January 2006

OPTIMAL TAXATION AND SOCIAL INSURANCE IN A LIFETIME PERSPECTIVE

A. Lans Bovenberg, Tilburg University, CEPR and CESifo*)

Peter Birch Sørensen, University of Copenhagen, EPRU and CESifo**)

Abstract

Advances in information technology have improved the administrative feasibility of redistribution based on lifetime earnings recorded at the time of retirement. We study optimal lifetime income taxation and social insurance in an economy in which redistributive taxation and social insurance serve to insure (ex ante) against skill heterogeneity as well as disability risk. Optimal disability benefits rise with previous earnings so that public transfers depend not only on current earnings but also on earnings in the past. Hence, lifetime taxation rather than annual taxation is optimal. The optimal tax-transfer system does not provide full disability insurance. 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 lifetime income taxation, optimal social insurance

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 TAXATION AND SOCIAL INSURANCE IN A LIFETIME PERSPECTIVE

A. Lans Bovenberg and Peter Birch Sørensen

1. Introduction

Much of the inequality in the distribution of annual incomes stems from people having different earnings capacities in various stages of the life cycle. Hence, in the presence of well-functioning capital markets enabling consumers to smooth consumption over the life cycle, redistributive taxes and transfers should address inequalities in the distribution of *lifetime* incomes. Yet, in practice, taxes and transfers are mostly conditioned on annual income, with little or no regard to a person's longer-run earnings capacity. The explanation is mainly administrative because governments rarely keep systematic records of the earnings histories of their citizens. Moreover, since a person's lifetime labor earnings are not fully known until the time he or she retires, the authorities cannot base taxes and transfers on lifetime income. However, it *is* possible to condition public retirement benefits on a person's previous earnings. The effective marginal and average tax rate on income earned earlier in life thus becomes dependent on earnings in other periods of life. In fact, retirement benefits in many countries do to some extent depend on previous earnings. Moreover, with modern information and communication technologies, information on individual earnings histories becomes much easier to gather and store. The question whether an optimal tax-transfer system should exploit information on lifetime earnings therefore becomes relevant.

This paper addresses this issue. In particular, we study whether social insurance benefits aimed at compensating for a loss of earnings capacity should depend on previous labor income. Although for the sake of concreteness we label the shock to earnings capacity as disability, our analysis applies also to other types of idiosyncratic shocks to human capital. In our model, people participate in the labor market for two periods, but some people become disabled in the second period. The government wants to redistribute 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. In the late stage of life, able individuals receive an ordinary retirement benefit, while the disabled collect a special disability benefit. Both types of benefits may be conditioned on previous earnings.

We show that the optimal disability benefit should increase more strongly with previous income than the ordinary retirement benefit. 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. By thus basing second-period transfers on first-period earnings, the optimal tax-transfer system involves lifetime taxation rather than annual taxation. 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 labor supply, workers can improve their insurance against disability because the disability benefit increases more strongly with previous income than the ordinary retirement benefit collected by able workers. Our analysis thus shows that full disability insurance is not optimal. Thus, even though the private market could implement full disability insurance (since moral hazard is absent in our model), this would not be optimal because private insurers would fail to internalize the external effects of additional disability insurance on the base of the redistributive labor tax. The government thus faces an incentive to prevent private insurance companies from fully insuring disability. Indeed, a mix of a public tax-transfer system offering less than full insurance and self insurance through precautionary saving is optimal.

The optimal tax literature has considered linear as well as non-linear tax systems. Real-world tax systems are typically piece-wise linear. In fact, recent decades have witnessed a trend towards more linearity, as governments have flattened their tax schedules and reduced the number of income brackets to simplify the tax system. Against this background, we consider a linear tax-transfer system with a constant marginal tax rate. However, by tying social insurance benefits to previous earnings, the policy maker in our model can differentiate the effective marginal tax rate on labor income according to lifetime earnings capacity. Our analysis shows that it is indeed optimal to exploit opportunities for such differentiation.

The literature on lifetime income taxation is quite sparse. Vickrey (1939, 1947) made early contributions to the normative theory of lifetime income taxation. He was concerned about the overtaxation of fluctuating as opposed to stable incomes under a progressive annual income tax with a marginal tax rate that rises with the level of income. Vickrey therefore proposed an income-averaging scheme in which annual income taxes are in fact collected as a form of withholding for lifetime income tax calculations that are completed only upon death.

Diamond (2003, ch. 3 and 4) analyzes lifetime income taxation in a two-period setting, but without allowing for early retirement due to disability. He finds that the optimal non-linear lifetime income tax tends to imply greater equality of consumption levels among retirees than among workers, assuming that the elderly tend to be more risk averse than younger people. Intuitively, when the marginal utility of consumption declines faster for the elderly, the social planner is more eager to avoid inequality of consumption opportunities among the elderly than among younger people.

A paper more closely related to the present one 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. In some respects, the model of Diamond and Mirrlees (op.cit.) is more general than the one presented here, since they allow for a fully non-linear tax scheme (including a capital income tax). However, whereas Diamond and Mirrlees assume that all able workers feature the same productivity, we allow for different skill levels. In our model, the government thus employs its redistributive policy instruments to 'insure' against not only skill heterogeneity but also disability risk. We thus integrate the conventional analysis of optimal redistributive taxation with the analysis of optimal social insurance. Moreover, by employing Epstein-Zin preferences (see Epstein and Zin (1989)), we are able to provide a detailed characterization of the optimal tax and subsidy rates.

Recent contributions to the literature on social insurance based on mandatory individual savings accounts also consider redistribution policy in a lifetime perspective (see, e.g., Fölster (1997, 1999), Orszag and Snower (1997), Feldstein and Altman (1998), Fölster et al. (2002), Stiglitz and Yun (2002), Sørensen (2003) and Bovenberg and Sørensen (2004)). These papers analyze policy schemes in which workers must contribute a fraction of their earnings to an individual savings account that is debited when the owner draws certain social insurance benefits. At the time of retirement, any surplus on the account is converted into an annuity and added to the ordinary public retirement benefit. If the account is negative, the owner is still guaranteed a minimum public pension. Bovenberg and Sørensen (op.cit.) show that the introduction of such a system as a supplement to the conventional tax-transfer system improves the equity-efficiency trade-off by reducing the distortionary impact of those taxes and transfers that mainly serve to redistribute income over the individual's own lifecycle.

Mandatory individual savings accounts for social insurance introduce an element of lifetime income taxation by effectively conditioning retirement benefits on the individual's prior labor market performance. Intertemporally optimizing agents who are able to accumulate a surplus on their account at the time of retirement face reduced marginal tax rates on labor effort. Individuals who end up with a surplus on their accounts – and who will therefore face stronger incentives to supply labor – tend to be concentrated in the low-risk segments of the working population. This is in contrast to the optimal tax-transfer system in the economy modelled here, where people who end up with a relatively low lifetime income due to disability actually face a lower marginal effective tax rate on labor income earned early in life. The contradiction is only superficial, however. The system of mandatory savings accounts is designed for social insurance benefits that involve a significant risk of moral hazard and relatively little redistribution from high to low lifetime incomes (as opposed to redistribution over the lifecycle). The present paper, however, focuses on optimal redistribution of lifetime incomes in a setting with exogenous idiosyncratic shocks to human capital. In any case, the individual accounts considered by Bovenberg and Sørensen (2004) and the social insurance scheme analyzed here are based on the same fundamental principle: net benefits received at a later stage in life vary positively with labor income earned earlier in life so as to reduce the distortions to labor supply caused by a redistributive tax-transfer system.

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. The leisure consumed by able workers in period 2 may be interpreted as time voluntarily spent in retirement. Larger second-period labor supply can thus be viewed as a higher retirement age. The government transfer collected by able workers in the second period corresponds to an ordinary retirement benefit. Also this benefit may be conditioned on previous earnings, and it 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. Also the real interest rate is exogenous. Indeed, our economy 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 ℓ_1, and his consumption during that period is $C_{1\ell}$. If he is able to work in the second period, he supplies labor ℓ_2 and consumes an amount $C_{2\ell}^{a}$. 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

$$U = U_1\left(C_{1\ell} - g\left(\ell_1\right)\right) + \delta f\left(E\left[U_2\right]\right), \qquad U_1' > 0, \quad U_1'' < 0, \quad f' > 0, \tag{2.1}$$

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

$$g' > 0, \qquad g'' > 0, \qquad h' > 0, \qquad h'' > 0,$$

where $U_1(\cdot)$ denotes utility during the first period of life, δ 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 $u(C_2^a - h(\ell_2))$ depends on his consumption corrected for the disutility of his second-period labor supply, $h(\ell_2)$. A disabled worker obtains utility $u\left(C_2^d\right)$. 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

$$U'_{d\ell} \equiv \frac{1}{\delta}\frac{1}{p}\frac{\partial U}{\partial C_{2\ell}^d} = f'\left(pu\left(C_{2\ell}^d\right) + (1-p)\,u\left(C_{2\ell}^a - h(\ell_2)\right)\right)\cdot u'\left(C_{2\ell}^d\right) > 0, \tag{2.2}$$

$$U'_{a\ell} \equiv \frac{1}{\delta}\frac{1}{(1-p)}\frac{\partial U}{\partial C_{2\ell}^a} = f'\left(pu\left(C_{2\ell}^d\right) + (1-p)\,u\left(C_{2\ell}^a - h(\ell_2)\right)\right)\cdot u'\left(C_{2\ell}^a - h(\ell_2)\right) > 0. \tag{2.3}$$

$$U''_{d\ell} \equiv \frac{1}{p}\frac{\partial U'_{d\ell}}{\partial C_{2\ell}^d} = f''\cdot\left[u'\left(C_{2\ell}^d\right)\right]^2 + \frac{f'\cdot u''\left(C_{2\ell}^d\right)}{p}, \tag{2.4}$$

$$U''_{a\ell} \equiv \frac{1}{1-p}\frac{\partial U'_{a\ell}}{\partial C_{2\ell}^a} = f''\cdot\left[u'\left(C_{2\ell}^a - h(\ell_2)\right)\right]^2 + \frac{f'\cdot u''\left(C_{2\ell}^a - h(\ell_2)\right)}{1-p}, \tag{2.5}$$

$$U''_{da\ell} \equiv \frac{1}{p}\frac{\partial U'_{a\ell}}{\partial C_{2\ell}^d} = \frac{1}{1-p}\frac{\partial U'_{d\ell}}{\partial C_{2\ell}^a} = f''\cdot u'\left(C_{2\ell}^d\right)\cdot u'\left(C_{2\ell}^a - h(\ell_2)\right). \tag{2.6}$$

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. $U''_{da\ell}=0$). f'' is positive (negative) if 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

$$C_{1\ell} = w(1-t)\,\ell_1 + G - S^\ell, \tag{2.7}$$

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^ℓ 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.

With r denoting the real interest rate, an able worker therefore faces the following second-period budget constraint:

$$C_{2\ell}^{a} = (1+r)\, S^{\ell} + w\,(1-t)\,\ell_2 + B + s^{a} w \ell_1. \tag{2.8}$$

A disabled worker receives a benefit equal to the constant b plus a fraction s^d of his previous labor income, so his second-period budget constraint is:

$$C_{2\ell}^{d} = (1+r)\, S^{\ell} + b + s^{d} w \ell_1. \tag{2.9}$$

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:

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

The first-order condition for optimal saving is given by

$$\delta(1+r)\left[pU_{d\ell}^{'} + (1-p)U_{a\ell}^{'}\right] - U_{1\ell}^{'} = 0, \tag{2.11}$$

where $U_{1\ell}^{'}$ represents the marginal utility of first-period consumption of the low-skilled worker. $U_{d\ell}^{'}$ and $U_{a\ell}^{'}$ are defined in (2.2) and (2.3), respectively.

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

$$\left[w(1-t) - g'(\ell_1)\right] U_{1\ell}^{'} + \delta w\left[ps^{d}U_{d\ell}^{'} + (1-p)s^{a}U_{a\ell}^{'}\right] = 0. \tag{2.12}$$

Part of the benefit of first-period labor supply accrues in the second period if disability and retirement benefits rise with earnings (i.e. $s^a, s^d > 0$). Substituting (2.11) into (2.12) to eliminate $U_{1\ell}^{'}$, we can write (2.12) as

$$w(1-\hat{t}_{1\ell}) = g'(\ell_1), \tag{2.13}$$

where

$$\hat{t}_{1\ell} = t - \left(\frac{\hat{p}^{\ell}s^{d} + (1-\hat{p}^{\ell})s^{a}}{1+r}\right), \tag{2.14}$$

with

$$\hat{p}^{\ell} = \frac{pU_{d\ell}^{'}}{pU_{d\ell}^{'} + (1-p)U_{a\ell}^{'}}. \tag{2.15}$$

The variable $\hat{p}^{\ell}$ can be viewed as the risk-neutral probability of becoming disabled for the low-skilled worker, so that $\hat{t}_{1\ell}$ 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 $U'_{d\ell} \neq U'_{a\ell}$). If, for example, $s^d > s^a$ and $U'_{d\ell} > U'_{a\ell}$, 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 $\left(t - \frac{s^d}{1+r}\right)$ 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. A key issue addressed in this paper is whether such lifetime income taxation $\left(s^d \neq s^a\right)$ is in fact optimal and if so, which factors determine the optimal gap between s^d and s^a.

For welfare analysis, we employ the consumer's indirect lifetime utility function, which exhibits the form

$$V^\ell = V^\ell \left(G, b, B, t, s^d, s^a\right), \tag{2.16}$$

with the derivatives (denoted by subscripts and found by applying the Envelope Theorem):

$$V_G^\ell = U'_{1\ell}, \qquad V_b^\ell = \delta p U'_{d\ell}, \qquad V_B^\ell = \delta \left(1 - p\right) U'_{a\ell}, \tag{2.17}$$

$$V_t^\ell = -w\ell_1 U'_{1\ell} - \delta w \ell_2 \left(1 - p\right) U'_{a\ell}, \quad V_{s^d}^\ell = \delta p w \ell_1 U'_{d\ell}, \quad V_{s^a}^\ell = \delta \left(1 - p\right) w \ell_1 U'_{a\ell}. \tag{2.18}$$

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 are paid the wage rate $W > w$, and a high-skilled worker's labor supply is denoted by 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, and using subscripts to indicate time periods, we can write the constraint that a cohort's generational account must be zero as

$$\alpha \overbrace{\left[tw\ell_1 + \left(\frac{1-p}{1+r}\right)\left(tw\ell_2 - B - s^a w\ell_1\right) - \left(\frac{p}{1+r}\right)\left(b + s^d w\ell_1\right) - G\right]}^{\text{generational account of a low-skilled worker}} +$$

$$(1-\alpha)\overbrace{\left[tWL_1+\left(\frac{1-p}{1+r}\right)(tWL_2-B-s^aWL_1)-\left(\frac{p}{1+r}\right)\left(b+s^dWL_1\right)-G\right]}^{\text{generational account of a high-skilled worker}}=0. \tag{2.19}$$

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.[1] For low-skilled workers, the resulting non-mimicking constraint is given by

$$u\left((1+r)S^\ell+w\ell_2(1-t)+B+s^aw\ell_1-h(\ell_2)\right)\geq u\left((1+r)S^\ell+b+s^dw\ell_1\right) \iff$$

$$Z^\ell\equiv w\ell_2(1-t)-h(\ell_2)+B-b+\left(s^a-s^d\right)w\ell_1\geq 0, \tag{2.20}$$

and for high-skilled workers the analogous constraint amounts to

$$Z^h\equiv WL_2(1-t)-h(L_2)+B-b+\left(s^a-s^d\right)WL_1\geq 0. \tag{2.21}$$

The government maximizes the utilitarian sum of expected lifetime utilities. 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

$$SWF=\alpha V^\ell\left(G,b,B,t,s^d,s^a\right)+(1-\alpha)V^h\left(G,b,B,t,s^d,s^a\right), \tag{2.22}$$

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. Optimal taxation and social insurance

3.1. The optimality of social insurance through lifetime income taxation

The first-order conditions for the solution to the policy problem stated in the previous section are given in section A.3 of the appendix. Before exploring the implications of these optimality conditions, we demonstrate that a lifetime income tax, rather than an annual income tax, is optimal. In particular, the government can generate a Pareto improvement

[1]Sub-section 3.4 shows that the non-mimicking constraint is typically met in the optimum.

by moving from a conventional tax-transfer system based on annual incomes only (i.e. $s^d = s^a$) towards lifetime income taxation with $s^d > s^a$. Indeed, with $s^d > s^a$, the ex-post effective marginal tax rate on first-period labor income depends on lifetime earnings capacity. Moreover, second-period transfers are based not only on the earnings in that period, but also on the earnings in the first period. Hence, the government implements lifetime income taxation.

To prove these results, we start out from a situation with annual income taxation ($s = s^d = s^a$), where the government has optimized the other policy instruments in a manner respecting the non-mimicking constraints (2.20) and (2.21). With annual income taxation, it is optimal to increase b and to reduce B in a balanced-budget manner such that the non-mimicking constraint for the low-skilled worker becomes binding. The reason is that enhancing disability insurance in this way does not affect labor-supply incentives if $s^a = s^d$ (since (A.12) and (A.13) in the appendix imply that labor supply does not respond to b and B with annual income taxation). In the absence of a trade-off between incentives and insurance, full disability insurance for the low-skilled is optimal. With $s^d = s^a$, only the low-skilled can be fully insured against disability (i.e. $Z^\ell \geq 0$ implies $Z^h > 0$ (and hence $U'_{dh} - U'_{ah} > 0$), since $WL_2\,(1-t) - h\,(L_2) > w\ell_2\,(1-t) - h\,(\ell_2)$).[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 annual taxation, we consider a policy experiment involving an increase in s^d and a decrease in s^a calibrated so as to keep the average subsidy rate $\widetilde{s} \equiv ps^d + (1-p)\,s^a$ constant, that is, a policy change satisfying

$$d\widetilde{s} = 0 \qquad \Longrightarrow \qquad ds^a = -\left(\frac{p}{1-p}\right) ds^d, \qquad ds^d > 0. \tag{3.1}$$

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

$$-db - w\ell_1\left(ds^d - ds^a\right) = 0 \qquad \Longrightarrow \qquad db = -\left(\frac{w\ell_1}{1-p}\right) ds^d. \tag{3.2}$$

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_G^ℓ and V_b^ℓ given in

[2] 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 $WL_2\,(1-t) - h\,(L_2) > w\ell_2\,(1-t) - h\,(\ell_2)$.

(2.17), and noting from (2.11) that full insurance implies that $U'_{d\ell} = U'_{a\ell} = U'_{1\ell}/\delta\,(1+r)$, we find that the required change in G is

$$dG = -\frac{p}{1+r}db. \tag{3.3}$$

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 only on labor supply responses. With a binding non-mimicking constraint (2.20) (and thus full disability insurance of the low-skilled (i.e. $U'_{d\ell} = U'_{a\ell}$)), (2.14) implies that the changes in s^d and s^a satisfying (3.1) will not affect the effective tax rate $\widehat{t}_{1\ell}$ and hence will not affect ℓ_1, according to (2.13). Furthermore, since t is unchanged, it follows from (2.10) that also ℓ_2 and L_2 are constant, while (A.12) and (A.14) in the appendix imply that $\frac{\partial L_1}{\partial b} = \frac{\partial L_1}{\partial G} = 0$ when $s^d = s^a$. 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:

$$\frac{\partial L_1}{\partial s^d} = -\left(\frac{\widehat{p}^h}{1+r}\right)\frac{\partial L_1^c}{\partial t}, \qquad \frac{\partial L_1}{\partial s^a} = -\left(\frac{1-\widehat{p}^h}{1+r}\right)\frac{\partial L_1^c}{\partial t}, \tag{3.4}$$

where $\frac{\partial L_1^c}{\partial t} < 0$ is the *compensated* response of first-period high-skilled labor supply to a change in the ordinary tax rate t. Using (3.1), (2.11), and (2.15), and recalling that $U'_{dh} - U'_{ah} > 0$, we can write the (uncompensated) labor-supply response as

$$dL_1 = \left[\frac{\partial L_1}{\partial s^d} + \frac{\partial L_1}{\partial s^a}\frac{ds^a}{ds^d}\right]ds^d = p\delta\left[\left(-\frac{\partial L_1^c}{\partial t}\right)\left(\frac{U'_{dh} - U'_{ah}}{U'_{1h}}\right)\right]ds^d > 0. \tag{3.5}$$

Thus, high-skilled labor supply expands. Intuitively, when disability insurance is linked more closely to first-period labor effort, high-skilled workers can enhance their disability insurance by working harder. The improved labor-supply incentives benefit the government budget as long as $t_1 = t - \frac{s}{1+r} > 0$.

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[3]

$$dV^h = p\delta(WL_1 - w\ell_1)\left(U'_{dh} - U'_{ah}\right)ds^d > 0. \tag{3.6}$$

We conclude that moving from annual to lifetime taxation in this way enhances both labor-market incentives and disability insurance for the high-skilled. Lifetime taxation

[3] 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).

thus improves the trade-off between insurance and incentives. Even without redistributive motives (i.e. $t_1 = 0$), lifetime income taxation dominates annual taxation because of the possibility to offer better disability insurance for the high-skilled without violating the non-mimicking constraint for the low-skilled. These arguments are strengthened if redistributive taxation distorts labor supply. In that case, lifetime taxation not only improves disability insurance, but also alleviates the labor-market distortions imposed 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 2), 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 more with earnings than the retirement benefit does, 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^ℓ and V_b^ℓ given in (2.17), along with the analogous expressions for the high-skilled group, and noting from (2.11) that full insurance (i.e., $U_d' = U_a'$) implies that $U_d' = U_1'/\delta\,(1+r)$ for both skill groups, we find that such a policy reform must satisfy expression (3.3). From the government's perspective, the effective marginal tax rate on first-period labor income is (see (2.19))

$$t_1 \equiv t - \frac{\widetilde{s}}{1+r}, \qquad \widetilde{s} \equiv ps^d + (1-p)\, s^a, \tag{3.8}$$

where $\widetilde{s}$ 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 $t_1 w[\alpha d\ell_1 + (1-\alpha)dL_1]$. While (3.3) ensures that the direct effect on the budget is zero, (2.10) implies that second-period labor supply remains constant because the tax rate t is unaffected. The government budget thus improves if the first-period labor supply of both skill types increases (under the assumption $t_1 > 0$; sub-section 3.3 below shows that t_1 is indeed typically positive in the optimum). Given the relationship between dG and db implied by (3.3), labor supply does in fact increase, because section A.2 of the appendix establishes that

$$\frac{p}{1+r}\frac{\partial \ell_1}{\partial G} - \frac{\partial \ell_1}{\partial b} > 0 \quad \text{and} \quad \frac{p}{1+r}\frac{\partial L_1}{\partial G} - \frac{\partial L_1}{\partial b} > 0 \quad \text{for} \quad s^d > s^a. \tag{3.9}$$

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, whereas the expansion of the labor income tax base generates a first-order welfare gain if $t_1 > 0$. 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 $s^d > s^a$). Although we do not model moral hazard in disability insurance, full insurance is thus not optimal. The reason is that private insurance against disability generates a negative fiscal externality on the base of the distortionary tax offering insurance against skill heterogeneity. With endogenous labor supply and lack of public information on individual work effort, this public insurance of skill heterogeneity does generate moral hazard.[4]

3.3. The optimal marginal tax rates

Expressions for the optimal (effective) marginal income tax rates are derived in section A.3 of the appendix. If the high-skilled are less than fully insured against disability, the optimal marginal tax rate on second-period labor income is given by[5]

$$\frac{t}{1-t} = \frac{(1-\beta_2)\left(1-\alpha_2^h\right)(1-\alpha)}{\overline{\varepsilon}_2}, \tag{3.10}$$

$$\beta_2 \equiv \frac{w\ell_2}{WL_2}, \quad \overline{\varepsilon}_2 \equiv \alpha\beta_2\varepsilon_{2\ell} + (1-\alpha)\,\varepsilon_{2h}, \quad \alpha_2^h \equiv \frac{\delta\,(1+r)\,U'_{ah}}{\lambda} + t_1 W\left(\frac{1+r}{1-p}\right)\frac{\partial L_1}{\partial B}.$$

The variable α_2^h in (3.10) measures the marginal social valuation of second-period income for an able high-skilled worker (accounting for the impact on the public budget through the induced income effect on labor supply). ε_2^ℓ and ε_2^h denote the wage elasticities of second-period labor supply for the low-skilled and the high-skilled, respectively, so that $\overline{\varepsilon}_2$ is a weighted elasticity of second-period labor supply. $1-\beta_2$ measures the degree of inequality in the distribution of second-period pre-tax labor income. The optimal value of t depends only on variables relating to the second period. The reason is that first-period labor supply is determined by $\widehat{t_1}$ rather than t. By varying s^d and s^a, the government can manipulate $\widehat{t_1}$ independently from t (see (2.13) and (2.14)).

The optimal effective marginal tax rate on first-period labor income (defined in (3.8)) is given by an analogous expression if both skill groups are less than fully insured against

[4]For the external effects between insurers in the presence of moral hazard, see Pauly (1974) and Greenwald and Stiglitz (1986).

[5]The next sub-section shows that the conditions for both skill groups to be less than fully insured in the optimum are weak. If the non-mimicking constraint for the high-skilled would nevertheless be binding, we must define $\alpha_2^h \equiv \frac{\delta(1+r)U'_{ah}}{\lambda} + t_1 W\left(\frac{1+r}{1-p}\right)\frac{\partial L_1}{\partial B} + \frac{\mu^h}{1-\alpha}\left(\frac{1+r}{1-p}\right)\left(1 - W(s^d - s^a)\frac{\partial L_1}{\partial B}\right)$, where μ^h is the shadow price associated with the non-mimicking constraint for the high-skilled. All other expressions are unaffected.

disability:[6]

$$\frac{t_1}{1-t_1} = \frac{(1-\beta_1)\left(1-\alpha_1^h\right)(1-\alpha)}{\overline{\varepsilon}_1^c}, \tag{3.11}$$

$$\beta_1 \equiv \frac{w\ell_1}{WL_1}, \quad \overline{\varepsilon}_1^c \equiv \alpha\beta_1\varepsilon_{1\ell}^c + (1-\alpha)\,\varepsilon_{1h}^c, \quad \alpha_1^h \equiv \frac{U'_{1h}}{\lambda} + t_1 W \frac{\partial L_1}{\partial G}.$$

The inequality in the distribution of first-period labor income enters through the variable β_1. During the first period, both income and substitution effects affect labor supply. Nevertheless, the optimal marginal tax rate depends only on substitution effects, captured by the weighted average ($\overline{\varepsilon}_1^c$) of the compensated skill-specific labor supply elasticities, $\varepsilon_{1\ell}^c$ and ε_{1h}^c. The variable α_1^h measures the marginal social evaluation of first-period income for a high-skilled worker taking the tax-base effect into account.

In the normal case, the government wishes to redistribute income so that $\alpha_i^h < 1$, $i = 1, 2$.[7] (3.10) and (3.11) then imply that the optimal marginal tax rates are positive. Moreover, *ceteris paribus* the elasticities and the marginal social evaluations, these optimal tax rates increase with the degree of inequality in the distribution of pre-tax income. Furthermore, a larger fraction of high-skilled workers in the labor force $1-\alpha$ broadens the base for redistribution, making it worthwhile to impose a higher marginal tax rate.[8]

According to (3.10) and (3.11), the government typically wants to impose different marginal effective tax rates on income in the two periods by choosing a non-zero value of

[6] As already mentioned, the next sub-section shows that the conditions are weak for both skill groups to be less than fully insured in the optimum.

[7] Expressions (A.16) and (A.18) in the appendix imply that the marginal social evaluation averaged over the low- and high-skilled is unity: $\alpha \cdot \alpha_i^\ell + (1-\alpha)\cdot\alpha_i^h = 1$ (where α_i^ℓ is defined analogously as α_i^h: $\alpha_1^\ell \equiv \frac{U'_{1\ell}}{\lambda} + t_1 w \frac{\partial \ell_1}{\partial G}$ and $\alpha_2^\ell \equiv \frac{\delta(1+r)U'_{a\ell}}{\lambda} + t_1 w \left(\frac{1+r}{1-p}\right)\frac{\partial \ell_1}{\partial B}$).

[8] Although derived in an intertemporal context, the formulas (3.10) and (3.11) are closely related to the formula for the optimal linear income tax obtained by Dixit and Sandmo (1977) for the case with many skill groups in a one-period setting. In the Dixit-Sandmo world, the optimal marginal tax rate on labor income is given by

$$\frac{t}{1-t} = -\frac{\operatorname{cov}\left[\alpha^i, w^i L^i\right]}{E\left(w^i L^i \varepsilon_i^c\right)}$$

where $\operatorname{cov}\left[\alpha^i, w^i L^i\right]$ is the covariance between the marginal social evaluation of income for skill group i (accounting for the impact on the public budget via the induced income effect on labor supply) and the pre-tax labor income $w^i L^i$ of that skill group, and $E\left(w^i L^i \varepsilon_i^c\right)$ is the income-weighted average compensated labor supply elasticity across skill groups. In fact, (3.10) and (3.11) can be written in this form by using expressions (A.16) and (A.18) in the Appendix, which imply that the marginal social evaluation averaged over the low- and high-skilled is unity (see the previous footnote).

the average subsidy rate $\widetilde{s}$ (since $t_1 \equiv t - \frac{\widetilde{s}}{1+r}$, while β_1, α_1^h and $\overline{\varepsilon}_1^c$ generally differ from β_2, α_2^h and $\overline{\varepsilon}_2$). *Ceteris paribus* β_i and α_i^h; $i = 1, 2$, if the labor supply of older workers is more wage elastic than that of younger workers (i.e $\varepsilon_2 > \varepsilon_1$), efficiency considerations cause the optimal t to be below the optimal t_1. *Ceteris paribus* the elasticities and the marginal social evaluations, α_i^h, distributional considerations reinforce this tendency if first-period labor income is more unequally distributed than second-period labor income (i.e. $\beta_1 < \beta_2$).

3.4. The optimal level of social insurance

The previous sub-section assumed that neither the low-skilled nor the high-skilled were fully insured. This sub-section states the conditions under which imperfect insurance of both skill groups is indeed optimal. Section A.4 in the appendix employs the first-order conditions for the solution to the optimal tax problem to derive expressions for the marginal utility differentials $U'_{d\ell} - U'_{a\ell}$ and $U'_{dh} - U'_{ah}$, 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, this validates the initial assumption of imperfect insurance.

For the low-skilled group, the assumption that no group faces a binding non-mimicking constraint gives rise to (see section A.4 of the appendix)

$$U'_{d\ell} - U'_{a\ell} = \left(s^d - s^a\right)\left(\frac{\lambda t_1}{\Psi}\right)\left\{(1-\beta_1)\, w\Omega^\ell + \left(\frac{\lambda t_1}{1-t_1}\right)\left(\frac{\varepsilon^c_{1h}}{U'_{1h}}\right)\left[w\Omega^\ell + \left(\frac{1-\alpha}{\alpha}\right) W\Omega^h\right]\right\}, \tag{3.12}$$

$$\Psi \equiv 1 - \beta_1 + \left(\frac{\lambda t_1}{1-t_1}\right)\left(\frac{\varepsilon^c_{1h}}{U'_{1h}} - \frac{\beta_1 \varepsilon^c_{1\ell}}{U'_{1\ell}}\right),$$

where Ω^ℓ and Ω^h are positive magnitudes that depend on the properties of the utility function (see eq. (A.37) in the appendix). Sub-section 3.1 demonstrated that the optimal policy involves $s^d > s^a$. The expression on the right-hand side of (3.12) is therefore positive if Ψ is positive. In view of the definition of Ψ, the conditions on ε^c_{1h} and $\varepsilon^c_{1\ell}$ for this to be the case are very weak, since $\beta_1 < 1$ and $U'_{1\ell} > U'_{1h}$. 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)

$$U'_{dh}-U'_{ah}=\left(s^{d}-s^{a}\right)\left(\frac{\lambda t_1}{\Psi}\right)\left\{\left(1-\beta_1\right)W\Omega^{h}-\left(\frac{\lambda t_1}{1-t_1}\right)\left(\frac{\beta_1\varepsilon^{c}_{1\ell}}{U'_{1\ell}}\right)\left[\left(\frac{\alpha}{1-\alpha}\right)w\Omega^{\ell}+W\Omega^{h}\right]\right\}. \tag{3.13}$$

Inserting (3.11) into (3.13) to eliminate $\frac{t_1}{1-t_1}$, we obtain

$$U'_{dh}-U'_{ah}=\left(s^{d}-s^{a}\right)\left(\frac{\lambda t_1}{\Psi}\right)\left(1-\beta_1\right)\left\{W\Omega^{h}-\left(1-\alpha^{h}_{1}\right)\left(\frac{\beta_1\lambda\varepsilon^{c}_{1\ell}}{U'_{1\ell}\bar{\varepsilon}^{c}_{1}}\right)\left[\alpha w\Omega^{\ell}+\left(1-\alpha\right)W\Omega^{h}\right]\right\}. \tag{3.14}$$

The conditions for the right-hand side of (3.14) to be positive are weak, since $W > w$, $1-\alpha^{h}_{1} \leq 1$, and $U'_{1\ell}/\lambda > 1$ (if $\frac{\partial L_1}{\partial G} \approx 0$). In particular, the condition is met if Ω^{ℓ} does not greatly exceed Ω^{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 β_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 in order to induce these agents to work harder in the first period so as to obtain better disability insurance in the second period. Indeed, equations (3.12) and (3.13) show that full disability insurance is optimal if the government does not employ distortionary taxes to redistribute across skills (i.e. if $t_1 = 0$ because $\beta_1 = 1$, $\alpha^{h}_{1} = 1$, or $\alpha = 1$). 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.

4. Concluding remarks

This paper studied optimal lifetime income 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 at its disposal sufficient policy instruments to insure both skill groups fully against disability, and even though moral hazard in disability is absent, full disability insurance is not optimal. Instead, by offering imperfect insurance and structuring disability benefits so as to enable workers to improve their insurance against disability by working harder, the government can alleviate the distortionary impact of the redistributive labor income tax. Specifically, optimal disability insurance should allow disability benefits to vary positively with previous earnings. Hence, the effective marginal tax rate depends on the taxpayer's lifetime earnings capacity, and redistribution is based on lifetime incomes. Lifetime taxation improves the trade-off between insurance and incentives. It provides better disability insurance for the high-skilled and enhances their incentives to supply labor, thereby alleviating the labor-market distortions imposed by redistributive taxation.

To allow a detailed characterization of the optimal tax and subsidy rates, we have restricted the analysis to a linear tax-transfer system with certain non-linear elements. We did not study the potential second-best role of capital income taxation in the overall tax-transfer system. Since precautionary saving allows people to partly insure against shocks to their human capital, the government may choose to distort saving. In future work we plan to extend the analysis to a fully non-linear tax system that also allows for capital income taxation distorting saving behavior.

Technical Appendix

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 tax 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 ℓ 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

$$\begin{pmatrix} -a_{1G}-(1+r)(a_{1b}+a_{1B}) & -a_{1G}\bar{s}^{\ell}w-a_{1b}s^{d}w-a_{1B}s^{a}w \\ -a_{2G}-(1+r)(a_{2b}+a_{2B}) & -g''(\ell_1)U_1'-a_{2G}\bar{s}^{\ell}w-a_{2b}s^{d}w-a_{2B}s^{a}w \end{pmatrix} \times \begin{pmatrix} dS \\ d\ell_1 \end{pmatrix}$$
$$= \begin{pmatrix} \Delta^S \\ \Delta^{\ell} \end{pmatrix}, \tag{A.1}$$

where

$$\Delta^S \equiv -a_{1G}dG+a_{1b}db+a_{1B}dB+(a_{1G}w\ell_1-a_{1B}w\ell_2)dt+a_{1b}w\ell_1ds^d+a_{1B}w\ell_1ds^a,$$

$$\begin{aligned} \Delta^L \quad &\equiv \quad -a_{2G}dG+a_{2b}db+a_{2B}dB+(wU_1'+a_{2G}w\ell_1-a_{2B}w\ell_2)dt \\ &\quad +(a_{2b}w\ell_1-\delta wpU_d')ds^d+(a_{2B}w\ell_1-\delta w(1-p)U_a')ds^a, \end{aligned}$$

$$\bar{s}^{\ell} \equiv \frac{\hat{p}^{\ell}s^d+(1-\hat{p}^{\ell})s^a}{1+r}, \qquad a_{1G} \equiv -U_1'', \qquad a_{2G} = -\bar{s}wU_1'',$$

$$a_{1b} \equiv -\delta(1+r)p\left[pU_d''+(1-p)U_{da}''\right], \qquad a_{2b} \equiv -\delta wp\left[ps^dU_d''+(1-p)s^aU_{da}''\right],$$

$$a_{1B} \equiv -\delta(1+r)(1-p)\left[pU_{da}''+(1-p)U_a''\right], \quad a_{2B} \equiv -\delta w(1-p)\left[ps^dU_{da}''+(1-p)s^aU_a''\right].$$

Applying Cramer's Rule to the system (A.1), we can find the various labor-supply effects from the system

$$
\begin{pmatrix} dS \\ d\ell_1 \end{pmatrix}
= \frac{1}{\Delta} \begin{pmatrix} -g''(\ell_1)U_1' - a_{2G}\bar{s}^\ell w - a_{2b}s^d w - a_{2B}s^a w & a_{1G}\bar{s}^\ell w + a_{1b}s^d w + a_{1B}s^a w \\ a_{2G} + (1+r)(a_{2b} + a_{2B}) & -a_{1G} - (1+r)(a_{1b} + a_{1B}) \end{pmatrix}
\times \begin{pmatrix} \Delta^S \\ \Delta^\ell \end{pmatrix} \tag{A.2}
$$

where the determinant Δ of the Jacobian is positive because of the second-order condition for individual optimization.

From this solution, we find

$$
\frac{\partial \ell_1}{\partial t} = \frac{\partial \ell_1^c}{\partial t} - w\ell_1 \frac{\partial \ell_1}{\partial G} - w\ell_2 \frac{\partial \ell_1}{\partial B}
$$

$$
\frac{\partial \ell_1}{\partial s_d} = \frac{\partial \ell_1^c}{\partial s_d} + w\ell_1 \frac{\partial \ell_1}{\partial b} \tag{A.3}
$$

$$
\frac{\partial \ell_1}{\partial s_a} = \frac{\partial \ell_1^c}{\partial s_a} + w\ell_1 \frac{\partial \ell_1}{\partial B} \tag{A.4}
$$

$$
\frac{\partial \ell_1^c}{\partial t} = -\frac{wU_1'[a_{1G} + (1+r)(a_{1b} + a_{1B})]}{\Delta} \tag{A.5}
$$

$$
\frac{\partial \ell_1^c}{\partial s^d} = \frac{\delta wpU_d'[a_{1G} + (1+r)(a_{1b} + a_{1B})]}{\Delta} = -\frac{\delta pU_d'}{U_1'} \frac{\partial \ell_1^c}{\partial t} = -\frac{\hat{p}^\ell}{1+r} \frac{\partial \ell_1^c}{\partial t}, \tag{A.6}
$$

where the last equality follows by substituting (2.11) to eliminate U_1' and using (2.15). Similarly, we find

$$
\frac{\partial \ell_1^c}{\partial s^a} = \frac{\delta w(1-p)U_a'[a_{1G} + (1+r)(a_{1b} + a_{1B})]}{\Delta} = -\frac{\delta(1-p)U_a'}{U_1'} \frac{\partial \ell_1^c}{\partial t} = -\frac{1-\hat{p}^\ell}{1+r} \frac{\partial \ell_1^c}{\partial t}, \tag{A.7}
$$

while the various income effects are given by

$$
\frac{\partial \ell_1}{\partial G} = \frac{(1+r)[a_{2G}(a_{1b} + a_{1B}) - a_{1G}(a_{2b} + a_{2B})]}{\Delta}, \tag{A.8}
$$

$$
\frac{\partial \ell_1}{\partial b} = \frac{a_{2G}a_{1b} - a_{1G}a_{2b} + (1+r)[a_{2B}a_{1b} - a_{1B}a_{2b}]}{\Delta}, \tag{A.9}
$$

$$
\frac{\partial \ell_1}{\partial B} = \frac{a_{2G}a_{1B} - a_{1G}a_{2B} - (1+r)[a_{2B}a_{1b} - a_{1B}a_{2b}]}{\Delta}, \tag{A.10}
$$

so that

$$
\frac{\partial \ell_1}{\partial G} = (1+r)\left(\frac{\partial \ell_1}{\partial b} + \frac{\partial \ell_1}{\partial B}\right). \tag{A.11}
$$

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 $\frac{\partial \ell_1}{\partial G} = \frac{\partial \ell_1}{\partial b} = \frac{\partial \ell_1}{\partial B} = 0$. 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:

$$\frac{\partial \ell_1}{\partial b} = -\frac{\delta w(s^d - s^a) p^2 (1-p)}{\Delta} \left\{ \begin{array}{c} a_{1G} \frac{U_a' U_d'}{(pU_d' + (1-p) U_a')} \left(\frac{U_{da}''}{U_a'} - \frac{U_d''}{U_d'} \right) + \\ \delta (1+r)^2 (1-p) \left[U_a'' U_d'' - (U_{da}'')^2 \right] \end{array} \right\} \tag{A.12}$$

$$\frac{\partial \ell_1}{\partial B} = \frac{\delta w(s^d - s^a) p (1-p)^2}{\Delta} \left\{ \begin{array}{c} a_{1G} \frac{U_a' U_d'}{(pU_d' + (1-p) U_a')} \left(\frac{U_{da}''}{U_d'} - \frac{U_a''}{U_a'} \right) + \\ \delta (1+r)^2 p \left[U_a'' U_d'' - (U_{da}'')^2 \right] \end{array} \right\} \tag{A.13}$$

$$\begin{aligned} \frac{\partial \ell_1}{\partial G} &= (1+r) \left(\frac{d\ell_1}{db} + \frac{d\ell_1}{dB} \right) \\ &= \frac{\delta (1+r) w (s^d - s^a)}{\Delta} \frac{a_{1G} p (1-p) U_a' U_d'}{(pU_d' + (1-p) U_a')} \\ &\quad \times \left\{ \frac{pU_d''}{U_d'} - \frac{(1-p) U_a''}{U_a'} - U_{da}'' \left[\frac{p}{U_a'} - \frac{(1-p)}{U_d'} \right] \right\} \end{aligned} \tag{A.14}$$

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

$$\frac{U_{da}''}{U_a'} - \frac{U_d''}{U_d'} = -\frac{u''\left(C_{2\ell}^d\right)}{pu'\left(C_{2\ell}^d\right)} > 0 \quad \text{and} \quad \frac{U_{da}''}{U_d'} - \frac{U_a''}{U_a'} = -\frac{u''\left(C_{2\ell}^a - h\left(\ell_2\right)\right)}{(1-p) u'\left(C_{2\ell}^a - h\left(\ell_2\right)\right)} > 0.$$

Moreover, concavity of the utility function implies that $U_a'' U_d'' - (U_{da}'')^2 > 0$. It then follows from (A.12) that a higher transfer to the disabled (b) 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 (B) 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). Note that there are two terms in the expressions for $\frac{d\ell_1}{db}$ and $\frac{d\ell_1}{dB}$. The term including a_{1G} depends on intertemporal substitution (and also on risk aversion), while the other term (including $U_a''U_d'' - (U_{da}'')^2$) depends only on risk aversion. With higher b, the consumer wants to spread the welfare gain to the able state if risk aversion is positive (this is the term with $U_a''U_d'' - (U_{da}'')^2$) and to the first period (via increased first-period consumption of leisure as well as material goods) if intertemporal substitution is finite. The latter effect is captured by the term with a_{1G}, which is positive only if risk aversion is correspondingly positive; otherwise, the consumer can better reallocate resources to the first period through dissaving rather than by lowering first-period labor supply.

A higher first-period transfer G depresses first-period labor supply if higher income boosts utility (especially in the disabled state ($U_d' > U_a'$)) and consumption in the two states are complements (i.e. $U_{da}'' > 0$ because risk aversion exceeds the inverse of intertemporal substitution), and the intertemporal substitution elasticity is finite (i.e., $a_{1G} > 0$). If U_{da}''=0, a higher transfer may actually raise first-period labor supply if additional second-period income especially leads to a rapid fall in utility in ability (i.e., $(-U_a'')$ is large compared to $(-U_d'')$) so that it becomes attractive to reallocate income to the disabled state. Note that the sign of the income effect on labor supply is different from normal. This is because labor supply has an insurance function.

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

$$\frac{p}{1+r}\frac{\partial \ell_1}{\partial G} - \frac{\partial \ell_1}{\partial b} = p\frac{\partial \ell_1}{\partial B} - (1-p)\frac{\partial \ell_1}{\partial b}$$

$$= \frac{\delta w(s^d - s^a)p^2(1-p)^2}{\Delta}\left\{\frac{a_{1G}U_a'U_d'X^\ell}{pU_d' + (1-p)U_a'} + \delta(1+r)^2\left[U_a''U_d'' - (U_{da}'')^2\right]\right\}, \qquad \text{(A.15)}$$

$$X^\ell \equiv \frac{U_{da}''}{U_d'} + \frac{U_{da}''}{U_a'} - \frac{U_d''}{U_d'} - \frac{U_a''}{U_a'}.$$

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

$$X^\ell = -\left[\frac{u''\left(C_{2\ell}^a - h\left(\ell_2\right)\right)}{(1-p)u'\left(C_{2\ell}^a - h\left(\ell_2\right)\right)} + \frac{u''\left(C_{2\ell}^d\right)}{pu'\left(C_{2\ell}^d\right)}\right] > 0.$$

Since concavity of the utility function implies $U_a''U_d'' - (U_{da}'')^2 > 0$, 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. The optimal labor income tax rates

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 ℓ (h) refers to the low-skilled (high-skilled), the superscript c indicates a compensated labor supply response, and λ, μ^ℓ, and μ^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)):[9]

$$G: \quad \alpha U'_{1\ell} + (1-\alpha) U'_{1h} + \lambda t_1 \left[\alpha w \frac{\partial \ell_1}{\partial G} + (1-\alpha) W \frac{\partial L_1}{\partial G} \right]$$

$$= \lambda + \mu^\ell w \left(s^d - s^a\right) \frac{\partial \ell_1}{\partial G} + \mu^h W \left(s^d - s^a\right) \frac{\partial L_1}{\partial G}, \tag{A.16}$$

$$b: \quad \delta p \left[\alpha U'_{d\ell} + (1-\alpha) U'_{dh}\right] + \lambda t_1 \left[\alpha w \frac{\partial \ell_1}{\partial b} + (1-\alpha) W \frac{\partial L_1}{\partial b} \right]$$

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

$$B: \quad \delta (1-p) \left[\alpha U'_{a\ell} + (1-\alpha) U'_{ah}\right] + \lambda t_1 \left[\alpha w \frac{\partial \ell_1}{\partial B} + (1-\alpha) W \frac{\partial L_1}{\partial B} \right]$$

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

$$t: \quad \alpha \left[w\ell_1 U'_{1\ell} + \delta w \ell_2 (1-p) U'_{a\ell}\right] + (1-\alpha)\left[W L_1 U'_{1h} + \delta W L_2 (1-p) U'_{ah}\right]$$

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

$$+ \lambda \alpha \left[t_1 w \left(\frac{\partial \ell_1^c}{\partial t} - w\ell_1 \frac{\partial \ell_1}{\partial G} - w \ell_2 \frac{\partial \ell_1}{\partial B}\right) + t w \left(\frac{1-p}{1+r}\right) \frac{\partial \ell_2}{\partial t} \right]$$

[9] (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.

$$+\lambda(1-\alpha)\left[t_1 W\left(\frac{\partial L_1^c}{\partial t}-WL_1\frac{\partial L_1}{\partial G}-WL_2\frac{\partial L_1}{\partial B}\right)+tW\left(\frac{1-p}{1+r}\right)\frac{\partial L_2}{\partial t}\right]$$
$$-\mu^\ell w\ell_2-\mu^\ell w\left(s^d-s^a\right)\left(\frac{\partial \ell_1^c}{\partial t}-w\ell_1\frac{\partial \ell_1}{\partial G}-w\ell_2\frac{\partial \ell_1}{\partial B}\right)$$
$$-\mu^h WL_2-\mu^h W\left(s^d-s^a\right)\left(\frac{\partial L_1^c}{\partial t}-WL_1\frac{\partial L_1}{\partial G}-WL_2\frac{\partial L_1}{\partial B}\right), \tag{A.19}$$

$$s^d\!:\quad \delta p\left[\alpha w\ell_1 U'_{d\ell}+(1-\alpha)WL_1U'_{dh}\right]+\lambda\alpha t_1 w\left[w\ell_1\frac{\partial \ell_1}{\partial b}-\left(\frac{\widehat{p}^\ell}{1+r}\right)\frac{\partial \ell_1^c}{\partial t}\right]$$
$$+\lambda(1-\alpha)t_1W\left[WL_1\frac{\partial L_1}{\partial b}-\left(\frac{\widehat{p}^h}{1+r}\right)\frac{\partial L_1^c}{\partial t}\right]$$
$$=\left(\frac{p\lambda}{1+r}\right)\left[\alpha w\ell_1+(1-\alpha)WL_1\right]+\mu^\ell w\ell_1+\mu^\ell w\left(s^d-s^a\right)\left[w\ell_1\frac{\partial \ell_1}{\partial b}-\left(\frac{\widehat{p}^\ell}{1+r}\right)\frac{\partial \ell_1^c}{\partial t}\right]$$
$$+\mu^h WL_1+\mu^h W\left(s^d-s^a\right)\left[WL_1\frac{\partial L_1}{\partial b}-\left(\frac{\widehat{p}^h}{1+r}\right)\frac{\partial L_1^c}{\partial t}\right], \tag{A.20}$$

$$s^a\!:\quad \delta(1-p)\left[\alpha w\ell_1 U'_{a\ell}+(1-\alpha)WL_1U'_{ah}\right]+\lambda\alpha t_1 w\left[w\ell_1\frac{\partial \ell_1}{\partial B}-\left(\frac{1-\widehat{p}^\ell}{1+r}\right)\frac{\partial \ell_1^c}{\partial t}\right]$$
$$+\lambda(1-\alpha)t_1W\left[WL_1\frac{\partial L_1}{\partial B}-\left(\frac{1-\widehat{p}^h}{1+r}\right)\frac{\partial L_1^c}{\partial t}\right]$$
$$=\left(\frac{(1-p)\lambda}{1+r}\right)\left[\alpha w\ell_1+(1-\alpha)WL_1\right]-\mu^\ell w\ell_1+\mu^\ell w\left(s^d-s^a\right)\left[w\ell_1\frac{\partial \ell_1}{\partial B}-\left(\frac{1-\widehat{p}^\ell}{1+r}\right)\frac{\partial \ell_1^c}{\partial t}\right]$$
$$-\mu^h WL_1+\mu^h W\left(s^d-s^a\right)\left[WL_1\frac{\partial L_1}{\partial B}-\left(\frac{1-\widehat{p}^h}{1+r}\right)\frac{\partial L_1^c}{\partial t}\right], \tag{A.21}$$

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:

$$\mu^\ell \geq 0, \qquad Z^\ell \geq 0, \qquad \mu^\ell Z^\ell = 0, \tag{A.22}$$

$$\mu^h \geq 0, \qquad Z^h \geq 0, \qquad \mu^h Z^h = 0. \tag{A.23}$$

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

$$\alpha U'_{1\ell}w\ell_1+(1-\alpha)U'_{1h}WL_1+\lambda t_1\left[\alpha w\frac{\partial \ell_1}{\partial G}w\ell_1+(1-\alpha)W\frac{\partial L_1}{\partial G}WL_1\right]$$
$$-\lambda t_1\left[\alpha w\frac{\partial \ell_1^c}{\partial t}+(1-\alpha)W\frac{\partial L_1^c}{\partial t}\right]=\lambda\left[\alpha w\ell_1+(1-\alpha)WL_1\right]$$

$$+\mu^{\ell} w\left(s^{d}-s^{a}\right)\left(w\ell_1\frac{\partial \ell_1}{\partial G}-\frac{\partial \ell_1^c}{\partial t}\right)+\mu^{h} W\left(s^{d}-s^{a}\right)\left(WL_1\frac{\partial L_1}{\partial G}-\frac{\partial L_1^c}{\partial t}\right). \tag{A.24}$$

Now insert (A.24) into (A.19) to find

$$\begin{aligned}
&\alpha w\ell_2\left[\delta\left(1-p\right)U'_{a\ell}-\lambda\left(\frac{1-p}{1+r}\right)+\lambda t_1 w\frac{\partial \ell_1}{\partial B}\right] \\
&+\left(1-\alpha\right)WL_2\left[\delta\left(1-p\right)U'_{ah}-\lambda\left(\frac{1-p}{1+r}\right)+\lambda t_1 W\frac{\partial L_1}{\partial B}\right] \\
&=-\lambda\left(\frac{1-p}{1+r}\right)\left(\frac{t}{1-t}\right)\left[\alpha w\ell_2\varepsilon_2^{\ell}+\left(1-\alpha\right)WL_2\varepsilon_2^{h}\right] \\
&-\mu^{\ell}w\ell_2\left[1-w\left(s^d-s^a\right)\frac{\partial \ell_1}{\partial B}\right]-\mu^{h}WL_2\left[1-W\left(s^d-s^a\right)\frac{\partial L_1}{\partial B}\right],
\end{aligned} \tag{A.25}$$

$$\varepsilon_2^{\ell}\equiv\frac{\partial \ell_2}{\partial w\left(1-t\right)}\frac{w\left(1-t\right)}{\ell_2}, \qquad \varepsilon_2^{h}\equiv\frac{\partial L_2}{\partial W\left(1-t\right)}\frac{W\left(1-t\right)}{L_2}.$$

Multiplying (A.18) by $w\ell_2$, we obtain

$$\begin{aligned}
&\alpha w\ell_2\left[\delta\left(1-p\right)U'_{a\ell}-\lambda\left(\frac{1-p}{1+r}\right)+\lambda t_1 w\frac{\partial \ell_1}{\partial B}\right] \\
&=-\left(1-\alpha\right)w\ell_2\left[\delta\left(1-p\right)U'_{ah}-\lambda\left(\frac{1-p}{1+r}\right)+\lambda t_1 W\frac{\partial L_1}{\partial B}\right]-w\ell_2\left(\mu^{\ell}+\mu^{h}\right) \\
&+w\ell_2\left(s^d-s^a\right)\left(\mu^{\ell}w\frac{\partial \ell_1}{\partial B}+\mu^{h}W\frac{\partial L_1}{\partial B}\right).
\end{aligned} \tag{A.26}$$

Substituting (A.26) into (A.25) to eliminate $U'_{a\ell}$, dividing through by $\frac{\lambda WL_2(1-\alpha)(1-p)}{1+r}$ and rearranging, we end up with

$$\frac{t}{1-t}=\frac{\left(1-\beta_2\right)\left(1-\alpha_2^h\right)\left(1-\alpha\right)}{\overline{\varepsilon}_2}+\left(\frac{\mu^{h}\left(1+r\right)\left(1-\beta_2\right)}{\lambda\left(1-p\right)\overline{\varepsilon}_2}\right)\left[W\left(s^d-s^a\right)\frac{\partial L_1}{\partial B}-1\right], \tag{A.27}$$

where α_2^h, β_2 and $\overline{\varepsilon}_2$ are defined in eq. (3.10) in the main text.

Next we derive the optimal effective marginal tax rate on first-period labor income (t_1). Multiplying (A.16) by $w\ell_1$, we obtain

$$\begin{aligned}
\alpha w\ell_1\left(U'_{1\ell}-\lambda+\lambda t_1 w\frac{\partial \ell_1}{\partial G}\right)&=-\left(1-\alpha\right)w\ell_1\left(U'_{1h}-\lambda+\lambda t_1 W\frac{\partial L_1}{\partial G}\right) \\
&+w\ell_1\left(s^d-s^a\right)\left(\mu^{\ell}w\frac{\partial \ell_1}{\partial G}+\mu^{h}W\frac{\partial L_1}{\partial G}\right),
\end{aligned} \tag{A.28}$$

while (A.24) implies

$$\alpha w\ell_1\left(U'_{1\ell}-\lambda+\lambda t_1 w\frac{\partial \ell_1}{\partial G}\right)=\lambda t_1\left[\alpha w\frac{\partial \ell_1^c}{\partial t}+\left(1-\alpha\right)W\frac{\partial L_1^c}{\partial t}\right]$$

$$
-\left(1-\alpha\right)WL_1\left(U'_{1h}-\lambda+\lambda t_1 W\frac{\partial L_1}{\partial G}\right)
$$

$$
+\mu^{\ell}w\left(s^d-s^a\right)\left(w\ell_1\frac{\partial \ell_1}{\partial G}-\frac{\partial \ell_1^c}{\partial t}\right)+\mu^h W\left(s^d-s^a\right)\left(WL_1\frac{\partial L_1}{\partial G}-\frac{\partial L_1^c}{\partial t}\right). \qquad \text{(A.29)}
$$

Equating the right-hand sides of (A.28) and (A.29), using the facts (from the definition of t_1) that

$$
\frac{\partial \ell_1^c}{\partial t}=\frac{\partial \ell_1^c}{\partial t_1}=-w\frac{\partial \ell_1^c}{\partial w\left(1-t_1\right)}, \qquad \frac{\partial L_1^c}{\partial t}=\frac{\partial L_1^c}{\partial t_1}=-W\frac{\partial L_1^c}{\partial W\left(1-t_1\right)}, \qquad \text{(A.30)}
$$

and dividing by WL_1, we get

$$
\left(1-\beta_1\right)\left(1-\alpha\right)\left(U'_{1h}-\lambda+\lambda t_1 W\frac{\partial L_1}{\partial G}\right)=\left(1-\beta_1\right)\mu^h W\left(s^d-s^a\right)\frac{\partial L_1}{\partial G}
$$

$$
+\left(\frac{s^d-s^a}{1-t_1}\right)\left(\mu^{\ell}\beta_1\varepsilon_{1\ell}^c+\mu^h\varepsilon_{1h}^c\right)-\lambda\left(\frac{t_1}{1-t_1}\right)\left[\alpha\beta_1\varepsilon_{1\ell}^c+\left(1-\alpha\right)\varepsilon_{1h}^c\right], \qquad \text{(A.31)}
$$

$$
\varepsilon_{1\ell}^c\equiv\frac{\partial \ell_1^c}{\partial w\left(1-t_1\right)}\frac{w\left(1-t_1\right)}{\ell_1}, \qquad \varepsilon_{1h}^c\equiv\frac{\partial L_1^c}{\partial W\left(1-t_1\right)}\frac{W\left(1-t_1\right)}{L_1}.
$$

Dividing through by $\lambda\left[\alpha\beta_1\varepsilon_{1\ell}^c+\left(1-\alpha\right)\varepsilon_{1h}^c\right]$ in (A.31) and rearranging, we find

$$
\frac{t_1}{1-t_1}=\frac{\left(1-\beta_1\right)\left(1-\alpha_1^h\right)\left(1-\alpha\right)}{\overline{\varepsilon}_1^c}
$$

$$
+\left(\frac{s^d-s^a}{\lambda\overline{\varepsilon}_1^c\left(1-t_1\right)}\right)\left(\mu^{\ell}\beta_1\varepsilon_{1\ell}^c+\mu^h\varepsilon_{1h}^c\right)+\frac{\mu^h W\left(1-\beta_1\right)\left(s^d-s^a\right)}{\lambda\overline{\varepsilon}_1^c}\frac{\partial L_1}{\partial G}, \qquad \text{(A.32)}
$$

where α_1^h and $\overline{\varepsilon}_1^c$ are defined in (3.11) in the main text. When none of the two non-mimicking constraints are binding (that is, when it is optimal to offer less than full insurance to both skill groups), we have $\mu^{\ell}=\mu^h=0$, and (A.27) and (A.32) then reduce to eqs. (3.10) and (3.11) in the text, respectively.

A.4. The optimal level of social insurance

Finally, we derive the expressions for the optimal level of social insurance reported in sub-section 3.4. To investigate the conditions under which the optimal insurance level is less than perfect, we set $\mu^{\ell}=\mu^h=0$. Dividing (A.20) by p and (A.21) by $1-p$, and subtracting the latter equation from the former, we obtain

$$
\delta\alpha w\ell_1\left(U'_{d\ell}-U'_{al}\right)+\delta\left(1-\alpha\right)WL_1\left(U'_{dh}-U'_{ah}\right)
$$

$$
+\lambda\alpha w t_1 w\ell_1\left[\frac{1}{p}\frac{\partial \ell_1}{\partial b}-\left(\frac{1}{1-p}\right)\frac{\partial \ell_1}{\partial B}\right]+\lambda\left(1-\alpha\right)W t_1 WL_1\left[\frac{1}{p}\frac{\partial L_1}{\partial b}-\left(\frac{1}{1-p}\right)\frac{\partial L_1}{\partial B}\right]
$$

$$+\frac{\lambda\alpha t_1 w}{1+r}\left[\frac{1-\widehat{p}^{\ell}}{1-p}-\frac{\widehat{p}^{\ell}}{p}\right]\frac{\partial \ell_1^c}{\partial t}+\frac{\lambda(1-\alpha)t_1 W}{1+r}\left[\frac{1-\widehat{p}^{h}}{1-p}-\frac{\widehat{p}^{h}}{p}\right]\frac{\partial L_1^c}{\partial t}=0. \quad (A.33)$$

From (2.11) and (2.15), we have

$$\frac{\widehat{p}^{i}}{p}=\frac{\delta(1+r)U'_{di}}{U'_{1i}}, \qquad \frac{1-\widehat{p}^{i}}{1-p}=\frac{\delta(1+r)U'_{ai}}{U'_{1i}}, \qquad i=\ell,h. \quad (A.34)$$

Inserting (A.34) into (A.33), dividing through by $\delta W L_1$, and using (A.30), we may write (A.33) as

$$\begin{aligned}&\alpha(U'_{d\ell}-U'_{al})+(1-\alpha)(U'_{dh}-U'_{ah})-(1-\beta_1)\alpha(U'_{d\ell}-U'_{al})\\&+\lambda\left(\frac{t_1}{1-t_1}\right)\left[\alpha\beta_1\varepsilon_{1\ell}^c\left(\frac{U'_{d\ell}-U'_{al}}{U'_{1\ell}}\right)+(1-\alpha)\varepsilon_{1h}^c\left(\frac{U'_{dh}-U'_{ah}}{U'_{1h}}\right)\right]\\=&\frac{\lambda t_1}{\delta}\left\{\alpha w\beta_1\left[\left(\frac{1}{1-p}\right)\frac{\partial \ell_1}{\partial B}-\frac{1}{p}\frac{\partial \ell_1}{\partial b}\right]+(1-\alpha)W\left[\left(\frac{1}{1-p}\right)\frac{\partial L_1}{\partial B}-\frac{1}{p}\frac{\partial L_1}{\partial b}\right]\right\}. \quad (A.35)\end{aligned}$$

Dividing (A.17) by δp and (A.18) by $\delta(1-p)$ (recalling that $\mu^{\ell}=\mu^{h}=0$) and subtracting the latter equation from the former, we obtain

$$\begin{aligned}&\alpha(U'_{d\ell}-U'_{al})+(1-\alpha)(U'_{dh}-U'_{ah})\\=&\frac{\lambda t_1}{\delta}\left\{\alpha w\left[\left(\frac{1}{1-p}\right)\frac{\partial \ell_1}{\partial B}-\frac{1}{p}\frac{\partial \ell_1}{\partial b}\right]+(1-\alpha)W\left[\left(\frac{1}{1-p}\right)\frac{\partial L_1}{\partial B}-\frac{1}{p}\frac{\partial L_1}{\partial b}\right]\right\}. \quad (A.36)\end{aligned}$$

Dividing through by $p(1-p)$ in (A.15), we find that

$$\left(\frac{1}{1-p}\right)\frac{\partial \ell_1}{\partial B}-\frac{1}{p}\frac{\partial \ell_1}{\partial b}=\delta\left(s^d-s^a\right)\Omega^{\ell}, \quad (A.37)$$

$$\Omega^{\ell}\equiv\frac{wp(1-p)(1+r)}{\Delta^{\ell}}\left\{\frac{a_{1G}^{\ell}U'_{a\ell}U'_{d\ell}X^{\ell}}{U'_{1\ell}}+(1+r)\left[U''_{a\ell}U''_{d\ell}-(U''_{da\ell})^2\right]\right\}>0,$$

and similarly we have

$$\left(\frac{1}{1-p}\right)\frac{\partial L_1}{\partial B}-\frac{1}{p}\frac{\partial L_1}{\partial b}=\delta\left(s^d-s^a\right)\Omega^{h}, \qquad \Omega^{h}>0, \quad (A.38)$$

where Ω^h is defined analogously to Ω^{ℓ}. Using (A.37) and (A.38), we can write (A.35) as

$$\begin{aligned}&\alpha\beta_1(U'_{d\ell}-U'_{al})\left[1+\frac{\lambda\varepsilon_{1\ell}^c}{U'_{1\ell}}\left(\frac{t_1}{1-t_1}\right)\right]+(1-\alpha)(U'_{dh}-U'_{ah})\left[1+\frac{\lambda\varepsilon_{1h}^c}{U'_{1h}}\left(\frac{t_1}{1-t_1}\right)\right]\\&=\lambda t_1\left(s^d-s^a\right)\left[\alpha w\beta_1\Omega^{\ell}+(1-\alpha)W\Omega^{h}\right], \quad (A.39)\end{aligned}$$

Using (A.37) and (A.38), we can write (A.36) as

$$\alpha(U'_{d\ell}-U'_{al})+(1-\alpha)(U'_{dh}-U'_{ah})=\lambda t_1\left(s^d-s^a\right)\left[\alpha w\Omega^{\ell}+(1-\alpha)W\Omega^{h}\right]. \quad (A.40)$$

Using (A.40) to eliminate $(1-\alpha)(U'_{dh} - U'_{ah})$ from (A.39), and solving for $U'_{d\ell} - U'_{dh}$, we arrive at eq. (3.12) in the main text. Alternatively, using (A.40) to eliminate $\alpha(U'_{d\ell} - U'_{al})$ from (A.39) and solving for $U'_{dh} - U'_{ah}$, we end up with eq. (3.13).

References

Bovenberg, A.L. and P.B. Sørensen (2004). Improving the Equity-Efficiency Trade-Off: Mandatory Savings Accounts for Social Insurance. *International Tax and Public Finance* 11, 507-529.

Diamond, P.A. (2003). *Taxation, Incomplete Markets, and Social Security.* The 2000 Munich Lectures in Economics, MIT Press.

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

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

Epstein and Zin (1989). Substitution, Risk Aversion and the Temporal Behavior of Asset Returns. *Journal of Political Economy* 99, 263-286.

Feldstein, M. and D. Altman (1998). Unemployment Insurance Savings Accounts. *NBER Working Paper* 6860.

Fölster, S. (1997). Social Insurance Based on Personal Savings Accounts: A Possible Reform for Overburdened Welfare States? *European Economy* no. 4, 1997, 81-100.

Fölster, S. (1999). Social Insurance Based on Personal Savings. *Economic Record* 75, 5-18.

Fölster, S., R. Gidehag, M. Orszag and D. Snower (2002). Assessing Welfare Accounts. *CEPR Discussion Paper* no. 3479.

Greenwald, B. and J.E. Stiglitz. (1986). Externalities in Economies with Imperfect Information and Incomplete Markets, *Quarterly Journal of Economics* 101, 229-264.

Orszag, J. and D. Snower (1997). Expanding the Welfare System: A Proposal for Reform. *European Economy* no. 4, 1997, 101-118.

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.

Stiglitz, J. and J. Yun (2002). Integration of Unemployment Insurance with Retirement Insurance. *NBER Working Paper* no. 9199.

Sørensen, P.B. (2003). Social Insurance Based on Individual Savings Accounts. Chapter 10 in S. Cnossen and H.-W. Sinn (eds.), *Public Finances and Public Policy in the New Century*, MIT Press.

Vickrey, W. (1939). Averaging of Income for Income-Tax Purposes. *Journal of Political Economy* 47, 379-397.

Vickrey, W. (1947). *Agenda for Progressive Taxation.* New York: The Ronald Press Company.