general formulation (2) or (3) does offer some problems for the behavioral interpretation of probability. The expected-utility theorem or hypothesis, especially in conjunction with the Bayesian concept of subjective probability, implies the meaningful separability of tastes (as represented by the utility function) from beliefs (as represented by probabilities). But in the form (3), this separation is no longer operational. If we multiply any one function U_s by a positive constant, divide the corresponding probability p_s by the same amount, and then renormalize all the probabilities so that they still add up to one, the observed behavior of the individual is unchanged. Hence, no set of observations can distinguish the probabilities from the utilities. This point was made by Herman Rubin about 1964 in an axiomatization of behavior under uncertainty leading to a result of the form (2), presented to the Joint Berkeley-Stanford Mathematical Economic Seminar but unpublished.

II. OPTIMAL CHOICE OF INSURANCE POLICY OF GIVEN EXPECTED VALUE AND PREMIUM: THE CASE OF STATE-INDEPENDENT UTILITY

The choice of insurance policies by the insured can be assumed to be made as follows: For any given premium, the insurer specifies an expected value; the insured can then choose any policy with that expected value of payments. The premium is assumed, of course, to be at least equal to the expected value; the excess is a loading due to administrative costs and safety considerations for the insurer. The premium is assumed to be proportional to the expected value. The decision of the insured can then be analyzed into two steps: (1) For any given premium and given expected value, what is the optimal policy--that is, the optimal choice of insurance payments--in each state of nature? (2) Given a proportionality relation between premium and expected value, what is the optimal scale of the policy?

We assume further that insurance payments are always nonnegative.

In symbols, let

a_s = income in state s before the policy is chosen,

y_s = income in state s after policy is chosen,

$U_s(y_s)$ = utility of income in state s,

p_s = probability of state s,

i_s = insurance payment if state s occurs,

P = premium,

E = expected value of insurance payments,

α = E/P = benefit-premium ratio.

From these definitions,

$$y_s = a_s + i_s - P \tag{4}$$

$$E = \sum_s p_s \, i_s \tag{5}$$

and

$$E = \alpha P \tag{6}$$

where

$$0 \leq \alpha \leq 1. \tag{7}$$

The nonnegativity of insurance payments is written

$$i_s \geq 0. \tag{8}$$

Then the optimal insurance policy is obtained by choosing i_s, E, and P to maximize (3)

$$\sum_s p_s U_s(y_s),$$

subject to (4), (5), (6), and (8).

As usual, we postulate the existence of risk aversion, which is to say that the functions $U_s(y_s)$ are all strictly concave functions, although hereafter "strictly" will be understood.

$$U_s''(y_s) < 0, \text{ all } s \text{ and all } y_s. \tag{9}$$

As noted, we can break up the problem into two parts: (1) the maximization of (3) with respect to i_s subject to (4), (5) and (8), with E and P given, and (2) the choice of an optimal E and P subject to (6). In this and the following section the first problem is considered. In this section the additional assumption is made:

$$U_s(y_s) = U(y_s) \tag{10}$$

that utility is in fact independent of the state.

The problem of this section can then be written: Maximize

$$\sum_s p_s U(a_s - P + i_s)$$

subject to (5) and (8). In view of (9), the Kuhn-Tucker theorem assures us that the optimum is characterized by the conditions

$$p_s U'(a_s - P + i_s) \leqq \lambda p_s \tag{11a}$$

$$p_s U'(a_s - P + i_s) < \lambda p_s \text{ implies } i_s = 0 \tag{11b}$$

$$\sum_s p_s i_s = E, \tag{11c}$$

where λ is the marginal utility of income.

These conditions can be stated more simply by defining

$$S = \left\{ s \mid U'(a_s - P + i_s) = \lambda \right\}.$$

Then the optimum is characterized by the existence of a number λ and a set S such that

$$U'(a_s - P + i_s) = \lambda \qquad (s \in S) \tag{12a}$$

$$i_s = 0 \text{ for } s \notin S \tag{12b}$$

$$U'(a_s - P) < \lambda \text{ for } s \notin S \tag{12c}$$

$$\sum_{s \in S} p_s i_s = E \tag{12d}$$

as can be seen by dividing through by p_s in (11a) and (11b).

But the function $U(y)$ was assumed concave, so that U' is strictly decreasing and hence has a unique inverse, $(U')^{-1}(\lambda)$. If we write

$$V(\lambda) = (U')^{-1}(\lambda) + P,$$

then the conditions for an optimum can be written

$$a_s + i_s = V(\lambda) (s \in S) \tag{13a}$$

$$i_s = 0 \text{ for } s \notin S \tag{13b}$$

$$a_s > V(\lambda) \text{ for } s \notin S \tag{13c}$$

$$\sum_{s \varepsilon S} p_s i_s = E. \tag{13d}$$

The probability of the set of states, S, is given by

$$p(S) = \sum_{s \varepsilon S} p_s. \tag{14}$$

Multiply through in (13a) by p_s, sum over $s \varepsilon S$, and substitute from (13d) and (14).

$$\sum_{s \varepsilon S} p_s a_s + E = V(\lambda)\, p(S)$$

If we solve this last equation for $V(\lambda)$ and substitute into (13a) and (13c) we have

$$a_s + i_s = (\sum_{s \varepsilon S} p_s a_s + E)/p(S) \text{ for } s \varepsilon S \tag{15a}$$

$$a_s > (\sum_{s \varepsilon S} p_s a_s + E)/p(S) \text{ for } s \notin S \tag{15b}$$

Note several remarkable implications of this characterization of the optimal solution. Neither the premium, P, nor the utility function appears in it. The optimal distribution of protection is completely known once the expected value of the policy is known. Let

$$\bar{a} = (\sum_{s \varepsilon S} p_s a_s + E)/p(S). \tag{16}$$

Then, since $i_s \geqq 0$, it follows from (15a) that $a_s \leqq \bar{a}$ for $s \varepsilon S$, and from (15b), only for such s. Thus, as already argued (Arrow, 1963), the optimal policy is defined as full protection beyond a deductible $\bar{a}$, in the sense that whenever the income falls below $\bar{a}$, the insurance payments just suffice to restore it to that level, whereas if income is above $\bar{a}$,

insurance payments are zero. Of course, income in both insured and uninsured states is reduced by the premium P.

The optimal policy is then completely characterized by $\bar{a}$. As has just been seen, S is determined by $\bar{a}$; more specifically,

$$S = \left\{ s \mid a_s \leqq \bar{a} \right\} \tag{17}$$

and then,

$$\begin{aligned} i_s &= \bar{a} - a_s \text{ if } s \in S, \\ &= 0 \qquad \text{if } s \notin S. \end{aligned}$$

If S is defined in terms of $\bar{a}$ by (17), then (16) becomes an equation for $\bar{a}$. Multiply both sides by p(S), use its definition (14), and transpose. Then,

$$F(\bar{a}) \equiv (\sum_{s \in S} p_s)\, \bar{a} - \sum_{s \in S} p_s\, a_s = E. \tag{18}$$

In Appendix A, note that the function $F(\bar{a})$, the lefthand side of (18), is a strictly increasing continuous function, with $F(0) = 0$, $F(+\infty) = +\infty$. It follows that equation (18) has a unique solution in $\bar{a}$ for any given level of expected value E.

<u>Theorem</u> 1. Suppose that utility depends only on income, and the individual can choose any insurance policy with nonnegative payments of a specified expected value, the premium being specified also. Then the policy has the form of specifying a critical income level $\bar{a}$ (gross of premium) and paying out an amount necessary to bring gross income up to that level if it falls short. The critical income level is just such as to make the expected payments have the specified value and is independent of the utility function and of the premium.

<u>Remark</u>. One difficulty that has been evaded above is that it is in no way guaranteed that post-insurance income be nonnegative; if the premium P is high enough, then y_s can indeed be negative, even though $i_s \geqq 0$. For the time being, we will avoid this issue by assuming both

in this and in the following two sections that

$$a_s \geqq P, \text{ all } s. \tag{19}$$

Then (4) and (8) guarantee that $y_s \geqq 0$, all s, for any feasible policy. In Section VII below, this assumption will be relaxed.

Appendix A

To show that $F(\bar{a})$, as defined in (18) above, is strictly increasing and continuous, order the states s in increasing order of a_s, so that $a_1 < a_2 < \ldots < a_n$. Then S consists of the states $1, \ldots, r$ if and only if $a_r \leq \bar{a} < a_{r+1}$; S contains all states if $\bar{a} \geq a_n$ and no states (that is, there is no insurance) if $\bar{a} < a_1$. Therefore $F(\bar{a}) = 0$ if $\bar{a} < a_1$; if $\bar{a} > a_n$,

$$F(\bar{a}) = \left(\sum_{s=1}^{n} p_s\right) \bar{a} - \sum_{s=1}^{n} p_s a_s = \bar{a} - \sum_{s=1}^{n} p_s a_s,$$

which certainly approaches infinity as $\bar{a}$ approaches infinity. For $a_r \leq \bar{a} < a_{r+1}$,

$$F(\bar{a}) = \left(\sum_{s=1}^{r} p_s\right) \bar{a} - \sum_{s=1}^{r} p_s a_s, \tag{20}$$

which is a linear function with a positive slope; hence, $F(\bar{a})$ is strictly increasing and continuous as $\bar{a}$ increases within any of the intervals (a_r, a_{r+1}). It remains only to show that $F(\bar{a})$ is continuous at any point a_r. As $\bar{a}$ approaches a_{r+1} from below, (20) holds, so that $F(\bar{a})$ approaches

$$\left(\sum_{s=1}^{r} p_s\right) a_{r+1} - \sum_{s=1}^{r} p_s a_s. \tag{21}$$

As $\bar{a}$ approaches a_{r+1} from above, (20) holds with r replaced by r+1, so that

$$F(\bar{a}) = \left(\sum_{s=1}^{r+1} p_s\right) \bar{a} - \sum_{s=1}^{r+1} p_s a_s = \left(\sum_{s=1}^{r} p_s\right) \bar{a} - \sum_{s=1}^{r} p_s a_s + p_{r+1}(\bar{a} - a_{r+1}),$$

so that as $\bar{a}$ approaches a_{r+1} from above, $F(\bar{a})$ again approaches (21). Hence, $F(\bar{a})$ is continuous and increasing everywhere.

III. OPTIMAL CHOICE OF INSURANCE POLICY OF GIVEN EXPECTED VALUE AND PREMIUM: THE CASE OF STATE-DEPENDENT UTILITY

Utility is now permitted to vary with the state of nature, along the lines discussed in Section I. Following the general formulation given at the beginning of Section II, when E and P are specified, and the choice is only that of i_s, maximize (3) subject to (4), (5), and (8). The problem can then be stated as that of maximizing

$$\sum_s p_s \, U_s(a_s - P + i_s)$$

subject to

$$\sum_s p_s \, i_s = E \tag{5}$$

$$i_s \geqq 0. \tag{8}$$

In view of (9), the concavity of the functions U_s, the Kuhn-Tucker theorem permits characterization of the optimal insurance policy by

$$p_s \, U_s'(y_s) \leqq \lambda \, p_s \tag{22a}$$

$$p_s \, U_s'(y_s) < \lambda \, p_s \text{ implies } i_s = 0 \tag{22b}$$

and (5) above, where $y_s = a_s - P + i_s$, from (4). Divide through in (22a-b) by p_s, and let

$$S = \left\{ s \,\middle|\, U_s'(y_s) = \lambda \right\}. \tag{23}$$

Then (22a-b) assert

$$U_s'(y_s) < \lambda \text{ and } i_s = 0 \text{ for } s \notin S. \tag{24}$$

Note that λ is the maximum marginal utility of post-insurance income.

For simplicity of exposition, make a further assumption usually made in this context, which will be relaxed in Section VI below.

$$U_s'(0) = +\infty; \; U_s'(+\infty) = 0. \tag{25}$$

Since U_s' is strictly decreasing, it has a well-defined inverse, $(U_s')^{-1}(\lambda)$, for each s; this inverse is also strictly decreasing and, by (25), defined for all $\lambda \geq 0$. Then from (23) and (24),

$$y_s = (U_s')^{-1}(\lambda) \text{ for } s \in S$$

$$y_s > (U_s')^{-1}(\lambda) \text{ for } s \notin S.$$

Since $y_s = a_s - P + i_s$ for all s, $y_s = a_s - P$ for $s \notin S$, by (24). Write

$$V_s(\lambda, P) = (U_s')^{-1}(\lambda) + P. \tag{26}$$

Then the optimal insurance policy satisfies the conditions

$$a_s + i_s = V_s(\lambda, P) \text{ for } s \in S \tag{27a}$$

$$a_s > V_s(\lambda, P) \text{ for } s \notin S. \tag{27b}$$

Since $i_s \geq 0$, $a_s \leq V_s(\lambda, P)$ for $s \in S$, from (27a). Hence, from (27b), we can characterize S by

$$S = \left\{ s \mid a_s \leq V_s(\lambda, P) \right\}. \tag{28}$$

Thus, for a given P, the optimal policy is determined by the *critical marginal utility*, λ; this determines S by (28), and then, by (27a) and (24),

$$i_s = V_s(\lambda, P) - a_s \text{ for } s \in S \tag{29a}$$

$$= 0 \text{ for } s \notin S. \qquad (29b)$$

Because of the definition of S in (28), the condition $i_s \geq 0$ will be satisfied.

Thus, the quantity V_s can be interpreted as the *deductible limit* for the state s. Whereas in the case of state-independent utility there was a single deductible limit, now the limit should vary from state to state according to the schedule relating utility to income. As (28) and (29) show, the insurance is paid in those states for which income falls below the corresponding deductible limit.

It might be objected that this interpretation says very little, since both income and deductible limit vary from state to state, and the insurance payments are jointly determined by the optimization procedure. However, some further considerations show that V_s is indeed meaningfully regarded as a deductible limit. For one thing, suppose that there are several states in which the utility function is the same but in which incomes may differ. Thus there may be two states in which medical care is equally efficacious and other commodities equally enjoyable but in one of which income is high and the other in which it is low. Then V_s will be the same in all such states, and the insurance payments will be just those needed to bring the income up to the deductible limit if it falls below.

For another thing, consider two different situations, in which the list of the states and the utility function in each state are the same but the pre-insurance incomes differ; let them be a_s and a'_s in the two situations, respectively. Let the optimal policy in the first situation with an expected value E be defined by a critical marginal utility of income, λ, so that the insurance payments are defined by (29). Now in the second situation consider the insurance policy defined by the same deductible limits, V_s, and therefore by the same critical marginal utility, λ. This policy will be optimal if it satisfies the new budget constraint--that is, if the expected value of the policy changes to E', which will just permit the new insurance policy to be bought. More specifically, if we define

$$S' = \left\{ s \mid a'_s \leqq V_s(\lambda, P) \right\}$$

$$a'_s + i'_s = V_s(\lambda, P) \text{ for } s \in S$$

$$i'_s = 0 \text{ for } s \notin S$$

and if we choose E' so that

$$E' = \sum_s p_s \, i'_s$$

then the payments i'_s will be optimal if the expected value of the policy changes in the income-compensating manner from E to E'. Hence, given compensating changes in income, the original policy as defined by deductible limits remains optimal.

Note briefly how the critical marginal utility of income, λ, can be determined from the budget constraint (5). Since $i_s = 0$ for $s \notin S$, this can be written

$$\sum_{s \in S} p_s \, i_s = E.$$

Multiply through in (29a) by p_s and sum over $s \in S$.

$$E = \sum_{s \in S} p_s \, V_s(\lambda, P) - \sum_{s \in S} p_s \, a_s \equiv \phi(\lambda, P) \tag{30}$$

The existence of a unique solution to this equation is argued along lines similar to that of (18) in Appendix B.

<u>Theorem 2</u>. Suppose that utility of income depends upon the state of the world and that the individual can choose any insurance policy with nonnegative payments of a specified expected value, the premium being specified also. Suppose also that (1) the specified premium does not exceed pre-insurance income for any state, and (2) the marginal utility of income in any state decreases from $+\infty$ to 0 as income increases from 0 to $+\infty$. Then the policy has the form of stating a critical marginal utility of income and paying out an amount sufficient to

bring the marginal utility of income down to that level if the marginal utility of post-premium pre-insurance income were higher. The critical marginal utility is just such as to make the expected payments have the specified value.

Appendix B

For each s, the function V_s is strictly decreasing in λ. Hence, the equation

$$V_s(\lambda, P) = a_s$$

has a unique solution $\lambda = \lambda_s$ for a given P, namely, $\lambda_s = U_s'(a_s - P)$. This number is well-defined since we are assuming $a_s \geqq P$, by (19); however, it will be $+\infty$ if the equality holds. Number the states so that $\lambda_1 > \lambda_2 > \dots > \lambda_n$. Then if $\lambda_r \geqq \lambda > \lambda_{r+1}$, S consists of the states $1,\dots,r$, which remains unchanged so long as λ remains in that interval. Since V_s is strictly decreasing in λ for all $s \in S$, Φ is strictly decreasing in such an interval. As λ approaches λ_r from below, Φ approaches

$$\Phi(\lambda_r, P) = \sum_{s=1}^{r} p_s V_s(\lambda_r, P) - \sum_{s=1}^{r} p_s a_s$$

$$= \sum_{s=1}^{r-1} p_s V_s(\lambda_r, P) - \sum_{s=1}^{r-1} p_s a_s + p_r \left| V_r(\lambda_r, P) - a_r \right|$$

$$= \sum_{s=1}^{r-1} p_s V_s(\lambda_r, P) - \sum_{s=1}^{r-1} p_s a_s$$

from the definition of λ_r. If λ is slightly greater than λ_r, S consists of the states $1,\dots,r-1$, so that,

$$\Phi(\lambda, P) = \sum_{s=1}^{r-1} p_s V_s(\lambda, P) - \sum_{s=1}^{r-1} p_s a_s;$$

as λ approaches λ_r from above, this expression approaches

$$\sum_{s=1}^{r-1} p_s V_s(\lambda_s, P) - \sum_{s=1}^{r-1} p_s a_s = \Phi(\lambda_r, P),$$

which is also the limit from below, as we have seen. Hence, $\Phi(\lambda, P)$

is continuous at each of the points λ_r, while it is obviously continuous between these points. Since Φ is decreasing on all the intermediate intervals, it is decreasing everywhere in the interval $\lambda_1 \geqq \lambda \geqq \lambda_n$. If $\lambda > \lambda_1$, then S is an empty set of states, and therefore $\Phi(\lambda, P) = 0$.

If $a_s = P$, then $\lambda_s = + \infty$, and therefore we would have to let $\lambda_1 = + \infty$. In this case, we can, of course, have no $\lambda > \lambda_1$. But for λ very large, S consists of the single state 1, and therefore,

$$\Phi(\lambda, P) = p_1 \left| V_1(\lambda, P) - a_1 \right| = p_1 \left| (U_s')^{-1}(\lambda) + P - a_1 \right|$$

$$= p_1 (U_s')^{-1}(\lambda),$$

which approaches 0 as λ approaches $+ \infty = \lambda_1$. Hence it remains true that $\Phi(\lambda_1, P) = 0$. If $\lambda \leqq \lambda_n$, then again $\Phi(\lambda, P)$ is a linear combination of strictly decreasing functions and therefore is strictly decreasing. Further, as λ approaches 0, $(U_s')^{-1}(\lambda)$ approaches $+ \infty$ from (25), so that $V_s(\lambda, P)$ approaches $+ \infty$, and hence $\Phi(\lambda, P)$ approaches $+ \infty$. Therefore, Φ ranges from 0 to $+ \infty$ as λ decreases from λ_1 to 0, and so the equation (30) has a unique solution in λ for any given E.

IV. EFFECTS OF VARYING PREMIUM, EXPECTED VALUE, AND OTHER PARAMETERS

So far we have taken E and P as given. In the next section, the optimal choice of these variables will be discussed. As a preliminary, we discuss the effects of changes in several magnitudes taken as parametric in the preceding section, in particular E and P.

Broadly speaking, we can assume S constant. That is, the changes considered may be regarded as leaving the set of insured states unchanged; if the changes are sufficiently small, S will in fact not change, and therefore derivatives can be calculated on that assumption. Further, it is easy to demonstrate that the derivatives so calculated are continuous across the boundaries at which the set of insured states, S, does change.

Note that the policy is completely characterized by the critical marginal utility, λ. Thus, if any change does take place, it is comparatively easy to compute the effects on post-insurance income in the insured states, since this is determined by the equation $U_s'(y_s) = \lambda$, so that,

$$dy_s/d\lambda = 1/U_s''(y_s) = (1/\lambda)\ U_s'(y_s)/U_s''(y_s). \tag{31}$$

Now the ratio $-\ U_s''/U_s'(y_s)$ has been introduced into the theory of risk bearing under the name *absolute risk aversion*. Its reciprocal has been named *risk-tolerance*,

$$T_s = -\ U_s'(y_s)/U_s''(y_s). \tag{32}$$

A closely related concept is the *relative* or *proportionate risk aversion*,

$$R_{Rs} = -\ y_s\ U_s''(y_s)/U_s'(y_s) = y_s/T_s$$

(Pratt, 1964; Arrow, 1965, reprinted in Arrow, 1971, p. 94; Wilson, 1968, p. 120). From (31) and (32), we note

> The changes in post-insurance income in insured states due to some parameter shift are proportional to the risk-tolerances. (34)

This can be restated using (33):

> The proportionate changes in post-insurance incomes in insured states due to some parameter shift are inversely proportional to the relative risk aversions. (35)

Since the shifts in post-insurance incomes in insured states are completely determined by λ, it is useful to consider the variation in λ with respect to different parameters. First we consider the effect of changing P. For simplicity of notation, define

$$\partial V_s/\partial\lambda = V_s'.$$

Note that from (26), V_s' is independent of P and indeed $V_s = y_s + P$, so that, from (31) and (32),

$$V_s' = - T_s/\lambda. \tag{36}$$

Differentiate the basic equation (30), which determines λ, with respect to P.

$$\sum_{s\varepsilon S} p_s[V_s'(d\lambda/dP) + 1] = 0$$

so that, from (36),

$$d\lambda/dP = \lambda p(S)\Big/\sum_{s\varepsilon S} p_s\, T_s,$$

where $p(S) = \sum_{s\varepsilon S} p_s$, as defined in (14). For any $s \varepsilon S$, $p_s/p(S)$ is the conditional probability of the state s, given that one of the states in S has occurred. Hence, if numerator and denominator are both divided by p(S),

$$d\lambda/dP = \lambda/E(T_s|S) > 0, \tag{37}$$

where $E(T_s|S)$ means the conditional expectation of T_s, given that one of the insured states has occurred. Strictly speaking, this formula has been derived only when S is constant for sufficiently small variations in P, which is to say at any point for which λ differs from all of the λ_s's. But by continuity (easily demonstrated as was done earlier), (37) also holds as the juncture points λ_s.

Equations (31) and (37) together show the variation of y_s as P rises, E remaining fixed, for $s \in S$. Clearly, y_s must fall for all s. Since $i_s = y_s + P - a_s$, we can also infer the behavior of the insurance payments. As P increases, it can happen that one or more of the i_s's decreases and eventually becomes 0; for still larger P, that state s ceases to be insured. On the other hand, it is possible for some previously uninsured state (state not in S) to enter S as P becomes larger.

It may be worth noting that on the average desired insurance payments are not altered by changes in the premium. This can be seen by differentiating the budget constraint

$$\sum_{s \in S} p_s \, i_s = E$$

with respect to P, yielding

$$\sum_{s \in S} p_s (di_s/dp) = 0.$$

If we divide through by $p(S)$, we can say that $E(i_s|S) = 0$. Thus an increase in premium, so long as S does not change, simply reallocates insurance payments to those states with low risk-tolerances.

For the analysis of the next section, we will need to know the effects of a change in P on the maximum expected utility attainable by the individual. Then, let $W(E, P)$ be the expected utility attainable by following the optimal insurance policy for a given E and P.

$$W(E, P) = \sum_{s=1}^{n} p_s \, U_s(y_s), \tag{38}$$

where

$$y_s = a_s + i_s - P \text{ is optimal for given E and P.} \qquad (39)$$

Then we can calculate $\partial W/\partial P$ easily by using Samuelson's envelope theorem (1947, p. 39): The marginal effect on W of changing a parameter is the same whether the policy variables are optimally adjusted to the change or remain unchanged. Hence, we can compute $\partial W/\partial P$ by calculating the effect of a change in P, holding i_s constant. Since $\partial y_s/\partial P = -1$,

$$\partial W/\partial P = - \sum_{s=1}^{n} p_s \, U_s'(y_s) = - E[U_s'(y_s)]. \qquad (40)$$

We now turn to a similar calculation of the effects on policy and welfare of changing E, the expected value of insurance payments, with constant P. Again differentiate (30), now with respect to E, in order to find the effect on λ.

$$\sum_{s \varepsilon S} p_s \, V_s'(d\lambda/dE) = 1,$$

so that

$$d\lambda/dE = - \lambda/(\sum_{s \varepsilon S} p_s \, T_s) = - \lambda/p(S) \; E(T_s | S) < 0. \qquad (41)$$

Thus an increase in E (in effect, in income) increases post-insurance income and the amount of insurance bought in every insured state, but more in those for which the risk-tolerance is greatest. Clearly, as E increases, the insurance in any insured state never becomes zero, but previously uninsured states may become insured. To see the latter point, notice that $\lambda - U_s'(y_s) > 0$ for an uninsured state, with $y_s = a_s - P$. As E increases, y_s is constant and therefore so is $U_s'(y_s)$, while λ decreases in accordance with (41). If the difference becomes 0, s passes over into the class S of insured states.

The effect of a change in E on W is easily derived by noting that W is the result of maximization under a budget constraint, with E as the magnitude of the constraint. Then, as is well known, the rate of

change of the maximand with respect to the magnitude of the constraint is simply the Lagrange multiplier.

$$\partial W/\partial E = \lambda. \tag{42}$$

For later reference, it will also be useful to consider the effects of changes in the initial incomes, a_s. Consider any one of them, say a_t. Following now-familiar lines, differentiate (30) with respect to a_t.

$$\sum_{s\varepsilon S} p_s V_s'(\partial\lambda/\partial a_t) - \sum_{s\varepsilon S} p_s(\partial a_s/\partial a_t) = 0.$$

But $\partial a_s/\partial a_t = 1$ if $s = t$ and 0 otherwise, so that

$$\begin{aligned}\sum_{s\varepsilon S} p_s(\partial a_s/\partial a_t) &= p_t \text{ if } t \varepsilon S\\ &= 0 \text{ if } t \not\varepsilon S,\end{aligned}$$

and therefore,

$$\begin{aligned}\partial\lambda/\partial a_t &= p_t/(\sum_{s\varepsilon S} p_s V_s') = -\lambda p_t/(\sum_{s\varepsilon S} p_s T_s) \text{ if } t \varepsilon S\\ &= 0 \text{ if } t \not\varepsilon S.\end{aligned} \tag{43}$$

(43) can be given a slightly more unified appearance if we introduce the conditional probablity $p(t|S)$. Note that $p(t|S) = p(t)/p(S)$ if $t \varepsilon S$, $= 0$ if $t \not\varepsilon S$. If we divide numerator and denominator of the right-hand side of (43) for the case $t \varepsilon S$ through by $p(S)$, we have

$$\partial\lambda/\partial a_t = -\lambda p(t|S)/E(T_s|S), \tag{44}$$

which is also valid for the case $t \not\varepsilon S$.

Finally, the effect of a change in a_t on W can be found by use of the envelope theorem. Since $y_s = a_s + i_s - P$,

$$\partial W/\partial a_t = p_t U_t'(y_t). \tag{45}$$

V. CHOICE OF THE OPTIMUM SCALE OF AN INSURANCE POLICY

We now consider the individual to be able to choose the scale E of his policy as well as the allocation of payments to states for a given scale. Of course, he can increase E only at the cost of increasing premium P. We will assume a proportional relation,

$$E = \alpha P. \tag{46}$$

Thus α is the (expected) *benefit-premium ratio.*

For any given E and P, we are assuming the insurance payments i_s optimally determined. Hence, E and P should be chosen to maximize W(E, P) subject to (46). Note that W(E, P) is a concave function; for the maximand, $\sum_{s=1}^{n} p_s U_s(a_s + i_s - P)$, is jointly concave in the variables i_s, P.[1] Hence, the first-order conditions for maximization are both necessary and sufficient. We select P as our independent variable and assume E determined by (46).

---

[1]The general theorem, which can be found in many places, is the following: Let $f(x_1,\ldots,x_n, y_1,\ldots,y_m)$, $g_j(x_1,\ldots,x_n, y_1,\ldots,y_m)$ $(j = 1,\ldots,r)$ be each jointly concave in the *decision variables* $x_1,\ldots,x_n$ and the *parameters* $y_1,\ldots,y_m$. For any fixed set of values of $y_1,\ldots,y_m$, let $F(y_1,\ldots,y_m)$ be the maximum of $f(x_1,\ldots,x_n, y_1,\ldots,y_m)$ among all values of $x_1,\ldots,x_n$ satisfying the constraints $g_j(x_1,\ldots,x_n, y_1,\ldots,y_m) \geqq 0$. Then $F(y_1,\ldots,y_m)$ is a concave function.

In the present application, the decision variables are i_s $(s=1,\ldots,n)$, the parameters are E and P, the function being maximized is

$$\sum_{s=1}^{n} p_s U_s(a_s + i_s - P),$$

and the single constraint is (5), which can be written as

$$E - \sum_{s=1}^{n} p_s i_s \geqq 0.$$

$$dW/dP = (\partial W/\partial P) + \alpha(\partial W/\partial E) = \alpha\lambda - E[U_s'(y_s)] \tag{47}$$

from (40) and (42).

We will continue to maintain the assumptions (21), that $a_s \geqq P$, all s, and (25), that U_s' decreases from infinity to zero. Since P is a variable here, the first of these assumptions will be taken to hold at the optimal insurance policy. Both of these assumptions will be relaxed in the following sections.

First, what is the condition that there be no insurance? This is achieved when $dW/dP \leqq 0$ at $P = 0$. Of course, when $P = 0$, $E = 0$, by (46), and therefore $i_s = 0$, all s, so that $y_s = a_s$, all s. Also, in general,

$$\lambda = \max_s U_s'(y_s).$$

If we set $y_s = a_s$ in (47), we see that the condition that no insurance be taken out is that

$$\alpha \leqq E[U_s'(a_s)]/\max_s U_s'(a_s). \tag{48}$$

In words, no insurance at all is taken out if the benefit-premium ratio does not exceed the ratio of expected to maximum marginal utility of pre-insurance income.

Now suppose some insurance is taken out, so that the maximum of W occurs at an interior point, where $dW/dP = 0$.

$$\alpha\lambda = E[U_s'(y_s)]. \tag{49}$$

Then the optimum insurance can be determined as follows. For any given P, E is determined by (46), then λ from (30), and therefore i_s by (29). By substitution into (49), the optimum P is determined.

Will there be insurance in all states? In that case, $U_s'(y_s) = \lambda$ in all states; (49) becomes $\alpha\lambda = \lambda$, or $\alpha = 1$. Thus, there will be some uninsured states unless the policy is actuarially fair. Conversely, if the policy is actuarially fair, optimal choice will in general call for

insurance in all states. To see this, first note that there must be some insurance if in the pre-insurance state the marginal utility varies from state to state (if it were initially constant, then, of course, insurance would be superfluous), for then the right-hand side of (48) is less than one, and therefore (48) cannot hold when $\alpha = 1$. There must then be an interior maximum, so that (49) holds. At the optimum, $U_s'(y_s) \leq \lambda$, all s, so that $E[U_s'(y_s)] \leq \lambda$, and the inequality is strict unless $U_s'(y_s) = \lambda$ for all s. Thus for $\alpha = 1$, (49) implies that S must consist of all states.

The value of y_s for each s is determined by the condition $U_s'(y_s) = \lambda$. We can find the appropriate value of λ by solving (30). Here S consists of all states, and E = P. Since $V_s(\lambda, P) = (U_s')^{-1}(\lambda) + P$, by (26), equation (30) reduces to

$$\sum_s p_s \left[(U_s')^{-1}(\lambda) - a_s\right] = 0. \tag{50}$$

The left-hand side is a strictly decreasing function of λ; further, since U_s' has been assumed to decrease from infinity to zero, it follows that $(U_s')^{-1}(\lambda)$ decreases from infinity to zero as λ increases from zero to infinity. Hence, the left-hand side of (30) decreases from infinity to $-\sum_s p_s a_s < 0$, so that (50) always has a unique positive solution.

There is a trivial nonuniqueness in the choice of optimal policy in the actuarially fair case. Let i_s be an optimal policy with premium P, and let $i_s' = i_s + h$, $P' = P + h$, $h \geq 0$. Under the assumption of actuarial fairness, the new policy is feasible. If y_s and y_s' are the net incomes in state s under the original and new policies, we see that $y_s' = a_s + i_s' - P' = a_s + i_s - P = y_s$, all s, so the new policy is also optimal. Indeed, it is possible to choose $h < 0$, so long as the condition $i_s' \geq 0$ is satisfied. That is, for any optimal policy,

$$P \geq \max_s \left[a_s - (U_s')^{-1}(\lambda)\right] \tag{51}$$

and we can choose indifferently any such P. The condition we are imposing, that $a_s \geq P$, all s, can be satisfied provided

$$\min_s a_s \geq \max_s \left| a_s - (U_s')^{-1}(\lambda) \right| . \tag{52}$$

In the actuarially unfair case, $\alpha < 1$, the probability of being in a state covered by insurance is less than 1. In fact, (49) implies a simple inequality, if it is recalled that $U_s'(y_s) = \lambda$ for $s \in S$, that $U_s'(y_s) > 0$, all s, and that there is at least one state not in S.

$$\alpha\lambda = \sum_s p_s U_s'(y_s) = \sum_{s\in S} p_s U_s'(y_s) + \sum_{s\notin S} p_s U_s'(y_s) > \lambda \sum_{s\in S} p_s = \lambda p(S)$$

so that

$$p(S) < \alpha \text{ if } \alpha < 1. \tag{53}$$

It will be recalled that we have assumed that the optimal solutions found satisfy condition (19), that $a_s \geq P$, all s. To complete the analysis, we wish to show when this condition holds. Let $\bar{P} = \min_s a_s$; then (19) is equivalent to the statement $P \leq \bar{P}$. For any such P, there is a corresponding choice of λ that satisfies (30). Then an optimal policy is defined by that $P \leq \bar{P}$ for which $dW/dP = 0$. Since W is concave, dW/dP is decreasing. If (48) does not hold, then $dW/dP > 0$ for $P = 0$; hence, $dW/dP = 0$ for some $P \leq \bar{P}$ if and only if $dW/dP \leq 0$ for $P = \bar{P}$. In view of (46), this can be stated as follows: Define $\bar{\lambda}$ to satisfy the equation

$$\Phi(\bar{\lambda}, \bar{P}) = \alpha\bar{P}; \tag{54}$$

then there is an optimal policy with $P \leq \bar{P}$ if, and only if,

$$\alpha\bar{\lambda} \leq E[U_s'(\bar{y}_s)], \tag{55}$$

where

$$\bar{y}_s = a_s + \bar{i}_s - \bar{P},\ \bar{i}_s \text{ is optimal insurance when } P = \bar{P},\ E = \alpha\bar{P}. \tag{56}$$

The last condition reduces to (52) in the case $\alpha = 1$.

As noted earlier, following (24), in the optimal policy for given P, with $E = \alpha P$, the maximum marginal utility of post-insurance income is λ. If $y_s(P)$ is the post-insurance income for the optimal policy for given P, then (48), (49), and (55) can all be stated in terms of the ratio of expected to maximum marginal utility of post-insurance income--that is,

$$R(P) = E[U_s'(y_s(P))]/\max_s U_s'(P)).$$

The condition (48) for no insurance becomes

$$\alpha \leqq R(0);$$

if any insurance is purchased, the amount is defined by (49),

$$R(P) = \alpha;$$

and the condition for a solution with $a_s \geqq P$, all s, becomes

$$\alpha \leqq R(\bar{P}).$$

Theorem 3. Suppose that the utility of income depends upon the state of the world and that the individual can choose any insurance policy with nonnegative payments with a given ratio, $\alpha \leqq 1$, of expected benefits to premium. Assume further that the marginal utility of income in any state decreases from $+\infty$ to 0 as income decreases from 0 to $+\infty$. Let $\bar{P} = \min_s a_s$. For any $P \leqq \bar{P}$, let R(P) be the ratio of expected to maximum marginal utility of post-insurance income according to the optimal policy specified in Theorem 2 for premium P and expected benefits αP.

(a) No insurance is taken out if and only if the expected benefit-premium ratio does not exceed the ratio of expected to maximum marginal utility of pre-insurance income--that is, $\alpha \leqq R(0)$.

(b) If $R(0) < \alpha \leqq R(\bar{P})$, then there is an optimal policy with premium $P \leqq \bar{P}$, which is optimal in the sense of Theorem 2 for that premium and expected benefits αP, and which satisfies the condition $R(P) = \alpha$.

(c) If the offering is not actuarially fair, so that $\alpha < 1$, then the probability of being in a state covered by the optimal policy is less than α.

(d) If the offering is actuarially fair ($\alpha = 1$), then the policy of equalizing the marginal utility of post-insurance income in all states at a level just insuring actuarial fairness (expected value of pre-insurance income equals expected value of post-insurance income) is optimal.

Remark. Strictly speaking, part (d) has been proved only under the additional condition (52) that the minimum pre-insurance income is at least equal to the maximum decrease from pre- to post-insurance income. But, as will be shown in Section VII below, the condition is in fact superfluous.

VI. THE OPTIMUM INSURANCE POLICY WITHOUT RESTRICTIONS ON THE RANGE OF MARGINAL UTILITIES

It would be desirable to remove two restrictions in the previous analysis. The first, to be discussed in this section, is that the marginal utility of income in each state varies from $+\infty$ down to 0; the second, to be covered in the following section, is that the choice of optimum premium has so far been restricted so that it does not exceed minimum pre-insurance income.

The first is probably of less consequence, since the assumption being removed is usually reasonable. The condition that the marginal utility approach 0 as income approaches infinity is, of course, implied if one assumed that utility functions are bounded, as some assert (Arrow, 1971, p. 69). Even apart from this argument, which is strongly contested by some, the idea that the marginal utility approaches a positive rather than a zero limit seems unreasonable. It would imply that the individual would be essentially risk-neutral with respect to fluctuations at high income levels. But this would imply that wealthy individuals hold no safe assets; indeed, they would hold only the riskiest assets with the highest expected values.

The assumption at the other end, that marginal utility becomes infinite as income approaches zero, is more disputable in certain contexts. Consider the case where one of the states is death. In other words, we are imagining a policy that includes life insurance as one aspect. Income still has utility if the individual has the desire to leave a bequest. But if the intended heir has other wealth, human or material, the marginal utility of the bequest at zero should not be infinite.

Since, however, the restrictions on the range of the marginal utilities can be removed completely, without excessive complications to the analysis, we will do so here.

Since $U_s'(y_s)$ is strictly decreasing, its range is in general bounded above by $\bar{\lambda}_s = \lim_{y_s \to 0} U_s'(y_s)$ and below by $\underline{\lambda}_s = \lim_{y_s \to +\infty} U_s'(y_s)$, where $\bar{\lambda}_s$

is either $+\infty$ or a finite positive number $\underline{\lambda}_s \geq 0$. Then $(U_s')^{-1}(\lambda)$ is defined only for $\bar{\lambda}_s \geq \lambda \geq \underline{\lambda}_s$; at the lower limit, the "definition" is somewhat metaphorical, since $(U_s')^{-1}(\underline{\lambda}_s) = +\infty$.

Let us first review the reasoning of Section IV, that is, we seek the optimal policy for given premium and expected benefits. The analysis through (24) remains unchanged. Since $U_s'(y_s) < \lambda$ for $s \notin S$, $\lambda > \underline{\lambda}_s$. However, for any given λ, it is possible that $\lambda > \bar{\lambda}_s$, so that $(U_s')^{-1}(\lambda)$ would be undefined. Then, for $s \notin S$, we have

$$\text{either } y_s > (U_s')^{-1}(\lambda) \qquad \text{or } \lambda > \bar{\lambda}_s.$$

In either case, $i_s = 0$, so that $y_s = a_s - P$. We are assuming $a_s \geq P$. Hence, if we extend (26) by defining

$$V_s(\lambda, P) = (U_s')^{-1}(\lambda) + P \text{ if } \underline{\lambda}_s \leq \lambda \leq \bar{\lambda}_s,$$

$$= P \text{ if } \lambda > \bar{\lambda}_s,$$

we can conclude $a_s \geq V_s(\lambda, P)$ with the strict inequality if either $a_s > P$ or $\lambda \leq \bar{\lambda}_s$. Thus the characterization of S in (28) should be slightly altered, to exclude the case where both $a_s = P$ and $\lambda > \bar{\lambda}_s$; however, this alteration is unimportant for defining the optimal policy, since we would have $i_s = 0$, according to (29a), even if we had included that state in S.

The existence of a unique solution to (30) remains valid but requires some alteration in the algorithm. Since $(U_s')^{-1}(\underline{\lambda}_s) = +\infty$, we must have $\lambda > \underline{\lambda}_s$ for all the states in S, or, equivalently,

$$\lambda > \max_{s \in S} \underline{\lambda}_s.$$

We define $\lambda_s = U_s'(a_s - P)$ as before, if, in particular, there is a state for which $a_s = P$, then $\lambda_s = \bar{\lambda}_s$. In any case, $\bar{\lambda}_s \geq \lambda_s \geq \underline{\lambda}_s$, all s. As before we number the states in decreasing order of λ_s. Now define $\tilde{\lambda}_s$ to be the largest of the numbers $\underline{\lambda}_1, \ldots, \underline{\lambda}_s$. Note that

$$\tilde{\lambda}_s = \max\,(\tilde{\lambda}_{s-1},\ \underline{\lambda}_s).$$

Let p be the smallest number s, if any, for which $\tilde{\lambda}_s \geqq \lambda_{s+1}$.

First we note that

$$\lambda_p > \tilde{\lambda}_p \geqq \lambda_{p+1}. \tag{57}$$

To see this, first suppose p = 1. But certainly $\lambda_1 > \underline{\lambda}_1 = \tilde{\lambda}_1$, by definition. Now suppose p > 1. Then by construction $\lambda_p > \tilde{\lambda}_{p-1}$. Since $\tilde{\lambda}_p = \max\,(\tilde{\lambda}_{p-1},\ \underline{\lambda}_p)$ and $\underline{\lambda}_p < \lambda_p$, $\tilde{\lambda}_p < \lambda_p$. The second inequality in (57) is immediate from the definition of p.

Now let λ decrease from λ_r to λ_{r+1}, with r < p. Then $\lambda_{r+1} \geqq \lambda_p > \tilde{\lambda}_p$, so that $\lambda > \underline{\lambda}_s$, s = 1,...,p, and, in particular, s = 1,...,r. Since $\lambda \leqq \lambda_r \leqq \lambda_s \leqq \bar{\lambda}_s$ (s = 1,...,r), $\underline{\lambda}_s < \lambda \leqq \bar{\lambda}_s$ (s = 1,...,r), so that $(U_s')^{-1}(\lambda)$ is defined for s = 1,...,r, and therefore $\phi(\lambda, P)$ is well defined and increases as λ decreases. However, $\tilde{\lambda}_p = \underline{\lambda}_s$ for some $s \leqq p$. Hence, as λ decreases from λ_p to $\tilde{\lambda}_p$, it must be true that $(U_s')^{-1}(\lambda)$ approaches infinity for at least one $s \leqq p$, and hence $\phi(\lambda, P)$ approaches infinity as λ approaches $\tilde{\lambda}_p$ from above.

If p is not defined, then $\lambda_{s+1} > \tilde{\lambda}_s$ (s = 1,...,n - 1). If we set s = n - 1, we have $\lambda_n > \tilde{\lambda}_{n-1}$, and therefore, by an argument just used, $\lambda_n > \tilde{\lambda}_n$. Then as λ decreases from λ_n to $\tilde{\lambda}_n$, $\phi(\lambda, P)$ approaches $+\infty$. Thus $\phi(\lambda, P)$ is strictly decreasing in λ, approaches $+\infty$ as λ approaches $\tilde{\lambda}_p$ or $\tilde{\lambda}_n$ (according as p is defined or not), and approaches 0 as λ approaches λ_1, as before. Thus (30) always has a solution.

It follows that Theorem 2 remains valid when the hypothesis (b) there is deleted. However, the algorithm for finding the optimal policy has to be modified as indicated.

(In the situation studied in Section II, where the utility function is independent of the state of the world, no use was made of the hypothesis that marginal utility declined from infinity to zero, and therefore Theorem 1 remains valid. This can also be seen from the just preceding reasoning. For then $\bar{\lambda}_s$ and $\underline{\lambda}_s$ are independent of s, say equal

to $\bar{\lambda}$ and $\underline{\lambda}$, respectively; then $\tilde{\lambda}_s$ is also equal to $\underline{\lambda}$ for all s. It can then be seen that $\lambda_n > \underline{\lambda} = \tilde{\lambda}_n$, so that the algorithm of Section III remains valid; but for the special case of identical utility functions in all states, this is the same as the algorithm of Section II.)

The steps leading from Theorem 2 to Theorem 3 did not make any use of the range of the marginal utilities, and therefore Theorem 3 also remains valid with no hypothesis on the range of the marginal utilities.

VII. OPTIMUM INSURANCE POLICIES WITH PREMIUM WAIVERS NEEDED

We now reconsider the optimum policy but remove the requirement (19) that the premium never exceed pre-insurance income. Clearly if (19) fails to hold, we must be prepared for the possiblity that in some states the insured could not afford to pay his premium without insurance. He therefore must have an insurance payment in such states simply to meet his premium obligations. We might refer to such payments as constituting *premium waivers*, to stretch an ordinary insurance term somewhat. That is, to the constraint that insurance payments be nonnegative, we add the constraint that the post-insurance income be nonnegative, $y_s \geqq 0$. Since $y_s = a_s + i_s - P$, this condition can be written, for given P, as

$$i_s \geqq P - a_s.$$

Since we also require $i_s \geqq 0$, we can write

$$i_s \geqq \max(P - a_s, 0). \tag{58}$$

We can then reconsider the problem of Section IV, the choice of an optimum policy for given P and E, with the nonnegativity constraint (8) replaced by (58): we seek to maximize

$$\sum_s p_s U_s(y_s)$$

subject to (58) and the budget constraint

$$\sum_s p_s i_s = E. \tag{5}$$

The simplest procedure turns out to be a reinterpretation of the variables to permit the direct application of Theorem 2. Let

$$b_s = a_s + \max(P - a_s, 0) \tag{59}$$

$$j_s = i_s - \max (P - a_s, 0). \tag{60}$$

Note that

$$y_s = a_s + i_s - P = b_s + j_s - P \tag{61}$$

and that the constraint (58) can be written

$$j_s \geqq 0. \tag{62}$$

Finally, from (5) and (60), the budget constraint can be written

$$\sum_s p_s j_s = F, \tag{63}$$

where

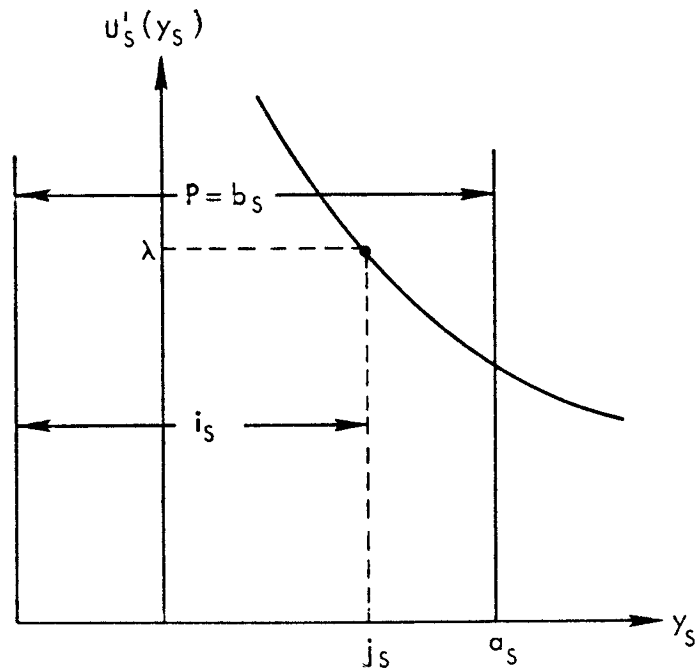
$$F = E - \sum_s p_s \max (P - a_s, 0). \tag{64}$$

We have to assume that $F \geqq 0$ for the problem to be feasible. Then the optimization problem is identical to that of Section IV for given E and P, with the parameters a_s being replaced by b_s and the variables i_s by j_s. (See figure on next page for the case $a_s > P$.) We can thus, as before, find an optimal policy defined by a critical marginal utility of income. Certain interpretive remarks can be made.

1. The set S is now the set of states for which $j_s > 0$. From (60), we see that the total insurance payment can be written as the sum of a *premium waiver*, max $(P - a_s, 0)$, and a *discretionary insurance payment*, j_s. If we write

$$Q = \{s \mid a_s < P\}, \tag{65}$$

we see that Q is the set of states in which premium waivers are paid, while S is the set of states in which discretionary insurance payments are made. In general, there is no necessary relation between these two sets. The set of states for which some insurance payment is made



is the *union* of these two sets, denoted by $S \cup Q$, namely the set of states that are in at least one of S and Q.

2. However, if we assume $U_s'(0) = +\infty$ for all s, then Q must be a subset of S. For if $s \in Q$ but $s \notin S$, then $b_s = a_s + (P - a_s) = P$, while $j_s = 0$, so that $y_s = 0$. For any s, $U_s'(y_s) \leq \lambda$, which is impossible if $y_s = 0$ and $U_s'(0) = +\infty$.

Theorem 4. Suppose that utility of income depends upon the state of the world and that the individual can choose any insurance policy with nonnegative payments of a specified expected value, the premium being specified also. Let the *premium waiver* for any state be the difference between premium and pre-insurance income if positive, and zero otherwise. Suppose further that the expected benefits are at least equal to the expected premium waiver. Then the optimal policy has the form of stating a critical marginal utility of income and paying (a) exactly the premium waiver if pre-insurance income falls short of the premium and the marginal utility at zero income does not exceed the critical marginal utility, (b) the amount needed to be added to post-premium pre-insurance income to bring the marginal utility to the critical level if the marginal utility at zero is greater than the critical

level and if the post-premium pre-insurance income is either negative or has a marginal utility higher than the critical, and (c) nothing in other states. The critical marginal utility is just such as to make the expected payments have the specified value.

If the marginal utility of zero income is always infinite, then possibility (a) is ruled out for all states.

In the special case where the utility functions in all states of the world are the same, it can be seen that (a) cannot occur. For if $U_s'(0) \leqq \lambda$ for some s, the same holds for all s when utility is independent of state. In that case, $U'(y_s) < \lambda$ when $y_s > 0$, which would mean that no insurance is paid at all, other than premium waivers. If expected benefits exceed expected premium waivers, this could hardly be optimal. Hence, Theorem 1 remains valid as stated.

Now we consider the full problem of optimization, with premium variable and expected benefits a prescribed fraction, α, of the premium. As before, we approach this through the intermediate problem of studying the effects of varying E and P. The transformed problem is slightly more complicated than before, because a change in P has not only the direct effect studied before but also effects through the fact that in part it determines the values of b_s and of F, as can be seen from (59) and (64). Let the maximum utility obtained in the transformed problem with parameters F, P, $b_1,\ldots,b_n$ be denoted by $\bar{W}(F, P, b_1,\ldots,b_n)$, while the maximum utility can also be expressed in terms of the original parameters as W(E, P), so that

$$W(E, P) \equiv \bar{W}(F, P, b_1,\ldots,b_n). \tag{66}$$

In view of (65), we can write (59) and (64) as

$$\begin{aligned} b_s &= P \text{ if } s \in Q \\ &= a_s \text{ if } s \notin Q \end{aligned}$$

$$F = E - \sum_{s \in Q} p_s (P - a_s) = E + \sum_{s \in Q} p_s a_s - P\, p(Q),$$

and therefore

$$db_s/dP = 1 \text{ if } s \in Q$$

$$= 0 \text{ if } s \notin Q \qquad (67)$$

$$\partial F/\partial P = -\, p(Q), \ \partial F/\partial E = 1. \qquad (68)$$

It should be noted, though, that db_s/dP and $\partial F/\partial P$ are not continuous functions of P. Specifically, as P increases through a value a_s, for some s, both of these magnitudes change discontinuously, the first changing from 0 to 1, the second decreasing by p_s when P passes beyond a_s, so that s is added to Q.

A change in E operates only through a change in F; hence from (66) and the second half of (68),

$$\partial W/\partial E = (\partial \bar{W}/\partial F)(\partial F/\partial E) = \partial \bar{W}/\partial F = \lambda \qquad (69)$$

by (42) applied to the transformed problem.

From (66),

$$\partial W/\partial P = (\partial \bar{W}/\partial F)(\partial F/\partial P) + (\partial \bar{W}/\partial P) + \sum_s (\partial \bar{W}/\partial b_s)(db_s/dP).$$

But from (40) and (45), applied to the transformed problem,

$$\partial \bar{W}/\partial P = -\sum_s p_s\, U_s'(y_s), \ \partial \bar{W}/\partial b_s = p_s\, U_s'(y_s).$$

Hence, with the aid of (67) through (69),

$$\partial W/\partial P = -\,\lambda\, p(Q) - \sum_s p_s\, U_s'(y_s) + \sum_{s \in Q} p_s\, U_s'(y_s)$$

$$= -\,\lambda\, p(Q) - \sum_{s \notin Q} p_s\, U_s'(y_s).$$

Among the states not in Q, consider separately those in S and those not in S. Let

$$M = S \cup Q, \tag{70}$$

that is, those states in either S or Q. For all states in S, and in particular those in $S \sim Q$ (that is, those in S but not in Q), $U_s'(y_s) = \lambda$. Those states in neither S nor Q are precisely those not in M.

$$\sum_{s \notin Q} p_s\, U_s'(y_s) = \lambda\, p(S \sim Q) + \sum_{s \notin M} p_s\, U_s'(y_s).$$

Hence,

$$\partial W/\partial P = -\,\lambda[p(Q) + p(S \sim Q)] - \sum_{s \notin M} p_s\, U_s'(y_s).$$

Since the sets Q and $S \sim Q$ are disjoint and their union is M, $p(Q) + p(S \sim Q) = p(M)$.

$$\partial W/\partial P = -\,\lambda\, p(M) - \sum_{s \notin M} p_s\, U_s'(y_s). \tag{71}$$

It is easy to see that, as before, W is a concave function of E and P. Hence, if we vary P with $E = \alpha P$, the total derivative is non-increasing. From (69) and (71),

$$dW/dP = [\alpha - p(M)]\lambda - \sum_{s \notin M} p_s\, U_s'(y_s). \tag{72}$$

We now parallel the discussion in Section V. For P sufficiently small, specifically when $P < \min_s a_s$, the set Q of waiver states is empty. The condition that there be no insurance at all therefore remains the same as before, as given by (48).

Can there be insurance in all states? Note that M is the set of insured states, whether through waivers or through discretionary insurance or both. Hence, the question is whether, at an optimum, M consists of all states. If it did, the second term in (72) vanishes, while

$p(M) = 1$. Clearly if $\alpha < 1$, then $dW/dP < 0$ at such a point, implying that it is not optimal. Therefore, as before, complete coverage is not optimal if the offering is actuarially unfair. The converse is also true, as can be seen rather trivially by raising all payments and the premium by the same amount, preserving post-insurance incomes. Then every state can be made into a waiver state. This is possible because the budget constraint for the actuarially fair case,

$$\sum_s p_s i_s = P, \tag{73}$$

will remain valid under the simultaneous and equal increase of insurance payments and premium. The optimal policy in the actuarially fair case is determined in a somewhat more general way than described in Section V, since we are no longer assuming that $U_s'(0) = +\infty$. What we do is choose y_s to maximize

$$\sum_s p_s U_s(y_s),$$

subject to $y_s \geqq 0$, and

$$\sum_s p_s y_s = \sum_s p_s a_s.$$

Then choose i_s, P so that $i_s = P - a_s + y_s$, the desired post-insurance incomes are achieved, and P is sufficiently large so that $i_s \geqq 0$, all s. The budget constraint (73) will then automatically be satisfied.

When $\alpha < 1$, there will be an interior maximum. In general, this will occur when $dW/dP = 0$, but since dW/dP is not continuous at values of P equal to some a_s, we have only the weaker condition that $dW/dP \geqq 0$ for P slightly smaller, $dW/dP \leqq 0$ for P slightly larger. If we define

$$R_1(P) = p(M) + [\sum_{s \neq M} p_s U_s'(y_s)/\lambda], \tag{74}$$

we can say that optimal P is determined by the condition

$$R_1(P) = \alpha$$

at any point of continuity of $R_1(P)$, and

$$R_1(P - 0) \leqq \alpha \leqq R_1(P + 0)$$

at a point of discontinuity, where $R_1(P - 0)$ means the limit of R_1 as P is approached from the left, and $R_1(P + 0)$ the limit as P is approached from the right.

A point of discontinuity of $R_1(P)$ can occur only when M changes. However, a change in M without a change in Q occurs at a point P not equal to any of the initial incomes a_s; there must therefore be a change in S. If a state s is added to S, then $p(M)$ is increased by p_s, while the term in brackets is reduced by $p_s\, U'_s(y_s)/\lambda$. But when a state not in S for smaller P enters into S, it must be true, by continuity, that $U'_s(y_s) = \lambda$, and we have an increase of p_s balanced by an equal decrease, and therefore no discontinuity. Hence, a discontinuity in $R_1(P)$ can occur only at a point of change in Q, that is, when P reaches one of the values a_t. But even in this case, if $t \in S$ at $P = a_t$, there would be no discontinuity. The only discontinuities possible then occur at points a_t for which $t \notin S$--that is, for which $U'_t(y_t) < \lambda$. Clearly, by continuity, if $t \notin S$ at $P = a_t$, $t \notin S$ in a neighborhood of a_t. Also, by the definition of Q, (65), $t \notin Q$ for $P \leqq a_t$, $t \in Q$ for $P > a_t$. Hence, $t \notin M$ for $P \leqq a_t$, $t \in M$ for $P > a_t$. Therefore, for all possible discontinuities,

$$R_1(P - 0) = p(M) + [p_t\, U'_t(y_t)/\lambda] + [\sum_{\substack{s \notin M \\ s \neq t}} p_s\, U'_s(y_s)/\lambda] = R_1(P)$$

$$R_1(P + 0) = p(M) + p_t + [\sum_{\substack{s \notin M \\ s \neq t}} p_s\, U'_s(y_s)/\lambda].$$

(In these formulas, the set M is understood to be that at $P = a_t$.)

Thus the premium level P is optimal if one of the following two conditions holds:

$$R_1(P) = \alpha \tag{75}$$

$$R_1(P) \leqq \alpha \leqq p(M) + p_t + [\sum_{\substack{s \notin M \\ s \neq t}} p_s U_s'(y_s)/\lambda], \ t \notin S, \text{ where}$$

$$S \text{ and } M \text{ are those corresponding to } P = a_t. \tag{76}$$

From either (75) or (76), $R_1(P) \leqq \alpha < 1$. From (74), it must be that $p(M) < 1$, so that there are some states s not in M, and therefore the second term of $R_1(P)$ is positive. The inequality $R_1(P) \leqq \alpha$ then implies

$$p(M) < \alpha \text{ if } \alpha < 1. \tag{77}$$

It will be helpful to give something of an interpretation of $R_1(P)$. M can be partitioned into the two sets S and $Q \sim S$, so that $p(M) = p(S) + p(Q \sim S)$. Also $U_s'(y_s) = \lambda$ for $s \in S$, so that

$$\lambda p(S) = \sum_{s \in S} p_s U_s'(y_s).$$

If $s \in Q \sim S$, then $P - a_s > 0$, while $j_s = 0$, so that

$$i_s = j_s + \max (P - a_s, 0) = P - a_s,$$

and therefore $y_s = a_s + i_s - P = 0$. At the same time, $\lambda > U_s'(y_s) = U_s'(0)$.

$$\lambda p(Q \sim S) = \sum_{s \in Q \sim S} p_s[\lambda - U_s'(0) + U_s'(y_s)]$$

$$= \sum_{s \in Q \sim S} p_s[\lambda - U_s'(0)] + \sum_{s \in Q \sim S} p_s U_s'(y_s)$$

$$= p(Q \sim S)\{\lambda - E[U_s'(0) | Q \sim S]\} + \sum_{s \in Q \sim S} p_s U_s'(y_s).$$

Substitution into (74) yields

$$R_1(P) = R(P) + p(Q \sim S) \left\{ 1 - E[U_s'(0)\ Q \sim S]/\lambda \right\},$$

where R(P) is defined in Theorem 3 as the ratio of expected to maximum marginal utility of income at the optimum policy. From an earlier remark recall that if $U_s'(0) = +\infty$ for all s, then Q is a subset of S, so that $Q \sim S$ can have no elements, and $R_1(P) = R(P)$, which is continuous, so that the optimal P is determined by (75).

Finally, is the policy that satisfies (75) or (76) in fact feasible? It may be recalled that one of the hypotheses of Theorem 4 was that expected benefits, which are αP, be at least equal to expected premium waivers. In the notation used, we want to make sure that $F \geqq 0$. But if $P = 0$, so that $E = \alpha P = 0$, and certainly $F = 0$. Further, as P increases, with $E = \alpha P$, the total derivative of F is given by

$$dF/dP = (\partial F/\partial P) + \alpha(\partial F/\partial E) = \alpha - p(Q)$$

from (68). But for P not exceeding the optimum $dW/dP \geqq 0$, which in turn is equivalent to $R_1(P) \leqq \alpha$, and therefore implies $p(M) < \alpha$. Since Q is a subset of M, $p(Q) \leqq p(M) < \alpha$, so that $dF/dP > 0$ at all P up to and indeed somewhat beyond the optimum P. Since $F = 0$ when $P = 0$, $F > 0$ throughout this interval, and the P chosen as optimum by (75) or (76) is indeed feasible.

Theorem 5. Suppose that the utility of income depends in general upon the state of the world and that the individual can choose any insurance policy with nonnegative payments with a given ratio, $\alpha \leqq 1$, of expected benefits to premium. For any P, consider the optimal policy defined in Theorem 4 for premium P and expected benefits αP. For that policy, let R(P) be the ratio of expected to maximum marginal utility of post-insurance income, $\lambda(P)$ be the critical marginal utility, and Q the set of states defined in Theorem 4(a), those for which the insurance payment is positive but exactly equals the premium waiver, and,

$$R_1(P) = R(P) + p(Q) \left\{ 1 - E[U_s'(0) | Q]/\lambda(P) \right\}.$$

(a) No insurance is taken out if and only if the expected benefit-premium ratio does not exceed the ratio of expected to maximum marginal utility of pre-insurance income--that is, $\alpha \leqq R(0)$.

(b) If $R(0) < \alpha$, then the optimal premium level always exists and is defined by the condition

$$R_1(P) \leqq \alpha \leqq R_1(P + 0).$$

(c) The function $R_1(P)$ is monotone increasing. Hence, if $R_1(P) = \alpha$, condition (b) certainly holds. Also, a discontinuity at P can occur only if both (a) $P = a_s$, the pre-insurance income, for some state s, and (b) $U_s'(0) < \lambda(P)$ when $P = a_s$. Hence, for any other P, the condition $R_1(P) = \alpha$ is necessary as well as sufficient that P be optimal.

(d) If any of the following conditions hold, then in every state where a premium waiver is paid, the insurance payment exceeds the premium waiver (Q is a subset of S), and therefore $R_1(P) = R(P)$ at the optimum: (a) $\alpha \leqq R(\min_s a_s)$; (b) the utility function for income is independent of the state of the world; (c) $U_s'(0) = +\infty$, all s. Since R(P) is increasing and continuous, the optimality condition would then be simply $R(P) = \alpha$.

(e) If the offering is not actuarially fair, so that $\alpha < 1$, then the probability of being in a state for which the insurance payment is positive is less than α.

(f) If the offering is actuarially fair ($\alpha = 1$), then the optimal policy is defined as the set of post-insurance incomes that maximizes expected utility subject to the condition of actuarial fairness (expected value of pre-insurance income equals expected value of post-insurance income). These post-insurance incomes can be realized with nonnegative insurance payments by raising the premium sufficiently high.

VIII. COMPARATIVE STATICS: THE EFFECT OF CHANGING THE BENEFIT-PREMIUM RATIO

It was shown how to determine the optimal insurance policy when the expected benefit-premium is some given α. It is of interest to observe how the optimal policy will change with changes in that parameter. To avoid unnecessary complications, we will essentially return to the situation of Theorem 3. That is, we ignore the possibility of premium waivers; or, to be more precise, in any state in which insurance is paid, it is assumed that the amount paid exceeds the premium waiver. Alternative conditions for validity of this assumption are given in Theorem 5(d).

The optimal policy for any given α is completely characterized by two parameters, the critical marginal utility of income, λ, and the premium, P. These can be regarded as defined by the two equations, (30), with $E = \alpha P$, and (49), provided, of course $\alpha > R(0)$, so that we are in the interior case. These equations are rewritten somewhat.

In (49), recall that $U_s'(y_s) = \lambda$ for $s \varepsilon S$. Then (49), or, equivalently, the equation $R(P) = \alpha$, can be written

$$\alpha\lambda = \sum_{s\varepsilon S} p_s\, U_s'(y_s) + \sum_{s\notin S} p_s\, U_s'(y_s) = \lambda p(S) + \sum_{s\notin S} p_s\, U_s'(y_s),$$

or,

$$[\alpha - p(S)]\lambda = \sum_{s\notin S} p_s\, U_s'(y_s). \tag{78}$$

In (30), substitute the definition of $V_s(\lambda, P)$ from (26).

$$\sum_{s\varepsilon S} p_s[(U_s')^{-1}(\lambda) + P - a_s] = \alpha P$$

or,

$$\sum_{s\varepsilon S} p_s[(U_s')^{-1}(\lambda) - a_s] + p(S)\ P = \alpha P,$$

or, finally,

$$\sum_{s \varepsilon S} p_s[(U_s')^{-1}(\lambda) - a_s] = [\alpha - p(S)]\ P. \tag{79}$$

Note that from Theorem 5(e) or Theorem 3(c), $\alpha - p(S) > 0$. We will study the dependence of λ and P on α through (78)-(79) over an interval in which S is constant. Since λ and P are continuous functions of α even at points where S changes, this analysis will give us a correct qualitative picture for all changes. For S fixed, let

$$\beta_S = \alpha - p(S) > 0 \tag{80}$$

$$\psi_S(P) = \sum_{s \notin S} p_s\ U_s'(y_s) \tag{81}$$

$$\phi_S(\lambda) = \sum_{s \varepsilon S} p_s[(U_s')^{-1}(\lambda) - a_s]. \tag{82}$$

Since $y_s = a_s - P$ for $s \notin S$, it is indeed true that $\psi_S(P)$ is a function of P alone (and not of λ or α, for S fixed). Since U_s' is a decreasing function, it follows from (81) and (82) that

$$\psi_S'(P) > 0 \tag{83}$$

$$\phi_S'(\lambda) < 0. \tag{84}$$

Equations (78) and (79) can be written in the compact form,

$$\beta_S \lambda - \psi_S(P) = 0 \tag{85}$$

$$\phi_S(\lambda) - \beta_S\ P = 0. \tag{86}$$

As long as S is fixed, a change in α is equivalent to a change in β_S. Hence differentiate the equations (85)-(86) with respect to β_S.

$$\beta_S(d\lambda/d\alpha) - \psi_S'(P)(dP/d\alpha) = -\lambda \tag{87}$$

$$\Phi'_S(\lambda)(d\lambda/d\alpha) - \beta_S(dP/d\alpha) = P \tag{88}$$

Equations (87)-(88) constitute a pair of linear equations in the comparative statics relations $d\lambda/d\alpha$ and $dP/d\alpha$. The determinant of this system is easily seen to be

$$D = -\beta_S^2 + \psi'_S(P)\ \Phi'_S(\lambda) < 0 \tag{89}$$

from (83)-(84). Solving by Cramer's rule yields

$$d\lambda/d\alpha = [\lambda\beta_S + P\psi'_S(P)]/D \tag{90}$$

$$dP/d\alpha = [\beta_S P + \lambda\Phi'_S(\lambda)]/D. \tag{91}$$

From (80), (83), and (89), $d\lambda/d\alpha < 0$ unequivocally. Hence, since $U'_s(y_s) = \lambda$ for $s \varepsilon S$, y_s will be increasing for all insured states.

The variation of P with respect to α is not completely defined as to sign by purely theoretical considerations. We know, of course, that for α sufficiently small, $P = 0$, while for $\alpha = 1$, P can be indefinitely large. Hence, broadly speaking, the premium demanded will increase with the ratio of expected benefits to premiums. However, there could in principle be intervals in which an increase in α is accompanied by a decrease in P, though only under unlikely conditions.

We first note that, from (80) and (84), the two terms in the numerator of (91) are of opposite signs. From (89), the sign of $dP/d\alpha$ is opposite to that of this numerator, which we now proceed to interpret. From (31) and (32), we see that

$$\lambda\ \Phi'_S(\lambda) = -\sum_{s\varepsilon S} p_s\ T_s.$$

From (80) and the budget equation,

$$\beta_S\ P = \alpha P - P\ p(S) = \sum_{s\varepsilon S} p_s(i_s - P).$$

Hence,

$$dP/d\alpha > 0 \text{ if and only if } \sum_{s\varepsilon S} p_s(T_s - i_s + P) > 0. \tag{92}$$

Certainly a sufficient condition for (92) to hold is that $T_s > i_s - P$ for all insured states s. Under the assumption of risk aversion, $T_s > 0$, so that certainly $T_s > i_s - P$ when $i_s \leq P$. For the case $i_s > P$ (the states where there is insurance beyond that needed to cover the premium), it is convenient to use the concept of relative risk aversion, introduced in (33). Then $T_s > i_s - P$ if and only if

$$R_{Rs}(y_s) < y_s/(i_s - P) = 1 + [a_s/(i_s - P)]. \tag{93}$$

Now if insurance is relatively minor compared with incomes, the right-hand side of (93) is very large, and the inequality is likely to hold. It can fail to hold, however, for states in which the insurance payment is large compared with pre-insurance income and for which the relative risk aversion is large. This can happen in the medical context. Consider a state of illness in which the marginal utility of income is high up to some relatively large figure and then drops off sharply; further income has little value for either medical or nonmedical purposes. Then the right-hand side may not be much above one while the left-hand side is large. But if this situation holds only for a set of states of relatively low total probability, we may expect (92) to hold even if (93) fails for a few states.

It is certainly true that (93) holds if relative risk aversion never exceeds one, but this seems to be a strong assumption.

If the marginal utility of income is independent of the state, then y_s, $T_s(y_s)$, and $R_{Rs}(y_s)$ are the same for all insured states. From (92), by an argument like that leading to (93), a necessary and sufficient condition for $dP/d\alpha > 0$ is that

$$R_R(y) < 1 + [E(a_s|S)/E(i_s - P|S)].$$

It is easy to see that (92) can fail to hold. Consider, for example, the case where S contains a single element. Then it is only necessary to construct a case in which the inequality in (93) is reversed, while (78) and (79) hold. Hence, as remarked, the pure theory does not exclude the possibility that P falls as α increases for some ranges. In those ranges, note that $y_s = a_s - P$ is increasing in the noninsured states, while in the standard case of increasing P, the individual is getting worse off in the noninsured states. In those states, the ratio $U_s'(y_s)/\lambda$ is less than 1, but the numerator is increasing if P increases, and the denominator is decreasing, so that eventually such a state moves into the insured category. Similarly, in an insured state, $y_s = a_s + i_s - P$ is increasing. If P is increasing, i_s must certainly be increasing and can never fall to zero, so that a state, once insured, cannot subsequently become uninsured in the standard case.

If, however, P decreases, it is conceivable that in an insured state i_s may fall to zero, and a state that is insured at one value becomes uninsured at a higher value. However, the post-insurance income associated with any state must be higher at higher α than it was when insured. Let $y_s(\alpha)$ and $\lambda(\alpha)$ be post-insurance income in state s and critical marginal utility as functions of α. Suppose state s is insured at some benefit-premium ratio α, and let $\alpha' > \alpha$. Then,

$$U_s'[y_s(\alpha)] = \lambda(\alpha), \; U_s'[y_s(\alpha')] \leqq \lambda(\alpha') < \lambda(\alpha)$$

so that $y_s(\alpha') > y_s(\alpha)$.

<u>Theorem 6</u>. Consider the optimal policy defined in Theorem 5 for any given expected benefit-premium ratio α. Assume that the premium waiver, if paid at all, is always smaller than the insurance payment (as would be true under any of the hypotheses of Theorem 5(d)). Then the critical marginal utility of income, λ, decreases as α increases, and therefore the post-insurance income for any insured state increases with α so long as that state remains insured. If it ceases to be insured at some higher value of α, it still must be true that post-premium income exceeds the post-insurance income at any lower benefit-premium ratio for which that state was insured.

The premium is zero for sufficiently small α and can be regarded as indefinitely large when $\alpha = 1$. It is not necessarily monotonic increasing with α. It increases if and only if

$$E(T_s - i_s + P|S) > 0.$$

A sufficient condition that P increase with α is that the relative risk aversion for any insured state s at income y_s be smaller than $1 + [a_s/(i_s - P)]$.

If the utility of income is independent of the state, then the necessary and sufficient condition for premium to increase with benefit-premium ratio is that the relative risk aversion in any insured state (it is the same for all such states) not exceed $1 + [E(a_s|S)/E(i_s - P|S]$.

IX. COMPARATIVE STATICS: THE EFFECT OF CHANGING PROBABILITIES

The benefit-premium ratio is not the only parameter of the optimal choice of insurance policy. The probabilities of the different states are also parameters, and we may consider how the optimal policy would be different if the parameters were different. Such effect may be useful to study for several reasons: It may be that, with changing knowledge or changing circumstances, such as alterations of medical techniques or public health measures, the probabilities of different states change; or one might want to know the sensitivity of the policy to errors in estimating the probabilities.

As in the preceding section, the basic technique is the differentiation of equations (85)-(86), defining the optimal policy with respect to the parameter under study, in this case p_s, the probability of state s. There is a slight complication; the parameters p_s being probabilities, their sum must always be equal to one. Hence, one probability cannot be changed without changing one or more others. In what follows, it will be understood that if p_s is altered, then all other probabilities $p_t (t \neq s)$ are changed so as to keep their values relative to each other, while insuring that the sum of the probabilities adds up correctly.

From a purely formal viewpoint, we can consider the parameters p_s in the insurance optimization to be independent parameters, if we ignore their interpretation as probabilities. It will be mathematically useful to follow this interpretation as a step in deriving the effects corrected by having other probabilities change appropriately. That is, we first consider the solution to the problem stated in (3)-(8), of maximizing

$$\sum_s p_s \, U_s(y_s) \tag{3}$$

subject to

$$\sum_s p_s \, i_s = \alpha P$$

as a function of the parameters p_s, taken to be independent. As we know, a solution can be completely characterized by the choice of the critical marginal utility, λ, and the premium, P, and we consider the variation of these two magnitudes with respect to the variables p_s. From these derivatives, in turn, we derive the derivatives when a change in any one p_s is offset by changes in others to preserve the sum at unity.

The first step, then, is to differentiate (85)-(86) with respect to p_s. The result depends on whether $s \in S$ or not. First suppose $s \in S$. In (85), p_s only appears in the factor $\beta_S = \alpha - p(S) = \alpha - \sum_{s \in S} p_s$. Hence, differentiation of (85) with respect to $p_s (s \in S)$ yields

$$\beta_S(\partial\lambda/\partial p_s) - \psi_S'(\partial P/\partial p_s) = \lambda (s \in S).$$

In (86), p_s appears in β_S and also, from (82), in one term of the sum defining ψ_S. In the latter, p_s is multiplied by

$$(U_s')^{-1}(\lambda) - a_s = y_s - a_s = i_s - P.$$

Hence, differentiation of (86) with respect to p_s yields,

$$\psi_S'(\partial\lambda/\partial p_s) - \beta_S(\partial P/\partial p_s) = -(y_s - a_s) - P (s \in S).$$

We can solve these two equations $\partial\lambda/\partial p_s$ and $\partial P/\partial p_s$ by Cramer's rule. By comparison with (87)-(90), it is easy to see that

$$\partial\lambda/\partial p_s = -(d\lambda/d\alpha) - (y_s - a_s)\,\psi_S'/D \ (s \in S) \tag{94a}$$

$$\partial P/\partial p_s = -(dP/d\alpha) - (y_s - a_s)\,\beta_S/D \ (s \in S). \tag{94b}$$

Now differentiate with respect to $p_s (s \notin S)$. In (85), p_s enters only in the term ψ_S, where it has coefficient U_s'.

$$\beta_S(\partial\lambda/\partial p_s) - \psi_S'(\partial P/\partial p_s) = U_s' \ (s \notin S).$$

In (86), p_s does not appear at all for $s \notin S$.

$$\Phi_S'(\partial\lambda/\partial p_s) - \beta_S(\partial P/\partial p_s) = 0 \quad (s \notin S).$$

Solution yields

$$\partial\lambda/\partial p_s = - \beta_S \, U_s'/D \quad (s \notin S) \tag{95a}$$

$$\partial P/\partial p_s = - U_s' \, \Phi_S'/D \quad (s \notin S). \tag{95b}$$

The next step is to adjust the effect of a change in one probability for the fact that others have to change simultaneously. Write

$$q_s = 1, \; q_t = p_t/(1 - p_s) \text{ for } t \neq s.$$

Then the condition that the relative values of $p_t (t \neq s)$ remain constant is equivalent to the condition that q_t remain constant, since $1 - p_s = \sum_{t \neq s} p_t$. Then

$$p_s = q_s \, p_s, \; p_t = q_t(1 - p_s) \text{ for } t \neq s.$$

For constant q_t's, we can consider the total derivative of λ with respect to p_s; call it λ_s'.

$$\begin{aligned} \lambda_s' &= \sum_t (\partial\lambda/\partial p_t)(dp_t/dp_s) = q_s(\partial\lambda/\partial p_s) - \sum_{t \neq s} q_t(\partial\lambda/\partial p_t) \\ &= (\partial\lambda/\partial p_s) - [1/(1 - p_s)] \sum_{t \neq s} p_t(\partial\lambda/\partial p_t) \\ &= [1/(1 - p_s)][(\partial\lambda/\partial p_s) - \sum_t p_t(\partial\lambda/\partial p_t)]. \end{aligned}$$

This result assumes slightly simpler form if the independent variable is taken to be not p_s, but its transform, $-\ln(1 - p_s)$. Let

$$\lambda_s = d\lambda/d[-\ln(1 - p_s)] = (1 - p_s)\,\lambda_s'$$

$$= (\partial\lambda/\partial p_s) - \sum_t p_t\ (\partial\lambda/\partial p_t). \tag{96}$$

Note that $-\ln(1 - p_s)$ is monotone increasing in p_s. Hence, the sign of λ_s indicates whether an increase in p_s with compensating proportional changes in all other probabilities increases or decreases λ. λ_s is the effect on λ of a given proportionate decrease in the probability that s will not occur.

One can calculate the term $\sum_t p_t(\partial\lambda/\partial p_t)$ from (94a) and (95b), but in fact it is easier and more revealing to argue more directly. The term in question is clearly the effect on λ of a shift in all p_s's in the same proportion. Such a shift has two effects: it multiplies the maximand (3) by a constant, and it multiplies the left-hand side of the budget constraint

$$\sum_t p_t\ i_t = \alpha P$$

by a constant. The first shift has no effect at all on the choice of the maximizing variables and, in particular, on λ. The second has exactly the same effect as a decrease of α in the same proportion. Hence,

$$\sum_t p_t(\partial\lambda/\partial p_t) = -\ \alpha(d\lambda/d\alpha). \tag{97}$$

We can therefore calculate λ_s by substituting (97) and (94a) or (95a) into (96).

$$\lambda_s = -\ [(1 - \alpha)(d\lambda/d\alpha) + (y_s - a_s)\ \psi'_S/D]\ (s \in S) \tag{98a}$$

$$\lambda_s = \alpha(d\lambda/d\alpha) - (\beta_S\ U'_s/D)\ (s \notin S). \tag{98b}$$

If we substitute (90) into (98a) and simplify, we find

$$\lambda_S = (-1/D) \left\{ (1 - \alpha)\lambda\beta_S + [(1 - \alpha)\ P + (i_s - P)]\ \psi'_S \right\} (s \in S) \tag{99}$$

From (89), $D < 0$; from (83), $\psi_S' > 0$, so that certainly $\lambda_s > 0$ if $(1 - \alpha) P + (i_s - P) = i_s - \alpha P \geqq 0$. Since an increase in λ implies a decrease in y_s for all $s \in S$,

> $i_s \geqq \alpha P$ implies that an increase in the probability of s, all other probabilities changing proportionately, leads to a fall in post-insurance income in every insured state. (100)

Note further that the right-hand side of (99) is an increasing function of i_s. Since $i_s \geqq 0$ for insured states,

$$\lambda_s \geqq (-1/D)[(1 - \alpha)\lambda\beta_S - \alpha P\psi_S'] \quad (s \in S). \tag{101}$$

If the right-hand side of (101) is positive, then certainly $\lambda_s > 0$ for all $s \in S$. We can also say that there is a critical level of i_s such that among insured states, $\lambda_s > 0$ if and only if i_s exceeds that level; this critical level is, of course, a function of all the parameters of the problem. Given the parameters, and therefore the value of P, we can equivalently say there is a critical level of i_s/P with the stated property. This critical level may be zero and, from (100), must be less than α.

Now substitute (90) into (98b) and simplify to find

$$\lambda_s = (-1/D)[(U_s' - \alpha\lambda)\beta_S - \alpha P\psi_S'] \quad (s \notin S). \tag{102}$$

We immediately note that $U_s' \leqq \alpha\lambda$ implies $\lambda_s < 0$. Also the right-hand side of (102) is increasing in U_s', so that there is a critical level of U_s' such that $\lambda_s < 0$ if and only if U_s' is below that limit. We can speak equivalently of a critical level for $(U_s'/\lambda) - 1$, and then we see that this critical level must be greater than $\alpha - 1$. Finally, since $U_s' \leqq \lambda$ for all $s \notin S$, it follows from (102) that

$$\lambda_s \leqq (-1/D)[1 - \alpha)\lambda\beta_S - \alpha P\psi_S'] \quad (s \notin S). \tag{103}$$

If the right-hand side of (101) and (103) is positive, then $\lambda_s > 0$ for all $s \in S$ and for all $s \notin S$ for which $(U_s'/\lambda) - 1$ is sufficiently close to zero. If that right-hand side is negative, then $\lambda_s < 0$ for all $s \notin S$ and for all $s \in S$ for which i_s/P is sufficiently small. Define therefore

$$k_s = i_s/P \text{ for } s \in S$$

$$= (U_s'/\lambda) - 1 \text{ for } s \notin S.$$

Note that $k_s \geqq 0$ for $s \in S$, $k_s \leqq 0$ for $s \notin S$. Then all the preceding can be summed up by saying that there is a critical level, k_o, such that $\lambda_s > 0$ if $k_s > k_o$, $\lambda_s < 0$ if $k_s < k_o$. Further, k_o may be positive or negative or zero but is necessarily bounded by

$$\alpha - 1 < k_o < \alpha.$$

The effects of changing probabilities on the premium P can be studied similarly. The following analogues of (96) and (97) hold:

$$P_s = dP/D[-\ln(1 - p_s)] = (\partial P/\partial p_s) - E(\partial P/\partial p_s),$$

$$E(\partial P/\partial p_s) = - \alpha(dP/d\alpha).$$

Substitute from (94b) and (91).

$$P_s = - [(1 - \alpha)(dP/d\alpha)] - \beta_S(y_s - a_s)/D$$

$$= (-1/D)[(1 - \alpha)\lambda\Phi_S' + \beta_S(i_s - \alpha P) \quad (s \in S). \qquad (104)$$

From (84), $\Phi_S' < 0$. Hence, $P_s < 0$ if $i_s \leqq \alpha P$. Since P_s increases with i_s, we can say, by reasoning like that just used, that there is a critical level, k_1, such that for insured states, $P_s > 0$ if $i_s/P > k_1$, $P_s < 0$ if $i_s/P < k_1$, and $k_1 > \alpha$.

From (95b) and (91),

$$P_s = \alpha(dP/d\alpha) - (U_s' \Phi_S/D) = (-1/D)[(U_s' - \alpha\lambda)\Phi_S' - \alpha\beta_S P] \quad (s \notin S). \quad (105)$$

If $U_s' \geqq \alpha\lambda$, then $P_s < 0$. Since P_s decreases as U_s' increases, it follows that there is a critical level, k_2, such that for uninsured states $P_s < 0$ if $(U_s'/\lambda) - 1 > k_2$, $P_s > 0$ in the opposite case; further, $k_2 < \alpha - 1$. Hence, we can summarize by saying that there are two critical levels, k_1 and k_2, such that $P_s < 0$ if $k_2 < k_s < k_1$, $P_s > 0$ if $k_s > k_1$ or $k_s < k_2$; further $k_1 > \alpha$, $k_2 < \alpha - 1$.

It can also be seen from (105), since $U_s' > 0$, that $P_s \leqq \alpha(dP/d\alpha)$ for $s \notin S$; hence, if $dP/d\alpha < 0$, $P_s < 0$ for all $s \notin S$ so that $k_2 = -1$.

The actuarially fair case ($\alpha = 1$) has not been covered by the discussion to this point. In this case, as we know, the optimal premium is essentially indeterminate, and all attention is concentrated on λ, which is determined by (50). Clearly, multiplication of all coefficient p_s by a common factor leaves the solution, λ, unchanged, so that

$$\sum_s p_s(\partial\lambda/\partial p_s) = 0.$$

Then, from (96), $\lambda_s = \partial\lambda/\partial p_s$. Differentiation of (50) yields

$$[\sum_t (p_t/U_t'')](\partial\lambda/\partial p_s) = -(i_s - P).$$

In this case, $\partial\lambda/\partial p_s$ has the same sign as $i_s - P$, so that $k_o = 1$.

There is one curious implication of this remark: If a state is net insured (that is, $i_s > P$) for some set of probabilities, it will remain net insured if its probability alters, all other probabilities changing in proportion. To see this, suppose state s is net insured for some set of probabilities. If p_s falls, λ falls, so that i_s rises, and $i_s - P$ increases, the state becoming even more heavily net insured. If p_s rises, i_s indeed falls. But it cannot fall to a value below P, for if it did, λ would start falling, and therefore i_s could not fall. Another way of seeing this is to note that if $i_s = P$ at some set of probabilities, then λ must be constant for any change in p_s. If we

write (50) as

$$p_s(i_s - P) + \sum_{t \neq s} p_t(y_t - a_t) = 0,$$

then, since the first term is zero, the second term is also. If p_s changes, and the p_t's change proportionately ($t \neq s$), then if λ remains constant, all y_t's remain constant, both terms remain zero, and therefore the equation remains satisfied. Hence, if $i_s > P$ (or $i_s < P$) for some set of probabilities, no change in p_s alone can bring i_s into equality with P.

Theorem 7. Consider the optimal policy defined in Theorem 5. Assume that the premium waiver, if paid at all, is always smaller than the insurance payment (as would be true under any of the hypotheses of Theorem 5(d)). In the following, when it is asserted that the probability of a state rises, it is understood that the probabilities of all remaining states fall in proportion to each other so as to preserve the sum of probabilities at 1. Let $k_s = i_s/P$ if s is an insured state, $= [U_s'(y_s)/\lambda] - 1$ if s is an uninsured state.

(a) There is a critical level, k_o, for which $\alpha - 1 < k_o < \alpha$, such that if $k_s > k_o$, then post-insurance incomes in all insured states fall if p_s increases, while if $k_s < k_o$, they rise with an increase in p_s.

(b) In the actuarially fair case, all states are insured and $k_o = 1$, so that post-insurance incomes fall with an increase in the probability of a net insured state (one for which the insurance payment exceeds the premium) and rise with the increase in the probability of a state with negative net insurance.

(c) There are two critical levels, k_1 and k_2, such that the premium paid increases with the probability of any state with either $k_s > k_1$ or $k_s < k_2$, where $k_1 > \alpha$, $k_2 < \alpha - 1$, and decreases with an increase in the probability of any state for which $k_1 < k_s < k_2$.

(d) Statements (a) and (c) can be partly restated as follows: Among insured states, an increase in the probability of any state for which the insurance payment is at least equal to the expected insurance payment decreases post-insurance incomes in all insured states, while an increase in the probability of any state for which the insurance

payment does not exceed the expected insurance payment decreases the premium paid. Among uninsured states, an increase in the probability of any state for which the ratio of the marginal utility of post-premium income to the critical marginal utility is at least equal to the benefit-premium ratio decreases the premium paid, while an increase in the probability of a state where the ratio of marginal utility of post-premium income to the critical marginal utility does not exceed the benefit-premium ratio increases post-insurance income in every insured state.

(e) If the premium decreases as the benefit-premium ratio increases, then the premium decreases with an increase in the probability of any uninsured state.

(f) In the actuarially fair case, any change in the probability of any given state s leaves unchanged the sign of $i_s - P$.

Remark. For completeness, I record the exact expressions for k_o, k_1, and k_2 implicit in the preceding analysis, with some substitutions from (85) and (86).

$$k_o = \alpha + (1 - \alpha)\ E(U_s'|\tilde{S})/PE(U_s''|\tilde{S}) \text{ if positive,}$$

$$= \alpha - 1 - \alpha P\ E(U_s''|\tilde{S})/E(U_s'|\tilde{S}) \text{ if negative,}$$

$$k_1 = \alpha + (1 - \alpha)\ E(T_s|S)/E(y_s - a_s|S),$$

$$k_2 = \alpha - 1 - \alpha\ E(y_s - a_s|S)/E(T_s|S).$$

X. COMPARATIVE STATICS: THE EFFECT OF CHANGING INITIAL INCOMES

To complete the discussion, consider the one remaining set of parameters, the initial incomes a_s. Again, differentiate expressions (85) and (86). First consider a_s for $s \varepsilon S$. This parameter does not enter (85) at all, so that its differentiation leads to

$$\beta_S(\partial\lambda/\partial a_s) - \psi_S'(\partial P/\partial a_s) = 0.$$

But a_s does appear in Φ_S, and $\partial\Phi_S/\partial a_s = - p_s$. Differentiation of (86) yields

$$\Phi_S'(\partial\lambda/\partial a_s) - \beta_S(\partial P/\partial a_s) = p_s,$$

so that we may easily calculate

$$\partial\lambda/\partial a_s = p_s \, \psi_S'/D < 0 \ (s \ \varepsilon \ S)$$

$$\partial P/\partial a_s = p_s \, \beta_S/D < 0 \ (s \ \varepsilon \ S).$$

If $s \notin S$, then a_s appears in ψ_S, since $y_s = a_s - P$, so that $\partial\psi_S/\partial a_s = p_s \, U_s''$; a_s does not appear at all in (86). Hence differentiation yields

$$\beta_S(\partial\lambda/\partial a_s) - \psi_S'(\partial P/\partial a_s) = p_s \, U_s''$$

$$\Phi_S'(\partial\lambda/\partial a_s \ - \beta_S(\partial P/\partial a_s) = 0,$$

which can be solved to obtain

$$\partial\lambda/\partial a_s = - p_s \, U_s'' \, \beta_S/D < 0 \ (s \notin S)$$

$$\partial P/\partial a_s = - p_s \, U_s'' \, \Phi_S'/D > 0 \ (s \notin S).$$

Clearly, $y_t = a_t - P$ falls as a_s rises in any uninsured state other than s. What about $y_s = a_s - P$?

$$\begin{aligned}
\partial y_s/\partial a_s &= 1 - (\partial P/\partial a_s) = (D + p_s\, U''_s\, \Phi'_S)/D \\
&= (-\,1/D)\left[\beta_S^2 + \Big(\sum_{t\notin S} p_t\, U''_t - p_s\, U''_s\Big)\, \Phi'_S\right] \\
&= (-\,1/D)\left(\beta_S^2 + \Phi'_S \sum_{\substack{t\notin S \\ t\neq s}} p_t\, U''_t\right) > 0
\end{aligned}$$

with the aid of (89), (81), (84), and the fact that $U''_t < 0$.

Finally, in the actuarially fair case, we use (50) in the form

$$\sum_s p_s (U'_s)^{-1}(\lambda) = \sum_s p_s\, a_s,$$

and it is obvious that an increase in any a_s decreases λ.

Theorem 8. Consider the optimal policy defined in Theorem 5. Assume that the premium waiver, if paid at all, is always smaller than the insurance payment (as would be true under any of the hypotheses of Theorem 5(d)).

(a) An increase in the initial (pre-insurance) income in any state increases post-insurance income in every insured state.

(b) An increase in the pre-insurance income in any insured state decreases the premium paid.

(c) An increase in the pre-insurance income in any uninsured state increases the premium but not so much that the post-premium income in that state is reduced.

REFERENCES

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

-----, 1965, *Aspects of the Theory of Risk-Bearing*, Helsinki: Yrjö Jahnssonin säätiö.

-----, 1971, *Essays in the Theory of Risk-Bearing*, Chicago: Markham; Amsterdam-London: North-Holland.

Ehrlich, I. and G. S. Becker, 1972, "Market Insurance, Self-Insurance, and Self-Protection," *Journal of Political Economy*, 80:623-648.

Eisner, R. and R. Strotz, 1961, "Flight Insurance and the Theory of Choice," *Journal of Political Economy*, 69:355-368.

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

Hirshleifer, J., 1970, *Investment, Interest, and Capital*, Englewood Cliffs, New Jersey: Prentice-Hall.

Pashigian, B. P., L. L. Schkade, and G. H. Menefee, 1966, "The Selection of an Optimal Deductible for a Given Insurance Policy," *Journal of Business*, 39:35-44.

Pratt, J. W., 1964, "Risk Aversion in the Small and in the Large," *Econometrica*, 32:122-136.

Ramsey, F. P., 1931, "Truth and Probability," *The Foundation of Mathematics and Other Essays*, London: K. Paul, Trench, Trubner & Co.

Samuelson, P. A., 1947, *The Foundations of Economic Analysis*, Cambridge, Massachusetts: Harvard University Press.

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

Wilson, R., 1968, "The Theory of Syndicates," *Econometrica*, 36:119-132.