The Scandinavian
Journal of Economics

Scand. J. of Economics 111(2), 251–275, 2009
DOI: 10.1111/j.1467-9442.2009.01563.x

Optimal Social Insurance with Linear Income Taxation*

Lans Bovenberg

Tilburg University, NL-5000 LE Tilburg, Netherlands
A.L.Bovenberg@uvt.nl

Peter Birch Sørensen

University of Copenhagen, DK-1455 Copenhagen, Denmark
Peter.Birch.Sorensen@econ.ku.dk

Abstract

We study optimal social insurance aimed at insuring disability risk in the presence of linear income taxation. Optimal disability insurance benefits rise with previous earnings. Optimal insurance is incomplete even though disability risks are exogenous and verifiable so that moral hazard in disability insurance is absent. Imperfect insurance is optimal because it encourages workers to insure themselves against disability by working and saving more, thereby alleviating the distortionary impact of the redistributive income tax on labor supply and savings.

Keywords: Disability insurance; optimal taxation; moral hazard; redistribution; labor supply; skill groups

JEL classification: *H*21; *H*55

I. Introduction

This paper explores optimal social insurance in the presence of redistributive taxation. The optimal tax-benefit system redistributes income for two reasons: first, to reduce inequalities stemming from exogenous differences in productivities at the beginning of the working life and, second, to compensate unlucky individuals who become disabled during their career. Although we label the adverse shock to earnings capacity as "disability", our analysis applies also to other types of idiosyncratic shocks to human capital. From an *ex-ante* perspective, a redistributive income tax provides a form of insurance for individuals who turn out to have a low earnings capacity. We show that in the presence of such a redistributive income tax, the government should offer less than full insurance against disability even if disability risks are purely exogenous and fully verifiable. We thus integrate

* We wish to thank Thomas Gaube, Espen Moen and three anonymous referees for comments on an earlier draft of this paper. All remaining errors and shortcomings are our own responsibility.

the conventional analysis of optimal redistributive taxation with the analysis of optimal social insurance.

In our theoretical framework, people participate in the labor market for two periods, but some people become disabled in the second period. If disabled, they receive disability benefits. We demonstrate that these disability benefits should rise with previous income. In this way, the government can provide better disability insurance while at the same time improving first-period labor-supply incentives. In the presence of distortionary labor taxes, optimal disability insurance is incomplete. The reason is that imperfect insurance that rises with previous income encourages young workers to improve their insurance by working more in the first period, thereby alleviating the labor-market distortion produced by redistributive taxation.[1]

A paper related to our analysis is that of Diamond and Mirrlees (1978) who analyze optimal social insurance in a two-period model in which agents can choose their retirement age endogenously, but may be forced to retire early due to an exogenous risk of disability. One of the results derived by Diamond and Mirrlees is that agents who suffer disability early in life should receive a larger net transfer from the government than those able to work until later in life. The optimal social insurance scheme subsidizes those who retire early, although only to the extent that it is compatible with maintaining incentives to work.

Just as we do in this paper, Lozachmeur (2006b) extends the model of Diamond and Mirrlees by incorporating two different skill levels, thereby allowing for an analysis of the interaction between optimal social insurance and redistributive taxation. Lozachmeur focuses on the case of non-linear income taxation. He finds that full disability insurance is optimal if substantial skill heterogeneity ensures that the low skilled do not want to mimic the high skilled so that high-skilled labor supply is not distorted by the income tax. In contrast to Lozachmeur, we explore linear income taxation, following a long tradition in public economics that studies optimal redistribution if the tax-transfer system is constrained to be linear; see e.g. Stern (1976) and Dixit and Sandmo (1977). The interest in linear taxation was

[1] Our model assumes that social insurance is provided by the public sector. Even though the private market could implement full disability insurance, our analysis shows that such insurance would not be socially optimal because private insurers would fail to internalize the negative external effects of additional disability insurance on the base of the redistributive income tax. The government therefore has an incentive to prevent private insurance companies from fully insuring disability, for example by taxing private disability insurance. With full disability insurance, the welfare loss from marginally reducing that insurance would be only second order, while the positive labor-supply response would generate a first-order social welfare gain because the distortionary labor income tax drives the marginal product of labor above the marginal disutility of work in the initial equilibrium. For the external effects between insurers in the presence of moral hazard, see Pauly (1974) and Greenwald and Stiglitz (1986).

stimulated by the simulations in Mirrlees (1971), suggesting that the optimal (unconstrained) labor income tax schedule should be approximately linear. Although later work by Tuomala (1984) and others has shown that this conclusion is not robust to reasonable respecifications of the Mirrlees model, the interest in linear income taxation remained. Indeed, many countries have recently moved closer to linear taxation by reducing the number of tax brackets in order to simplify administration and enhance the transparency of the tax system. In fact, several countries in Eastern Europe have gone all the way by implementing a pure linear (flat) income tax.

One can view our analysis as being complementary to that of Lozachmeur (2006b) in that we explore the robustness of his results with respect to alternative assumptions about the availability of policy instruments. In the context of a model with two skill levels, it is particularly important to investigate whether results from non-linear taxation survive in a setting with linear taxation. The reason is that, with two skill types, non-linear taxation may yield peculiar tax rules that are only local results with little relevance for the optimal tax schedule in the presence of a continuum of households. To illustrate, the well-known result that the marginal tax rate on the highest skill level is zero, and that high-skilled labor supply is thus not distorted may have little practical relevance because it is only a local result for the highest skill level; see Tuomala (1984) and Diamond (1998). Our key result in this paper—that full disability insurance is not optimal—depends on the fact that labor supply is distorted: imperfect disability insurance encourages high-skilled as well as low-skilled workers to work more, thereby alleviating labor-supply distortions. Disability insurance thus becomes more imperfect with more substantial skill heterogeneity calling for higher distortionary taxes on labor income. In practice, high-skilled labor supply is distorted as governments levy positive rather than zero marginal tax rates on high-skilled workers. Hence, imperfect disability insurance can help alleviate the resulting labor-market distortions. In Lozachmeur (2006b), in contrast, imperfect disability insurance does not provide any efficiency gains by raising high-skilled labor supply because the marginal tax rate on the labor supply of this group is zero. This explains why full disability insurance may be optimal in Lozachmeur (2006b) in some cases, while it is never optimal in our setting with linear taxation. More generally, our formulation in terms of linear taxation yields more clear-cut, unambiguous results on the interaction between disability insurance and redistribution across skills compared to the analysis of non-linear taxation.

Our paper is also related to the literature on optimal dynamic taxation. Brito, Hamilton, Slutsky and Stiglitz (1991) study optimal dynamic non-linear taxation but abstract from disability risk. Dillén and Lundholm (1996) consider optimal linear taxation if the government cannot commit to future tax policy. Golosov, Kocherlakota and Tsyvinski (2003) investigate a setting in which the productivity levels of agents follow any arbitrary stochastic process. They focus on optimal non-linear taxation of saving rather than the taxation of labor income. Albanesi and Sleet (2006) explore optimal non-linear labor taxation with idiosyncratic productivity shocks that are identically and independently distributed. Gaube (2007) studies optimal non-linear taxation of labor income in a two-period setting with two skill types, finding that a zero marginal tax rate at the top is no longer optimal when taxation can be conditioned only on current (annual) income.

Another related strand of literature is the analysis of optimal age-specific income taxation. Non-linear age-dependent taxes are advocated by Kremer (2001) and by Blomquist and Micheletto (2008) as a means of improving the equity–efficiency trade-off. Erosa and Gervais (2002) and Lozachmeur (2006a) defend linear age-dependent income taxes on the basis of efficiency arguments in the setting of optimal Ramsey taxation in life-cycle models. Our paper considers the question of optimal age-dependent linear taxation when the social planner features a preference for redistribution.

Like us, several earlier writers have found that the optimal tax-transfer system involves imperfect social insurance combined with history-dependent social insurance benefits; see the review by Ljungqvist and Sargent (2004, Chs. 19 and 21). However, the previous literature derived these results in models with commitment problems and/or asymmetric information in social insurance in the presence of non-linear taxation. Such a setting gives rise to incentive-compatibility constraints that limit the scope for insurance. To the best of our knowledge, we are the first to demonstrate that imperfect social insurance with history-dependent benefits can be optimal even in the absence of commitment or information problems in social insurance. To arrive at our main result, that social insurance should be incomplete, all that is needed is the presence of a redistributive linear income tax.

Our finding may be interpreted in light of the well-known analysis of optimal commodity taxation by Corlett and Hague (1953), who found that substitutes for labor should be taxed while complements to labor should be subsidized. In our setting, labor and disability insurance can be viewed as substitutes in the sense that better disability insurance induces less work effort. Accordingly, our optimal tax-transfer system involves an implicit tax on disability insurance by offering only imperfect insurance.

Section II sets up our basic model, describing individual and government behavior in a setting in which only labor income can be taxed. Section III employs the model to demonstrate that perfect social insurance cannot be optimal, and Section IV shows that optimal disability benefits rise with previous earnings. Section V characterizes the optimal linear tax rate and the optimal degree of disability insurance. Section VI demonstrates that the presence of a tax on (income from) wealth is likely to reinforce our conclusion that imperfect social insurance is optimal. Section VII summarizes our main conclusions.

II. The Model

Individuals live for two periods. Everybody is able to work in the first period, but in the second period individuals face a risk of becoming disabled. Disabled individuals must finance their consumption by saving undertaken in the first period and by a public transfer that may be conditioned on their previous earnings. Able individuals work during (part of) the second period. The risk of disability is exogenous, and the government can perfectly observe whether a person is truly disabled or not.[2] We distinguish two skill groups (the low-skilled and the high-skilled) earning different real wage rates reflecting exogenous differences in labor productivity. Pre-tax factor prices are fixed by a linear technology. Without loss of generality, the real interest rate (and the individual subjective utility discount rate) is normalized to zero.

The government can observe a person's total income, but not his hourly wage rate.[3] In the main version of the model, we assume that the government can tax only wage income. In an extension of the model, we show that our main result on the suboptimality of full social insurance is likely to be reinforced if the government can tax not only labor income but also wealth (or, equivalently, capital income).

Individual Behavior

Expected lifetime utility of an individual (whether low-skilled or high-skilled) is given by the utility function[4]

$$U = u(C_1 - g(L_1)) + pu(C_{2d}) + (1-p)u(C_{2a} - g(L_2)), \tag{1}$$

$$u' > 0, \quad u'' < 0, \quad g' > 0, \quad g'' > 0, \quad 0 < p < 1,$$

where L_1 is labor supply during the first period of life, L_2 is the second-period labor supply of an able worker, C_1 represents first-period

[2] We make this assumption in order to show that even in the absence of moral hazard in social insurance, the government does not want to offer full insurance in the presence of redistributive taxation. In Bovenberg and Sørensen (2007), we show that the results reported in this paper continue to hold if the government cannot verify whether a person is disabled, so that public policy must respect certain incentive-compatibility constraints.

[3] If it could observe wage rates, the government could rely solely on non-distortionary redistributive taxes conditioned on wage rates in the present setting with exogenous skills.

[4] We choose a simple specification of preferences for expositional convenience. In Bovenberg and Sørensen (2007), we show that all our results continue to hold with a more general specification of preferences that allows the degree of intertemporal substitutability in consumption to be distinguished from the degree of risk aversion.

consumption, C_{2a} and C_{2d} denote the second-period consumption of an able person and a disabled person, respectively, and p stands for the exogenous probability of becoming disabled in the second period.

During the first period, the consumer's budget constraint amounts to

$$C_1 = W(1-t_1)L_1 + G - S, \tag{2}$$

where W is the real wage rate, t_1 is the constant marginal tax rate on first-period labor income, G is a lump-sum transfer received during the first period, and S is the amount of saving. In the second period, an able worker faces the following budget constraint:

$$C_{2a} = S + W(1-t_2)L_2, \tag{3}$$

where t_2 is the second-period tax rate, which may deviate from the marginal tax rate during the first period.[5] A disabled worker receives a disability benefit equal to the constant b plus a fraction s of previous labor income so that the individual can enhance the insurance against disability risk by raising first-period labor supply if $s > 0$. Hence, the second-period budget constraint is

$$C_{2d} = S + b + sWL_1. \tag{4}$$

The consumer maximizes (1) subject to (2) through (4). Optimal second-period labor supply implies that the marginal disutility of work equals the marginal after-tax real wage:

$$g'(L_2) = W(1-t_2). \tag{5}$$

The first-order condition for optimal saving requires that the marginal utility of first-period consumption equals the expected marginal utility of consumption during the second period:

$$\begin{gathered} u_1' = pu_d' + (1-p)u_a', \\ u_1' \equiv u'(C_1 - g(L_1)), \quad u_d' \equiv u'(C_{2d}), \quad u_a' \equiv u'(C_{2a} - g(L_2)). \end{gathered} \tag{6}$$

The first-order condition for optimal first-period labor supply amounts to

$$\left[W(1-t_1) - g'(L_1)\right]u_1' + psWu_d' = 0. \tag{7}$$

Part of the reward for first-period labor supply accrues in the second period if the disability benefit rises with previous earnings (i.e. $s > 0$). Substituting

[5] In an overlapping-generations context in which the government is constrained to levy the same tax rate on young and old workers in any given historical time period, the effective marginal tax rates on young and old workers may be differentiated from each other through second-period transfers that depend on first-period earnings, as demonstrated in Bovenberg and Sørensen (2007).

(6) into (7) to eliminate u_1', we can write (7) as

$$g'(L_1) = W(1 - (t_1 - s\hat{p})), \quad \hat{p} \equiv \frac{pu_d'}{pu_d' + (1-p)u_a'}. \tag{8}$$

The variable $\hat{p}$ can be viewed as the risk-adjusted (certainty-equivalent) probability of becoming disabled, and $t_1 - s\hat{p}$ may be interpreted as a risk-adjusted marginal effective tax rate on first-period labor income. The certainty-equivalent probability $\hat{p}$ exceeds the actual disability risk p if agents are risk-averse and not perfectly insured (so that $u_d' > u_a'$).

Equation (5) implies that second-period labor supply depends only on the second-period income tax rate. The Appendix explains how (6) and (8) give rise to the following policy effects on first-period labor supply:

$$\frac{\partial L_1}{\partial G} = -\left(\frac{sWu_1''p(1-p)}{u_1'\Delta}\right)\left(u_a'u_d'' - u_d'u_a''\right) \gtrless 0, \tag{9}$$

$$\frac{\partial L_1}{\partial b} = -\left(\frac{sWu_d''p(1-p)}{u_1'\Delta}\right)\left(u_a'u_1'' + u_1'u_a''\right) < 0, \tag{10}$$

$$\frac{\partial L_1}{\partial s} = \overbrace{\frac{\partial L_1^c}{\partial s}}^{\text{substitution effect}} + \overbrace{WL_1\frac{\partial L_1}{\partial b}}^{\text{income effect}}, \tag{11}$$

$$\frac{\partial L_1}{\partial t_1} = \overbrace{\frac{\partial L_1^c}{\partial t_1}}^{\text{substitution effect}} - \overbrace{WL_1\frac{\partial L_1}{\partial G}}^{\text{income effect}}, \tag{12}$$

where Δ is a positive magnitude defined in the Appendix, and $\partial L_1^c/\partial t_1$ and $\partial L_1^c/\partial s$ are compensated labor-supply responses given by

$$\frac{\partial L_1^c}{\partial t_1} = \left(\frac{Wu_1'}{\Delta}\right)\left[u_1'' + pu_d'' + (1-p)u_a''\right] < 0, \tag{13}$$

$$\frac{\partial L_1^c}{\partial s} = -\left(\frac{pWu_d'}{\Delta}\right)\left[u_1'' + pu_d'' + (1-p)u_a''\right] = -\hat{p}\cdot\frac{\partial L_1^c}{\partial t_1} > 0. \tag{14}$$

To understand the labor-supply effects of the transfers G and b, we consider condition (8) for optimal first-period labor supply: by expanding the budget set of a disabled person and thus lowering his marginal utility of consumption (u_d'), a rise in b reduces the certainty-equivalent disability risk $\hat{p}$, thereby increasing the effective marginal tax rate $t_1 - s\hat{p}$ on first-period labor income if $s > 0$. First-period labor supply consequently falls, as shown by (10). Intuitively, with a higher lump-sum component of the disability benefit, the individual faces a weaker incentive to raise disability income through first-period work effort.

A rise in the first-period transfer G raises saving, as the consumer seeks to smooth consumption over time.[6] The resulting increase in second-period wealth reduces u'_d as well as u'_a, so the effect on the certainty-equivalent disability risk $\hat{p}$ is ambiguous. Hence, the effect on first-period labor supply is also ambiguous, as implied by (9).

The last equality in (14) implies that t_1 and s are not equivalent policy instruments. In particular, if agents are risk-averse and not perfectly insured (so that $u'_d > u'_a$), (8) implies that $\hat{p} > p$, so that a rise in s exerts a more powerful impact on compensated labor supply than does a cut in t_1 with the same budgetary cost.[7] The reason is that while the reward for additional work can be raised through a cut in t_1, as well as through a rise in s, the increase in s allows the taxpayer to spend more of the additional income from higher labor supply on consumption in the disabled state. Higher consumption in the disabled state cannot be achieved as efficiently through an increase in precautionary saving out of a cut in t_1, because savings raise consumption also in the able state. Hence, a higher level of s makes work in the first period a more effective instrument for providing insurance against disability.

As we shall see below, the labor-supply responses reported above will allow us to characterize the optimal government policy.

The Government

Setting aside issues of intergenerational redistribution, we assume that the present value of the taxes levied on each generation equals the present value of the transfers paid to that generation. This implies that the *generational account* of each cohort is zero. We use the superscript l to denote a low-skilled worker while the superscript h indicates a high-skilled worker so that $W^h > W^l$. The exogenous fraction of low-skilled individuals in each cohort is α. Both skill types face the same probability p of disability in the second period of life. Normalizing the size of the cohort to unity, we can write the constraint that a cohort's generational account must be

[6] Equation (A8) in the Appendix implies that

$$\frac{\partial S}{\partial G} = \left(\frac{u''_1}{u'_1 \Delta}\right)\left[p(1-p)u'_a u''_d (sW)^2 - g''_1 \left(u'_1\right)^2\right] > 0,$$

where we have used the fact that $1-(1-\tau)\tilde{p} = (1-p)u'_a/u'_1$ if the wealth tax τ equals zero.

[7] With a frequency of disability equal to p, a rise in s of the magnitude ds has the same direct budgetary cost as an income tax cut of the absolute magnitude $|dt_1| = p \cdot ds$. With $\partial L_1^c/\partial s = -\hat{p} \cdot (\partial L_1^c/\partial t_1)$ (see (14)) and $\hat{p} > p$, it then follows that the compensated labor-supply response to a rise in s will be larger than the compensated supply response to an equally costly tax cut.

zero as

$$\alpha \overbrace{\left[t_1 W^l L_1^l - G + (1-p)t_2 W^l L_2^l - p\left(b + sW^l L_1^l\right)\right]}^{\text{generational account of a low-skilled worker}} +$$

$$(1-\alpha)\overbrace{\left[t_1 W^h L_1^h - G + (1-p)t_2 W^h L_2^h - p\left(b + sW^h L_1^h\right)\right]}^{\text{generational account of a high-skilled worker}} = 0. \quad (15)$$

The government maximizes the utilitarian sum of expected lifetime utilities, committing to its policy before individuals take any decisions. With V^l and V^h indicating the indirect (expected) utilities of a low-skilled worker and a high-skilled worker, respectively, we write the utilitarian social welfare function (*SWF*) as

$$SWF = \alpha V^l(G, b, s, t_1, t_2) + (1-\alpha)V^h(G, b, s, t_1, t_2), \quad (16)$$

which must be maximized with respect to the policy instruments G, b, s, t_1, t_2, subject to the government budget constraint (15).

To obtain the derivatives of the indirect utility functions, we insert (2) through (4) into (1), differentiate with respect to the policy instruments (denoted by subscripts), and apply the envelope theorem to find (for $i = l, h$):

$$V_G^i = u'_{1i}, \quad V_b^i = pu'_{di}, \quad V_s^i = pW^i L_1^i u'_{di}, \quad (17)$$

$$V_{t_1}^i = -W^i L_1^i u'_{1i}, \quad V_{t_2}^i = -(1-p)W^i L_2^i u'_{ai}. \quad (18)$$

The government can offer full insurance against disability to both skill groups if it wishes to do so. Full insurance requires that the second-period utility of a disabled worker equals that of an able worker. For low-skilled workers, this implies that

$$u\left(\overbrace{S^l + W^l L_2^l(1-t_2)}^{C_{2a}^l} - g\left(L_2^l\right)\right) = u\left(\overbrace{S^l + b + sW^l L_1^l}^{C_{2d}^l}\right) \Longleftrightarrow$$

$$W^l L_2^l(1-t_2) - g\left(L_2^l\right) - b - sW^l L_1^l = 0, \quad (19)$$

and for high-skilled workers the analogous condition for full insurance amounts to

$$W^h L_2^h(1-t_2) - g\left(L_2^h\right) - b - sW^h L_1^h = 0. \quad (20)$$

With the five non-equivalent policy instruments (G, b, s, t_1, t_2) available, the government will generally be able to satisfy (19) and (20) simultaneously. However, the next section shows that the government will not find it optimal to offer full disability insurance, although disability is strictly exogenous and perfectly verifiable so that disability insurance does not suffer from moral hazard.

III. The Suboptimality of Full Insurance

This section establishes that full disability insurance of both skill groups can never be optimal. To that end, we show that, starting from an equilibrium with full insurance, we can design a policy reform that leaves the utility levels of both groups unaffected, while at the same time increasing public revenue.

We start by noting that if both skill groups are fully insured, we may subtract (19) from (20) to obtain

$$W^h L_2^h(1-t_2) - g\left(L_2^h\right) - \left[W^l L_2^l(1-t_2) - g\left(L_2^l\right)\right] = s\left(W^h L_1^h - W^l L_1^l\right). \tag{21}$$

Since the left-hand side is positive,[8] and first-period skilled earnings exceed the corresponding unskilled earnings (i.e. $W^h L_1^h > W^l L_1^l$), this expression implies that $s > 0$. Intuitively, compared to the low skilled, the high skilled face a larger income loss if they become disabled. If low-skilled agents are fully insured against disability risk, the disability benefit must therefore rise with previous earnings to ensure that also the high-skilled agents are not hurt, should they become disabled.

We now make disability insurance less than perfect by decreasing b and increasing G. We reduce disability insurance in such a way that the expected lifetime utility of both households remains constant. Using the expressions for V_G^i and V_b^i given in (17), and noting from (6) that full insurance (i.e. $u_d' = u_a'$) implies that $u_d' = u_1'$ for both skill groups, we find that such a policy reform must satisfy

$$dG = -p \cdot db > 0. \tag{22}$$

According to (15), the policy reform (22) would yield no net effect on the government budget in the absence of changes in labor supply. At the same time, (5) implies that second-period labor supply remains constant because the tax rate t_2 is unaffected. From the perspective of the government's budget constraint, the net marginal tax rate on first-period labor income is $\hat{t}_1 \equiv t_1 - ps$. Accordingly, the overall impact of the policy reform on the government budget (15) can be written as $\hat{t}_1[\alpha W^l dL_1^l + (1-\alpha) W^h dL_1^h]$. The government budget thus improves if first-period labor supply of both

[8] The envelope theorem implies that the surpluses $W^h L_2^h(1-t_2) - g(L_2^h)$ and $W^l L_2^l(1-t_2) - g(L_2^l)$ are increasing in the pre-tax wage rate. $W^h > W^l$ thus implies that $W^h L_2^h(1-t_2) - g(L_2^h) > W^l L_2^l(1-t_2) - g(L_2^l)$.

skill types increases in the presence of any positive redistributive tax $\hat{t}_1 > 0$.[9] The policy reform (22) does in fact raise first-period labor supply, because (9) and (10) imply that[10]

$$p \cdot \frac{\partial L_1}{\partial G} - \frac{\partial L_1}{\partial b} = \left(\frac{sWp(1-p)}{u'_1 \Delta}\right)\left[u'_a u''_1 u''_d (1-p) + u''_a \left(u'_1 u''_d + p u'_d u''_1\right)\right] > 0. \tag{23}$$

The improvement of the public budget resulting from the utility-preserving policy reform (22) would enable the government to engineer a Pareto improvement, say, by raising G by more than implied by (22). This shows that the starting point characterized by full insurance of both skill groups cannot be a social optimum.

The intuition for full disability insurance not being optimal is that incomplete disability insurance helps to alleviate the labor-market distortions implied by redistribution. In particular, by reducing disability insurance through a cut in b, the government stimulates labor supply, and thus expands the base of the labor tax, because agents can partly undo the worsening of disability insurance by working harder in the first period if $s > 0$, which is a condition that must be met in the initial equilibrium with full insurance. Given an initial equilibrium with full disability insurance, the welfare loss from reduced insurance is only second order, but the expansion of the labor income tax base generates a first-order welfare gain if $\hat{t}_1 > 0$. To be sure, the envelope theorem implies that a small increase in labor supply yields no direct first-order effect on private welfare when workers are initially in a private optimum. The positive marginal tax rate, however, drives the marginal productivity of labor above the marginal disutility of work in the initial equilibrium so that the rise in labor supply does create a first-order social welfare gain showing up as an increase in public revenue. Accordingly, disability insurance should be less than perfect if the government wants to redistribute resources from high-skilled to low-skilled agents through a positive labor-income tax rate.

[9] Section V shows that $\hat{t}_1 > 0$ under the optimal redistributive policy.

[10] Note that an actuarially neutral policy reform, which yields no direct effect on the government budget (i.e. structured in accordance with (22)), enhances labor supply also if the initial equilibrium does not feature full insurance (i.e. $u'_1 \neq u'_a \neq u'_d$). Indeed, the labor-supply response to less disability insurance is triggered by the lower income in the disabled state compared to the able state, and does not depend on the direct welfare effects of reduced insurance. Intuitively, if $s > 0$ (so that labor supply provides additional disability insurance), higher labor supply helps to offset the impact of less insurance—even if agents are over-insured or underinsured in the initial equilibrium and an actuarially neutral reform implies a direct effect on private welfare. The sign of the labor-supply response thus does not depend on the sign of the direct private welfare effect of reduced insurance, but on the fact that the policy reduces insurance and thus makes anything that provides insurance (including labor supply in the case of $s > 0$) more attractive (or less unattractive if people are initially overinsured).

This result that disability insurance should be imperfect is related to the well-known analysis of optimal commodity taxation by Corlett and Hague (1953). Condition (23) implies that with positive s labor supply and disability insurance can be viewed as substitutes: less disability insurance—implemented through policy reform (22)—raises labor supply. According to Corlett and Hague (1953), substitutes for labor should be taxed in order to alleviate labor-market distortions implied by a distortionary tax on labor. This intuition explains why the optimal tax-transfer system involves an implicit tax on disability insurance by offering only imperfect insurance.

IV. Disability Insurance and Previous Income

The previous section showed that the policy variable s is positive if disability insurance is complete. However, we did not demonstrate that s is necessarily positive under the optimal policy. This section establishes that whenever disability insurance is less than perfect for at least one skill group—as Section III showed to be the case under the optimal policy—the government can generate a Pareto improvement by moving from a situation without any income-dependent disability benefits (i.e. $s = 0$) towards disability benefits that rise with previously earned income ($s > 0$). Hence, optimal policy involves disability benefits that increase with previous earnings.

Starting from an equilibrium with $s = 0$, consider a policy reform that combines an increase in s and t_1 satisfying

$$dt_1 = p \cdot ds, \quad ds > 0. \tag{24}$$

Using (6), (17) and (18), the welfare effects of the policy reform (24) amount to

$$dV^l = p(1-p)W^l L_1^l\left(u'_{dl} - u'_{al}\right)ds, \quad dV^h = p(1-p)W^h L_1^h\left(u'_{dh} - u'_{ah}\right)ds. \tag{25}$$

The policy experiment thus raises the welfare of at least one skill group whenever at least one group is imperfectly insured, so that $u'_d > u'_a$.[11] If we can prove that public revenue increases, we may thus conclude that our policy reform yields a Pareto improvement.

Policy experiment (24) generates no direct impact on net government revenue (see (15)), so the revenue effect of the reform depends

[11] If the low skilled are fully insured and $s = 0$, the high skilled will necessarily be imperfectly insured, since they have higher (potential) earnings but receive the same disability benefit as the low skilled do. We disregard the unlikely scenario in which the optimal policy involves *over*insurance (i.e. $u'_d < u'_a$) of one skill group. As shown in Bovenberg and Sørensen (2007), this possibility is ruled out by incentive-compatibility constraints if the government cannot perfectly verify disability. Even with perfectly verifiable disability, overinsurance is highly unlikely to be optimal, as Section V demonstrates.

entirely on labor-supply responses. Equations (9) through (12) imply that $\partial L_1/\partial s = \partial L_1^c/\partial s$ and $\partial L_1/\partial t_1 = \partial L_1^c/\partial t_1$ if $s=0$ initially. Using (13) and (14), we therefore obtain the following labor-supply reaction from both skill groups (where we drop the skill-group index for convenience):

$$\frac{\partial L_1}{\partial s} + p\frac{\partial L_1}{\partial t_1} = \left(\frac{p(1-p)W}{\Delta}\right)\left[u_1'' + pu_d'' + (1-p)u_a''\right]\left(u_a' - u_d'\right) > 0 \quad \text{for } u_d' > u_a'. \tag{26}$$

Labor supply thus expands. Intuitively, if disability benefits become linked to first-period labor effort, workers can obtain better disability insurance by working harder. The enhanced labor-supply incentives improve the government budget as long as the net redistributive tax rate $\hat{t}_1 \equiv t_1 - ps$ is positive, which will in fact be the case under the optimal redistributive policy described in Section V below. Accordingly, in addition to enhancing disability insurance, income-dependent disability insurance alleviates the labor-market distortions generated by redistributive taxation and is thus optimal.

V. Optimal Fiscal Policy

Section III showed that it cannot be optimal to offer full insurance to *both* groups of workers, but it did not prove that *neither* skill group should be fully insured. This section explores the conditions for optimal insurance and shows that, under weak conditions, the optimal policy does in fact involve imperfect insurance for both skill groups. In addition, we prove that the optimal tax policy does indeed involve a positive marginal effective tax rate on first-period labor income, as we assumed above.

The Appendix explains how the first-order conditions for the solution to the government's optimal policy problem may be used to derive the following expression for the optimal net marginal tax rate on first-period income:

$$\frac{\hat{t}_1}{1-\hat{t}_1} = \frac{(1-\beta_1)\left(1-\theta_1^h\right)(1-\alpha)}{\bar{\varepsilon}_1^c} > 0, \tag{27}$$

$$\beta_1 \equiv \frac{W^l L_1^l}{W^h L_1^h} < 1, \quad \theta_1^h \equiv \frac{u_{1h}'}{\lambda} + \hat{t}_1 W^h \frac{\partial L_1^h}{\partial G}, \quad \bar{\varepsilon}_1^c \equiv \alpha\beta_1\varepsilon_{1l}^c + (1-\alpha)\varepsilon_{1h}^c > 0,$$

where ε_{1l}^c and ε_{1h}^c denote the compensated net wage elasticities of first-period labor supply for the low skilled and the high skilled, respectively, so that $\bar{\varepsilon}_1^c$ represents a weighted compensated elasticity of first-period labor supply. The variable θ_1^h measures the marginal social valuation of a unit of first-period income for a high-skilled worker (accounting for the behavioral effect of more income on the tax base), where λ stands for the marginal social value of public funds. The Appendix shows that the government's preference for redistribution implies $\theta_1^h < 1$. Equation (27) is a special case of the classic formula for the optimal linear income tax derived by Dixit and Sandmo (1977). It confirms our previous claim that the optimal policy implies $0 < \hat{t}_1 < 1$. *Ceteris paribus*, the optimal marginal tax rate rises with lower compensated elasticities of labor supply (i.e. a lower value of $\bar{\varepsilon}_1^c$), a greater degree of inequality (i.e. a lower value of β_1) and a stronger preference for equality (i.e. a lower value of θ_1^h). Furthermore, a larger fraction of high-skilled workers in the labor force $1-\alpha$ broadens the base for redistribution, making it attractive to impose a higher marginal tax rate.

The Appendix employs the conditions for the optimal fiscal policy to derive expressions for the marginal utility differentials $u'_{dl} - u'_{al}$ and $u'_{dh} - u'_{ah}$, thereby implicitly characterizing the optimal degree of disability insurance for the two skill groups. We find that if $\hat{t}_1 > 0$, the following condition is *sufficient* (but far from necessary) to ensure that imperfect insurance for the low-skilled group ($u'_{dl} > u'_{al}$) is optimal:

$$\varepsilon_{1h}^c \geq \beta_1 \varepsilon_{1l}^c. \tag{28}$$

Since $\beta_1 \equiv W^l L_1^l / W^h L_1^h < 1$, this condition is met unless the labor supply of high-skilled workers is much less elastic than that of the low skilled.

For the high-skilled group, the Appendix shows that if the labor-supply responses of the two skill groups are symmetric and the effective tax rate $\hat{t}_1$ is set at its optimal positive level, the condition

$$\frac{2\theta_1^h - 1}{1 - \theta_1^h} + \eta(1 - \beta_1) > 0, \tag{29}$$

$$\eta \equiv 1 - (1-\alpha)\left(\frac{\gamma \varepsilon_{1h}^G}{\bar{\varepsilon}_1^c}\right), \quad \gamma \equiv \frac{W^h L_1^h (1 - \hat{t}_1)}{G}, \quad \varepsilon_{1h}^G \equiv \frac{\partial L_1^h}{\partial G} \frac{G}{L_1^h}$$

is necessary *and* sufficient to ensure that imperfect insurance of the high skilled ($u'_{dh} > u'_{ah}$) is optimal. If consumers have constant relative risk aversion, (9) implies that the income elasticity ε_{1h}^G is negative whenever $u'_{dh} > u'_{ah}$. Intuitively, a rise in G allows the consumer to undertake more precautionary saving, thereby reducing the need to insure disability through a strong first-period work effort. Since $\gamma > 0$ and $\beta_1 < 1$, it follows that the term $\eta(1 - \beta_1)$ in (29) is positive. Hence, $u'_{dh} > u'_{ah}$ is surely the optimal policy for any $\theta_1^h \geq 0.5$, and for values of β_1 substantially below unity condition (29) will be met even when the social valuation of a unit of high-skilled income (θ_1^h) is considerably below 0.5. Indeed, (29) will be violated only under an implausible combination of strong aversion to inequality (a very small value of θ_1^h) and little actual inequality (a value of β_1 close to unity). Accordingly, imperfect insurance of the high skilled is highly likely to be optimal if labor-supply responses are reasonably symmetric across skills.

From (29) and the associated definition of η we see that, *ceteris paribus*, imperfect insurance is more likely to be optimal the greater the degree of inequality (i.e. the lower the value of β_1), the broader the high-skilled tax base $(1-\alpha)$, and the lower the compensated net wage elasticity of labor supply ($\bar{\varepsilon}_1^c$). The reason is that lower values of α, β_1 and $\bar{\varepsilon}_1^c$ all drive up the marginal tax rate (cf. (27)), thereby distorting labor supply. To offset this distortion, the government finds it optimal to offer only imperfect disability insurance to skilled agents in order to induce them to insure disability by working harder in the first period.[12]

This analysis highlights the fact that, even in a setting with exogenous and fully verifiable earnings shocks, the presence of a redistributive income tax is both necessary and sufficient to ensure the optimality of imperfect social insurance with history-dependent social transfers. In the limiting case of perfect wage equality ($W^l = W^h$), (8) implies that $L_1^l = L_1^h$ so that $\beta_1 = 1$, and thus (from (27)) $\hat{t}_1 = 0$. The Appendix shows that in this case the optimal policy involves full disability insurance as well as $t_2 = 0$. With effectively only one skill group, full disability insurance can be achieved by proper choice of the lump-sum instrument b without resorting to s. In the absence of distortionary redistributive taxation ($t_1 = t_2 = s = 0$), the government does not face any second-best arguments for distorting social insurance so that social insurance is designed to equalize the marginal utility of consumption across states.

VI. Optimal Social Insurance with a Wealth Tax

We have so far assumed that the government can only observe and tax income from labor. This section shows that if the government can also impose a linear tax on wealth (or, equivalently, on income from capital), imperfect social insurance is even more likely to be optimal.

In the presence of a linear wealth tax levied at the rate $\tau < 1$, the second-period individual budget constraints (3) and (4) modify to

$$C_{2a}^i = S^i(1-\tau) + W^i(1-t_2)L_2^i, \quad C_{2d}^i = S^i(1-\tau) + b + sW^iL_1^i, \quad (30)$$

[12] The presence of the term $\gamma\varepsilon_{1h}^G$ in the definition of η also reflects that a higher (numerical) income elasticity of labor supply drives up the marginal tax rate by lowering θ_1^h (see (27)). However, since a simultaneous increase in $\gamma\varepsilon_{1h}^G$ and a fall in θ_1^h has offsetting effects on the left-hand side of (29), the effect on the optimal insurance policy is ambiguous.

and the government budget constraint (15) changes to

$$\alpha\big[t_1 W^l L_1^l - G + (1-p)t_2 W^l L_2^l - p\big(b + s W^l L_1^l\big) + \tau S^l\big] +$$
$$(1-\alpha)\big[t_1 W^h L_1^h - G + (1-p)t_2 W^h L_2^h - p\big(b + s W^h L_1^h\big) + \tau S^h\big] = 0. \tag{31}$$

We now show that if both skill groups were fully insured against disability, the government could almost certainly implement a Pareto improvement, implying that full insurance cannot be optimal. To show this, consider again the policy reform analyzed in Section IV, which involved a cut in b and a rise in G satisfying $dG = -p \cdot db$. According to (17), this reform still leaves the welfare of both skill groups unaffected if they are initially both fully insured. Moreover, (31) ensures that this reform exerts no direct impact on the public budget. Hence, the reform will allow the government to engineer a Pareto improvement if it induces a rise in labor supply and/or in savings so that the tax base expands and tax revenues thus increase. The first section in the Appendix, *The effects of taxes and transfers on labor supply and savings*, shows that with $\tau \neq 0$, we still have (for both skill groups and starting from an equilibrium with full insurance so that $s > 0$):

$$p \cdot \frac{\partial L_1}{\partial G} - \frac{\partial L_1}{\partial b} > 0. \tag{32}$$

The reform thus expands the labor-income tax base.

The effect of the reform on the wealth tax base is in principle ambiguous (see the Appendix). This may seem surprising because one would expect an increase in precautionary saving if disability insurance is worsened through a cut in the disability benefit. The explanation for this ambiguity is that workers may want to insure themselves against disability by increasing their first-period work effort rather than their financial saving if both the wealth tax and the subsidy rate s are very high and labor supply is elastic, while saving is inelastic, on account of a small elasticity of intertemporal substitution. However, the first section in the Appendix shows that such a counterintuitive scenario is highly unlikely. For example, suppose the disutility-of-work function $g(L)$ takes the constant-elasticity form $g(L) = (L^{1+1/\varepsilon})/(1+1/\varepsilon)$, so that the uncompensated net wage elasticity of labor supply implied by (5) and (8) equals the constant ε. Suppose further that the instantaneous utility function $u(x)$ features constant relative risk aversion with a coefficient of relative risk aversion equal to σ, implying that the constant elasticity of intertemporal substitution is $1/\sigma$. With these popular assumptions on preferences, the Appendix shows that whenever saving (and, hence, the wealth tax revenue) is initially non-negative, the following condition is sufficient (but not necessary) to ensure that the utility-preserving policy reform $dG = -p \cdot db$ will raise the savings of both skill groups in the benchmark case where $\hat{t}_1 = t_2$:

$$1 + (1-\tau)^{-\frac{1}{\sigma}} > \sigma p(1-p)\left(\frac{\varepsilon}{1+\varepsilon}\right)\frac{\tau}{(1-\tau)^2}. \tag{33}$$

Condition (33) will be met for any plausible parameter values, given that the factor $p(1-p)$ reaches a maximum of 0.25 at $p = 0.5$. For example, with logarithmic utility ($\sigma = 1$), a labor-supply elasticity $\varepsilon = 0.25$ and $p = 0.5$, (33) will be satisfied for any wealth tax rate below the punitive level of $\tau = 0.954$. Indeed, except in the unlikely case where the wealth tax is extremely high at the same time as labor supply is very elastic compared to saving (so that ε, σ and τ are all large), the utility-preserving move from perfect to imperfect disability insurance will expand the wealth tax base along with the labor-income tax base. With a positive wealth tax rate, the policy reform will then increase the revenue from the wealth tax, as well as the labor-income tax revenue, thus enhancing the government's ability to create a Pareto improvement by, say, raising G by more than what is needed to keep consumer utility constant.[13]

Accordingly, imperfect social insurance is optimal because it encourages agents to work and save more, thereby alleviating not only labor-market distortions associated with labor taxation, but also savings distortions implied by positive taxes on wealth (or on capital income).

VII. Conclusions

This paper studied optimal taxation and social insurance in an economy in which public policy aims to insure (from behind the "veil of ignorance") both skill heterogeneity and exogenous disability risk. Although the government has sufficient policy instruments to completely insure both skill groups against disability, full insurance is not optimal. This contrasts with the result of optimal full disability insurance in Diamond and Mirrlees (1978) and also in Lozachmeur (2006b) for the case with substantial skill heterogeneity. In particular, we find that the government can alleviate the distortionary impact of redistribution across skills by offering imperfect insurance and structuring disability benefits so as to encourage workers to improve their insurance against disability by working harder and saving more. Moreover, disability insurance should rise with previous earnings. This not only provides better disability insurance for the high skilled, but also enhances the incentives for all workers to supply labor, thereby alleviating the labor-market distortions caused by redistributive taxation.

[13] Even in the unlikely case in which the reform reduces savings, it will still allow a Pareto improvement as long as the fall in wealth tax revenue does not exceed the revenue gain from higher labor-income taxes.

Appendix

This appendix derives the effects of the various policy instruments on individual behavior and the first-order conditions for the solution to the optimal policy problem. We then use these relationships to prove some results reported in the main text.

The Effects of Taxes and Transfers on Labor Supply and Savings

To derive the individual behavioral responses to the policy variables, we start by inserting the budget constraints (2) and (30) into (1) to arrive at the following expression for expected lifetime utility in the extended model with a wealth tax:

$$\begin{aligned} U = {} & u\big(WL_1(1-t_1)+G-S-g(L_1)\big)+pu(S(1-\tau)+b+sWL_1) \\ & +(1-p)u\big(WL_2(1-t_2)+S(1-\tau)-g(L_2)\big). \end{aligned} \tag{A1}$$

Optimization of utility yields the following first-order conditions for optimal savings and first-period labor supply:

$$\begin{aligned} \partial U/\partial S=0 \Longrightarrow & -u'(WL_1(1-t_1)+G-S-g(L_1)) + (1-\tau)pu'\big(S(1-\tau)+b \\ & +sWL_1\big)+(1-\tau)(1-p)u'(S(1-\tau)+WL_2(1-t_2)-g(L_2))=0, \end{aligned} \tag{A2}$$

$$\begin{aligned} \partial U/\partial L_1=0 \Longrightarrow & \big[W(1-t_1)-g'(L_1)\big]u'\big(WL_1(1-t_1)+G-S-g(L_1)\big) \\ & +psWu'\big(S(1-\tau)+b+sWL_1\big)=0. \end{aligned} \tag{A3}$$

Differentiating (A2) and (A3), we obtain:[14]

$$\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \times \begin{bmatrix} dS \\ dL_1 \end{bmatrix} = \begin{bmatrix} D^S \\ D^L \end{bmatrix}, \tag{A4}$$

where

$$a_{11} \equiv u_1''+(1-\tau)^2\big[pu_d''+(1-p)u_a''\big], \quad a_{22} \equiv -g_1''u_1'+u_1''\big(sW\tilde{p}\big)^2+p(sW)^2u_d'',$$

$$a_{12}=a_{21}=sW\big[u_1''\tilde{p}+(1-\tau)pu_d''\big], \quad \tilde{p} \equiv \frac{pu_d'}{(1-\tau)\big[pu_d'+(1-p)u_a'\big]},$$

$$D^S \equiv u_1''dG-p(1-\tau)u_d''db-p(1-\tau)WL_1u_d''ds-WL_1u_1''dt_1,$$

$$D^L \equiv sW\tilde{p}u_1''dG-psWu_d''db-pW\big(u_d' + sWL_1u_d''\big)ds + W\big(u_1'+u_1''sWL_1\tilde{p}\big)dt_1.$$

The second-order condition for individual optimization ensures that

$$\Delta = a_{11}a_{22}-a_{12}a_{21}>0, \tag{A5}$$

[14] In deriving (A4), we use the fact that (A2) and (A3) imply

$$g'(L_1)-W(1-t_1)=\frac{sWpu_d'}{(1-\tau)\big[pu_d'+(1-p)u_a'\big]}=sW\tilde{p}.$$

where Δ is the determinant of the Jacobian in the system (A4). Solving this system by Cramer's rule and setting $\tau = 0$, we obtain the labor-supply effects reported in (9) through (14). With $\tau \neq 0$, the solution to (A4) implies[15]

$$\frac{\partial L_1}{\partial G} = \left(\frac{sWp(1-p)(1-\tau)^2 u_1''}{u_1' \Delta}\right)\left(u_d' u_a'' - u_a' u_d''\right), \tag{A6}$$

$$\frac{\partial L_1}{\partial b} = -\left(\frac{sWp(1-p)(1-\tau)u_d''}{u_1' \Delta}\right)\left[u_a' u_1'' + u_1'(1-\tau)u_a''\right], \tag{A7}$$

$$\frac{\partial S}{\partial G} = \left(\frac{u_1''}{\Delta}\right)\left[p u_d''(sW)^2\left(1-(1-\tau)\tilde{p}\right) - g_1'' u_1'\right], \tag{A8}$$

$$\frac{\partial S}{\partial b} = \left(\frac{p u_d''}{\Delta}\right)\left[u_1''(sW)^2 \tilde{p}\left(1-(1-\tau)\tilde{p}\right) + g_1'' u_1'(1-\tau)\right]. \tag{A9}$$

To derive the result reported in (32) in Section VI, we use (A6) and (A7) to find

$$p\frac{\partial L_1}{\partial G} - \frac{\partial L_1}{\partial b} = \left(\frac{sWp(1-p)(1-\tau)}{\Delta u_1'}\right)\left\{u_a' u_1'' u_d''[1-p(1-\tau)] + u_a''(1-\tau)\left(u_1' u_d'' + p u_d' u_1''\right)\right\}. \tag{A10}$$

This expression is positive if $s > 0$.

Section VI also discusses the effect of the policy reform $dG = -p \cdot db$ on savings when one starts out from an equilibrium with full insurance where $(1-\tau)\, u_a' = (1-\tau) \times u_d' = u_1'$ so that $\tilde{p} = p/(1-\tau)$. In that case we find from (A8) and (A9) that

$$p\frac{\partial S}{\partial G} - \frac{\partial S}{\partial b} = \left(\frac{p u_1'' u_d''}{\Delta}\right) \cdot X, \tag{A11a}$$

$$X \equiv -\left[g_1'' u_1' \left(\frac{1-\tau}{u_1''} + \frac{1}{u_d''}\right) + (sW)^2 p(1-p)\left(\frac{\tau}{1-\tau}\right)\right]. \tag{A11b}$$

Since the term $(p u_1'' u_d''/\Delta)$ in (A11a) is positive, the reform will raise savings if $X > 0$. The purpose of the derivations below is to evaluate the likely sign of X. Recalling that $u_1' = (1-\tau)u_d'$ with full insurance, we can write (A11b) as

$$X = -g_1''(1-\tau)C_{2d}\left[\left(\frac{u_1'}{(C_1 - g(L_1))\, u_1''}\right)\left(\frac{C_1 - g(L_1)}{C_{2d}}\right) + \frac{u_d'}{C_{2d} u_d''}\right] - (sW)^2 p(1-p)\left(\frac{\tau}{1-\tau}\right). \tag{A12}$$

[15] In deriving these expressions, we use that (A3) and the definition of $\tilde{p}$ imply

$$(1-(1-\tau)\tilde{p}) = \frac{u_a'(1-\tau)(1-p)}{u_1'} \quad \text{and} \quad \tilde{p} = \frac{u_d' p}{u_1'}.$$

If consumers feature constant relative risk aversion so that utility in the different states (indicated by subscripts) is given by

$$u_1 = \frac{(C_1 - g(L_1))^{1-\sigma}}{1-\sigma}, \quad u_a = \frac{(C_{2a} - g(L_2))^{1-\sigma}}{1-\sigma}, \quad u_d = \frac{C_{2d}^{1-\sigma}}{1-\sigma}, \tag{A13}$$

it follows from the assumption of full insurance (i.e. $u_1' = (1-\tau)u_d'$) that

$$(C_1 - g(L_1))^{-\sigma} = (1-\tau)C_{2d}^{-\sigma} \Longleftrightarrow \frac{C_1 - g(L_1)}{C_{2d}} = (1-\tau)^{-\frac{1}{\sigma}}. \tag{A14}$$

From (A2) and (A3), we find that

$$g'(L_1) = W\left(1-\tilde{t}_1\right), \quad \tilde{t}_1 \equiv t_1 - s\tilde{p}, \tag{A15}$$

where $\tilde{t}_1$ is the risk-adjusted marginal effective tax rate on first-period labor income in the presence of a wealth tax. If the disutility function $g(L)$ takes the constant elasticity form $g(L) = (L^{1+1/\varepsilon})/(1+1/\varepsilon)$, (A15) and (5) imply that $g'(L_1) = L_1^{1/\varepsilon} = W\left(1-\tilde{t}_1\right)$ and $g'(L_2) = L_2^{1/\varepsilon} = W(1-t_2)$, so that

$$L_1 = \left[W\left(1-\tilde{t}_1\right)\right]^{\varepsilon}, \quad L_2 = [W(1-t_2)]^{\varepsilon}, \tag{A16}$$

$$g_1'' = \frac{W\left(1-\tilde{t}_1\right)}{\varepsilon L_1}, \tag{A17}$$

where ε is the (constant) uncompensated net wage elasticity of labor supply. Noting from (A13) that $u_1'/(C_1 - g(L_1))u_1'' = u_d'/C_{2d}u_d'' = -1/\sigma$ and recalling that $C_{2d} = S(1-\tau) + b + sWL_1$, we may use (A14) and (A17) to rewrite (A12) as

$$X = W^2(1-\tau)Z, \tag{A18a}$$

$$Z \equiv \left(\frac{S(1-\tau) + b + sWL_1}{WL_1}\right)\left(\frac{1-\tilde{t}_1}{\sigma\varepsilon}\right)\left[1 + (1-\tau)^{-\frac{1}{\sigma}}\right] - \tau p(1-p)\left(\frac{s}{1-\tau}\right)^2, \tag{A18b}$$

where $X > 0$ iff $Z > 0$. To evaluate the likely sign of Z, note that when $g(L) = (L^{1+1/\varepsilon})/(1+1/\varepsilon)$, it follows from (21) and (A16) that full insurance implies

$$\left(\frac{1}{1+\varepsilon}\right)\left\{[W^h(1-t_2)]^{1+\varepsilon} - [W^l(1-t_2)]^{1+\varepsilon}\right\} = s(1-\tilde{t}_1)^{\varepsilon}\left[(W^h)^{1+\varepsilon} - (W^l)^{1+\varepsilon}\right]. \tag{A19}$$

In the benchmark case discussed in Section VI where $\tilde{t}_1 = t_2$, we find from (A19) that

$$s = \frac{(1-t_2)}{1+\varepsilon}. \tag{A20}$$

Full insurance ($u_d' = u_a'$) also implies that $b + sWL_1 = WL_2(1-t_2) - g(L_2)$. Assuming once again that $\tilde{t}_1 = t_2$, and using $g(L) = (L^{1+1/\varepsilon})/(1+1/\varepsilon)$, (A16) and (A20), we can rewrite this condition as follows (which holds for both skill groups):

$$b = \frac{[W(1-t_2)]^{1+\varepsilon}}{1+\varepsilon}\left[1 - \left(\frac{1-\tilde{t}_1}{1-t_2}\right)^{\varepsilon}\right] = 0 \quad \text{for } \tilde{t}_1 = t_2. \tag{A21}$$

With non-negative savings, it follows from (A18b) and (A21) that the following condition is sufficient (but not necessary) to ensure that $Z > 0$:

$$\left(\frac{1-\tilde{t}_1}{\sigma\varepsilon}\right)\left[1+(1-\tau)^{-\frac{1}{\sigma}}\right] > sp(1-p)\left(\frac{\tau}{(1-\tau)^2}\right). \tag{A22}$$

Inserting (A20) into (A22) along with $\tilde{t}_1 = t_2$, one ends up with condition (33) in Section VI.

Optimal Taxes and Insurance

We now focus on the basic model without a wealth tax ($\tau = 0$). The government's policy problem is to maximize the social welfare function (16), subject to the government budget constraint (15). Using (17) and (18) together with (11) through (14), and recalling that $\hat{t}_1 \equiv t_1 - ps$ and $\hat{p}^i \equiv (pu'_{di})/(pu'_{di} + (1-p)u'_{ai})$, we may write the first-order conditions for the solution to this problem as follows, where λ is the shadow price associated with the government budget constraint:

$$G\colon \quad \alpha u'_{1l} + (1-\alpha)u'_{1h} + \lambda\hat{t}_1\left[\alpha W^l \frac{\partial L_1^l}{\partial G} + (1-\alpha)W^h \frac{\partial L_1^h}{\partial G}\right] = \lambda, \tag{A23}$$

$$b\colon \quad p\left[\alpha u'_{dl} + (1-\alpha)u'_{dh}\right] + \lambda\hat{t}_1\left[\alpha W^l \frac{\partial L_1^l}{\partial b} + (1-\alpha)W^h \frac{\partial L_1^h}{\partial b}\right] = p\lambda, \tag{A24}$$

$$\begin{aligned} s\colon \quad & p\left[\alpha W^l L_1^l u'_{dl} + (1-\alpha)W^h L_1^h u'_{dh}\right] + \lambda\hat{t}_1\alpha W^l\left(W^l L_1^l \frac{\partial L_1^l}{\partial b} - \hat{p}^l \frac{\partial L_1^{cl}}{\partial t_1}\right) \\ & + \lambda\hat{t}_1(1-\alpha)W^h\left(W^h L_1^h \frac{\partial L_1^h}{\partial b} - \hat{p}^h \frac{\partial L_1^{ch}}{\partial t_1}\right) = p\lambda\left[\alpha W^l L_1^l + (1-\alpha)W^h L_1^h\right], \end{aligned} \tag{A25}$$

$$\begin{aligned} t_1\colon \quad & \alpha W^l L_1^l u'_{1l} + (1-\alpha)W^h L_1^h u'_{1h} - \lambda\hat{t}_1\alpha W^l\left(\frac{\partial L_1^{cl}}{\partial t_1} - W^l L_1^l \frac{\partial L_1^l}{\partial G}\right) \\ & - \lambda\hat{t}_1(1-\alpha)W^h\left(\frac{\partial L_1^{ch}}{\partial t_1} - W^h L_1^h \frac{\partial L_1^h}{\partial G}\right) = \lambda\left[\alpha W^l L_1^l + (1-\alpha)W^h L_1^h\right], \end{aligned} \tag{A26}$$

$$\begin{aligned} t_2\colon \quad & \alpha W^l L_2^l u'_{al} + (1-\alpha)W^h L_2^h u'_{ah} - \lambda t_2\left[\alpha W^l \frac{dL_2^l}{dt_2} + (1-\alpha)W^h \frac{dL_2^h}{dt_2}\right] \\ & = \lambda\left[\alpha W^l L_2^l + (1-\alpha)W^h L_2^h\right]. \end{aligned} \tag{A27}$$

Dividing (A23) by λ and using the definition of the marginal social evaluation of income, $\theta_1^i \equiv u'_{1i}/\lambda + t_1 W^i(\partial L_1^i/(\partial G)$, we obtain

$$\alpha \cdot \theta_1^l + (1-\alpha) \cdot \theta_1^h = 1. \tag{A28}$$

When the government wishes to redistribute income, we have $\theta_1^l > \theta_1^h$, in which case (A28) implies $\theta_1^h < 1$, as reported in Section V.

To find the optimal effective marginal tax rate on first-period labor income ($\hat{t}_1$), we multiply (A28) by $W^l L_1^l$ and rearrange to obtain

$$\alpha W^l L_1^l \cdot \theta_1^l = W^l L_1^l - (1-\alpha) W^l L_1^l \cdot \theta_1^h. \tag{A29}$$

Noting that $\hat{t}_1 \equiv t_1 - ps$ implies $\partial L_1^c/\partial t_1 = \partial L_1^c/\partial \hat{t}_1 = -W(\partial L_1^c)/(\partial W(1-\hat{t}_1))$, we may rewrite (A26) as

$$\begin{aligned} \alpha W^l L_1^l \cdot \theta_1^l &= \alpha W^l L_1^l + (1-\alpha) W^h L_1^h - (1-\alpha) W^h L_1^h \cdot \theta_1^h \\ &\quad - \left(\frac{\hat{t}_1}{1-\hat{t}_1}\right) \left[\alpha W^l L_1^l \frac{W^l(1-\hat{t}_1)}{L_1^l} \frac{\partial L_1^{cl}}{\partial W^l(1-\hat{t}_1)} \right. \\ &\quad \left. + (1-\alpha) W^h L_1^h \frac{W^h(1-\hat{t}_1)}{L_1^h} \frac{\partial L_1^{ch}}{\partial W^h(1-\hat{t}_1)} \right]. \end{aligned} \tag{A30}$$

Inserting (A29) into (A30), dividing by $W^h L_1^h$ and using the definition of the compensated labor-supply elasticity, $\varepsilon_{1i}^c \equiv (W^i(1-\hat{t}_1))/(L_1^i) \cdot (\partial L_1^{ci})/(\partial W^i(1-\hat{t}_1))$, along with $\beta_1 \equiv W^l L_1^l / W^h L_1^h$, we end up with equation (27) in Section V.

To derive the optimal degree of disability insurance, we start by adding (A25) and (A26) and insert the consumer's first-order condition $u'_{1i} = pu'_{di} + (1-p)u'_{ai}$ to arrive at

$$\begin{aligned} &(1-p)\left[\alpha W^l L_1^l \left(u'_{dl} - u'_{al}\right) + (1-\alpha) W^h L_1^h \left(u'_{dh} - u'_{ah}\right)\right] \\ &+ \lambda \hat{t}_1 \alpha W^l \left[\left(\frac{p - \hat{p}^l}{p}\right) \frac{\partial L_1^{cl}}{\partial t_1} - W^l L_1^l \left(\frac{\partial L_1^l}{\partial G} - \frac{1}{p} \frac{\partial L_1^l}{\partial b} \right) \right] \\ &+ \lambda \hat{t}_1 (1-\alpha) W^h \left[\left(\frac{p - \hat{p}^h}{p}\right) \frac{\partial L_1^{ch}}{\partial t_1} - W^h L_1^h \left(\frac{\partial L_1^h}{\partial G} - \frac{1}{p} \frac{\partial L_1^h}{\partial b} \right) \right] = 0. \end{aligned} \tag{A31}$$

Defining

$$\Omega^i \equiv \frac{1}{(1-p)L_1^i} \left(\frac{\partial L_1^i}{\partial G} - \frac{1}{p} \frac{\partial L_1^i}{\partial b} \right), \tag{A32}$$

noting from (6) and (8) that $(p - \hat{p}^i)/p = ((1-p)(u'_{ai} - u'_{di}))/(u'_{1i})$, and recalling that $\partial L_1^c/\partial t_1 = \partial L_1^c/\partial \hat{t}_1 = -W(\partial L_1^c)/(\partial W(1-\hat{t}_1))$, we may rewrite (A31) as

$$\begin{aligned} &\alpha\beta_1 \left[1 + \frac{\lambda \varepsilon_{1l}^c}{u'_{1l}} \left(\frac{\hat{t}_1}{1-\hat{t}_1} \right) \right] \left(u'_{dl} - u'_{al}\right) + (1-\alpha) \left[1 + \frac{\lambda \varepsilon_{1h}^c}{u'_{1h}} \left(\frac{\hat{t}_1}{1-\hat{t}_1} \right) \right] \left(u'_{dh} - u'_{ah}\right) \\ &= \lambda \hat{t}_1 W^h L^h \left[\alpha \beta_1^2 \Omega^l + (1-\alpha)\Omega^h\right]. \end{aligned} \tag{A33}$$

Subtracting (A24) from (A23), we also find

$$\alpha\left(u'_{dl} - u'_{al}\right) + (1-\alpha)\left(u'_{dh} - u'_{ah}\right) = \lambda \hat{t}_1 W^h L^h \left[\alpha \beta_1 \Omega^l + (1-\alpha)\Omega^h\right]. \tag{A34}$$

Solving (A33) and (A34) for $(u'_{dl} - u'_{al})$ and $(u'_{dh} - u'_{ah})$ and inserting (27) to eliminate $\hat{t}_1/(1-\hat{t}_1)$, we arrive at

$$u'_{dl} - u'_{al} = \left(\frac{\lambda \hat{t}_1 W^h L_1^h}{\alpha \Psi}\right)\left\{\alpha\beta_1 \Omega^l + (1-\alpha)\big(1-\theta_1^h\big)\left(\frac{\lambda \varepsilon_{1h}^c}{u'_{1h}\bar{\varepsilon}_1^c}\right)\big[\alpha\beta_1\Omega^l + (1-\alpha)\Omega^h\big]\right\}, \tag{A35}$$

$$\Psi \equiv 1 + (1-\alpha)\big(1-\theta_1^h\big)\left(\frac{\lambda}{\bar{\varepsilon}_1^c}\right)\left(\frac{\varepsilon_{1h}^c}{u'_{1h}} - \frac{\beta_1 \varepsilon_{1l}^c}{u'_{1l}}\right),$$

$$u'_{dh} - u'_{ah} = \left(\frac{\lambda \hat{t}_1 W^h L_1^h}{\Psi}\right)\left\{\Omega^h - \beta_1\big(1-\theta_1^h\big)\left(\frac{\lambda \varepsilon_{1h}^c}{u'_{1h}\bar{\varepsilon}_1^c}\right)\big[\alpha\beta_1\Omega^l + (1-\alpha)\Omega^h\big]\right\}. \tag{A36}$$

From the definition in (A32), we see that the magnitudes Ω^l and Ω^h in (A35) and (A36) are (proportional to) the relative changes in labor supply induced by the policy reform $dG = -p \cdot db$. Section IV showed that $s > 0$ in the optimum so that (23) implies $\partial L_1^i/\partial G - 1/p \cdot \partial L_1^i/\partial b > 0$. From (A32) we thus have $\Omega^i > 0$. Moreover, since $\theta_1^h < 1$, we see from the definition of Ψ in (A35) that condition (23) in Section V is sufficient (but not necessary) to ensure that $\Psi > 0$. From (A35) and the fact that the optimal value of $\hat{t}_1$ is positive, it then follows that (28) is also sufficient to ensure that the low skilled should be imperfectly insured (i.e. $u'_{dl} - u'_{al} > 0$).

The optimal degree of disability insurance for high-skilled workers is implicitly given by (A36). In the benchmark case in which the labor-supply responses of the two skill groups are symmetric so that $\Omega^l = \Omega^h = \Omega$ and $\varepsilon_{1l}^c = \varepsilon_{1h}^c$, we can use the definitions of θ_1^h and $\bar{\varepsilon}_1^c$ in (27) to rewrite (A36) as

$$u'_{dh} - u'_{ah} = \left(\frac{\lambda \hat{t}_1 W^h L_1^h \Omega}{\Psi}\right)\left\{1 - \left[\frac{\beta_1\big(1-\theta_1^h\big)}{\theta_1^h - \hat{t}_1 (W^h L^h/G)\varepsilon_{1h}^G}\right]\right\}, \quad \varepsilon_{1h}^G \equiv \frac{\partial L_1^h}{\partial G}\frac{G}{L_1^h}, \tag{A37}$$

where ε_{1h}^G is the income elasticity of first-period labor supply.[16] With positive values of $\hat{t}_1$, Ω and Ψ, it follows from (A37) that imperfect insurance of the high skilled (i.e. $u'_{dh} - u'_{ah} > 0$) is optimal iff

$$\theta_1^h - \beta_1\big(1-\theta_1^h\big) - \gamma\varepsilon_{1h}^G\left(\frac{\hat{t}_1}{1-\hat{t}_1}\right) > 0, \quad \gamma \equiv \frac{W^h L^h \big(1-\hat{t}_1\big)}{G}. \tag{A38}$$

Inserting (27) into (A38) to eliminate $\hat{t}_1/(1-\hat{t}_1)$ and rearranging, one ends up with condition (29) in Section V.

To prove the result reported in Section V that $t_2 = 0$ with homogeneous agents, we start by defining the social marginal valuation of second-period income for an able worker (note from (5) that there is no income effect on second-period labor supply, so the definitions below do not include tax base effects):

$$\theta_a^l \equiv \frac{u'_{al}}{\lambda}, \quad \theta_a^h \equiv \frac{u'_{ah}}{\lambda}, \tag{A39}$$

[16] To derive (A37), we use the fact that $u'_{1h}/\lambda = \theta_1^h - \hat{t}_1\big(W^h L_1^h/G\big)\varepsilon_{1h}^G$ and observe from (27) that $\varepsilon_{1h}^c/\bar{\varepsilon}_1^c = \alpha\beta_1 + 1 - \alpha$ when $\varepsilon_{1l}^c = \varepsilon_{1h}^c$.

Using (6) and (A39), we may rewrite (A27) as

$$\frac{t_2}{1-t_2} = \frac{\alpha\beta_2\left(1-\theta_a^l\right)+(1-\alpha)\left(1-\theta_a^h\right)}{\bar{\varepsilon}_2}, \quad \beta_2 \equiv \frac{W^l L_2^l}{W^h L_2^h}, \tag{A40}$$

$$\bar{\varepsilon}_2 \equiv \alpha\beta_2\varepsilon_2^l + (1-\alpha)\varepsilon_2^h, \quad \varepsilon_2^i \equiv \frac{dL_2^i}{dW^i(1-t_2)}\frac{W^i(1-t_2)}{L_2^i}, \quad i=l,h,$$

where ε_2^i is the (compensated and uncompensated) net wage elasticity of second-period labor supply, and $\bar{\varepsilon}_2$ is the corresponding weighted average (compensated and uncompensated) labor-supply elasticity. In the limiting case of equal wages ($W^l = W^h$), we have $L_1^l = L_1^h$ and $L_2^l = L_2^h$ so that $\beta_1 = \beta_2 = 1$. In that case, it follows from (6), (27), (A35) and (A36) that $\hat{t}_1 = 0$ and $u'_{ai} = u'_{di} = u'_{1i}$ which in turn implies

$$\theta_a^l = \theta_1^l, \quad \theta_a^h = \theta_1^h, \tag{A41}$$

given the definitions stated in (27) and (A39). Substituting (A41) and $\beta_2 = 1$ into (A40), we find

$$\frac{t_2}{1-t_2} = \frac{1-\left[\alpha\cdot\theta_1^l + (1-\alpha)\cdot\theta_1^h\right]}{\bar{\varepsilon}_2} = 0, \tag{A42}$$

where the last equality follows from (A28). Thus, the optimal value of t_2 is indeed zero if agents are homogeneous.

References

Albanesi, S. and Sleet, C. (2006), Dynamic Optimal Taxation with Private Information, *Review of Economic Studies 73*, 1–30.

Blomquist, S. and Micheletto, L. (2008), Age-related Optimal Income Taxation, *Scandinavian Journal of Economics 110*, 47–71.

Bovenberg, A. L. and Sørensen, P. B. (2007), Optimal Social Insurance with Redistributive Taxation. Working paper downloadable from www.econ.ku.dk/pbs/.

Brito, D. L., Hamilton, J. H., Slutsky, S. M. and Stiglitz, J. E. (1991), Dynamic Optimal Income Taxation with Government Commitment, *Journal of Public Economics 44*, 15–35.

Corlett, W. J. and Hague, D. C. (1953), Complementarity and the Excess Burden of Taxation, *Review of Economic Studies 21*, 21–30.

Diamond, P. A. (1998), Optimal Income Taxation: An Example with a U-shaped Pattern of Optimal Marginal Tax Rates, *American Economic Review 88*, 83–95.

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

Dillén, M. and Lundholm, M. (1996), Dynamic Income Taxation, Redistribution, and the Rachet Effect, *Journal of Public Economics 59*, 69–93.

Dixit, A. and Sandmo, A. (1977), Some Simplified Formulae for Optimal Income Taxation, *Scandinavian Journal of Economics 79*, 417–423.

Erosa, A. and Gervais, M. (2002), Optimal Taxation in Life-cycle Economies, *Journal of Economic Theory 105*, 338–369.

Gaube, T. (2007), Optimum Taxation of Each Year's Income, *Journal of Public Economic Theory 9*, 127–150.

Golosov, M., Kocherlakota, N. and Tsyvinski, A. (2003), Optimal Indirect and Capital Taxation, *Review of Economic Studies 70*, 569–587.

Greenwald, B. and Stiglitz, J. E. (1986), Externalities in Economies with Imperfect Information and Incomplete Markets, *Quarterly Journal of Economics 101*, 229–264.
Kremer, M. (2001), Should Taxes be Independent of Age?, manuscript, Harvard University.
Ljungqvist, L. and Sargent, T. (2004), *Recursive Macroeconomic Theory*, 2nd edn., MIT Press, Cambridge, MA.
Lozachmeur, J. M. (2006a), Optimal Age-specific Income Taxation, *Journal of Public Economic Theory 8*, 709–723.
Lozachmeur, J. M. (2006b), Disability Insurance and Optimal Income Taxation, *International Tax and Public Finance 13*, 717–732.
Mirrlees, J. A. (1971), An Exploration in the Theory of Optimal Income Taxation, *Review of Economic Studies 38*, 175–208.
Pauly, M. V. (1974), Over Insurance and Public Provision of Insurance. The Roles of Moral Hazard and Adverse Selection, *Quarterly Journal of Economics 88*, 44–62.
Stern, N. B. (1976), On the Specification of Models of Optimum Income Taxation, *Journal of Public Economics 6*, 123–162.
Tuomala, M. (1984), On the Optimal Income Taxation: Some Further Numerical Results, *Journal of Public Economics 23*, 351–366.

First version submitted November 2007;
final version received January 2009.