OPTIMAL SOCIAL INSURANCE
WITH REDISTRIBUTIVE TAXATION

A. Lans Bovenberg, Tilburg University, Netspar, CEPR and CESifo
Peter Birch Sørensen, University of Copenhagen, EPRU and CESifo

Abstract

We study optimal social insurance aimed at insuring disability risk in the presence of linear income taxation. Optimal disability benefits rise with previous earnings so that public transfers depend not only on current earnings but also on earnings in the past. Moreover, disability insurance is incomplete. By offering imperfect insurance and structuring disability benefits so as to enable workers to insure against disability by working harder, social insurance is designed to offset the distortionary impact of the redistributive labor income tax on labor supply.

Keywords: Optimal social insurance; optimal lifetime income taxation.
JEL Code: H21, H55

*) Corresponding author: CentER and Netspar, Tilburg University, P.O. Box 90153, 5000 LE Tilburg, The Netherlands. E-Mail: A.L.Bovenberg@uvt.nl

**) Department of Economics, University of Copenhagen, Studiestraede 6, 1455 Copenhagen K, Denmark. E-mail: Peter.Birch.Sorensen@econ.ku.dk
OPTIMAL SOCIAL INSURANCE
WITH REDISTRIBUTIVE TAXATION

A. Lans Bovenberg and Peter Birch Sørensen

1. 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. For concreteness we label the adverse shock to earnings capacity as ‘disability’, but 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 when the government finds it optimal to ‘insure’ against skill heterogeneity through a redistributive income tax, it becomes optimal to offer less than full insurance against disability even when disability risks are purely exogenous. We thus integrate 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 in which case they receive a special disability benefit. We show that disability benefits should increase with previous income. In this way the government can provide disability insurance to not only the low-skilled but also to the high-skilled, while at the same time improving the first-period labor-supply incentives of the high-skilled. In the presence of distortionary labor taxes aimed at redistribution from the high-skilled to the low-skilled, optimal disability insurance is only imperfect. The reason is that imperfect disability insurance encourages young workers to increase their first-period earnings by working harder. By raising their first-period labor supply, workers can improve their insurance against disability because the

---

1 We wish to thank Thomas Gaube for comments on an earlier draft of this paper. All remaining errors and shortcomings are our own responsibility.
disability benefit rises with previous income. Thus, even though the private market could implement full disability insurance, our analysis shows that such insurance would not be optimal because private insurers would fail to internalize the negative external effects of additional disability insurance on the base of the redistributive labor tax. The government therefore faces an incentive to prevent private insurance companies from fully insuring disability. 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.

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 also 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 compatible with maintaining incentives to work. This result is consistent with the analysis in the present paper.

Just as we do in this paper, Lozachmeur (2006b) extends the model of Diamond and Mirrlees by allowing for two different skill levels. However, unlike us, Lozachmeur focusses on the case of non-linear taxation of labor and capital income and finds that full disability insurance is optimal when 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. We, in contrast, deal with linear labor income taxation and abstract from taxes on capital income. In our framework, the government thus has access to a smaller set of tax instruments. Indeed, in a small open economy, the government may find it difficult to actually observe wealth invested abroad. Furthermore, labor income taxes tend to be piece-wise linear while many countries are moving towards flat taxes on labor income.

One can view our analysis as being complementary to that of Lozachmeur (2006b) in the sense that we explore the robustness of his results with respect to different assump-
tions about the availability of tax 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 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 for 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 since it is only a local result for the highest skill level in case the skill distribution has finite support (see Diamond (1998)). Our key result in this paper that full disability insurance for the high-skilled is not optimal depends crucially on the fact that high-skilled labor supply is distorted: imperfect disability insurance encourages high-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 play a role to alleviate the resulting labor-market distortions. More generally, our formulation in terms of linear taxation yields more clear-cut, unambiguous results on the interaction between redistribution across skills and disability insurance compared to the analysis of non-linear taxation.

Our paper is related to the literature on optimal dynamic taxation. Brito et al. (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. Alabanesi and Sleet (2006) explore optimal non-linear labor taxation with idiosyncratic productivity shocks that are identically and independently distributed. Gaube (2006) 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 tax-
ation. Non-linear age-dependent taxes are advocated by Kremer (2001) on the basis of
distributional arguments. Erosa and Gervais (2002) and Lozachmeur (2006a) defend lin-
ear age-dependent income taxes by 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 has a preference for redistribution.

2. The model

Individuals live for two periods. Everybody is able to work in the first period, but in the
second period individuals face the 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. They also receive a transfer which may be conditioned on previous
earnings and may be differentiated from the disability benefit. We distinguish two skill
groups (the low-skilled and the-high skilled) earning different real wage rates reflecting
exogenous differences in labor productivity. The real interest rate is also exogenous, so
our setting can be viewed as a small open economy with perfect capital mobility.

2.1. Individual behavior

This section describes the behavior of a low-skilled worker; the behavior of the high-
skilled is given by fully analogous relationships. A low-skilled worker’s labor supply in
the first period is \( \ell_1 \), and his consumption during that period is \( C_{1\ell} \). If he is able to work
in the second period, he supplies labor \( \ell_2 \) and consumes an amount \( C_{2\ell}^{d} \). If he becomes
disabled in period 2, his consumption is \( C_{2\ell}^{d} \). His expected lifetime utility \( U \) is given by
the nested utility function

\[
U = U_1 + \delta f \left( E \left[ U_2 \right] \right), \quad f' > 0, \quad (2.1)
\]

\[
U_1 = U_1 \left( C_{1\ell} - g \left( \ell_1 \right) \right), \quad U_1' > 0, \quad U_1'' < 0,
\]

\[
E \left[ U_2 \right] = pu \left( C_{2\ell}^{d} \right) + (1 - p) u \left( C_{2\ell}^{a} - h \left( \ell_2 \right) \right), \quad 0 < p < 1,
\]

\[
u' > 0, \quad u'' < 0, \quad g' > 0, \quad g'' > 0, \quad h' > 0, \quad h'' > 0,
\]
where $U_1(\cdot)$ denotes utility during the first period of life, $\delta$ a discount factor, $E[U_2]$ expected utility during the second period, and $p$ the probability of becoming disabled in the second period. Utility during the first period depends on first-period consumption, adjusted for the disutility of first-period work effort, $g(\ell_1)$. Similarly, for an able worker, the second-period utility $u(C^a_{2t} - h(\ell_2))$ depends on his consumption corrected for the disutility of his second-period labor supply, $h(\ell_2)$. A disabled worker obtains utility $u(C^d_{2t})$. The specification in (2.1) is sufficiently flexible to allow the degree of intertemporal substitutability in consumption to deviate from the reciprocal of the degree of relative risk aversion, as suggested by Epstein and Zin (1989). For later purposes, we define

$$U_{dt}^0 \equiv \frac{1}{\delta p} \frac{\partial U}{\partial C^d_{2t}} = f'(pu(C^d_{2t}) + (1-p)u(C^a_{2t} - h(\ell_2))) \cdot u'(C^d_{2t}) > 0,$$

$$U_{at}^0 \equiv \frac{1}{\delta (1-p)} \frac{\partial U}{\partial C^a_{2t}} = f'(pu(C^a_{2t}) + (1-p)u(C^a_{2t} - h(\ell_2))) \cdot u'(C^a_{2t} - h(\ell_2)) > 0,$$

$$U_{at}'' \equiv \frac{1}{1-p} \frac{\partial U_{at}'}{\partial C^a_{2t}} = f'' \left[ u'(C^a_{2t} - h(\ell_2)) \right]^2 + \frac{f' \cdot u''(C^a_{2t} - h(\ell_2))}{1-p},$$

$$U_{dat}'' \equiv \frac{1}{1-p} \frac{\partial U_{dat}'}{\partial C^d_{2t}} = \frac{f'' \cdot u'(C^d_{2t}) \cdot u'(C^a_{2t} - h(\ell_2))}{1-p}.\quad (2.6)$$

In the special case in which the reciprocal of the intertemporal substitution elasticity coincides with the coefficient of relative risk aversion, $f'' = 0$ so that the (ex ante) marginal utility of disabled consumption does not depend on able consumption (i.e. $U_{dat}'' = 0$). If $f''$ is positive (negative), the degree of risk aversion is greater (smaller) than the inverse of the intertemporal substitution elasticity so that the marginal utility of disabled consumption rises (falls) with able consumption.

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

$$C_{1t} = w(1-t)\ell_1 + G - S^t,$$  

(2.7)

where $w$ represents the real wage rate of a low-skilled worker, $t$ the constant marginal tax rate on labor income, $G$ a lump-sum transfer, and $S^t$ saving of the low-skilled worker. In
the second period, an able worker receives a benefit consisting of a lump-sum component $B$ plus a component amounting to a fraction $s^a$ of his earnings during the first period. We abstract from taxes on capital income, since governments of small open economies face serious difficulties in observing savings invested abroad. With $r$ denoting the real interest rate, an able worker therefore faces the following second-period budget constraint:

$$C^a_{2t} = (1 + r) S^t + w (1 - t) \ell_2 + B + s^a w \ell_1.$$  \hspace{1cm} (2.8)

A disabled worker receives a benefit equal to the constant $b$ plus a fraction $s^d$ of his previous labor income, so that he faces the following second-period budget constraint:\footnote{Of the three lump-sum instruments $G$, $b$, and $B$, one is redundant. For example, we can set the lump-sum benefit for able workers equal to zero (i.e. $B = 0$ ). $s_a$ and $s_d$ are not redundant because the marginal tax rate $t$ is the same in both periods. However, if we differentiate between the marginal tax rate in the first period and the tax rate in the second period, we can set $s_a = 0$ and $B = 0$ so that the government does not have to provide benefits to able workers in the second period of life but can limit itself to providing disability benefits in that period.}

$$C^d_{2t} = (1 + r) S^t + b + s^d w \ell_1.$$  \hspace{1cm} (2.9)

The consumer maximizes (2.1) subject to (2.7) through (2.9). Optimal second-period labor supply implies that the marginal disutility of work equals the marginal after-tax real wage:

$$h' (\ell_2) = w (1 - t).$$  \hspace{1cm} (2.10)

The first-order condition for optimal saving is given by

$$\delta(1 + r) \left[ p U'_{at} + (1 - p) U'_{at} \right] - U'_{at} = 0,$$  \hspace{1cm} (2.11)

where $U'_{at}$ represents the marginal utility of first-period consumption of the low-skilled worker. $U'_{at}$ and $U'_{at}$ are defined in (2.2) and (2.3), respectively.

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

$$[w(1 - t) - g' (\ell_1)] U'_{1t} + \delta w \left[ p s^d U'_{dt} + (1 - p) s^d U'_{at} \right] = 0.$$  \hspace{1cm} (2.12)

Part of the reward for first-period labor supply accrues in the second period if benefits in that period rise with earnings (i.e. $s^a, s^d > 0$). Substituting (2.11) into (2.12) to eliminate $U'_{1t}$, we can write (2.12) as

$$g' (\ell_1) = w(1 - 1_{1t}),$$  \hspace{1cm} (2.13)
where
\[
\hat{t}_{1c} = t - \left( \hat{p}^t s^d + (1 - \hat{p}^t) s^a \right) \frac{1}{1 + r},
\]  
(2.14)
with
\[
\hat{p}^t = \frac{pU_{at}'}{pU_{dt}' + (1 - p)U_{at}'}. 
\]  
(2.15)

The variable \( \hat{p}^t \) can be viewed as the risk-neutral probability of becoming disabled for the low-skilled worker, so that \( \hat{t}_{1c} \) may be interpreted as a risk-adjusted (certainty-equivalent) marginal effective tax rate on first-period labor income for the low-skilled worker. The risk-neutral probabilities differ from real-world probabilities if agents are risk-averse and not perfectly insured (so that \( U_{dt}' \neq U_{at}' \)). For example, if \( s^d > s^a \) and \( U_{dt}' > U_{at}' \), the individual can enhance the insurance against disability risk by raising first-period labor supply. Ex post, the effective marginal tax rate on first-period income for a disabled worker \( \left( t - \frac{\hat{p}^t s^d}{1 + r} \right) \) then differs from the corresponding effective marginal tax rate for an able worker \( \left( t - \frac{s^a}{1 + r} \right) \). By differentiating \( s^d \) from \( s^a \), the government thus makes the marginal tax rate on first-period income depend on second-period income. In other words, marginal and average tax rates depend on lifetime earnings.

For welfare analysis, we employ the consumer’s indirect lifetime utility function which takes the form
\[
V^\ell = V^\ell \left( G, b, B, t, s^d, s^a \right),
\]  
(2.16)
with the derivatives (denoted by subscripts and found by applying the Envelope Theorem):
\[
V^\ell_G = U_{1t}', \quad V^\ell_b = \delta pU_{dt}', \quad V^\ell_B = \delta (1 - p) U_{at}',
\]  
(2.17)
\[
V^\ell_t = -w\ell_1 U_{1t}' - \delta w\ell_2 (1 - p) U_{at}', \quad V^\ell_{s^d} = \delta pwl_1 U_{dt}', \quad V^\ell_{s^a} = \delta (1 - p) w\ell_1 U_{at}'.
\]  
(2.18)

2.2. 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 transfers paid to that generation. This implies that the generational account of each cohort is zero. The high-skilled earn the wage rate \( W > w \), and a high-skilled worker’s labor supply is denoted by
The exogenous fraction of low-skilled individuals in each cohort is $\alpha$. Both skill types face the same probability $p$ of disability in the second period of life. Normalizing the size of the cohort to unity, and using subscripts to indicate time periods, we can write the constraint that a cohort’s generational account must be zero as

\[ \alpha \left[ tw_1 + \left( \frac{1-p}{1+r} \right) (tw_2 - B - s^aw_1) - \left( \frac{p}{1+r} \right) (b + s^d w_1) - G \right] + \]

\[ (1 - \alpha) \left[ tW_1 + \left( \frac{1-p}{1+r} \right) (tW_2 - B - s^aW_1) - \left( \frac{p}{1+r} \right) (b + s^d W_1) - G \right] = 0. \]

(2.19)

Assuming that disability cannot be verified, the government also faces the incentive compatibility constraint that an able worker should have no incentive to mimic a disabled worker. In other words, the second-period utility of a mimicker should be no higher than the second-period utility of a non-mimicker.\(^3\) For low-skilled workers, the resulting non-mimicking constraint is given by

\[ u\left( (1 + r) S^e + w_2 (1 - t) + B + s^aw_1 - h\left( \ell_2 \right) \right) \geq u\left( (1 + r) S^e + b + s^d w_1 \right) \iff Z^e \equiv w_2 (1 - t) - h\left( \ell_2 \right) + B - b + \left( s^a - s^d \right) w_1 \geq 0, \]

(2.20)

and for high-skilled workers the analogous constraint amounts to

\[ Z^h \equiv W_2 (1 - t) - h\left( \ell_2 \right) + B - b + \left( s^a - s^d \right) W_1 \geq 0. \]

(2.21)

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 utility of a low-skilled and that of a high-skilled worker, respectively, we write the utilitarian social welfare function (SWF) as

\[ SWF = \alpha V^l \left( G, b, B, t, s^d, s^a \right) + (1 - \alpha) V^h \left( G, b, B, t, s^d, s^a \right), \]

(2.22)

which must be maximized with respect to the policy instruments $G, b, B, t, s^d, s^a$, subject to the constraints (2.19), (2.20) and (2.21).

\(^3\)Section 3.3 shows that the non-mimicking constraint is typically not binding in the optimum.
3. Optimal social insurance

3.1. Disability insurance rises with previous income

This section demonstrates that the government can generate a Pareto improvement by moving from a situation without any income-dependent disability benefits (i.e. \( s^d = s^a \)) towards disability benefits that rise with previously earned income (i.e. \( s^d > s^a \)). This strongly suggests that the optimal disability benefit should increase with previous income. Hence, second-period transfers are based not only on the earnings in that period, but also on the earnings in the first period.

To prove this result, we start out from a situation with \( s = s^d = s^a \) in which the government has optimized the other policy instruments in a manner respecting the non-mimicking constraints (2.20) and (2.21). In this situation, the non-mimicking constraint for the low-skilled worker becomes binding. The reason is that enhancing disability insurance of the low-skilled (by increasing \( b \) and reducing \( B \) in a balanced-budget manner) does not affect labor-supply incentives if \( s^a = s^d \), as shown by (A.12) and (A.13) in the appendix. In the absence of a trade-off between incentives and insurance, full disability insurance for the low-skilled is optimal. With \( s^d = s^a \), it follows from (2.20) and (2.21) that only the low-skilled can be fully insured against disability, i.e. \( Z^l = 0 \) implies \( Z^h > 0 \) (and hence \( U'_{dh} - U'_{ah} > 0 \)), since \( WL_2 (1 - t) - h (L_2) > w \ell_2 (1 - t) - h (\ell_2) \).

Intuitively, compared to the low-skilled, the high-skilled lose more earnings in case of disability, but receive the same compensation \( b - B \) if \( s^d = s^a \).

Starting from an equilibrium with \( s^d = s^a \), we consider a policy experiment involving an increase in \( s^d \) and a decrease in \( s^a \) calibrated to keep the average subsidy rate \( \tilde{s} \equiv ps^d + (1 - p) s^a \) constant, that is, a policy change satisfying

\[
d\tilde{s} = 0 \quad \implies \quad ds^a = -\left( \frac{p}{1-p} \right) ds^d, \quad ds^d > 0.
\]

At the same time, the government adjusts the policy instrument \( b \) to satisfy the binding non-mimicking constraint (2.20). Recalling that \( s^d = s^a \) initially, and using (3.1) to eliminate \( ds^a \), this requires

\[
-db - w \ell_1 (ds^d - ds^a) = 0 \quad \implies \quad db = -\left( \frac{w \ell_1}{1-p} \right) ds^d.
\]

\(4\) The Envelope Theorem implies that the surpluses \( WL_2 (1 - t) - h (L_2) \) and \( w \ell_2 (1 - t) - h (\ell_2) \) are increasing in the pre-tax wage rate. \( W > w \) thus implies that \( WL_2 (1 - t) - h (L_2) > w \ell_2 (1 - t) - h (\ell_2) \).
Finally, $G$ is adjusted to keep the utility of the low-skilled agents constant, given the policy changes specified in (3.1) and (3.2). Using the expressions for $V^c_G$ and $V^c_b$ given in (2.17), and noting from (2.11) that full insurance implies $U'_{d\ell} = U'_{a\ell} = U'_{1\ell}/\delta (1 + r)$, we find that the required change in $G$ is

$$dG = -\frac{p}{1 + r} db. \quad (3.3)$$

At the same time, the changes in $b, G, s_a,$ and $s_d$ increase the lifetime utility of high-skilled workers, since it follows from (2.17), (2.18) and (3.1) through (3.3) that

$$dV^h = p\delta (WL_1 - W\ell_1) (U'_{dh} - U'_{ah}) \, ds^d > 0. \quad (3.4)$$

If we can prove that public revenue increases, we may thus conclude that our policy experiment creates a Pareto-improvement. Using (2.19), one can easily show that the policy changes described by (3.1) through (3.3) have no direct impact on net government revenue so that the revenue effect of the policy reform depends on labor supply responses.

With a binding non-mimicking constraint (2.20) implying full disability insurance of the low-skilled (i.e. $U'_{d\ell} = U'_{a\ell}$), it follows from (2.14) that the changes in $s^d$ and $s^a$ satisfying (3.1) will not affect the effective tax rate $\tilde{\ell}_{1\ell}$ and hence will not affect $\ell_1$, according to (2.13). Furthermore, since $t$ is unchanged, it follows from (2.10) that also $\ell_2$ and $L_2$ are constant, while (A.12) and (A.14) in the appendix imply that $\frac{\partial L_1}{\partial b} = \frac{\partial L_1}{\partial a} = 0$ when $s^d = s^a$ initially. According to (A.6) and (A.7) in the appendix, the changes in $s^d$ and $s^a$ will affect the first-period labor supply of high-skilled workers in the following manner:

$$\frac{\partial L_1}{\partial s^d} = -\left(\frac{\hat{\rho}^h}{1 + r}\right) \frac{\partial L^c_1}{\partial t}, \quad \frac{\partial L_1}{\partial s^a} = -\left(\frac{1 - \hat{\rho}^h}{1 + r}\right) \frac{\partial L^c_1}{\partial t}, \quad (3.5)$$

where $\frac{\partial L^c_1}{\partial t} < 0$ is the compensated response of first-period high-skilled labor supply to a change in the ordinary tax rate $t$. Using (3.1), (3.5), (2.11), and (2.15), and recalling that $U'_{dh} - U'_{ah} > 0$, we can write the (uncompensated) labor-supply response as

$$dL_1 = \left[\frac{\partial L_1}{\partial s^d} + \frac{\partial L_1}{\partial s^a} \frac{ds^a}{ds^d}\right] ds^d = p\delta \left[-\frac{\partial L^c_1}{\partial t} \left(\frac{U'_{dh} - U'_{ah}}{U'_{1h}}\right)\right] ds^d > 0. \quad (3.6)$$

Thus, high-skilled labor supply expands. Intuitively, when disability benefits are linked more closely to first-period labor effort, high-skilled workers can obtain better disability benefits.

---

5 We use the fact that the derivatives of the indirect utility function of the high-skilled are given by expressions analogous to (2.17) and (2.18).
insurance by working harder. The improved labor-supply incentives benefit the government budget as long as \( t_1 = t - \frac{\bar{h}}{1+r} > 0 \).

Linking disability insurance to first-period labor effort therefore enhances both labor-market incentives and disability insurance for the high-skilled. It improves the trade-off between insurance and incentives. Even without redistributive motives (i.e. \( t_1 = 0 \)), disability insurance should depend on previous income because it provides the possibility to offer better disability insurance to the high-skilled without violating the non-mimicking constraint of the low-skilled. These arguments are strengthened if redistributive taxation distorts labor supply. In that case, income-dependent disability insurance not only improves disability insurance but also alleviates the labor-market distortions generated by redistributive taxation.

3.2. The suboptimality of full insurance

We now proceed to show that full disability insurance of both skill groups can never be optimal, even though separate linear tax schedules for the high-skilled and the low-skilled allow for full insurance. To prove this result, we show that starting from an equilibrium with full insurance of both skill groups, we can design a policy reform that leaves the utility levels of both groups unaffected, while at the same time raising public revenue.

We start by noting that if both skill groups are fully insured (so that the non-mimicking constraints are both met with equality), we may add (2.20) and (2.21) to obtain

\[
(WL_2 - w\ell_2)(1 - t) - [h(L_2) - h(\ell_2)] = (s^d - s^a)(WL_1 - w\ell_1). \tag{3.7}
\]

Since the left-hand side is positive (see footnote 4), and first-period skilled earnings exceed the corresponding unskilled earnings (i.e. \( WL_1 > w\ell_1 \)), this expression implies that \( s^d > s^a \). Intuitively, compared to the low-skilled, high-skilled households face a larger income loss if they become disabled. Hence, if low-skilled agents are fully insured against disability risk, the disability benefit must rise with previous earnings so as to ensure that also the high-skilled agents are not hurt should they become disabled.

We now make disability insurance less than perfect by reducing \( b \) and increasing \( G \). We reduce disability insurance in such a way that the lifetime utility of both households remains constant. Using the expressions for \( V_G^{\ell} \) and \( V_b^{\ell} \) given in (2.17), along with the
analogous expressions for the high-skilled group, and noting from (2.11) that full insurance (i.e., \( U_d = U_d^a \)) implies that \( U_d^* = U_d^a / \delta (1 + r) \) for both skill groups, we find that such a policy reform must satisfy expression (3.3) which ensures that the direct effect (i.e. the impact with a fixed tax base) on the government budget is zero. Further, (2.10) implies that second-period labor supply remains constant because the tax rate \( t \) is unaffected. From the government’s perspective, the effective marginal tax rate on first-period labor income is (see (2.19))

\[
t_1 \equiv t - \frac{\tilde{s}}{1 + r}, \quad \tilde{s} \equiv ps^d + (1 - p) s^a,
\]

where \( \tilde{s} \) denotes the expected second-period subsidy rate on first-period income. With this definition of the first-period marginal tax rate, the overall impact of the policy reform on the government budget (2.19) can be written as \( t_1 w[\alpha d \ell_1 + (1 - \alpha) d L_1] \). The government budget thus improves if the first-period labor supply of both skill types increases (under the assumption \( t_1 > 0 \); (A.29) in the appendix implies that \( t_1 \) is indeed typically positive in the optimum).

Given the relationship between \( dG \) and \( db \) implied by (3.3), first-period labor supply does in fact increase, because section A.2 of the appendix establishes that

\[
\frac{p}{1 + r} \frac{\partial \ell_1}{\partial G} - \frac{\partial \ell_1}{\partial b} > 0 \quad \text{and} \quad \frac{p}{1 + r} \frac{\partial L_1}{\partial G} - \frac{\partial L_1}{\partial b} > 0 \quad \text{for} \quad s^d > s^a.
\]

The improvement of the public budget resulting from the utility-preserving policy reform (3.3) would enable the government to engineer a Pareto improvement, say, by raising \( G \) by more than implied by (3.3). This shows that the starting point characterized by full insurance of both skill groups cannot be a social optimum.

The intuition for this result is the following. 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 \( s^d > s^a \) – 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 \( t_1 > 0 \). To be sure, the Envelope Theorem implies that a small increase in labor supply has no direct first-order effect on private welfare when workers are initially in a private optimum, but since the positive
marginal tax rate drives the marginal productivity of labor above the marginal disutility of work in the initial equilibrium, the rise in labor supply does create a first-order social welfare gain which shows up as an increase in public revenue.

In other words, disability insurance should be less than perfect if the government also wants to insure against skill heterogeneity through a positive labor income tax rate redistributing resources from high-skilled to low-skilled agents. The government thus faces an incentive to prevent private insurance companies from fully insuring disability. This encourages individuals to self-insure through precautionary individual saving and to improve their benefits from public disability insurance through additional work effort when young (if $s^d > s^a$). Hence full disability insurance is not optimal. The reason is that private insurance against disability generates a negative fiscal externality on the base of the distortionary income tax which is used to insure against skill heterogeneity. This result on the optimality of imperfect insurance holds for any arbitrary tax rate $t_1 > 0$, including the optimal tax rate derived in (A.29) in the appendix.

3.3. Conditions for optimal insurance

The previous sub-section 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 fairly weak conditions, the optimal policy does in fact involve imperfect insurance for both skill groups.

Section A.4 in the appendix employs the first-order conditions for the solution to the optimal insurance problem to derive expressions for the marginal utility differentials $U'_d - U'_a$ and $U'_h - U'_a$, assuming that no skill group faces a binding non-mimicking constraint, i.e., that no group is fully insured. If the resulting expressions are positive, the initial assumption of imperfect insurance is validated.

For the low-skilled group, the assumption that no group faces a binding non-mimicking

---

6For the external effects between insurers in the presence of moral hazard, see Pauly (1974) and Greenwald and Stiglitz (1986).
constraint gives rise to the optimality condition (see section A.4 of the appendix)
\[ U_{dt} - U_{at} = (s^d - s^a) \left( \frac{\lambda t_1}{\Psi} \right) \left\{ (1 - \beta_1) w\Omega^t + \left( \frac{\lambda t_1}{1 - t_1} \right) \left( \frac{\epsilon_{1t}^c}{U_{1t}^t} \right) \left[ w\Omega^t + \left( \frac{1 - \alpha}{1 - \alpha} \right) W\Omega^h \right] \right\}, \]
(3.10)

\[ \beta_1 \equiv \frac{w\ell_1}{WL_1}, \quad \epsilon_1 \equiv \alpha\beta_1\epsilon_{1t}^c + (1 - \alpha) \epsilon_{1h}^c, \]
\[ \Psi \equiv 1 - \beta_1 + \left( \frac{\lambda t_1}{1 - t_1} \right) \left( \frac{\epsilon_{1h}^c}{U_{1h}^h} - \beta_1\epsilon_{1t}^c \right), \]
where \( \epsilon_{1t}^c \) and \( \epsilon_{1h}^c \) denote the compensated wage elasticities of second-period labor supply for the low-skilled and the high-skilled, respectively, so that \( \epsilon_{1t}^c > 0 \) is a weighted compensated elasticity of first-period labor supply, and where \( \Omega^t \) and \( \Omega^h \) are positive magnitudes that depend on the risk aversion properties of the utility function (see eq. (A.34) in the appendix). Section 3.1 indicated that the optimal policy involves \( s^d > s^a \).

The expression on the right-hand side of (3.10) is therefore positive if \( \Psi \) is positive. In view of the definition of \( \Psi \), the conditions on \( \epsilon_{1h}^c \) and \( \epsilon_{1t}^c \) for this to be the case are very weak, since \( \beta_1 < 1 \) and \( U_{1t}^t > U_{1h}^h \). Accordingly, the low-skilled are imperfectly insured against disability as long as \( t_1 > 0 \). Redistributive taxation thus makes imperfect disability insurance optimal.

For high-skilled workers, the assumption that no skill group faces a binding non-mimicking constraint implies that (see section A.4 of the appendix)
\[ U_{dh} - U_{ah} = (s^d - s^a) \left( \frac{\lambda t_1}{\Psi} \right) \left\{ (1 - \beta_1) W\Omega^h - \left( \frac{\lambda t_1}{1 - t_1} \right) \left( \frac{\beta_1\epsilon_{1t}^c}{U_{1t}^t} \right) \left[ \left( \frac{\alpha}{1 - \alpha} \right) w\Omega^t + W\Omega^h \right] \right\}, \]
(3.11)

Inserting the optimal tax formula (A.29) from the appendix to eliminate \( \frac{\ell_1}{1 - t_1} \), we obtain
\[ U_{dh} - U_{ah} = (s^d - s^a) \left( \frac{\lambda t_1}{\Psi} \right) (1 - \beta_1) \left\{ W\Omega^h - (1 - \alpha_1^h) \left( \frac{\beta_1\chi_{1t}^c}{U_{1t}^t\epsilon_{1t}^c} \right) \left[ \alpha w\Omega^t + (1 - \alpha) W\Omega^h \right] \right\}, \]
(3.12)

\[ \alpha_1^h \equiv \frac{U_{1h}^h}{\lambda} + t_1 W \frac{\partial L_1}{\partial G}, \]
where \( \alpha_1^h \) measures the marginal social evaluation of first-period income for a high-skilled worker, accounting for the tax-base effect. In the normal case, the government wishes to redistribute income so that \( \alpha_1^h < 1 \).\(^7\) The conditions for the right-hand side of (3.12) to

\(^7\)Equation (A.16) in the appendix implies that marginal social evaluations averaged over the low- and high-skilled is unity: \( \alpha \cdot \alpha^l + (1 - \alpha) \cdot \alpha^h = 1 \), where \( \alpha^l \) is defined analogously to \( \alpha_1^h \): \( \alpha^l \equiv \frac{U_{1t}^t}{\lambda} + t_1 w \frac{\partial G}{\partial \ell} \). If the government wishes to redistribute from the high-skilled to the low-skilled, we must therefore have \( \alpha^l > 1 \) and \( \alpha_1^h < 1 \).
be positive are nevertheless weak, since $W > w$, $1 - \alpha^h_1 \leq 1$, $\beta_1 < 1$, and since $U_{i\ell}/\lambda > 1$ if $\partial\Omega c_1 / \partial G \approx 0$ (see footnote 7). In particular, the condition is met if $\Omega c$ does not greatly exceed $\Omega h$ (implying that imperfect insurance of the low-skilled does not provide much stronger incentives than imperfect insurance of the high-skilled) and inequality is high so that $\beta_1$ is small. Intuitively, high inequality drives up the marginal tax rate, thus distorting labor supply. To offset this distortion, the government finds it optimal to offer only imperfect disability insurance to skilled agents to induce them to work harder in the first period so as to obtain better disability insurance in the second period.

Indeed, equations (3.10) and (3.11) show that full disability insurance is optimal if the government does not employ distortionary taxes to redistribute across skills (i.e. if $t_1 = 0$). Hence, disability insurance is imperfect to the extent that it helps to alleviate the labor-market distortions imposed by redistributive taxation. In the absence of these distortions, the government would structure its public transfers so as to provide full disability insurance to both skills, given that disability risks are exogenous and hence do not give rise to moral hazard.

4. Conclusions

This paper studied optimal taxation and social insurance in an economy where public policy insures (from behind the 'veil of ignorance') both skill heterogeneity and exogenous disability risk. Although the government has sufficient policy instruments to insure both skill groups fully against disability, full disability 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 enable workers to improve their insurance against disability by working harder. 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 the high-skilled to supply labor, thereby alleviating the labor-market distortions caused by redistributive taxation.
This appendix derives the effects of the various policy instruments on individual labor supply 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.

A.1. The effects of taxes and transfers on labor supply

We consider the labor supply of the low-skilled group; the labor supply of high-skilled workers is characterized by completely analogous expressions. For convenience, we drop the subscript \( \ell \) in terms involving derivatives of the utility function. To find the elasticities of first-period labor supply and saving with respect to the policy variables, we totally differentiate (2.11) and (2.12) to arrive at

\[
\begin{pmatrix}
-a_1G - (1 + r)(a_{1b} + a_{1B}) & -a_1G \bar{s}^d w - a_{1b} s^d w - a_{1B} s^a w \\
-a_2G - (1 + r)(a_{2b} + a_{2B}) & -g''(}\ell_1)U_1' - a_2G \bar{s}^d w - a_{2b} s^d w - a_{2B} s^a w
\end{pmatrix} \times \begin{pmatrix}
dS \\
\Delta \ell
\end{pmatrix}
\]

\[
= \begin{pmatrix}
\Delta^S \\
\Delta^\ell
\end{pmatrix},
\]

where

\[
\Delta^S \equiv -a_1G dG + a_{1b} db + a_{1B} dB + (a_{1G} w\ell_1 - a_{1B} w\ell_2) dt + a_{1b} w\ell_1 ds^d + a_{1B} w\ell_1 ds^a,
\]

\[
\Delta^L \equiv -a_2G dG + a_{2b} db + a_{2B} dB + (wU_1' + a_{2G} w\ell_1 - a_{2B} w\ell_2) dt + (a_{2b} w\ell_1 - \delta wp U_0') ds^d + (a_{2B} w\ell_1 - \delta w(1 - p) U_0') ds^a,
\]

\[
\bar{s}^d \equiv \frac{\bar{p}^d s^d + (1 - \bar{p}^d)s^a}{1 + r}, \quad a_{1G} \equiv -U_1'', \quad a_{2G} = -\bar{s} w U_1'',
\]

\[
a_{1b} \equiv -\delta (1 + r)p [p U_d'' + (1 - p) U_d''], \quad a_{2b} \equiv -\delta wp [ps^d U_d'' + (1 - p)s^a U_d''],
\]

\[
a_{1B} \equiv -\delta (1 + r)(1 - p) [p U_d'' + (1 - p) U_d''], \quad a_{2B} \equiv -\delta w(1 - p) [ps^d U_d'' + (1 - p)s^a U_d''].
\]

Applying Cramer’s Rule to the system (A.1), we can find the various labor-supply effects from the system.
\[
\begin{pmatrix}
  \frac{dS}{dt} \\
  \frac{d\ell_1}{dt}
\end{pmatrix} = \frac{1}{\Delta}
\begin{pmatrix}
  -g''(\ell_1)U_1' - a_{2G}s^d w - a_{2b}s^d w - a_{2B}s^a w \\
  a_{2G} + (1 + r)(a_{2b} + a_{2B})
\end{pmatrix}
\begin{pmatrix}
  a_{1G}s^d w + a_{1b}s^d w + a_{1B}s^a w \\
  -a_{1G} - (1 + r)(a_{1b} + a_{1B})
\end{pmatrix}
\times
\begin{pmatrix}
  \Delta s \\
  \Delta t
\end{pmatrix}
\]  

(A.2)

where the determinant \( \Delta \) of the Jacobian is positive because of the second-order condition for individual optimization.

From this solution, we find

\[
\frac{\partial c_1}{\partial t} = \frac{\partial c_1}{\partial s^d} - w \frac{\partial c_1}{\partial G} - w \frac{\partial c_1}{\partial B}
\]

\[
\frac{\partial c_1}{\partial s^d} = \frac{\partial c_1}{\partial s^d} + w \frac{\partial c_1}{\partial b}
\]

\[
\frac{\partial c_1}{\partial s^a} = \frac{\partial c_1}{\partial s^a} + w \frac{\partial c_1}{\partial B}
\]

\[
\frac{\partial c_1}{\partial t} = -\frac{wU_1'[a_{1G} + (1 + r)(a_{1b} + a_{1B})]}{\Delta}
\]

\[
\frac{\partial c_1}{\partial s^d} = -\frac{\delta wU_1'[a_{1G} + (1 + r)(a_{1b} + a_{1B})]}{\Delta}
\]

(A.3)

(A.4)

(A.5)

where the last equality follows by substituting (2.11) to eliminate \( U_1' \) and using (2.15).

Similarly, we find

\[
\frac{\partial c_1}{\partial s^a} = \frac{\delta w(1 - p)U_a'[a_{1G} + (1 + r)(a_{1b} + a_{1B})]}{\Delta}
\]

\[
\frac{\partial c_1}{\partial s^a} = \frac{-\delta(1 - p)U_a'[a_{1G} + (1 + r)(a_{1b} + a_{1B})]}{\Delta}
\]

\[
\frac{\partial c_1}{\partial s^d} = \frac{-\delta pU_1'[a_{1G} + (1 + r)(a_{1b} + a_{1B})]}{\Delta}
\]

(A.6)

(A.7)

(A.8)

(A.9)

(A.10)

so that

\[
\frac{\partial c_1}{\partial G} = (1 + r)\left(\frac{\partial c_1}{\partial b} + \frac{\partial c_1}{\partial B}\right)
\]

(A.11)
Note that with \( s = s^a = s^d \) (so that first-period labor supply does not act as insurance against disability), we have \( a_{2i} = \frac{sw}{(1 + r)}a_{1i}, i = G, b, B \) and thus \( \frac{\partial a_i}{\partial b} = \frac{\partial a_i}{\partial B} = \frac{\partial a_i}{\partial G} = 0. \)

Intuitively, saving rather than labor supply is adjusted to reallocate consumption intertemporally. This is also the intuition behind (A.11): if the consumer receives additional lump-sum income in both states in the second period (i.e. \( db = dB > 0 \)), she will respond in the same way as if that income comes in the first period (discounted properly with \( 1 + r \) so that \( dG = \frac{db}{1 + r} = \frac{dB}{1 + r} \)). The consumer will simply undo reallocation of lump-sum income \( dG = -\frac{db}{1 + r} = -\frac{dB}{1 + r} \) over the life cycle through saving behavior as long as the generational account is not affected.

By substituting the definitions of \( a_{ij} \) into the solutions for the income effects on labor supply, we find:

\[
\frac{\partial \ell_1}{\partial b} = -\frac{\delta w(s^d - s^a)p^2(1 - p)}{\Delta} \left\{ a_{1G} \frac{U''_{ad}U''_{ad}}{pU'_{ad} + (1 - p)U'_{ad}} \left( \frac{U''_{da}}{U''_{ad}} - \frac{U''_{da}}{U''_{ad}} \right) + \right\} \tag{A.12}
\]

\[
\frac{\partial \ell_1}{\partial B} = \frac{\delta w(s^d - s^a)p(1 - p)^2}{\Delta} \left\{ a_{1G} \frac{U''_{ad}U''_{ad}}{pU'_{ad} + (1 - p)U'_{ad}} \left( \frac{U''_{da}}{U''_{ad}} - \frac{U''_{da}}{U''_{ad}} \right) + \right\} \tag{A.13}
\]

\[
\frac{\partial \ell_1}{\partial G} = (1 + r) \left( \frac{\partial \ell_1}{\partial b} + \frac{\partial \ell_1}{\partial B} \right) = \frac{\delta(1 + r)w(s^d - s^a)}{\Delta} a_{1G} p(1 - p)U''_{ad} \left( pU'_{ad} + (1 - p)U'_{ad} \right) \times \left\{ \frac{pU''_{ad}}{U''_{ad}} - \frac{(1 - p)U''_{da}}{U''_{ad}} - U''_{da} - (1 - p)U''_{ad} \right\} \tag{A.14}
\]

From (2.2) through (2.6) one can show that

\[
\frac{U''_{da}}{U''_{ad}} - \frac{U''_{d}}{U''_{ad}} = -\frac{u''(C_{2d})}{pU''(C_{2d})} > 0 \quad \text{and} \quad \frac{U''_{da}}{U''_{ad}} - \frac{U''_{d}}{U''_{ad}} = -\frac{u''(C_{2d} - h(\ell_2))}{(1 - p)u''(C_{2d} - h(\ell_2))} > 0.
\]

Moreover, concavity of the utility function implies that \( U''_{ad}U''_{ad} - (U''_{da})^2 > 0 \). It then follows from (A.12) that a higher transfer to the disabled reduces labor supply if \( s^d > s^a \). Intuitively, labor supply helps to insure disability if \( s^d > s^a \). In that case, more insurance through a higher \( b \) makes labor supply less attractive. Similarly, a higher transfer to the able implies that disability is less well insured, and according to (A.13) labor supply therefore increases to better insure disability (if \( s^d > s^a \) so that labor supply helps to insure disability).
A.2. The suboptimality of full insurance

We may now derive the result stated in eq. (3.9) which was used to demonstrate that full insurance of both skill groups cannot be optimal. From (A.12) through (A.14), we have

\[
\frac{p}{1+r} \frac{\partial c_1}{\partial G} - \frac{\partial c_1}{\partial b} = p \frac{\partial \ell_1}{\partial B} - (1-p) \frac{\partial \ell_1}{\partial b} = \frac{\delta w (s^d - s^a) p^2 (1-p)^2}{\Delta} 
\]

\[
X^\ell = \frac{\partial c_1}{\partial G} \frac{U_a^m}{U_a^0} + \frac{\partial c_1}{\partial b} \frac{U_a^m}{U_a^0} - \frac{U_a^m}{U_a^0} \frac{\lambda t}{(1-p)w} \left[ \frac{U_a^m}{U_a^0} + \frac{U_a^m}{U_a^0} \right],
\]

(A.15)

Using the definitions in (2.2) through (2.6), we find that

\[
X^\ell = - \left[ \frac{u'' (C_2^d - h (\ell_2))}{(1-p)u' (C_2^d - h (\ell_2))} + \frac{u'' (C_2^d)}{pu' (C_2^d)} \right] > 0.
\]

Since concavity of the utility function implies \( U_a^m U_a^m - (U_a^m)^2 > 0 \), it then follows from (A.15) that \( \frac{p}{1+r} \frac{\partial \ell_1}{\partial G} - \frac{\partial \ell_1}{\partial b} > 0 \) for \( s^d > s^a \). A similar result holds for the high-skilled group, as reported in (3.9).

A.3. Optimal taxes and insurance

The optimal tax problem is to maximize the social welfare function (2.22), subject to the constraints (2.19), (2.20) and (2.21). Using (2.17) and (2.18) together with the results (A.3) through (A.7), we may write the first-order conditions for the solution to this problem as follows (where the subscript \( \ell \) (\( h \)) refers to the low skilled (high skilled), the superscript \( c \) indicates a compensated labor supply response, and \( \lambda, \mu^c, \) and \( \mu^h \) are the shadow prices associated with the government budget constraint and the non-mimicking constraints for the low-skilled and the high-skilled, respectively (note that second-period labor supply is not affected by income effects)):\(^8\)

\[
G: \alpha U_{1\ell} + (1-\alpha) U_{1h} + \lambda_1 \left[ \alpha w \frac{\partial \ell_1}{\partial G} + (1-\alpha) W \frac{\partial L_1}{\partial G} \right] = \lambda + \mu^c W (s^d - s^a) \frac{\partial \ell_1}{\partial G} + \mu^h W (s^d - s^a) \frac{\partial L_1}{\partial G},
\]

(A.16)

\(^8\) (A.16), (A.17) and (A.18) are not independent equations. To see this, add (A.17) and (A.18), multiply the result by \((1+r)\), and use (A.11) and (2.11) to arrive at (A.16). The government thus has only two independent lump-sum instruments.
b: \[ \delta p [\alpha U'_{at} + (1 - \alpha) U'_{dh}] + \lambda t_1 \left[ \alpha w \frac{\partial \ell_1}{\partial b} + (1 - \alpha) W \frac{\partial L_1}{\partial b} \right] \]

\[ = \frac{p \lambda}{1 + r} + \mu^f \left[ 1 + w \left( s^d - s^a \right) \frac{\partial \ell_1}{\partial b} \right] + \mu^h \left[ 1 + W \left( s^d - s^a \right) \frac{\partial L_1}{\partial b} \right], \quad (A.17) \]

\[ = \frac{(1 - p) \lambda}{1 + r} - \mu^f \left[ 1 - w \left( s^d - s^a \right) \frac{\partial \ell_1}{\partial b} \right] - \mu^h \left[ 1 - W \left( s^d - s^a \right) \frac{\partial L_1}{\partial b} \right], \quad (A.18) \]

\[ t: \alpha [w \ell_1 U'_{at} + \delta \ell_2 (1 - p) U'_{at}] + (1 - \alpha) [W L_1 U'_{1h} + \delta W L_2 (1 - p) U'_{ah}] \]

\[ = \lambda \left\{ \alpha w \ell_1 + (1 - \alpha) W L_1 + \left( \frac{1 - p}{1 + r} \right) [\alpha w \ell_2 + (1 - \alpha) W L_2] \right\} \]

\[ + \lambda (1 - \alpha) \left[ t_1 W \left( \frac{\partial L_1^c}{\partial t} - W L_1 \frac{\partial L_1}{\partial G} - W L_2 \frac{\partial L_1}{\partial B} \right) + t W \left( \frac{1 - p}{1 + r} \right) \frac{\partial L_2}{\partial t} \right] \]

\[ - \mu^f w \ell_2 - \mu^f w \left( s^d - s^a \right) \left( \frac{\partial \ell_1^c}{\partial t} - w \ell_1 \frac{\partial \ell_1}{\partial G} - w \ell_2 \frac{\partial \ell_1}{\partial B} \right) \]

\[ - \mu^h W L_2 - \mu^h W \left( s^d - s^a \right) \left( \frac{\partial L_1^c}{\partial t} - W L_1 \frac{\partial L_1}{\partial G} - W L_2 \frac{\partial L_1}{\partial B} \right), \quad (A.19) \]

\[ s^d: \delta p [\alpha w \ell_1 U'_{at} + (1 - \alpha) W L_1 U'_{ah}] + \lambda t_1 w \left[ w \ell_1 \frac{\partial \ell_1}{\partial b} - \left( \frac{\hat{p}^h}{1 + r} \right) \frac{\partial \ell_1^c}{\partial t} \right] \]

\[ = \left( \frac{p \lambda}{1 + r} \right) [\alpha w \ell_1 + (1 - \alpha) W L_1] + \mu^f w \ell_1 + \mu^f w \left( s^d - s^a \right) \left[ w \ell_1 \frac{\partial \ell_1}{\partial b} - \left( \frac{\hat{p}^h}{1 + r} \right) \frac{\partial \ell_1^c}{\partial t} \right] \]

\[ + \mu^h W L_1 + \mu^h W \left( s^d - s^a \right) \left[ W L_1 \frac{\partial L_1}{\partial b} - \left( \frac{\hat{p}^h}{1 + r} \right) \frac{\partial L_1^c}{\partial t} \right], \quad (A.20) \]

\[ s^a: \delta (1 - p) [\alpha w \ell_1 U'_{at} + (1 - \alpha) W L_1 U'_{ah}] + \lambda t_1 w \left[ w \ell_1 \frac{\partial \ell_1}{\partial B} - \left( \frac{1 - \hat{p}^h}{1 + r} \right) \frac{\partial \ell_1^c}{\partial t} \right] \]

\[ + \lambda (1 - \alpha) t_1 W \left[ W L_1 \frac{\partial L_1}{\partial B} - \left( \frac{1 - \hat{p}^h}{1 + r} \right) \frac{\partial L_1^c}{\partial t} \right] \]
\[
\left(\frac{(1-p)\lambda}{1+r}\right) \left[\alpha w \ell_1 + (1-\alpha) WL_1\right] - \mu^t w \ell_1 + \mu^t w \left(s^d - s^a\right) \left[w \ell_1 \frac{\partial \ell_1}{\partial B} - \left(1 - \frac{\hat{p}_t^a}{1+r}\right) \frac{\partial \ell_1^c}{\partial t}\right] \\
- \mu^h W L_1 + \mu^h W \left(s^d - s^a\right) \left[W L_1 \frac{\partial L_1}{\partial B} - \left(1 - \frac{\hat{p}_t^b}{1+r}\right) \frac{\partial L_1^c}{\partial t}\right],
\] (A.21)

where \(t_1\) is defined in (3.8). In addition to meeting these first-order conditions, the solution to the optimal tax problem must also satisfy the complementary slackness conditions:

\[
\begin{align*}
\mu^t &\geq 0, \quad Z^t \geq 0, \quad \mu^t Z^t = 0, \quad \text{(A.22)} \\
\mu^h &\geq 0, \quad Z^h \geq 0, \quad \mu^h Z^h = 0. \quad \text{(A.23)}
\end{align*}
\]

To find the optimal marginal tax rate on second-period labor income \((t)\), we start by adding the first-order conditions (A.20) and (A.21), multiplying by \(1 + r\), and using (A.11) and (2.11) (for both households) to obtain

\[
\alpha U'_t w \ell_1 + (1-\alpha) U'_{1h} WL_1 + \lambda t_1 \left[\alpha w \frac{\partial \ell_1}{\partial G} w \ell_1 + (1-\alpha) W \frac{\partial L_1}{\partial G} WL_1\right] \\
- \lambda t_1 \left[\alpha w \frac{\partial \ell_1}{\partial t} + (1-\alpha) W \frac{\partial L_1}{\partial t}\right] = \lambda \left[\alpha w \ell_1 + (1-\alpha) W L_1\right] \\
+ \mu^t w \left(s^d - s^a\right) \left(w \ell_1 \frac{\partial \ell_1}{\partial G} - \frac{\partial \ell_1}{\partial t}\right) + \mu^h W \left(s^d - s^a\right) \left(W L_1 \frac{\partial L_1}{\partial G} - \frac{\partial L_1}{\partial t}\right). \quad \text{(A.24)}
\]

We may now characterize the optimal effective marginal tax rate on first-period labor income \((t_1)\). Multiplying (A.16) by \(w \ell_1\), we obtain

\[
\alpha w \ell_1 \left(U'_t - \lambda + \lambda t_1 w \frac{\partial \ell_1}{\partial G}\right) = - (1-\alpha) w \ell_1 \left(U'_{1h} - \lambda + \lambda t_1 W \frac{\partial L_1}{\partial G}\right) \\
+ w \ell_1 \left(s^d - s^a\right) \left(\mu^t w \frac{\partial \ell_1}{\partial G} + \mu^h W \frac{\partial L_1}{\partial G}\right), \quad \text{(A.25)}
\]

while (A.24) implies

\[
\alpha w \ell_1 \left(U'_t - \lambda + \lambda t_1 w \frac{\partial \ell_1}{\partial G}\right) = \lambda t_1 \left[\alpha w \frac{\partial \ell_1}{\partial t} + (1-\alpha) W \frac{\partial L_1}{\partial t}\right] \\
- (1-\alpha) W L_1 \left(U'_{1h} - \lambda + \lambda t_1 W \frac{\partial L_1}{\partial G}\right) \\
+ \mu^t w \left(s^d - s^a\right) \left(w \ell_1 \frac{\partial \ell_1}{\partial G} - \frac{\partial \ell_1}{\partial t}\right) + \mu^h W \left(s^d - s^a\right) \left(W L_1 \frac{\partial L_1}{\partial G} - \frac{\partial L_1}{\partial t}\right). \quad \text{(A.26)}
\]

Equating the right-hand sides of (A.25) and (A.26), using the facts (from the definition of \(t_1\)) that

\[
\frac{\partial \ell_1^c}{\partial t} = \frac{\partial \ell_1^c}{\partial t_1} = -w \frac{\partial \ell_1}{\partial w (1-t_1)}, \quad \frac{\partial L_1^c}{\partial t} = \frac{\partial L_1^c}{\partial t_1} = -W \frac{\partial L_1}{\partial W (1-t_1)}, \quad \text{(A.27)}
\]
and dividing by \( WL_1 \), we get
\[
(1 - \beta_1) (1 - \alpha) \left( \frac{U'_{lh} - \lambda + \lambda t_1 W}{G} \frac{\partial L_1}{\partial G} \right) = (1 - \beta_1) \mu^h W \left( \frac{s^d}{s^a} \right) \frac{\partial L_1}{\partial G}
\]
\[
+ \left( \frac{s^d - s^a}{1 - t_1} \right) \left( \mu^\ell \beta_1 \varepsilon_{1t}^\ell + \mu^h \varepsilon_{1h}^\ell \right) - \lambda \left( \frac{t_1}{1 - t_1} \right) [\alpha \beta_1 \varepsilon_{1t}^\ell + (1 - \alpha) \varepsilon_{1h}^\ell], \quad (A.28)
\]
where \( \varepsilon_{1t}^\ell \) and \( \varepsilon_{1h}^\ell \) are defined in (3.10) and (3.12) in the main text. When no skill group faces a binding non-mimicking constraint (i.e. \( \mu^\ell = \mu^h = 0 \)), (A.29) collapses to the term in the first line which is a special case of the classical formula for the optimal linear income tax derived by Dixit and Sandmo (1977).

**A.4. The conditions for optimal imperfect insurance**

We now derive the conditions for optimal imperfect insurance stated in eqs. (3.10) and (3.12) which assume that no skill group faces a binding non-mimicking constraint \((\mu^\ell = \mu^h = 0)\) so that no group is fully insured.

Dividing (A.20) by \( p \) and (A.21) by \( 1 - p \), and subtracting the latter equation from the former, we obtain
\[
\delta \alpha \omega \ell_1 (U'_{dl} - U'_{al}) + \delta (1 - \alpha) W L_1 (U'_{dh} - U'_{ah})
\]
\[
+ \lambda \omega t_1 \ell_1 \left[ \frac{1}{p} \frac{\partial \ell_1}{\partial b} - \left( \frac{1}{1 - p} \right) \frac{\partial \ell_1}{\partial B} \right] + \lambda (1 - \alpha) W t_1 W L_1 \left[ \frac{1}{p} \frac{\partial L_1}{\partial b} - \left( \frac{1}{1 - p} \right) \frac{\partial L_1}{\partial B} \right]
\]
\[
+ \lambda \ell_1 w \left[ \frac{1}{1 + r} - \frac{\tilde{p}^d}{p} - \frac{\tilde{p}^h}{p} \right] \frac{\partial \ell_1}{\partial t} + \lambda (1 - \alpha) t_1 W \left[ \frac{1}{1 - p} - \tilde{p}^h \right] \frac{\partial L_1}{\partial t} = 0. \quad (A.30)
\]

From (2.11) and (2.15), we have
\[
\frac{\tilde{p}^d}{p} = \frac{\delta (1 + r)}{U''_{dl}}, \quad \frac{\tilde{p}^h}{p} = \frac{\delta (1 + r) U''_{ah}}{U''_{al}}, \quad i = \ell, h. \quad (A.31)
\]

Inserting (A.31) into (A.30), dividing through by \( \delta W L_1 \), and using (A.27), we may write (A.30) as
\[
\alpha (U'_{dl} - U'_{al}) + (1 - \alpha) (U'_{dh} - U'_{ah}) - (1 - \beta_1) \alpha (U'_{dl} - U'_{al})
\]
\[ +\lambda \left( \frac{t_1}{1-t_1} \right) \left[ \alpha \beta_1 \epsilon_{i\ell} (U_{dt}' - U_{at}') + (1-\alpha) \epsilon_{ih} \left( U_{dh}' - U_{ah}' \right) \right] \]
\[ = \frac{\lambda t_1}{\delta} \left\{ \alpha w \beta_1 \left[ \left( \frac{1}{1-p} \right) \frac{\partial \ell_1}{\partial B} - \frac{1}{p} \frac{\partial \ell_1}{\partial b} \right] + (1-\alpha) W \left[ \left( \frac{1}{1-p} \right) \frac{\partial L_1}{\partial B} - \frac{1}{p} \frac{\partial L_1}{\partial b} \right] \right\}. \quad (A.32) \]
Dividing (A.17) by \( \delta p \) and (A.18) by \( \delta (1-p) \) (recalling that \( \mu^\ell = \mu^h = 0 \)) and subtracting the latter equation from the former, we obtain
\[ \alpha (U_{dt}' - U_{at}') + (1-\alpha) (U_{dh}' - U_{ah}') \]
\[ = \frac{\lambda t_1}{\delta} \left\{ \alpha w \beta_1 \left[ \left( \frac{1}{1-p} \right) \frac{\partial \ell_1}{\partial B} - \frac{1}{p} \frac{\partial \ell_1}{\partial b} \right] + (1-\alpha) W \left[ \left( \frac{1}{1-p} \right) \frac{\partial L_1}{\partial B} - \frac{1}{p} \frac{\partial L_1}{\partial b} \right] \right\}. \quad (A.33) \]
Dividing through by \( p (1-p) \) in (A.15), we find that
\[ \left( \frac{1}{1-p} \right) \frac{\partial \ell_1}{\partial B} - \frac{1}{p} \frac{\partial \ell_1}{\partial b} = \delta (s^d - s^a) \Omega^\ell, \quad (A.34) \]
\[ \Omega^\ell \equiv \frac{wp}{\Delta^\ell} \left[ \alpha \beta_1 \left( \frac{t_1}{1-t_1} \right) + (1-\alpha) (U_{dh}' - U_{ah}') \right] \left[ \left( \frac{1}{1-p} \right) \frac{\partial L_1}{\partial B} - \frac{1}{p} \frac{\partial L_1}{\partial b} \right] > 0, \]
and similarly we have
\[ \left( \frac{1}{1-p} \right) \frac{\partial L_1}{\partial B} - \frac{1}{p} \frac{\partial L_1}{\partial b} = \delta (s^d - s^a) \Omega^h, \quad \Omega^h > 0, \quad (A.35) \]
where \( \Omega^h \) is defined analogously to \( \Omega^\ell \). Using (A.34) and (A.35), we can write (A.32) as
\[ \alpha \beta_1 \left( U_{dt}' - U_{at}' \right) + \lambda \left[ \epsilon_{i\ell} \frac{t_1}{1-t_1} \right] + (1-\alpha) (U_{dh}' - U_{ah}') \left[ 1 + \frac{\epsilon_{ih}}{\ell_{1\ell}} \left( \frac{t_1}{1-t_1} \right) \right] \]
\[ = \lambda t_1 \left[ \frac{s^d - s^a}{\alpha \beta_1 (s^d - s^a) + (1-\alpha) W \Omega^h} \right]. \quad (A.36) \]
In a similar way, we can use (A.34) and (A.35) to rewrite (A.33) as
\[ \alpha (U_{dt}' - U_{at}') + (1-\alpha) (U_{dh}' - U_{ah}') = \lambda t_1 \left[ \frac{s^d - s^a}{\alpha \beta_1 (s^d - s^a) + (1-\alpha) W \Omega^h} \right]. \quad (A.37) \]
Using (A.37) to eliminate \( 1-\alpha \) \( (U_{dh}' - U_{ah}') \) from (A.36), and solving for \( U_{dt}' - U_{dh}' \), we arrive at equation (3.10) in the text. Further, by using (A.37) to eliminate \( \alpha \) \( (U_{dt}' - U_{at}') \) from (A.36) and solving for \( U_{dh}' - U_{ah}' \), we obtain equation (3.11) in the text. To arrive at (3.12), we finally insert (A.29) with \( \mu^\ell = \mu^h = 0 \) into (3.11).
References


