The Geneva Papers on Risk and Insurance, 8 (No 26, January 1983), 4-33

Risk, Incentives and Insurance : The Pure Theory of Moral Hazard *

by Joseph E. Stiglitz **

1. Introduction

In this paper, I wish to explore the relationship between risk, insurance, incentives, and imperfect information. Understanding the relationship between these is fundamental, and not only to an understanding of the functioning of insurance markets. The phenomenon of risk and insurance, and the problems which they pose, are pervasive throughout the economy, often in quite disguised form ; there are many features of the economy that can best be understood as an institutional adaptation to the problems of risk and incentives. Somewhat surprisingly what has become the standard paradigm for analyzing the market economy, the competitive model (as exposited, say by Arrow [1964] and Debreu [1959], systematically ignores these considerations : As a result not only does it fail to provide insights into some of the important features of modern capitalist economies, but also it reaches conclusions, e.g. concerning the

---

* Fifth Annual Invited Lecture of the Geneva Association, delivered at Zurich, March 18, 1982. This paper reports on results obtained in a long term research program on incentives, risk, and imperfect information, undertaken jointly with Richard Arnott of Queen's University, Bruce Greenwald and Andy Weiss at Bell Labs, Barry Nalebuff at Oxford, Avi Braverman at the World Bank, and Carl Shapiro at Princeton. Most of the results reported here represent work undertaken jointly with Richard Arnott and are described at greater length in R. Arnott and J. E. Stiglitz [1981 a]. For the most part, in this exposition I omit all formal proofs, and a number of technical qualifications to the central argument. Interested readers are urged to consult the more complete and detailed development of these ideas in Arnott and Stiglitz [1981].

Applications of these ideas to the labor market are described in Arnott and Stiglitz [1981 b], and an application to land and credit markets in l.d.c.'s is described in Braverman-Stiglitz [1981, 1982]. The issues of incentives in credit markets is discussed in Stiglitz-Weiss [1981, 1982]. A general treatment of the welfare economics with imperfect information and incomplete markets is contained in Greenwald-Stiglitz [1982].

I am indebted to R. Arnott, D. Abreu, O. Hart, M. Hellwig, C. Shorin, T. Tan and T. Ungern and to the participants at the Seminar in Zurich where an earlier version of this paper was discussed. I am particularly grateful to Martin Hellwig for pointing out how some of our results of section 3.4 may be altered by changes in the equilibrium concepts and strategy spaces employed. These points are discussed at greater length in Hellwig's comments which follow.

Financial support from the National Science Foundation is gratefully acknowledged.

** Professor of economics, Princeton University.

efficiency of the market and the existence of equilibrium, which are necessarily suspect. Indeed, recent research which I will discuss today has shown them not to be generally valid. Equally important, policies designed to alleviate what have been perceived as market failures, to provide social insurance where the market has provided inadequate insurance, have failed to grapple with the underlying reasons that the markets failed to provide adequate insurance ; the consequences of this are gradually becoming clear, with country after country facing the unpleasant task of choosing between increasing tax burdens and the associated losses of economic incentives, or an alteration in the nature of the social insurance they provide.

The basic principles underlying risk and insurance with symmetric information have been well understood for the past two decades. One of Arrows's [1964] great contributions was to show that these, in themselves, require no significant alteration in basic competitive analysis. This is no longer true, however, when we combine the problems of risk and insurance with asymmetric information. Two major problems have been uncovered, referred to as " adverse selection " and " moral hazard ".[1,2] The first is concerned with imperfect information about the attributes of those who apply for insurance ; changing the terms of the insurance contract (the price, the co-insurance clause, the deductibility provisions) affects the mix of those who purchase insurance. In this paper, I will ignore the problems posed by adverse selection and focus instead on those presented by moral hazard.[3] Moral hazard problems arise when there is imperfect information concerning the actions of these who purchase insurance, because those actions cannot be perfectly monitored[4] and the insurance contract cannot specify all of the actions which the insured is to undertake. Under those circumstances, the provision of insurance implies that individuals do not bear fully the consequences of their actions. If an individual is in an accident for which he is insured, he personally must pay only a fraction of the costs of the damage. If an individual has hospital

---

[1] Arrow, in his Johnsson lectures, was among the first to point to the importance of these phenomena (Arrow [1965]).

[2] These are not, of course, the only problems. There is a *verifiability* problem, i.e. ascertaining whether the insured contingency has actually occurred, and an *enforceability* problem, i.e. ensuring that the insurance company actually fulfills its contracts.

Though insurance regulations (e.g. reserve requirements) and the legal system are intended to deal with the enforceability problem, transactions costs are such as to necessitate partial reliance on other enforcement mechanisms, such as reputation, and even these may have only limited efficacy. See Smallwood [1975]. The enforceability problems are central to an understanding of *implicit* contracts, where, say, an employer provides insurance to his employees, against cyclical risks, but not through a formal contract which is enforceable by the courts. See Newbery and Stiglitz [1982].

[3] This is not because these adverse selection problems are not important. Rather, it is because these fundamental problems have been dealt with extensively elsewhere (Stiglitz [forthcoming], Rothschild-Stiglitz [1976], Wilson [1977]. Some important insights are lost by treating adverse selection and moral hazard problems in isolation. In subsequent work, I hope to examine the two phenomena together.

[4] In some contexts, one can make inferences about the actions taken by observing the outcomes. Here, we assume only whether, say an accident has or has not occurred is observable. This information is not sufficient to infer the care level of the individual.

insurance but smokes, he personally has to pay only a fraction of the costs resulting from the cancer which may develop as a consequence. If the state provides unemployment insurance, the individuals pays only a fraction of the loss in output resulting from his failure to seek employment. If a landowner provides " insurance " to his tenant, through a sharecropping arrangement, the tenant pays only a part of the costs associated with taking inadequate precautions against an early frost. In each of these cases, we could have rephrased our statements to read, " the individual only receives a fraction of the benefits that accrue from his actions ". Thus under a sharecropping arrangement, the worker only receives a fraction of the increased output resulting from his taking precautions against an early frost ; the company manager who takes some action that reduces the probability of bankruptcy may receive only a fraction of the benefits which accrue (and even the shareholders receive only a fraction of the gain).

This, then, is the fundamental conflict : the more and better insurance that is provided against some contingency, the less incentive individuals have to avoid the insured event, because the less they bear the full consequences of their actions.

At the onset, there are two points which I wish to emphasize. First, the incentive problem arises from the presence of all three factors : risk, insurance, and imperfect (asymmetric) information. If there is risk, but individuals are not risk averse, so there is no insurance, there is no incentive problem. Contracts will be designed to make the individual bear the full consequences of his actions. In agriculture, this means that there will only be rental contracts ; [5] in credit markets, this means there will only be loan contracts ; in manufacturing, all individuals will work on a strictly piece rate basis.

If there is risk and insurance, but symmetric information, the contract would specify all the actions to be undertaken : there would again be no incentive problem.

A contract which required the individual to undertake certain accident avoidance activities would have premiums that appropriately reflected the differences in accident probabilities that result. In such contracts, individuals are rewarded for taking greater care.

Secondly, as my earlier remarks have suggested, I am concerned not only with the incentive problems associated with explicit insurance contracts, but also with those associated with the insurance implicit in a large variety of contractual arrangements. The risk sharing which occurs between a landowner and his tenant under sharecropping can be viewed as a form of insurance, and the nature of that insurance contract is critically affected by the difficulties the landlord has in monitoring the inputs of the tenant (Stiglitz [1974]). When suppliers of credit provide capital through equity arrangements, there again is a risk sharing arrangement which can be viewed as a form of insurance (see Stiglitz-Weiss [1981]). When employers make part of the compensation which they provide their employees independent of the output of their employees,

---

[5] Even with such contracts, moral hazard problems may arise if there is a chance of bankruptcy, and if the probability of bankruptcy can be affected by the action of the tenant (in the agricultural contexts) or the borrower. See Stiglitz-Weiss [1981].

they are providing a kind of insurance to their workers (Stiglitz [1975]). [6] When firms provide a guarantee on the products they sell, they are, in part, providing an insurance contract. In each of these cases, one party to the contract (often referred to as the principal) is affected by the actions of the other (the agent) which he cannot control directly, but can affect through the design of the contract. Thus, with sharecropping the landlord would like his tenant to work hard, since that increases the landlord's income ; the bank would like the borrower to undertake safe projects, since that reduces the probability of default ; an insurance company would like the insured to exercise care in avoiding accidents, and the firm that offers a guarantee on its product would like the purchaser not to abuse the product.

Similar risk-incentive problems arise in a variety of other contexts as well. In the recent macro-economic literature, there have been extensive discussions of " implicit contracts " in which the employer does not vary the wage, in spite of the variability in the workers' productivity resulting from changing market conditions (see Azariadis [1975] or Baily [1974]) ; the firm is thus providing the worker with a kind of insurance. With perfect information (as was implicitly assumed in the earlier implicit contracts literature) there is no distortion associated with these insurance contracts ; but with imperfect information, these insurance contracts interfere with the efficient allocation of labor. When an individual's productivity fall, as a result of a decline in the demand for the product he produces, he does not have an incentive to seek employment elsewhere. [7] See Arnott, Hosios, and Stiglitz [1980].

Similarly, in labor markets in which specific training is important, in the absence of risk aversion an individual would pay for his own training, but would receive a wage commensurate with his marginal product (which would normally be high, because of the training). But this would imply that, if an individual turned out to be badly

---

[6] In each of these instances we could divide the contractual relationship into two parts, one of which entails the provision of insurance. Thus, in the landlord-tenant relationship, we could imagine the landlord renting the land to the tenant for a fixed rent, and then the tenant purchasing an insurance policy, with the provision that the insurance company (here the landlord) will pay the tenant if his income (output) falls below some critical level, an amount which is some fraction of the difference between his output and that critical level ; in return, the tenant agrees to pay the insurance company if the output exceeds the critical level, an amount which again is proportional to the deviation between his output and the critical level. Thus, if the price of output is normalized at unity, if Q is the level of output, R is the fixed rent, α is the equilibrium share, then the income of the tenant under the sharecropping contract αQ, can be rewritten as

$$Q - R + (1-\alpha)(\hat{Q}-Q)$$

where $\hat{Q} = \frac{R}{1-\alpha}$. The first part represents the rental contract, the second part the insurance contract. In principle, the insurance contract could be provided by a separate party. In the case of the particular situation under investigation here, there are strong transactions costs reasons for the landlord (employer, etc.) to provide the insurance.

[7] These distortions are quite apart from those which arise if there is asymmetric information between the employer and employee concerning which state has occurred, whether, for instance demand for the firm's product is very low. This verifiability problem is resolved by restricting the firms in such a way that they have an incentive to state honestly the state of nature. There is, of course, a lose of output from the imposition of these (self-selection) constraints. See Azariadis and Stiglitz [1982], Azariadis [1982] or Grossman and Hart [1981]. For a critique of the approach, see Newbery and Stiglitz [1982].

matched with a firm or if he had to leave for one of a variety of reasons, he would not be able to recoup his investment in his specific training. As a result, firms are induced to provide insurance : to pay for a fraction of the costs of training, but at the same time to pay individuals less than their marginal product. But this again leads to incorrect incentives with respect to labor allocation : workers may leave when it may be socially inefficient for them to do so, since they are not receiving their true marginal product. See Arnott and Stiglitz [1981 b].

I could cite numerous other instances but the point I think is clear : almost all economic relations are affected by risk, and by the problems of insurance and incentives to which this gives rise. Thus, an understanding of these phenomena are central to an understanding of any economy. We now turn to a review of the central concerns in the design of contractual arrangements.

2. The design of contracts

The design of insurance contracts must represent a balancing of the risk and incentive effects : the greater the degree of risk aversion and the greater the risks faced by the individual [8], the more important is it to the individual to have the risks alleviated. In some cases the incentive effects may be negligible. The major incentive that individuals have for avoiding large automobile accidents is probably saving their own life ; even complete payment of the monetary losses does not alter in a significant way the incentives for accident avoidance for most (non-suicidal) individuals. This, however, may not be true for small accidents.

There is a third aspect to the design of contracts : what might be an optimal contract for one environment is not for another. Again, if there were perfect information the contract would specify how the benefits and payments would alter with the change in the environment. But there is commonly *asymmetric* information ; one party to the contract (the insured) has different information from the other, and as a result, a contract which alters the terms with each change in the environment is not feasible. [9] Thus, if the level of driving care were observable, an optimal insurance contract would (with perfect information) specify the level of care to be taken in each situation. Even if the landowner could specify the level of effort or input of fertilizers to be used by the tenant, the optimal contract (again with perfect information) would need to vary this amount with each change in the weather conditions. A piece rate contract needs

---

[8] In Stiglitz [1975], I derive the optimal piece rate, relating it to a) the magnitude of risk faced ; b) the degree of risk aversion (as measured by the Arrow-Prott measure of relative risk aversion) ; and c) the elasticity of effort supply. Similarly, in Stiglitz [1974], I derive the optimal sharecropping contract, where, in addition to the parameters listed above, the elasticity of substitution between labor and land is shown to be important.

[9] There are other reasons that the contract may not be written to reflect fully changes in the environment : transactions costs would make such contracts prohibitively expensive to write, and even when the state of nature is observable to both parties of the contract, to be enforceable, the state must be *verifiable* by a third party.

to alter the piece rate with each minor change in the technology which makes the task easier or more difficult. The appropriate incentive pay (bonus) for a salesman may differ, depending on how hard it is to sell what he has to sell (see Nalebuff-Stiglitz [1982]). These conditions necessitate taking into account a third property of the contract, which I have called its *flexibility*. [10]

Thus, when the input of fertilizer is observable, although one could write a contract which required the tenant to provide a particular level of fertilizer, such a contract would not be flexible. A cost sharing arrangement provides the tenant, who has better information concerning the returns to fertilizer, with incentives to adjust the input. At the same time, it may impose a slightly greater risk on him (see Braverman-Stiglitz [1982]).

Similarly, an insurance policy could specify what accident avoidance actions are to be taken ; but since it could not specify how these were to vary as circumstances varied, much of the time the actions specified would not be appropriate. Thus, by making the individual bear some of the risk himself, the appropriate adjustments to the changes in environment are induced ; the resource savings may more than compensate for the additional risk which he must bear.

These remarks, I think, provide us with some insight into why contractual relations differ so markedly from the simple kinds envisioned with the traditional competitive paradigm. It provides us with some insight, too, into why markets do not provide complete insurance, and warns us of some of the consequences which might be expected if the government goes where the market fears to tread. But these remarks are, still, only suggestive, for so far, we have not presented a theory of market equilibrium, within which we can then assess the efficiency of the market. This is the task to which we now turn.

3. Market equilibrium

3.1. The basic insurance model

We investigate these questions in the context of the simplest possible model. There is a single accident to be insured ; it occurs with probability p. This probability is a function of the effort, e, spent at accident avoidance. [11] In the absence of insurance the individual's ex ante utility is a function of his income in the two states, no accident and accident :

$$y_0 = w$$
$$y_1 = w - d$$

---

[10] Limitations of space do not allow me to pursue further the implications of flexibility in the more formal analysis of market equilibrium that we undertake in the next section.

[11] We assume that increases in effort reduce the likelihood of an accident, $p'(e) \leqslant 0$; but at a diminishing rate, $p'' \geqslant 0$.

where w is the initial wealth and d is the damage of the accident. Thus his expected utility can be written

(1) $$EU = EU(y_0, y_1, e).$$

For simplicity, for most of the analysis, we assume a separable event-independent utility function, with constant marginal disutility of effort. Few of our results depend on this ; we shall comment briefly on what happens when they do.

(2) $$EU = u(y_0)\,(1 - p) + u(y_1)\,p - e.$$

With insurance, individual income, in the event of an accident is increased, while if there is no accident, it is reduced. Let β be the premium, α the (net) benefit, so

$$y_0 = w - \beta$$
$$y_1 = w - d + \alpha$$

3.2. Fundamental non-concavity of the indifference curves

We now depict individuals' indifference curves between benefits, α, and premia, β. Clearly, individuals are willing to pay a higher premium for a greater benefit. If e were not variable, then with diminishing returns, the indifference curves would appear as in Figure 1. Assume, however, that e is variable. For instance, individuals can choose to smoke or not to smoke ; the probability of an accident (fire) depends on which action is taken.

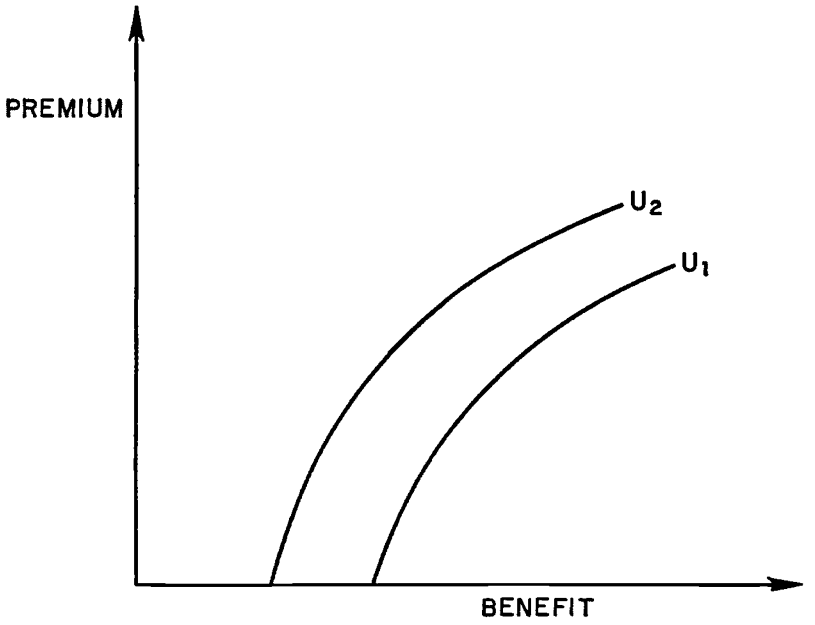

Figure 1 :

Indifference curves between benefit and premia with fixed effort are quasi-concave

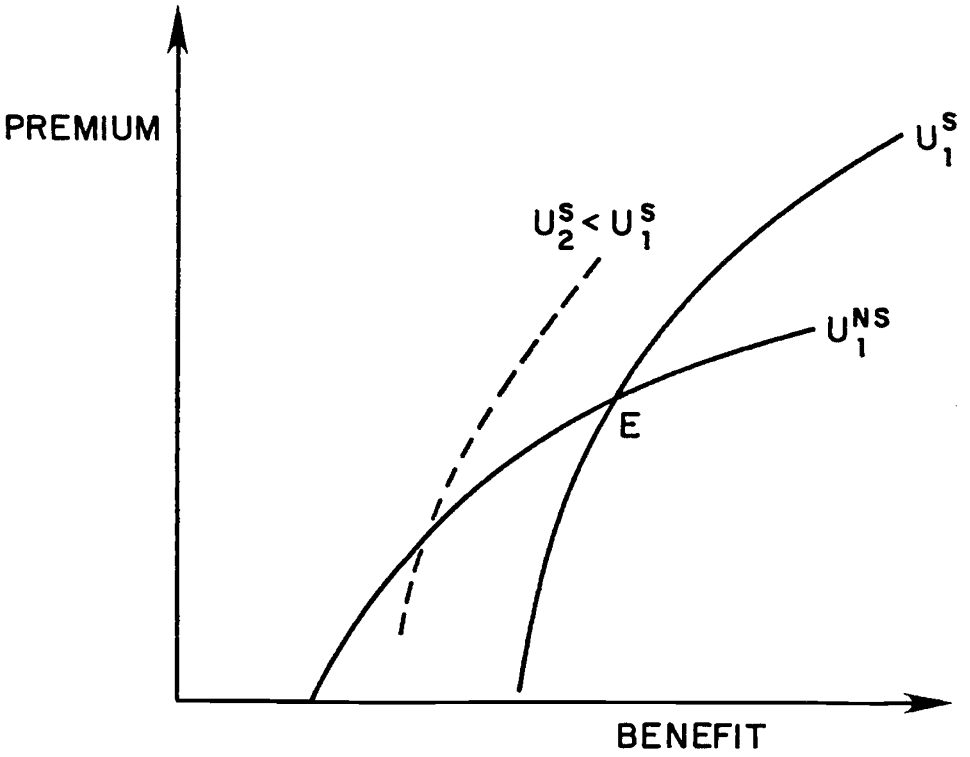

Figure 2 : With variable effort indifference curves are not quasi-concave

There are thus two " families " of indifference curves, one for smoking, one for not smoking. The individual chooses which is relevant (i.e. whether to smoke or not smoke) ; for any insurance policy this depends, in turn, simply on whether the no smoking indifference curve through the corresponding point has a higher utility level than the smoking indifference curve through that point. In Figure 2, U_1^S and U_1^{NS} give the same level of utility. [12] We have depicted the smoking indifference curve through any point to be steeper than the no smoking indifference curve, reflecting the higher probability of an accident. It is thus apparent that along the indifference curve U_1^{NS} below E, the point of intersection of the two indifference curves, the utility to not smoking exceeds that to smoking ; conversely, along the indifference curve U^S above E. Thus, the indifference

[12] It is possible, of course, that the indifference curves U_1^S and U_1^{NS} never cross. This implies that, regardless of the level of insurance, the individual smokes. There is then no moral hazard problem : undertaking the accident avoidance activity is not affected by the provision of insurance. Our concern here is with those situations where effort is affected.

curve, taking into account the adjustment of effort as the insurance is altered, is not quasi-concave. This is a general result, and has profound implications for the nature [13] of competitive equilibrium. [14]

It implies, in particular, that both price and income consumption curves may be discontinuous, as illustrated in Figures 3 and 4.

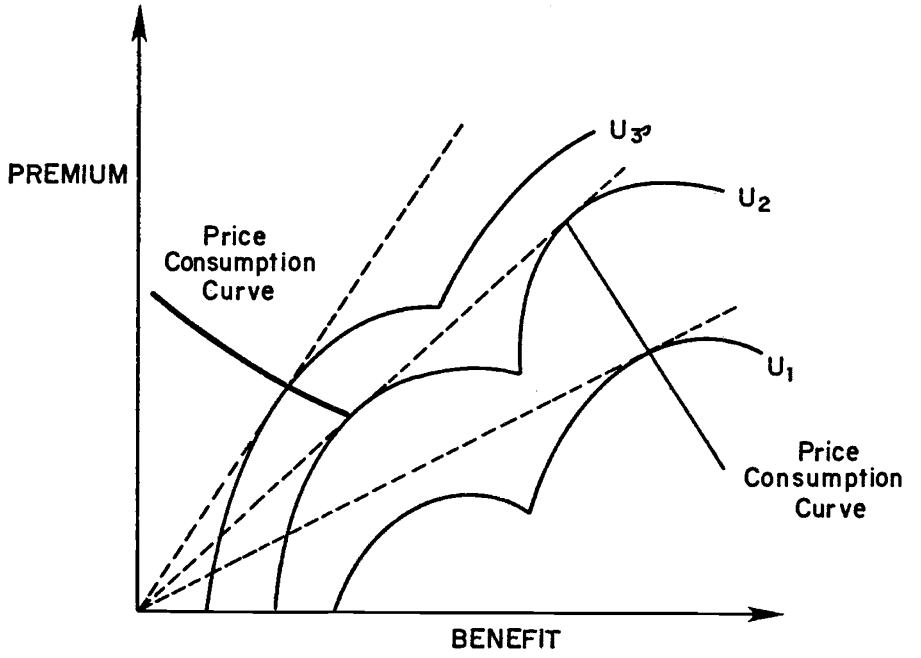

Figure 3 : Price consumption curve is discontinuous

---

[13] We assume, moreover, that the individual can smoke a fraction λ, of the time, in which case the probability of an accident is simply $\lambda p_S + (1-\lambda)p_{NS}$, where p_S is the probability of an accident with smoking and p_{NS} is the probability of an accident with no smoking. One implication of our analysis is that, except at points such as E, λ will either be 1 or 0. At E it is indeterminate.

[14] Whenever there are a discrete number of activities (as in this example), the indifference curves will never be quasi-concave. If there are a continuum of activities, and we assume p is a continuous, thrice differentiable function of e, then, if

$$\lim_{e \to 0} \frac{(p')^3}{p''} = -\infty$$

(where $e = 0$ represents the minimum level of effort), indifference curves will never be quasi-concave. If the above condition is not satisfied, indifference curves may be quasi-concave, but they need not be. See Arnott-Stiglitz [1981 a], Part I.

Notice that in the discrete activity case, we need not quantify, in any way, what we mean by effort.

In the continuous case, even with seemingly well behaved $p(e)$ and utility functions, e may be a discontinuous function of the amount of insurance provided when the utility function is not separable.

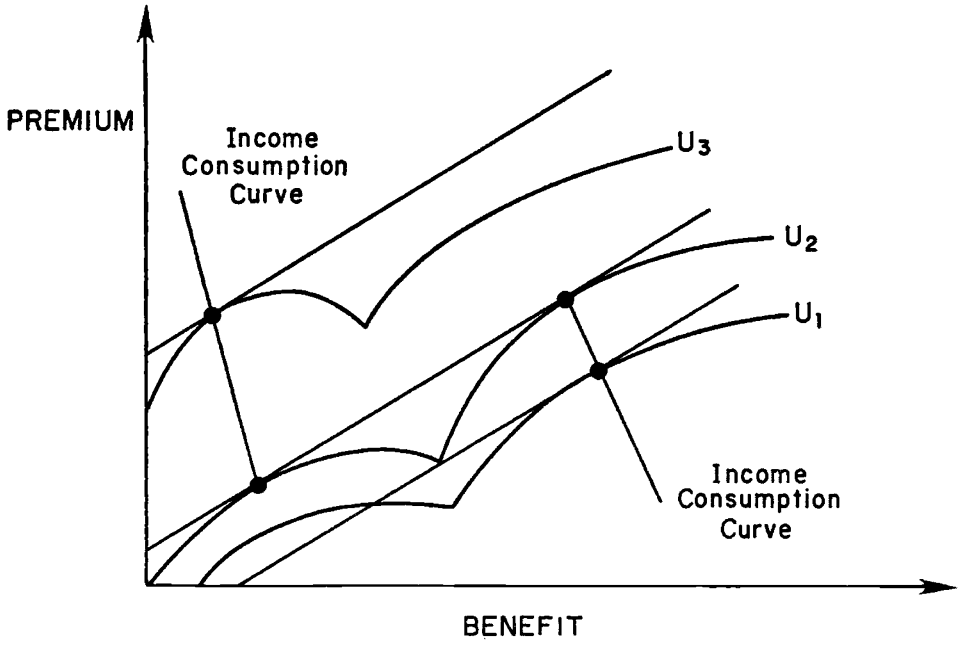

Figure 4 : Income consumption curves are discontinuous

3.3. Market equilibrium

A competitive market equilibrium is a set of insurance contracts, all of which make non-negative profits and are purchased, such that there does not exist another contract which could be offered which would make a profit. [15]

This definition is the same as that employed by Rothschild and Stiglitz [1976] in their study of adverse selection. It is markedly different from conventional definitions of competitive equilibrium which typically *assume* market clearing, zero profits, and the existence of prices. In the approach taken here, these are characteristics of equilibrium which need to be established ; it turns out that in the presence of imperfect information, competitive equilibrium may have none of these characteristics (see Rothschild-Stiglitz [1976], Stiglitz-Weiss [1981, 1982], and Stiglitz [forthcoming]).

The nature of the equilibrium depends critically on what are feasible contracts.

---

[15] We thus rule out what we refer to as latent policies, policies which are offered, not purchased, but which might be purchased if a new firm offered an additional policy. Such latent policies may serve to deter entry. See Arnott and Stiglitz [1981 a], Part II.

Figures 5 : Even with well-behaved utility functions the feasibility locus may have irregular shapes

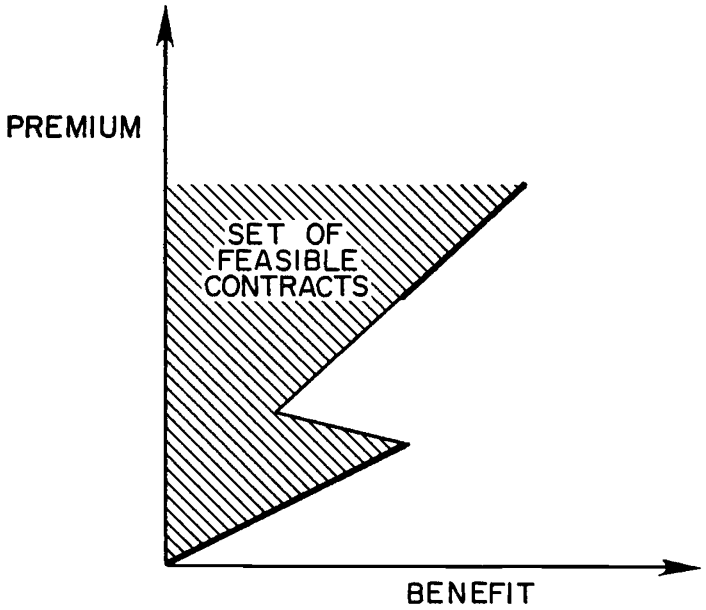

Figure 5 a : Feasibility locus : two-activity case

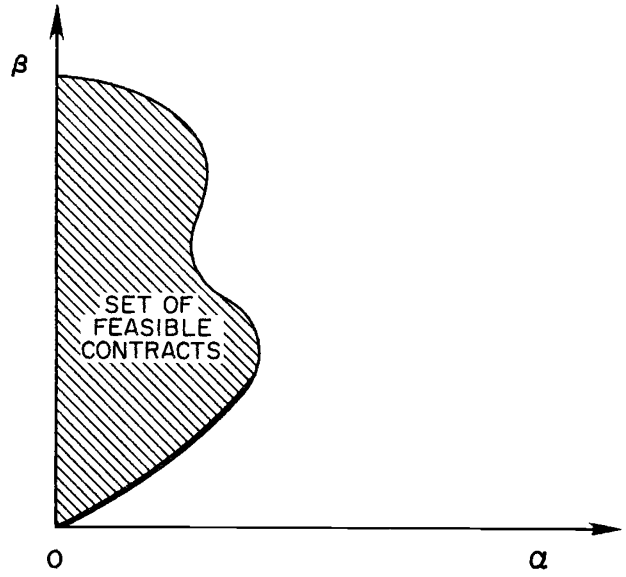

Figure 5 b : Feasibility locus : with continuum of techniques, the resource constraint may be positively or negatively sloped, and may be convex or concave

What are *feasible* contracts depends, as we have already noted, on what is observable. The issue of incentives, with which we are concerned here, arises largely because the actions aimed at accident avoidance are not observable ; i.e. the insurance contracts cannot specify what action the individual is to undertake.

The nature of the equilibrium turns out to depend critically on whether insurance purchases from *other* insurance companies are observable. We will consider here only two polar cases, where the quantities purchased are and are not observable. In the general case, some insurance purchases may be observable, while it may be difficult to quantify the insurance provided implicitly by employment contracts, friends, and social institutions. Moreover, whether particular insurance purchases are or are not observable may be an endogenous variable. See, e.g. Arnott and Stiglitz [1981 a] and Hellwig [1983]. See Jaynes [1978] for a discussion of same point in the context of adverse selection. If they are, then the equilibrium may easily be characterized.

In our earlier discussion, we argued that the level of effort that an individual undertakes will be a function of the insurance provided : [16] we thus write

$$e = e(\alpha,\beta). \tag{3}$$

We can now define the set of feasible contracts : those $\{\alpha,\beta\}$ for which expected profits are non-negative, i.e.

$$\beta\,[1 - p(e(\alpha,\beta))] \geqslant \alpha p(e(\alpha,\beta)).$$

The set of $\{\alpha,\beta\}$ for which this holds with equality is called the resource constraint. We define the feasibility locus,

$$\beta = \beta(\alpha)$$

giving the minimum premium, consistent with any benefit, yielding non-negative profits.

For the two-activity case, it appears as in Figure 5, while more generally, it can take on a variety of irregular shapes.

It is then easy to establish that the market equilibrium is characterized by the tangency of the indifference curve to the feasibility locus and entails exclusive insurance provided by a single insurer. Moreover, since along the feasibility locus

$$\frac{d\beta}{d\alpha} = \frac{p}{1-p} + \frac{\alpha}{(1-p)^2}\frac{dp}{d\alpha}$$

and $\frac{dp}{d\alpha} > 0$ (more insurance leads to less care) the optimal insurance contract always

---

[16] If p is a continuous function of e, the first order condition for effort is

$$(u_1 - u_0)p' = 1 \tag{4}$$

where $u_1 \equiv u(y_1)$, $u_0 \equiv u(y_0)$.

In the two-activity case, clearly e is a well-defined function everywhere, except for those values of α,β for which the individual is indifferent between the two activities. More generally, e will not be a continuous function of α and β.

entails quantity rationing,[17, 18] i.e. if individuals could purchase additional insurance at the same " price " ratio of premium to benefit, they would always wish to do so, as illustrated in Figure 6. This result (noted earlier by Pauly [1974]) has some important implications, to which we return later.

For now, we note two direct consequences of our analysis. First, a natural institutional method for implementing quantity rationing is to impose exclusivity requirements : the individual is allowed to purchase from only one insurance company. As we remark later, these exclusivity clauses may have some important implications for the functioning of the competitive economy.

Secondly, the irregularity of the shapes of both the feasibility locus and the indifference curves implies that small changes in parameters (e.g. the cost of automobile repair) may lead to a discontinuous change in the demand for insurance (and hence in the demand for other commodities).

As usual, this implies that, in general, competitive equilibria need not exist. Note that in obtaining this result, we have assumed that all the underlying functions in our analysis (e.g. utility functions) are well behaved ; they satisfy all the conditions typically imposed (e.g. by Debreu [1959]). The only change in the analysis is to assume that individuals' actions are not perfectly observable by the insurance firm, an assumption which seems eminently reasonable.

How serious this existence problem is, is a moot question : as usual, if there are a continuum of individuals, then, under reasonable conditions, equilibrium will exist. This is *not* true of the existence problem uncovered in the next section.

Quantity rationing is not the only way that market equilibrium will differ from the conventional price equilibrium. Under a variety of circumstances, one can show

---

[17] Let q be the " price of insurance ", the premium paid per dollar of benefit, $q = \frac{\beta}{\alpha}$. In a zero profit equilibrium, $q = \frac{p}{1-p}$. Assume individuals are not quantity rationed. Let $V(\alpha,\beta) \equiv \max_{\{e\}} EU$, i.e. the maximized level of expected utility is just a function of α and β. Individuals who are not quantity rationed will choose α to max $V(\alpha, \alpha q)$, i.e.

$$-\frac{V_\alpha}{V_\beta} = q = \frac{p}{1-p}.$$

In contrast, the tangency between the indifference curve and the feasibility locus is characterized by

$$-\frac{V_\alpha}{V_\beta} = \frac{p}{1-p} + \frac{\alpha}{(1-p)^2}\frac{dp}{d\alpha}.$$

This argument only establishes that individual's utility would *locally* be increased by an increase in insurance. There may be a global maximum entailing a reduction in insurance, as we shall see below. This cannot, however, be the case at the optimal insurance contract.

[18] To see that $\frac{dp}{d\alpha} > 0$ assume the contrary, i.e. that increases in α lead to a higher level of care. Recall the first order condition for effort

$$-p'[u_0(w-\beta) - u_1(w-d+\alpha)] = 1$$

By the definition of the feasibility locus, the minimum premium for any benefit, it is clear that if α increases, β must increase. As α increases, u_1 increases ; as β increases u_0 decreases. Hence $u_0 - u_1$ decreases. But then e could not have increased.

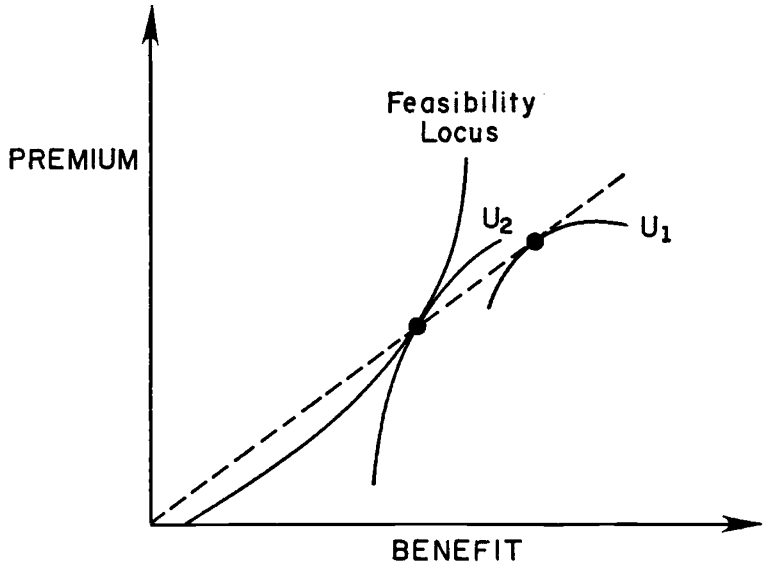

Figure 6 : The optimum insurance contract entails quantity rationing, if quantities are observable.

that it is desirable to randomize benefits and/or premia. The randomization imposes greater risk, but in certain circumstances, induces greater effort, and the gains in increased effort may more than offset the losses from the greater risk. [19]

3.4. Equilibrium when insurance purchases are not observable

Matters, are, however, far more complicated when insurance purchases from other firms are not observable. Since there cannot be quantity rationing, [20] in this case, we shall, for simplicity, refer to the equilibrium (when it exists) as a price equilibrium.

---

[19] The conditions under which randomization is desirable are described in Arnott-Stiglitz [1981 a], Part II. There, we distinguish two forms of randomization, ex ante and ex post, i.e. randomization before the individual decides on e and after. Both kinds of randomization may be desirable.

The desirability of randomization was earlier noted in the context of taxation in Atkinson and Stiglitz [1976], and Stiglitz [1982].

[20] This is not completely accurate, since any firm can ration the amount of insurance it sells to any of its customers. What it cannot do is ration the total quantity of insurance purchased, since what the individual purchases from other firms is unobservable. We show that if any firm attempted to ration the amount of insurance, by restricting the quantity of insurance which he sold, it would always pay some other firm to offer supplemental insurance, thus making the rationing ineffective. This analysis assumes, however, the absence of latent policies, policies which are offered but not purchased, but none the less serve to deter entry (see fn. 15). With such latent policies, there may be quantity rationed equilibria even with unobservability of policy purchases ; indeed, in some circumstances, it may be possible to sustain the equilibrium described in the previous section. See Arnott and Stiglitz [1981] and Hellwig [1983].

An equilibrium if it is to exist, must be of one of three forms : (a) zero profits with positive insurance ; (b) zero insurance ; and (c) positive profits. We shall first characterize the zero profit equilibrium and show that no such equilibrium may exist. We then turn to an investigation of the other two possibilities. It is easy to see that a zero profit price equilibrium, *if* it exists, is characterized by complete insurance, in the sense that the marginal utility of incomes must be equated in the two states ; if e^* is the equilibrium level of effort, the zero profit condition implies that q, the premium per dollar of benefit (the price of insurance)

$$q = \frac{\beta}{\alpha} = \frac{p(e^*)}{1 - p(e^*)} \cdot \tag{5}$$

Individuals will purchase insurance to the point where marginal rate of substitution equals $\frac{1}{q}$ [21] i.e.

$$\frac{u_0'(y_0)(1-p)}{u_1'(y_1)p} = \frac{1}{q} \tag{6}$$

or

$$u_0'(y_0) = u_1'(y_1) \tag{7}$$

(If the individual's marginal rate of substitution is not equal to the market odds, someone could offer a small insurance policy which would leave effort unchanged and make a profit. Hence, the original allocation could not have been an equilibrium.)

Thus, the zero profit equilibrium, when it exists, is characterized by the intersection of the price consumption locus (defined by equation (6), the feasibility locus, and the complete insurance locus (defined by (7)). See Figure 7.

This, in turn, has one important implication : if utility functions are event independent (as we assumed earlier) then equating marginal utilities in the two states is equivalent to equating utility levels ; but if utility levels are equated, individuals have *no* incentive at accident avoidance ; effort is set at its minimal level. [22] (Cf. equation (4) above.)

The consequence of this, in turn, is that the premium may have to be very high ; and at these high premia, individuals may prefer to self-insure (i.e. have no insurance) or to have only partial insurance. Such a situation is depicted in Figure 8. It is clear that there is no zero profit equilibrium with insurance.

---

[21] The marginal rate of substitution is found from (2):

$$\left(\frac{\partial y_1}{\partial y_0}\right)_{\overline{EU}} = \frac{\partial EU/\partial y_0}{\partial EU/\partial y_1} = \frac{dEU/dy_0}{dEU/dy_1}$$

since

$$\frac{dEU}{dy_i} = \frac{\partial EU}{\partial y_i} + \frac{\partial EU}{\partial e}\frac{\partial e}{\partial y_i}, \text{ and } e \text{ is chosen to maximize expected utility, so}$$

$\frac{\partial EU}{\partial e} = 0$. See also fn. 17 above.

[22] When accidents reduce the marginal utility of insurance, equating the marginal utility of insurance in each state implies that utility is still higher in the event of no accident, and hence care levels may be positive.

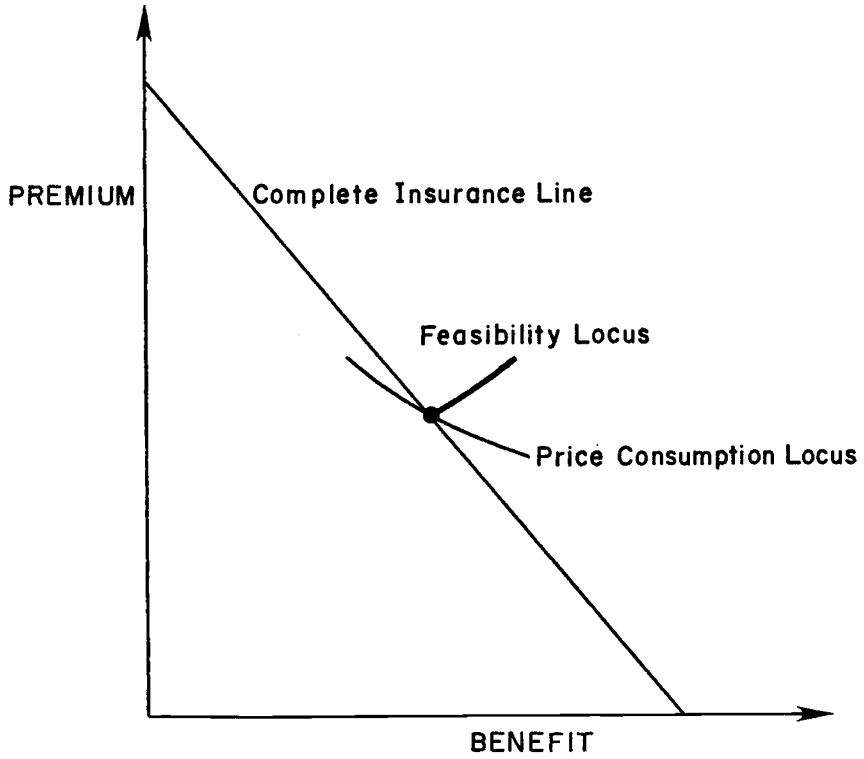

Figure 7 : The zero profit price equilibrium occurs at the intersection of the feasibility locus (zero profit locus), the complete insurance locus, and the price consumption curve

This, however, is not the only situation in which a *zero profit* equilibrium may not exist. Two other problems have been identified. The first occurs whenever any insurance firm believes that, given the current price of insurance, it can offer a sufficiently large policy (on which it can make a profit) that no one will wish to supplement it. (This policy itself is not an equilibrium, since were firms to offer only this policy, it would pay some firms to offer a new, small policy, at a price lower than that of the original " equilibrium ".) Formally, this simply requires that the income consumption curve through E enter the feasibility set at some point " beyond " E, i.e. at points corresponding to higher incomes (utility) than at E. Figure 9 illustrates one such example. [23] The final set of circumstances in which there does not exist a zero profit

---

[23] Figure 9 presents an example where the income consumption curve near but below E lies outside the feasibility locus, but examples can also be constructed where the income consumption locus near but below E lies inside the feasibility locus, in which case there again does not exist a zero profit Nash equilibrium.

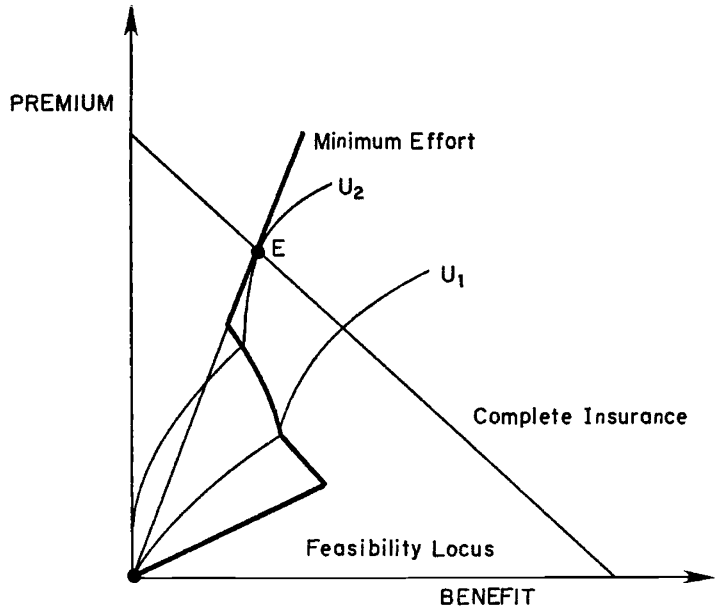

Figure 8 : Zero insurance may be preferred to the zero profit price equilibrium.

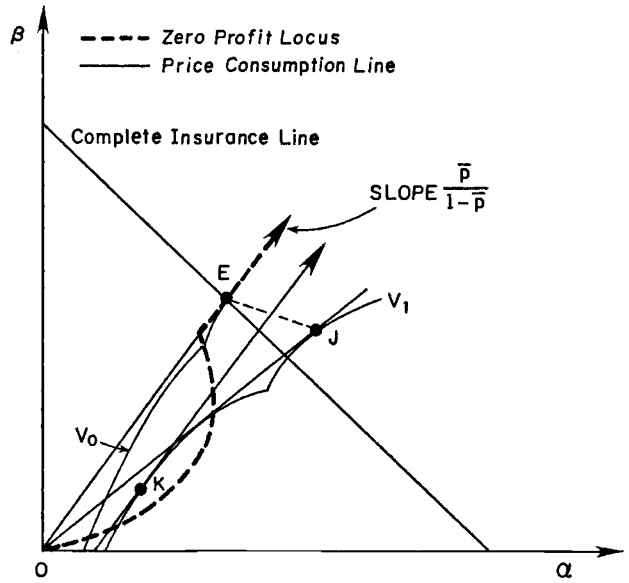

Figure 9 : With a zero profit pure price equilibrium, part of the relevant income-consumption line beyond E may lie inside the zero profit locus. (K is on income consumption curve through E and lies inside feasibility locus.)

equilibrium is when the price consumption locus is not continuous, a possibility we noted earlier. As a result, the price consumption locus may not intersect the feasibility locus (Figure 10). It is clear then that there cannot exist a zero profit equilibrium. [24]

What happens when there does not exist a zero profit equilibrium ? As we noted earlier, there would appear to be two possibilities, a zero insurance equilibrium or a positive profits equilibrium. We consider each of these in turn.

We first show that there cannot exist a zero insurance Nash equilibrium. For if no insurance is offered, it will pay some firm to offer a small policy so long as the indifference curves through the origin lies somewhere inside the feasibility set, as depicted in Figure 8. Thus, in this situation, there is no zero profit (Nash) equilibrium.

The entering insurance firm may realize however that he will not be alone in offering an insurance policy ; and that if others do offer insurance contracts, he will either make a loss, or have to raise his premiums to the point where no one wishes to purchase from him. If these considerations are taken into account, one might argue no firm will wish to enter, and argue that the zero insurance point is an equilibrium. [25]

Though there will not, in general, exist a zero insurance equilibrium, there may exist a positive profits equilibrium, in the situation depicted in Figure 10 a where the price consumption locus is discontinuous, and fails to intersect the feasibility locus.

In that case, the equilibrium if it exists, is the lowest point on the price consumption locus which is interior to the feasibility set. Such an equilibrium is characterized by positive profits, partial insurance, and positive effort.

Such a point may, however, not be an equilibrium. To see this, consider the income consumption curve through the point F. If that portion of the income consumption curve which is associated with higher levels of utility than that attained at F lies anywhere inside the feasibility locus (as it may), then there does not exist an equilibrium. For clearly a point such as G is preferred to F ; and clearly at G, individuals would purchase no additional insurance from the other insurance firms on the market. Finally, at G, the firm is making a profit. (But G itself cannot be an equilibrium. At G the individual's marginal rate of substitution is not equal to the slope of the zero profit line, at the given level of effort, so there is an incentive for a firm to introduce a small amount of additional insurance.)

Conversely, if no portion of the income consumption curve which lies below F lies inside the feasibility locus, the equilibrium is in fact at F. It is obvious that any firm which offered a lower price, even by an epsilon, would find itself making a loss.

---

[24] Actually, this point is not separate from the other two discussed above. Whenever the individual prefers partial or no insurance to full insurance at the zero effort premium-benefit ratio, the price consumption curve cannot intersect the zero profit locus.

[25] This is a kind of " reaction equilibrium ". Note, however, that these conjectures are very different from the kinds of conjectured reactions that have been discussed in the adverse selection literature. These, in general, entail existing firms dropping the policies which they presently offer. For a small firm to believe that it can upset a market equilibrium by introducing a policy which it sells to a few customers seems inconsistent with the basic spirit of competitive analysis. For a further discussion of these reaction equilibria, see, for instance, Rothschild-Stiglitz [1976], Wilson [1977], and Stiglitz [forthcoming].

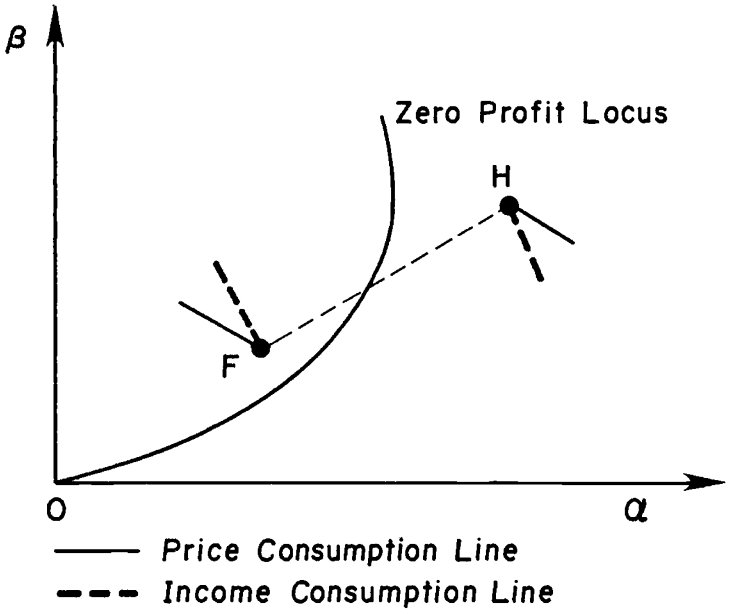

Figure a

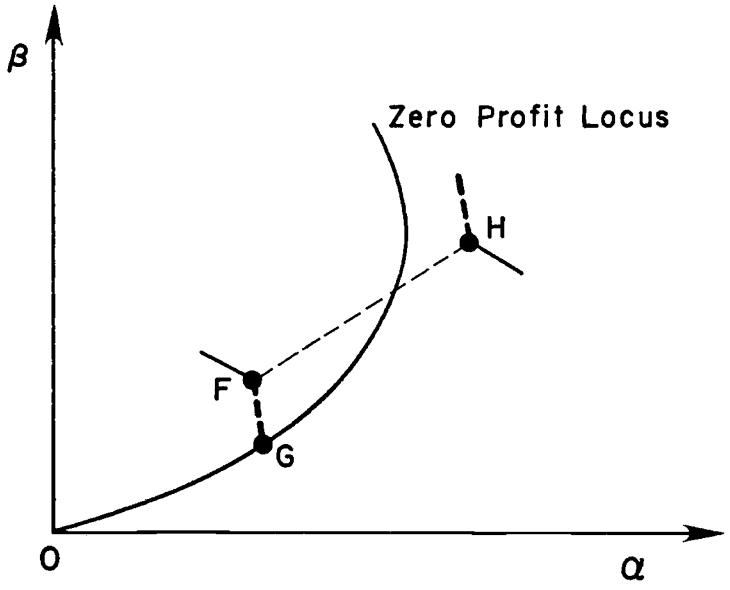

Figure b

Figures 10

Figure 10 a : Price- and income-consumption lines jump in the same direction

Figure 10 b : Price- and income-consumption lines jump in opposite directions

But even stronger, any firm which offered a small policy would find itself making a loss. For individuals would only purchase the policy if it offered more favorable terms. But then they would supplement this by buying insurance at the market terms, placing themselves somewhere along their income consumption curve below F. But all such points entail losses for the set of policies as a whole ; and since the new policy has more favorable terms than the old, a fortiori it must be making a loss.

We can summarize our results concerning equilibrium in insurance markets where the quantity of insurance purchased from other firms is unobservable as follows : price equilibrium, if it exists, either entails zero profit, with zero effort at accident avoidance [26] or positive profits, with positive effort. There may not exist an equilibrium if (a) individuals prefer to have no insurance, to paying the high premiums associated with the low levels of care which individuals undertake when they purchase complete insurance ; or (b) if insurance firms believe that they can sell a sufficiently large policy (but still providing less than complete insurance) that individuals will not wish to supplement it with other policies presently available. [27]

We have examined Nash equilibria with *deterministic* policies. In most contexts, this is the natural equilibrium concept for competitive markets. In the present context, one might reasonably take issue with the Nash equilibrium on three counts. First, in the positive profits Nash equilibrium, it is assumed that a potential entrant knows that he will make a loss if he enters. It might be argued that firms behave more naively, and simply enter an industry if profits are positive ; in this case, the positive profits equilibrium will not exist. There will exist no equilibrium to the market, since as firms compete by lowering their price, they discover that they are making a loss, in which case they exit. Secondly, we have already argued that in the situation where no insurance is being offered, an entering firm might well reason that were it to offer insurance, it would pay others to do so as well, and that in the resulting situation, he may take a loss. Thirdly, there is nothing to exclude firms offering random insurance (e.g. policies which pay a random fraction of the damage). We noted earlier that under a variety of situations, introducing such randomness is desirable ; although risk averse individuals are slightly worse off because of the direct effect of the variability, the random insurance leads them (under certain circumstances) to work so much harder, thus reducing the premium by more than enough to compensate for the increase in risk. [28] In the case where there is a positive profits equilibrium with deterministic policies, there is a zero profits equilibrium with random insurance which dominates the deterministic non-random positive profits equilibrium.

---

[26] In the case of state independent utility functions.

[27] There exist alternative specifications of equilibrium, entailing latent policies, in which even under these circumstances there may exist a positive profits equilibrium. The argument is similar to that presented earlier : any additional policy which, if purchased, would lead the individual to be better off will result in the purchase of the latent policy, and a discontinuous decrease in effort, so that the supplemental policy makes a loss. Hellwig has shown that even with latent policies, an equilibrium may not exist.

[28] In the case where the quantity of insurance is observable, randomness will never be desirable if the utility function is separable in income and effect.

3.5. Some lessons from the analysis

There are several important morals that we can learn from the analysis as presented so far :

The first set are directed at conventional economic theory. It is standard to take both the absence of quantity rationing (the use of prices) and the zero profit condition as part of the *definition* of equilibrium. We have contended, in contrast, that whether equilibrium will be characterized by a simple price system and by all firms making zero profits is a theorem to be proved, rather than an assumption ; and we have shown that in the kinds of situations with which we are concerned here, equilibrium may well be characterized by positive profits and (where quantities purchased from other firms are observable) by quantity rationing.

Not only have we shown that the traditional formulation of what is an equilibrium is incorrect [29] and that its characteristics differ markedly from equilibria in situations without moral hazard, but we have also shown that, even under the kinds of restrictive assumptions conventionally employed, equilibrium may not exist. (In the next section, we shall show, further, that the standard welfare theorems do not extend to situations where moral hazard is present.)

The second lesson to be learned from this analysis — a lesson which is perhaps obvious but seems to have gone unheeded in the design of many social insurance programs — is that economic efficiency requires that, where feasible, there must be quantity rationing of insurance ; in equilibrium, individuals should want to purchase more insurance than they have at the going market odds. In the absence of quantity rationing, individuals will take no care, and in most situations, this will clearly be inefficient. The perception that markets fail to provide " adequate " or sufficient insurance is not, in itself, evidence of market failure. Whether the market provides the correct amount of insurance is a difficult question, to which we now turn.

4. Welfare economics and the impossibility of efficient decentralization

One of the central results of modern competitive analysis is the fundamental theorem of welfare economics : every competitive equilibrium is Pareto optimal, and every Pareto optimum can be sustained by a competitive equilibrium, with the appro-

---

[29] These are not the only ways in which competitive equilibria with imperfect information and incentive problems differ from that of the standard competitive paradigm. Since the terms of the contract, e.g. the wage which the individual receives, or the interest rate which a borrower must pay, affects the actions he takes, it may be in the interest of the employer to pay a wage higher than the wage at which the individual is willing to work or the bank may be willing to lend at an interest rate below the market clearing interest rate. In both cases, the market equilibrium may be characterized by supply not equalling demand. Thus, market clearing should not be imposed as part of the definition of equilibrium ; rather it should be proved to be the consequence of more fundamental equilibrium conditions ; and while it is true for markets with perfect information, market clearing need not characterize equilibrium with imperfect information. See Stiglitz-Weiss [1981, 1982] and Shapiro-Stiglitz [1982].

priate lump sum redistributions. This result represents the precise sense in which Adam Smith's invisible hand works. It is a remarkable result, which provides the foundations, not only for the reliance on markets within capitalist economies, but also for the use of prices (and decentralized allocation mechanisms) within socialist economies.

The standard proofs of the Fundamental Theorem ignore the informational problems with which we have been concerned here. And when these informational considerations are introduced, the theorem is no longer valid. Competitive equilibria will no longer be Pareto efficient. Moreover, the particular forms in which the market failure occur provide some insight into the scope of government intervention.

In analyzing the efficiency of the market equilibrium, one should not, of course, compare the perfect information equilibrium with a perfect information planning solution : obviously, the latter generates a higher level of utility than the former. The comparison between the two needs to be made assuming the planner has the same information (or access to that same information technology) as private agents.

We thus ask, does the market generate a Pareto efficient allocation of resources, taking into account the limitations on information and the costs of acquiring additional information ? Does there exist a set of taxes/subsidies which can make some individual or group of individuals better off, without making anyone worse off ?

The inefficiencies which we are about to discuss may come as a surprise to some, since there are several papers in the literature (e.g. Shavell [1979]) alleging that the market is efficient, when there is moral hazard (but no adverse selection). Where have these studies gone wrong ? How do they differ from the results reported here ?

In my 1974 paper on sharecropping, I showed how the competitive equilibrium contract in the presence of moral hazard (what has since come to be called the principal agent problem) should be described as the solution to a simple (indirect control) maximization problem : the landowner maximized his (expected) utility of profits, subject to the constraint of being able to obtain workers (i.e. subject to an expected utility constraint), recognizing that effort, choice of technique, etc. of the tenant would depend on the terms of the contract. The competitive equilibrium contract would thus appear to be on the Edgeworth " contract " curve : landlords cannot be made better off without making workers worse off (lowering the utility level of workers). But this inference is incorrect : in making their choice of contract, both landlords and tenants take prices, incomes from other sources, etc. as given. The central issue is, in a general equilibrium, where these are determined simultaneously, is the market equilibrium efficient ?

To pose the question more clearly in the traditional insurance context : when the insurance company and the insured decide on the extent of co-insurance, they take the cost of automobile repair as given. But when all insurance companies together change the amount of coinsurance, the total demand for accident repair services may change, and this alters the price of accident repair, and profits in the repair industry.

These changes in prices do of course occur in traditional economic models, and the remarkable result of the Fundamental Theorem of Welfare Economics is that they can be ignored. Pecuniary externalities, as they are commonly called, make no

difference, so long as individuals are sufficiently small that they fail to take into account their effect on the market prices. The reason for this is that in the market equilibrium, social marginal cost = private marginal cost = private marginal benefit = social marginal benefit. Thus, the small perturbations to action induced by small changes in prices lead to no changes in welfare. Whenever there is insurance, however, individuals do not equate the social marginal cost to the social marginal benefit. If an individual decides to decrease the amount of care slightly, some of the costs will be borne by the insurance company, [30] or, in a competitive equilibrium, where the insurance company must adjust its premium to break even, by all the other customers of the insurance company. Now, if a change in the price induces the individual to take more care, there is a real external benefit. Of course, it is not only price changes which alter individuals' behavior. Our analysis has uncovered at least four important categories of what we may refer to as " externalities " arising in the presence of insurance, which result in the inefficiency of the market equilibrium.

1. The Interdependence of Seemingly Unrelated Insurance Contracts (and other risk markets). We argued earlier that if the amount of insurance purchased from other insurance companies were observable, equilibrium would be characterized by exclusivity : each individual would purchase his insurance from a single insurance company. The reason for this is that the individual's behavior, in the given insurance market — i.e. the effort he expended at accident avoidance — depended on the total amount of insurance purchased, not just on the amount of insurance purchased from the particular insurance company. Purchases of insurance from a second company imposed a kind of externality on the first : because the individual would, as a result, take less care, the profits of the first company would be reduced if the individual purchased more insurance.

The same argument extends to seemingly unrelated insurance markets. The amount of hospital insurance I have may effect the care I take in driving. This kind of interaction seems clear. But there are interactions even between, say, fire insurance and automobile insurance. Because of moral hazard considerations, we have argued individuals will not be able to purchase complete insurance. They must, to some extent at least, self-insure. The way they do this is to save (this used to be referred to as the precautionary motive for savings). But savings serve equally well for any of a variety of accidents against which the individual would like to purchase insurance. Thus, increasing the availability of fire insurance may reduce individuals' demand for savings, and this reduced savings may induce individuals to take greater care in driving.

Indeed, when posed this way, one realizes that it is virtually impossible to write down a model in which there are not important externalities of the form we have described.

In some situations, these interactions may be far more important than in others. One of the phenomenon which has been a source of some discussion in recent years is the inter-linkage between credit markets and land markets in less developed countries. Some have contended that this is a manifestation of monopoly power on the part of landlords ; they attempt to use the credit market to squeeze the tenants still further.

[30] Provided that the level of care is not observable. If it is, premiums will depend on the level of care, and there will be no externalities.

The logical basis of these arguments is suspect, but it has, nonetheless, had some political impact. Our analysis has provided an alternative rationale for this phenomenon, one which is perfectly consistent with competitive behavior : share tenancy contracts can be viewed, as we have noted, as a form of insurance contract ; the landlords rents the land to the tenant, but then insures the tenant against a bad crop. There is thus a moral hazard problem : the tenant has insufficient incentives to avoid a bad crop. The availability of credit (the presence of debt) will affect the actions which the tenant undertakes, both his level of effort and his choice of technique. Moreover, the terms of the tenancy contract will affect the probability of a default on a loan. There are thus important externalties running both ways, and interlinkage is a method by which these externalities may be internalized (see Braverman-Stiglitz [1982]).

2. Intertemporal Interdependence. The interdependence of seemingly unrelated insurance contracts holds not only across markets within any period, but within the same market over time. The accident avoidance actions taken today may depend on the nature of the insurance coverage provided in future periods. Insurance companies may take advantage of this by making the terms of the contract in future periods be a function of performance (e.g. whether an accident has occurred) in earlier periods. Similarly, banks may make the terms of future loans depend on the performance of borrowers in earlier periods, even though the investment projects themselves are statistically independent, and even though previous performance conveys no information about the characteristics of the borrowers. Indeed, banks may find it optimal to cut off credit from those who default on their loans in earlier periods. [31] These interdependencies, like those discussed earlier, imply again (under the strong hypothesis that consumers are perfectly informed about insurance companies) that equilibrium will be characterized by the same insurance firm covering individuals over several periods. This result, like the exclusivity result described in the previous paragraph, has important implications for the nature of competition in the presence of moral hazard.

3. The Interdependence Through Price Effects. A similar argument shows that, in general, an individual's actions will be affected by the prices of a whole variety of goods which the individual purchases. Even with separable utility functions, demand functions (action functions) will not be independent. Some of these interdependencies are, of course, clearer than others. Thus, an individual's purchases of fire extinguishers (clearly affected by the price of fire extinguishers) may have a significant effect on the risk of a fire. One's first inclination is to suspect that it would decrease it, and although there is some presumption that this is true, the case is far from unambiguous. For the presence of a fire extinguisher may induce individuals to take less care, and the net effect may be an actual increase in the probability of a fire. This may occur even with well behaved utility functions.

The internalization of these externalities is far from straightforward. While in those cases where there is a large positive externality, the insurance firm can attempt to subsidize the good in question (say through coupons) when the externality is negative, it may be difficult for the firm (by itself) to restrict consumption, in the way the government might, say, through a tax.

---

[31] For a more extensive discussion of this, see Stiglitz and Weiss [1982].

4. The Interdependence Through Income Effects. Again, a similar argument shows that, in general, an individual's actions will be affected by his income, and again these effects generate a kind of externality. The individual's income in turn is affected by the level of profits, and the level of profits are affected by the composition of demand. This effect occurs even in an economy with a single kind of accident, if the cost of repair (d) is a function of the level of demand for repair services (i.e. there is not constant costs).

There is a fifth source of market failure, what Arnott and Stiglitz refer to as the " cross subsidization market failure ". In conventional models, so long as there is decreasing or constant returns to scale, the efficient resource allocation can be sustained with competitive prices, without any firm receiving lump sum subsidies. We show that this is not true in insurance markets characterized by moral hazard ; cross-subsidies are required. [32]

These intuitive arguments can be formalized, and they show clearly that the market equilibrium is Pareto efficient only under extremely restrictive conditions (perfect information, or no risk). Though these arguments destroy the widespread presumption in favor of the efficiency of market economies, they do not create a presumption for government intervention.

The standard analysis of the efficiency of market institutions compares the market allocation with that generated by a highly idealized " central planner ". Since the major result is that even such a central planner could do not better than the market, the fact that the central planner bears little resemblance to any actual government is of little concern. Here, however, we have obtained a quite different result : the market does not do as well as our " idealized " central planner. It is thus conceivable that government intervention might improve welfare, but to make any such assertion requires more accurate analysis of government behavior (including an analysis of incentive structures within the public sector). This we have not done here. Indeed, the kinds of corrective taxes and subsidies required here are quite complicated, and depend on parameters about which the government is unlikely to have reliable information. When there has been government intervention, it is not clear that it has been motivated by the concerns which we have raised here or indeed, that its interventions have been welfare increasing. We shall turn in a minute to a discussion of the implications of our analysis for the design of social insurance programs. First, however, I wish to say a few words about another kind of response to what are perceived as market failures.

5. Social Intervention

There is a crude functionalist view that suggests that, when markets fail, societies find ways of adapting to the market failure ; institutions are created to alleviate, at least partially, the consequences of these market failures. Thus, when the market fails

[32] This list of problems is not meant to be exhaustive. In our more extended treatment, we identify further problems with decentralization of the efficient resource allocation particularly in situations where the amount of insurance purchased from other firms is unobservable.

to provide complete insurance, as we have suggested it will, there are incentives for social adaptation : formal and informal insurance is provided by a host of mechanisms within our society (see K. Arrow [1963]). That there exist incentives for supplementing the insurance provided by the market when the market rations the amount of insurance that an individual can purchase seems clear from our earlier analysis. What does not seem obvious is that these social interventions lead to an increase in welfare, for the reason that the market restricted the amount of insurance was that with greater insurance, individuals took less care, less than is socially desirable, since they do not bear the full costs of their actions. We have investigated this question in some detail. [33] Two results emerge :

If these " supplemental " insurance contracts do not specify (implicitly or explicitly) the actions to be undertaken by the individual (i.e. his level of care) then welfare is unambiguously reduced. The intuition is simple : all of the possibilities of risk sharing have already been taken into account in the equilibrium insurance contract with risk neutral insurance firms. The benefits from additional risk sharing are exceeded by the costs resulting from the decrease in care. The only reason that supplemental insurance (which generally involves less efficient risk sharing) is thus desired is that some part of the costs of decreased care are imposed on the primary insurance company, rather than on those providing the supplemental insurance.

If the supplemental insurance contracts do specify (implicitly or explicitly) the actions to be undertaken by the individual, then the effect on welfare is ambiguous. There is the possibility that welfare may be increased, but we have not yet derived simple necessary or sufficient conditions characterizing those situations.

There are a variety of circumstances in which social intervention takes the form not only of providing additional insurance, but also controlling the actions of individuals. In many jobs, the direct financial incentives are limited — piece rates are not employed at all ; it is social coercion, of one form or another, which induces individuals to exert effort. Those who work together often have a far better knowledge of the level of effort expended than does the employer. If they realize that the wages that the firm can pay depend on the level of output of the group, as a group they have an incentive to work harder ; but each individual has an incentive to be a free rider. What stops him are the variety of social sanctions that his peers can impose. But the effectiveness of these social sanctions may depend on the social cohesiveness of the group : if an individual is unconcerned with what his fellow workers think of him, these social sanctions may be ineffective.

This in turn has two implications. First, social sanctions may be more effective within homogeneous groups and stable populations ; secondly, social sanctions may be more effective in smaller units of organization within the society (e.g. families) than in larger units. Thus, while individuals may be reluctant to cheat on their family, to claim disability when they are really capable of working, many individuals would show little compunction at cheating the State, if they could get away with it. Indeed, the " moral " obligations which individuals commonly feel to members of their family, and less commonly to members of their work unit, are seldom evidenced in individual's tax contributions, where tax avoidance has received widespread social approval.

---

[33] See R. Arnott and J. E. Stiglitz [1981 a], Part III.

Thus, the shift in the primary locus of " social " insurance which has occurred over the past century, from the family and local community, to the State, has undoubtedly been accompanied by an increase in the distortions and inefficiencies associated from what we have called the moral hazard problem.

6. Implications for social policy

Western governments have now almost universally accepted the principle that it is their responsibility to provide some kind of social insurance to their citizens. Some aspects of this social insurance are mislabeled : they are really redistribution programs. But some, such as disability insurance, old age insurance, and health insurance are at least in part real insurance programs. As coverage under these programs has been increased from their often initially modest levels, costs have risen, seemingly out of control. It would be well for governments to bear in mind the kinds of considerations with which we have been concerned here ; the fundamental trade-off between risk and incentives implies that complete insurance, as " just " and " fair " as it might seem, is not desirable. By changing the structure of these social programs, governments may be able to improve the incentives, without altering in a significant way the nature of the real risks being borne. For instance, for " small risks " individuals are approximately risk neutral. Thus, if the government is to provide medical insurance, it is important that it focus on major medical insurance (i.e. have large deductibility clauses).

Social security can be viewed partly as insurance against old-age disability and/or obsolescence — being unable to work effectively at 65 or older. The reason that payments are not explicitly made conditional on disability is that disability, particularly, partial disability, and obsolescence are hard to monitor. With complete insurance, one's income would be independent of whether one could work or not ; but with complete insurance, no one would have an incentive to work. With incomplete insurance, those with a greater ability to work (higher wages) continue working, while those with more significant disabilities or greater obsolescence do not. There is some evidence that the level of insurance implicit in the payments structure of social security — the incentives that it provides for those who are still productive to retire — may be too high ; the gains in additional risk reduction are outweighed by the social losses from the incorrect incentives for retirement that it provides.

More generally, the fact that the electorate feels that private markets have failed to provide adequate insurance — that they would like to reduce risks beyond the point which they can at present — should not by itself be taken as evidence of a market failure. [34]

7. Conclusions

In this paper, I have attempted to survey the fundamental relations between risk, incentives, insurance, and imperfect information. These relationships have been

[34] Our analysis here has focused on the problems raised by moral hazard. No analysis of insurance markets can be complete without, at the same time, dealing with the problems posed by adverse selection. Our conclusions must, accordingly, be treated with caution.

discussed in a variety of guises within the literature : under the rubric of the Theory of Moral Hazard within the insurance literature, and under the rubric of the Theory of Principal-Agent Relations in some of the recent theoretical literature. We have argued that understanding these relationships is important for three reasons.

First, it provides us with considerable insight into the nature of many of the institutional structures which we observe, and which cannot be explained within the context of the standard economic theory. Among these are the persistence of share-cropping as a form of tenancy over a large fraction of the world ; the interlinkage of credit and land markets in many l.d.c.'s ; the variety of forms of compensation which are observed in modern capitalist economies — the strict reliance on piece rates, as suggested by traditional economy theory, being seldom observed ; the complicated contractual arrangements between suppliers of credit and entrepreneurs ; and the use of co-insurance clauses and deductibility provisions and quantity restrictions (exclusivity provisions) in insurance markets.

Secondly, it forces us to re-examine the nature of competitive equilibrium and the assumptions which are conventionally employed in competitive analysis. We have shown that indifference curves will, in general, not be quasi-concave, and as a result, income consumption and price consumption curves will in general not be continuous. Competitive equilibrium may well not exist even when price consumption curves are continuous. When it does exist, it may look considerably different from standard competitive equilibria : if quantities of insurance are observable, there will be quantity rationing ; if they are not, there may be positive profit equilibria, while at the zero profit equilibrium, the level of care (effort) may be set at its minimal level. In either case, market equilibrium does not have the optimality properties we usually associate with competitive markets : we have identified five distinct reasons that markets will not, in general, be efficient. At the same time, the nature of the interventions that may be required by the government are such that we are not sanguine about the possibilities of improvements. Moreover, we have shown that the kinds of social institutions which have frequently developed, in response to the limitations on insurance provided by the market, may actually lower welfare rather than increase it.

Thirdly, our analysis has, we think, important implications for the design of social insurance programs. The same considerations which make it desirable for the market to restrict the quantity of insurance provided make it also desirable that the government provide only limited insurance and suggest that the government ought to be more concerned in the design of these insurance programs with the fundamental incentive problems, which are inherent in the provision of any insurance, than it has been in the past.

REFERENCES

ARNOTT, R. J., and STIGLITZ, J. E. [1981 a] : " Equilibrium in competitive insurance markets : The welfare economics of moral hazard ", mimeo, August 1981. (Part I : The basic analytics of consumer behavior in insurance markets with moral hazard ; Part II : Competitive equilibrium ; Part III : Welfare economics.)

ARNOTT, R. J., and STIGLITZ, J. E. [1981 b] : " Labor turnover, wage structures and moral hazard : The inefficiency of competitive markets ", Econometric Research Programme, Research Memorandum 289, September 1981.

ARNOTT, R. J., HOSIOS, A., and STIGLITZ, J. E. [1980] : " Implicit contracts, labor mobility and unemployment ", mimeo, December 1980.

ARROW, K. J. [1963] : " Uncertainty and the welfare economics of medical care ", *American Economic Review*, 53 (December 1963), 941-973.

ARROW, K. J. [1964] : " The role of securities in the optimal allocation of risk bearing ", *Review of Economic Studies*, 31 (April 1964), 91-96.

ARROW, K. J. [1965] : *Aspects of the Theory of Risk Bearing*, Yrjo Jahnssönin Säätio, Helsinki.

ATKINSON, A. B., and STIGLITZ, J. E. [1976] : " The design of tax structure : Direct vs. indirect taxation ", *Journal of Public Economics*, 6 (July-August 1976), 55-75.

AZARIADIS, C. [1975] : " Implicit contracts and underemployment equilibria ", *Journal of Political Economy*, 83 (December 1975), 1183-1202.

AZARIADIS, C. [1982] : " Employment with asymmetric information ", *Quarterly Journal of Economics*, (forthcoming).

AZARIADIS, C., and STIGLITZ, J. E. [1982] : " Introduction " to *Quarterly Journal of Economics*, Symposium on Rigid Wages and Unemployment, (forthcoming).

BAILY, M. N. [1974] : " Wages and employment under uncertain demand ", *Review of Economic Studies*, 41 (January 1974), 37-50.

BRAVERMAN, A., and STIGLITZ, J. E. [1981] : " Cost sharing arrangements under sharecropping : A multidimensional incentive problem ", World Bank Discussion Paper.

BRAVERMAN, A., and STIGLITZ, J. E. [1982] : " Sharecropping and the interlinking of agrarian markets ", *American Economic Review*, 72 (September 1982), 695-715.

DEBREU, G. [1959] : *Theory of Value*, John Wiley and Sons, New York.

GREENWALD, B., and STIGLITZ, J. E. [1982] : " The welfare economics of imperfect information and incomplete markets ", Bell Labs, mimeo.

GROSSMAN, S., and HART, O. [1981] : " Implicit contracts, moral hazard and unemployment ", *American Economic Review*, 71 (May 1981), 301-307.

HELLWIG, M. [1983] : " Moral hazard and monopolistically competitive insurance markets ", *Geneva Papers on Risk and Insurance*, this issue.

JAYNES, G. G. [1978] : " Equilibria in monopolistically competitive insurance markets " *Journal of Economic Theory*, 19 (1978), 394-422.

NALEBUFF, B., and STIGLITZ, J. E. [1982] : " Prizes and incentives ", Econometric Research Programme, Research Memorandum 298.

NEWBERY, D., and STIGLITZ, J. E. [1982] : " Wage rigidity, implicit contracts, and economic efficiency : Are market wages too flexible ? ", Princeton University, mimeo.

PAULY, M. [1974] : " Overprovision and public provision of insurance ", *Quarterly Journal of Economics*, 88 (February 1974), 44-62.

ROTHSCHILD, M., and STIGLITZ, J. E. [1976] : " Equilibrium in competitive insurance markets : An essay on the economics of imperfect information ", *Quarterly Journal of Economics*, 90 (November 1976), 629-649.

SHAPIRO, C., and STIGLITZ, J. E. [1982] : " Equilibrium unemployment as a worker discipline device ", Woodrow Wilson School, Princeton University.

SHAVELL, S. [1979] : " On moral hazard and insurance ", *Quarterly Journal of Economics*, 93 (November 1979), 541-562.

SMALLWOOD, D. [1975] : " Competition, regulation, and production quality in the automobile insurance industry ", in *Promoting Competition in Regulated Markets*, A. Phillips (Ed.), Brookings Institution, Washington D.C., 241-300.

STIGLITZ, J. E. [1974] : " Incentives and risk sharing in sharecropping ", *Review of Economic Studies*, 41 (April 1974), 219-255.

STIGLITZ, J. E. [1975] : " Incentives, risk, and information : Notes towards a theory of hierarchy ", *Bell Journal of Economics and Management Science*, 6 (Autumn 1975), 552-579.

STIGLITZ, J. E. [1982] : " Self-selection and Pareto efficient taxation ", *Journal of Public Economics*, 17 (... 1982), 213-240.

STIGLITZ, J. E. [forthcoming] : *Information and Economic Analysis.*

STIGLITZ, J. E., and WEISS, A. [1981] : " Credit rationing in markets with imperfect information : Part I ", *American Economic Review*, 71 (June 1981), 393-410.

STIGLITZ, J. E., and WEISS, A. [1982] : " Cutting off credit : An application of constraints as incentive devices ", Bell Labs, mimeo.

WILSON, C. [1977] : " A model of insurance markets with incomplete information ", *Journal of Economic Theory*, 16 (December 1977), 167-207.