A Model of Income Insurance and Social Norms

Assar Lindbeck and Mats Persson

A Model of Income Insurance and Social Norms[*]

by

Assar Lindbeck[♣] and Mats Persson[♠]

Abstract:

A large literature on *ex ante* moral hazard in income insurance emphasizes that the individual can affect the probability of an income loss by choice of lifestyle and hence, the degree of risk-taking. The much smaller literature on moral hazard *ex post* mainly analyzes how a "moral hazard constraint" can make the individual abstain from fraud ("mimicking"). The present paper instead presents a model of moral hazard ex post *without* a moral hazard constraint; the individual's ability and willingness to work is represented by a continuous stochastic variable in the utility function, and the extent of moral hazard depends on the generosity of the insurance system. Our model is also well suited for analyzing social norms concerning work and benefit dependency.

Key words: Moral hazard, sick pay insurance, labor supply, asymmetric information

JEL classification: G22, H53, I38, J21

A Model of Income Insurance and Social Norms

1. Introduction

Several European countries have recently experienced high and rising work absence, often financed by sick pay ("temporary disability") insurance or early retirement ("permanent disability") insurance. This development raises the issue of the determinants of work absence. The earliest contributions to the literature on this issue were made in the fields of social psychology, industrial relations, and social medicine. Absence was then regarded as a way for employees of compensating for, or recovering from, arduous working conditions. A related approach is found in the labor supply literature. In a seminal paper, Allen (1981) makes the assumption that the contracted number of hours of work is larger than the number of hours actually desired by the individual. Absence is then a way for the individual of adjusting to his desired hours of work; in this framework, absence would not be a problem if each individual were able to decide the desired number of hours of work in the labor contract.

Insurance aspects are not dealt with in that literature. It is, however, not possible to understand the determinants of absenteeism in today's developed countries without considering moral hazard in income insurance. The distinction between moral hazard *ex ante* and *ex post* then becomes important.[1] Moral hazard *ex ante* implies that because of the insurance, the individual changes his behavior in such a way that the probability of an insured event increases. In the case of sick pay insurance, this would be reflected in a less healthy lifestyle (see e.g. Arnott, 1992). We do not think, however, that moral hazard *ex ante* is quantitatively important in the case of income insurance. It is not plausible that individuals choose to smoke, or avoid exercise, as a result of having income insurance. Since it is quite unpleasant to be sick, the (uninsurable) pain, so to speak, functions as a kind of coinsurance, discouraging *ex ante* moral hazard.

By contrast, moral hazard *ex post* is highly relevant in the context of income insurance, since an individual can increase his leisure without much loss of income by

---

[1] For an alternative classification of types of moral hazard, see Sinn (1983, p 315-326).

exploiting the insurance system. In the seminal model of Diamond and Mirrlees (1978) on moral hazard, such exploitation takes place when a healthy individual pretends to be sick (mimicking). Diamond and Mirrlees and their followers then assume the individual's health status to be a well-defined dichotomous variable; either you are sick or not. The government is assumed to impose a so-called "moral hazard constraint" on the design of the optimal insurance system, implying that a healthy person should not find such mimicking worthwhile. According to this approach, full insurance will not arise, since the moral hazard constraint would then be violated.

Whinston (1983) has extended the analysis to the case of many groups of individuals who differ with respect to their probability of being unable to work. By imposing moral hazard constraints on all groups, the optimal system guarantees that all individuals will find it advantageous to work when being able to do so, and hence nobody will cheat. Although such a model is more realistic than the original Diamond and Mirrlees (1978) setting, it is still simplistic in the sense of the individual's ability to work being dichotomous, and in the sense of identifying moral hazard with fraud.

We believe that a realistic model should deal with sickness absence in a more complex way. Indeed, Cochrane (1972) has argued that health, from a medical point of view, is usually a continuous variable that cannot be approximated to just two alternative states ("sick" and "not sick"). Along this line, Barmby et al. (1994) assume that, besides consumption and leisure, the individual's utility function includes a continuous index variable reflecting the individual's health status.[2] Neither Cochrane (1972) nor Barmby et al. (1994) deal with insurance issues, however (the latter paper analyzes optimal wage setting in an efficiency wage context, when shirking takes the form of absence).

The purpose of the present paper is to give a more realistic, and yet analytically tractable, picture of sickness absence than in the existing insurance literature. In particular, we acknowledge that the individual's ability and willingness to work is a matter of degree, and that the individual's choice depends on the pain or pleasure

---

[2] For a survey of some representative studies in this tradition, see Brown and Sessions (1996).

derived from working, which we represent by a continuous, stochastic variable in the individual's utility function.

2. The Basic Model

It is useful to take a modified version of the traditional labor supply model as a point of departure, according to which the individual maximizes a utility function in consumption (c), working time ($1-\ell$) and leisure time (ℓ). For simplicity, we assume the utility function to be additively separable in these three arguments:

$$U(c, 1-\ell, \ell) = u(c) + f(1-\ell) + g(\ell),$$

where $u'(c) > 0, u''(c) < 0$. While the evaluation of work is usually assumed to be negative in the standard labor supply literature, we use a general enough formulation to also encompass the cases where work is regarded as pleasant.[3]

We simplify matters by assuming that the *f* and *g* functions are linear and that the evaluation of work and leisure can be represented by the stochastic preference variables a and b, respectively:

$$u(c, \ell) = u(c) + a \cdot (1-\ell) + b \cdot \ell.$$

While a positive realization of a means that the individual enjoys going to work during a specific period, a negative realization implies that it is unpleasant to do so (perhaps due to health problems, or just because work is boring). A similar inter-pretation of positive and negative values of b holds for the evaluation of leisure; a particularly high value of b may occur on a day when the weather is conducive for leisure activities, or when interesting sports events are shown on television. Naturally,

---

[3] The first-order condition for utility maximization is then

$$\frac{g'(\ell) - f'(1-\ell)}{u'(w\ell)} = w,$$

i.e., the wage must compensate for lost leisure adjusted for the value (or disutility) attached to increased work. It is obvious that this first-order condition can be satisfied even with a positive marginal utility of work, provided that the marginal utility of leisure is even higher. This formulation of the work/leisure choice is, for instance, used in Layard and Walters (1978, pp 308-310).

a and b may be correlated in different ways. For instance, a negative realization of both a and b could represent a spell of sickness making the individual feel discomfort from work without being able to enjoy leisure; obvious examples are a spell of flu, or a broken leg. Similarly, a negative a together with a positive b could reflect a health problem that makes it difficult to work without impairing one's possibilities to enjoy leisure (for example, an allergic reaction to chemical substances used at the workplace, perhaps at the same time as the Summer Olympics are broadcast on television).

While the individual's evaluations of work and leisure are expressed as continuous taste variables, we will treat work itself as a dichotomous variable, taking either the value $1 - \ell = 1$ (working) or $1 - \ell = 0$ (not working). In other words, we confine the analysis of labor supply to the extensive margin. Indeed, this is the most relevant margin when studying income insurance, which pays benefits when the individual does not work. (The model could easily be extended to the case where the individual is allowed to work part time and obtain an insurance benefit for the other part).

At the beginning of each period, the individual observes his/her preference parameters, and then decides whether to go to work or not. Denoting consumption when working by c^w, the individual's utility in this case is[4]

$$u^w = u(c^w) + a . \tag{1}$$

Denoting consumption when absent from work by c^a, the individual's utility in that case is

$$u^a = u(c^a) + b . \tag{2}$$

---

[4] An alternative approach would have been to model the individual's health status in terms of his labor productivity, rather than in the context of his utility function. This would, however, have been much more complicated, since it would then have been necessary to explain how a change in the individual's productivity of work actually influences his work decision. An interesting issue is whether, or to what extent, these two approaches are equivalent, and whether they can be combined.

In each period, values of a and b are drawn from a probability distribution; the individual chooses to work if $u^w > u^a$, and to be absent in the opposite case. While separate work and leisure parameters would allow us to study implications of different correlations between a and b, the exposition would become somewhat cumbersome. The basic insurance-theoretic points in our paper can be effectively made if instead of a and b, we use a variable α interpreted as $a - b$. Thus, we write functions (1) and (2) as $u^w = u(c^w) + \alpha$ and $u^a = u(c^a)$, respectively.

To make the insurance problem non-trivial, we assume that the insurer cannot observe the realization of the stochastic variable α, although he knows its probability distribution. The insurer can also observe whether the individual goes to work or not. In the real world, it is true that some health deficiencies, such as a broken leg, can easily be verified by external observers. However, such verification is often not possible for ailments where the insured individual's self-assessment is crucial. Important examples are neuro-muscular problems and psycho-social problems, which constitute the bulk of diagnosed cases in connection with work absenteeism today in the context of sick-pay insurance.

3. Individual Behavior

Let α_i, $i = 1, \ldots, n$, denote the parameter α of individual i. We assume that α_i is drawn from a probability distribution with mean $\overline{\alpha}_i$, which may differ across individuals. In most of the discussion, we will assume that individuals are identical in all other respects (for instance, in terms of wage rate, risk aversion, etc.). If, for a particular realization of α_i, agent i chooses to work, his consumption is $c^w = w(1 - p)$, where w is the wage rate, and p is the sick-pay insurance premium rate.[5] Thus, equation (1) becomes

$$u_i^w = u\big(w(1 - p)\big) + \alpha_i. \qquad (1')$$

---

[5] Non-labor income is easily included in the model; however, such income would not add any insights, but would complicate the notation.

If, for some other realization of α_i, the agent chooses to be absent from work, his consumption is $c^a = bw$, where b is the replacement rate in the sick pay insurance. (For simplicity, we abstract from intertemporal links; the agent can thus not self-insure by saving income from one period to another.) Equation (2) then becomes

$$u_i^a = u(bw).\tag{2'}$$

The cut-off value of α_i, for which the individual is indifferent between work and non-work, satisfies $u_i^w = u_i^a$. This cut-off point is given by

$$\alpha_i^* = \alpha^* = u(bw) - u\big(w(1-p)\big).\tag{3}$$

The probability that an individual is absent from work is then equal to the probability that $\alpha_i < \alpha^*$. Note that α^* is the same for all individuals, since no terms on the right-hand side have an index i. The intuition is that we have assumed that individuals may only differ in terms of the distribution of the stochastic preference parameter α, and not with respect to the consumption utility function u and productivity w.[6] The absence rate then depends on how often the individual's α_i falls below the cut-off point α^* which, in turn, depends on the individual's mean $\overline{\alpha}_i$. We see from (3) that the monotonicity of $u(c)$ implies

$$1 - p > b \quad \Leftrightarrow \quad \alpha^* < 0.\tag{4}$$

Thus, with less than full insurance (i.e., $1 - p > b$), the cut-off point α^* will be a negative number; when it is costly for the individual to be absent, he chooses a low cut-off point (α^*) to prevent that the realization of α_i too often falls below α^*. In other words, he stays home only when the realized α_i is exceptionally bad. A positive

---

[6] If the individuals were to differ with respect to w, the cut-off α^*, and thereby the absence rate, would be a function of w. Depending on the curvature of the utility function $u(c)$, we could then get either the result that absence is lower for high-wage individuals (which seems to be the most realistic feature of today's absence data) or the other way round.

α^* would mean overfull insurance (which, although perhaps not a very common case in the real world, might be a possible market outcome in our model; see below). The intuition for this specific case is that if the individual has a higher income when being absent, he goes to work only when the realized α_i is very high (i.e., when work happens to be very pleasant).

To derive an expression for the probability of being absent, we must specify the probability distribution of α_i. For simplicity, we choose a uniform distribution with mean $\bar{\alpha}_i$ and spread s; see *Figure 1*. Thus, $\alpha_i \in [\bar{\alpha}_i - s, \ \bar{\alpha}_i + s]$. If we had allowed the spread s to differ across individuals, the model would have become richer – but it would have complicated the exposition without yielding much additional insight.

(Figure 1)

In terms of Figure 1, the probability of being absent from work is equal to the area of the leftmost rectangle. Clearly, that area is[7]

$$\pi_i = \begin{cases} 1 & \text{if } \bar{\alpha}_i < \alpha^* - s \\ \dfrac{\alpha^* - \bar{\alpha}_i + s}{2s} & \text{if } \bar{\alpha}_i \in [\alpha^* - s, \alpha^* + s], \\ 0 & \text{if } \bar{\alpha}_i > \alpha^* + s, \end{cases} \tag{5}$$

where $\alpha^* < \bar{\alpha}_i \iff \pi_i < 1/2$. The intuition is obvious. If the cut-off point is less than the mean, it quite rarely happens that the realized value of α_i falls below the cut-off point and hence π_i is relatively small (a low absence rate). On the other hand, if the cut-off point is larger than the mean, the realization is often below the cut-off point and hence π_i is large (a high absence rate).

---

[7] In *Figure 1*, we have drawn the distribution of α_i, such that $\bar{\alpha}_i > 0$. Needless to say, $\bar{\alpha}_i$ might as well be negative, without any consequences for our analysis.

We assume that π_i in (5) is strictly inside the interval (0, 1); in other words, we abstract from individuals who are always (or never) on their jobs ($\pi_i = 0$ or $\pi_i = 1$). Clearly, such individuals are not interesting from an insurance point of view.

Some properties of the model can now be illustrated by the following comparative-statics results:

$$\frac{\partial \pi_i}{\partial \bar{\alpha}_i} < 0; \quad \frac{\partial \pi_i}{\partial b} > 0; \quad \frac{\partial \pi_i}{\partial p} > 0. \tag{6}$$

The interpretation of the first derivative is straightforward: if work becomes more pleasant on average, the individual will be less absent. Moreover, from the two other derivatives, we see that absence, on average, is higher with than without an insurance system, and that absence is increasing by the size of the insurance scheme.[8] As we shall see, this means that we can regard moral hazard as a continuous variable; the individual adjusts his/her behavior in a continuous way in response to the rules of the insurance system.

In other words, we measure moral hazard as the frequency of absence from work in excess of what would prevail if the individual were uninsured.[9] Accordingly, in our model, some moral hazard will always exist, and its size depends on various parameters of the model, including the insurance system. We believe this definition of moral hazard to correspond quite well to the common usage of the term in economics.

Aggregating the expression in equation (5), with *n* individuals in society, the total absence rate is

---

[8] This is, however, just a partial effect on individual behavior; general equilibrium effects will be discussed later on, when we have introduced the supply side in the insurance market. The unambiguously positive sign of $\partial \pi_i / \partial p$ may seem surprising. The explanation is simply that the model only deals with decisions on the extensive margin, for which there is only a substitution effect (like in studies of labor force participation in the standard labor supply literature). Here, we assume that π_i is interior, i.e., $0 < \pi_i < 0$. If the parameters are such that π_i is at a corner solution (i.e., $\pi_i = 0$ or $\pi_i = 1$), some of the above derivatives might be zero.

[9] As mentioned earlier, Diamond and Mirrlees (1978) define moral hazard as fraud, in the sense that a healthy individual pretends to be sick.

$$\pi = \frac{1}{n}\sum_{i=1}^{n}\pi_i = \frac{\alpha^* - \bar{\alpha} + s}{2s}, \tag{7}$$

where $\bar{\alpha} \equiv (1/n)\sum\bar{\alpha}_i$. For the total absence rate in the economy, we also have the

relation $\alpha^* < \bar{\alpha} \iff \pi < 1/2$. The partial derivatives of aggregate work absence, π,

look the same as the individual derivatives in (6).

For a given insurance system, (b, p), the individual's expected utility is

$$EU_i = (1-\pi_i)\left[u(w(1-p)) + E(\alpha_i|\alpha_i \geq \alpha^*)\right] + \pi_i \cdot u(wb), \tag{8}$$

where $E(\alpha_i|\alpha_i \geq \alpha*) = (s + \alpha^* + \bar{\alpha}_i)/2$. Let us substitute this expression, together with

the expressions for π_i in (5) and α^* in (3), into (8). Setting EU_i equal to some

constant, k, we obtain an expression for an indifference curve in the (p, b) plane.

Differentiating this expression and rearranging terms yields the slope, Q, of the

indifference curve:

$$Q \equiv \frac{db}{dp}\bigg|_{EU_i = k} = -\frac{u'(w(1-p))}{u'(wb)} \cdot \frac{(\alpha^* - \bar{\alpha}_i - s)}{(\alpha^* - \bar{\alpha}_i + s)}. \tag{9}$$

From (5), we see that $(\alpha^* - \bar{\alpha}_i - s) < 0$ since $\pi_i < 1$, and that $(\alpha^* - \bar{\alpha}_i + s) > 0$ since

$\pi_i > 0$. Thus, the second term on the right-hand side of (9) is always negative, which

means that the indifference curves are always upward-sloping in the (p, b) plane,

which is what we would also intuitively expect.

While the slope of the indifference curve is thus unambiguous, the curvature is not. It

is, however, easily seen that the slope $Q \to \infty$ as $\alpha^* \to \bar{\alpha}_i - s$. The interpretation is

that for a combination (p, b) such that the individual very seldom chooses to be absent

(hence, when $\pi_i \to 0$ according to equation (5)), the indifference curve is steep. The

intuition is simple: For an individual who almost always works, there is no rise in b

high enough to compensate for an increase in p – simply because he hardly ever

receives any *b*. Similarly, $Q \to 0$ as $\alpha^* \to \overline{\alpha}_i + s$. For an individual who seldom works ($\pi_i \to 1$), the indifference curve is rather flat.[10] Such an individual does not require any higher *b* as a compensation for an increase in *p* – simply because he (asymptotically) never pays any *p*.

Thus, the indifference curve has one steep segment for small values of *p*, and one flat segment for large values of *p*, as shown in *Figure 2*. It is also easily shown that the indifference curves must be concave close to $\overline{\alpha}_i - s$ and $\overline{\alpha}_i + s$, respectively. However, the indifference curves may or may not be concave over the whole interval $[\overline{\alpha}_i - s, \overline{\alpha}_i + s]$.

Figure 2.

The fact that indifference curves in income insurance may contain both concave and convex segments has earlier been observed in the literature on *ex ante* moral hazard (see, for instance, Arnott, 1992). Thus, this observation concerning *ex ante* moral hazard carries over to our model of *ex post* moral hazard. While no clear-cut results concerning the shape of the indifference curves seem to have been derived in the *ex ante* moral hazard literature, the simple structure of our model, in fact, permits such a result: a necessary condition for the indifference curve to have a convex segment (as illustrated by the dashed curve in Figure 2) is that

$$(\alpha^* - \overline{\alpha}_i - s) \cdot \pi_i \cdot \frac{u''(w(1-p))}{u'(w(1-p))} - u'(w(1-p)) > 0.$$

Hence, the larger is u'' relative to u', that is, the larger the absolute risk aversion, the more likely is the indifference curve to have a convex segment. In other words, for the indifference curve to have a convex segment, the consumption utility function $u(c)$ must be "sufficiently" concave.

---

[10] In the extreme case where an individual always works, the indifference curve is vertical, while it is horizontal for an individual who never works – cases ruled out by our assumption that $0 < \pi_i < 1$.

For the special case where consumption utility is linear, hence $u'' = 0$, the indifference curve is everywhere concave (as illustrated by the solid curve in *Figure 2*). The conventional view is, of course, that the insurance is not relevant for this case. As we shall subsequently see, this view does not carry over to the present model, where the individual may be interested in income insurance, even in the presence of linear consumption utility.

4. The Insurer's Behavior

Let us now look at the insurance problem from the insurer's point of view. For the time being, we assume that the insurer only offers one policy (p, b). Abstracting from administration cost, the net profit is

$$p \cdot (1 - \pi) - b \cdot \pi \ , \tag{10}$$

where π is given by (8). Substituting and rearranging, we see that a balanced-budget insurance scheme (zero profit) must satisfy

$$\alpha^* = \overline{\alpha} + s\frac{p - b}{p + b} . \tag{11}$$

This expression looks deceptively simple; it should be born in mind, however, that the left-hand side, by (3), is a non-linear function of p and b. In fact, the combinations of p and b that satisfy the zero-profit constraint (11) can be plotted as a "Laffer curve" like that in *Figure 3*, where the maximum point is reached when the disincentive effects on average labor supply, in the form of a higher cut-off rate α^*, are so strong that the replacement b rate cannot rise any further in connection with an increase in p.

Figure 3.

The slope of the zero-profit curve,

$$\left.\frac{db}{dp}\right|_{\substack{zero \\ profit}} = \frac{2sb - wu'(w(1 - p))(b + p)^2}{2sp + wu'(bw)(b + p)^2} , \tag{12}$$

is positive for *p* close to zero, and negative when the zero-profit curve intersects the horizontal axis further to the right. It can be shown (with some tedious algebra) that the curve is a well-behaved Laffer-type curve in the sense of only having one maximum in between, provided that $0 < \pi_i < 1$.[11]

An interesting question is whether full insurance is consistent with budget balance. In the present model, full insurance means that $b = 1 - p$, represented by a straight line in the *p, b* plane; see *Figure 3*. All points below the line mean less than, and all points above mean more than, full insurance. Full insurance is possible if the straight line intersects the curve representing the zero profit constraint where it slopes upwards, as we may see from the figure.[12]

5. Market Equilibrium

It is now time to confront the zero-profit condition with the indifference map. To begin with, we assume that there is only one representative individual. *Figure 4* illustrates three possible market equilibria in this case. In *Figure 4a*, there is an internal equilibrium on a convex segment of the individual's indifference curve. In *Figure 4b*, the highest indifference curve is instead attained at a corner solution with no insurance (*p* = *b* = 0).

Figure 4.

In *Figure 4c*, finally, we illustrate a situation when the indifference curve is everywhere concave. Even in such a case, we may, of course, obtain an internal market equilibrium *p* > 0, *b* > 0. As noted in Section 3 above, a special case resulting in a concave indifference curve occurs when the consumption utility function $u(c)$ is linear. It may seem counterintuitive that insurance is warranted in this case, since it

---

[11] The proof consists of acknowledging that the left-hand side of (12) is an increasing, convex function of p, while the right-hand side is an increasing, concave function of p. These two functions can intersect at most twice, which can be shown to imply a well-behaved Laffer curve (utilizing our previous assumption that $0 < \pi < 1$).
[12] In *Figure 3*, we have drawn the zero profit curve in such a way that it intersects the horizontal axis at a point *p* < 1. Intersection points *p* > 1 are, in principle, possible, but probably of limited empirical interest.

might be considered that insurance is relevant, only if the consumption utility function is strictly concave. The reason for our result is that the probability of an insurance-generating event, π_i, is endogenous and depends on consumption $c^w = w(1-p)$ and $c^a = wb$, respectively, via equation (5). As a consequence, the expected utility EU_i, as given by (8), in fact becomes non-linear in consumption, and may thus lead to a demand for insurance, even if the consumption utility $u(c)$ is linear (or even convex).

Let us now turn to comparative statics by looking at a shift in the individual's preferences concerning work as compared to non-work, expressed by $\overline{\alpha}_i$. Since this parameter only shows up in the second term of equation (10), the change in the slope of the indifference curve is

$$\frac{\partial Q}{\partial \overline{\alpha}_i} = -\frac{u'\big(w(1-p)\big)}{u'(bw)} \cdot \frac{\partial\left(\dfrac{\alpha^* - \overline{\alpha}_i - s}{\alpha^* - \overline{\alpha}_i + s}\right)}{\partial \overline{\alpha}_i} = \frac{u'\big(w(1-p)\big)}{u'(bw)} \cdot \frac{2s}{(\alpha^* - \overline{\alpha}_i + s)^2} > 0 .$$

Thus, a higher $\overline{\alpha}_i$ makes the indifference curves steeper. The intuition is that if $\overline{\alpha}_i$ rises, i.e., the individual finds working more pleasant on average, he requires a higher increase in b than earlier to be compensated for a rise in p, since he will now receive b very seldom.

The zero-profit curve is, however, also affected; a higher $\overline{\alpha}_i$ shifts this curve upwards.[13] Intuitively, a higher $\overline{\alpha}_i$ means that people evaluate work higher (or enjoy leisure less). Due to the ensuing lower absence rate, the insurer can then pay a higher benefit for a given premium.

Thus, while the indifference curves become steeper, the zero-profit curve shifts upwards. Without putting more structure on the model, we cannot say whether the

---

[13] For a given value of b, the right-hand side of (11) is an increasing, convex function of p, and the left-hand side is an increasing, concave function of p. These two functions will intersect at two points, corresponding to two values of p for each value of b. Thus, the solution set of (12) can be depicted as the inverted-U, Laffer-curve shape in *Figure 3*. Increasing $\overline{\alpha}_i$ will shift the right-hand side of (11) upward, while the left-hand side will be unaffected; thus, the two points of intersection will diverge. This means that the zero-profit curve will shift upwards.

new equilibrium implies more or less insurance (i.e., whether p and b increase or fall). The intuition is that we do not know, in general, whether the rise in b that the insurer can now provide (for a given p) is large enough to satisfy the individual's demand for a higher b.

So far, we have assumed individuals to be homogenous *ex ante* (i.e., they have the same probability distribution of α_i) although they are heterogeneous *ex post* (i.e., the realization of α_i differs across individuals). In reality, individuals are, of course, heterogeneous also *ex ante*, in the sense that the probability distribution of α_i differs across individuals, for instance by a different mean, $\overline{\alpha}_i$. As a result, for a given insurance system (p, b), different individuals would systematically choose different absence rates.

Further, individuals who enjoy work more than others (i.e., who have relatively high values of $\overline{\alpha}_i$ and consequently, relatively steep indifference curves in the p, b plane) would not like to share the insurance system with those who enjoy work less (i.e., those having low values of $\overline{\alpha}_i$ and flat indifference curves). If the insurance company does not know to which group a specific individual belongs, we enter the world of asymmetric information as dealt with in the insurance literature following the seminal paper by Rothschild and Stiglitz (1976). One of their points is that a pooling equilibrium is not possible in this situation, because some insurer will offer low-risk individuals a better deal (offering a separating equilibrium). The authors also argue, however, that a separating equilibrium can only occur under very restrictive assumptions, since high-risk individuals would mimic those with low risk, hence undermining the financial sustainability of a separating equilibrium. Thus, while pooling equilibria would be impossible, separating equilibria would be unlikely.[14]

We are, however, not convinced that asymmetric information in the Rothschild-Stiglitz sense is a major problem in real-world income insurance markets. One reason is that, in many cases, it is possible to confine a particular insurance policy to a group

---

[14] Rothschild and Stiglitz seem to abstract from rational expectations. With rational expectations in the case when no separating equilibrium is possible, a pooling equilibrium would be stable, since nobody would then challenge this equilibrium by trying to create a separating one.

of individuals with an observable characteristic (for instance, profession), where the within-group variations in risk are modest. Indeed, such insurance schemes were very common before the emergence of mandatory social insurance, and they are still prevalent in many countries as supplements. Another reason why adverse selection may not be a serious problem in the case of income insurance is that, contrary to the prediction of Rothschild and Stiglitz (1976), real-world low-risk individuals are often particularly interested in buying insurance policies, while high-risk individuals often choose not to do so. The reason may be that a personal trait of "prudence" often takes the form of not only a low-risk lifestyle, but also a strong desire to be insured – a feature generating "advantageous selection".[15]

What are, then, the rationales for government intervention in the context of our model? It is useful to distinguish between three rationales. First, even in the case of a homogenous population, the market equilibrium may be one of zero insurance (as depicted in *Figure 4b*). A paternalistic government may introduce a mandatory system in this case. Second, even if a separating equilibrium could be obtained, the government may not like it, mainly for distributional reasons. Some groups will be offered more generous insurance policies than others, and some individuals may not be able to get any insurance at all.[16] Third, as mentioned above, while low-risk groups are often willing and able to buy insurance, high-risk people may choose not to do so (for instance, because of low incomes, or because of limited "prudence"). For distributional reasons the government may want to transfer resources to the latter group, and for reasons of paternalism the government may choose to do so by way of mandatory insurance rather than by cash transfers.

6. What is Moral Hazard?

In this paper, we have dealt with moral hazard *ex post*, that is, we have discussed how the individual behaves after an insured event has occurred. In terms of our model, this means that after a realization of the random variable α_i, the individual decides

---

[15] This point has been developed in Barsky et al. (1997) and de Meza and Webb (2001).
[16] Our model highlights the possibilities of drastically different separating equilibria for different groups of citizens. The reason is that the attitude to work (in our notation, $\overline{\alpha}_i$) is likely to differ strongly across occupations. Indeed, as an extreme case, for groups with a relatively high $\overline{\alpha}_i$, full insurance (and even overfull insurance) is possible. While the notion of such a high enjoyment from work is not realistic for most people, it may not be far-fetched for some readers of this paper.

whether to call sick – a decision that, in turn, depends on the insurance parameters p and b. We measure the amount of moral hazard for individual i as work absence above what would prevail if he/she were not insured, i.e.,

$$MH_i \equiv \pi_i - \pi_i^0 \equiv \frac{u(wb) - u(w(1-p)) - \overline{\alpha}_i + s}{2s} - \frac{u(0) - u(w) - \overline{\alpha}_i + s}{2s},$$

where π_i^0 is work absence if there were no insurance.[17] The aggregate amount of moral hazard, MH, is, of course, ΣMH_i. When absent from work, the difference in utility with and without insurance is $\Delta u^a \equiv u(wb) - u(0)$. Similarly, when working, the corresponding difference is $\Delta u^w \equiv u(w(1-p)) - u(w)$. Moral hazard may then be compactly written as[18]

$$MH_i \equiv \frac{1}{2s}(\Delta u^a - \Delta u^w).$$ (13)

From the inequalities in (7), we see that moral hazard is strictly positive when there is an insurance system (i.e., when $b > 0$, $p > 0$).

In our model, moral hazard is not just mimicking (like it is in the Diamond-Mirrlees model). The reason is that the individual's ability and willingness to work is a continuum, rather than a dichotomous variable. In each time period, the individual weighs the pros and cons of not going to work, and concepts like "mimicking" and "fraud" are too blunt to characterize that choice.

In the tradition of Diamond and Mirrlees (1978), the analysis of optimal insurance has typically been based on the maximization of a social welfare function (usually a Utilitarian one) subject to a moral hazard constraint and a budget constraint for the insurer. The moral hazard constraint implies that no individual should find it worth-

---

[17] Work absence π_i^0 is defined by (5) and (3), with $p = b = 0$.

[18] Here, every individual has the same amount of moral hazard, which means that

$MI = \Sigma MH_i = n \cdot \frac{1}{2s}(\Delta u^a - \Delta u^w).$ This is a consequence of our simplifying assumption that individuals only differ with respect to $\overline{\alpha}_i$.

while to pretend to be sick when he is not. Since Diamond and Mirrlees analyze the case of homogenous individuals, it may seem natural to impose such a constraint, since there are only two possible equilibria in this framework: either everybody cheats, or nobody cheats. It is, of course, possible to impose moral hazard constraints also in the case of heterogeneous individuals. For instance, Whinston (1983) imposes moral hazard constraints in a model with two groups of individuals, each with a different probability of being disabled. In that model, he concludes that the optimal policy of a mandatory insurance (when individual characteristics cannot be observed by the insurer) is a single, pooling policy. The policy must be strict enough for the moral hazard constraint to be satisfied for the group with the highest probability of being unable to work, which implies that it will be satisfied also for the other group. In other words, the individuals in both groups will prefer to work when being able to do so. Whinston also shows that the same conclusion holds for an arbitrary number of groups, differing in terms of the probability of being unable to work.

This is an extremist solution, however, illustrating the unsuitability of imposing moral hazard constraints on the government's optimization. Removing the moral hazard constraints, and maximizing social welfare subject to the insurer's budget constraint only, will normally result in higher welfare (or, as a special case, the same welfare). This becomes particularly evident in a society with a large number of widely differing groups in term of the risk of being unable to work. Then, the system must be very meager indeed, having to be adjusted to the group with the highest probability of being unable to work – thereby providing very little insurance for other groups.

Without any moral hazard constraints, a better insurance could be provided. Although some individuals will then cheat, others will get better insurance, and social welfare – based on all individual utilities – will increase. The number of cheaters will then be endogenously determined in the maximization process – and zero cheaters would be a special and quite unlikely outcome. Indeed, this is the obvious trade-off that confronts the designer of a mandatory income insurance system.

A similar argument can be applied to the original Diamond and Mirrlees (1978) analysis for homogenous individuals. The moral hazard constraint is simply redundant in that framework. The reason is that the insurer's budget constraint will, by itself,

ensure that everybody works, and nobody cheats. (Naturally, if everybody were to cheat, the income of households, and the revenue of the insurer, would be very low and hence inconsistent with welfare maximization.)

By contrast, in our model, the issue of such constraints never arises. The reason is that we analyze the individual's choice as a trade-off between the pains and pleasures of going to work, rather than as an issue of cheating or not cheating. The individual can himself decide under which conditions he can utilize the benefits of the insurance system – and the issue is not a clear-cut case of honesty dishonesty. We believe that our approach gives a more realistic picture of the individual's choice situation, and, thereby, of the government's choice of insurance system.

Our approach to moral hazard is relevant not only for sick pay and early retirement (permanent disability) insurance, but also for unemployment insurance. The stochastic parameter α_i in our model would, in the latter case, reflect the attractiveness of a specific job offer as compared to continuing search and enjoying leisure (relative to the offered job).[19] The probability that an individual will accept a given job offer, rather than continuing to be unemployed, then depends on the realization of α_i as compared to α^* and hence, on parameters p and b.

In real-world insurance systems, p and b are, of course, not the only parameters available for the insurer. The latter can also exert stronger or weaker administrative control to ensure that the individual is sufficiently sick to qualify for sick pay or early retirement, and that he is searching for jobs in a serious manner when claiming unemployment benefits. Naturally, this approach corresponds to the established view that there is a trade-off between insurance, incentives, and controls. The need for administrative controls and strong work incentives in the insurance system is, of course, smaller if there also exists informal social control, as a result of social norms in society in favor of working or against living on benefits. Indeed, our model is well suited for dealing with this issue.

---

[19] For a related model, see Dionne (1984).

7. Social Norms

7.1 Individual Behavior

So far, we have abstracted from the possibility that one person's absence behavior is influenced by how this behavior is judged by others – that is, we have abstracted from social norms. Our model is, however, well suited for integrating social norms into an insurance-theoretic framework with moral hazard. The reason is that our model endogenously determines the number of individuals who are absent from work, and that it is natural to assume that the strength of a social norm concerning individual absence depends on this number. When a large number of individuals are absent from work, such behavior is likely to be more legitimate than what would otherwise be the case. [20]

To integrate social norms into the model, we assume there to be a stigma from being absent from work and living on benefits, and that the individual is less stigmatized the larger is the number of people who are also absent. The utility u^a may then be written

$$u^a = u(bw) - f(\pi^n); \quad f(\pi^n) > 0; \quad f'(\pi^n) \leq 0, \qquad (2'')$$

where the superscript "n" on π indicates that we now deal with the absence rate in the presence of social norms, in contrast to the old π, given by (5). Here, the function $f(\cdot)$ indicates the stigma of being absent from work. Comparing (2'') to the original version (2), we see that $f(\pi^n)$ here plays the same role as the taste parameter β, if β had been a deterministic function instead of a stochastic parameter. The new cut-off point α^n is given by

$$\begin{aligned} \alpha^n &= u(bw) - u\big(w(1-p)\big) - f(\pi^n) \\ &= \alpha^* - f(\pi^n), \end{aligned} \qquad (3')$$

where α^* is the old cut-off point in a model without social norms. Naturally, introducing social norms reduces the cut-off point and, thereby, the absence rate:

---

[20] An alternative interpretation is that, when absent from work, the individual can enjoy leisure more if there are other non-working individuals with whom to interact.

$$\pi_i^n = \frac{\alpha^n - \bar{\alpha}_i + s}{2s}$$

$$= \frac{\alpha^* - (\bar{\alpha}_i + f(\pi^n)) + s}{2s}.$$

(5')

The aggregate version of (5') is

$$\pi^n = \frac{\alpha^* - (\bar{\alpha} + f(\pi^n)) + s}{2s}.$$

(7')

We see that introducing a linear norm term $f(\pi^n)$ into the utility function is formally equivalent to assuming that the mean of the α_i distribution is not $\bar{\alpha}_i$, but $\bar{\alpha}_i + f(\pi^n)$, which means that working becomes more attractive as compared to leisure, but that this mean is decreasing in π^n.

Equation (7') may have multiple solutions for π^n if the $f(\pi^n)$ function is non-linear. Indeed, there is no *a priori* reason for a particular curvature; both concave and convex $f(\pi^n)$ functions, and combinations of such functions, are reasonable. It follows from (7') that if $f(\pi^n)$ is convex, so that $-f(\pi^n)$ is concave, there is at most one root π^n. By contrast, if $f(\pi^n)$ has concave sections, there may be multiple roots.

When studying these issues, it is instructive to start with the simple case of a linear $f(\cdot)$ function. Then, the solution is unique, and it is easy to derive an expression for a "social multiplier". We simply assume $f(\pi^n) \equiv \gamma \cdot (1 - \pi^n)$, where γ is a positive constant. With this parameterization, (7') yields the unique closed-form solution

$$\pi^n = \frac{\alpha^* - (\bar{\alpha} + \gamma) + s}{2s - \gamma}.$$

(5')

Using the expression for π (i.e., the aggregate absence rate without social norms, given by (5)) and rearranging terms, we can write π^n as a linear function of π:

$$\pi^n = \frac{2s}{2s - \gamma} \cdot \pi - \frac{\gamma}{2s - \gamma} . \qquad (7'')$$

Suppose that the government changes some parameter x (for instance, p or b), and that as a result the without-norms absence rate π changes by $\partial \pi / \partial x$. With norms, the change in the absence rate will be larger:

$$\frac{\partial \pi^n}{\partial x} = \frac{2s}{2s - \gamma} \cdot \frac{\partial \pi}{\partial x},$$

where $2s/(2s - \gamma) > 1$. Thus, π^n is more sensitive to policy changes than π, and we can regard $2s/(2s - \gamma)$ as a "social multiplier" in the sense of Glaeser, Sacerdote and Scheinkman (2003).

The solution to equation (7') in the case of a linear norms term $f(\pi^n) \equiv \gamma \cdot (1 - \pi^n)$ is illustrated in *Figure 5a*. Here, the right-hand side of (7') is represented by an upward-sloping line, while the left-hand side is represented by the 45-degree line. Those two lines intersect at the equilibrium point A. If government policy results in an upward shift of the line, the new solution occurs at point B. Thus, a given shift of the line results in a larger change in π^n than the vertical shift.

(Figure 5)

In *Figure 5b*, the solid curve represents a non-linear $f(\pi^n)$ function with three potential solutions, where the middle solution is unstable under reasonable assumptions about the dynamics.[21] Assume, for instance, that the economy is initially located at the "good" equilibrium A, and that government policy shifts the curve to the broken line with a new, still rather good, equilibrium B. This point is still locally stable, but it is dangerously close to the new, unstable equilibrium. This means that a modest shock may push the system all the way to the bad equilibrium, C.

---

[21] For instance, it is reasonable to assume that the individual observes aggregate absence only with a time lag. As a result, aggregate π at time t becomes a function of aggregate π at time $t - 1$.

7.2 Market Equilibrium and Moral Hazard

When analyzing the role of social norms in the insurance market, a crucial question is what we assume about the rationality of the agents. More specifically, we have to take a stand on the issue of whether the individual realizes that the aggregate absence rate, π^n, changes as he chooses to change his own absence rate, π_i^n, in response to changes in parameters such as *p* and *b*.

It is useful to study two polar cases. One is that the individual does not realize the consequences for aggregate π^n; he simply takes some historically determined $\bar{\pi}$ as given. We will call this case "non-rational expectations", and it may be realistic in a short and medium term perspective. The other polar case is that of rational expectations, in the sense that the individual fully realizes how aggregate absence changes. This case may be realistic in a long-term perspective, when individuals have learnt how others change their behavior.

What, then, would be a realistic assumption regarding the expectations of *insurance providers*? They will immediately discover changes in aggregate absence, since such changes influence their revenues and expenditures. Therefore, it is reasonable to assume that they do have rational expectations, in the sense that the actual π enters their profit function $(1 - \pi^n) \cdot p - \pi^n \cdot b$. This holds regardless of whether individuals have rational or non-rational expectations.

With non-rational expectations among individuals, introducing social norms into the model is formally equivalent to increasing the mean of the taste parameter distribution, $\bar{\alpha}_i$. As we pointed out in Section 5, this means that the indifference curves become steeper and the zero-profit curve shifts upwards – but that we cannot say whether the equilibrium size of the insurance scheme, i.e., the size of *p* and *b*, will increase or decrease.

The qualitative effects are the same in the case of rational expectations. The quantitative effects will differ, however; the indifference curves will not become quite as steep as without rational expectations, and the zero-profit curve will not shift upwards to the same extent. Naturally, the reason is that the individual will realize that when

he increases his absence π_i^n, others will do the same. As a result, non-work will be more attractive than without rational expectations.

Applying the definition of moral hazard in Section 6 to the case with social norms, we have

$$MH_i^n \equiv \pi_i^n - \pi_i^{n0},$$

where the last term denotes the absence rate in the presence of norms, but without insurance. Inserting expression (5') for the absence rate, and assuming a linear norm term $f(\pi^n) \equiv \gamma \cdot (1 - \pi^n)$, we obtain the expression for moral hazard:

$$MH^n \equiv \frac{1}{2s - \gamma}(\Delta u^a - \Delta u^w), \tag{13'}$$

where, as usual, we assume that $2s > \gamma$. Equation (13') can be compared to the corresponding expression for moral hazard in the absence of social norms, (13); we then see that

$$MH^n \equiv \frac{2s}{2s - \gamma} MH,$$

where $2s / (2s - \gamma)$ is the "social multiplier" discussed in Section 7.1. In other words, social norms function as a multiplier on moral hazard.

8. Concluding Comments

We have developed a model of income insurance with moral hazard *ex post*, without using the traditional concept of a "moral hazard constraint". In fact, in a model with homogenous agents, such a constraint turns out to be redundant. In a model with heterogeneous agents, imposing a moral hazard constraint is inconsistent with the maximization of social welfare (at least if the social welfare function is of the type used in the normative literature on income insurance).

By contrast to the traditional insurance literature, we have treated the individual's ability and willingness to work as a continuous variable, which means that his/her decision to work is affected by a broad set of health-related and psychological factors. These factors are represented by a continuous stochastic variable, α_i, in the individual's utility function. As a result, absence from work also becomes a continuous variable that depends on the parameters of the insurance system. In this framework, moral hazard will always exist in the sense that the individual changes his behavior as a result of being insured. In our model, the extent of moral hazard is then reflected in the difference between the absence rate with and without insurance. We also show that the model is suitable for the analysis of the interplay between moral hazard and social norms, the strength of which is assumed to depend on the total number of individuals in society who live on benefits rather than earnings from work. Indeed, social norms create a multiplier effect on work absence and moral hazard.

References

Allen, S. G. (1981): "An Empirical Model of Worker Attendance", *Review of Economics and Statistics*, Vol. 71, No. 1, pp 1-17.

Arnott, R. (1992): "Moral Hazard and Competitive Insurance Markets", in G. Dionne (ed.), *Contributions to Insurance Economics*, Kluwer Academic Publishers, Boston.

Barmby, T., J. Sessions and J. Treble (1994): "Absenteeism, Efficiency Wages and Shirking", *Scandinavian Journal of Economics*, Vol. 96, No. 4, pp 561-566.

Barsky, R. B., F. T. Juster, M. S. Kimball and M. D. Shapiro (1997): "Preference Parameters and Behavioral Heterogeneity: An Experimental Approach in the Health and Retirement Study", *Quarterly Journal of Economics*, Vol. CXII, No. 2, pp 537-579.

Brown, S. and J. G. Sessions (1996): "The Economics of Absence: Theory and Evidence", *Journal of Economic Surveys*, Vol. 10, No. 1, pp 23-53.

Cochrane, A. L. (1972): "The Measurement of Ill Health", *International Journal of Epidemiology*, Vol. 1, pp 89-92.

Diamond, P. A. and J. A. Mirrlees (1978): "A Model of Social Insurance with Variable Retirement", *Journal of Public Economics*, Vol. 10, pp 295-336.

Dionne, G. (1984): "Search and Insurance", *International Economic Review*, Vol. 25, No. 2, pp 357-367.

Eek, D. and K. Rikner (2005): "What determines people's decisions whether or not to report sick?", *Applied Economics*, Vol. 37, No. 5, pp 533-543.

Glaeser, E. L., J. Scheinkman and B. I. Sacerdote (2003): "The Social Multiplier", *Journal of the European Economic Association*, Vol. 1, No. 2, pp 345-353.

Layard, P. R. G., and A. A. Walters (1978): *Microeconomic Theory*, McGraw-Hill, New York.

De Mesa, D. And D. C. Webb (2001): "Advantageous Selection in Insurance Markets", *The RAND Journal of Economics*, Vol. 32, No. 2, pp 249-262.

Rothschild, M. and J. Stiglitz (1976): "Equilibrium in Competitive Insurance Markets: An Essay on the Economics of Imperfect Information", *Quarterly Journal of Economic*s, Vol. 90, No. 4, pp 629-649.

Sinn, H.-W. (1983): *Economic Decisions under Uncertainty*, North-Holland, Amsterdam and New York.

Stiglitz, J. (1983): "Risk, Incentives and Insurance: The Pure Theory of Moral Hazard", *The Geneva Papers on Risk and Insurance*, Vol. 8, No. 26, pp 4-33.

Whinston, M. D. (1983): "Moral Hazard, Adverse Selection, and the Optimal Provision of Social Insurance", *Journal of Public Economics*, Vol. 22, pp 49-71.
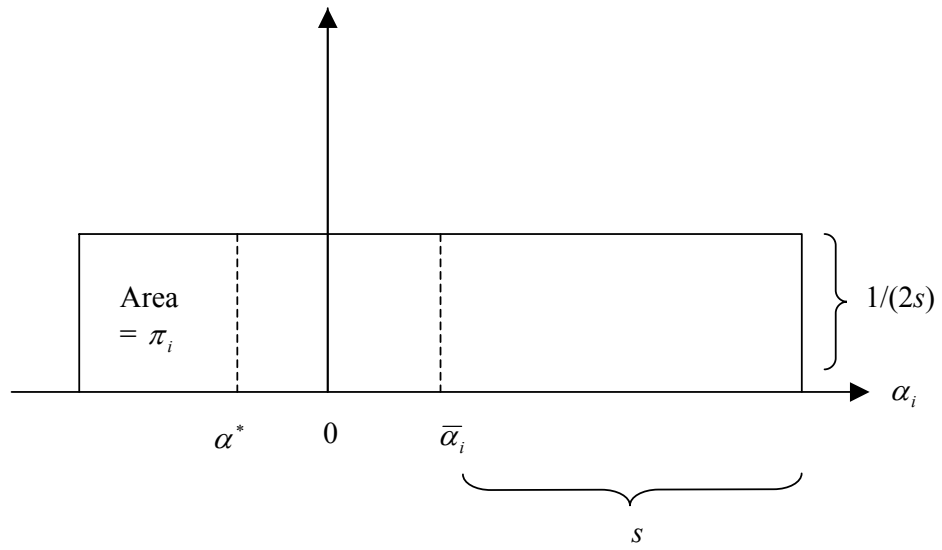
Figure 1: The density function of α.
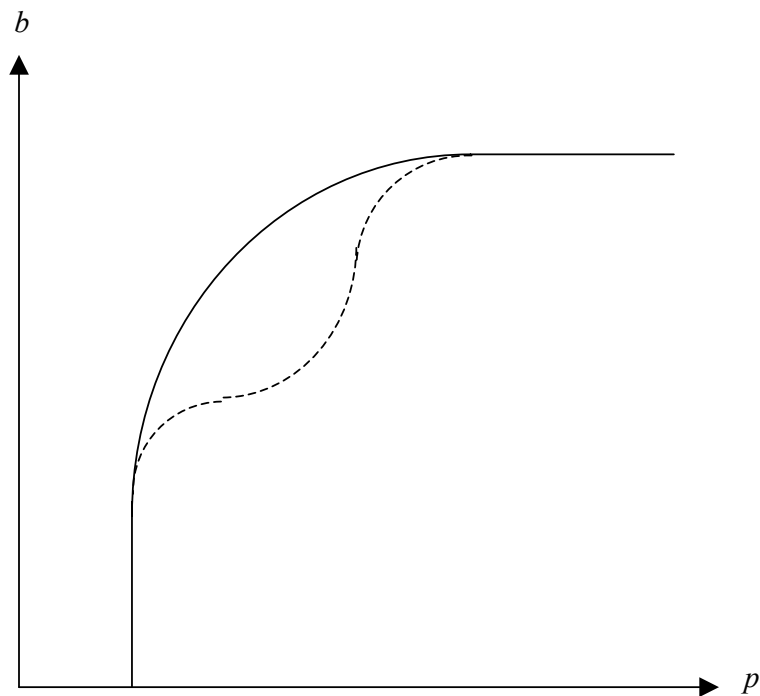
Figure 2: Possible shapes of the indifference curve
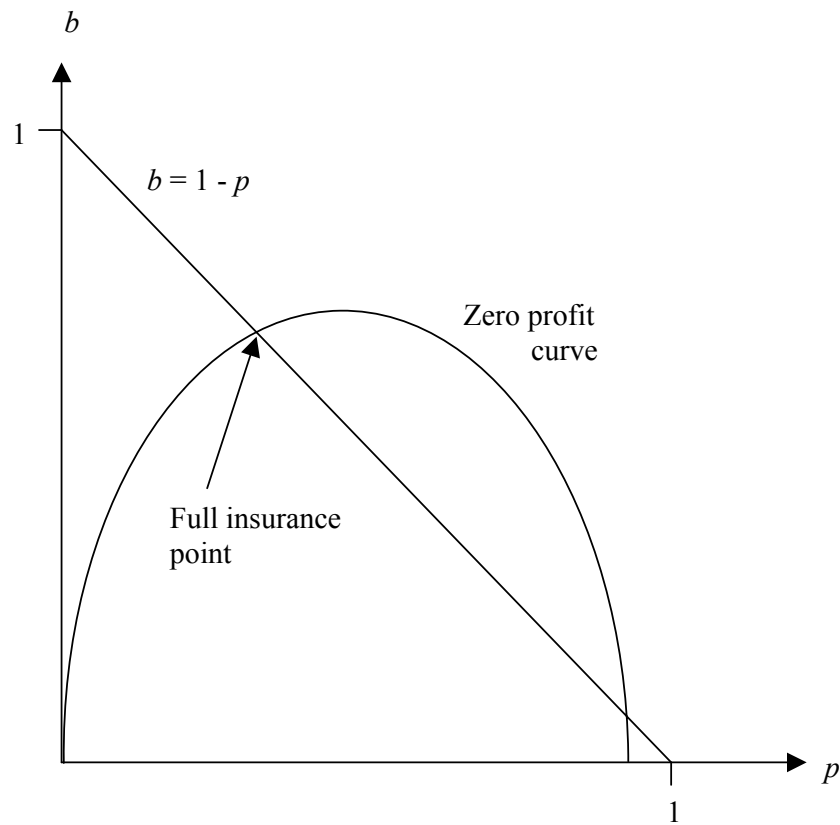
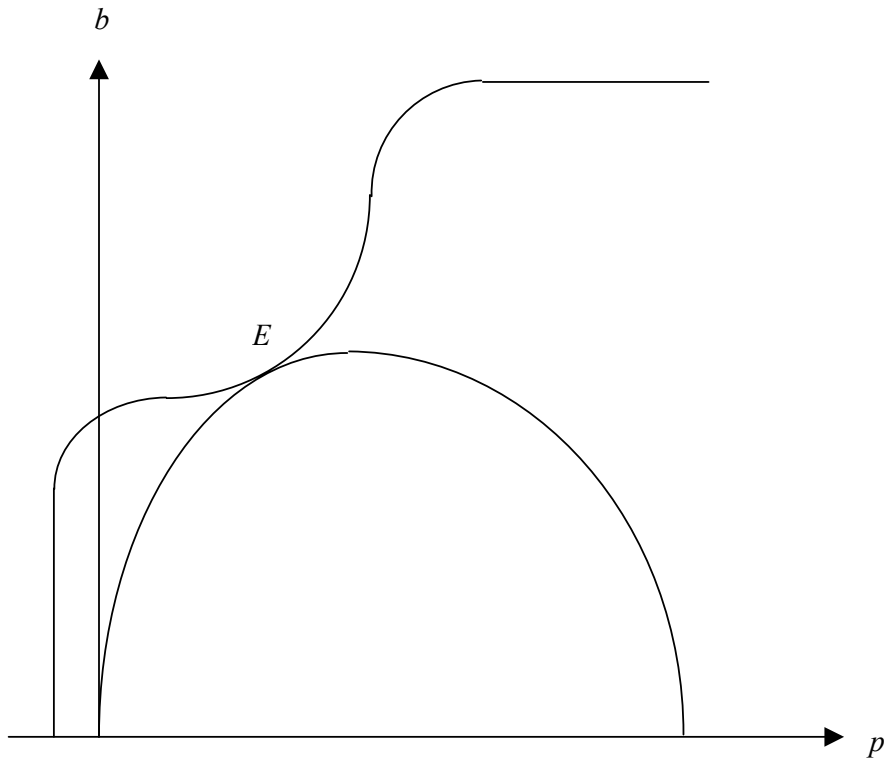Figure 3: The locus of admissible combinations of p and b

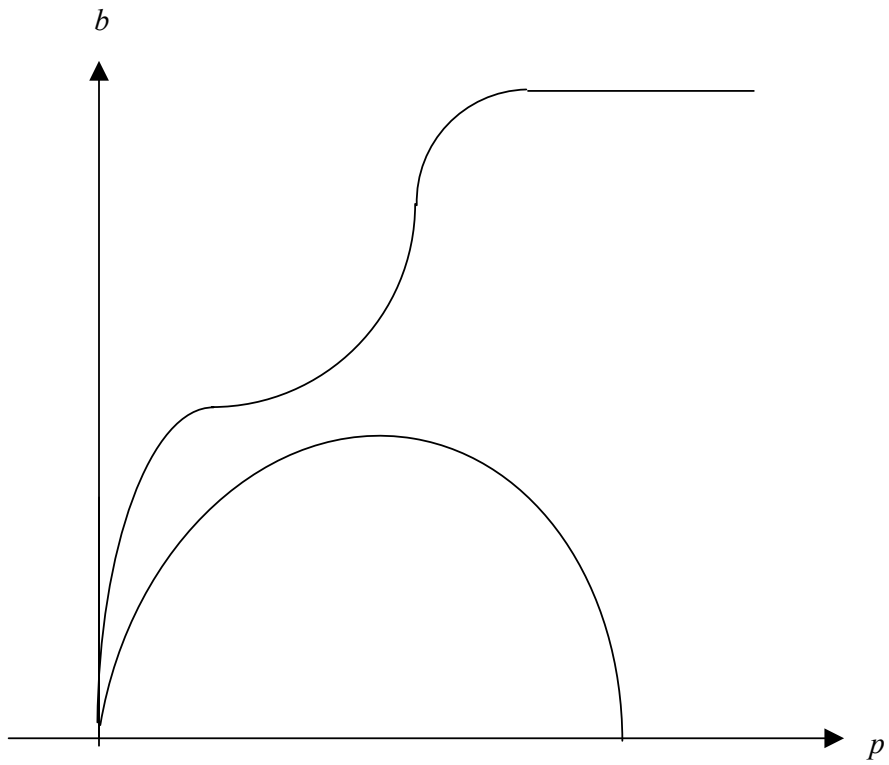Figure 4a: An interior equilibrium at a convex segment of the indifference curve

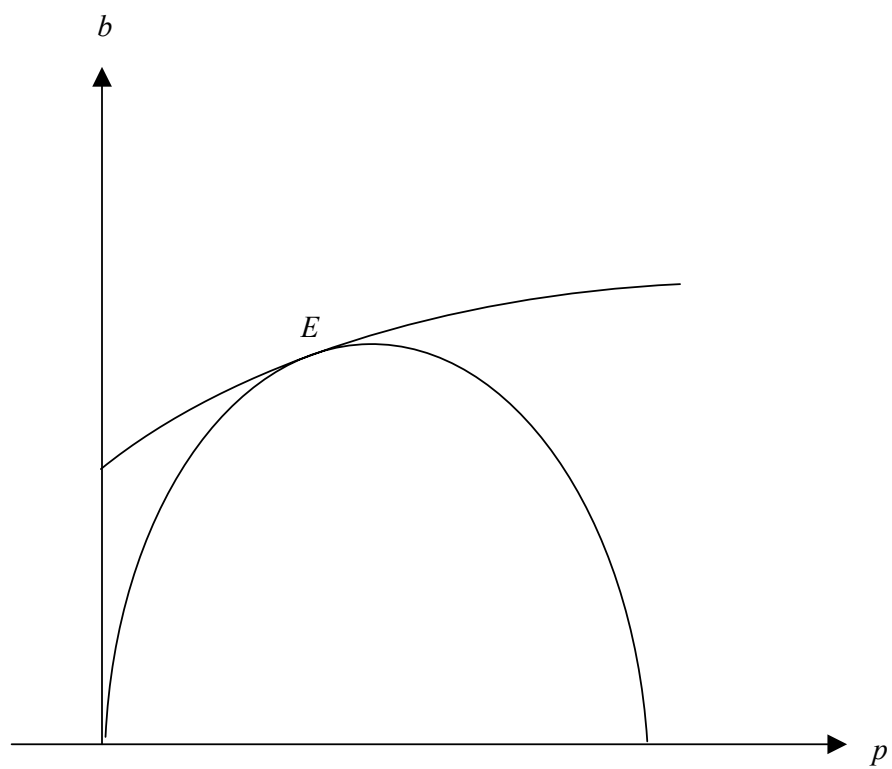Figure 4b: A corner equilibrium at (0, 0)

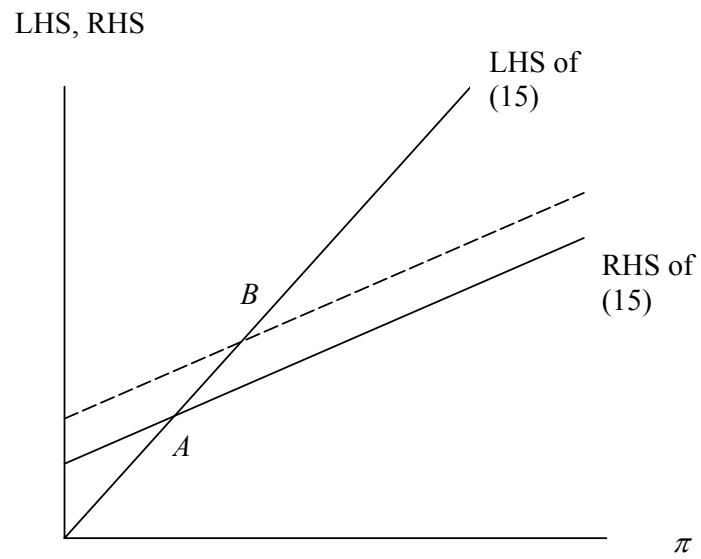Figure 4c: An interior equilibrium with a concave indifference curve

LHS, RHS

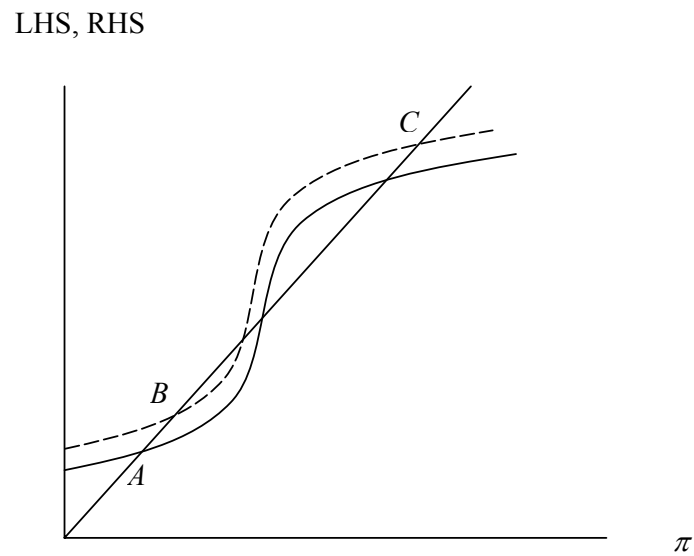LHS of
(15)

B

RHS of
(15)

A

π

Figure 5a: Linear norms

LHS, RHS

C

B

A

π

Figure 5b: Non-linear norms