Int Tax Public Finance (2013) 20:938–960
DOI 10.1007/s10797-012-9253-9

A continuous model of income insurance

Assar Lindbeck · Mats Persson

Published online: 4 October 2012

Abstract In this paper we treat an individual's health as a continuous variable, in contrast to the traditional literature on income insurance, where it is assumed that the individual is either able or unable to work. A continuous treatment of an individual's health sheds new light on the role of income insurance and makes it possible to capture a number of real-world phenomena that are not easily captured in the traditional, dichotomous models. In particular, we show that moral hazard is not necessarily outright fraud, but a gradual adjustment of the willingness to work, depending on preferences and the conditions stated in the insurance contract. Further, the model can easily encompass phenomena such as administrative rejection of claims, and it clarifies the conditions for the desirability of insurance in the first place.

Keywords Moral hazard · Disability insurance · Sick pay · Work absence · Tax wedge

JEL Classification G22 · H53 · I38 · J21

1 Introduction

The health state of an individual is most realistically regarded as a continuous phenomenon. Nevertheless, the academic literature on sick-pay insurance is based on

A. Lindbeck · M. Persson (✉)
Institute for International Economic Studies, Stockholm University, Stockholm, Sweden
e-mail: mp@iies.su.se

A. Lindbeck
e-mail: assar@iies.su.se

A. Lindbeck
IFN, Stockholm, Sweden

a dichotomous view of health. In particular, in the large literature that follows the seminal paper by Diamond and Mirrlees (1978) it is simply assumed that health is dichotomous: the individual is either able or unable to work.[1] Moral hazard then occurs when perfectly healthy individuals pretend to be unable to work, hence mimicking individuals who are unable to work. In the real world, however, such fraud is not the most important form of moral hazard. Rather, like health itself moral hazard is basically a gradual phenomenon, reflecting the everyday observation that individuals who feel more discomfort from work than usual, apply for and receive benefits without actually being unable to work. Moreover, an individual's decision to call in sick does not depend only on his health, but also on his preferences for leisure, as well as on other aspects of his private life. These variables are also continuous in nature.

For these reasons, important aspects of income insurance are lost in a dichotomous approach. One such aspect is how individuals adjust their labor supply to gradual variations in their health status as well as to gradual changes in the rules of the insurance contract. Another is the insight that some moral hazard is unavoidable in an optimum insurance contract if the individual's health is only imperfectly observable by the insurer. To make the analysis of the consequences of income insurance realistic we therefore treat the individual's ability and willingness to work as a continuous variable. However, our results do not depend on continuity *per se*. The results could alternatively be derived by treating health as a discrete variable with a large number of realizations for which the individual would, in principle, be able to work—although the pain of doing so would vary between the states. However, a continuous treatment of health is more analytically convenient.

Insurance contracts are quite simple in the real world. They are characterized by four parameters: a premium (usually, a fixed proportion of income), a benefit (also a fixed proportion of income), a criterion for eligibility, and an administrative procedure for deciding whether an individual's benefit claim should be accepted or not. In theory, one could conceive of much more complex contracts, where both the insurance premium and the benefits are non-linear (and possibly random) functions of an observable variable such as age and/or income.[2] In this paper, we do not discuss such hypothetical systems.

While the bulk of the literature has followed the dichotomous approach, there are a few insurance models with a continuous representation of an individual's health—although these studies deal with specific policy issues. In particular, Diamond and Sheshinski (1995) use a continuous approach when allowing a subgroup of retirees to replace their normal old-age pension with a more generous disability pension.[3]

[1]For expositions of the traditional dichotomous approach to insurance theory, see Rees (1989), Rees and Wambach (2008) and Zweifel (2007). Whinston (1983) and Golosov and Tsyvinski (2006) have elaborated on the Diamond–Mirrlees model in various ways.

[2]The general, non-linear income tax system of Mirrlees (1971) may be interpreted as an insurance system, where the insurer has information about the individual's income, but not about the realization of his productivity.

[3]Diamond (2003, Chap. 6) also uses a continuous approach to analyze the design of optimal retirement incentives. Moreover, Engström and Holmlund (2007) use a continuous representation of the individual's health when asking whether the benefit levels in unemployment and sick-pay insurance should differ or be the same. Outside the insurance literature, there are several papers on absence from work using a utility

We analyze a different set of questions: the consequences for aggregate labor supply when insurance is introduced, the distinction between moral hazard and tax wedges, and whether insurance is desirable for risk-averse individuals in the first place. We express our results in the form of "propositions" only when our conclusions differ from, or add to, results in the previous literature.

2 The basic model

Like Diamond and Sheshinski (1995) and Diamond (2003), we write the individual's utility in the simplest possible way:

$$u^W = u(c^W) + \theta \quad \text{when working} \tag{1}$$

$$u^A = u(c^A) \quad \text{when absent from work,} \tag{2}$$

where $u'(\cdot) > 0$ and $u''(\cdot) < 0$, and where θ is a random variable with a distribution function $F(\theta)$. We regard θ as an expression for an individual's willingness and ability to work (i.e., the disutility of work), which depends on factors such as his health, work environment and available leisure activities.

It should be pointed out that all our results do not depend on continuity *per se*. Formally, some of our results could be derived also in a situation with two health states for which the individual would in principle be able to work, although the pain of doing so might be considerable.[4] However, this does not hold for the effects on labor supply which is necessarily affected in our continuous model, but which may be unaffected in a dichotomous model. Since health in reality is continuous, a model like ours highlights the individual's adjustment to gradual variations in his health status. In addition, it turns out that a continuous treatment of the individual's health is analytically convenient.

One purpose of our paper is to analyze the determinants of sickness absence—in contrast to the traditional literature, where absence is exogenously given by the fraction of the population that has experienced the least favorable of the two possible health outcomes. However, we also discuss moral hazard and the desirability of insurance in the first place. We want to avoid complications that distract us from these purposes. Therefore, we make a number of strategic simplifications. First, the analysis is static, rather than dynamic. This means that we abstract from saving. Second, we assume additive separability in the utility function, as specified by (1), although we discuss the implications of dropping this assumption in some cases.[5] Third, while

function with a continuous index variable reflecting the individual's health status; see the survey by Brown and Sessions (1996).

[4]The traditional, dichotomous framework means that θ can take only two values, one of which is so negative (for instance, $-\infty$) that the individual is completely unable to work.

[5]There is a literature on how the marginal utility of consumption is influenced by the individual's health; see for instance Kremslehner and Muermann (2009) and Finkelstein et al. (2009). This literature often asserts that the cross derivative $u_{c\theta}$ is positive. However, it could well be argued that the cross derivative is negative. It is true that if you are sick, your marginal utility of many consumption goods (food, cars,

the disutility of work is represented by the continuous variable θ, labor supply is formally analyzed at the extensive margin, as usual in the literature on income insurance. One rationale for this simplification is that the extensive margin is particularly relevant when studying income insurance which mainly pays benefits to individuals who do not work at all. However, it is straightforward (but tedious) to work out the model for the case of part-time work and part-time income insurance. These simplifying assumptions may be worth relaxing in future work.

In the absence of insurance, the individual's utility may be written as $u^W = u(1) + \theta$ when working, with the wage rate normalized to unity, and $u^A = u(0)$ when absent from work.[6] The cut-off point at which he is indifferent between work and non-work is obtained by setting $u^W = u^A$ and yields

$$\theta_0^* \equiv u(0) - u(1) < 0. \tag{3}$$

Hence, in a world without insurance, the individual stays at home for all realizations $\theta \le \theta_0^*$ and goes to work otherwise.[7] Although our model formally deals with labor supply at the extensive margin, we may interpret $F(\theta_0^*)$ as the frequency of absence from work. Thus, the analysis also covers labor supply at the intensive margin. Intuitively speaking, while the individual during a specific day can only be wholly present at, or wholly absent from, work, his total absence during a year can be expressed as the continuous quantity $F(\theta_0^*)$.

Let us now introduce insurance into the model. At an abstract level, insurance can be defined as a contract conditioning a payment ψ on a random event s. For some values of s, the individual pays money to the insurer (i.e., $\psi(s)$ is negative, called a "premium") while for other values, the insurer pays money to the individual (i.e., $\psi(s)$ is positive, called a "benefit"). The individual's utility is $u^W = u(1+\psi(s)) + \theta$ for values of $\psi(s)$ and θ that induce the individual to work. The utility is $u^A = u(\psi(s))$ for all other values of $\psi(s)$ and θ. The optimal insurance system can be found by maximizing expected utility with respect to $\psi(s)$, subject to a zero-profit constraint for the insurer.

We derive the optimal insurance system under alternative assumptions about the information structure of the model. In one case, we assume that θ is *fully observable*, i.e., $s = \theta$. In another case, we assume that θ is *completely unobservable* for the insurer. There is also the intermediate (and most realistic) case, where θ is *partly observable*, i.e., where the payment ψ is conditioned on a noisy signal $s = \theta + \varepsilon$. The paper is organized around these three cases.

We interpret our model as describing the behavior of a large number of *ex ante* identical individuals, each with an i.i.d. stochastic taste parameter θ drawn from a

---

vacation trips) is low. But on the other hand, your marginal utility of certain other consumption goods might be very high. Obvious examples are nursing services, medical equipment, wheelchairs, etc. It should also be noticed that in contrast to this literature, the stochastic variable θ in our model enters the utility function only when the individual works; when he chooses not to work, his utility function is $u(c)$.

[6] 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 $u(\cdot)$.

[7] Dropping additive separability, we have $u(c,\theta)$. Instead of (3), the cut-off is then given by $u(1,\theta_0^*) = u(0,0)$. Provided that $u(c,\theta)$ is monotone in both arguments, the solution $\theta_0^* < 0$ is unique.

distribution $F(\theta)$. According to this interpretation, individuals differ *ex post*, i.e., after the realization of the stochastic taste parameters. One can interpret our model either as a model of a competitive insurance market where individuals freely can choose to be insured, or as a model of social insurance, where a benevolent government maximizes social welfare. As we assume an *ex ante*-homogeneous population, these two interpretations are equivalent.

3 Insurance under full observability

3.1 Optimal insurance

Although basically unrealistic, the case of full observability yields some basic insights about income insurance. Let W denote the set of realizations of θ for which the individual chooses to work, given the insurance system ψ. We assume that the insurance contract is conditioned on both the observable θ and the individual's (also observable) work decision $L = 0, 1$. We thus have $\psi(\theta, L)$, and we will show that under full observability, such a contract is equivalent to one conditioned on θ only, $\psi(\theta)$. While W is the set of realizations of θ for which the individual prefers working to living on benefits, A is the set of realizations of θ for which the individual prefers to be absent from work (living on benefits):

$$W \equiv \{\theta | u(1 + \psi(\theta, 1)) + \theta > u(\psi(\theta, 0))\}$$
$$A \equiv \{\theta | u(1 + \psi(\theta, 1)) + \theta \leq u(\psi(\theta, 0))\}.$$

Note that these sets depend on the insurance system; if $\psi(\theta, L)$ changes, then W and A will in general also change. The optimal insurance system is found by maximizing expected utility subject to the insurer's budget constraint. The Lagrangian thus is

$$L = \int_W [u(1 + \psi(\theta, 1)) + \theta] \, dF(\theta) + \int_A u(\psi(\theta, 0)) \, dF(\theta) \\ - \lambda \left[\int_W \psi(\theta, 1) \, dF(\theta) + \int_A \psi(\theta, 0) \, dF(\theta) \right]. \quad (4)$$

In the case of an interior solution $\psi(\theta, L) \neq 0$, the first-order conditions are

$$u'(1 + \psi(\theta, 1)) = \lambda, \quad \forall \theta \in W, \\ u'(\psi(\theta, 0)) = \lambda, \quad \forall \theta \in A. \quad (5)$$

Since these conditions imply $1 + \psi(\theta, 1) = \psi(\theta, 0)$, the optimal contract implies full insurance—a well-known property of optimal insurance under full observability. The first-order conditions also imply that the payments are independent of θ both when $L = 1$ and when $L = 0$. Hence, both $\psi(\theta, 1)$ and $\psi(\theta, 0)$ are constants which we denote

$$\psi(\theta, 1) \equiv -p, \quad \forall \theta \in W; \qquad \psi(\theta, 0) \equiv b, \quad \forall \theta \in A.$$

The optimal insurance contract, characterized by (5), is incentive-compatible by definition, since it has been derived from the sets W and A. The reason is that these sets define the values of θ for which the individual chooses to work, and be absent from work, respectively. As a result, we can drop the second argument in the $\psi(\theta, L)$ function and write the insurance system as $\psi(\theta)$.

Under full observability, the optimal insurance system thus implies a constant premium p and a constant benefit b.[8] There is also a value of θ, denoted by $\hat{\theta}$, at and below which the individual receives b, and above which he pays p: $\psi(\theta) = -p$ for $\theta > \hat{\theta}$ and $\psi(\theta) = b$ for $\theta \leq \hat{\theta}$. The optimal insurance contract may thus be written as a triplet $(p_F, b_F, \hat{\theta}_F)$, where the subscript denotes full observability. $\hat{\theta}_F$ is the critical value of θ below which a benefit b_F is received, and above which a premium p_F is paid.

With this insight, the Lagrangian (4) may be written

$$L = \big(1 - F(\hat{\theta})\big) \cdot \big[u(1 - p) + E\big(\theta | \theta > \hat{\theta}\big)\big] + F(\hat{\theta}) \cdot u(b) \\ + \lambda \cdot \big[\big(1 - F(\hat{\theta})\big) \cdot p - F(\hat{\theta}) \cdot b\big]. \tag{6}$$

The solution to the optimization is surprisingly simple:

$$p_F = F(\hat{\theta}_F), \tag{7}$$

$$b_F = 1 - F(\hat{\theta}_F), \tag{8}$$

$$\hat{\theta}_F = -\lambda = -u'\big(1 - F(\hat{\theta}_F)\big). \tag{9}$$

The system (7)–(9) is recursive. Since the right-hand side of (9) is monotonically decreasing in $\hat{\theta}_F$, it has a unique solution. Inserting this solution into (7) and (8), we obtain closed-form expressions for p_F and b_F. From (9) we see that the number of people not working under the optimal insurance system depends on preferences $u(\cdot)$ and the distribution function $F(\cdot)$—in contrast to the traditional, dichotomous model, where the number of people not working is simply the number of people who have experienced the less favorable of the two health states.

In the following, we emphasize two questions. One is whether insurance is desirable, and the other is how the individual's labor supply is affected by insurance.

3.2 Is insurance desirable?

In the literature on income insurance, it is usually claimed that insurance is always desirable (abstracting from administration costs) if the utility function is concave. However, in our analytical framework, insurance may not be desirable even if the representative individual is risk-averse. Whether or not insurance is desirable depends on the utility function $u(\cdot)$ and on the distribution function $F(\cdot)$.

We define the lower and upper support of the distribution as $\theta_{\text{lower}} \equiv \inf(\theta | f(\theta) > 0)$ and $\theta_{\text{upper}} \equiv \sup(\theta | f(\theta) > 0)$, and we have

[8] With a non-separable utility function $u(1 + \psi(\theta, 1), \theta)$ when working and $u(\psi(\theta, 0), 0)$ when not working, the optimal benefit under full observability should still be independent of θ while the optimal premium should vary with θ.

Proposition 1 *Assuming a concave consumption utility function, and abstracting from administrative costs, the equation system* (7)–(9) *has an interior solution* $p_F > 0$, $b_F > 0$ *if and only if the two following conditions are satisfied*:

(i) $\theta_{\text{upper}} > -u'(0)$
(ii) $\theta_{\text{lower}} < -u'(1)$.

In other words, insurance is desirable if and only if the distribution of θ *is such that there is a positive mass between the points* $-u'(0)$ *and* $-u'(1)$.

Proof See Appendix A. □

Thus, concavity of the utility function is not sufficient for insurance to be desirable. If conditions (i) and (ii) in the proposition are not satisfied, the individual will not demand insurance in a competitive market, and there is no reason for a benevolent planner who respects the individual's preferences to introduce a compulsory social insurance.[9]

The basic intuition behind the proposition is that the individual chooses to be insured only if the pain relief (i.e., the possibility to avoid all realizations $\theta \leq \hat{\theta}_F$) is greater than the loss of consumption utility $u'(1 - p_F)$ when paying the contribution. This happens only if the worst possible outcome is sufficiently bad to make it worthwhile to insure against it. Thus the pain of going to work (θ_{lower}) has to be greater than loss of consumption utility, $u'(1)$, when paying an arbitrarily small premium. Similarly, insurance can be financed only if the best possible outcome (θ_{upper}) is sufficiently large, so that the individual will work at least some time with insurance. This happens only if $\theta_{\text{upper}} > -u'(0)$, when receiving an arbitrarily small benefit.

One might be tempted to argue that these conditions for insurance to be desirable are trivial. For instance, one may think that an individual who never stays home from work in the absence of insurance does not need insurance. However, this would be an incorrect inference. The case for insurance does not depend on the individual's behavior in the absence of insurance, but on the new possibilities that arise if insurance is introduced. An individual who never stays home in the absence of insurance may nevertheless be interested in insurance. The reason is that he can then afford to stay home in the case of relatively unfavorable health states (relatively low realizations of θ). Proposition 1 says that it is not the cut-off point θ_0^* that is relevant for his willingness to be insured, but the marginal utilities $u'(1)$ and $u'(0)$.

One might also be tempted to assume that θ_{lower} in reality is extremely low—possibly, minus infinity. If that were the case, condition (ii) in Proposition 1 would be trivially satisfied in reality, and the individual would always buy insurance. But

[9] There are some similar results concerning the desirability of insurance in other strands of the literature. In the case of sick care insurance, assuming a dichotomous distribution of health, Strohmenger and Wambach (2000) analyze whether there exists some probability of sickness for which such insurance is not desirable, and show that an individual with concave consumption utility may abstain from buying insurance. Moreover, in the case of insurance of irreplaceable commodities (such as family heirloom, and the like) Cook and Graham (1977) point out the possibility that an individual may choose not to buy insurance. That result is driven by the assumption that money cannot compensate for irreplaceable losses (i.e., that there is a cross derivative $u_{c\theta} > 0$).

this assumption is not necessarily true. The reason is that θ does not reflect the individual's health situation in general, but the extra discomfort from working rather than staying home. Thus θ_{lower} could in reality take a wide range of different values between zero and minus infinity. Therefore, condition (ii) is not trivially satisfied in reality.

Proposition 1 may explain the empirical observation that individuals in poor countries often are not insured. To clarify this issue, let us for the moment drop the simplification that the wage rate is identically equal to unity. With a general wage rate w, condition (ii) in Proposition 1 should be written $\theta_{\text{lower}} < -u'(w)$. Assume that there are two societies with identical health distributions and identical preferences, but with different wage rates w_1 and w_2, where $w_1 > w_2$. Assume also that $-u'(w_2) < \theta_{\text{lower}} < -u'(w_1)$. By Proposition 1, the representative individual in society 1 (the rich country) will find insurance desirable, while the representative individual in society 2 (the poor country) will not. Heuristically, the same argument explains why poor groups of citizens in Western countries tend to be uninsured in the absence of mandatory insurance, while middle- and high-income groups tend to be insured. In our model, low-income earners will remain uninsured.

So far, we have abstracted from differences in risk across individuals. By contrast, the traditional insurance literature has studied the consequences of such differences, but have abstracted from differences in income.[10] The prediction from that literature is that high-risk individuals want buy insurance while low-risk individuals do not (adverse selection). Let us now combine income and risk differences, and realistically assume that low-income groups are more exposed to health risks than others. Then there are two counteracting influences on the demand for insurance across socio-economic groups. Since low-income earners have a higher marginal utility of consumption, they tend to be uninsured (by Proposition 1). But since they are exposed to higher risk, they are more interested than other groups in being insured. Which of these counteracting effects that dominates is an empirical question. Since in reality low-income earners are more often uninsured than high-income earners, it seems as if differences in income (via the influence of the marginal utility of consumption) dominate over differences in risk.

3.3 The effect of insurance on labor supply

One might think that the introduction of insurance makes the individual more picky, so that he stays home also for moderately bad realizations of θ. However, this intuition does not necessarily hold under full observability. Without insurance, the individual's absence rate is $F(\theta_0^*)$, while with insurance, it is $F(\hat{\theta}_F)$. If $\hat{\theta}_F > \theta_0^*$, the introduction of insurance would make the individual work less on average, and if $\hat{\theta}_F < \theta_0^*$, he would work more. We have

Proposition 2 *Under full observability, the effect on aggregate labor supply of introducing insurance is ambiguous.*

[10]Cf. Rothschild and Stiglitz (1976).

Proof By (3) and (9), $\hat{\theta}_F > \theta_0^*$, if and only if $-u'[1 - F(\hat{\theta}_F)] > u(0) - u(1)$, where $\hat{\theta}_F$ is the solution to (9). Since $0 < 1 - F(\hat{\theta}_F) < 1$, the derivative $-u'[1 - F(\hat{\theta}_F)]$ might be either larger or smaller than the difference $u(0) - u(1)$. Thus, we cannot say whether labor supply increases or decreases if optimal insurance is introduced. □

The intuition why the introduction of insurance has an ambiguous effect is that under full observability, the effect on labor supply is the result of an exogenous change in the individual's income. This change could be either positive (b) or negative ($-p$), depending on the realization of θ relative to the optimal cut-offs $\hat{\theta}_F$ and θ_0^*. With a positive income change, the individual would prefer more pain relief, hence reducing labor supply. With a negative change, he would opt for less pain relief and accordingly increase his labor supply. Which effect dominates depends on the utility function and the distribution of θ. In Appendix B we clarify the conditions under which the income change is positive or negative.

Thus, insurance causes a behavioral adjustment even in the case of full observability of θ although, as we have seen, the direction of the adjustment is undetermined. Since no asymmetric information is involved, this behavioral adjustment cannot be interpreted as moral hazard; it is the effect of lump-sum change in income.

Thus, already in the case of full observability, several properties of income insurance stand out that deviate from the traditional, dichotomous approach. First, although there is no moral hazard and no tax wedge, labor supply will change when optimal insurance is introduced. Second, it is possible to derive intuitively plausible conditions (in addition to concavity) for insurance to be desirable.

4 Insurance under non-observability

4.1 A baseline case

Let us now turn to the other polar case, namely when the realization θ cannot be observed at all by the insurer—a case that dominates the literature. While under full observability, it is possible to tie the payments ψ to the realization of θ, this cannot be done when θ is unobservable. Assuming that the individual's absence from work is observable, it is natural to tie benefits to that observation. We consider an insurance system where the individual pays a constant premium p when working and receives a constant benefit b when not working. In principle, the values of p and b could be stochastic, i.e., determined by a lottery. Indeed, as emphasized in the literature on mechanism design, the individual's behavioral response to such a lottery could reveal information about the unobserved θ; we return to this issue in Sect. 4.2. However, since we want to analyze real-world insurance systems we do not consider complex lotteries of this type. The only lottery we consider (in Sects. 4.2 and 5) is one where the claims of some applicants are rejected. Indeed, such rejections of claims are an essential part of real-world insurance systems.

With constant p and b, the individual has a subjective cut-off $\theta^* = u(b) - u(1 - p)$ such that he will prefer to live on benefits for all realizations $\theta \leq \theta^*$ and will prefer to work if $\theta > \theta^*$. Under non-observability, the contract must be incentive-compatible. This means that the insurer's cut-off $\hat{\theta}$ must be equal to the individual's cut-off θ^*:

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

Equation (10) corresponds to the "moral hazard constraint" in the Diamond and Mirrlees (1978) dichotomous model.

Maximizing the Lagrangian (6) subject to (10) and to the non-negativity constraints $p_N \geq 0$, $b_N \geq 0$ yields the first-order conditions

$$\big(\lambda - u'(1-p_N)\big)\big(1 - F(\hat{\theta}_N)\big) \leq f(\hat{\theta}_N) \cdot (p_N + b_N) \cdot u'(1-p_N) \cdot \lambda, \tag{11}$$

$$\big(u'(b_N) - \lambda\big) F(\hat{\theta}_N) \leq f(\hat{\theta}_N) \cdot (p_N + b_N) \cdot u'(b_N) \cdot \lambda, \tag{12}$$

where $\hat{\theta}_N \equiv u(b_N) - u(1-p_N)$. Assuming an interior solution, (11) and (12) are satisfied as equalities. The solution may be written as the triplet $(p_N, b_N, \hat{\theta}_N)$. It follows that $u'(1-p_N) < \lambda < u'(b_N)$ which means less than full insurance. The social value of resources (λ) deviates from the marginal utility of consumption both when working and when not working. The optimum under non-observability is thus a second-best optimum.

What, then, are the conditions for insurance to be desirable? We have

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

(i) *a necessary condition for an interior solution* ($p_N > 0$, $b_N > 0$) *is that* $\theta_{\text{upper}} > u(0) - u(1) \equiv \theta_0^*$;
(ii) *given* (i), *a sufficient condition is that* $\theta_{\text{lower}} < u(0) - u(1) \equiv \theta_0^*$.

Thus, in the case of non-observability, insurance is desirable if the distribution of θ *is such that there is a positive mass around the point* $\theta_0^* \equiv u(0) - u(1)$.

Proof Condition (i) is necessary since if $\theta_{\text{upper}} \leq u(0) - u(1) \equiv \theta_0^* < u(b_N) - u(1 - p_N) \equiv \hat{\theta}_N$, then no one will work, and no insurance can be financed. Thus, $\theta_{\text{upper}} > \theta_0^*$ is necessary for insurance to be feasible. When proving that condition (ii) is sufficient, given (i), we note that sufficiency means that $\theta_{\text{lower}} < u(0) - u(1) \;\Rightarrow\; p, b > 0$. We show by contradiction that this implication holds. Assume $\theta_{\text{lower}} < u(0) - u(1)$ and $p = b = 0$. The absence rate would then be $F(\theta_0^*)$. Since $\theta_{\text{lower}} < \theta_0^* < \theta_{\text{upper}}$, we have $0 < F(\theta_0^*) < 1$. Thus (11) and (12) imply $u'(1) - \lambda \geq 0$ and $\lambda - u'(0) \geq 0$. These two inequalities imply that $u'(1) \geq u'(0)$, which is impossible with a strictly concave utility function.[11] Thus, $p = b = 0$ cannot be optimal if $\theta_{\text{lower}} < \theta_0^*$. □

As in the case of full observability (Proposition 1) risk-averse individuals will abstain from insurance if the sacrifice in terms of consumption utility is large relative

[11] 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).

to the value of pain relief.[12] (Observe that pain relief is the difference in pain between non-work and work.) Hence, also in the case of non-observability, individuals in low-income countries may abstain from buying insurance.

The effect on labor supply can be summarized as

Proposition 4 *The introduction of insurance under non-observability leads to lower aggregate labor supply.*

Proof Labor supply is lower if $\hat{\theta}_N > \theta_0^*$. This inequality follows trivially from the fact that $\hat{\theta}_N = u(b_N) - u(1 - p_N) > u(0) - u(1) = \theta_0^*$ for all $p > 0, b > 0$. □

Thus insurance will make an individual stay home for less severe outcomes of θ than he would without insurance, and labor supply (and thus aggregate production) will therefore be lower. It turns out that this also holds if the utility function is non-separable.[13]

We have found that the effect of insurance on labor supply differs between the full-information and no-information case. As we noted earlier, the change in labor supply under full information is due to the effect of a lump-sum change in income (for a given θ). By contrast, under non-observability there is only a substitution effect:

Proposition 5 *Under non-observability, the labor supply is distorted by a tax wedge* $(p_N + b_N)$ *in the optimal insurance contract.*

Proof The income increase for an individual going from non-work to work is $(1 - p_N) - b_N = 1 - (p_N + b_N)$, where $(p_N + b_N)$ is the tax wedge. □

The results in Propositions 4 and 5 stand in contrast to the Diamond–Mirrlees model. In that model, an optimal insurance contract implies that labor supply is constant and equal to the number of people who are objectively able to work. Such models can therefore not contribute to our understanding of high sickness absence in European countries with generous insurance systems. The tax wedge $(p + b)$ in our model is quantitatively important since it should be added to the usual tax wedge, t, when assessing real-world incentives to work: the total tax wedge on labor earnings is thus $(t + p + b)$. For many European countries, realistic figures are of the magnitude $t = 0.25$, $p = 0.10$, and $b = 0.5$, which add up to 0.85. This is the relevant tax wedge when individuals choose whether to work or live on benefits.

We may say that income insurance has two rationales: income smoothing and pain relief (in the sense that insurance makes it affordable for the individual to stay home

[12] Condition (ii) in Proposition 3 is not necessary, but only sufficient. Even individuals who always work in the absence of insurance (i.e., individuals with $\theta_{\text{lower}} > \theta_0^*$) may be willing to buy insurance to be able to stay home for relatively unfavorable realizations of θ.

[13] With utility $u(c, \theta)$, the cut-off in the absence of insurance is given by $u(0, 0) = u(1, \theta_0^*)$. Condition (10) then is $u(b_N, 0) = u(1 - p_N, \hat{\theta}_N)$. Since the left-hand side of the latter equation is larger than the left-hand side of the former, and since $1 - p_N < 1$, it must hold that $\hat{\theta}_N > \theta_0^*$. Hence, the introduction of insurance leads to a fall in labor supply also with non-separable utility.

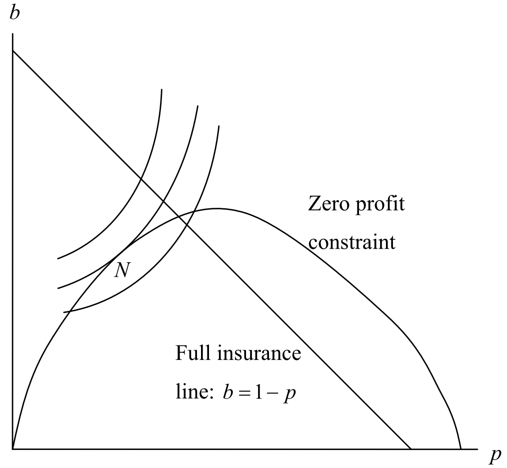


Fig. 1 Equilibrium for the case of non-observability

when working is particularly painful). The cost of insurance is a fall in labor supply and a corresponding loss in production (and hence consumption). The optimal contract thus implies a trade-off between income smoothing, pain relief, and consumption. In the traditional dichotomous model, there is no such trade-off, since optimal insurance in that model implies that individuals with the most favorable of the two health outcomes will work.[14]

A geometrical representation of the optimum insurance contract under non-observability may be useful. In Fig. 1 we have depicted a set of indifference curves in the (p, b) plane. [15] The insurer's zero-profit constraint in the (p, b) plane looks like a Laffer curve, as depicted in the figure.[16] The optimal contract is then represented by the point N. Since optimal insurance is less than full under non-observability, that point is located below the straight line representing full insurance, $b = 1 - p$.

4.2 Administrative rejection of claims

The optimal contract is second-best since the individual has a higher marginal utility of income when living on benefits than when working: $u'(1 - p_N) < \lambda < u'(b_N)$. It may therefore be tempting to redistribute income between the two states—for instance, by raising b_N and finance it by a higher p_N. However, this would increase

[14]In the Diamond and Mirrlees (1978) model, the disutility of work is infinite for a sick individual. In this sense, the tax wedge $p + b$ does not bite in that model. In other dichotomous models, disutility of work may be finite. In both cases, aggregate labor supply will be a step function.

[15]It is easy to prove that the indifference curves are upward-sloping, as shown in the figure. This also holds for non-separable utility functions. Although we have depicted the curves as convex, they may contain both concave and convex segments. Such an ambiguity concerning the curvature of indifference curves is found also in other strands of the insurance literature; cf. Stiglitz (1983) and Arnott (1992).

[16]It can be shown that a sufficient condition for the zero-profit constraint to look like a well-behaved Laffer curve, with a unique maximum, is that $F(\theta^*)/(1 - F(\theta^*))$ is an increasing, convex function of θ^*. This property characterizes many distribution functions, although there are some exceptions (for instance, a Student's t distribution with less than one "degree of freedom", i.e., with thick tails and an infinite mean).

sickness absenteeism in a non-optimal fashion. Is it possible instead to raise benefits for *some* individuals, by rejecting the claims of others? Since θ is non-observable, such rejection has to be completely random.

At first sight, one might think that exposing risk-averse individuals to additional randomness could never increase expected utility. However, we know from the general theory of economic policy that introducing an additional policy instrument—in this case a random rejection rate—may improve welfare. Indeed, the literature on mechanism design tells us that in a situation with asymmetric information, introducing a lottery may increase welfare. Random rejection of claims may function as such a lottery. However, the lotteries studied in the mechanism design literature are normally much more complicated than simple rejection of claims, and therefore often difficult (or even impossible) to implement in reality. This holds, for instance, for an insurance contract proposed by Prescott and Townsend (1984) which includes a lottery on how much an individual should consume and work. In our paper, we do not deal with insurance involving lotteries on work assignments, because we do not consider such lotteries enforceable.[17] We therefore confine our analysis to the effects of simple administrative rejection of insurance claims, as found in the real-world insurance.[18]

It turns out that the welfare consequences of introducing the possibility of rejection into insurance contracts depends crucially on whether an individual whose claim has been rejected can return to work or not. It can be shown that expected utility will always fall if a rejected individual cannot return to work. The intuition is simply that the individual in such a case will be exposed to a higher income risk without any compensating gain. Let us therefore concentrate on the case where an individual whose claim has been rejected *is* able to go back to work after the rejection. To analyze this case, first note that total absence now consists of two groups: those whose claims have been accepted and those who chose to stay at home after their claims have been rejected:

$$\text{Total absence} = (1-q)\cdot F\big(\theta^*\big) + q\cdot F\big(\theta^{**}\big), \tag{13}$$

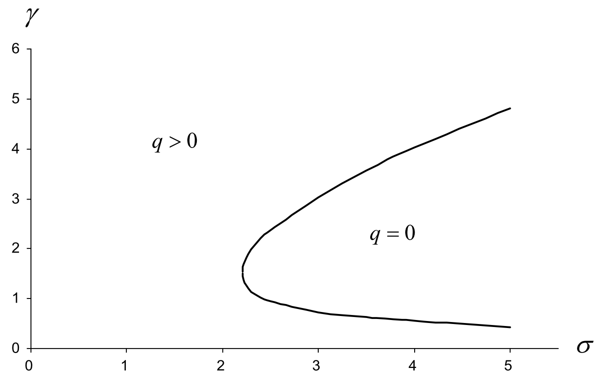
where q is the probability of being rejected, and where θ^{**} is the cut-off at which the individual is indifferent between staying at home with no benefits and working: $\theta^{**} \equiv u(0) - u(1-p)$. Expected utility is

$$\begin{aligned} EU_q \equiv {} & \int_{\theta^*}^{\infty} \big[u(1-p)+\theta\big]\,dF(\theta) + \int_{-\infty}^{\theta^*} (1-q)\cdot u(b)\,dF(\theta) \\ & + \int_{-\infty}^{\theta^{**}} q\cdot u(0)\,dF(\theta) + \int_{\theta^{**}}^{\theta^*} q\cdot \big(u(1-p)+\theta\big)\,dF(\theta). \end{aligned} \tag{14}$$

[17] More specifically, the Prescott and Townsend (1984) lottery implies that with probability π_1, you should work h_1 hours, and with probability π_2, you should work h_2 hours, regardless of your health status. If e.g. h_2 is very large, that outcome of the lottery may be impossible to implement in a society with no slave labor.

[18] By definition, all lotteries violate the principle of horizontal equity. Nevertheless, the world is full of lotteries, in many cases sponsored by the government. We return to the issue of rejection of claims in a more realistic framework in Sect. 5.

Fig. 2 The region in (σ, γ) space where the optimal rejection rate $q > 0$



We maximize EU_q with respect to p, b and q (with $\hat{\theta} = \theta^*$ to achieve incentive compatibility) subject to the non-negativity constraints $p \geq 0$, $b \geq 0$ and to the budget constraint

$$p \cdot \left[\int_{\theta^*}^{\infty} dF(\theta) + q \int_{\theta^{**}}^{\theta^*} dF(\theta) \right] - b \cdot (1-q) \int_{\theta^*}^{\infty} dF(\theta) = 0. \tag{15}$$

We denote the solution to this problem $(p_q, b_q, \hat{\theta}_q = \theta_q^*, q)$. It follows from the first-order conditions that $u'(1 - p_q) < \lambda < u'(b_q)$, illustrating the second-best character of the contract with less-than-full insurance. By contrast to the case without rejection, we now have two different tax wedges: $p_q + b_q$ for individuals who choose between working and applying for benefits, and p_q for individuals who, after having been rejected, choose between working and staying at home without benefits.

We cannot analytically determine whether q should be zero or positive in general. Therefore, we simulate the model numerically. For this purpose, we assume a utility function with constant relative risk aversion, i.e., $u(x) = x^{1-\gamma}/(1-\gamma)$. For $u(0)$ to be finite, we introduce an exogenous non-wage income, k. Consumption utility is now $u(1 - p + k)$ when working, $u(b + k)$ when absent from work and living on benefits, and $u(k)$ when absent without benefits. The results of the simulations are reported in Fig. 2 for a normal distribution $\theta \sim N(m, \sigma)$ of the taste parameter. The figure is based on the parameter values $k = 0.25$ and $m = 0$, but the results are qualitatively similar for a large set of values.[19] Combinations of sigma and gamma for which $q > 0$ are located outside the convex set in the figure.

Thus, our simulations prove that it is possible to find plausible parameter configurations for which a positive rejection rate is optimal.[20] The curve in the figure shows that for a given degree of risk aversion, the individual prefers a positive rejection rate for low values of the variance of θ. Intuitively speaking, the individual is willing to

[19] For instance, we have studied the consequences of variations in $k > 0$. It turns out that with our parameterization, q is a decreasing function of k.

[20] By "plausible" we mean values that are of an order of magnitude similar to those observed in the real world. In the simulations reported in Fig. 2, the absence rate, as given by (13), varies between 0.2 and 0.25. This is a realistic figure for many European countries if both sick-pay insurance and disability pensions are included.

take the risk of having his claim rejected when the probability of very negative outcomes of θ is small. As the variance increases, the individual will sooner or later be better off without a rejection rate.

Hence, the simulations provide a numerical proof of the following proposition.

Proposition 6 *If a rejected individual can return to work, a positive rejection rate q will increase the expected utility for some parameter constellations—in particular, when the variance of θ is small.*

The economic intuition for why an additional policy instrument in the form of a random rejection rate may be welfare-enhancing is that the rejection rate opens the possibility of better income smoothing for those whose claims are accepted. Of course, this is achieved at the price of less income smoothing for those whose claims are rejected. However, this cost is mitigated by the possibility for them to go back to work—an option that is chosen by those with relatively favorable realizations of θ (i.e., relatively low disutility of work). There is thus a self-selection back into work among those whose claims have been rejected.[21] This self-selection is conducive to allocative efficiency and is facilitated by the fact that the tax wedge for those who are rejected is only p_q, while the tax wedge in an insurance system without a rejection rate is $p_N + b_N$ for everybody.

We also have the following two results.

Proposition 7 *The conditions for insurance with a rejection rate to be desirable are the same as those for insurance without a rejection rate (Proposition 3).*

Proof The proof is parallel to that of Proposition 3. When proving Proposition 3, we showed that assuming a corner solution with $p = b = 0$ leads to a contradiction. The same holds in this case. Setting $p = b = 0$ in the first-order conditions to the maximization of (14) subject to (15) and the incentive compatibility constraint yields $\lambda \le u'(1)$ and $\lambda \ge u'(0)$ which cannot hold for a strictly concave utility function. Thus the conditions for insurance with a rejection rate to be desirable are the same as for insurance without a rejection rate. □

Proposition 8 *The introduction of insurance with a rejection rate $q > 0$ has the same qualitative effect on absence as the introduction of insurance without a rejection rate, i.e., aggregate labor supply will fall.*

Proof Everyone with a realization $\theta < \hat{\theta}_q = \theta_q^* = u(b_q) - u(1 - p_q)$ will apply for benefits. If all these claims were accepted, the fall in labor supply (as compared to the case with no insurance) would have been $\int_{\theta_0^*}^{\hat{\theta}_q} dF(\theta)$. However, a fraction q of these

[21] Thus, in our lottery, we do not eliminate the individual's freedom of choice: after the lottery has been executed, the individual is free to decide whether he should work or not. This contrasts to the work-assignment lottery in the paper by Prescott and Townsend (1984) mentioned in footnote 17 above; in that lottery, the individual is obliged to perform the number of hours of work specified by the outcome of the lottery.

claims are rejected, and rejected individuals with realizations $\theta > \theta^{**} \equiv u(0) - u(1 - p_q)$ will go back to work, rather than staying home without benefits. Thus, the net fall in labor supply is only $\int_{\theta_0^*}^{\hat{\theta}_q} dF(\theta) - q \int_{\theta^{**}}^{\hat{\theta}_q} dF(\theta)$. Since $0 < q < 1$ and $\theta_0^* < \theta^{**}$, the first term in this expression is larger than the second term. Hence labor supply and production will fall. □

Finally, we note that since the contract is incentive-compatible, there are no *Type I errors*: no one will receive benefits without being qualified. In this sense, moral hazard (= Type I errors) is ruled out in optimum. By contrast, a rejection rate $q > 0$ necessarily causes *Type II errors*: some individuals who qualify for benefits will not receive them. Such a rejection rate may nevertheless increase expected utility due to more generous insurance for those whose claims were accepted, and due to self-selection back to work among those whose claims were rejected.[22]

5 Partial observability

We now turn to the case where θ is *partially* observable. Clearly, this may be regarded as an intermediate case between full observability and non-observability. Moreover, it is the most realistic case. It turns out that several features of the optimal contracts under the two unrealistic assumptions concerning observability also appear, under various guises, in the more realistic case.

We assume that the insurer can observe a noisy signal $s \equiv \theta + \varepsilon$, where the noise ε has a cumulative distribution function $G(\varepsilon)$ with $0 < \text{var}(\varepsilon) < \infty$.[23] As in the case of non-observability, we limit the study to insurance contracts that are represented by a triplet $(p, b, \hat{\theta})$.[24] In the following, we discuss two versions of this contract. In the first version, which is the most common type of income insurance in the real world, the individual receives the benefit b if two conditions are satisfied: the signal s is smaller than or equal to $\hat{\theta}$, and the individual does not go to work. Symmetrically, the individual pays the premium p if s is greater than $\hat{\theta}$ and he goes to work. In the second type of contract, the benefit is conditioned only on the signal: the individual receives b if $s \leq \hat{\theta}$, regardless of whether he goes to work or not, and he pays p if $s > \hat{\theta}$.

Consider a particular individual, with health status θ. Like in the previous section, we let q denote the probability that an applicant who sends a signal s to the insurer will not be granted the benefit b. But under partial observability, q is not exogenous; it is instead given by

$$q \equiv \Pr(s > \hat{\theta}) \equiv \Pr(\varepsilon > \hat{\theta} - \theta) \equiv 1 - G(\hat{\theta} - \theta) \equiv q(\hat{\theta} - \theta). \tag{16}$$

[22] In the standard dichotomous model, a rejection rate can never increase expected utility. The reason is that there is no heterogeneity among those who apply for benefits in the case of an optimum contract; they are all unable to work.

[23] This general formulation nests the informational setups in Sects. 3 and 4. Under full observability, we would have $\text{var}(\varepsilon) = 0$, while in the case of non-observability, we would have $\text{var}(\varepsilon) = \infty$.

[24] An interesting topic for future study is a contract where the size of the benefit may vary with the signal $s = \theta + \varepsilon$ (the contribution cannot vary with the signal, since the premium has to be paid in advance).

Since distribution functions are always non-decreasing it follows that $q'(\hat{\theta} - \theta) \leq 0$ and hence q is a non-decreasing function of the true θ: $\partial q / \partial \theta \geq 0$. This property has an intuitive appeal; an individual with severe health problems (i.e., a very low θ) is less likely to be denied benefits than an individual who is healthier.

5.1 Benefits conditioned on the signal and on non-work

Let us start with the type of contract where payments are conditioned on both the signal and the individual's work decision. Diamond and Sheshinski (1995) studied a similar contract when asking whether an existing insurance (social security) should be supplemented by an additional insurance (disability pension) in the case of particularly bad outcomes. They concluded that such a supplement is warranted under certain conditions.[25] In this paper, we ask under what conditions (in addition to concavity) insurance is desirable in the first place, and we study the consequences for aggregate labor supply of introducing insurance.

With such a contract, the individual observes his realization θ and decides to apply for a benefit if $\theta \leq \theta^* \equiv u(b) - u(1-p)$. The expected utility is

$$EU_P \equiv \int_{\theta^*}^{\infty} \big[u(1-p) + \theta\big]\, dF(\theta) + \int_{-\infty}^{\theta^*} \big(1 - q(\hat{\theta} - \theta)\big) u(b)\, dF(\theta) + \int_{-\infty}^{\theta^{**}} q(\hat{\theta} - \theta) u(0)\, dF(\theta) + \int_{\theta^{**}}^{\theta^*} q(\hat{\theta} - \theta) \cdot \big(u(1-p) + \theta\big)\, dF(\theta), \tag{17}$$

where $\theta^{**} \equiv u(0) - u(1-p)$, i.e., the value of θ for which a rejected individual is indifferent between working and staying home without benefits. The insurer's budget constraint is

$$p \cdot \left[\int_{\theta^*}^{\infty} dF(\theta) + \int_{\theta^{**}}^{\theta^*} q(\hat{\theta} - \theta)\, dF(\theta)\right] - b \cdot \int_{-\infty}^{\theta^*} \big(1 - q(\hat{\theta} - \theta)\big)\, dF(\theta) = 0. \tag{18}$$

Since the insurer's decision to grant benefits is based on an observable signal $s \equiv \theta + \varepsilon$, the insurance contract does not have to be incentive-compatible: the insurer's cut-off $\hat{\theta}$ does not have to be equal to the individual's cut-off θ^*. The first-order conditions imply $u'(b_P) > \lambda > u'(1 - p_P)$ which means that the optimal solution $(p_P, b_P, \hat{\theta}_P)$ is second-best, just as in the case of non-observability. However, there will be a more efficient allocation between work and non-work among claimants than in the case of non-observability. In the latter case, the only mechanism for improving the allocation is self-selection among those whose claims have been rejected. Under partial observability, there are two selection mechanisms. One is self-selection, like in the non-observability case. The other is selection by the insurer, since the rejection

[25] The conditions are similar to the condition for a moral-hazard problem to emerge in the dichotomous model under non-observability of Diamond and Mirrlees (1978).

rate now is endogenous rather than constant: individuals with very low realizations of θ are now less likely than others to be rejected.

As in the preceding sections, we ask under what conditions insurance is desirable, and we also ask how the introduction of insurance affects labor supply. The answers are straightforward. First, the conditions for desirability are the same as in the case of non-observability, and thus concavity is not sufficient for insurance to be desirable.[26] Second, aggregate labor supply will fall if insurance is introduced.[27] These two results may seem surprising. One might have expected that partial observability would combine properties of full observability and non-observability, but this turns out not to be the case; this type of optimal contract under partial observability is more like the contract under non-observability.

While there is no moral hazard (in the sense of Type I errors) in the case of an optimal contract under non-observability, moral hazard may arise under partial observability: some (lucky) individuals will receive benefits even though their actual health status θ would not make them qualified for benefits. This holds for individuals whose *signals* are smaller than the insurer's cut-off (i.e., $s < \hat{\theta}$) at the same time as their *actual* θ is larger than $\hat{\theta}$. Such individuals appear sick in the eyes of the insurer, although in reality they are quite healthy. There will also be Type II errors; for some realizations of the disturbance, the individual looks healthy in the eyes of the insurer—although he is in fact sick. Thus, under partial observability, there will be both Type I and Type II errors.

5.2 Payments tied to the signal only

An alternative type of income insurance is when payments are tied to the (distorted) signal $s = \theta + \varepsilon$ of the individual's health, regardless of whether he works or not. If $s \leq \hat{\theta}$, the individual receives a benefit b from the insurer, and if $s > \hat{\theta}$, he pays a premium p. Thus, according to (16), the probability that the individual has to pay p to the insurer is $q(\hat{\theta} - \theta)$. Similarly, the probability that he will receive b from the insurer is $1 - q(\hat{\theta} - \theta)$. Since payments between the insurer and the insured are not contingent on work decisions, all individuals in the population participate in the "lottery" defined by the probability q. After the payments have been determined, the individual decides whether or not to go to work.

This type of contract, which to the best of our knowledge has not been discussed in the insurance literature, may be relevant if the variance of ε is very small. An example is accident insurance, including workers' compensation for work injuries. In this case, payments are often tied to the signal of the injury, independently of whether the individual chooses to work or not. Other examples where the signal is quite precise, and where such contracts could be implemented, are easily detected diseases such heart failure, cancer and diabetes. An advantage of not tying the benefits to an individual's work decision is that a tax wedge is thereby avoided. A disadvantage

[26] As in the case of Proposition 7, we use proof by contradiction. Let p and b go to zero in the first-order conditions for the maximization of (17) subject to (18). Provided that $0 \leq q < 1$, these inequalities yield $u'(1) \geq \lambda$ and $u'(0) \leq \lambda$, which is inconsistent with a strictly concave utility function.

[27] The proof is parallel to that of Proposition 8 above and is therefore not reported here.

is that there is not much income smoothing; some people might receive both wage income and benefits, $1+b$, while others may have to pay a contribution p although they feel such a strong discomfort from work that they stay home. The range for the income distribution is $[-p,\quad 1+b]$. With this type of contract, expected utility is

$$
\begin{aligned}
EU \equiv &\int_{\tilde{\theta}}^{\infty} \big[1-q(\hat{\theta}-\theta)\big] \cdot \big[u(1+b)+\theta\big]\, dF(\theta) \\
&+ \int_{-\infty}^{\tilde{\theta}} \big[1-q(\hat{\theta}-\theta)\big] \cdot u(b)\, dF(\theta) \\
&+ \int_{\tilde{\tilde{\theta}}}^{\infty} q(\hat{\theta}-\theta) \cdot \big[u(1-p)+\theta\big]\, dF(\theta) + \int_{-\infty}^{\tilde{\tilde{\theta}}} q(\hat{\theta}-\theta) \cdot u(-p)\, dF(\theta),
\end{aligned}
$$

where $\tilde{\theta} \equiv u(b) - u(1+b)$ is the cut-off between non-work and work for individuals with good luck in the lottery, and $\tilde{\tilde{\theta}} \equiv u(-p) - u(1-p)$ is the cut-off for individuals with bad luck. It follows from concavity that $\tilde{\tilde{\theta}} < \tilde{\theta}$. The budget constraint is

$$
b \int_{-\infty}^{\infty} \big[1-q(\hat{\theta}-\theta)\big]\, dF(\theta) = p \int_{-\infty}^{\infty} q(\hat{\theta}-\theta)\, dF(\theta).
$$

As in Sects. 4 and 5.1, the optimal contract in this section implies less than full insurance (for a proof, see Appendix A). This may seem surprising, since there is no tax wedge, and the situation in this sense is similar to that of full observability. The reason why full insurance is not optimal is that insurance under partial observability is plagued by Type I and Type II errors.

As with the other assumptions about the information structure, concavity is not sufficient for insurance to be desirable. The conditions are, however, more involved than in the previous sections (see Appendix A). Moreover, the effect on labor supply of introducing insurance is ambiguous, as in the case of full observability (for proof, see Appendix A).

Could this type of contract be preferable to the type of contract discussed in Sect. 5.1, where the payment is conditioned on both the signal and the work decision? Intuitively, we may expect that it is preferable if the variance of the disturbance term ε is small, since then the advantage of avoiding the tax wedge may dominate over the disadvantage of making Type I and Type II errors. To illustrate the relative merits of the contracts of Sects. 5.1 and 5.2, we have carried out simulations of the two models. For simplicity, we assume that the stochastic variables θ and ε are independently distributed. Figure 3 shows combinations of σ and σ_ε for which one contract dominates the other. The parameterization is the same as the one underlying Fig. 2. We have assumed that $\theta \sim N(0,\sigma)$, $\varepsilon \sim N(0,\sigma_\varepsilon)$, $k = 0.25$ and $\gamma = 2$. For all combinations of σ and σ_ε below the curve, the contract that is based only on the signal yields the highest expected utility. For all combination above the curve, the contract that is based on both the signal and the work decision (Sect. 5.1) yields the highest utility.

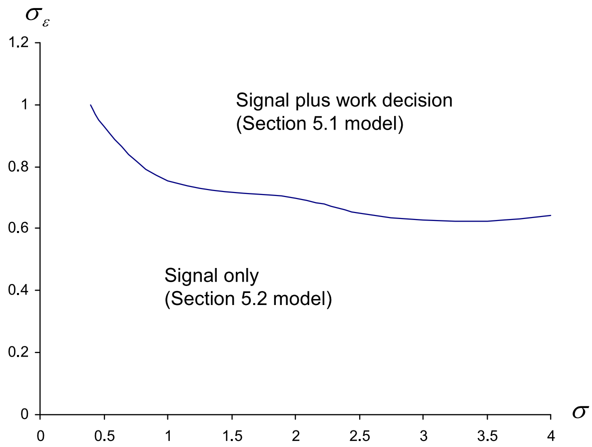


Fig. 3 The region in $(\sigma, \sigma_\varepsilon)$ space where the contract of Sect. 5.2 yields a higher expected utility than the contract of Sect. 5.1

Thus our hypothesis is confirmed: a contract based on the signal only is preferable to the contract of Sect. 5.1 if the variance of the disturbance ε is small. In such a case, the advantage of avoiding the tax wedge dominates over the disadvantage of a poor income smoothing caused by Type I and Type II errors.

6 Concluding remarks

In this paper, we have developed a model of income insurance to highlight several real-world features that are not well dealt with in the traditional, dichotomous model. In particular, our treatment of the individual's health as a continuous variable highlights the complex trade-offs between income smoothing, pain relief, and aggregate labor supply. These trade-offs, although often referred to in the general policy discussion, are not well reflected in the academic literature on income insurance, where the individual is assumed to be either able or unable to work.

The model can be extended in various ways. One possibility could be to modify the model in order to include part-time work and part-time benefits. Such an extension is straightforward, although tedious. Another extension is to incorporate social norms into the analysis. Indeed, social norms in the form of a stigmatization connected with living on benefits is an alternative to low benefits and a strict rejection rate as a means for reducing moral hazard. Finally, the model could be modified to address problems related to *ex ante* moral hazard, i.e., behavioral adjustment by the individual *before* a 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 as a situation where the introduction of insurance affects the probability distribution of the health shock.

Acknowledgements We are grateful to Per Engström, 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.

Appendix A: Proofs

Proof of Proposition 1

Condition (i) in the proposition is necessary since if $\theta_{\text{upper}} \leq \hat{\theta}_F = -u'(1 - F(\hat{\theta}_F))$, no one will ever work in the presence of insurance and thus, $F(\hat{\theta}_F) = 1$. In this case, the inequality can be written $\theta_{\text{upper}} \leq -u'(0)$ and no insurance can be financed. Thus, $\theta_{\text{upper}} > -u'(0)$ is a necessary condition for insurance to be feasible. Condition (ii) in the proposition is also necessary since if $\theta_{\text{lower}} \geq \hat{\theta}_F = -u'(1 - F(\hat{\theta}_F))$, everyone will always work in the presence of insurance and thus insurance will not be utilized. In that case, $F(\hat{\theta}_F) = 0$ and as a result the inequality can be written $\theta_{\text{lower}} \geq -u'(1)$. Thus, $\theta_{\text{lower}} < -u'(1)$ is necessary for insurance to be desirable.

To prove sufficiency, we note that, by definition, an interior solution $\hat{\theta}_F$ satisfies $\theta_{\text{lower}} < \hat{\theta}_F < \theta_{\text{upper}}$. To prove that a utility function u and a distribution function F satisfying (i) and (ii) must also satisfy this inequality, we define the function $\varphi(\theta) \equiv \theta + u'(1 - F(\theta))$. We have $\varphi(\theta_{\text{lower}}) = \theta_{\text{lower}} + u'(1)$ which, by (ii), is negative. We also have $\varphi(\theta_{\text{upper}}) = \theta_{\text{upper}} + u'(0)$ which, by (i), is positive. The continuous and monotone function $\varphi(\theta)$ must therefore take the value of zero for one (unique) value of θ somewhere in the open interval $(\theta_{\text{lower}}, \theta_{\text{upper}})$. By (9), the $\varphi(\theta)$ is zero for $\theta = \hat{\theta}_F$; thus $\hat{\theta}_F$ is located in the interval $(\theta_{\text{lower}}, \theta_{\text{upper}})$. □

Proofs of some properties of optimal insurance in Sect. 5.2

(i) An optimal contract implies less than full insurance

The first-order conditions of the maximization of *EU* subject to the budget constraint, with respect to p and b, can be written

$$\alpha \cdot u'(1-p) + (1-\alpha) \cdot u'(-p) \geq \lambda, \qquad \beta \cdot u'(1+b) + (1-\beta) \cdot u'(b) \leq \lambda,$$

where

$$\alpha \equiv \frac{\int_{\tilde{\tilde{\theta}}}^{\infty} q(\hat{\theta}-\theta)\, dF(\theta)}{\int_{-\infty}^{\infty} q(\hat{\theta}-\theta)\, dF(\theta)}, \qquad \beta \equiv \frac{\int_{\tilde{\theta}}^{\infty} [1-q(\hat{\theta}-\theta)]\, dF(\theta)}{\int_{-\infty}^{\infty} [1-q(\hat{\theta}-\theta)]\, dF(\theta)}.$$

With an interior solution, these conditions are satisfied as equalities, implying that the weighted average of $u'(1-p)$ and $u'(-p)$ should be equal to the weighted average of $u'(1+b)$ and $u'(b)$. Since $u'(-p) > u'(1+b)$, the two weighted averages can be equal only if $u'(b) > u'(1-p)$, i.e., only if $1-p > b$. □

(ii) The effects on labor supply are ambiguous

The introduction of insurance will induce some individuals with a lucky outcome to choose to stay home from work even though they would have gone to work in the absence of a lottery. These individuals will reduce their labor supply. The number of such individuals is $\int_{\theta_0^*}^{\tilde{\theta}} dF(\theta)$. Similarly, a number of individuals with a bad outcome in θ and bad luck in the lottery will choose to work, even though they would have stayed home in the absence of insurance; such individuals contribute to an increase in the labor supply. The number of these individuals is $\int_{\tilde{\tilde{\theta}}}^{\theta_0^*} dF(\theta)$. Without adding more structure to the model, it is not possible to determine whether $\int_{\tilde{\tilde{\theta}}}^{\theta_0^*} dF(\theta)$ is smaller or larger than $\int_{\theta_0^*}^{\tilde{\theta}} dF(\theta)$. □

(iii) Desirability of insurance

A necessary condition for insurance, based only on the signal s, to be desirable is that there is some mass between $\tilde{\tilde{\theta}}$ and $\tilde{\theta}$, where $\tilde{\theta} \equiv u(b) - u(1 + b)$ and $\tilde{\tilde{\theta}} \equiv u(-p) - u(1 - p)$. This condition is satisfied if there is some mass around θ_0^*.

To prove this, we first note that $\tilde{\tilde{\theta}} < \theta_0^* < \tilde{\theta}$. Assume that all mass of the distribution is to the right of $\tilde{\theta}$. Then everyone will always work, regardless of the outcome of the lottery. The lottery would thus only cause variability in income, and would therefore be undesirable to a risk-averse individual. Thus $\theta_{\text{lower}} < \tilde{\theta}$ is a necessary condition for the lottery to be desirable. Assume now that all mass of the distribution is to the left of $\tilde{\tilde{\theta}}$. Then no one would ever work. In such a case, the lottery would only cause income variability, which is undesirable. Hence $\theta_{\text{upper}} > \tilde{\tilde{\theta}}$ is also a necessary condition for the lottery to be desirable. These two conditions combined imply that some mass between $\tilde{\tilde{\theta}}$ and $\tilde{\theta}$ is necessary for the lottery to be desirable. Since $\tilde{\tilde{\theta}} < \theta_0^* < \tilde{\theta}$, some mass around θ_0^* is sufficient for this to occur. □

Appendix B: Effects on labor supply of insurance under full observability

Under what conditions does the introduction of insurance result in a positive (b) or a negative ($-p$) income change, and hence a fall or an increase in labor supply? It depends on whether $\theta_0^* = u(0) - u(1)$ is larger than, or smaller than $\hat{\theta}_F = -u'(1 - F(\hat{\theta}_F))$, which in turn depends on the $u(\cdot)$ and $F(\cdot)$ functions.

Assume first that $\theta^* < \hat{\theta}_F$. The individual would earn 0 if not working, and 1 if working, in the absence of insurance. With insurance, and with a realization $\theta^* < \theta < \hat{\theta}_F$, he would earn b if not working, and $1 + b$ if working (since the benefit is not conditioned on his labor supply decision, but only on the realization θ). Thus, the introduction of insurance provides a positive income increase b regardless of whether the individual works or not. As a result, the individual will demand more pain relief, hence reducing his labor supply.

This reasoning applies for a realization $\theta^* < \theta < \hat{\theta}_F$. For a realization outside that interval, the introduction of insurance will not have any effect on labor supply; an individual who does not work in the absence of insurance has even stronger reasons not to work when he gets an income transfer b. Correspondingly, an individual who works in the absence of insurance will have even stronger reasons to work when he is exposed to a lump-sum income reduction $-p$. The configuration $\theta^* < \hat{\theta}_F$ therefore implies that the introduction of insurance will reduce aggregate labor supply.

Assume instead that $\hat{\theta}_F < \theta^*$. As always, the individual would earn 0 if not working, and 1 if working, in the absence of insurance. With insurance, and with a realization $\hat{\theta}_F < \theta < \theta^*$, he would earn $-p$ if not working, and $1 - p$ if working. Thus, the introduction of insurance provides a negative income change, regardless of whether the individual works or not. As a result, he will settle for less pain relief, hence increasing his labor supply.

This reasoning applies for a realization $\theta^* < \theta < \hat{\theta}_F$. With the same argument as above for realizations outside the interval, we conclude that the configuration $\hat{\theta}_F < \theta^*$ implies that the introduction of insurance will increase aggregate labor supply.

The analysis shows that under full observability, the effect of insurance on labor supply is driven only by exogenous lump-sum changes in income, and not by any price distortions. These lump-sum changes are positive or negative, depending on the $u(\cdot)$ and $F(\cdot)$ functions.

References

Arnott, R. J. (1992). Moral hazard and competitive insurance markets. In G. Dionne (Ed.), *Contributions to insurance economics* (pp. 325–359). Boston: Kluwer Academic.
Brown, S., & Sessions, J. G. (1996). The economics of absence: theory and evidence. *Journal of Economic Surveys*, *10*(1), 23–53.
Cook, P. J., & Graham, D. A. (1977). The demand for insurance and protection: the case of irreplaceable commodities. *Quarterly Journal of Economics*, *91*(1), 143–156.
Diamond, P. A. (2003). *Taxation, incomplete markets, and social security. The 2000 Munich lectures*. Cambridge: MIT Press.
Diamond, P. A., & Mirrlees, J. A. (1978). A model of social insurance with variable retirement. *Journal of Public Economics*, *10*(3), 295–336.
Diamond, P. A., & Sheshinski, E. (1995). Economic aspects of optimal disability benefits. *Journal of Public Economics*, *57*(1), 1–23.
Engström, P., & Holmlund, B. (2007). Worker absenteeism in search equilibrium. *Scandinavian Journal of Economics*, *109*(3), 439–467.
Finkelstein, A., Luttmer, E., & Notowidigdo, M. (2009). Approaches to estimating the health state dependence of the utility function. *American Economic Review: Papers and Proceedings*, *99*(2), 116–121.
Golosov, M., & Tsyvinski, A. (2006). Designing optimal disability insurance: a case for asset testing. *Journal of Political Economy*, *114*(2), 257–279.
Kremslehner, R., & Muermann, A. (2009). *State-dependent preferences and insurance demand. Vienna university of economics and business*. Working paper. http://ssrn.com/abstract=1466719 or doi:10.2139/ssrn.1466719.
Mirrlees, J. A. (1971). An exploration in the theory of optimum income taxation. *Review of Economic Studies*, *38*(2), 175–208.
Prescott, E. C., & Townsend, R. M. (1984). Pareto optima and competitive equilibria with adverse selection and moral hazard. *Econometrica*, *52*(1), 21–45.
Rees, R. (1989). Uncertainty, information and insurance. In J. D. Hey (Ed.), *Current issues in microeconomics*, London: Macmillan.
Rees, R., & Wambach, A. (2008). The microeconomics of insurance. *Foundations and Trends in Microeconomics*, *4*(1–2), 1–163.
Rothschild, M., & Stiglitz, J. E. (1976). Equilibrium in competitive insurance markets: an essay on the economics of imperfect information. *Quarterly Journal of Economics*, *90*(4), 629–649.
Stiglitz, J. E. (1983). Risk, incentives and insurance: the pure theory of moral hazard. *The Geneva Papers on Risk and Insurance. Issues and Practice*, *8*(26), 4–33.
Strohmenger, R., & Wambach, A. (2000). Adverse selection and categorical discrimination in the health insurance markets: the effects of genetic tests. *Journal of Health Economics*, *19*(2), 197–218.
Whinston, M. D. (1983). Moral hazard, adverse selection, and the optimal provision of social insurance. *Journal of Public Economics*, *22*(1), 49–71.
Zweifel, P. (2007). The theory of social health insurance. *Foundations and Trends in Microeconomics*, *3*(3), 183–273. Adverse selection and categorical discrimination in the health insurance markets.