

NBER WORKING PAPERS SERIES

EQUILIBRIUM IN COMPETITIVE INSURANCE MARKETS  
WITH MORAL HAZARD

Richard Arnott

Joseph Stiglitz

Working Paper No. 3588

NATIONAL BUREAU OF ECONOMIC RESEARCH  
1050 Massachusetts Avenue  
Cambridge, MA 02138  
January 1991

Any opinions expressed are those of the authors and not those of  
the National Bureau of Economic Research.

NBER Working Paper #3588  
January 1991

EQUILIBRIUM IN COMPETITIVE INSURANCE MARKETS  
WITH MORAL HAZARD

ABSTRACT

This paper examines the existence and nature of competitive equilibrium with moral hazard. The more insurance an individual has, the less care will he take. Consequently, insurance firms attempt to restrict their clients' aggregate insurance purchases. If individuals' aggregate insurance purchases are observable, each firm will ration the amount of insurance its clients can purchase and insist that they purchase no insurance from other firms.

This paper focuses on the alternative situation where firms cannot observe their clients' aggregate insurance purchases. We show that firms will still attempt to restrict their clients' aggregate purchases, but now they must do so indirectly. One possibility is that all firms sell only policies with a sufficiently large amount of coverage that individuals choose to purchase insurance from only one firm. Another possibility is that each firm offers a latent policy in addition to its regular policy. Latent policies are not purchased in equilibrium, but serve to restrict entry. If an entering firm offers a supplementary policy, an individual will purchase not only this policy plus his previous policy but also the latent policy. The latent policy is designed so that the individual reduces effort by enough to render any entering policy unprofitable.

Richard Arnott  
Department of Economics  
Boston College  
Chestnut Hill, MA 02167

Joseph Stiglitz  
Department of Economics  
Stanford University  
Stanford, CA 94305

# Equilibrium in Competitive Insurance Markets with Moral Hazard\*

## I. Introduction

It is now widely recognized that moral hazard-incentive problems are pervasive in the economy. In situations where moral hazard problems arise, firms make use of *indirect control* mechanisms to alter behavior in the desired way. Among the most important are *quantitative limits*. Insurers limit the amount of insurance which an insured can purchase, recognizing that the more insurance he has, the less care will he take (Pauly, 1974). Lenders limit the extent of indebtedness of a borrower, recognizing that the greater his indebtedness, the more likely he is to take actions which result in default (Eaton and Gersovitz (1981)). Employers, such as universities, limit their employees' outside employment (e.g. to one day of consulting a week), recognizing that the level of effort exerted on behalf of the university will be reduced by moonlighting. Landlords attempt to limit the employment of tenants working under sharecropping contracts for similar reasons.

The effectiveness of these quantitative limitations depends on the information available.<sup>1</sup> Universities often find it impossible to monitor outside consulting. Insurers find it difficult to monitor insurance provided by others.<sup>2</sup> Where firms and principals are unable to enforce quantitative limits, there are other mechanisms they can employ to control indirectly quantities. Insurers design their contracts to deter their customers from purchasing insurance from others, and to deter other insurers from selling insurance to their customers. Similarly, employers design their labor contracts to deter their employees from wanting to moonlight, and to deter other employers from providing moonlighting opportunities.

The objective of this paper is to examine the existence and nature of market equilibrium

---

\* We are grateful to the National Science Foundation, the Olin Foundation, the Hoover Institution, and the Social Sciences and Humanities Research Council of Canada for financial support. We would like to thank Oliver Hart, Martin Hellwig, Bentley MacLeod, Urs Schweizer, Thomas Von Ungern-Sternberg, and seminar participants at the University of Pennsylvania for helpful comments. Remaining errors are our responsibility.

<sup>1</sup> This in turn depends on the cost of acquiring information, legal restrictions concerning what information can be collected and how it may be acquired, the extent of communication between principals concerning their agents' activities, etc.

<sup>2</sup> This is particularly true because insurance is provided not only by insurance firms, but also by other institutions in our society. Sick leave provided by firms for their employees is a form of health insurance. For a more extended discussion, see Arnott and Stiglitz (1990a).

under these circumstances.<sup>3</sup> We focus our analysis on the insurance market, because it provides the simplest canonical model in which these questions can be examined. It should be clear, however, that the general results obtained here apply in other contexts in which moral hazard arises and direct quantitative restrictions are unenforceable.

Even when an insurer cannot monitor his clients' outside purchases of insurance, he can monitor his own sales to them. By requiring his customers to purchase sufficiently large policies, he might discourage them from purchasing insurance from other firms. We show that, while firms will try to do this, this mechanism does not in general solve the over-insurance problem which arises when quantities are unobservable. Also, the attempt to limit the quantity of insurance may result in there being no Nash equilibrium in the market.

The actions which an individual takes to prevent the insured-against accident depend on his total purchases of insurance; and his total purchases depend on the entire set of insurance policies which are available to him. This introduces what turns out to be a powerful commitment for deterring entry: An incumbent insurer can offer a set of *latent* insurance policies, in addition to the policy he sells in equilibrium. These latent policies are not purchased in equilibrium. But if an entrant offers a new policy which the incumbent's clients find attractive enough to buy, they will also purchase the latent policies. Having purchased a lot of additional insurance, they will reduce care, which will increase the probability of accident by enough that the new policy makes a loss. We show that such latent insurance policies are extremely effective in deterring entry, so effective that there exist positive profit equilibria. Thus, while attempts to control the moral hazard problem by offering large policies aimed at discouraging insurance purchases from other firms turn out not to be successful, attempts to deter entry by offering latent policies turn out to be, in a sense, too successful.

## II. Model and Notation

We employ a very simple model in which individuals are identical (to abstract from the adverse selection problem). There is a single, fixed-damage accident, the probability of

---

<sup>3</sup> Important contributions to the literature on the topic include Pauly [1974], Helpman and Laffont [1975], Stiglitz [1983], and Hellwig [1983a,b].

which,  $p$ , is a function of the level of effort devoted to accident avoidance,  $e$ . Moral hazard arises because the insurer is unable to observe individuals' effort, and is hence unable to write contracts contingent on effort; instead he must write insurance contracts contingent on which of the two events, "accident" or "no accident," occurs. Then the more insurance he offers, the lower the individual's effort; hence, there is a tradeoff between risk-bearing and incentives, which is the hallmark of the moral hazard problem. In the absence of insurance and in the event of no accident, the individual's consumption is  $w$ , while if there is an accident it is  $w - d$ , where  $d$  is the accident damage. The insured individual receives  $\alpha$  (the "(net) payout" or "benefit") if an accident occurs, and if it does not he pays  $\beta$  (the "premium"). Thus,

$$y_0 = w - \beta, \quad y_1 = w - d + \alpha, \quad (II.1)$$

where  $y_0$  is consumption in the event of no accident, and  $y_1$  consumption in the event of accident. The individual's expected utility is

$$EU = (1 - p(e))u_0(y_0, e) + p(e)u_1(y_1, e). \quad (II.2)$$

We make the simplifying assumption that the expected utility has the form

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

This function is *separable* in consumption and effort, and is *event-independent*—which event occurs does not affect the utility from consumption (the accident does not cause pain, nor does it alter tastes). There is positive but diminishing marginal utility from consumption. Also, we measure effort so that  $e = 0$  corresponds to minimum effort and assume that  $p' < 0$  and  $p'' > 0$  for  $e > 0$ , and  $p(0) \equiv \bar{p} < 1$  (which rules out the possibility that the insured will deliberately cause the accident.)

We shall develop our analysis in  $\alpha$ - $\beta$  space. For the purpose of illustration, suppose that insurance firms are offering *quantity* policies, which specify both a premium and a payout. Diagrammatically, a positive quantity insurance policy is represented by a positively-sloped, upward-pointing arrow (such as  $KK_1$  in Figure II.1), and a negative quantity insurance policy (one in which the individual pays if the accident occurs and receives a payout if it does not) by a positively-sloped, downward-pointing arrow (such as  $KK_2$  in Figure II.1). An

individual's aggregate insurance purchases are obtained as the vector sum of all the policies he purchases; thus, in terms of Figure II.1,  $OK + KK_1 = OK_1$  for example. We define  $q \equiv \frac{\beta}{\alpha}$  to be the price of insurance corresponding to  $(\alpha, \beta)$  (geometrically, it is the slope of a ray from the origin to  $(\alpha, \beta)$ ).  $\alpha$  is the "quantity" of insurance.

An individual with aggregate insurance purchases characterized by  $(\alpha, \beta)$  chooses effort to maximize expected utility; i.e.,

$$\max_e EU = (1 - p(e))u_0 + p(e)u_1 - e \equiv V(\alpha, \beta), \quad (II.3)$$

where  $u_0 \equiv u(w - \beta)$  and  $u_1 \equiv u(w - d + \alpha)$ . The first-order condition is

$$-(u_0 - u_1)p'(e) - 1 = 0 \quad \text{if } e > 0. \quad (II.4)$$

This gives  $e$  as a function of  $\alpha$  and  $\beta$ ,  $e(\alpha, \beta)$ , which, under the assumption of separable utility, is a continuous function of  $(\alpha, \beta)$ .<sup>4</sup> Furthermore,  $\frac{\partial e}{\partial \alpha} < 0$  and  $\frac{\partial e}{\partial \beta} < 0$  for  $e > 0$ ; as more insurance is provided, the individual reduces effort. Substituting  $e(\alpha, \beta)$  into the expression for expected utility gives  $V(\alpha, \beta)$ , the indifference curves corresponding to which can be plotted in  $\alpha$ - $\beta$  space.

The individual's marginal rate of substitution between the premium and net payout is

$$\left. \frac{d\beta}{d\alpha} \right|_V = -\frac{V_\alpha}{V_\beta} = \frac{u'_1 p}{u'_0(1-p)}. \quad (II.5)$$

As more insurance is provided, diminishing marginal utility causes  $\frac{u'_1}{u'_0}$  to decrease, but individuals take less care, and as a result  $\frac{p}{1-p}$  increases. For this reason, indifference curves between net payout and premium will not in general be convex<sup>5</sup> (see Figure II.1). Because of this, even though, under our assumptions, effort is a continuous function of  $(\alpha, \beta)$ , an individual's purchases of insurance may not be a continuous function of, say, the price of insurance. This greatly complicates the analysis to follow.

The set of aggregate, non-negative insurance purchases for which expected returns are non-negative,

$$\pi(\alpha, \beta) \equiv \beta(1 - p(e(\alpha, \beta))) - \alpha p(e(\alpha, \beta)) \geq 0,$$

<sup>4</sup> The separability assumption was made to circumvent analytical complexities which arise when effort is not a continuous function of  $(\alpha, \beta)$ , which may be the case with a nonseparable utility function; see Arnott and Stiglitz (1988a).

<sup>5</sup> We say that indifference curves are convex when they have the normal shape, even though in  $\alpha - \beta$  space normally-shaped indifference curves are concave.

is referred to as the *feasibility set* and denoted by  $\mathcal{F}$ . The boundary of this set is referred to as the *resource constraint* or *zero profit locus* (ZPL) (see Figure II.1). As one moves up the zero profit locus, effort falls, the probability of accident increases, and to maintain zero profits the ratio  $\frac{\beta}{\alpha}$  must increase. The feasibility set is never convex. It is the lack of convexity of the feasibility set, combined with the lack of quasiconcavity of  $V(\alpha, \beta)$ , which give rise to the problems that are the concern of this paper.

It will be useful, before proceeding, to consider some aspects of the geometry of the problem. First, because of the possible nonconvexity of indifference curves, price-consumption lines and income-consumption lines can be discontinuous.<sup>6</sup> This has important implications for the analysis that follows. Second, because our assumption of separable utility implies that with  $e > 0$  effort falls as  $\alpha$  or  $\beta$  increases, the probability of accident rises monotonically as one moves up any ray from the origin. Third, the zero effort line (ZEL) is the locus of  $(\alpha, \beta)$  such that  $(u_0 - u_1) \lim_{e \downarrow 0} p'(e) + 1 = 0$  (to simplify notation, we shall subsequently use  $p'(0)$  to denote  $\lim_{e \downarrow 0} p'(e)$ , etc.) and has slope  $\frac{d\beta}{d\alpha} \Big|_{ZEL} = -\frac{u'_1}{u'_0}$ . Effort is zero everywhere beyond the zero effort line. Fourth, the full insurance line (FIL), the locus of  $(\alpha, \beta)$  corresponding to which the marginal utility of consumption is the same in the accident and no-accident events, is defined implicitly by  $u'_0 = u'_1$  which with event-independence implies  $d = \alpha + \beta$ . If  $p'(0) = -\infty$ , then  $u_0 = u_1$  along the zero effort line, which, under our assumption of event-independent utility, implies that the zero effort line coincides with the full insurance line. If, however,  $p'(0)$  is finite, then  $u_0 > u_1$  along the zero effort line, which with event-independent utility implies that less than full insurance is provided along the zero effort line and that the full insurance line lies outside the zero effort line, the case depicted in Figure II.2.

A limiting case<sup>7</sup> of this general model is that where there are two activities, a safe activity and a risky activity;  $e$  represents the fraction of the individual's time engaged in the safe activity, and  $p(e) = ep^s + (1 - e)p^r$ , where  $p^s(p^r)$  is the accident probability associated with the safe (risky) activity. In this case, the indifference curves are *never* convex; they appear as in Figure II.3, where we have also drawn the zero profit locus.

<sup>6</sup> The price-consumption line is defined in the usual way as the locus of points of maximum utility on price lines (rays through the origin). The income-consumption line is the locus of points of maximum utility on parallel budget constraints.

<sup>7</sup> This is a limiting case since  $p''(e) = 0$ .

### III. Definition of Equilibrium and the Nature of the Problem

#### III.1 *Definition of equilibrium*

Equilibrium is defined in a similar way to Rothschild-Stiglitz [1976], which examined equilibrium in insurance markets with adverse selection. Each firm offers a single insurance contract in a *class of admissible contracts*, or does nothing. A Nash equilibrium in admissible insurance contracts is defined as a set of admissible contracts such that: i) all contracts offered at least break even; ii) taking as given the contracts offered by incumbent firms (those offering contracts) there is no additional contract which if offered by an entering firm (one not offering a contract) can make a strictly positive profit; and iii) taking as given the set of contracts offered by other incumbent firms, no incumbent firm can increase its profits by altering the contract it offers.

We term such contracts competitive because each insurance firm treats other firms anonymously and believes it will have no effect on other firms' actions (the standard Nash assumption), because entry and exit are free, and because no barrier to entry or scale economies are present.

It should be emphasized that equilibrium is defined conditional on the set of admissible contracts. Thus, determining the appropriate equilibrium concept entails identifying the appropriate set of admissible contracts. We shall define various equilibrium concepts, some in which the set of admissible contracts is restricted exogeneously, and some in which the set is restricted by the information available to insurers. Among the restrictions we consider are those where the firm can restrict the *total* quantity of insurance a client purchases, where the firm can restrict the quantity of insurance a client purchases from itself, and where the firm can impose no quantity restriction. Other characteristics of the set of admissible contracts which need to be specified are whether negative or stochastic (where, for example, the payout is stochastic) insurance is allowed. Furthermore, it matters whether a contract contains only a single, simple policy or instead a set of policies from which the client may mix and match, and whether inactive (not purchased in equilibrium) policies are admitted.

We start out by imposing considerable restrictions on the set of admissible contracts, and then gradually loosen these restrictions. To begin, we disallow negative and stochastic

insurance, permit each company to offer only a single, simple policy in its contract, and exclude inactive policies.

### III.2 *Exclusive contract and price equilibria: two limiting cases*

If effort were observable, the natural set of admissible contracts would include those in which payouts and premia are contingent on effort. The equilibrium in this case would be (equivalent to) the standard competitive equilibrium. But we are concerned with situations in which effort is not observable.

Though effort cannot be controlled directly, firms know that effort is affected by the quantity of insurance. Hence, *when firms can monitor either their clients' purchases of insurance or payouts from other insurance firms,*<sup>8</sup> *they will ration clients' purchases by insisting that they be exclusive agents,*<sup>9</sup> not permitting their clients to purchase insurance from other firms. We term the corresponding equilibria, *exclusive contract equilibria.*<sup>10</sup> This type of equilibrium was first described by Pauly [1974] and is examined in Arnott and Stiglitz [1988a].

In the other polar case, *firms offer price contracts, imposing no quantity restrictions on insurance purchases, except that negative purchases are disallowed. In the resulting price equilibrium, if profits are zero and insurance purchases strictly positive, there will be no effort at accident avoidance.* We call this the *zero profit price equilibrium.*<sup>11</sup> Since individuals can choose how much insurance to purchase at the equilibrium price, equilibrium lies at the intersection of the price-consumption line and the zero profit locus (see Figure III.1). With strictly positive insurance, at any point on the price-consumption line, the price line with slope  $q$  is tangent to an indifference curve with slope  $\frac{u'_1 p}{u'_0(1-p)}$ . Since zero profits imply  $q = \frac{p}{1-p}$ , at the intersection point  $u'_1 = u'_0$  — there is full insurance. Thus, equilibrium is at the point of intersection of the full insurance line and the zero profit locus, which we

<sup>8</sup> It should be noted that to enforce exclusivity provisions, an insurance company does not actually have to monitor its clients' purchases; all it has to do is make sure that its clients do not receive payouts from other insurance companies.

<sup>9</sup> Insurers could also quote customers their zero profit locus, which is essentially what Helpman and Laffont [1975] have insurers do, but this is unnecessarily complicated.

<sup>10</sup> If the quantities of goods and services consumed by the individual are observable, the insurance policy will specify the levels of all goods or services which affect the probability of the insured-against event, either directly or, indirectly, through an effect on  $e$ .

<sup>11</sup> This is the price equilibrium identified in Pauly [1974].

label throughout the paper as  $E$ . Furthermore, with full insurance, effort is zero. Hence, the equilibrium price of insurance is  $q^* \equiv \frac{p(0)}{1-p(0)}$  (see Figure III.1).

Because indifference curves need not be convex, there may not be an intersection of the zero profit locus and price-consumption line,<sup>12</sup> in which case the price equilibrium cannot be at the point  $E$ . A full characterization of the price equilibrium is provided in Arnott and Stiglitz (1990b). For the purposes of this paper, it suffices to note that a unique price equilibrium always exists, at the lowest point on the price-consumption line consistent with non-negative profits. The equilibrium may be one of three types – a zero profit price equilibrium (zero profit, zero effort, full insurance), a positive profit price equilibrium (positive profit, positive effort, partial insurance),<sup>13</sup> or a zero insurance price equilibrium. Both the positive profit and zero insurance price equilibria occur because the price-consumption line has a discontinuity across the zero profit locus. A positive profit price equilibrium is shown in Figure III.2, and a zero insurance price equilibrium in Figure III.3.

We can use the diagrammatic apparatus to show the gains from exclusivity. In Figure III.1, the exclusive contract equilibrium<sup>14</sup> occurs at  $\theta$ , the point of highest utility on the zero profit locus. The exclusive equilibrium contract is (constrained) efficient,<sup>15</sup> conditional on the unobservability of effort, and clearly dominates all the price equilibria.

<sup>12</sup> The price-consumption line is the locus of points of maximum utility for the family of price lines  $\beta = q\alpha$  with  $q > 0$  and  $\alpha \geq 0$ . A point of tangency of an indifference curve and a price line  $\beta = q\alpha$  need not be on the price-consumption line. First, because of the possible nonconvexity of indifference curves, there may be multiple points of tangency – i.e., the tangency condition picks out local extrema while only the global maximum is on the price-consumption line. Second, a corner point, the origin, may be the point of maximal utility on a price line.

<sup>13</sup> See Figure III.2. Since effort decreases in the amount of insurance offered, if effort is zero at  $H$ , it is zero beyond  $H$ . This implies that the indifference curve passing through  $H$  is convex beyond  $H$ , which is inconsistent with a jump discontinuity from  $H$  to  $H'$ . Since  $H$  is on the price-consumption line

$$\left. \frac{d\beta}{d\alpha} \right|_V = \frac{u'_1 p^H}{u'_1 (1-p^H)} = q^H > \frac{p^H}{1-p^H}$$

(positive profits)  $\Rightarrow u'_1 > u'_0 \Rightarrow$  partial insurance.

<sup>14</sup> Because of nonconvexities, a profitable random exclusive contract may, in some circumstances, upset the deterministic exclusive contract equilibrium. The reason is that even though contract randomization *per se*, by exposing the individual to additional uncertainty, is unambiguously harmful, it may cause the individual to increase effort by so much that the firm is able to offer insurance at sufficiently more favorable terms that utility is increased. This point is developed in Arnott and Stiglitz [1988b].

<sup>15</sup> This result depends on the fact that there is a single good. With more than one good, constrained efficiency requires that insurance contracts must, in addition, specify *levels* of consumption of all goods and services; see Arnott and Stiglitz (1990c).

As noted in the introduction, when an insurance firm cannot directly control its clients' purchases of insurance, it may try to control them indirectly. Thus, limiting the admissible contracts to only price contracts when insurance purchases are unobservable is unduly restrictive. It remains to be seen whether, or under what circumstances, equilibrium with unobservable purchases coincides with the price equilibrium. Though an insurance firm cannot monitor its clients' insurance purchases from *other* firms, it typically does know its own sales and the set of insurance contracts offered by other firms, as well as clients' tastes. On the basis of this information, it can compute how the contract it offers will affect its clients' aggregate purchases of insurance and accordingly their effort, from which it can derive the profitability of its contract. It is this complicated indirect control problem which is the subject of inquiry for the remainder of the paper. We proceed with the analysis in three steps. In the next section, we assume that *all* firms offer *only* quantity contracts; we refer to an equilibrium in which only such contracts are admissible as a *Q-equilibrium*. In Section V, both price and quantity contracts are admissible, and the resulting equilibria are termed *PQ-equilibria*. In both these sections, we assume that only active policies are admissible – the only policies allowed in equilibrium are those that are actually purchased in equilibrium. Then in Section VI, we consider *PQ-equilibria* with latent policies (*LPQ-equilibria*).

We summarize the results of this section in Proposition 1 and 2:

**Proposition 1:** When firms can restrict their clients' aggregate insurance purchases, the form of competitive equilibrium is the *exclusive contract equilibrium*. In this equilibrium, firms sell a quantity contract and do not permit their clients to purchase insurance from other firms. The equilibrium quantity contract is Pareto efficient, conditional on the unobservability of effort, and rations the quantity of insurance individuals may purchase at the equilibrium price.

**Proposition 2:** When the class of admissible contracts is restricted to include only *price contracts*, in which individuals can purchase as much insurance as they wish at the contract price, the form of competitive equilibrium is a *price equilibrium*. The price equilibrium occurs at the lowest point on the

price-consumption line consistent with non-negative profits. The price equilibrium is unique and is one of three types. If indifference curves are convex (in  $\alpha$ - $\beta$  space), the price equilibrium is a zero profit price equilibrium (zero profit, zero effort, full insurance). When indifference curves are nonconvex, the price equilibrium may be of this type. But if the price-consumption line jumps across the zero profit line, it may also be a positive profit price equilibrium (positive profit, positive effort, partial insurance) or a zero insurance price equilibrium.

#### IV. (Non-Exclusive) Quantity Equilibria

We now consider the situation in which the class of admissible contracts includes only quantity contracts. In a quantity contract, a firm may offer only one policy which specifies both the quantity and price of insurance and which is active. Furthermore, firms cannot monitor their clients' aggregate purchases of insurance. As a result, if an equilibrium exists, it is a *non-exclusive* quantity equilibrium.

##### IV.1 *The nature of the problem*

To provide insight, we first investigate possible *quantity equilibria* in which there is a single incumbent firm. First, we shall demonstrate that, at the exclusive contract equilibrium, as long as effort is positive – the normal case – there are profitable supplementary quantity contracts which would be bought. As a result, in the normal case, the exclusive contract equilibrium is not a quantity equilibrium. Turn to Figure IV.1, in which a ray from the origin through the exclusive contract equilibrium point is drawn. Since at  $\theta$  the individual is rationed in the amount of insurance he can purchase at the price  $q^\theta = \frac{p^\theta}{1-p^\theta}$ , a small supplementary contract ( $\theta Z$  in the diagram) offered at a price  $q$  such that  $\frac{d\beta}{d\alpha} \Big|_{V^\theta} > q > \frac{p^\theta}{1-p^\theta}$  will be bought (since  $\frac{d\beta}{d\alpha} \Big|_{V^\theta} > q$ ) and make a profit (since  $q > \frac{p^\theta}{1-p^\theta}$ ). But since  $Z$  lies outside the zero profit locus, the individual's aggregate insurance purchases generate a loss. Since the small supplementary contract  $\theta Z$  makes a profit, the equilibrium exclusive contract makes a loss. The reason is that, with the additional purchase of  $\theta Z$ , the individual reduces his effort, so that while  $\theta$  broke even with effort level  $e^\theta$ , it makes a loss with the effort level

corresponding to the higher level of insurance with the supplementary policy  $\theta Z$ .

By a similar argument (see Figure IV.2), when only quantity contracts are admissible and exclusive contracts are unenforceable, a point on the zero profit locus can be a quantity equilibrium only if the indifference curve through the point is tangent to the ray from the origin through the point. Consider, for example, the point  $A'$  in the Figure; it can be upset as an equilibrium by the small supplementary policy  $A'Z'$ . The slope of an indifference curve through a point on the zero profit locus is  $\frac{u'_1 p}{u'_0(1-p)}$ , while the slope of a ray from the origin to the point is  $\frac{p}{1-p}$ . There is only one point on the zero profit locus at which the two slopes are equal, where  $u'_1 = u'_0$  - where the full insurance line intersects the zero profit locus, point  $E$ . It can also be shown that any point in the interior of the feasibility set can be upset by a small, supplementary quantity contract.

Is  $E$  then a (non-exclusive) quantity equilibrium, and, if it is, is it the unique equilibrium? The answer to both questions is: Not necessarily. What we have argued is that  $E$  is the only point which can be a quantity equilibrium if there is a single incumbent firm, since it is the only one that cannot be upset by a small, supplementary quantity contract. We have not shown, however, that  $E$  cannot be upset by a *large* quantity contract, nor that other equilibria are not possible with more than one incumbent firm, where individuals combine the contracts offered by different firms.

#### IV.2 Definition of equilibrium

Suppose there are  $J$  firms offering (non-exclusive) quantity contracts, with firm  $j$  offering contract  $(\alpha_j, \beta_j)$ . The individual is permitted to purchase any subset of contracts, but only one contract per firm. Where  $K$  is the index set of contracts chosen by the individual, we define his aggregate insurance purchases to be  $(\alpha, \beta) \equiv (\sum_{k \in K} \alpha_k, \sum_{k \in K} \beta_k)$ . Then a *quantity* or *Q-equilibrium* is defined to be a set of active (i.e. purchased in equilibrium) contracts,  $\{\alpha_j, \beta_j\}$ , such that: (i) all contracts at least break even; (ii) individuals may purchase any subset of  $\{\alpha_j, \beta_j\}$  but only one contract per firm; (iii) taking the set of  $\{\alpha_j, \beta_j\}$  as given, there is no additional contract which, if offered by an entering firm, would make a profit; and (iv) taking the set  $\{\alpha_j, \beta_j\}_{j \neq i}$  as given, firm  $i$  cannot increase its profits by offering contract  $(\alpha'_i, \beta'_i)$  instead of  $(\alpha_i, \beta_i)$ .

There is potentially a vast number of such equilibria. For one thing, there may be an infinity of market configurations which can sustain a given aggregate equilibrium. For example, if, in equilibrium, everyone purchases  $(\alpha^*, \beta^*)$ , there may be: (i) only one firm selling  $(\alpha^*, \beta^*)$  from which everyone buys; (ii) as many firms as individuals, each selling  $(\alpha^*, \beta^*)$  with each firm selling a contract to only one individual; (iii) two firms, each selling contract  $[\frac{\alpha^*}{2}, \frac{\beta^*}{2}]$  with everyone buying from both firms, etc. This by itself poses no great difficulty, since one can define equilibrium in terms of individuals' aggregate purchases. But, more problematically, there is potentially a vast number of aggregate equilibria. The set of equilibria may be different when there are two active firms each servicing half the population than when there is only one firm in the market, or when there are two active firms each providing the entire population with half its aggregate purchases. And so on. We term the number of firms in the market along with the contract each offers and the number of clients each has, the "market structure".

Since there is no way we can characterize all possible quantity equilibria, we have chosen to treat the case in which *in equilibrium*: (i) all firms supply the same contract; (ii) there are a large number of firms with each firm supplying all of a particular individual's insurance; and (iii) individuals are *able* to purchase integer multiples of the equilibrium contract (though, by (ii), they will choose to purchase only a single contract in equilibrium). In this case, the characterizations of equilibrium in terms of the set of contracts offered and in terms of an individual's aggregate insurance purchases are equivalent. Thus

$(\alpha^*, \beta^*)$  is a  $Q$ -equilibrium iff:

(a)  $\beta^* [1 - p(e(\alpha^*, \beta^*))] - \alpha^* p(e(\alpha^*, \beta^*)) \geq 0$

(b)  $V(\alpha^*, \beta^*) \geq V(n\alpha^*, n\beta^*)$  for  $n = 2, \dots, N$  (the population)

(c)  $\nexists$  a policy  $(\tilde{\alpha}, \tilde{\beta})$  such that:

$\exists$  integer  $n = 0, \dots, N$  such that

$$V(n\alpha^* + \tilde{\alpha}, n\beta^* + \tilde{\beta}) > V(\alpha^*, \beta^*) \text{ and}$$

$$\tilde{\beta} [1 - p(e(n\alpha^* + \tilde{\alpha}, n\beta^* + \tilde{\beta}))] - \tilde{\alpha} p(e(n\alpha^* + \tilde{\alpha}, n\beta^* + \tilde{\beta})) \geq 0.$$

In words,  $(\alpha^*, \beta^*)$  is a  $Q$ -equilibrium iff: (a) it makes non-negative profits; (b) it is weakly preferred to any integer multiple of itself ( $n = 2, \dots, N$ ); and (c) there is no contract

which by itself or in combination with an integer multiple of  $(\alpha^*, \beta^*)$ , would increase utility and would not make a loss. Note that negative quantity contracts are admissible.

### IV.3 *Multiple purchases disallowed*

To simplify the discussion, we begin by disallowing multiple purchases of the candidate equilibrium contract. It is clear that equilibrium must lie along the zero profit locus, since if the candidate equilibrium were strictly inside the feasibility set, there would exist a small, supplementary policy which would be bought and be profitable. It is also easy to verify that any equilibrium must also lie along the extended price-consumption line, the locus of *all* tangencies of the indifference curves with rays through the origin, plus the origin. Suppose not. Then there always exists either a small, positive supplementary policy at a "price" strictly higher than the candidate equilibrium price, which will be bought and make a profit (though it will result in the candidate equilibrium contract making a loss) or a small, negative<sup>16</sup> supplementary policy at a price lower than the candidate equilibrium price, which will be bought and make a profit (see Figure IV.2). Thus, if equilibrium exists, it must be at a point of intersection of the extended price-consumption line and the zero profit locus. There are only two such points – the origin and the point  $E$ . Since at the origin the indifference curve is steeper than the zero profit locus, there are profitable small policies which would upset the zero insurance point as an equilibrium. Thus, when multiple purchases are disallowed, the only candidate  $Q$ -equilibrium is the point  $E$ .

But  $E$  may not be an equilibrium. As drawn in Figure IV.2, the individual derives a higher level of utility from purchasing the contract  $B$ , which lies below the indifference curve through  $E$  but above the zero profit locus, than from purchasing  $E$  alone or  $B$  plus  $E$ . Hence, if all incumbent firms were offering the contract  $E$ , an entering firm offering contract  $B$  would attract all customers and make a profit. In this case, the contract  $E$  is not an equilibrium because it can be upset by a *large* quantity contract. And by the argument of the previous paragraph, the contract  $B$  cannot be an equilibrium either. Hence, in this case,

<sup>16</sup> It may be that all large policies which are preferred to the candidate equilibrium policy and which are profitable when purchased by themselves, would be combined with the candidate equilibrium policy and would as a result be rendered unprofitable. For this reason, small negative quantity policies are required to *necessarily* upset a point on the zero profit locus at which the price line is steeper than the indifference curve (i.e., where the individual would locally prefer to purchase less insurance at that price.)

no equilibrium exists with only quantity contracts.

It is immediate that, with multiple purchases disallowed, *if there exists a point  $G$  on the  $q^*$ -income-consumption line (the income-consumption line corresponding to price  $q^*$ ) which lies below  $E$  within the feasibility set, there does not exist a quantity equilibrium.  $G$  is strictly preferred to  $E$  or to  $G$  combined with any quantity contract at price  $q^*$ , including  $E$ .*

The nonexistence problem we have identified here is a property of the Nash equilibrium: When calculating whether a new policy is viable, an entering firm assumes that offering the new policy will not give rise to still further additions (or deletions) in the set of policies. The problem is analogous to that discussed by Rothschild-Stiglitz [1976] and Wilson [1977] in their analyses of insurance markets with adverse selection.

#### IV.4 *Multiple purchases allowed*

In the previous subsection, multiple purchases of the candidate equilibrium insurance contract were excluded by assumption. Allowing such purchases requires only a minor modification in the analysis. With such purchases, there may be equilibria on the boundary of the set for which individuals choose to purchase multiples. We term these *boundary quantity equilibria* or  $Q_B$ -equilibria. Any small supplementary policy attempting to break such an equilibrium will be unprofitable, because when that policy is offered, individuals purchase multiples of the equilibrium contract and lower their level of effort substantially, rendering the small, supplementary policy unprofitable. Thus, though effort is a continuous function of insurance, insurance purchased may not be a continuous function of the set of insurance policies offered. An analysis of boundary equilibria is contained in Appendix 1. In the text, we present some of the results in:

**Proposition 3:** There may be zero, one, or many  $Q_B$ -equilibria. Boundary quantity equilibria may be characterized by zero or positive profits. Two necessary conditions for a boundary point,  $C$ , to be such an equilibrium are that there be no point on the  $q^c$ -income-consumption line below  $C$  in  $\mathcal{F}$ , and that  $q^c$  be less than the price of insurance in the corresponding price equilibrium. Utility may be lower in a  $Q_B$ -equilibrium than in the corresponding price equilibrium.

By definition, an insurance contract for which an individual would like to purchase multiples cannot be a  $Q$ -equilibrium. Boundary equilibria entail contracts for which the individual is indifferent to purchasing multiples. That leaves contracts for which the individual strictly prefers a single contract to multiples. We term an equilibrium corresponding to this set of contracts an interior quantity equilibrium or a  $Q_I$ -equilibrium.

It is easy to extend the arguments of the previous subsection to prove

**Proposition 4:** The only possible  $Q_I$ -equilibrium is  $E$ , the point of intersection of the zero profit locus and the full insurance line. Necessary conditions for  $E$  to be a  $Q_I$ -equilibrium are that it be on the price-consumption line,<sup>17</sup> and that all the points on the  $\dot{q}^*$ -income-consumption line below  $E$  lie outside  $\mathcal{F}$ .

An important implication of Propositions 3 and 4 together is:

**Proposition 5:** A  $Q$ -equilibrium may not exist.

It is also worth noting that if indifference curves are convex, which will be the case if risk aversion is strong and the moral hazard relatively weak, the  $Q$ -equilibrium is unique, is at  $E$ , the point of intersection of the zero profit locus and the full insurance line, and coincides with the corresponding price equilibrium. Thus, restrictions by firms on the quantity of insurance sold to a client (as opposed to restrictions on his aggregate insurance purchases) are ineffective.

#### IV.5 *In retrospect*

It will be helpful to draw together the principal results obtained thus far.

With exclusive contracts, equilibrium always exists, entails rationing and zero profits and is (constrained) efficient conditional on the unobservability of effort.

A price equilibrium always exists and is unique. Equilibrium either (a) is at the point of

<sup>17</sup> We earlier established that the  $Q_I$ -equilibrium must be at an intersection of the extended price-consumption line and the zero profit locus, and cannot be at the origin. Our analysis showed that the unique point satisfying those conditions is  $E$ . If  $E$  is on the extended price-consumption line, but not on the price-consumption line, there is an insurance contract,  $W$ , at the same price, on the price-consumption line, which is preferred to  $E$  and which makes a profit. ( $W$  offers less insurance than  $E$ , since at  $E$  and beyond ZEL all indifference curves are well behaved.) Since indifference curves are convex beyond  $E$ ,  $E$  is preferred to  $W + nE$ ,  $n \geq 1$ . *A fortiori*,  $W$  is preferred to  $W + nE$ . Thus,  $W$  will not be supplemented by  $E$  or multiples of  $E$ , and therefore will upset  $E$  as an equilibrium.

intersection of the price-consumption line and the zero profit locus, (b) entails zero insurance, or (c) entails positive profits.

A quantity equilibrium takes one of two forms: An interior equilibrium may not exist; if it does, it coincides with the zero profit price equilibrium. There may be zero, one, or many boundary equilibria, which do not in general lie on the price-consumption line and may entail positive profits.

Thus, neither the price equilibrium nor quantity equilibria coincide with the exclusive contract equilibrium; the price equilibrium may or may not “be” (coincide with) a quantity equilibrium; and a quantity equilibrium may or may not be the price equilibrium.

We have not yet commented on the efficiency of the price and quantity equilibria, conditional on the unobservability of both effort and each individual’s aggregate insurance purchases. In another paper, we have investigated the efficiency of the price equilibrium (Arnott and Stiglitz [1990b]). If the government has direct control, it can obviously make a Pareto improvement by nationalizing insurance and offering only the exclusive contract. But even when the government has only indirect control, it can in most cases make a Pareto improvement through the linear taxation of insurance.

## V. Price-cum-(non-exclusive) Quantity Equilibria

In the previous section, each firm was allowed to sell only one unit of a single quantity policy to each client. In this section, we expand the set of admissible contracts to include both price policies and (non-exclusive) quantity policies. We allow a firm to sell only one type of policy in its contract, and in the case of quantity contracts to sell only one policy to each client. Again, we require that all policies be active in equilibrium. We term an equilibrium satisfying these requirements a price *cum* (non-exclusive) quantity equilibrium or *PQ*-equilibrium.

### V.1 *Not all P-equilibria are PQ-equilibria*

A small quantity policy may upset a price equilibrium. This is an important result. The restriction of the admissible set of contracts to only price contracts is not a natural one. As soon as it is recognized that firms can restrict the quantity of insurance they sell to

any customer, the relevant equilibrium concept is the  $PQ$ -equilibrium. And when quantity contracts are admitted, some of the price equilibria cannot be sustained.

To see when a  $P$ -equilibrium is not a  $PQ$  equilibrium consider first the zero profit price equilibrium, which we have denoted by  $E$ . Is  $E$  a  $PQ$ -equilibrium? It cannot be upset by other price contracts; higher-price contracts would not be bought, and lower-price contracts would operate at a loss. Can a quantity contract such as  $B$  in Figure V.1 upset  $E$  as an equilibrium? Perhaps. Perhaps not. It depends on whether the individual would choose to supplement  $B$  with insurance at the price  $q^*$  and, if so, by how much. Since the individual can purchase both  $B$  and any quantity of insurance at the price  $q^*$ , he faces the budget constraint  $RS$ . If faced with  $RS$ , the individual chooses  $B$ , then  $B$  upsets the equilibrium point  $E$ . If, however, the individual chooses  $C$ , for example, then since  $C$  makes a loss and the price of insurance for  $B$  is less than  $q^*$ , *a fortiori*  $B$  makes a loss. It follows that when both price contracts and quantity contracts are admitted, a zero profit price equilibrium can be upset by quantity contracts if and only if part of the  $q^*$ -income-consumption line below  $E$  lies in  $\mathcal{F}$ . An analogous result holds for the positive profit equilibrium (with price of insurance  $q^H$ ).

Consider now a zero insurance price equilibrium. At the origin, with separable, event-independent utility, the indifference curve is steeper than the zero profit locus. Hence, there is a small, profitable quantity contract that would be bought and upset the origin as an equilibrium. Thus, we have

**Proposition 6:** Zero profit and positive profit price equilibria are  $PQ$ -equilibria iff no point on the  $q^*$ - and  $q^H$ -income-consumption lines respectively, below  $E$  and  $H$  respectively, lies in the interior of  $\mathcal{F}$ . Zero insurance price equilibria are not  $PQ$ -equilibria.

## V.2 All $Q$ -equilibria are $PQ$ -equilibria

Since  $Q$ -equilibria must be able to survive any new quantity policy, they surely must be able to survive a price policy. Assume otherwise – that a price policy  $\hat{q}$  breaks the equilibrium. In breaking the equilibrium a particular quantity of that price policy would be purchased; the corresponding quantity policy would then have broken the equilibrium as

well.

### V.3 Other $PQ$ -equilibria

Are there other possible  $PQ$ -equilibria? If there are, any such equilibrium must entail individuals purchasing *both* a quantity policy  $(\hat{\alpha}, \hat{\beta})$  and a price policy  $\hat{q}$ . Denote the purchases of the price policy by  $(\tilde{\alpha}, \tilde{\beta})$ . If a  $PQ$ -equilibrium exists which is neither a  $P$ -equilibrium nor a  $Q$ -equilibrium, the quantity policy must have a strictly lower price than  $\hat{q}$ . Thus, if the quantity policy makes non-negative profits, the price policy must make strictly positive profits. This obviously provides an incentive for a firm to offer a lower price policy. For  $(\alpha^*, \beta^*) \equiv (\hat{\alpha} + \tilde{\alpha}, \hat{\beta} + \tilde{\beta})$  to be a  $PQ$ -equilibrium, it must therefore be a *boundary equilibrium* in the sense that reducing the price of the price policy causes the individual to purchase multiples of  $(\hat{\alpha}, \hat{\beta})$  and thereby renders the price policy unprofitable. Furthermore, there can be no small, utility-improving supplementary policy which is profitable.

Turn to Figure V.2. It displays the insurance possibility set as well as a variety of contracts.  $(\alpha^*, \beta^*)$  is the candidate  $PQ$ -equilibrium.  $(\alpha', \beta')$  is the individual's aggregate insurance purchase when he combines two quantity contracts with his preferred amount of the price insurance. As drawn, the individual is indifferent between  $(\alpha^*, \beta^*)$  and  $(\alpha', \beta')$  and strictly prefers either point to any other point in the insurance possibility set. Assume that the utility gradient is steeper in all directions at  $(\alpha', \beta')$  than at  $(\alpha^*, \beta^*)$ . Then if the price of insurance is lowered slightly from  $\hat{q}$ , the individual will switch from purchasing one unit of  $(\hat{\alpha}, \hat{\beta})$  to two. Assume that this renders the price insurance unprofitable. Furthermore if any small, utility-improving (price less than  $\hat{q}$ ) supplementary policy is offered, the individual will switch from purchasing one unit of  $(\hat{\alpha}, \hat{\beta})$  to two. Assume that this renders the small, supplementary policy unprofitable. Hence, there can be  $PQ$ -equilibria that are neither  $P$ -equilibria nor  $Q$ -equilibria. We term these mixed equilibria or  $M$ -equilibria since they entail individuals mixing price and quantity contracts. We shall omit a formal characterization of mixed equilibria. It should be evident, however, that there may be many such equilibria or there may be none.

We have thus proved

**Proposition 7:** All  $Q$ -equilibria are  $PQ$ -equilibria. The set of  $PQ$ -equilibria includes  $Q$ -

equilibria (when they exist), plus the positive profits price equilibrium (when it exists) provided the  $q^H$ -income-consumption line below  $H$  lies everywhere outside  $\mathcal{F}$ , plus mixed equilibria.

## VI. Income-Consumption Lines and the Feasibility Set

A central requirement for the existence of  $Q$ - and  $PQ$ -equilibria involved the relevant income-consumption line below the candidate equilibrium point never being inside the feasibility set. In this section, we investigate this requirement.

### VI.1 $Q_I$ -equilibrium

The  $Q_I$ -equilibrium, when it exists, is at the point we have denoted by  $E$ , the intersection of the zero profit locus and the full insurance line.

Since at  $E$ ,  $q = q^* \equiv \frac{p(0)}{1-p(0)}$  and effort is zero along the full insurance line, all points on the full insurance line are on the extended  $q^*$ -income-consumption line, the locus of all tangency points of lines with slope  $q^*$  to indifference curves. Since  $E$  is on the full insurance line and is, by assumption, on the  $q^*$ -income-consumption line, a necessary condition for part of the  $q^*$ -income-consumption line below  $E$  to lie in  $\mathcal{F}$  is that there be a discontinuity in the  $q^*$ -income-consumption line below  $E$ . Such a discontinuity is shown in Figure VI.1, where, as the price line with slope  $q^*$  moves to the right, the  $q^*$ -income-consumption line jumps from  $G'$  to  $G$ . Arnott and Stiglitz [1988a] show that the jump will occur in this direction if  $p^G(u'_1)^G > p^{G'}(u'_1)^{G'}$ .

Diagrammatically, we can see that the direction of the discontinuity depends on whether the indifference curves are closer together below  $G$  than below  $G'$ . If they are, a given rightward parallel shift in the budget constraint will increase utility more near  $G$  than near  $G'$ .<sup>18</sup>

### VI.2 $PQ$ -equilibrium corresponding to the positive profit price equilibrium

When it exists, the  $PQ$ -equilibrium corresponding to the positive profit price equilibrium occurs at the lower jump point at the price,  $q^H$ , at which: (i) the price-consumption line

<sup>18</sup> Formally, we need only calculate the derivative of utility along the (extended) income-consumption line near  $G$  and  $G'$ . If it is greater near  $G$  than near  $G'$ , then there is an inward discontinuity. The derivative of utility along the (extended) income-consumption line is just  $pu'_1$ .

jumps across the zero profit locus; and (ii) for all lower prices, the price-consumption line lies strictly outside the feasibility set. Points of discontinuity of the price-consumption line and the corresponding (i.e., corresponding to the price at which the discontinuity occurs) income-consumption line coincide. In Figure III.2  $H$  is the positive profits price equilibrium. For  $H$  to be a  $PQ$ -equilibrium, the corresponding income-consumption line below  $H$  must lie strictly outside the feasibility set; thus, both the price- and income-consumption lines must jump outwards, from  $H$  to  $H'$ . In the Figure, the price-consumption line jumps from  $H$  to  $H'$ . Arnott and Stiglitz [1988] show that the price-consumption line jumps outwards<sup>19</sup> if  $p^H(u'_1)^H \alpha^H < p^{H'}(u'_1)^{H'} \alpha^{H'}$  and in the opposite direction if the inequality is reversed. Since  $\alpha^{H'} > \alpha^H$ , the condition for an inward movement in the income-consumption line,  $p^H(u'_1)^H > p^{H'}(u'_1)^{H'}$ , and an outward movement in the price-consumption line are not inconsistent. Hence, a positive profits price equilibrium may or may not be a  $PQ$ -equilibrium. Figure VI.2 provides an example in which there is an outward discontinuity in the price-consumption line, but an inward discontinuity in the income-consumption line.

The intuition underlying this result is that for an outward discontinuity in the price-consumption line, the indifference curves near  $H'$  need not be quite as close as they would be for an outward discontinuity in the income-consumption line.

Note that the direction of the discontinuities in both lines depend on global rather than local properties, i.e. on a comparison of the values of accident probabilities and marginal utilities at two, quite distinct, points in  $\alpha - \beta$  space.

## VII. Inactive and Latent Policies

In the preceding sections, we required that all policies present in equilibrium be active, i.e. be purchased. This might seem a reasonable requirement, particularly when account is taken of the costs associated with drawing up policies. We also required that each insurance firm may offer only one policy. However, when both assumptions are relaxed, a new class of equilibria arises in which firms use inactive policies to supplement their active policies so as to deter entry; we have termed inactive policies employed for this purpose latent policies.

<sup>19</sup> The condition for an inward or outward discontinuity in the price-consumption line is derived in the same manner as that for the income-consumption line. We simply determine whether the derivative of utility along the extended price-consumption line near  $H$  is greater or less than that near  $H'$ .

### VII.1 *Inactive policies*

To start, we admit inactive policies but continue to require that each firm may offer only a single policy. We now show that the set of equilibria may be expanded when inactive policies are admitted. To simplify the argument, we assume that the market structure is such that there is only one incumbent firm in the market offering a quantity contract. We proceed by demonstrating an  $(\alpha, \beta)$  that can be an equilibrium when inactive policies are admitted, but not otherwise.

Turn to Figure VII.1.  $\hat{B} = (\hat{\alpha}, \hat{\beta})$  is any point having the properties that: (i) it is the lowest point on the income-consumption line corresponding to price  $\hat{q}$  in the interior of  $\mathcal{F}$ ; (ii) there is a jump discontinuity in the income-consumption line at  $\hat{B}$  (the upper jump point is labelled  $\hat{\hat{B}}$ ); (iii) insurance at price  $\hat{q}$  loses money at  $\hat{B}$  and at all points beyond  $\hat{B}$  on the  $\hat{q}$ -income-consumption line; and (iv)  $\hat{q} > \frac{\hat{\beta}}{\hat{\alpha}}$ .

Suppose that, as assumed, there is a single firm offering the quantity policy  $(\hat{\alpha}, \hat{\beta})$ , with the remaining incumbent firms offering a price policy at price  $\hat{q}$ . Since  $(\hat{\alpha}, \hat{\beta})$  is on the  $\hat{q}$ -income-consumption line and  $\hat{q} > \frac{\hat{\beta}}{\hat{\alpha}}$ , individuals will purchase the quantity policy but not the price policy – the price policy is inactive. We know from before that  $(\hat{\alpha}, \hat{\beta})$  cannot be a  $Q$ -equilibrium, since when only quantity contracts are admitted,  $(\hat{\alpha}, \hat{\beta})$  can be upset by a small, supplementary policy. We wish to show that  $(\hat{\alpha}, \hat{\beta})$  is, however, an equilibrium when the inactive  $\hat{q}$ -price policies are present. We need to demonstrate that  $\hat{B}$  can be upset by neither an entering quantity policy nor an entering price policy. If an entering quantity policy is bought, it must have price less than  $\hat{q}$ , say  $\bar{q}$ , and the individual's aggregate insurance purchases,  $\hat{\bar{B}}$ , must lie on the  $\bar{q}$ -income-consumption line beyond  $\hat{B}$  (if the individual purchases a positive amount of the price insurance).<sup>20</sup> Since  $\frac{p(\hat{\bar{B}})}{1-p(\hat{\bar{B}})} > \hat{q} > \bar{q}$ , the entering quantity policy loses money. If an entering price policy upsets  $\hat{B}$ , it must be bought in some quantity, in which case the corresponding quantity policy would also upset  $\hat{B}$ . But this cannot be. Hence, no entering price policy can upset  $\hat{B}$  either, and  $\hat{B}$  is therefore an equilibrium.

<sup>20</sup> A slight modification of the argument handles the case in which the individual would like to purchase a negative amount of price insurance, but is unable to do so because of the non-negativity requirement on purchases of price insurance.

## VII.2 *Latent policies*

In the previous subsection, we restricted firms to offering a single policy. It should be apparent from the above analysis that as soon as this restriction is dropped, firms may choose to use inactive policies to protect their active policies; for instance, in the example of the previous subsection, a firm could use the price policy  $\hat{q}$  to protect the quantity policy  $(\hat{\alpha}, \hat{\beta})$ . We term inactive policies used for this purpose latent policies, and equilibria in which they are employed latent policy or LPQ-equilibrium. A latent policy may be either a price or quantity policy, and firms may employ more than one latent policy. To simplify, we assume that each firm allows individuals to purchase the policies it offers separately, without restriction.

### VII.2.1 *Monopoly equilibrium*

We begin the analysis by considering a market structure in which a single insurance firm sells insurance to all individuals, but there are no barriers to entry—other than those posed by the firm's latent policies. Consider a policy  $B^*$  in  $\mathcal{F}$  (see Figure VII.2). Assume that the firm offers this policy, plus a latent quantity policy  $B^*L^*$  such that  $L^*$  lies on the indifference curve through  $B^*$ , but beyond  $\mathcal{F}$ . Since  $B^*L^*$  is latent,  $B^*$  is preferred to  $B^*L^*$ . *A necessary condition for  $B^*$  to be an LPQ-equilibrium – i.e., for  $B^*$  to be protected by  $B^*L^*$  – is that every small, supplementary policy, which when combined with  $B^*$  is profitable and utility-improving, is in fact combined with  $L^*$  and thereby rendered unprofitable.* Small, supplementary policies which when combined with  $B^*$  are both profitable and utility-improving have price  $\tau$  satisfying

$$\left[ \frac{pu'_1}{(1-p)u'_0} \right]_{B^*} > \tau > \left[ \frac{p}{1-p} \right]_{B^*} \quad (VII.1)$$

Thus,  $B^*L^*$  protects  $B^*$  against small, supplementary contracts if:

(i) The utility gradient at  $L^*$  is steeper than that at  $B^*$ , i.e.

$$(V_\alpha + \tau V_\beta)_{L^*} > (V_\alpha + \tau V_\beta)_{B^*}, \quad (VII.2)$$

for all  $\tau$  satisfying (VII.1). If (VII.2) is satisfied, when a small, supplementary policy satisfying (VII.1) is offered, the individual will purchase the small, supplementary policy plus the

original active policy plus the latent policy. A sufficient condition for (VII.2) to be satisfied is that

$$\frac{d(V_\alpha + \tau V_\beta)}{d\alpha} \Big|_{V_0} > 0 \quad (VII.3)$$

between  $B^*$  and  $L^*$  and for all  $\tau$  satisfying (VII.1). Equation (VII.3) may be rewritten as

$$\frac{-(p')^3}{p''(1-p)p} (u'_1 + \tau u'_0) > A_1 + \tau A_0 \quad (VII.3')$$

where  $A_i \equiv -\frac{u''_i}{u'_i}$ . Hence (VII.3) is more likely to hold the more rapidly effort falls as one moves up the indifference curve, and the lower the degree of absolute risk aversion. Furthermore, (VII.3) holds if the indifference curve,  $V_0$ , is weakly concave, and  $\frac{u'_1}{u'_0} > \frac{A_1}{A_0}$ , everywhere between  $B^*$  and  $L^*$ .

(ii) All small, supplementary policies satisfying (VII.1) make a loss at  $L^*$ , i.e.

$$(\tau <) \left[ \frac{pu'_1}{(1-p)u'_0} \right]_{B^*} < \left[ \frac{p}{1-p} \right]_{L^*}. \quad (VII.4)$$

This condition will be satisfied if, for example,  $(\left(\frac{d\beta}{d\alpha}\right) |_{V})_{B^*} < q^*$  and  $L^*$  lies beyond the zero effort line, so that  $\left[\frac{p}{1-p}\right]_{L^*} = q^*$ . Appendix 2 provides a simple constructive method for determining the set of  $B^*L^*$  which protect  $B^*$  against small, supplementary contracts for a particular class of utility functions.

Several comments are in order. First, even if  $B^*L^*$  protects  $B^*$  against small, supplementary policies, it may not protect  $B^*$  against large policies (or against price policies). Second, if  $B^*L^*$  protects  $B^*$  against only a subset of small, supplementary policies, it may be possible to protect  $B^*$  against the remaining small, supplementary policies by introducing additional latent policies along  $V_0$ . Third, if  $B^*$  is "strictly" protected by  $B^*L^*$  in the sense that no small policy "very nearly" upsets  $B^*$ , then all the points in the neighborhood of  $B^*$  are also protected against entry by small policies.

We now investigate conditions under which a policy may be protected from the entry of a large policy. New, large entering policies may upset  $B^*$  either as supplements to  $B^*$  or as replacements for it. To simplify the analysis, we assume that  $B^*$  is at the exclusive contract equilibrium, in which case there is no profitable policy which *replaces*  $B^*$ . This allows us to focus on large, supplementary policies.

We start by constructing the set of large policies that are profitable when combined with  $\theta$ . Turn to Figure VII.3. Draw a line through  $\theta$  with slope between  $(\frac{p}{1-p})_\theta$  and  $(\frac{d\beta}{d\alpha}|_{V_\theta})_\theta$ , such as  $WX$  in the diagram. Next draw the ray from the origin parallel to  $WX$  and mark its point of intersection with the ZPL by  $S$ . Now draw the constant effort locus through  $S$  and mark its point of intersection with  $WX$  by  $T$ . The claim is that  $\theta T$  is the largest policy at the price given by the slope of  $WX$ ,  $\bar{q}$ , that is profitable when combined with  $\theta$ . The proof is as follows: At  $S$ ,  $\bar{q} = (\frac{\beta}{\alpha})_S = (\frac{p}{1-p})_S$  since  $S$  is on the ZPL. Since  $T$  and  $S$  are on the same constant effort line,  $(\frac{p}{1-p})_S = (\frac{p}{1-p})_T$ . Hence,  $(\frac{p}{1-p})_T = \bar{q}$ , and therefore zero profits are made on the policy  $\theta T$ , positive profits on smaller policies at the same price, and negative profits on larger policies. Repeating the procedure for all  $q \in ((\frac{p}{1-p})_\theta, (\frac{d\beta}{d\alpha}|_{V_\theta})_\theta)$  generates the set of large policies<sup>21</sup> that are profitable when combined with  $\theta$ , which is the lozenge in the Figure.

Next slide the lozenge along the indifference curve until one finds a point say  $Z$  at which: (i) the utility gradient is steeper at every point in the moved lozenge than in the original lozenge; and (ii) every supplementary policy in the moved lozenge makes a loss. Then the latent policy  $\theta Z$  will protect  $\theta$  from entry by large, supplementary policies. If there is no single latent policy which satisfies these conditions, then more than one latent policy may protect  $\theta$ .

It is evidently difficult to specify conditions under which  $\theta$  is protected by latent policies, since these conditions depend on global rather than local properties of the utility and probability-of-accident functions. It is even more difficult to specify conditions under which points other than  $\theta$  are protected by latent policies, since one has to consider large replacement policies as well as large supplementary policies. However, the conditions for entry-detering latent policies do not seem very restrictive.

A further point is that if there is a latent policy which strongly protects  $\theta$  (in the sense that all large policies that are profitable when combined with  $\theta$  make a strictly positive loss with the latent policy present), then by continuity there is a dense set of points inside the feasibility set near  $\theta$  that can be protected by latent policies, including points at which

<sup>21</sup> The procedure is slightly more complicated if  $V_\theta$  is nonconvex near  $\theta$  since then there may be large profitable supplementary policies at a price greater than  $(\frac{d\beta}{d\alpha}|_{V_\theta})_\theta$ .

positive profits are made. Label the set of policies in the feasibility set that are protected by latent policies  $\mathcal{P}$ . Then the monopoly insurer will choose the most profitable policy in  $\mathcal{P}$ . Thus, *in insurance markets with effective latent policies, potential entry ensures neither economic efficiency nor zero profits*. Effective latent policies serve to discourage entry, even when the single insurance carrier is making strictly positive profits.

### VII.2.2 *Duopoly equilibrium*

With two firms, there is a much richer set of possible LPQ equilibria. Indeed, any point in the set  $\mathcal{P}$  is an LPQ equilibrium. And in general, the set  $\mathcal{P}$  is a dense set—in terms of Fig. VII.2, if  $B^*$  is strictly protected by  $B^*L^*$  in the sense that no policy very nearly upsets  $B^*$ , then all the points in the neighborhood of  $B^*$  are also protected. Hence, there will generically be a dense set of LPQ-equilibria.

Assume that both firms are offering the policy  $B^*$  and the associated latent policy  $B^*L^*$ , and that at  $B^*$  there are strictly positive profits. By the definition of  $B^*$ , no firm can offer a new policy and make a profit. Assume that the second firm tried to steal customers away from the first by offering a more attractive policy. Then the first firm's latent policy would come into effect, resulting in the second firm making a loss. On the other hand, if the second firm tried to increase its profits by offering a less attractive policy to its customers, they would all switch to the rival.

### VII.2.3 *The discrete activity case*

Turn to Figure VII.4. In this case the individual undertakes the safe activity for low levels of insurance and the risky activity for high levels —  $\phi^{r,s}$  separates the two regions and lies below the full insurance line. Now suppose a monopoly firm offers both the active policy  $B'$  which is in  $\mathcal{F}$  “just below”  $\phi^{r,s}$  on  $V_0$ , and the latent policy  $B'M'$  such that  $M'$  is outside  $\mathcal{F}$  “just beyond”  $\phi^{r,s}$  on  $V_0$ . Along  $V_0$  from  $B'$  to  $M'$ ,  $(V_\alpha + rV_\beta)V_0$  increases discontinuously across  $\phi^{r,s}$ . Hence, if  $B'$  and  $M'$  are sufficiently close, the utility gradient is unambiguously steeper at  $M'$  than at  $B'$ . Thus, any small, supplementary policy which would be profitable and utility-improving when combined with  $B'$ , would in fact be combined with  $M'$ . Furthermore, since profit also falls discontinuously across  $\phi^{r,s}$ ,  $B'$  and  $M'$  can be

chosen so that any such small, supplementary policy would be rendered unprofitable. Since the same argument applies to finite replacement policies near  $B'$ , it is possible to choose  $B'$  and  $M'$  so that any quantity policy which is utility-improving compared to  $B'$  and by itself is profitable (the cross-hatched area in Figure VII.4) is rendered unprofitable by the latent policy. By an earlier argument, if  $B'$  cannot be upset by a quantity policy, it cannot be upset by a price policy either. Similar results hold provided that there is a discrete number of activities.

We summarize this section in

**Proposition 8:** Latent policies can serve to deter entry. If there is a discrete number of activities, there is always a dense set  $\mathcal{P}$  of active policies that can be protected by latent policies. If there is a single firm in the market, there is a unique LPQ-equilibrium – the profit-maximizing policy in  $\mathcal{P}$  – in which the firm makes monopoly profit. If there is more than one firm in the market, every point in  $\mathcal{P}$  is an LPQ-equilibrium and in such equilibria most firms make positive profits. With continuous effort in the model, the set  $\mathcal{P}$  may be empty, but if it is not, the characteristics of LPQ-equilibria are the same as in the discrete activity case.

Latent policies can be thought of as a form of commitment by the incumbent firm to respond either to the entry of a new firm or to a deviation from equilibrium behavior by a rival. It is a form of response which, in this generalized contract space, gives rise to what may be viewed as a kinked demand curve.

It should be observed, however, that even if the firm could renege on its latent policy offers, it is not necessarily the case that it would want to. That is, some of the equilibria we have described are perfect equilibria. This is seen most clearly in the two activity case. We showed that any contract near the switchline (point  $B'$ ) is in the set  $\mathcal{P}$ , and there always exist latent contracts  $B'M'$  with slope steeper than  $p^*/(1 - p^*)$ , which serve to deter entry and which would make a profit should entry occur.

## VIII. Extensions and Comments

Two of the central results of our paper have been to show that, in the absence of latent policies there may be no Nash equilibrium, while with latent policies there may be a multiplicity of equilibria, some of which entail strictly positive profits.

Whenever one obtains a nonexistence result, one needs to question the behavioral and technological assumptions employed in the model as well as the equilibrium concept. Similarly, whenever one obtains a multiplicity of equilibria, one needs to examine carefully the equilibrium concept, to see whether there is a missing equilibrium condition, or a natural refinement of the equilibrium concept.

Much of this paper has been concerned with exploring these issues. In this section, we comment on the three essential ingredients of our analysis, the set of admissible contracts, the market structure, and the equilibrium concept.

#### VIII.1 *The set of admissible contracts*

Our analysis has shown that the existence and nature of equilibrium is dependent on the set of admissible contracts. We have proceeded by analyzing equilibrium under alternative assumptions concerning this set. We do not, however, take an agnostic view concerning which of these assumptions is the most relevant. We would argue that: (i) in most insurance markets, there are important dimensions of effort (accident avoidance activities) which cannot be directly monitored, and accordingly moral hazard (incentive) issues are important; (ii) in many insurance markets exclusive contracts are not feasible; and (iii) restricting the set to include only price policies is unjustifiably restrictive. That is why we have focused on quantity equilibria; at least in many instances, a firm can and does make inferences concerning what other insurance a client might buy from other firms on the basis of the quantity of insurance which he purchases from itself.

We are less sure concerning the relevance of latent policy equilibria, but to the extent that they are not important, a question is posed: What accounts for this?<sup>22,23</sup>

---

<sup>22</sup> One possible suggestion is that to the extent that latent policies reduce entry, they are providing a public good. If there are any costs associated with offering such policies, a representative firm may choose to free ride. Relatedly, if there is any cost to offering such policies, they cannot be part of a Nash equilibrium in pure strategies, since with such policies, there is no entry, and with no entry, it does not pay to offer such policies.

<sup>23</sup> In an earlier version of this paper, we examined the consequences of expanding the set of admissible

### VIII.2 *Market structure*

The inference that a firm can draw from an individual's purchase of its insurance policy depends, as in Rothschild-Stiglitz [1976], on the set of other insurance policies being offered. For example, if it were the only insurance firm in the market, it would not need to worry about individuals purchasing multiples of any quantity which it offered; it would, of course, still have to be concerned with the entry of new firms.

We have investigated the market structure which we believe to be the most relevant, where individuals typically buy all their insurance from one firm,<sup>24</sup> but in which they could purchase additional (multiple) insurance contracts from other firms. This, we suspect, produces the most efficient outcome, subject to the non-observability constraints. Other market structures yield different equilibria, as Hellwig's [1983b] analysis emphasized.

The market structure should, itself, be endogenous. One of the interesting and important results of Section VII.2 is that there are multiple equilibrium market structures. We characterized, for instance, both a monopoly equilibrium (in which latent policies protected the monopolist from the threat of entry) as well as a dense set of duopoly equilibria.

With many insurance firms, one needs to enquire into firms' incentives for communicating information concerning quantities of insurance purchased. Hellwig [1983a] showed that in the absence of communication costs, a full communication equilibrium is not possible. If it were, it would be the exclusive contract equilibrium. But in this candidate equilibrium each firm would profit by cheating.

### VIII.3 *Equilibrium concept*

contracts beyond that considered here, by the introduction of stochastic insurance policies and negative price insurance. In Arnott and Stiglitz [1988b] we noted that the non-convexities associated with moral hazard (and other principal-agent problems) may, under not too restrictive conditions, lead to stochastic insurance policies of a variety of forms. Again, one needs to ask why we do not observe such policies. What implications does this fact have for our theories? Here, we simply note that one form of randomisation of price insurance—where effort and the quantity of insurance are decided upon after the price of insurance is announced—sufficiently convexifies the problem that positive profit price equilibria are eliminated. This may not be true, however, of other forms of randomization.

Though individuals will never purchase negative insurance in the aggregate, the possibility of negative insurance policies may alter the set of equilibria, because it may expand both the set of potentially profitable entering policies and the set of deterrent policies.

<sup>24</sup> In addition, for most of the analysis, we have assumed that any insurance firm can offer only one (active) insurance policy. This restriction was made mainly to simplify the analysis; so long as latent policies are allowed, the structure of the equilibrium is not affected if firms can offer multiple active policies.

We have employed the standard Nash equilibrium concept with a one-period model in which all insurance firms make their offers simultaneously. Insurance firm A cannot react to firm B's offers by, for example, withdrawing a policy it had previously offered, or offering additional policies. These reactions were critical in the earlier analysis of insurance with self-selection. While Rothschild and Stiglitz [1976] discussed a variety of possible reaction equilibria, Wilson [1977] focused on one particular reaction equilibrium in which insurance firms were allowed to eliminate existing policies but not add new ones. Two criticisms were levelled at these reaction equilibria: They seemed outside the spirit of competitive markets; any small firm could induce large reactions by all other firms, and firms took this into account in their initial action. And the restrictions on the admissible (re)actions seemed unpersuasive.

Similar issues arise here. Clearly, a firm could react to the offer of a supplementary policy by another carrier (if this were observable) by withdrawing its original policy. While in the model with adverse selection this makes the new policy unprofitable, here this reaction increases the profitability of the new policy. Hence, this form of reaction makes it easier to break the equilibrium, making the set of equilibria smaller.

Latent policies can be interpreted as a mechanism for introducing reactions into a static model, since a latent policy becomes active only when other firms take actions attempting to break the initial equilibrium.

More recently, dynamic games, in which these reactions can be modeled as part of equilibrium strategies, have been formulated by Hellwig [1987] and by Stiglitz and Weiss [1987]. Similar dynamic games could presumably be formulated here. Though this eliminates some of the formal objections to these reaction function equilibria, the economic issues, e.g. whether these equilibria are consistent with a competitive market, remain.

The existence problems which we have encountered are associated with the nonconvexity of indifference curves (in benefit-premium space). While the assumption of convex indifference curves is natural in the context of choice *over goods*, with moral hazard the standard convexity assumptions do not rule out nonconvex indifference curves in benefit-premium space. Indeed, when there is a finite number of different effort levels which individuals

choose (or any linear combination thereof), indifference curves in benefit-premium space are *never* convex. The standard way to remedy nonconvex indifference curves is to assume a continuum of traders, which smooths out the nonconvexities by aggregation. That will not work here, however. Our analysis is completely unaffected by aggregation if all individuals are identical. Nor will assuming that individuals differ help. Either individual differences are observable, in which case contracts will be written contingent on these characteristics, and our analysis applies to each separate group with a particular set of characteristics; or else these differences are not observable, and problems of adverse selection arise, which alter the nature of the problem. Moreover, the existence problems discussed in this paper are distinct from those uncovered by Rothschild and Stiglitz [1976] and Wilson [1977] in the context of adverse selection. There, possible nonexistence arose when the quantity of insurance purchased by each individual was observable. Nonexistence in this paper, however, occurs only when clients' aggregate insurance purchases are unobservable. The mechanism causing nonexistence is that, for every feasible equilibrium configuration, there is always a profitable new contract that can be introduced, and is in this sense similar to that of Rothschild-Stiglitz (1976) and Wilson (1977).

#### VIII.4 *Other critical assumptions*

We should comment briefly on some of the structural assumptions of our model. We have abstracted from adverse selection and assumed a single period. Extending the model to incorporate these features will alter the structure of the problem. For example, in a multi-period model, the terms at which insurance is offered in later periods may be made contingent upon performance in earlier periods. While this mitigates the moral hazard, it does not, with discounting, eliminate it, and the possible nonexistence of equilibrium we have identified remains.

The simultaneous treatment of adverse selection and moral hazard, meanwhile, compounds the existence problems. We also assumed that consumers are perfectly informed concerning the set of available insurance policies. When consumers are imperfectly informed, the market may become monopolistically competitive and equilibrium may always exist.<sup>25</sup>

---

<sup>25</sup> In some versions of sequential search models (Diamond [1971]) equilibrium (contract or price) is identical

Finally, we assumed that firms are perfectly informed concerning profit opportunities and would enter, and perhaps upset, equilibrium, even when the profitable niche and the profits to be gained were arbitrarily small. Introducing noise into firms' information sets and/or costs of altering policies might also remedy nonexistence.

### Concluding Remarks

This paper belongs to the literature addressing the question of how to design contracts to mitigate the effects of moral hazard and the implications of such contracts for equilibrium and welfare. This literature includes the extensive principal-agent literature (in which the principal maximizes his expected utility subject to the incentive compatibility constraints and the individual rationality or reservation utility constraints of the agents).

In insurance markets, even though insurance firms cannot directly control the actions of the insured, they can affect his actions by, for instance, limiting the amount of insurance or threatening to terminate his insurance if he has too bad an accident record. Such *quantitative* restrictions play an important role in other markets, including capital markets, where lenders may try to limit the amount of funds borrowed by any borrower, and labor markets. But the effectiveness of these quantitative restrictions imposed by a firm on its customers depends on its ability to enforce them, which in turn depends, at least in part, on the market environment and on the observability of purchases.

This paper is concerned with showing how moral hazard considerations affect market *equilibrium*, in the central case in which a firm cannot observe the insurance that an individual purchases from *others*, but in which it can, quite naturally, observe the purchases an individual makes from itself. We assume that there are a large number of potential entrants, and that there are no costs of entry.

Market equilibrium, when it exists, looks distinctly different from that depicted by the standard Arrow-Debreu model. Equilibrium will be characterized by firms offering fixed quantity policies, rather than quoting prices and allowing customers to choose the quantity of insurance purchased. The intent of these quantitative restrictions is rather different from that analyzed in the earlier moral hazard literature; it is not so much to ration insurance to that of pure monopoly.

purchases as to discourage the purchase of policies from other insurance firms (in effect, to ration the total purchases of insurance.) In most cases, the quantitative restrictions take the form of requiring large purchases. Yet, though equilibrium always entails these quantitative constraints, in the absence of latent policies the equilibrium (when it exists) normally coincides with the zero profit price equilibrium (when that is the equilibrium price contract.)

The nonexistence problem which we have noted depends on the nonconvexity of the indifference curves, and arises whenever the discontinuity in income-consumption curves takes on a particular form, which, we have argued, it may quite plausibly have.

But while without latent policies equilibrium may plausibly not exist, with latent policies there are likely to be multiple equilibria, most of which entail strictly positive profits.

What light does our analysis cast on actual insurance markets and other markets in which moral hazard is important? Several stylized facts seem pertinent. First, many types of insurance (in particular fire insurance and health insurance, but not all accident insurance) in which moral hazard plays an important role are characterized by an exclusivity provision. Similarly, labor contracts in which moral hazard is important attempt to mitigate its effects by imposing restrictions on the quantity of insurance provided. Exclusivity provisions are absent only for insurance against those accidents for which moral hazard is unimportant, notably death (life insurance) and air flight accidents (air flight insurance). Second, such restrictions are only partly successful. While health insurance contracts may be able to restrict the insurance an individual can purchase from other insurance carriers, they cannot effectively restrict the insurance provided by employers (time off without loss of pay) or the implicit insurance provided by family and friends.<sup>26</sup> One context in which exclusivity provisions are particularly difficult to monitor is the credit market. Even though in the case of consumer credit, an attempt is made to monitor borrowing behavior and the extent of current liabilities, little attempt is made to restrict future access to credit; and since the typical loans are small, the price-equilibrium model may be the most relevant to this market. For other forms of credit, however, institutional arrangements in banking often serve

---

<sup>26</sup> In Arnott and Stiglitz [1990a] we analyze the conditions under which such insurance is welfare-enhancing.

to maintain exclusivity, while in bond markets seniority provisions may serve as at least a partial substitute.

Moral hazard is a pervasive phenomenon in our economy, arising not only in formal insurance markets, but also in labor, capital, and product markets. A persuasive analysis of competitive equilibrium should take it into account. In this paper, we investigated competitive equilibrium with moral hazard, and found that it differs in important respects from the conventional analysis. In particular, we have seen how the price mechanism of conventional market analysis no longer applies. We have analyzed these problems in the context of the simplest model in which moral hazard/incentive issues can arise. The problems that arise in that model have real-life counterparts; there is a rich set of institutional arrangements which have evolved in a variety of contexts to deal with them. An investigation of these arrangements will serve both to enrich our understanding of economics and of the relevance of alternative theoretical constructs.

## Appendix 1

### Boundary $Q$ -Equilibria

Analyzing such equilibria requires some new definitions. Define

$$S = \{(\alpha, \beta) | V(\alpha, \beta) > V(n\alpha, n\beta) \quad \forall \text{ integer } n > 1\}$$

to be the set of contracts such that the individual *strictly* prefers the contract to any integer multiple of that contract. Define  $\bar{S}$  to be the corresponding weakly-preferred set, and  $\tilde{S} \equiv \bar{S} - S$ . Note that (i)  $S$ ,  $\bar{S}$  and  $\tilde{S}$  need not be connected sets; (ii)  $(0, 0) \in \bar{S}$ ; and (iii) no point not in  $\bar{S}$  can be an equilibrium. We call  $Q$ -equilibria which lie in  $\tilde{S}$  "boundary equilibria" or  $Q_B$ -equilibria. Similarly, we call equilibria which lie in  $S$  "interior equilibria" or  $Q_I$ -equilibria.

Figure A.1 illustrates a policy  $M$  for which the individual is indifferent between buying one ( $M$ ) and two ( $2M$ ) policies. Note that if the indifference curves are not bunched together much more loosely at  $2M$  than they are at  $M$ , and if a policy  $M'$ , with slightly more insurance than the policy  $M$ , were to be offered, the individual would prefer  $2M'$  to  $M'$ . That is why  $M$  is on the boundary of the set  $\bar{S}$ .

It is *possible* that there are equilibria on the boundary of the set at which individuals purchase multiple insurance policies. Turn to Figure A.2.  $\tilde{S}$  is the set of boundary points of the connected subsets of  $\bar{S}$ . The only points in  $\tilde{S}$  which cannot necessarily be upset by either a small profitable supplementary contract or a profitable replacement contract are the points of maximal utility in each subset of  $\bar{S} \cap \mathcal{F}$ , such as  $D$  and  $D'$  in Figure A.2.\* Thus, one necessary condition for a point  $C$  to be a boundary  $Q$ -equilibrium is that it be a point of maximal utility in a connected subset of  $\bar{S} \cap \mathcal{F}$ .

Another necessary condition is that there be no point in the interior of  $\mathcal{F}$  on the  $q^c$ -income-consumption line below  $C$ . Suppose there were such a point,  $R$ . Then if contract  $R$  were offered, it would be bought since it provides a higher level of utility than  $C$ . Furthermore, since  $R$  is on the  $q^c$ -income-consumption line, the individual would not supplement it with  $C$  or a multiple of  $C$ , and  $R$  would be profitable. Hence,  $R$  would upset  $C$ .

\* If the indifference curve is tangent to the boundary, it implies that, in Figure A.1, the slope of the indifference curve at  $2M$  is identical to that at  $M$ . ( $\max_{\alpha, \beta} V(\alpha, \beta)$  s. t.  $V(\alpha, \beta) = V(2\alpha, 2\beta)$ ).

A third necessary condition is that the implicit price in any  $Q_B$ -equilibrium must be no higher than in the corresponding  $P$ -equilibrium. Assume the contrary, and let  $z$  denote the  $Q_B$ -equilibrium contract and  $y$  aggregate purchases in the  $P$ -equilibrium. Suppose the economy is at the  $Q_B$ -equilibrium with only quantity contracts admitted, and that an entering firm offers  $y$  as a quantity contract. The individual would prefer  $y$  to  $z$  or to  $y + nz$ ,  $n \geq 1$ , because  $y$  is the most preferred point on a lower price line. Since  $y$  would also break even or make a profit, it would upset  $z$ .

While the price associated with a  $Q_B$ -equilibrium can be no higher than that in the  $P$ -equilibrium, welfare may be lower. A  $Q_B$ -equilibrium may not be broken by the  $P$ -equilibrium contract (the  $(\alpha, \beta)$  purchased in the  $P$ -equilibrium) even though the latter is associated with a higher level of welfare, since the individual may choose to supplement the  $P$ -equilibrium contract with multiple purchases of the  $Q_B$ -equilibrium contract resulting in the  $P$ -equilibrium contract making a loss. Thus, while exclusive contracts which restrict the *aggregate* quantities of insurance purchased are welfare-enhancing, quantity restrictions by firms on only their own sales of insurance need not be.

We summarize the results of the Appendix in Proposition 3 in the text.

## Appendix 2

### Construction of latent policies against small, supplementary contracts

We assume that the utility function has the form  $-\frac{u''}{(u')^2} = \text{a constant}$ . Then (VII.3') reduces to the condition that the indifference curve be concave, and (VII.2) to the condition that the line tangent\* to  $B^*$  lie below  $L^*$ . Turn to Figure A.3. The set of points on  $V_0$  satisfying (VII.2) and (VII.4) is  $XYZ$ . To demonstrate this: First, draw the ray from the origin parallel to the tangency line at  $B^*$  and label its intersection point with the ZPL,  $W$ . By construction,  $(\frac{d\beta}{d\alpha}|_{V_0})_{B^*} = (\frac{pu'}{(1-p)u'_0})_{B^*} = (\frac{\beta}{\alpha})_W = (\frac{p}{1-p})_W$ . Next, draw the constant effort line through  $W$  and extend it back to  $V_0$ , cutting  $V_0$  at  $T$ . By construction  $(\frac{p}{1-p})_T = (\frac{p}{1-p})_W$ . Since effort falls as one moves out along an indifference curve, for any point beyond  $T$  on  $V_0$ , (VII.4) is satisfied. Thus, both (VII.2) and (VII.4) are satisfied for all points on  $V_0$  between  $X$  and  $Z$ .

---

\*  $V_\alpha + \tau V_\beta = (V_\alpha + \tau V_\beta)_{B^*}$  along the line tangent to the indifference curve at  $B^*$ .

## Bibliography

- Arnott, R.J. and J.E. Stiglitz, "Equilibrium in Competitive Insurance Markets with Moral Hazard", discussion paper #4, John M. Olin Program for the Study of Economic Organizations and Public Policy, Princeton University, 1987.
- Arnott, R.J. and J.E. Stiglitz, "The Basic Analytics of Moral Hazard," *Scandinavian Journal of Economics* 90, 1988a, 383-413.
- Arnott, R.J. and J.E. Stiglitz, "Randomization with Asymmetric Information," *Rand Journal of Economics* 19, 1988b, 344-362.
- Arnott, R.J. and J.E. Stiglitz, "Moral Hazard and Non-market Institutions: Dysfunctional Crowding Out or Peer Monitoring?", forthcoming, *American Economic Review*, 1990a.
- Arnott, R.J. and J.E. Stiglitz, "Price Equilibrium, Efficiency and Decentralizability in Insurance Markets," mimeo, 1990b.
- Arnott, R.J. and J.E. Stiglitz, "The Welfare Economics of Moral Hazard," in H. Loubergé, ed., *Risk, Information, and Insurance: Essays in the Memory of Karl H. Borch* (Norwell, MA: Kluwer Academic Publishers, 1990c).
- Diamond, P., "A Model of Price Adjustment," *Journal of Economic Theory* 3, 1971, 156-168.
- Eaton, Jonathan and Gersovitz, Mark, "Debt with Potential Repudiation: Theoretical and Empirical Analysis," *Review of Economic Studies*, 42 (2), April 1981, pp. 289-309.
- Hellwig, M.F., "Moral Hazard and Monopolistically Competitive Insurance Markets," (1983a), University of Bonn, discussion paper No. 105.
- Hellwig, M.F., "On Moral Hazard and Non-Price Equilibrium in Competitive Insurance Markets," (1983b), University of Bonn, discussion paper No. 109.
- Hellwig, Martin, "Some Recent Developments in the Theory of Competition in Markets with Adverse Selection," *European Economic Review*, February/March 1987, Vol. 31, pp. 319-325.
- Helpman, E., and J.J. Laffont, "On Moral Hazard in General Equilibrium," *Journal of Economic Theory* 10, 1975, 8-23.

- Pauly, M., "Overprovision and Public Provision of Insurance," *Quarterly Journal of Economics* 88, 1974, 44-62.
- Rothschild, M. and J. Stiglitz, "Equilibrium in Competitive Insurance Markets: An Essay on the Economics of Imperfect Information," *Quarterly Journal of Economics* 90, 1976, 629-649.
- Stiglitz, J.E., "Risk, Incentives and Insurance: The Pure Theory of Moral Hazard," *The Geneva Papers on Risk and Insurance* 8, 1983, 4-33.
- Stiglitz, J.E. and A. Weiss, "Macroeconomic Equilibrium with Credit Rationing," 1987, mimeo.
- Wilson, C., "A Model of Insurance Markets with Incomplete Information," *Journal of Economic Theory* 15, 1977, 167-207.

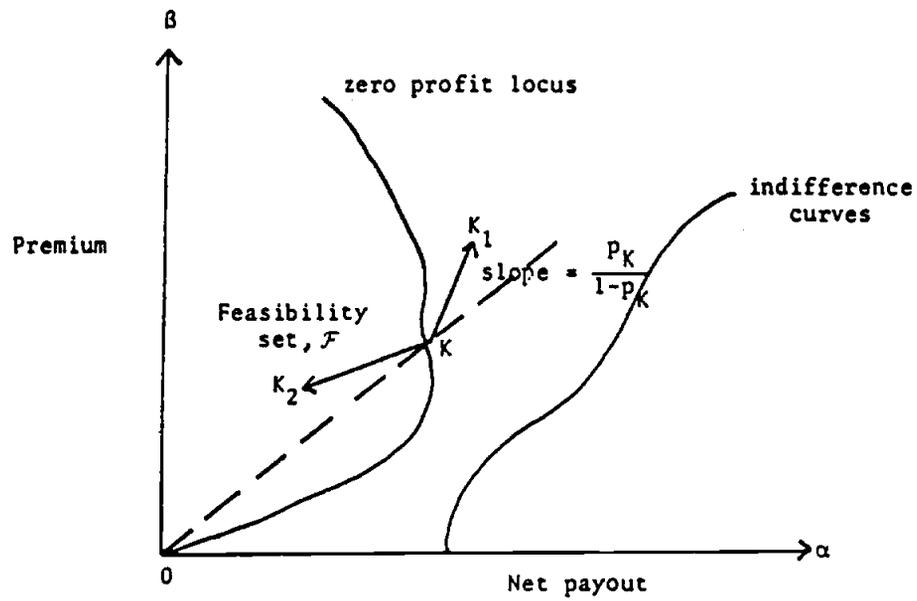


Figure II.1: Basic diagram 1, continuum of effort levels  
 i) the zero profit locus is  $\beta(1-p) - \alpha p = 0$   
 ii) the feasibility set is never convex  
 iii) indifference curves are not necessarily convex

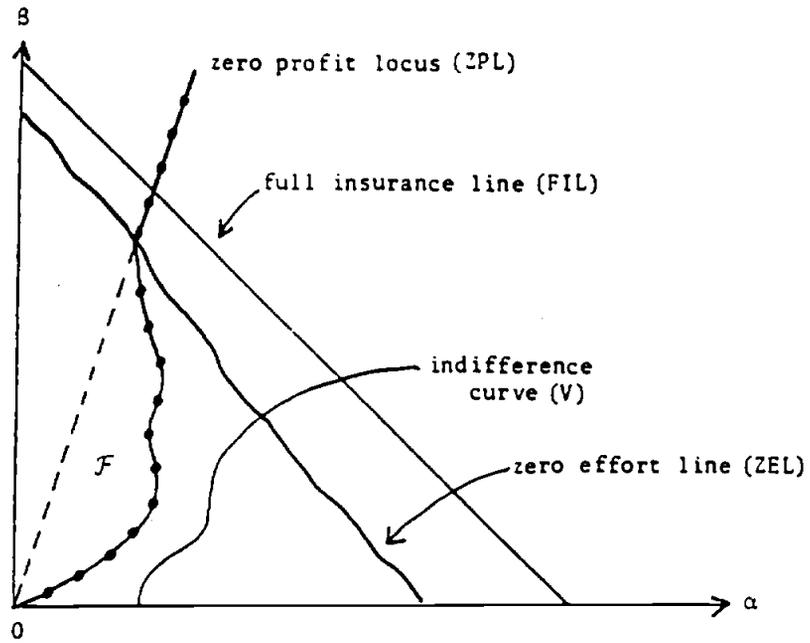


Figure II.2: Basic diagram 2, continuum of effort levels

- i) effort is zero beyond the zero effort line
- ii) beyond the zero effort line, the zero profit locus is  $\frac{\beta}{\alpha} = \frac{p(0)}{1-p(0)}$ , a ray going into the origin with slope  $p(0)/(1-p(0))$
- iii) beyond the zero effort line, indifference curves are strictly convex

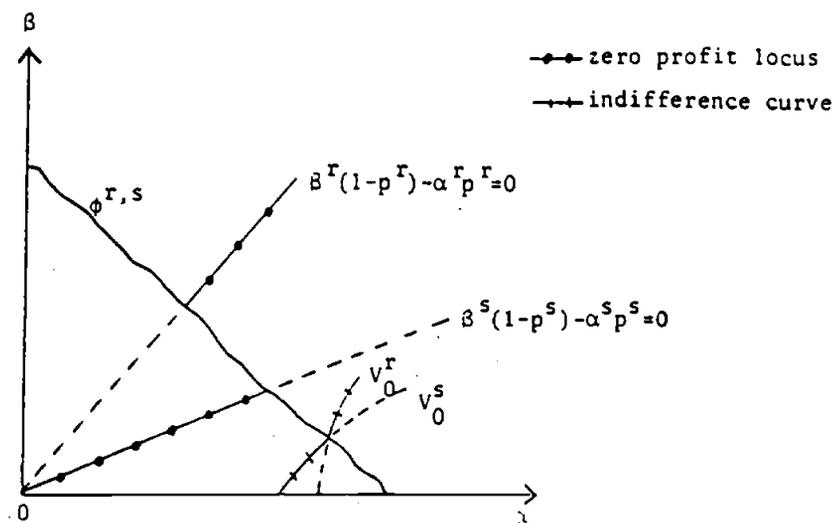


Figure II.3: Basic diagram 3, two effort levels

Observe: The safe activity is undertaken for  $(\alpha, \beta)$  below  $\phi^{r,s}$ , the risky activity above

- :  $V_0^s$  is the set of  $(\alpha, \beta)$  which generate utility  $V_0$ , contingent on the safe activity, etc.
- :  $\beta^s(1-p^s) - \alpha^s p^s = 0$  is the zero profit locus, contingent on the safe activity

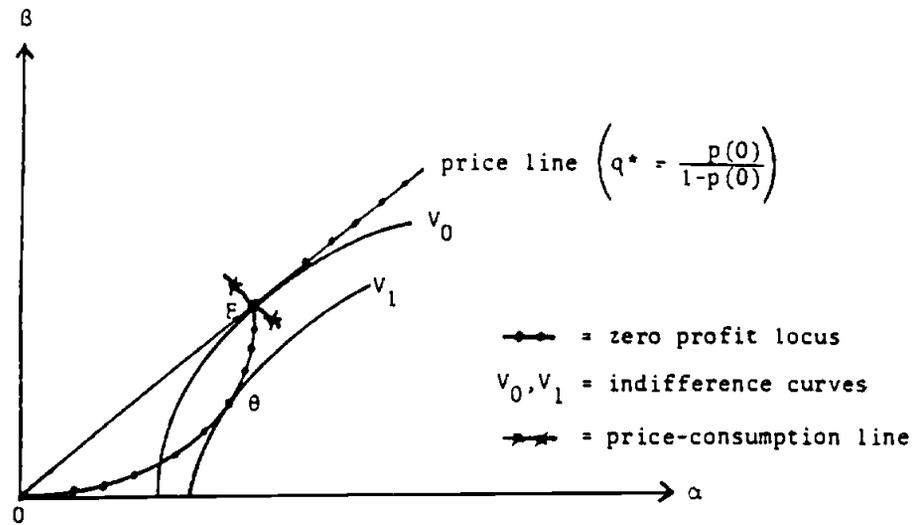


Figure III.1: Even where the price-consumption line is continuous and intersects the zero profit locus, the price "equilibrium" is dominated by the quantity contract  $\theta$  - the exclusive contract equilibrium.

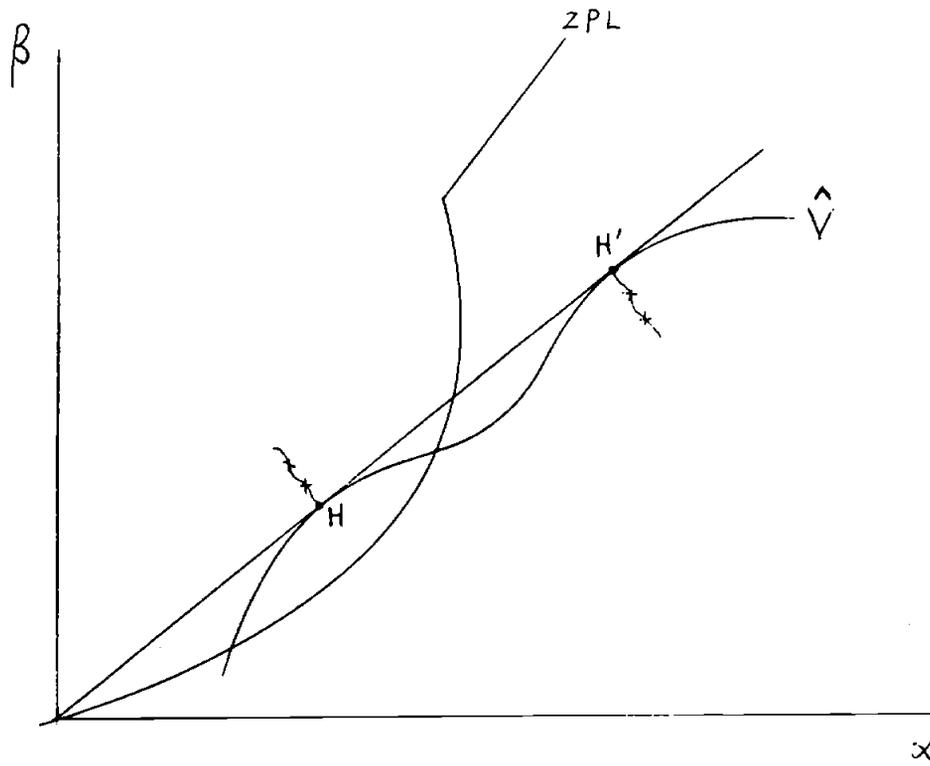


Figure III.2: A positive profit price equilibrium.

\*\*\* price-consumption line

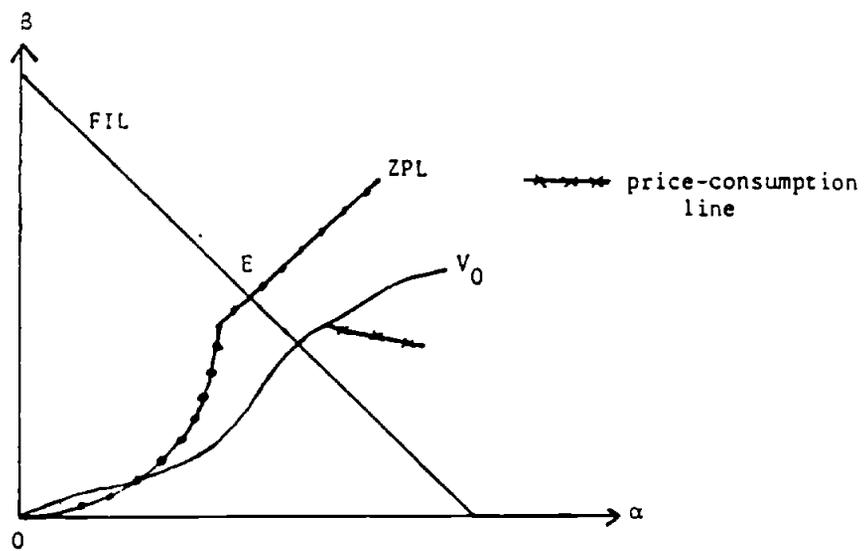


Figure III.3: Zero insurance price equilibrium. Observe that the individual's utility is higher at 0 than at E

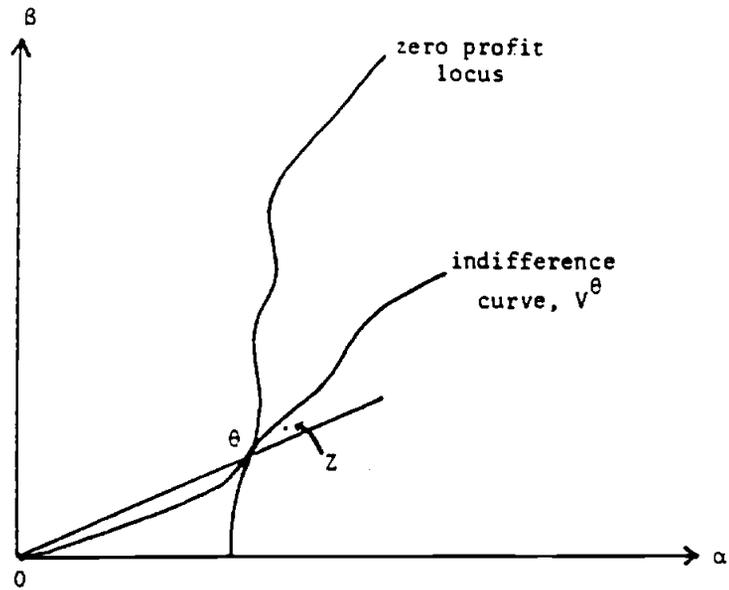


Figure IV.1: The exclusive contract equilibrium cannot be sustained when exclusive contracts are unenforceable

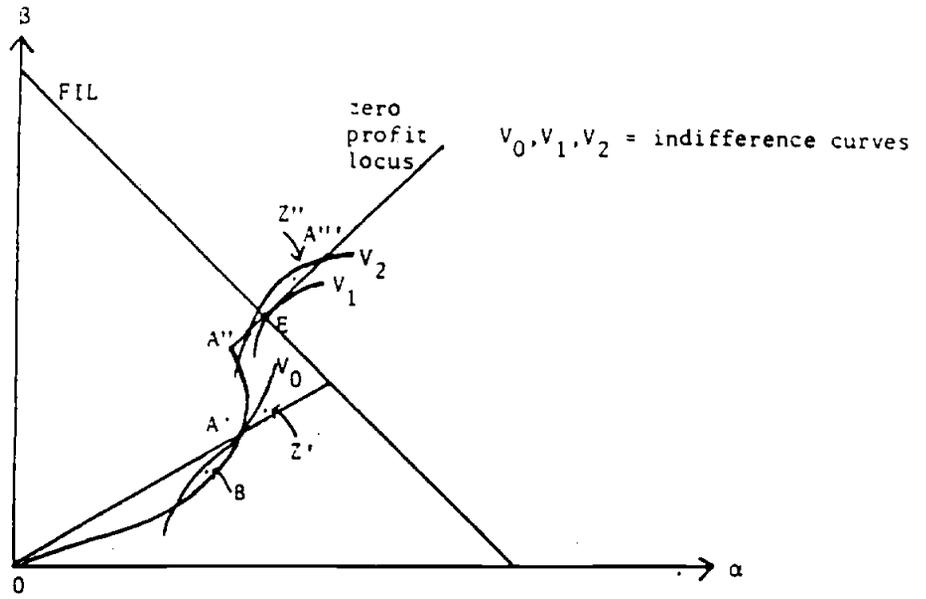


Figure IV.2: No point on the zero profit locus for which the indifference curve through the point is not tangent to the ray from the origin through the point can be an equilibrium when only quantity contracts are admissible and exclusive contracts are unenforceable. The individual would choose to supplement  $OA'$  with the profitable contract  $A'Z'$ , and  $OA''$  and  $OA'''$  with the profitable contracts,  $A''Z''$  and  $A'''Z'''$  respectively.

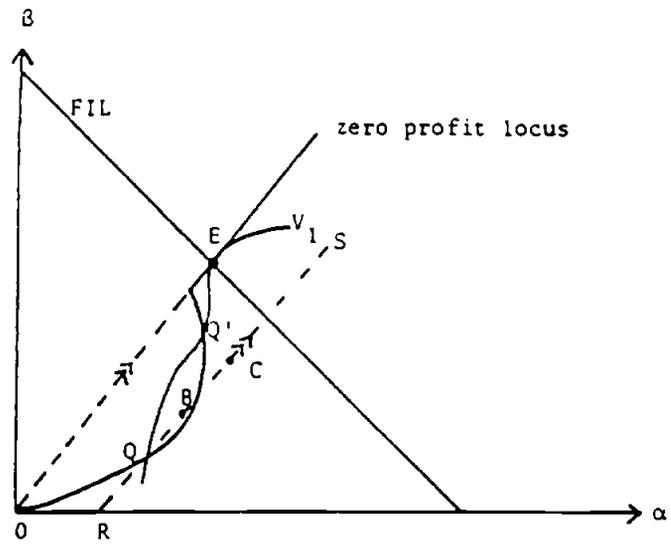


Figure V.1: A zero profit price equilibrium need not be a PQ-equilibrium.

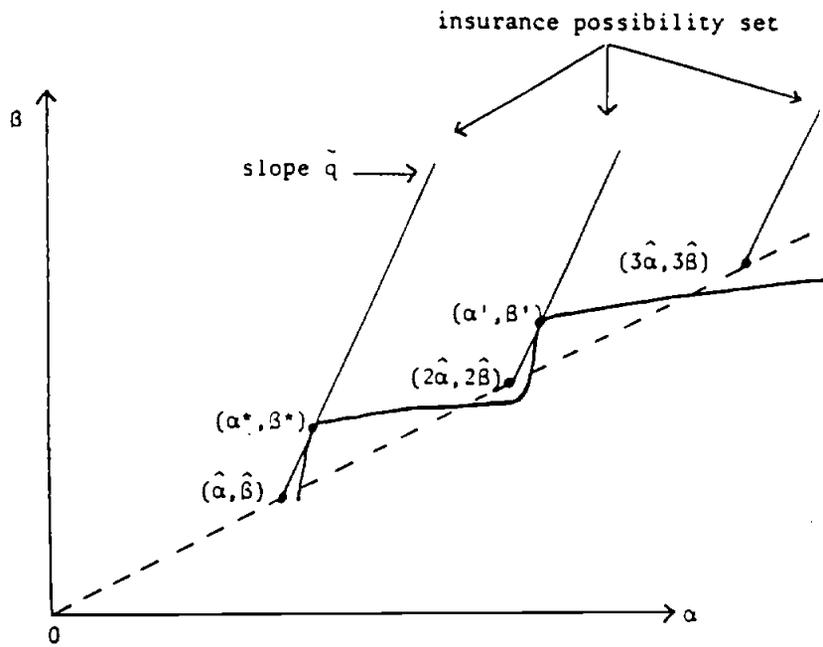


Figure V.2:  $(\alpha^*, \beta^*)$  can be a mixed equilibrium.

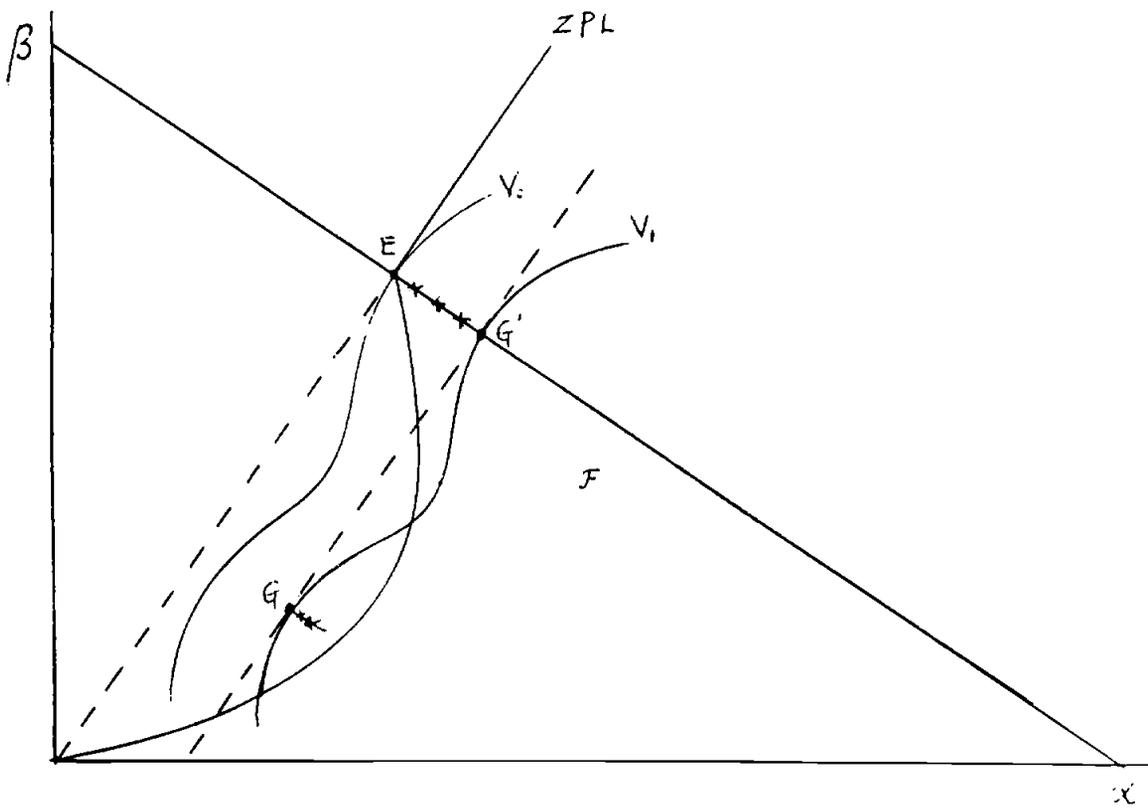


Figure VI.1: Part of the  $q^*$ -income-consumption line lies below  $E$  in  $\mathcal{F}$ .

-x-x-  $q^*$ -income-consumption line

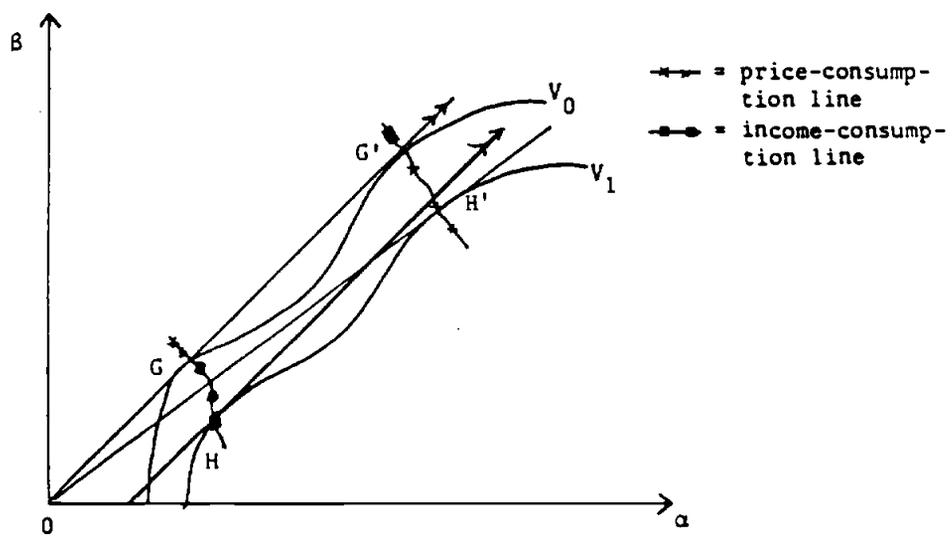


Figure VI.2: The price-consumption line can jump outwards when the income-consumption line jumps inwards.

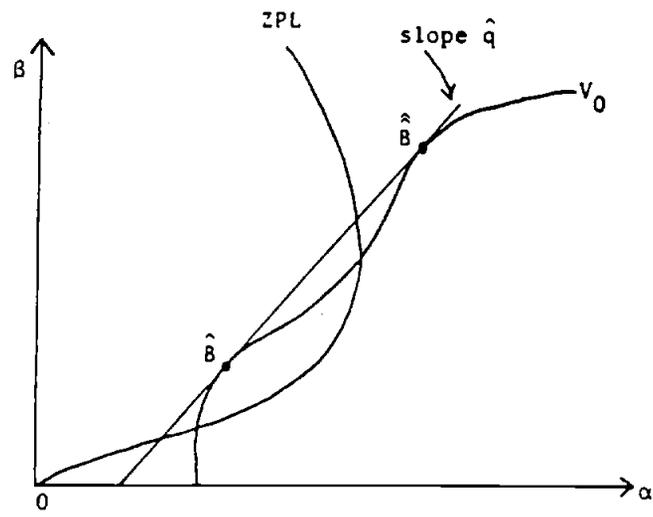


Figure VII.1:  $\hat{B}$  is an equilibrium with inactive policies admitted.

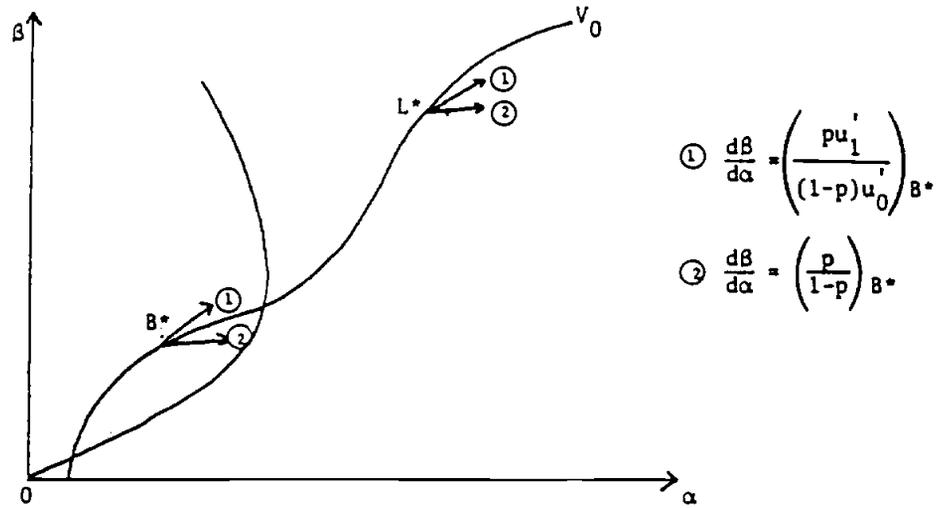


Figure VII.2: Investigation of latent policies. For  $B^*L^*$  to protect  $B^*$  against small, supplementary policies, the utility gradient between the directions characterized by arrows ① and ② must be steeper at  $L^*$  than at  $B^*$ .

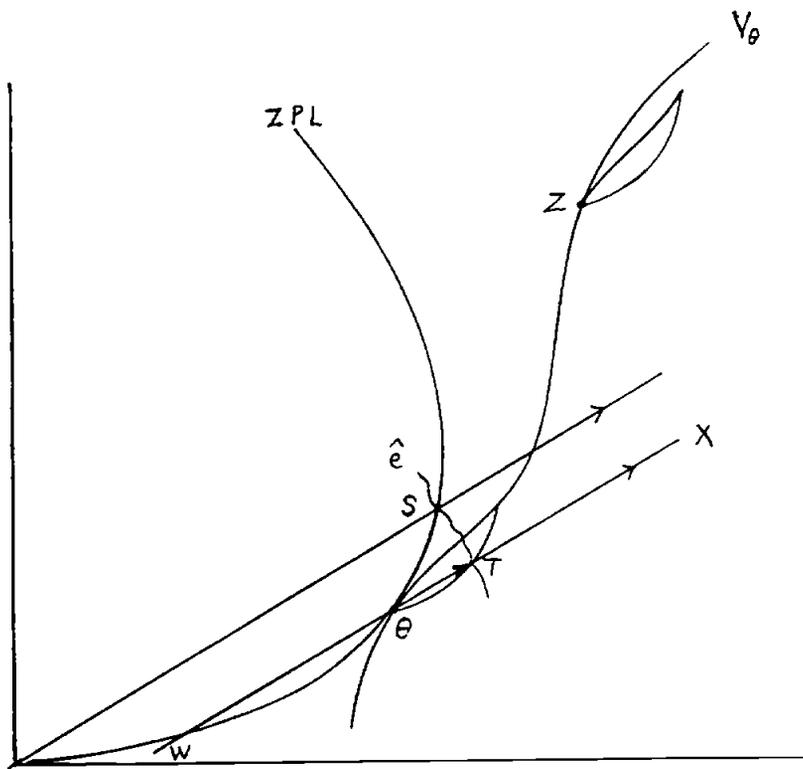


Figure VII.3: Construction of the set of large, profitable supplementary policies.



set of large, profitable supplementary policies

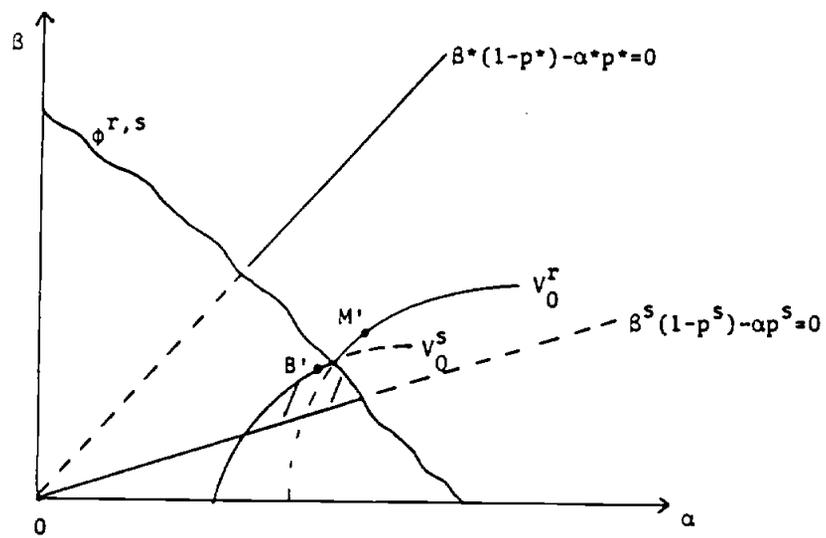


Figure VII.4: Latent policies in the two-activity case.

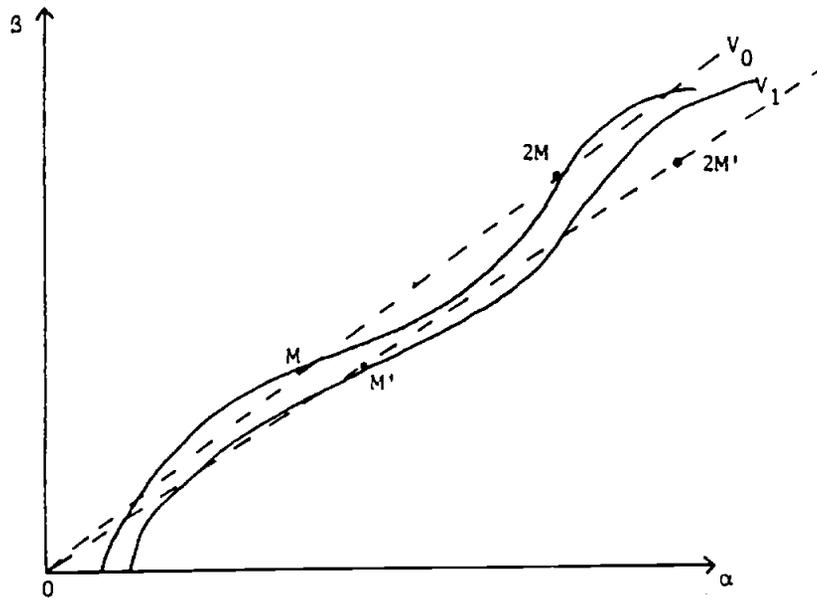


Figure A.1:  $M$  is a representative point in the set  $\mathcal{A}$ .

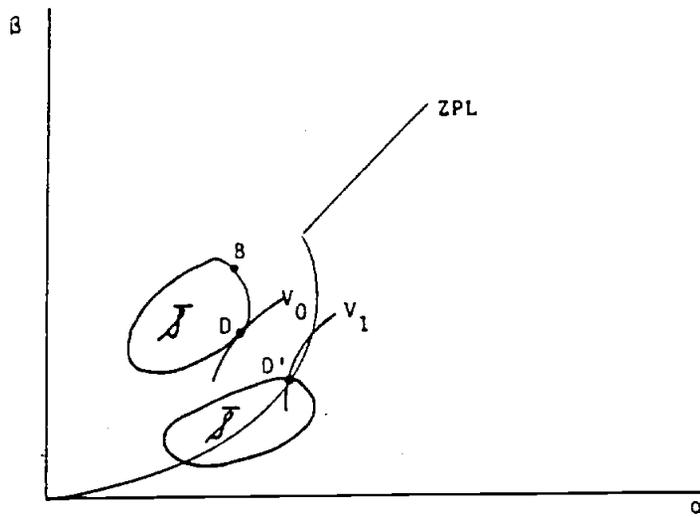


Figure A.2:  $D$ , a point of maximal utility on a connected subset of  $\mathcal{J} \cap \mathcal{Z}$ , may be a  $Q_B$ -equilibrium even though it is on neither the ZPL nor the price-consumption line.  $D'$  may also be a  $Q_B$ -equilibrium.

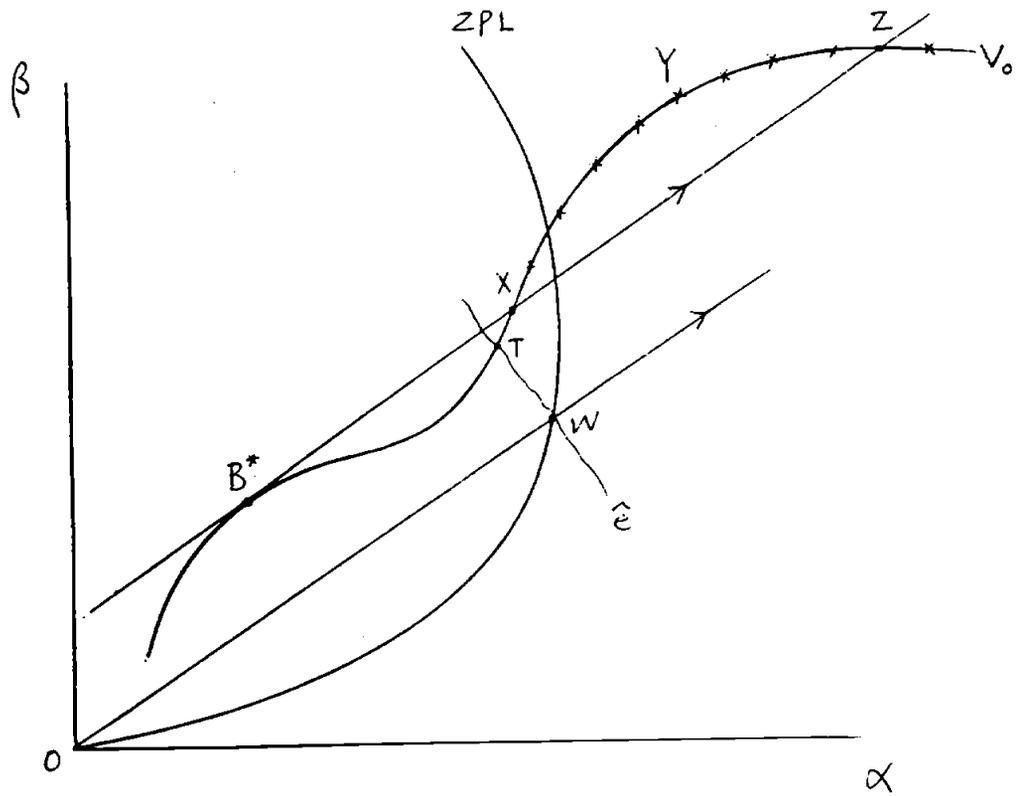


Figure A.3: Construction of the set of latent policies which protect  $B^*$  against small, supplementary policies.

—xxx— set of latent policies for  $B^*$