RESEARCH INSTITUTE OF INDUSTRIAL ECONOMICS

IFN Working Paper No. 763, 2008

A Continuous Model of Income Insurance

Assar Lindbeck and Mats Persson

Research Institute of Industrial Economics
P.O. Box 55665
SE-102 15 Stockholm, Sweden
info@ifn.se
www.ifn.se

2008-08-29

A Continuous Model of Income Insurance*

by

Assar Lindbeck♣ and Mats Persson♠

Abstract:

We develop a simple yet realistic model of income insurance, where the individual's ability and willingness to work is treated as a continuous variable. In this framework, income insurance not only provides income smoothing, it also relieves the individual from particularly burdensome work. As a result, the individual adjusts his labor supply in a continuous fashion to the implicit tax wedge of the insurance system. Moral hazard, in the sense that an individual receives insurance benefits without actually being fully qualified, also becomes a matter of degree. Moreover, our continuous framework makes it easy to analyze both the role of administrative rejection of claims, and the role of social norms, for the utilization of insurance.

Key words: Moral hazard, disability insurance, work absence, administrative rejection, asymmetric information, social norms.

JEL classification: G22, H53, I38, J21,

---

* We are grateful to Mathias Herzing, Harald Lang, Tomas Sjöström, Johan Stennek and Jens Svensson for valuable comments and suggestions on an earlier version of this paper. Assar Lindbeck gratefully acknowledges research support from The Catarina and Sven Hagströmer Foundation.

♣ Institute for International Economic Studies, Stockholm University, and IFN, Stockholm.
E-mail: assar@iies.su.se.

♠ Institute for International Economic Studies, Stockholm University. E-mail: mp@iies.su.se.

1. Introduction

The modern literature on income insurance originates mainly from the seminal work of Rothschild and Stiglitz (1976) and Diamond and Mirrlees (1978), who develop models where an individual suffers an income loss due to an exogenous, binary event. This event may be interpreted as a health shock, which can take one of two values: the individual is either able to work (healthy) or unable to do so (sick), where the latter state implies that the individual will necessarily be absent from work.[1] It is, however, more realistic to treat health as a matter of degree. For instance, Cochrane (1972) has emphasized that health is usually a continuous variable that cannot realistically be depicted in terms of two alternative states (sick and healthy). Moreover, an individual's absence from work may depend on other factors which are not purely medical, but which should realistically also be regarded in terms of degree. Examples include attitudes towards work, leisure, and social interaction at the workplace, as well as aspects of the individual's private life that may influence his ability and willingness to work (such as conflicts within the family).

Against this background, we model the individual's ability and willingness to work as a stochastic variable with a continuous distribution. This allows for a more comprehensive analysis of income insurance than the traditional approach based on a binary distribution of the individual's state of health. This may seem like a minor technical change in the model. However, the continuous approach substantially alters the character of insurance. In the binary model, the purpose of insurance is only to smooth income across two different states of health: when unable to work, the individual is (more or less) compensated for his income loss. In our approach, the purpose of insurance is both to smooth income and to make it feasible for the individual to stay home when work is particularly burdensome. For a given realization of the health variable, the individual has a choice of whether to go to work or to apply for benefits, and his actual choice depends on the generosity of the insurance system.

[1] In the Diamond-Mirrlees tradition, Whinston (1983) has analyzed the case with *ex ante* heterogeneous agents, and Gosolov and Tsyvinski (2006) have elaborated on the multi-period aspects of the model. The Rothschild-Stiglitz approach (as well as that of Wilson, 1977), can also be interpreted as dealing with binary stochastic health shocks. However, instead of studying a government insurance monopoly, these authors highlight the functioning, and weaknesses, of competitive insurance markets when individuals are heterogeneous (*ex ante* and not just *ex post*) in terms of risk. Prescott and Townsend (1984) also study a binary health shock; they show that with a representative agent, the social optimum can be achieved as the outcome of a decentralized, competitive market. (See Rees, 1989, Rees and Wambach, 2008, and Zweifel, 2007, for more recent expositions of the traditional binary approach in insurance theory.).

In addition to its realism, our approach has other advantages. It allows us to take administrative rejection of benefit claims into account in a coherent way. It is also well suited for analyzing the role of social norms in terms of the functioning of income insurance. Generally speaking, our approach enables us to capture a number of real-world phenomena within a simple analytical framework. The model is particularly suited for the analysis of temporary and permanent disability insurance, but after appropriate modifications also of other arrangements of income insurance.

2. The Individual

To build an insurance theory that accommodates a continuous approach to an individual's ability and willingness to work, we write utility as a linearly separable function of consumption *c* and labor supply ℓ: $u = u(c) + \theta \cdot \ell$, where θ is a taste parameter that may be negative as well as positive. We apply the standard assumptions $u'(\cdot) > 0$ and $u''(\cdot) < 0$. We also assume that ℓ can take only the values 0 and 1. Thus utility when working is $u^W = u(c) + \theta$, while utility when absent from work is $u^A = u(c)$.[2] Hence we assume that the consumption utility is the same regardless of whether the individual goes to work or not. The basic reason is that we want to avoid distractions from our ambition to investigate the consequences of a continuous treatment of θ. Moreover, it is not clear whether the marginal utility of consumption should be assumed to be increasing or decreasing in θ. As we proceed, the implications of dropping the assumption of separability will be discussed in several footnotes.

The parameter θ is drawn from an arbitrary probability distribution;[3] it takes negative values when the individual experiences disutility from work and positive values when he enjoys work *per se*. While in the traditional labor supply literature it is usually assumed that the

[2] There is a literature on absence from work that uses a utility function with a continuous index variable reflecting the individual's health status; see, for instance, Barmby *et al*. (1994) and the survey by Brown and Sessions (1996). However, this literature deals with efficiency wages rather than insurance theory. By contrast, Diamond and Sheshinski (1995), when studying the possibility of supplementing old-age pensions with disability pensions, assume a utility function with a continuous health parameter in an optimal insurance setting.

[3] Although we treat θ as a continuous variable in this paper, our formal analysis also covers the case when the distribution of θ is discrete.

individual always experiences disutility from work as compared to leisure (i. e., in the context of our framework $\theta < 0$) this is an unnecessarily narrow view. It is more realistic to assume that individuals sometimes enjoy work more than leisure – partly due to social interaction with coworkers.[4]

While we treat θ as a continuous variable, we follow the tradition in much of the income insurance literature and regard the choice between work and leisure as binary: the individual either works or stays home. Such a treatment of labor supply at the extensive margin not only simplifies the exposition; it is also highly relevant when studying income insurance, which mainly pays benefits to individuals who do not work at all. However, it is straightforward to extend the analysis to part-time work and part-time benefits.

The case without insurance is a useful starting point for our analysis. The individual's utility may now be written $u^W = u(1) + \theta$ when working, with the wage rate normalized to unity. Similarly, utility is $u^A = u(0)$ when absent from work.[5] Here, the "zero" does not necessarily mean that the individual is subject to starvation when not working. He may have other resources than labor income to support himself; these are suppressed in the notation $u(\cdot)$. The cut-off point, at which he is indifferent between work and non-work in a world without insurance, obtained by setting $u^W = u^A$, is given by

$$\theta_0^* \equiv u(0) - u(1). \qquad (1)$$

The individual stays home for all realizations $\theta \leq \theta_0^*$, and he goes to work otherwise.[6]

Let us now introduce insurance into the model. We confine our discussion to a simple type of insurance that schematically reflects income insurance in the real world. More specifically,

---

[4] An alternative to a stochastic taste parameter in the utility function would be a stochastic productivity parameter, expressing the relation between the individual's effort and output (cf. Albanesi and Sleet, 2006).

[5] The stochastic variable θ expresses the discomfort of going to work rather than staying home. Alternatively, we could have introduced two separate stochastic parameters in the utility function: $u^W = u(1) + \theta_1$ and $u^A = u(0) + \theta_2$, respectively. With additive separability, our variable θ could then be interpreted as $\theta_1 - \theta_2$.

[6] With a general utility function $u(c, \theta)$, the cut-off corresponding to equation (1) is given by $u(1, \theta_0^*) = u(0, 0)$. Provided $u(c, \theta)$ is monotone in both arguments, the solution θ_0^* is unique.

we represent an insurance contract by the triplet $(p, b, \hat{\theta})$, where p is the premium paid by the individual, b is the benefit, and $\hat{\theta}$ is the insurer's cut-off, i.e., the value of θ below which the insurer is willing to pay a benefit when the individual does not work. (We use a hat (^) to denote cut-offs that are set by the insurer, and a star (*) for cut-offs chosen by the individual.) Such a contract permits the individual to enjoy utility $u^W = u(1-p) + \theta$ when working. If his realization of θ is smaller than or equal to the cut-off $\hat{\theta}$ specified in the contract, he may instead stay home and enjoy utility $u^A = u(b)$. We assume that the stochastic taste parameter θ is drawn from a general probability distribution $F(\theta)$ with density $f(\theta)$. The probability that the individual will live on benefit now is

$$\pi \equiv \Pr(\theta \leq \hat{\theta}) \equiv F(\hat{\theta}) .$$

We interpret the model as describing the behavior of a large number of *ex ante* identical individuals, with i. i. d. stochastic taste parameters θ drawn from the distribution $F(\theta)$. With this interpretation, individuals differ *ex post*, i.e., after realization of the stochastic taste parameters, and the variable $\pi \equiv F(\hat{\theta})$ denotes the total absence rate in society. The reason for assuming *ex ante* identical individuals is that we want to study issues related to *ex post* moral hazard, rather than problems of adverse selection and cream-skimming, as thoroughly analyzed by Rothschild and Stiglitz (1976).[7]

Whether the insurer can observe the individual realization of θ or not is crucial for the functioning of insurance contracts. We therefore organize the subsequent analysis along the dimension of the insurer's ability to observe θ, starting with the polar cases when θ is either fully observable or entirely non-observable. Subsequently, we consider the more realistic case of *partial* observability.

[7] In an earlier version of this paper (Lindbeck and Persson, 2006) we allowed for *ex ante* different individuals; the difficulties of establishing pooling and separating equilibria in a competitive market were similar to those of Rothschild and Stiglitz (1976).

3. Market Equilibrium Under Full Observability

When the realization of θ is fully observable, we could conceive of very elaborate insurance contracts. For instance, both the premium *p* and the benefit *b* could be made contingent on the realization of θ. Such a contract may then state that if θ takes a particular value θ_i, the individual pays an amount p_i, while if θ takes a value θ_j, the individual receives an amount b_j – regardless of whether he works or not at that particular realization of θ. However, we are not going to study contracts of this general type. The reason is that the only purpose of the present section (on full observability of θ) is to provide a benchmark for the subsequent analysis, which deals with more realistic types of contracts. We therefore confine the analysis in this section to insurance contracts for which the triplet $(p, b, \hat{\theta})$ is independent of each individual's realization of θ, and where the individual pays a premium only when working, and where benefits are received when not working. It can be shown that individuals in this framework will either support themselves from labor earnings or live on benefits. We thus have the following lemma, which turns out to be useful in the subsequent analysis:

Lemma 1: Under full observability, an optimal contract $(p_F, b_F, \hat{\theta}_F)$ implies that no one will choose to stay home without benefits.

Proof: See Appendix 1.

Due to this Lemma, the expression for the individual's expected utility has only two terms:

$$EU \equiv F(\hat{\theta}) \cdot u(b) + \left(1 - F(\hat{\theta})\right) \cdot \left(u(1-p) + E(\theta \,|\, \theta > \hat{\theta})\right), \qquad (2)$$

where the first term represents the utility of those who stay home with benefit *b* and the second term the utility of those who work and earn a net wage 1 – *p.* An optimal contract maximizes the expected utility with respect to $(p, b, \hat{\theta})$, subject to the insurer's budget constraint

$$\left(1 - F(\hat{\theta})\right) \cdot p - F(\hat{\theta}) \cdot b = 0, \qquad (3)$$

and the non-negativity constraints

$$p \geq 0 \qquad \text{and} \qquad b \geq 0\,. \tag{4}$$

The Lagrangean is

$$\begin{aligned} L \equiv\; & F(\hat{\theta}) \cdot u(b) + \left(1 - F(\hat{\theta})\right) \cdot \left(u(1-p) + E(\theta \,\middle|\, \theta > \hat{\theta})\right) + \\ & + \lambda \cdot \left[\left(1 - F(\hat{\theta})\right) \cdot p - F(\hat{\theta}) \cdot b\right] + \mu \cdot p + \nu \cdot b \end{aligned} \tag{5}$$

with the first-order conditions

$$\text{w. r. t. } p\text{:} \quad -\left(1 - F(\hat{\theta})\right) \cdot u'(1-p) + \lambda \cdot \left(1 - F(\hat{\theta})\right) + \mu = 0\,, \tag{6}$$

$$\text{w. r. t. } b\text{:} \quad F(\hat{\theta}) \cdot u'(b) - \lambda \cdot F(\hat{\theta}) + \nu = 0\,, \tag{7}$$

$$\text{w. r. t. } \hat{\theta}\text{:} \quad u(b) - u(1-p) - \hat{\theta} - \lambda \cdot (p + b) = 0\,. \tag{8}$$

From these three conditions, we can derive a number of properties of an optimal insurance system when θ is fully observable by the insurer. Some of the properties are wellknown from the traditional (binary) insurance literature. The most obvious example is that if insurance is desirable at all, full insurance is optimal (abstracting from administrative costs).[8] The intuition is that if an individual's health status is fully observable, moral hazard in the sense that people misrepresent their health condition (θ) cannot emerge. An insurance that permits full income smoothing is therefore optimal – as in the traditional, binary insurance models (provided the health variable enters additively in the utility function).[9]

---

[8] Proof: With an interior solution $p_F > 0, b_F > 0$, the multipliers associated with the non-negativity constraints (4) are zero, and thus (6) and (7) can be written as $u'(1-p) = \lambda$ and $u'(b) = \lambda$, respectively. We therefore have $u'(1-p) = u'(b)$, which implies that $1 - p = b$. *Q. E. D.*

[9] With a non-separable utility function $u(c, \theta)$, full insurance is generally not optimal. The expression corresponding to $u'(1-p) = u'(b)$ then is $E(u_1(1-p, \theta) \,\middle|\, \theta > \hat{\theta}) = u_1(b, 0)$. This expression has a simple intuitive interpretation: the expected marginal utility of income should be the same across the two states of working and not working. It can be satisfied for full, less-than-full and overfull insurance, depending on the cross derivative u_{12} and the distribution of θ.

An attractive property of our model is that it allows a simple explicit solution for the optimal contract $(p_F, b_F, \hat{\theta}_F)$. From (6)-(8) and the budget constraint (3), we obtain

$$p_F = F(\hat{\theta}_F), \tag{9a}$$

$$b_F = 1 - F(\hat{\theta}_F), \tag{9b}$$

$$\hat{\theta}_F = -u'\left(1 - F(\hat{\theta}_F)\right). \tag{9c}$$

This system is recursive. Since the right-hand side of (9c) is monotonically decreasing in $\hat{\theta}_F$, it has a unique solution. Inserting this solution into (9a) and (9b), we obtain closed-form expressions for p_F and b_F.

Such a contract maximizes the individual's expected utility *ex ante*, hence before a specific θ has been realized. However, after the realization of θ, the individual might want to stay home and receive benefits even for some realizations of θ above $\hat{\theta}_F$. Indeed, the individual's subjective cut-off θ^* in the case of insurance is the point at which he is indifferent between working and living on benefits:

$$\theta^* \equiv u(b) - u(1 - p). \tag{10}$$

Since full insurance ($b_F = 1 - p_F$) is optimal under full observability, we conclude that $\theta_F^* \equiv u(b_F) - u(1 - p_F) \equiv 0$. Hence, individuals would like to stay home and receive benefits for all negative realizations of θ. However, since θ is fully observable, they will, in fact, receive benefits only if θ is smaller than or equal to the cut-off $\hat{\theta}_F$, as stipulated in the contract.

Some properties of the optimal insurance contract are illustrated in Figure 1. Since $\hat{\theta}_F < \theta_F^*$, only a fraction of those who would like to stay home and live on benefits are actually allowed to do so (i.e., those for whom $\theta \le \hat{\theta}_F$).[10]

[10] Proof that $\hat{\theta}_F < \theta_F^*$: $\hat{\theta}_F < 0$ by (9c). With full insurance, $\theta_F^* = 0$ by (10). Thus $\hat{\theta}_F < \theta_F^*$.

(Figure 1)

In general, insurance in a model with a continuous taste parameter achieves two goals: income smoothing and relief from particularly burdensome work. The cost is a loss in average production and hence consumption. While with an additively separable utility function $u(c, \theta) = u(c) + \theta$ the first goal is driven to its extreme in the sense that complete income smoothing (full insurance) is optimal,[11] there is a trade-off between the second goal (relief from burdensome work) and average consumption. We state this insight by the following proposition.

Proposition 1: With additive separability, hence $u(c, \theta) = u(c) + \theta$, optimal insurance implies a trade-off between relief from particularly burdensome work, on the one hand, and average consumption on the other hand.

Proof: By Lemma 1, $\hat{\theta}_F > \theta^{**} \equiv u(0) - u(1-p) > u(0) - u(1) \equiv \theta_0^*$. Thus $\hat{\theta}_F > \theta_0^*$ in the case of a continuous distribution of θ, which means that the individual will be absent more often with insurance than without. *Q. E. D.*

As we would expect, insurance thus makes the individual choosier when deciding whether to work or not, and the insurer also allows him to stay home (with benefit) more often. There is a corresponding loss in production (and consumption) as compared to the case with no insurance.[12]

So far, we have discussed a number of properties of an optimal insurance system, provided insurance is desirable in the first place. The next question is whether insurance actually *is* desirable. In traditional analysis, insurance is always desirable if the utility function is concave (abstracting from administrative costs). It turns out that our model has stricter conditions for insurance to be desirable. To see this we define the lower and upper support of the distribution of θ:

[11] As pointed out in footnote 9, this property depends on the assumption of additive separability. In the case of a non-separable utility function $u(c, \ell)$, insurance implies a simultaneous trade-off between income smoothing and pain relief on one hand, and consumption on the other hand.

[12] The size of the loss in production would be expected to depend on how much the individual's productivity is harmed by negative outcomes of θ - an issue that we do not model.

$$\theta_{lower} \equiv \inf(\theta | F(\theta) > 0),$$
$$\theta_{upper} \equiv \sup(\theta | 1 - F(\theta) > 0).$$

We then have

Proposition 2: Assuming a concave consumption utility function, and abstracting from administrative costs, insurance is desirable under full observability if and only if the distribution of θ is such that (*i*) $\theta_{upper} > u(0) - u(1) \equiv \theta_0^*$ and (*ii*) $\theta_{lower} < -u'(1)$.

Proof: See Appendix 2.

Whether insurance is desirable or not under full observability thus depends on whether the distribution of θ has positive density (*i*) above $u(0) - u(1)$ and (*ii*) below $-u'(1)$, where $u(0) - u(1) < -u'(1)$ by concavity. Proposition (2) can be equivalently phrased in the following way: Insurance is never desirable if either $\theta_{upper} < u(0) - u(1)$ or $\theta_{lower} > -u'(1)$, regardless of whether consumption utility is concave or not. The intuition is straightforward for our non-traditional conclusion that concavity of the consumption utility function is not sufficient for insurance to be desirable. Condition (*i*) says that insurance can be financed only if θ can take sufficiently high values for the individual to be willing to work (and pay an insurance premium) at least some time. Condition (*ii*) says that it is worthwhile to pay an insurance premium, and hence to abstain from some consumption, only if θ can take sufficiently negative values. Intuitively speaking, insurance is desirable only if the reduction in pain from burdensome work (θ) more than compensates for the reduced consumption when paying a premium with the utility cost $u'(1)$.

A graphical representation of the optimal interior solution ($p_F > 0, b_F > 0$) may be instructive. First, the insurer's budget constraint (3) can be written

$$b = \frac{1 - F(\hat{\theta}_F)}{F(\hat{\theta}_F)} \cdot p, \qquad (11)$$

depicted by the straight line *OF* in Figure 2 for a given cut-off $\hat{\theta}_F$.

(Figure 2)

Second, the slope of an indifference curve is

$$\left.\frac{dp}{db}\right|_{EU_F=const.} = \frac{1-F(\hat{\theta}_F)}{F(\hat{\theta}_F)} \cdot \frac{u'(1-p)}{u'(b)}. \qquad (12)$$

Clearly, the curve is upward-sloping and convex, as illustrated by the indifference curves in Figure 2. The obvious intuition is that the individual is willing to pay a higher premium only if he receives a higher benefit. Moreover, the required benefit increases progressively. Assuming that the conditions in Proposition 2 are satisfied, the optimal insurance contract is represented by point *F* in the figure, located on the "full insurance line" $b = 1 - p$.

In conclusion, income insurance provides two advantages: income smoothing and "pain avoidance". The latter, which is not dealt with in the traditional literature, means that the individual can afford to stay home from work during periods when working is particularly painful (formally, when $\theta \le \hat{\theta}_F$). The cost of enjoying these two benefits is a reduction in average production and hence consumption, as compared to the case with no insurance.

4. Market Equilibrium Under Non-Observability

We now turn to the opposite extreme, where the realization of the taste parameter is completely non-observable for the insurer. In this case, the insurer's cut-off in the triplet $(p, b, \hat{\theta})$ must be incentive compatible, i.e.,

$$\hat{\theta} = \theta^* = u(b) - u(1-p). \qquad (13)$$

The reason why the contract has to be incentive-compatible is straightforward. If $\theta^* > \hat{\theta}$, the individual would have an incentive to misrepresent his realization of θ, hence mimicking individuals with low realizations. If instead $\theta^* < \hat{\theta}$, the latter would be irrelevant for the individual's behaviour. In principle, our restriction (13) corresponds to the "moral hazard constraint" in Diamond and Mirrlees (1978).[13]

From (13) we immediately see that $\theta^* > \theta_0^*$ for all $p > 0, b > 0$. In other words, Proposition 1 holds also when θ is not observable. Introducing insurance will necessarily impose a cost in the form of lower aggregate labor supply and production. Optimal insurance will trade off this loss against the gain in terms of income smoothing and relief from burdensome work (pain avoidance).

The optimal insurance contract under non-observability is obtained by maximizing expected utility (2) subject to both the budget constraint (3) and the incentive-compatibility constraint (13). As a result, expected utility in the optimal contract is lower than (or, as a special case, the same as) under full observability: $EU_N \leq EU_F$. The first-order conditions with respect to *p* and *b* are (after some rearranging)[14]

$$\text{w. r. t. } p\text{: } -f(\theta^*)\cdot u'(1-p)\cdot\lambda\cdot(p+b) = \left(1-F(\theta^*)\right)\cdot\left(u'(1-p)-\lambda\right)-\mu, \qquad (14)$$

$$\text{w. r. t. } b\text{: } -f(\theta^*)\cdot u'(b)\cdot\lambda\cdot(p+b) = F(\theta^*)\cdot\left(\lambda-u'(b)\right)-\nu, \qquad (15)$$

where θ^* is shorthand for $u(b)-u(1-p)$. It can be shown that, just as in traditional theory of income insurance, less than full insurance is optimal under non-observability: $b_N < 1-p_N$ (see Appendix 3 for proof). However, our model also generates some new insights. For instance, we have

[13] We may alternatively write the same contract as a duplet (*p*, *b*) and drop the incentive-compatibility constraint (13), replacing $\hat{\theta}$ in equations (2) and (3) by θ^*.

[14] Here we have substituted (13) into (2) and (3), instead of entering equation (13) as a separate constraint.

Proposition 3: Assuming a concave consumption utility function, and abstracting from administrative costs,

(*i*) a necessary condition for insurance to be desirable under non-observability is that $\theta_{upper} > u(0) - u(1) \equiv \theta_0^*$;

(*ii*) given (*i*), a sufficient condition for insurance to be desirable under non-observability is that $\theta_{lower} < u(0) - u(1) \equiv \theta_0^*$.

Proof: See Appendix 4.

Thus, while both Proposition 3 and Proposition 2 convey the same basic message, namely that concavity is not sufficient for insurance to be desirable, they differ in other respects. [15] Conditions (*i*) are identical in the two Propositions, but conditions (*ii*) imply different critical values for θ_{lower}. Morever, the two Propositions differ as to whether the conditions are necessary or sufficient.

In Section 3 we pointed out that the consequences for labor supply differ between the binary and the continuous models in the case when θ is fully observable: introducing optimal insurance will never cause a fall in labor supply in the binary model, but may cause such a fall in our model. These properties carry over to the non-observability case when individuals are identical *ex ante*:

Proposition 4:

Assume that insurance is desirable by Proposition 3, i.e., that $\theta_{lower} < \theta_0^* < \theta_{upper}$. It then holds that

(*i*) if the distribution function $F(\theta)$ is such that θ can take only two values, θ_1 $(= \theta_{lower})$ and θ_2 $(= \theta_{upper})$, the introduction of optimal insurance will not cause any production loss.

[15] Why, then, is concavity sufficient for insurance to be desirable in the Diamond-Mirrlees binary model? The reason is simply that in such a model, the individual can only experience two health states: $\theta_1 < \theta_2$. Furthermore, θ_1 has to be so negative that the individual is simply unable to work, which means that $\theta_1 < \theta_0^*$ (as a special case, $\theta_1 \to -\infty$). Thus the conditions in Proposition 3 are, in fact, automatically satisfied in the binary model, and insurance will always be desired provided $u(c)$ is concave.

(*ii*) By contrast, if the distribution $F(\theta)$ is continuous and connected, the introduction of optimal insurance will always cause a production loss (provided insurance is desirable).[16]

Proof:

(*i*) If insurance is warranted by Proposition 3, it must be incentive-compatible. This means that there is no production loss: people work when $\theta = \theta_2$ and stay home when $\theta = \theta_1$, regardless of whether there is insurance or not. *Q. E. D.*

(*ii*) In the case of a distribution that is continuous and connected the optimal cut-off with insurance is $\theta_N^* \equiv u(b_N) - u(1 - p_N)$. By definition, this is larger than $\theta_0^* \equiv u(0) - u(1)$ for all $p_N, b_N > 0$. Thus, the introduction of insurance will always cause a production loss in this case. *Q. E. D.*

The main differences between our approach and the traditional binary approach to income insurance may be summarized as follows. As a result of our treatment of the individual's willingness and ability to work as a continuous stochastic variable, θ, the introduction of income insurance induces the individual to reduce his labor supply. The reason is the emergence of an implicit tax wedge, which is the sum of the insurance premium and the benefit rate, hence $p + b$.[17] (Indeed, this incentive effect also emerges if the realization of θ is fully observable for the insurer.)

By contrast, in the static version of the Diamond-Mirrlees' model, where individuals (as in our model) are *ex ante* identical, labor supply is unaffected by the introduction of optimal insurance. The explanation is that the "moral hazard constraint", which is basically an incentive-compatibility constraint, will bind in the case of an optimal social-insurance contract: sick individuals cannot work under any circumstances, and the optimal insurance arrangement implies that healthy individuals will not gain anything by pretending to be unable to work. Indeed, in a binary model with *ex ante* identical individuals, aggregate labor supply

[16] While (*ii*) holds for all continuous and connected distributions, it is easy to show that an optimal insurance contract may cause production losses also in the case of discrete distributions, if there are more than two outcomes – depending on how the realizations are located relative to the cut-off points θ_0^* and θ_N^*.

[17] The implicit tax wedge may be derived from the difference in disposable labor earnings and social-insurance benefits: $w(1-t-p) - wb = w\left[1-(t+p+b)\right]$, where w is labor income, t the tax rate for labor income (outside the social insurance system), p the compulsory insurance fee and b the (after-tax) replacement rate. While the implicit tax wedge in the insurance system is $p + b$, the total tax wedge is $t + p + b$. For many European countries, realistic figures are t=0.25, p=0.10, and b=0.5, which altogether add up to 0.85.

will in optimum always be determined by the number of individuals who are objectively able to work.

In Whinston's (1983) disaggregated version of the Diamond-Mirrlees model, with a number of *ex ante* different "types" of individuals, a utilitarian optimum requires redistributions of income in favour of types of individuals with particularly large probabilities of becoming sick (unable to work). In this disaggregated framework, it may (for certain parameter constellations) be optimal to lift the moral hazard constraint for some of these types.[18]

Hence, the introduction of an optimal insurance arrangement may result in a fall in aggregate labor supply in models with both a continuous and a binary treatment of the individual's health status. However, in the latter type of model, this can occur only if it includes *ex ante* different types of individuals in terms of the probability of becoming unable to work. Moreover, the mechanisms are fundamentally different in these two kinds of models. In a model with a continuous treatment of the individual's willingness and ability to work, people will adjust their behavior to the implicit tax wedges of the insurance system, in the sense that they will abstain from work more often than when there is no insurance. No such adjustments take place in binary models, since people are either fully able or wholly unable to work. In such models, labor supply will not be affected unless the authorities remove the moral hazard constraint for some types of individuals, hence allowing such types to stay home regardless of whether they are able to work or not. Thus, while in binary models with heterogeneous individuals, income insurance implies that certain groups of individuals may withdraw completely from the labor market, in our continuous model the incentives to supply labor are reduced for everyone [19] We believe that such a "gradual" treatment of individuals' willingness and ability to work, and hence a gradual adjustment of labor supply to income insurance, is much more realistic than the notion that the insurer lifts the moral hazard constraint for some particular *ex ante* type of individual, who then drops out of the labor market entirely.

---

[18] In a continuous-time version of their model, where each age group is regarded as a separate " type", Diamond and Mirrlees reached a similar conclusion as Whinston. If the probability of becoming unable to work increases by age, an optimal insurance policy in a continuous-time model with a binary health variable may be to abolish the moral hazard constraint for all individuals above a certain age. As pointed out by Diamond and Mirrlees, this is equivalent to introducing a pension system that permits both unhealthy and healthy elderly to live on benefits.

[19] A more technical way of stating the difference between the models is to point out that the consequences for aggregate labor supply can occur in the binary model only if we assume *ex ante* different types. In our continuous model, it is sufficient that individuals are different *ex post* as a result of different realizations of the stochastic parameter θ.

Let us end this section with a geometrical representation of the optimum insurance contract under non-observability. By differentiating (2) with respect to *p* and *b*, taking (13) into account, and making appropriate substitutions, the slope of the indifference curve in the (*p*, *b*) plane becomes

$$\left.\frac{db}{dp}\right|_{EU_N=const.} = \frac{1-F(\theta^*)}{F(\theta^*)} \cdot \frac{u'(1-p)}{u'(b)}. \qquad (16)$$

Although it looks similar to (12) (with $\hat{\theta}_F$ replaced by θ^*), equation (16) describes a different function in the (*p*, *b*) plane than (12) since θ^* is endogenous. Since the marginal utility of consumption is always positive, the indifference curve is again upward-sloping in the (*p*, *b*) plane for all $\pi \in (0,1)$.[20] While the slope of the indifference curve is thus unambiguous, its curvature is not. Indeed, the indifference curves may have both concave and convex segments (although we have chosen to depict a well-behaved, convex curve in Figure 3).[21]

(Figure 3)

The budget constraint now looks as follows:

$$b = \frac{1-F(\theta^*)}{F(\theta^*)} \cdot p, \qquad (17)$$

Rather than a straight line as in Figure 2, the budget constraint (17) forms a non-linear curve in the (*p*, *b*) plane. Such a curve necessarily passes through the origin (*p* = 0, *b* = 0), since that point trivially satisfies (17). It can be shown that for a wide class of distribution functions $F(\theta)$, the budget constraint (17) generates a well-behaved "Laffer-type" curve in the (*p*, *b*) plane; see Figure 3. For this class of distribution functions, we can state

Proposition 5: For all distributions for which $F(\theta)/(1-F(\theta))$ is convex in θ, the insurer's budget constraint under non-observability is a single-peaked curve in the (*p*, *b*) plane.

[20] This also holds for the case of a non-separable utility function.
[21] This property contrasts with the strict convexity of the indifference curves in the full-information case (12). The observation that indifference curves in insurance models may contain both concave and convex segments has been made earlier in a different analytical framework; cf. Stiglitz (1983) and Arnott (1992).

Proof: Expression (17) can be written in the form $b \cdot F(\theta^*)/(1-F(\theta^*)) = p$. The right-hand side of this equation is a linear, positively sloped function of *p*. If the left-hand side is a convex, positively sloped function of *p*, the two functions can intersect at most twice. Thus, for a given *b*, the equation can have at most two roots *p*. A sufficient condition for this to hold is that θ^* is an increasing, convex function of *p* (which obviously is the case) and that $F(\theta^*)/(1-F(\theta^*))$ is an increasing, convex function of θ^*. *Q. E. D.*

Is it then reasonable to assume that $F(\theta^*)/(1-F(\theta^*))$ is convex? In fact, this ratio, which in the statistical literature is often called the odds function (the logarithm of which is the logit function), is convex for a wide class of distributions. It is easy to show this analytically for the rectangular distribution, and numerically for a number of other distributions, such as the normal, log-normal, Weibull etc. distributions.[22] Thus the zero-profit constraint normally forms a well-behaved Laffer curve as in Figure 3.[23]

The slope of the Laffer curve, obtained by differentiating (17), is:[24]

$$\left.\frac{db}{dp}\right|_{\substack{zero\\profit}} = \frac{b - f(\theta^*)u'(1-p)(p+b)^2}{p + f(\theta^*)u'(b)(p+b)^2}. \qquad (18)$$

In the case of an interior solution, (p_N, b_N) is a tangency point like *N* in Figure 3 (where we have depicted the indifference curves as convex). Since less than full insurance is optimal, point *N* is located beneath the full insurance line.

As emphasized earlier, the equilibrium point is incentive compatible: confronted with the insurance policy (p_N, b_N), the individual has no reason to pretend that his realization θ is different from its true value. If there is no administrative rejection of claims, an incentive-

[22] It does not, however, hold for all distributions. An example where it does not hold is Student's *t* distribution with less than one "degree of freedom", i.e., with thick tails and an infinite mean.

[23] In the case of a distribution function $F(\theta)$ where the support has a limited domain (for instance, with a rectangular distribution), the Laffer curve intersects the horizontal axes at a finite value of $p > 0$, as illustrated in Figure 3. In the case of an unlimited domain (for instance, with the normal distribution), the curve instead approaches the horizontal axes asymptotically.

[24] It is easy to show that if we start at the origin and move to the right along the Laffer curve, the absence rate π_N increases.

compatible insurance contract means that the individual himself can choose whether to live on work or on benefits. This may seem to be quite an odd case, but it is a logical consequence of the assumption that θ is unobservable for the insurer. Indeed, until recently, in some countries sick-pay insurance has functioned in approximately this way, since the authorities have been reluctant to reject individual claims. However, the analysis becomes more realistic if we assume that the insurer will, in fact, reject some benefit claims. We now turn to this case.

5. Non-Observability with Administrative Rejection

With administrative rejection of claims, the insurance contract would be characterized by a quadruplet $(p, b, \hat{\theta} = \theta^*, q)$, rather than the triplet $(p, b, \hat{\theta} = \theta^*)$, where q is the probability of rejection.[25] Can such a contract, with $0 < q < 1$, constitute a welfare improvement?

If the insurer cannot observe θ, rejection of claims has to be purely random. At first glace, it may seem impossible that welfare (expected utility) could increase if such randomness is introduced, since a new type of income risk would then emerge. However, on further reflection, the issue turns out to be rather complicated. The reason is that administrative rejection allows for a more generous insurance contract for those whose claims are accepted. To analyze this issue, we assume that an individual whose claim has been rejected can choose either to continue working or to stay home without receiving benefits. The individual is thus confronted with the following situation:

$$\begin{aligned} &\textit{If} \quad \theta > \theta^* : \textit{utility} \quad \textit{is} \quad u(1-p) + \theta. \\ &\textit{If} \quad \theta \le \theta^* : \textit{utility} \quad \textit{is} \quad \begin{cases} u(b) \quad \textit{with} \quad \textit{prob.} \ (1-q) \\ \max \left\{ u(0),\ u(1-p) + \theta \right\} \quad \textit{with} \quad \textit{prob.} \ q, \end{cases} \end{aligned}$$

where the insurer's cut-off must be incentive-compatible: $\hat{\theta} = \theta^* = u(b) - u(1-p)$.

Individuals with a realization $\theta > \theta^*$ will not apply for benefits, but instead go to work. By contrast, all individuals with a realization $\theta \le \theta^*$ will apply for benefits, and they will have to

[25] Such lotteries have been extensively discussed in the literature on mechanism design; we return to that issue in Section 6.

participate in a kind of lottery where the probability of winning is $1 - q$. If lucky, they enjoy utility $u(b)$; otherwise they enjoy utility $\max\{u(0), u(1-p)+\theta)\}$. Total absence now consists of those whose claims were accepted *plus* those who chose to stay home even though their claims were rejected:

$$\pi = (1-q)\cdot F(\theta^{*}) + q\cdot F(\theta^{**}), \tag{19}$$

where θ^{**} is the cut-off at which the individual is indifferent between staying home with no benefits and working: $\theta^{**} \equiv u(0) - u(1-p)$. The welfare consequences of such insurance arrangements may be expressed as follows.

Proposition 6: If, after a rejection, the individual can return to work, a positive q will increase expected utility for certain parameter constellations.[26]

Proof: See Appendix 5.

Naturally the introduction of a rejection rate q has different welfare consequences for different realizations of θ and different outcomes of the "lottery". However, the consequences also depend on the adjustments of p and b made by the insurer to maintain a balanced budget. Clearly, those with a bad realization of θ and a bad outcome in the lottery are losers. The gainers are either those who have chosen to work (provided there will be a reduction in p), or those who have applied for, and received, benefits (provided there will be an increase in b). Due to these complications, the net effect on expected utility cannot, in general, be signed analytically.

Owing to the ambiguous analytical consequences for expected utility, we simulated the model numerically. For this purpose, we assumed a utility function with constant relative risk aversion, i.e., $u(x) = (x^{1-\gamma} - 1)/(1-\gamma)$. For $u(0)$ to be defined, we introduced an exogenous non-wage income, k; this also has the advantage that we can study the consequences of variations in non-wage income. Consumption utility then is $u(1-p+k)$ when working,

[26] An alternative assumption could be that the individual, after a rejection of his claim, is not able to return to work and thus has to be satisfied with utility $u(0)$. It is easy to show that in such a case, the introduction of a positive rejection rate cannot increase expected utility. We consider such a setup as less realistic than the one where the individual *can* return to work.

$u(b+k)$ when absent from work and receiving benefits, and $u(k)$ when absent without benefits. For a normal distribution $\theta \sim N(m, \sigma)$ of the taste parameter, the results of the simulations are reported in Figure 4. We chose the parameter values $\gamma = 4$, $m = 0$, and $\sigma = 2$, but the results are qualitatively similar for a large set of values.

(Figure 4)

Our simulations show that it is possible to find plausible parameter configurations for which a positive random rejection rate is optimal. By "plausible" we mean values that are of an order of magnitude similar to those observed in the real world.[27]

In this context, how is the optimal rejection rate related to the level of non-labor income *k*? Intuitively, the optimal *q* might be expected to rise with *k*, since a relatively large *k* makes rejection of claims less harmful for the individual. However, a large *k* also means that the individual has less need for the higher *b* that could be financed as a result of introducing a rejection rate; this mechanism tends to make the relation between *k* and *q* negative. The net effect of these counteracting forces is rather complex. Somewhat surprisingly, in all our simulations, the parameter configurations used in Figure 4 give rise to a monotone negative relation between *k* and *q*.

Summing up, the introduction of random administrative rejection into the model has two counteracting consequences for expected utility. On one hand, it is advantageous to have access to an additional policy instrument. On the other hand, the specific policy tool *q* is in conflict with the basic purpose of insurance, since it introduces a new type of income risk for the individual. It is therefore not surprising that the consequences for expected utility are ambiguous.

6. Partial Observability With Administrative Rejection

[27] For the simulations in Figure 4, the absence rate π_q varies between 0.2 and 0.25. This is a realistic figure for many European countries, considering that it includes both sick-pay insurance and disability pensions.

So far, we have dealt with the polar cases when θ is either fully observable or wholly non-observable for the insurer. The most realistic assumption is that θ is *imperfectly* observable. We model this case by assuming that an outsider (the insurer and/or a legal authority) can observe only a distorted signal $s \equiv \theta + \varepsilon$ rather than θ itself. For simplicity, we further assume that the stochastic variable ε is independent of θ, and that it has the cumulative distribution function $G(\varepsilon)$. Clearly, this case nests the two special cases discussed earlier, since a distribution with $\text{var}(\varepsilon) = 0$ implies full observability (Section 3), while $\text{var}(\varepsilon) \to \infty$ implies non-observability (Sections 4-5).

While rejection has to be completely random in the case of non-observability, the insurer can make the rejection rate conditional on the observed signal s when θ is partially observable. As before, the individual applies for sick-leave benefits iff $\theta \le \theta^*$, where, again, $\theta^* \equiv u(b) - u(1-p)$. The insurer proclaims a cut-off $\hat{\theta}$ and accepts the benefit claim iff $s \le \hat{\theta}$. The probability of rejection then is

$$q \equiv \Pr(s > \hat{\theta}) \equiv \Pr(\varepsilon > \hat{\theta} - \theta) \equiv 1 - G(\hat{\theta} - \theta) \equiv q(\hat{\theta} - \theta) . \qquad (20)$$

Thus, q now becomes endogenous and a function of θ. Since all distribution functions are non-decreasing, it follows that q is also a non-decreasing function of the true θ: $\partial q / \partial \theta \ge 0$. This property has an intuitive appeal; the claim of an individual with severe health problems (i.e., a very low θ) is less likely to be rejected than a claim from someone who is more healthy.[28]

With partial observability, the insurance contract may be written $(p, b, \hat{\theta}, q = q(\hat{\theta} - \theta))$. The expression for total absence from work in society is now a generalization of (19):

$$\pi_p = \int_{-\infty}^{\theta^*} \left(1 - q(\hat{\theta} - \theta)\right) f(\theta) d\theta + \int_{-\infty}^{\theta^{**}} q(\hat{\theta} - \theta) f(\theta) d\theta \qquad (21)$$

where, as before, $\theta^{**} \equiv u(0) - u(1-p)$.

[28] This property is consistent with Diamond's and Sheshinski's (1995) assumption that the probability of rejection of a claim for disability pension (which is higher than the early retirement pension) is an increasing function of the individual's true health status.

An important distinction between a constant rejection rate *q* (independent of θ), as discussed in Section 5, and an endogenous one is that the distribution of absence across individuals will differ. Since an endogenous *q* implies that individuals with very low realizations of θ are less likely than others to be rejected, there will be a more efficient allocation between work and non-work among claimants than if *q* were exogenous. This also means that the income distribution will be more favorable for individuals with a relatively unfavorable outcome of the θ variable.

As in the case of a constant *q*, we are not able to determine analytically whether the introduction of a rejection rate is desirable or not (see the Lagrangean in Appendix 4). We therefore, again, carried out a numerical simulation with the same utility and distribution functions as in Figure 4, but now with *q* defined by the function (20), with the disturbance $\varepsilon \sim N(0, 2)$; see Figure 5.

(Figure 5)

As in the case with an exogenous *q* (Figure 4), it is easy to find parameter configurations that yield realistic values of *p* and *b*. For most values of the non-wage income *k*, we also get values of the insurer's cut-off point $\hat{\theta}$ that yield an endogenous rejection rate $q = q(\hat{\theta} - \theta) > 0$. Since a higher $\hat{\theta}$ implies a lower rejection rate, *q* falls with *k*, just as in the case of an exogenous *q*, illustrated in Figure 4. (With a normal distribution for the disturbance ε, the rejection rate $q \to 0$ as $\hat{\theta} \to \infty$). We may note that θ^* may be smaller or larger than $\hat{\theta}$, depending on *k*.

Exogenous and endogenous rejection of benefit claims may be regarded as two alternative types of "lotteries" among individuals who claim benefits. We could think of more elaborate lotteries. For instance, the lottery may not only determine whether the individual should receive a benefit, but also how large the benefit will be; the individual may receive b_1 with probability q_1, b_2 with probability q_2, etc. Although such elaborate lotteries may be worth further examination, we confine our discussion to the simple case of only two outcomes:

either rejection or approval of the claims.[29] In fact, this simple type of lottery seems to reflect the way sick-pay insurance and disability pensions are usually constructed in the real world.[30]

The variant of our model discussed in this section, with partial observability and an endogenous rejection rate, is much more realistic than the cases of full observability and non-observability discussed in earlier sections. Clearly, in the real world, an individual's health condition *is* partially observable, and some claims *are* rejected on the basis of imperfect signals.

7. Tax Wedges and Moral Hazard

We have emphasized that income insurance in our analytical framework not only provides income smoothing, but also relief from work when it would be particularly burdensome. These advantages come at a cost in the form of two disincentive problems. One is a direct consequence of the unavoidable tax wedge, which emerge regardless of whether the individual's willingness and ability to work are observable or not for the insurer. Indeed, a reduction in labor supply is part of the *purpose* of income insurance in our framework, namely to make it more feasible for the individual to finance absence from work during periods when work is associated with severe discomfort.

The other form of a disincentive effect on work is moral hazard. Clearly, there is no moral hazard problem at all in the case of full observability. Although there is such a problem if θ is not fully observable, it takes different forms in the cases of non-observability and partial observability. With *non-observability*, the insurer has to offer an incentive-compatible

[29] Presumably, the more elaborate the lottery, the more problematic it may be from the point of view of legitimacy.

[30] Prescott and Townsend (1984) have argued that quite another type of lottery is optimal in income insurance. More specifically, every individual who applies for benefits will receive them, but the applicant will be exposed to a lottery of whether he has to work or not. With probability ρ he has to stay home, thereby enjoying utility $u(b)$, while with probability $(1-\rho)$ he has to work, although only receiving *b*, thus enjoying utility $u(b)+\theta$. A lottery of this type cannot, however, be implemented in the real world. In a competitive labor market without slavery-like contracts, an insurer can hardly force an individual with an "unlucky" outcome of the lottery to work at the remuneration *b* (that is below what he can obtain in the market, 1 – *p*), or to stay home without benefits. And even if slavery-like contracts *were* allowed, such contracts cannot be implemented in a modern economy, where the production process is complex (often with strong complementarities among individual workers), and where the possibilities for shirking are abundant.

contract $(p_N, b_N, \hat{\theta}_N = \theta_N^*)$ and the individual wants to work if the realization of θ is greater than the cut-off $\theta_N^* = u(b_N) - u(1 - p_N)$. Clearly, with an optimal contract, the problem of moral hazard in the form of misrepresentation of θ is solved in this case.[31] However, the solution of the problem comes at a cost in the form of less than full income smoothing. Thus, such a contract is second-best in the special sense that a better contract could have been achieved if θ had been observable. With *a rejection rate q*, the optimal contract $(p_q, b_q, \hat{\theta}_q = \theta_q^*, q)$ also precludes misrepresentation of θ (since $\hat{\theta}_q = \theta_q^*$). Although a positive rejection rate might increase expected utility, some "type 2 errors" (when eligible persons are denied benefits) are unavoidable.

Under *partial observability*, with the contract $(p_p, b_p, \hat{\theta}_p, q = q(\hat{\theta}_p - \theta))$, the insurer's cut-off $\hat{\theta}_p$ becomes an operative parameter, by contrast to the case of non-observability. The rejection rate $q(\hat{\theta}_p - \theta)$, given by equation (20), implies a better targeting of benefits than the exogenous rejection rate q in the case of non-observability. As with non-observability, the tax wedge is $p_p + b_p$ when the individual decides whether to claim benefits or not, while it is p_p for an individual whose claim has been rejected. However, by contrast to the case of non-observability, misrepresentation is not eliminated by the optimal contract. Everyone with a realization of θ such that $\theta \le \theta_p^*$ will apply for benefits, and a fraction $1 - q(\hat{\theta}_p - \theta)$ of individuals in the interval $\hat{\theta}_p < \theta \le \theta_p^*$ will actually receive benefits, even though their true θ is larger than the insurer's cut-off $\hat{\theta}_p$. This may be labeled "type 1 errors" in the sense that these individuals will receive benefits because they appear to be more sick than they actually are. Thus, there will be some moral hazard also with an optimal insurance contract, in the form of type 1 errors. There will also be "type 2 errors" since a fraction $q(\hat{\theta}_p - \theta)$ of individuals with realizations $\theta \le \hat{\theta}_p$ will be denied benefits although they are actually qualified.

[31] In the literature, moral hazard is often defined as "hidden action". In our model, the individual's action, i.e., going to work or not, is perfectly observable by the insurer, while the realization of the individual's state θ may not be observable. Hence, since the connotation of moral hazard as "hidden action" is too narrow, we suggest defining *ex post* moral hazard as "hidden action or hidden state".

8. Social Norms

So far, we have analyzed how traditional economic factors, such as prices (p and b) and rationing (administrative rejection) affect the utilization of income insurance. However, the functioning of such insurance also depends on social norms concerning the utilization of the benefit system. Our model turns out to be well suited for incorporating such considerations into the analysis. To clarify this issue in the simplest possible way, we return to the basic model of non-observability (dropping the subscript "N"), without a rejection rate q.

A simple way of incorporating social norms into our model is to add a "stigmatization variable" $\phi \geq 0$ to the individual's utility when he is absent from work: $u^a \equiv u(b) - \phi$. (We follow the formalization of the role of social norms for benefit dependency in Lindbeck, Nyberg and Weibull, 1999.[32]) One possible specification of this variable is to regard the norm as a constant: $\phi \equiv \overline{\phi}$, i. e., as *exogenous* (for instance, inherited from the past). Another possibility is to treat norms as *endogenous*, with the strength of the norm depending on the number of individuals who live on benefits: when a large number of individuals are absent from work, absence is likely to be more legitimate than if only a few individuals are absent. Hence, when the norm is endogenous, we have $\phi \equiv \phi(\overline{\pi})$ with $\partial\phi / \partial\overline{\pi} < 0$, where $\overline{\pi}$ stands for the average absence rate in society.

The individual's cut-off point in the presence of norms, θ^{*n}, is the value of θ for which $u(1-p) + \theta = u(b) - \phi$:

$$\begin{aligned} \theta^{*n} &= u(b) - u(1-p) - \phi \\ &= \theta^* - \phi, \end{aligned} \tag{22}$$

where θ^* is the same cut-off point as in a model without social norms (2). From (22) it follows that both exogenous and endogenous social norms reduce the cut-off point: $\theta^{*n} \leq \theta^*$. Since the individual chooses to stay home whenever $\theta \leq \theta^{*n}$, his absence rate with norms is

[32] See Moffitt (1983) for an early analysis of stigmatization from living on welfare payments. Brock and Durlauf (2001) give a systematic discussion of alternative ways of specifying various types of social interaction among individuals, including the role of social norms.

$\pi^n = F(\theta^* - \phi)$, hence lower than without norms. In the case of exogenous norms, with $\phi = \bar{\phi}$, this expression yields the closed-form solution

$$\pi^n_{ex} = F(\theta^* - \bar{\phi}). \tag{23}$$

In the case of endogenous norms, we instead have $\pi^n_{end} = F\left(\theta^* - \phi(\bar{\pi})\right)$. Since the average absence rate $\bar{\pi}$ is the same as the absence rate π^n_{end} of the representative individual, we may write

$$\pi^n_{end} = F\left(\theta^* - \phi(\pi^n_{end})\right). \tag{24}$$

The right-hand side of (24) is increasing in π^n_{end} and may be non-linear. The equation may therefore have multiple solutions, as is often the case in models of social interaction. Hence there may be several alternative absence rates (for a given insurance system) in a society with endogenous social norms.

We achieve a particularly simple analysis by looking at a linear version of the model, assuming that θ is uniformly distributed on $\left[\bar{\theta} - s, \bar{\theta} + s\right]$. For the case of endogenous norms, we further assume the linear stigmatization function $\phi(\pi^n_{end}) \equiv \gamma \cdot (1 - \pi^n_{end})$, where γ is a positive constant. With these functional forms, we obtain the following simple expressions for the absence rate in the cases of no norms, exogenous norms, and endogenous norms, respectively:

$$\begin{aligned}
\pi &= \frac{\theta^* - \bar{\theta} + s}{2s}, \\
\pi^n_{ex} &= \frac{\theta^* - \bar{\theta} + s - \bar{\phi}}{2s}, \\
\pi^n_{end} &= \frac{\theta^* - \bar{\theta} + s - \gamma}{2s - \gamma}.
\end{aligned}$$

For π^n_{end} to be non-negative, we assume that $2s - \gamma \geq 0$. Substituting the expression for π into the last two functions, we obtain π^n_{ex} and π^n_{end} as linear functions of π, i.e., of the absence rate without norms:

$$\pi^n_{ex} = \pi - \frac{\bar{\phi}}{2s} \tag{25}$$

$$\pi^n_{end} = \frac{2s}{2s-\gamma} \cdot \pi - \frac{\gamma}{2s-\gamma} \tag{26}$$

We may summarize the properties of our linear model in the form of the following proposition.

Proposition 7:

(i) The absence level is lower with than without norms.

(ii) Parameter changes in the insurance system (for instance, in *p* or *b*) will have the same effect on the aggregate absence rate in the case of exogenous norms as without norms.[33]

(iii) Parameter changes in the insurance system (for instance, in *p* or *b*) will result in larger changes in aggregate absence in the case of endogenous norms than with exogenous norms or no norms at all. In other words, endogenous norms create a "social multiplier" as defined by Glaeser, Sacerdote and Scheinkman (2003).

Proof: These properties follow immediately from the previous analysis.

In policy discussions, the consequences of tax wedges and moral hazard are often measured in terms of either monetary costs (in our notation $\pi \cdot b$) or absence rates (π). When measured by absence rates, one possibility is to look at the difference in absence rates with and without

[33] This result relies on our simplifying assumption of a rectangular distribution $F(\theta)$. For an arbitrary distribution, $\partial \pi^n_{ex} / \partial x < \partial \pi / \partial x$ iff $f(\theta^* - \bar{\phi}) < f(\theta^*)$. A sufficient condition for this to hold is that both cut-offs θ^{*n} and θ^* are located on the upward-sloping part of the density function $f(\theta)$.

insurance. For instance, in the case of non-observability the cost in terms of absence rates of introducing optimal insurance is[34]

$$C_1 \equiv \pi_N - \pi_0 . \qquad (27)$$

Of course, such a measure could also be applied in the presence of norms: $C_1^n = \pi^n - \pi_0^n$. In the case of exogenous norms, we have by (25) that

$$C_{1,ex}^n = C_1 ,$$

i.e., the cost (as defined here) is the same with and without norms. By contrast, it becomes different in the case of endogenous norms (by equation (26)):

$$C_{1,end}^n = \frac{2s}{2s-\gamma} C_1 .$$[35]

Thus, the cost of introducing insurance is the same with exogenous norms as with no norms.[36] By contrast, the cost of introducing insurance is higher (due to the multiplier $2s/(2s-\gamma)$) in an economy with endogenous norms than in an economy without norms.[37]

9. Concluding Remarks

[34] An alternative quantification of the cost of insurance would use the absence rate under full observability as the benchmark: $C_2 \equiv \pi_N - \pi_F$. While C_1 reflects total behavioral adjustment due to the insurance arrangement, C_2 reflects moral hazard as a result of non-observability.

[35] These expressions are easily obtained by noting that

$$\pi_{0,ex}^n = \frac{\theta_0^* - \bar{\theta} + s - \bar{\phi}}{2s} \quad \text{and} \quad \pi_{0,end}^n = \frac{2s}{2s-\gamma} \cdot \pi_0 - \frac{\gamma}{2s-\gamma} .$$

[36] The intuitive reason is that with a rectangular distribution, exogenous norms have the same consequences for π as for π_0.

[37] As in the end of Section 7, we could also isolate the cost associated with the moral-hazard problem: $C_2 = \pi - \pi_F$.

We have treated an individual's ability and willingness to work as a continuous variable, rather than a binary variable as is usual in the literature on income insurance. What are then the gains from our continuous approach? Basically, there are three gains. First, we obtain a more realistic view of the purpose of income insurance. It is not only a question of income smoothing but also of pain avoidance. Insurance helps the individual to finance absence when work is particularly burdensome. Optimal insurance requires the best possible trade-off between income smoothing and pain relief on one hand, and a fall in average consumption on the other hand.

Second, our broader view of the purpose of income insurance means that the conditions for insurance to be desirable turn out to be stricter than in the traditional theory. In addition to concavity, the desirability of insurance also depends on the relation between the distribution of the individual's health shock and the loss in consumption utility tied to reduced labor supply.

Third, our framework is conducive to the analysis of not only traditional incentives in income insurance (i.e., prices and incomes) but also the role of administrative rejection as well as the role of social norms. By treating an individual's ability and willingness to work as a continuous variable, both these features can easily be integrated into the analysis.

Other extensions also seem promising, although we have not pursued them in this paper. One is to consider whether part-time benefits should be allowed – a policy issue that has become increasingly important in recent years. Such a reform creates incentives for individuals to shift either from full-time benefits, or from full-time work, to part-time benefits. The net effect on total work absence would then be expected to depend on factors such as the distribution of the individual's state of health, whether the insurer restores budget balance after the reform by changing the contribution rate or the benefit rate, etc. In such a framework, administrative rejection could take the form of granting part-time sickness absence for some individuals who apply for full-time benefits.

Another extension could be to include the effects on production of so-called "presenteeism", i.e., a situation when individuals go to work even when they have health problems that reduce

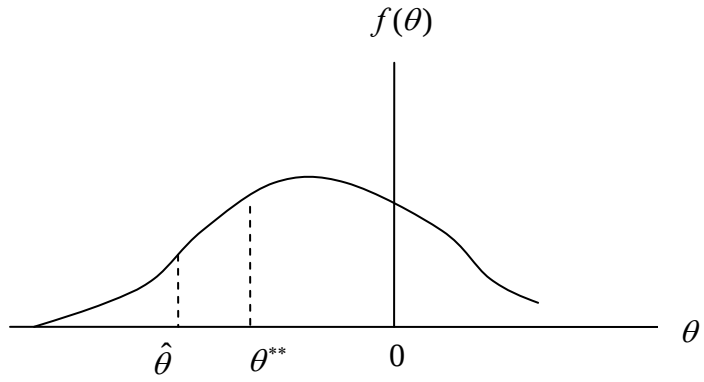
labor productivity at the workplace.[38] Finally, the model may be modified to address problems connected with *ex ante* moral hazard, i.e., behavioral adjustment by the individual *before* the random health shock has been realized (for instance, when the insured individual chooses a less prudent lifestyle). In our framework, *ex ante* moral hazard can be analyzed in terms of induced changes in the probability distribution of the health shock.

[38] See, for instance, Chatterji and Tilley (2002).

Appendix 1: Proof of Lemma 1

To prove Lemma 1 by contradiction, we assume that at the optimum, there *is* a group of people who stay home from work although they do not receive any benefits. If such a group exists, there must be an interval $\left(\hat{\theta}, \theta^{**}\right)$, with $\hat{\theta} < \theta^{**}$, where $\hat{\theta}$ is the cut-off specified in the contract and $\theta^{**} \equiv u(0) - u(1-p)$; see figure A1.

Figure A1



With a realization $\theta \le \hat{\theta}$, the individual stays home and receives a benefit *b*. With a realization between $\hat{\theta}$ and θ^{**}, the individual prefers to stay home even if he does not receive any benefit, while with a realization $\theta \ge \theta^{**}$, the individual chooses to work. The insurer's budget constraint is $F(\hat{\theta}) \cdot b = \left(1 - F(\theta^{**})\right) \cdot p$. The Lagrangean then is

$$L = F(\hat{\theta}) \cdot u(b) + \left(F(\theta^{**}) - F(\hat{\theta})\right) \cdot u(0) + \left(1 - F(\theta^{**})\right) \cdot \left(u(1-p) + E(\theta \,|\, \theta > \theta^{**})\right) + \\ + \lambda \cdot \left[\left(1 - F(\theta^{**})\right) \cdot p - F(\hat{\theta}) \cdot b\right]$$

and we obtain the first-order conditions with respect to *p* and *b*:

$$u'(b) = \lambda \qquad \text{and} \qquad u(b) - u(0) = \lambda \cdot b .$$

Hence, $u(b) - u(0) = b \cdot u'(b)$, which is a contradiction of the assumption of a strictly concave utility function $u(\cdot)$. Thus, no optimal contract $(p_F, b_F, \hat{\theta}_F)$ exists such that $\hat{\theta}_F < \theta^{**}$.

Q. E. D.

Appendix 2: Proof of Proposition 2

Condition (*i*) is necessary, since if $\theta_{upper} \leq u(0) - u(1)$ no one will work in the absence of insurance. Thus no one will work in the presence of insurance, either, and therefore insurance cannot be financed. Similarly, Condition (*ii*) is necessary, since if $\theta_{lower} \geq -u'(1)$ everybody will always work, and there is no need for insurance. To prove sufficiency, we note that, by definition, an interior solution $\hat{\theta}_F$ satisfies

$$\theta_{lower} < \hat{\theta}_F < \theta_{upper} . \qquad \text{(A1)}$$

To prove that a solution to (9c) satisfying (*i*) and (*ii*) also must satisfy (A1), we write (9c) as $\phi(\hat{\theta}) \equiv \hat{\theta} + u'\left(1 - F(\hat{\theta})\right) = 0$. We have $\phi(\theta_{lower}) = \theta_{lower} + u'(1)$ which, by (*ii*), is negative. We also have $\phi(\theta_{upper}) = \theta_{upper} + u'(0) > \theta_{upper} + u(1) - u(0)$, where the inequality follows from concavity. Thus, by (*i*), $\phi(\theta_{upper}) > 0$. The continuous and monotone function $\phi(\hat{\theta})$ must therefore have a zero in the open interval $(\theta_{lower}, \theta_{upper})$. *Q. E. D.*

Appendix 3: Proof that $b_N < 1 - p_N$

Assume an interior solution ($p_N > 0, b_N > 0, \mu = 0, \nu = 0$). Dividing (14) by (15) then yields

$$\frac{u'(1-p)}{u'(b)} = -\frac{1 - F(\hat{\theta})}{F(\hat{\theta})} \cdot \frac{u'(1-p) - \lambda}{u'(b) - \lambda} . \qquad \text{(A1)}$$

Since the left-hand side is always positive, the minus sign on the right-hand side, together with the fact that we have assumed an interior solution, means that the second term on the right-hand side is negative. We have three possible cases: (*i*) $u'(1-p) > \lambda$ and $u'(b) < \lambda$; (*ii*) $u'(1-p) < \lambda$ and $u'(b) > \lambda$; (*iii*) $u'(1-p) = \lambda$ and $u'(b) = \lambda$.

Case (*i*) implies that $1-p<b$, i.e., overfull insurance. Case (*ii*) implies that $1-p>b$, i.e., less than full insurance, while case (*iii*) implies $1-p=b$, i.e., full insurance. We now prove by contradiction that only case (*ii*) can apply.

Assume that case (*i*) applies. Then the right-hand sides of (14) and (15) are positive (recall that $\mu=\nu=0$) which is inconsistent with the left-hand sides being negative. Thus overfull insurance cannot be optimal. Assume then that case (*ii*) applies. Then the right-hand sides of (14) and (15) are negative; no contradiction occurs in this case. Assume finally that case (*iii*) applies. Then the right-hand sides of (14) and (15) are zero, which is inconsistent with the left-hand sides being negative. Thus only less-than-full insurance can apply. *Q. E. D.*

Let us instead assume a general, non-separable utility function $u(c,\theta)$. Instead of cases (*i*)-(*iii*) under equation (A1), we now obtain

$$E\left(u_1(1-p,\theta)\,\middle|\,\theta>\hat{\theta}\right)<\lambda \quad \text{and} \quad u_1(b,0)>\lambda$$

as the only possible configuration. Thus at the optimum, the expected marginal utility of consumption should be lower when working than when living on benefits. If it had not been for the cross-derivative u_{12}, this would translate into the conclusion of less-than-full insurance being optimal.

Appendix 4: Proof of Proposition 3

Condition (*i*) is necessary for the same reason as in the proof of Proposition 2. To prove that condition (*ii*) is sufficient, given that (*i*) is satisfied, we note that sufficiency means that $\theta_{lower}<u(0)-u(1) \;\Rightarrow\; p,b>0$. We will show by contradiction that this holds. Assume $\theta_{lower}<u(0)-u(1)$ and $p=b=0$. Since $\mu,\nu\geq 0$ for such a corner solution, (14) and (15) can be written $u'(1)-\lambda\geq 0$ and $\lambda-u'(0)\geq 0$. But these two inequalities imply $u'(1)\geq u'(0)$,

which is a contradiction in the case of a strictly increasing, concave utility function.[39]

Q. E. D.

Appendix 5:Proof of Proposition 6

With an exogenous *q* (Section 5), the Lagrangean

$$\begin{aligned} L = {} & \left(1 - F(\theta^*)\right) \cdot \left(u(1-p) + E(\theta \,\middle|\, \theta > \theta^*\right) + F(\theta^*) \cdot (1-q) \cdot u(b) + \\ & + F(\theta^{**}) \cdot q \cdot u(0) + \left(F(\theta^*) - F(\theta^{**})\right) \cdot q \cdot \left(u(1-p) + E(\theta \,\middle|\, \theta^{**} < \theta < \theta^*\right) + \\ & + \lambda \cdot \left\{\left[1 - F(\theta^*) + q \cdot \left(F(\theta^*) - F(\theta^{**})\right)\right] \cdot p - F(\theta^*) \cdot (1-q) \cdot b\right\} + \eta \cdot q \end{aligned}$$

is maximized with respect to *p*, *b* and *q*. From the first-order conditions, it is not possible to determine whether the optimal *q* should be zero or positive. To further clarify the issue, we simulated the model numerically (Figure 4).

With an endogenous *q* (Section 6), the Lagrangean is

$$\begin{aligned} L = {} & \left(1 - F(\theta^*)\right) \cdot u(1-p) + \left(1 - F(\theta^*)\right) \cdot E(\theta \,|\, \theta > \theta^*) + \int_{-\infty}^{\theta^*} f(\theta) \cdot \left(1 - q(\hat{\theta} - \theta)\right) d\theta + \\ & + \int_{-\infty}^{\theta^{**}} f(\theta) \cdot q(\hat{\theta} - \theta) d\theta \cdot u(0) + \int_{\theta^{**}}^{\theta^*} f(\theta) \cdot q(\hat{\theta} - \theta) \cdot \left(u(1-p) + \theta\right) d\theta + \\ & + \lambda \cdot \left[\left(1 - F(\theta^*) + \int_{\theta^{**}}^{\theta^*} f(\theta) \cdot q(\hat{\theta} - \theta) d\theta\right) \cdot p - \int_{-\infty}^{\theta^*} f(\theta) \cdot \left(1 - q(\hat{\theta} - \theta)\right) d\theta \cdot b\right], \end{aligned}$$

where $q(\hat{\theta} - \theta)$ is defined by equation (20). A maximization with respect to *p*, *b* and $\hat{\theta}$ yields three first-order conditions. Again, from these conditions, it is not possible to establish

[39] With a non-separable utility function, the inequality corresponding to $u'(1) \geq u'(0)$ becomes $u_1(0, 0) \leq E(u_1(1, \theta) \,|\, \theta > \theta_0^*)$. Whether this inequality is consistent with a concave utility function depends not only on the distribution of θ, as in the case of a separable utility function, but also on the cross derivative u_{12} (which may be positive or negative).

whether the optimal q is zero or positive. Therefore, the model was numerically simulated (Figure 5).

References

Albanesi**,** Stefania and Christopher Sleet (2006): "Dynamic Optimal Taxation with Private Information", *Review of Economic Studies*, Vol. 73, No. 1, pp. 1-30.

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

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

Brock, William A. and Steven N. Durlauf (2001): "Interactions-Based Models", in Heckman, J.J. and E. Leamer (eds.), *Handbook of Econometrics, Vol. 5*, Elsevier Science, New York.

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

Chatterji, Monojit and Colin J. Tilley (2002): "Sickness, Absenteeism, Presenteeism, and Sick Pay", *Oxford Economic Papers*, Vol. 54, pp. 669-687.

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

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

Diamond, Peter A. and Eytan Sheshinski (1995): "Economic Aspects of Optimal Disability Benefits", *Journal of Public Economics*, Vol. 57, No. 1, pp. 1-23.

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

Gosolov, Mikhail and Aleh Tsyvinski (2006): "Designing Optimal Disability Insurance: A Case for Asset Testing", *Journal of Political Economy*, Vol. 114, No. 2, pp. 257-279.

Lindbeck, Assar, Lars Nyberg and Jörgen W. Weibull (1999): "Social Norms and Economic Incentives in the Welfare State", *Quarterly Journal of Economics*, Vol. 114, No. 1, pp. 1-35.

Lindbeck, Assar and Mats Persson (2006): "A Model of Income Insurance and Social Norms", CESifo Working Paper No. 1675.

Moffitt, Robert (1983), "An Economic Model of Welfare Stigma", *The American Economic Review*, Vol. 73, No. 5, pp. 1023-1035.

Prescott, Edward C. and Robert M. Townsend (1984): "Pareto Optima and Competitive Equilibria with Adverse Selection and Moral Hazard" *Econometrica*, Vol. 52, No. 1, pp. 21-45.

Rees, Ray (1989): "Uncertainty, Information and Insurance", in J. D. Hey (ed.), *Current Issues in Microeconomics*, Macmillan, London.

Rees, Ray, and Achim Wambach (2008): "The Microeconomics of Insurance", *Foundations and Trends in Microeconomics*, Vol. 4, No. 1-2, pp. 1-163.

Rothschild, Michael and Joseph Stiglitz (1976): "Equilibrium in Competitive Insurance Markets: An Essay on the Economics of Imperfect Information", *Quarterly Journal of Economics*, Vol. 90, N0. 4, pp. 629-649.

Stiglitz, Joseph E. (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, no. 1, pp. 49-71.

Wilson, Charles A. (1977): "A Model of Insurance Markets with Incomplete Information", *Journal of Economic Theory*, Vol. 16, No. 2, pp 167-207.

Zweifel, Peter (2007): "The Theory of Social Health Insurance", *Foundations and Trends in Microeconomics*, Vol. 3, No. 3, pp. 183-273.

Figure 1: Location of the cut-offs under full observability: $\theta_0^*, \hat{\theta}_F$ and θ_F^*.

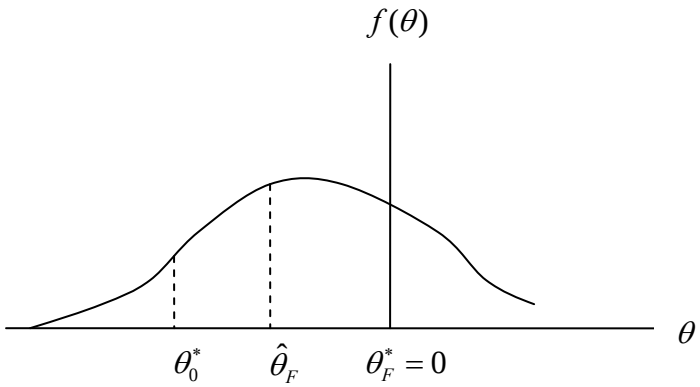


Figure 2: Equilibrium for the case of full observability

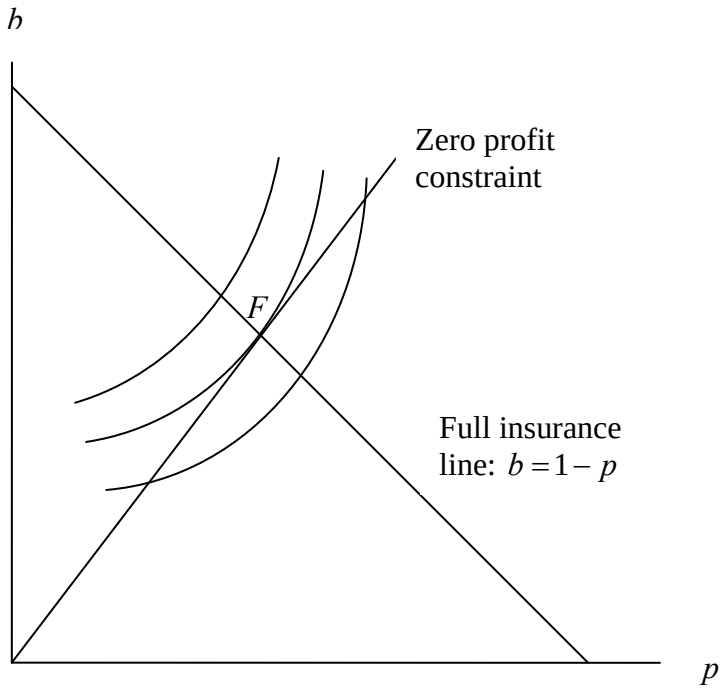


Figure 3: Equilibrium for the case of no observability

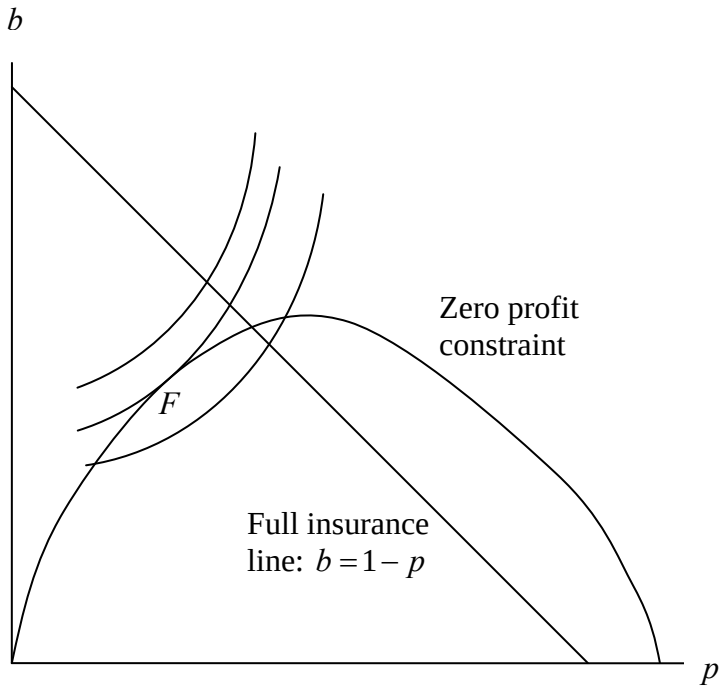


Figure 4: Optimal values of *p*, *b* and *q* for different values of *k*.

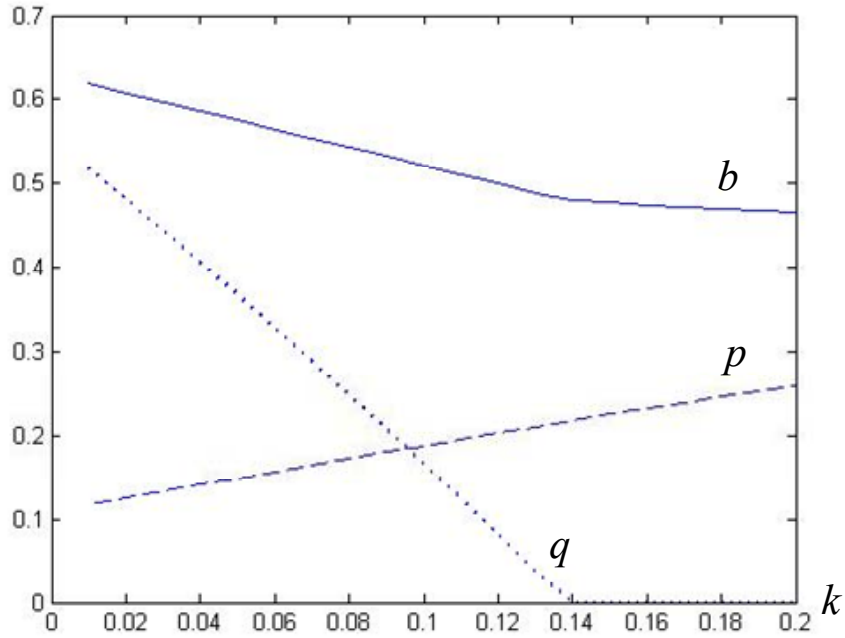


Figure 5: Optimal values of *p*, *b*, $\hat{\theta}$ and θ^* for different values of *k*.