

Human Capital and Optimal Positive Taxation of Capital Income*

Bas Jacobs^{†◊} and A. Lans Bovenberg[‡]

December 10, 2008

Abstract

This paper analyzes optimal linear and non-linear taxes on capital and labor incomes in a life-cycle model of human capital investment, financial savings, and labor supply with heterogeneous individuals. A dual income tax with a positive marginal tax rate on not only labor income but also capital income is optimal. The positive tax on capital income serves to alleviate the distortions of the labor tax on human capital accumulation. The optimal marginal tax rate on capital income is lower than that on labor income if savings are elastic compared to investment in human capital, substitution between verifiable and non-verifiable inputs in human capital formation is difficult, and most investments in human capital are verifiable so that education subsidies can directly reduce the tax wedge on learning. Numerical calculations suggest that the optimal marginal tax rate on capital income is substantial

Keywords: human capital, labor income taxation, capital income taxation, life-cycle, education subsidies.

JEL codes: H2, H5, I2, J2.

*The authors thank Richard Blundell, Martin Gervais, Søren Bo Nielsen, Rick van der Ploeg, and seminar participants at the University of Amsterdam, the London School of Economics, the Institute for Fiscal Studies, and members of the CESifo area conference on Public Economics, May 7–9, 2004, München for useful comments and suggestions. Bas Jacobs gratefully acknowledges financial support from the NWO Priority Program ‘Scholar’ and for financial support from the Dutch Organization for Sciences (NWO Vidi Grant No. 452-07-013, ‘Skill Formation in Distorted Labor Markets’). This paper previously circulated under the title “Optimal Capital Income Taxation with Endogenous Human Capital Formation”.

[†]Erasmus University Rotterdam, Tinbergen Institute, Netspar, and CESifo. Email: bjacobs@few.eur.nl
[◊]Corresponding author: Erasmus School of Economics, Erasmus University Rotterdam, PO box 1738, 3000 DR Rotterdam, The Netherlands. Phone: +31 – 10 – 408 1481. Fax: +31 – 10 – 408 9166. [‡]Tilburg University, CentER, Netspar, and CEPR. E-mail: a.l.bovenberg@uvt.nl.

“... A tax under which consumption and saving are placed on an equal footing is superior to one, as with a general income tax, under which an extra impost is laid on savings. [...] It would be possible, if it were so desired, to make a rough adjustment for the fact that some expenditures upon consumption are, in effect, investments yielding income in the future, and, therefore, liable under an ordinary income tax to a double impost. Thus charges incurred for [...] education might be exempted *eo nomine*.” Pigou (1928, p.122/123)

1 Introduction

Should capital income be taxed? This has always been an important question in public finance. Since Pigou (1928), many papers have shown that under certain conditions capital income should not be taxed. However, the public finance literature has traditionally focussed either on optimal taxation of savings in dynamic models with representative agents or on optimal labor income taxation in static models with endogenous labor supply and heterogeneous agents. Reminiscent of Pigou (1928), models with infinitely lived individuals without endogenous human capital formation typically find that a zero tax on capital income is optimal in the long run (Chamley, 1986; Judd 1985). This result also holds true in models with heterogeneous or finitely lived agents as long as preferences are weakly separable between consumption and leisure (Atkinson and Stiglitz, 1976; Ordoover and Phelps, 1979).¹ Some do also analyze human capital formation in infinitely lived representative agent models, and the typical result is that optimal taxes on human and financial capital should be zero in finite time (Jones et al. 1993, 1997; Judd 1999).²

This paper explores how redistributive governments should simultaneously optimize taxes on financial savings, labor incomes and human capital investments in the presence of heterogeneous finitely lived households. As the quote by Pigou suggests, the desirability of zero capital taxes critically relies on the possibility to expense all costs of human capital investments at the rate of the labor tax. This paper demonstrates that the government should optimally employ positive capital income taxes to alleviate the distortionary impact of labor income taxes on human capital formation if the costs of learning are not all verifiable to the government and thus cannot all be expensed. Direct educational expenditures on books, computers and traveling are important examples of non-verifiable investments in human capital. Tuition costs cannot be deducted for income-tax purposes in many countries either, so that these costs are also effectively non-verifiable. Moreover, costs of effort while enrolled in education, such as studying hard, sacrificing leisure activities, and preparing exams, are important immaterial costs that the government cannot verify easily.³ Simple simulations suggest that the optimal tax rate on capital income is substantial – even if a relatively large part of human capital investments are verifiable and thus can be subsidized. Indeed, the indirect costs of education, foregone labor earnings while enrolled in education, are in effect deductible from labor taxation as lower labor earnings reduce the labor tax bill.

¹The optimal zero capital tax result breaks down in models with heterogeneous agents and finite lives when different labor types are imperfect substitutes in production (Pirttilä and Tuomala, 2001).

²See Boskin (1975) and Heckman (1976) for early treatments of the effects of capital income taxes on human capital formation. These treatments are not put in an optimal tax setting.

³Education may also generate non-pecuniary benefits, such as the fun of studying, nicer jobs, additional status, more freedom of occupational choice, et cetera. These non-pecuniary benefits, however, are typically much less important than immaterial costs in view of the observed high returns on (higher) education, which exceed returns on safe investments and approach those on equity. Whereas these high returns can be due to market failures, they also compensate investors for non-pecuniary costs that exceed non-pecuniary benefits (see also Judd, 2000; and, Palacios-Huerta, 2006).

To simultaneously analyze optimal redistributive taxes on financial saving and labor income, we formulate a two-period life-cycle model of human capital formation, financial savings, and labor supply with heterogenous agents who differ in their abilities to acquire human capital. Individuals invest in human capital in the first period of their lives. In the second period, they work, enjoy leisure and consume all their assets. We show that the optimal tax system features a dual-income tax with positive marginal tax rates on both capital and labor incomes. This result applies to both linear and non-linear tax systems.

In contrast to the earlier literature, positive capital taxes are optimal in order to mitigate the distortionary effects of the labor tax on investments in human capital. Whereas labor taxes encourage individuals to substitute human by financial assets, the capital tax offsets these distortions in the composition of saving. Since capital income taxes distort the overall level of saving, the optimal capital tax strikes a balance between distorting the composition and the level of saving. Indeed, the government faces a fundamental trade-off between efficiency in human capital formation and allocative efficiency in the intertemporal allocation of consumption. The optimal tax rate on capital income is relatively large compared to the tax on labor income if aggregate saving is inelastic compared to learning so that learning distortions dominate saving distortions.

Our case for substantially positive capital income taxes relies only on the presence of positive labor income taxes and does not directly depend on the redistributive preferences of the government. Optimal labor taxes increase with stronger redistributive desires and capital income taxes should increase accordingly – but only for efficiency reasons. In contrast to representative agent models (see, e.g., Atkinson and Sandmo, 1980; Nielsen and Sørensen, 1997; Jones et al. 1993, 1997; and Judd, 1999), we do not have to arbitrarily exclude lump-sum taxes as a policy instrument to prevent the optimal tax problem from becoming trivial.

Our finding that capital income needs to be taxed does not depend on non-separable utility, heterogenous preferences, or financial market failures.⁴ Also Erosa and Gervais (2002), Golosov et al. (2006), and Diamond (2006) analyze optimal capital income taxes in life-cycle models rather than models with infinitely lived individuals, but they allow for non-separable preferences in consumption and leisure. They demonstrate that optimal capital taxes are positive if leisure and consumption are more complementary later in life than they are earlier in life.⁵ Saez (2002) and Diamond (2006) incorporate heterogenous preferences and find that capital income should optimally be taxed for redistributive reasons if high-ability individuals feature a lower discount rate than low-ability individuals do.

Another strand of literature relies on incomplete financial markets to derive optimal capital income taxes. Aiyagari (1995) finds that positive capital income taxes are optimal because they redistribute resources from unconstrained towards liquidity-constrained phases in the life-cycle and from high-income states towards low-income states of nature. Hence, a positive tax on capital income helps to complete the missing capital and insurance markets. Also Golosov et al. (2003), Golosov et al. (2006) and Diamond (2006) study optimal taxation in the presence of non-insurable idiosyncratic skill shocks. They demonstrate that a positive tax wedge on saving may help to relax incentive constraints.⁶

⁴All the papers mentioned, including our own, assume that the government can commit to announced policies. A lack of commitment can also result in positive capital income taxes (see Kydland and Prescott, 1980; and, Fischer, 1980). We show that capital income taxes are optimally positive even if the government can credibly commit to tax policies.

⁵In life-cycle models, Ordober and Phelps (1979) and Atkinson and Sandmo (1980) showed earlier that the optimal capital tax is zero if leisure is weakly separable from consumption. See also Bernheim (2002) for a more elaborate discussion.

⁶In a similar vein, Grochulski and Piskorski (2005), da Costa and Maestri (2007) and Anderberg (2008) show that

Our paper is also related to the literature on the optimal taxation of capital incomes with infinitely lived households (Chamley, 1985; Judd, 1986, 1999; Jones et al. 1993, 1997). The results of these papers critically hinge on the availability of government debt to control the intertemporal allocation of resources (see also Atkinson and Sandmo, 1980). In particular, the government must tax income only initially, run large budget surpluses and build up a substantial stock of assets in initial periods so as to finance all future spending from the returns of these assets and distortionary taxes to zero in finite time. Indeed, if our model would be extended to an infinite horizon setting in which the government would have full access to public debt, we also would obtain zero optimal taxes on both labor and capital incomes in finite time. However, in more realistic and practical settings, taxation is a fact of life at any moment in time. As long as the government needs positive tax revenues, we show that positive capital income taxes are optimal if labor income is taxed at positive rates.

The rest of this paper is organized as follows. Section 2 describes individual behavior. Section 3 derives optimal linear tax policy if all educational efforts are non-verifiable. Subsequently, section 4 introduces verifiable educational efforts, which the government can subsidize. Section 5 derives optimal non-linear tax policies. Section 6 performs some numerical simulations. Section 7 concludes and discusses the policy implications of the analysis. Four appendices contain the technical details of our analysis.

2 The model

2.1 Preferences and technologies

We consider a partial equilibrium two-period life-cycle model without uncertainty. Before-tax wage rates and interest rates are exogenously fixed.⁷ The mass of agents is normalized to one. In the first period, agents supply unskilled labor and devote resources to learning. In the second period, agents supply skilled labor and spend time on leisure, which can be interpreted as early retirement. Perfect capital markets allow individuals to freely transfer resources across both periods.⁸

Individuals are heterogeneous in exogenous ability n . The cumulative distribution of ability is $F(n)$. $f(n)$ is the corresponding density function with support $[\underline{n}, \bar{n}]$, $\underline{n}, \bar{n} > 0$. The government knows the distribution of abilities, but does not observe individual ability. Accordingly, it cannot levy individual-specific lump-sum taxes to redistribute incomes, but must rely on distortionary taxes instead.

In the first period of their lives, individuals invest $p_e e_n$ in education, where p_e represents the unit costs of education e_n . Initially, we assume that the government cannot observe any of these educational investments so that it cannot subsidize them. Educational investment therefore consists only of direct expenditures and (monetized) effort costs. Section 4 shows that our main results continue to hold if verifiable costs of education are allowed for, such as tax-deductible foregone earnings, as long as some non-verifiable costs remain.

tax wedges on savings or wealth are optimal in risky environments with endogenous human capital formation.

⁷The model can thus be viewed as a model of a small open economy in which the international capital market fixes the real interest rate. Endogenous factor prices determined in general equilibrium in a closed economy do not affect the characterization of optimal tax rules derived below if labor types are perfect substitutes in production and production displays constant returns to scale (see Diamond and Mirrlees, 1971). If labor types are imperfect substitutes, however, Pirttilä and Tuomala (2001) show that the production efficiency theorem fails and capital income is optimally taxed or subsidized.

⁸Cameron and Heckman (2001) and Carneiro and Heckman (2002) argue that liquidity constraints are only of minor importance empirically.

Ability n can be viewed as the productivity of education, so that more able individuals produce more human capital with the same educational effort. The production function for human capital features a constant elasticity:

$$h_n = n\phi(e_n) \equiv ne_n^\beta, \quad (1)$$

where h_n denotes human capital of agent n . Human capital accumulation exhibits decreasing returns with respect to educational effort e_n (i.e., $\beta < 1$). Ability and educational investments are complementary inputs in producing human capital (i.e., $\frac{\partial^2 h_n}{\partial e_n \partial n} = \beta e_n^{\beta-1} \geq 0$).

In the second period, human capital is supplied to the labor market in the form of skilled labor. Gross labor income z_n is the product of the number of efficiency units of human capital, h_n , and hours worked l_n , i.e., $z_n \equiv h_n l_n = n\phi(e_n)l_n$. Furthermore, individuals consume their net labor earnings plus the net returns on their assets (which can be negative if individuals repay their debts).

Individuals feature a common, concave, and twice differentiable utility function defined over consumption in the first period c_n^1 , leisure $\ell_n \equiv 1 - l_n$, and consumption in the second period c_n^2 :

$$u(v(c_n^1, c_n^2), \ell_n), \quad (2)$$

where $u_{c^1} u_{c^2}, u_\ell > 0$, and $u_{c^1 c^1}, u_{c^2 c^2}, u_{\ell\ell} \leq 0$. The sub-utility function $v(c_n^1, c_n^2)$ is homothetic and weakly separable from leisure ℓ_n . With this particular utility function, the optimal capital income tax would be zero in the absence of human capital formation (see Bernheim, 2002). This specification thus most clearly shows how endogenous human capital formation affects the optimal capital income tax.

2.2 Budget constraints

The first-period budget constraint is given by

$$c_n^1 + p_e e_n = a_o - a_n, \quad (3)$$

where a_o is the common level of initial wealth.⁹

In the second period, individuals consume, work and consume their assets or repay their debts. Hence, with linear taxes, the second-period budget constraint amounts to¹⁰

$$c_n^2 = (1 - t)l_n n\phi(e_n) + Ra_n + g, \quad (4)$$

where $R \equiv 1 + (1 - \tau)r$ is the discount factor, r stands for the exogenous real interest rate, and τ denotes the tax rate on capital income. The lump-sum transfer g and the marginal tax rate t characterize the linear labor tax.

We arrive at the life-time budget constraint by substituting the first-period budget constraint (i.e., (3)) into the second-period budget constraint (4) to eliminate a_n :

$$Rc_n^1 + c_n^2 = (1 - t)l_n n\phi(e_n) - Rp_e e_n + Ra_o + g. \quad (5)$$

⁹We abstract from heterogeneity in the initial level of wealth a_o to eliminate distributional reasons for taxing saving so as to focus on the efficiency-enhancing properties of capital income taxation. High-ability agents feature the lowest life-time savings, because they borrow more than low-ability agents to finance their education in the first period. Jacobs and Bovenberg (2005) include a third period to allow for the possibility that high-ability agents save more than low-ability agents do. They show that the main results do not rely on any particular relationship of financial savings with ability.

¹⁰Section 5 models non-linear taxation and modifies the budget constraints accordingly.

2.3 Individual optimization

Individuals maximize their utilities by choosing c_n^1 , c_n^2 , l_n , and e_n , subject to their life-time budget constraints, and taking the policy instruments of the government as given. The resulting first-order condition for the optimal choice of education e_n amounts to

$$(1-t)l_n n \phi'(e_n) = p_e (1 + (1-\tau)r) = p_e R. \quad (6)$$

Marginal benefits of education (the left-hand side) should equal marginal costs (the right-hand side). The labor tax harms learning by depressing marginal benefits (since $\phi''(e_n) < 0$). The capital income tax, in contrast, boosts education because it raises the present value of investments in human capital by reducing the rate of return on alternative investments R . Indeed, capital income taxes induce individuals to substitute human capital for financial savings in their portfolio of human and financial assets.

The first-order condition for e_n (6) and the production function of human capital (1) imply that gross labor income z_n is proportional to e_n :

$$z_n = l_n n \phi(e_n) = \frac{p_e R}{(1-t)\beta} e_n, \quad (7)$$

The proportionality factor $p_e R / ((1-t)\beta)$ does not depend on ability n and is thus the same for all agents.

The first-order condition for labor supply amounts to

$$\frac{u_\ell}{u_v v_{c^2}} = w_n \equiv (1-t)n\phi(e_n), \quad (8)$$

while the Euler equation for savings is

$$\frac{v_{c^1}}{v_{c^2}} = 1 + (1-\tau)r = R. \quad (9)$$

The second-order condition for utility maximization implies (see Appendix A)

$$\mu_n \equiv 1 - \beta(1 + \epsilon_n) > 0, \quad \epsilon_n \equiv \frac{\partial l_n}{\partial w_n} \frac{w_n}{l_n}, \quad (10)$$

where ϵ_n denotes the compensated wage elasticity of labor supply with respect to the after-tax wage $w_n \equiv (1-t)n\phi(e_n)$. In view of a positive feedback between human capital and labor supply, decreasing returns in human capital accumulation ($\beta < 1$) are not sufficient for the second-order condition to be met. In particular, more learning raises the wage rate. The associated substitution effect boosts labor supply, which in turn makes learning more attractive. In order to prevent corner solutions, decreasing returns in the production of human capital must offset this positive feedback effect.

Substituting the link between e_n and z_n (7) into the life-time budget constraint (5), we find that the discounted value of life-time consumption $Rc_n^1 + c_n^2$ is linear in gross income (and in view of (7) also linear in learning):

$$Rc_n^1 + c_n^2 = (1-\beta)(1-t)z_n + Ra_o + g. \quad (11)$$

Weak separability between leisure and consumption in utility and homotheticity of $v(c_n^1, c_n^2)$ ensure that the shares of first- and second-period consumption in after-tax labor income do not depend on ability n , i.e.,

$$1 - \omega \equiv \frac{Rc_n^1}{(1-\beta)(1-t)l_n n \phi(e_n) + Ra_o + g}, \quad (12)$$

$$\omega \equiv \frac{c_n^2}{(1-\beta)(1-t)l_n n \phi(e_n) + Ra_o + g}. \quad (13)$$

For later reference, we define indirect utility $v(\cdot)$ as

$$v(g, t, R, n) \equiv u(v(\hat{c}_n^1, \hat{c}_n^2), \hat{\ell}_n), \quad (14)$$

where hats denote the optimized values for consumption in both periods and leisure.

2.4 Government

The government maximizes a social welfare function Γ defined over individuals' indirect utilities $v(g, t, R, n)$:

$$\Gamma \equiv \int_n^{\bar{n}} \Psi(v(g, t, R, n)) dF(n), \quad (15)$$

where $\Psi'(\cdot) > 0$, and $\Psi''(\cdot) \leq 0$. With $\Psi'(\cdot) = 1$, the social welfare function is utilitarian.

The government taxes labor at rate t and capital incomes at rate τ to finance exogenously given public spending Λ and the endogenous uniform lump-sum transfer g . The fundamental informational assumptions are that the government must be able to verify aggregate labor incomes $\int_n^{\bar{n}} l_n n \phi(e_n) dF(n)$ and aggregate capital incomes $\int_n^{\bar{n}} r a_n dF(n)$.

The government can freely borrow and lend at the capital market at rate r . By using the definition for $R \equiv 1 + r(1 - \tau)$, and the first-period budget household constraint, we can express the government budget constraint in terms of t and R as¹¹

$$\int_n^{\bar{n}} [t l_n n \phi(e_n) + (1 + r - R)(a_o - c_n^1 - p_e e_n) - g - \Lambda] dF(n) = 0. \quad (16)$$

3 Optimal linear taxation with non-verifiable learning

The Lagrangian \mathcal{L} for maximizing social welfare is given by

$$\begin{aligned} \max_{\{g, t, R\}} \mathcal{L} &= \int_n^{\bar{n}} \Psi(v(g, t, R, n)) dF(n) \\ &+ \eta \int_n^{\bar{n}} [t l_n n \phi(e_n) + (1 + r - R)(a_o - c_n^1 - p_e e_n) - g - \Lambda] dF(n), \end{aligned} \quad (17)$$

where η represents the Lagrange multiplier of the government budget constraint.

In order to characterize the first-order conditions, we denote the overall tax wedge on human capital investment by

$$\Delta \equiv \frac{t n l_n \phi'(e_n)}{p_e} - \tau r = \frac{t}{1-t} R - \tau r, \quad (18)$$

where the second equality is derived by using the first-order condition for learning (6). Whereas the labor tax implies an additional tax on learning, a positive capital tax gives rise to a subsidy on education by raising the present value of the marginal benefits of learning.

¹¹As in Nielsen and Sørensen (1997), one can interpret our model as the steady state of an overlapping-generations economy where optimal taxes are equivalent to the optimal taxes of a Pareto efficient tax reform where the government insulates the existing generations from the transition by employing public-debt policy.

The first-order conditions for maximizing social welfare are given by

$$\frac{\partial \mathcal{L}}{\partial g} = \int_{\underline{n}}^{\bar{n}} \left[(\Psi' \lambda_n - \eta) + \eta t n \phi(e_n) \frac{\partial l_n}{\partial g} + \eta \Delta p_e \frac{\partial e_n}{\partial g} - \eta \tau r \frac{\partial c_n^1}{\partial g} \right] dF(n) = 0, \quad (19)$$

$$\frac{\partial \mathcal{L}}{\partial t} = \int_{\underline{n}}^{\bar{n}} \left[-(\Psi' \lambda_n - \eta) z_n + \eta t n \phi(e_n) \frac{\partial l_n}{\partial t} + \eta \Delta p_e \frac{\partial e_n}{\partial t} - \eta \tau r \frac{\partial c_n^1}{\partial t} \right] dF(n) = 0, \quad (20)$$

$$\frac{\partial \mathcal{L}}{\partial R} = \int_{\underline{n}}^{\bar{n}} \left[(\Psi' \lambda_n - \eta) a_n + \eta t n \phi(e_n) \frac{\partial l_n}{\partial R} + \eta \Delta p_e \frac{\partial e_n}{\partial R} - \eta \tau r \frac{\partial c_n^1}{\partial R} \right] dF(n) = 0, \quad (21)$$

where we have employed Roy's lemma in all equations.

We follow Diamond (1975) and define the net social marginal value of income of an individual with ability n , including the income effects on the tax base, as

$$b_n \equiv \frac{\Psi' \lambda_n}{\eta} + \Delta p_e \frac{\partial e_n}{\partial g} + t n \phi(e_n) \frac{\partial l_n}{\partial g} - \tau r \frac{\partial c_n^1}{\partial g}. \quad (22)$$

Hence, the first-order condition (19) implies that the average social marginal benefit equals the unit marginal cost of a higher g :

$$\int_{\underline{n}}^{\bar{n}} b_n dF(n) = 1. \quad (23)$$

The first-order conditions yield the following relationship between the optimal distortions on human capital investment and saving (see Appendix C):

$$\Delta \beta = \tau r (1 - \omega) (1 - \beta + \bar{\gamma}) \sigma, \quad (24)$$

where $\sigma \equiv d \ln(c_n^2/c_n^1) / d \ln(u_{c^1}/u_{c^2})$ represents the intertemporal elasticity of substitution in consumption and $\gamma_n \equiv (R a_o + g) / (1 - t) z_n$. A bar denotes an income-weighted average of a skill-specific variable, e.g., $\bar{\gamma} \equiv \left[\int_{\underline{n}}^{\bar{n}} \gamma_n z_n dF(n) \right] \left[\int_{\underline{n}}^{\bar{n}} z_n dF(n) \right]^{-1}$.

This expression clearly shows the role of the capital income tax in reducing the tax distortions on learning caused by the labor income tax. The optimal capital income tax is zero if human capital formation is not distorted by the labor tax (i.e., $t = 0$ so that $\Delta = -\tau r$). The government optimally employs positive capital income taxes only if learning is endogenous ($\beta > 0$) and positive labor taxes distort learning (i.e., if $t > 0$ and thus $\Delta > 0$). Intuitively, by raising the net present value of investments in human capital, a positive capital income tax alleviates the tax distortions imposed by the labor income tax, as indicated by the left-hand side of (24). At the same time, a capital income tax distorts the intertemporal allocation of consumption, as indicated by the right-hand side of (24). At small capital income taxes, the welfare costs of distorted saving behavior are only second order, while the welfare benefits of lowering the tax distortions on learning are first order. Hence, the introduction of a small capital income tax enhances welfare. At the optimal capital tax, the marginal welfare benefits of alleviating the learning distortions balance the marginal welfare costs in terms of a distorted intertemporal allocation of consumption. The optimal capital tax thus trades off efficiency in the *composition* of saving (i.e., a level playing field between financial and human capital) and efficiency in the *level* of saving. Whereas the intertemporal substitution elasticity σ impacts the welfare losses of capital taxes on the level of saving, the learning elasticity β determines the welfare gains of capital taxes on the composition of saving.

Using $\Delta = tR/(1 - t) - \tau r$, we find the optimal dual-income tax structure of capital and labor taxes from (24):

$$\frac{\tau r}{1 + (1 - \tau)r} = (1 - \delta) \frac{t}{1 - t}, \quad \delta \equiv \frac{(1 - \beta + \bar{\gamma})(1 - \omega)\sigma}{\beta + (1 - \beta + \bar{\gamma})(1 - \omega)\sigma}, \quad (25)$$

where $0 < \delta < 1$. The capital tax is large compared to the labor tax if saving is inelastic (small σ) compared to learning (large β). In that case, labor taxes cause large distortions on human capital investment, while capital taxes do not distort saving much. Hence, capital taxes are an attractive instrument to fight labor-tax distortions on learning. δ measures the extent to which capital taxes leave the labor-tax distortion on human capital accumulation intact. If capital taxes are very distortionary (i.e., a large σ), capital taxes should optimally be lower. With infinite intertemporal substitution ($\sigma = \infty$), capital taxes are zero and the full learning distortion remains, i.e., $\delta = 1$. In contrast, if capital taxes are not distortionary at all ($\sigma = 0$), the capital tax optimally eliminates the entire labor-tax distortion on learning (i.e., $\Delta = 0$) so that $\delta = 0$.

In order to facilitate the discussion of the optimal tax schedules, we define the distributional characteristic ξ of labor income as

$$\xi \equiv \frac{\int_{\underline{n}}^{\bar{n}} (1 - b_n) z_n dF(n)}{\int_{\underline{n}}^{\bar{n}} z_n dF(n) \int_{\underline{n}}^{\bar{n}} b_n dF(n)} > 0. \quad (26)$$

ξ is the (negative) normalized covariance between the social value the government attaches to the income of a particular ability b_n and to gross second-period labor income z_n (see also Atkinson and Stiglitz, 1980). ξ represents the marginal distributional benefits, measured in monetary units, of redistributing a marginal euro through the income tax. With a positive distributional characteristic ξ , the base of the labor tax is larger for high-ability agents (who feature relatively low welfare weights) than for low-ability agents (who feature relatively high welfare weights), so that taxing labor income yields distributional benefits. The magnitude of the distributional characteristic depends not only on the correlation between ability and the tax base, but also on the correlation between ability and the welfare weights. Indeed, a zero distributional characteristic implies either that the government is not interested in redistribution (so that the welfare weight b_n is the same for all n) or that all ability types feature the same labor income (taxable income is the same for all n).

By using the distributional characteristic, we find the following expression for the optimal tax on labor income (see Appendix C):¹²

$$\frac{t}{1-t} = \frac{\xi}{(\omega + (1-\omega)\delta) [\overline{\epsilon/\mu}] + \delta\beta\overline{[(1+\epsilon)/\mu]}}, \quad (27)$$

This expression illustrates the fundamental trade-off between equity and efficiency. The numerator of (27) represents the distributional benefits associated with the labor tax. If redistributive concerns become more important (as indicated by a larger distributional characteristic ξ), the optimal marginal tax rate rises (*ceteris paribus* the income-weighted elasticities). The welfare losses of a higher labor-tax rate are captured by the denominator of (27), which measures the total elasticity of the labor-tax base. Welfare losses arise because labor taxes distort labor supply (i.e., the first term in the denominator, where $\overline{[\epsilon/\mu]}$ represents the compensated elasticity of labor supply with respect to the tax rate) and human capital investment (i.e. the second term in the denominator, where $\beta\overline{[(1+\epsilon)/\mu]}$ stands for the compensated elasticity of human capital accumulation with respect to the tax rate). If large compensated elasticities (in absolute value) indicate that redistributive taxes substantially distort labor supply and human capital accumulation, positive marginal taxes are costly and the optimal marginal labor tax is low (*ceteris paribus* the distributional characteristic ξ).

¹²These expressions are not closed-form solutions because the elasticities (ϵ_n and μ_n), the variables γ_n and the distributional characteristic ξ are endogenously determined.

By combining (25) and (27), we observe that the capital tax is directly related to the distributional characteristic of labor income rather than capital income. Indeed, the optimal capital tax is targeted only at alleviating learning distortions and is used neither for redistributive reasons nor to cut tax distortions on labor supply. With weakly separable and homothetic preferences and a constant elasticity in education, saving is proportional to labor income. Hence, a tax on saving in fact acts as a tax on additional labor earnings. A capital tax therefore gives rise to the same labor-supply distortions and the same distributional benefits as a labor-income tax, but introduces additional intertemporal distortions, which can be avoided by taxing labor rather than capital income.

The overall tax wedge on labor supply is $t/(1-t) - (1-\omega)\tau r/R$. The term $(1-\omega)\tau r/R$ features negatively in this tax wedge because the capital tax acts as an implicit subsidy on labor supply. The total tax wedge on labor supply can be written as (see Appendix C)

$$\frac{t}{1-t} - \frac{(1-\omega)\tau r}{1+(1-\tau)r} = \frac{\xi}{[\epsilon/\mu] + \frac{\delta}{\delta(1-\omega)+\omega}\beta[(1+\epsilon)/\mu]}. \quad (28)$$

The optimal overall tax wedge on labor supply (28) is lower if behavior becomes more sensitive to taxes, *ceteris paribus* the distributional characteristic. A higher wage elasticity of labor supply ϵ_n or human capital elasticity β reduces the optimal labor tax and therefore the need for capital taxes to mitigate the distortions of the labor tax on learning. By boosting the efficiency costs of the capital tax, a higher intertemporal substitution elasticity σ decreases the optimal labor-supply wedge, as the optimal capital tax can correct for only a small part of the labor-tax distortions on learning (so that δ and $0 \leq \delta/(\delta(1-\omega) + \omega) \leq 1$ are large).

The impact of the behavioral margins can be illustrated with some special cases. Exogenous learning ($\beta = 0$) implies a zero optimal capital income tax ($\tau = 0$). This is a familiar result from the standard model of optimal linear labor taxation with weakly separable utility (2), which is homothetic in consumption (see e.g., Bernheim, 2002). The optimal linear labor tax is then (see Atkinson and Stiglitz, 1980)

$$\frac{t}{1-t} = \frac{\xi}{\bar{\epsilon}}. \quad (29)$$

In the absence of intertemporal substitution in consumption (i.e., $\sigma = 0$), the capital income tax can costlessly accomplish efficient learning (i.e., $\delta = \Delta = 0$ and $t/(1-t) = \tau r/R$). In this case, the optimal overall tax wedge is given by

$$\frac{t}{1-t} - \frac{(1-\omega)\tau r}{1+(1-\tau)r} = \omega \frac{t}{1-t} = \frac{\xi}{[\epsilon/\mu]}. \quad (30)$$

Although investment in human capital and financial saving are not distorted, tax rates remain finite because the labor tax continues to distort labor supply. Endogenous learning raises the effective elasticity of the tax base only by increasing the absolute value of the effective elasticity of labor supply $[\epsilon/\mu]$, because the learning elasticity $\beta[1 + \epsilon/\mu]$ drops out of the denominator of (30) (compare (30) with (28)).

With infinite intertemporal substitution in consumption (i.e., $\sigma \rightarrow \infty$ so that $\delta = 1$), in contrast, the capital income tax is absent (i.e., $\tau = 0$) and the entire learning elasticity adds to elasticity of the labor-tax base so that the optimal tax wedge on labor supply amounts to (see also Bovenberg and Jacobs, 2005)

$$\frac{t}{1-t} = \frac{\xi}{[\epsilon/\mu] + \beta[(1+\epsilon)/\mu]}. \quad (31)$$

Exogenous labor supply ($\epsilon_n = 0$) does not directly affect the optimal tax structure (25) but only raises overall tax levels. The case for taxing capital income thus depends on endogenous learning ($\beta > 0$) rather than on endogenous labor supply. The overall tax wedge on labor supply,

$$\frac{t}{1-t} - \frac{(1-\omega)\tau r}{1+(1-\tau)r} = \frac{\xi(1-\beta)}{\frac{\delta\beta}{\delta(1-\omega)+\omega}}, \quad (32)$$

remains finite because the labor tax causes learning distortions that cannot be completely offset by the capital tax (if $\sigma > 0$ and thus $\delta > 0$).

4 Optimal linear taxation with partly verifiable learning

The previous section assumed that all investments in human capital were not verifiable. Hence, the government could not directly subsidize learning and the government had to rely on the capital income tax to alleviate the distortions of the labor-income tax on human capital formation. This section explores how verifiable educational investments, which can be subsidized by the government, affect our results. Time invested in education is arguably the most important verifiable investment in human capital, since foregone earnings are tax deductible and enrollment in (higher) education is widely subsidized across the Western world.

With verifiable learning the government has two instruments at its disposal to offset the labor-tax distortion on human capital accumulation: education subsidies and capital income taxes. This raises the question whether the government still wants to employ capital income taxes to alleviate the labor-tax distortions on human capital.

To answer this question, let educational efforts e_n consist of both a verifiable part x_n and a non-verifiable part y_n . x_n can be interpreted as the years spent in formal education and as foregone labor time invested in human capital. y_n can be viewed as direct costs and (monetized) effort costs. We assume that the verifiable input x_n is already tax-deductible, while after-tax expenditures are subsidized at rate s so that the marginal cost of investing x_n amounts to $(1-t)(1-s)p_x$.

x_n and y_n produce aggregate investment in human capital e_n through a constant-returns-to-scale sub-production function ψ :

$$e_n \equiv \psi(x_n, y_n), \quad (33)$$

where $\psi_x, \psi_y > 0$, $\psi_{xx}, \psi_{yy} \leq 0$ and $\psi_{xy} \geq 0$. The intertemporal budget constraint of households is now given by

$$Rc_n^1 + c_n^2 = (1-t)l_n n \phi(\psi(x_n, y_n)) - R((1-t)(1-s)p_x x_n + p_y y_n) + Ra_o + g, \quad (34)$$

where p_x and p_y denote the exogenous prices of x_n and y_n , respectively. p_x can be interpreted as foregone unskilled labor earnings when learning.

The first-order conditions for choice of the two educational inputs x_n and y_n amount to

$$l_n n \phi'(\cdot) \psi_x(x_n, y_n) = R(1-s)p_x, \quad (35)$$

$$(1-t)l_n n \phi'(\cdot) \psi_y(x_n, y_n) = R p_y. \quad (36)$$

The tax rate t does not enter (35) because tax deductibility of x_n implies that the tax rate reduces equally the marginal benefits and the marginal costs of verifiable learning x_n . In contrast, the tax rate leaves the costs of non-verifiable investments y_n unaffected and depresses only the benefits of these investments (see (36)). The investment subsidy s decreases the costs of verifiable investments

x_n . A capital income tax $\tau > 0$ reduces the costs of verifiable and non-verifiable learning alike by decreasing the opportunity return R .

When maximizing social welfare, the government faces the following budget constraint

$$\begin{aligned} & \int_{\underline{n}}^{\bar{n}} [t(nl_n\phi(e_n) - (1-s)p_x x_n(1+r)) + (1+r-R)((a_o - c_n^1 - (1-t)(1-s)p_x x_n - p_y y_n)] dF(n) \\ &= \int_{\underline{n}}^{\bar{n}} [(1+r)sp_x x_n + g + \Lambda] dF(n). \end{aligned} \quad (37)$$

Appendix C sets up the maximization program and finds that the optimal education subsidy satisfies

$$\frac{s + \frac{\tau r}{1+(1-\tau)r}}{(1-s)(1-t)} = \left(\frac{(1-\alpha)(1-\rho)}{1-(1-\alpha)(1-\rho)} \right) \frac{\Delta}{R}, \quad (38)$$

where $\rho \equiv d \ln(x_n/y_n) / d \ln(\psi_y/\psi_x)$ stands for the elasticity of substitution between x_n and y_n in the composite of aggregate investment in human capital e_n (33), and $\alpha \equiv x_n \psi_x / \psi$, is the constant share of verifiable inputs in the production of human capital (see Appendix B).

To interpret this expression, we note that a positive total subsidy on verifiable investments in human capital, $s + \tau r/R$, reduces the distortionary effect of the labor tax ($t > 0$) on aggregate investment in human capital. At the same time, however, it exacerbates the labor-tax distortions on the composition of learning. In particular, the labor-tax system boosts the demand for x_n at the expense of y_n because only x_n is tax deductible. An education subsidy results in even more substitution away from y_n to the tax deductible and subsidized inputs x_n . The sign of the optimal education subsidy thus depends on the relative impact of the subsidy on the *level* and *composition* of human capital investment. With a Leontief production function of human capital ($\rho = 0$), a positive subsidy on verifiable investments does not distort the composition of learning. Hence, the government can completely offset the tax distortions on non-verifiable learning y_n by subsidizing verifiable inputs x_n . If individuals cannot easily substitute y_n for x_n (i.e., $\rho < 1$), x_n is subsidized rather than taxed (i.e., $s + \tau r/R > 0$) in order to alleviate the distortionary effect of the labor tax on aggregate learning. A large substitution elasticity ($\rho > 1$), in contrast, implies that tax distortions on the composition of learning dominate tax distortions on the aggregate level of learning. Hence, tax-deductible investments should be taxed so as to combat substitution between the two educational inputs on account of the tax deductibility of verifiable inputs. In the extreme case of infinite substitution (i.e., $\rho \rightarrow \infty$), education subsidies do not reduce the learning distortion compared to the case in which all inputs are non verifiable. Indeed, the government then finds it optimal to tax x_n (i.e., $s + \tau r/R < 0$) so as make x_n effectively non-tax deductible. This ensures a level playing field with y_n .

With optimal education policies, Appendix C finds the following relationship between the optimal capital tax and the learning distortion Δ :

$$\Delta\varphi\beta = \tau r(1-\omega)(1-\beta+\bar{\gamma})\sigma, \quad \varphi \equiv \frac{\rho(1-\alpha)}{\alpha + \rho(1-\alpha)}, \quad (39)$$

where $0 \leq \varphi \leq 1$ measures the extent to which education subsidies leave the education distortion intact. Except for the presence of φ , this expression is identical to the corresponding expression (24) for the case without verifiable learning. Capital income taxes are smaller (larger) if education subsidies are more (less) powerful in alleviating learning distortions, i.e., if φ is small (large). Education subsidies largely eliminate the learning distortions if the share of non-verifiable learning

in aggregate learning $(1 - \alpha)$ and the substitution elasticity between verifiable and non-verifiable learning ρ are small.

We find the following for the optimal dual-income tax structure (see Appendix C):

$$\frac{\tau r}{1 + (1 - \tau)r} = (1 - \delta^*) \frac{t}{1 - t}, \quad \delta^* \equiv \frac{(1 - \beta + \bar{\gamma})(1 - \omega)\sigma}{\varphi\beta + (1 - \beta + \bar{\gamma})(1 - \omega)\sigma}, \quad (40)$$

$$\frac{t}{1 - t} = \frac{\xi}{(\delta^*(1 - \omega) + \omega) \overline{[\epsilon/\mu]} + \delta^*\varphi\beta \overline{[(1 + \epsilon)/\mu]}}, \quad (41)$$

$$\frac{t}{1 - t} - \frac{\tau r(1 - \omega)}{1 + (1 - \tau)r} = \frac{\xi}{\overline{[\epsilon/\mu]} + \frac{\delta^*}{(\delta^*(1 - \omega) + \omega)} \varphi\beta \overline{[(1 + \epsilon)/\mu]}}, \quad (42)$$

where $0 \leq \delta^* \leq 1$ is the slightly modified equivalent of δ . Compared to the case without verifiable learning, subsidized verifiable learning in effect reduces the elasticity of human capital investment with respect to the tax rate from $\delta\beta$ to $\delta^*\varphi\beta$ (compare (40) with (25), (41) with (27), and (42) with (28)). Optimal capital taxes remain positive ($\tau > 0$). The only two exceptions are the cases in which either all learning is verifiable ($\alpha = 1$) or education subsidies do not distort the composition of learning ($\rho = 0$). Education subsidies then costlessly eliminate the entire labor-tax distortion on human capital accumulation ($\varphi = 0$, and $\delta^* = 1$) so that capital income taxes are not needed anymore.

The insights from section 3 remain valid, except that verifiable learning introduces an additional behavioral margin: substitution between verifiable and non-verifiable learning. Both capital taxes and education subsidies are imperfect instruments for alleviating the learning distortions imposed by labor taxes. Whereas education subsidies distort the composition of learning, the capital income tax distorts aggregate saving. Just as the other behavioral margins, more elastic behavior on account of a higher substitution elasticity ρ reduces the optimal overall tax wedge (42). Moreover, just as the aggregate learning and saving margin, the learning composition margin affects the composition of the tax burden (40). At the optimum, the government balances distortions on aggregate learning with those on the composition of learning and the intertemporal allocation of consumption. The capital tax becomes a more important instrument for fighting the labor-tax distortions on learning if a high elasticity ρ makes the educational subsidy a relatively weak instrument for this purpose.

5 Optimal non-linear taxation with non-verifiable learning

This section introduces non-linear policy instruments in the model of section 2 with non-verifiable learning. The crucial informational requirements for non-linear instruments are that the government can verify labor earnings z_n and savings a_n at the individual level. In particular, the government levies a non-linear labor-income tax $T_z(z_n)$ and a non-linear capital-income tax $T_a(ra_n)$, where marginal tax rates are denoted by $T'_z(z_n) \equiv dT(z_n)/dz_n$ and $T'_a(ra_n) \equiv dT_a(ra_n)/d(ra_n)$. The first-period budget constraint (3) is unaffected, but the second-period household budget constraint (4) now becomes

$$c_n^2 = z_n - T_z(z_n) + (1 + r)a_n - T_a(ra_n). \quad (43)$$

Furthermore, in the presence of non-linear tax instruments, we no longer have to assume that the production function for human capital $h_n = n\phi(e_n)$ exhibits a constant elasticity, and we impose only $\phi' > 0$, and $\phi'' < 0$. Hence, the learning elasticity $\beta_n \equiv \phi'(e_n)e_n/\phi(e_n)$ may depend on skill n .

The first-order conditions for individual optimization are given by

$$\frac{u_\ell}{u_v v_{c^2}} = w_n \equiv (1 - T'_z) n \phi(e_n), \quad (44)$$

$$\frac{v_{c^1}}{v_{c^2}} = R_n, \quad (45)$$

$$(1 - T'_z) l_n n \phi'(e_n) = p_e R_n, \quad (46)$$

where the interest factor $R_n \equiv 1 + (1 - T'_a)r$ is no longer necessarily uniform across households.

We can write the incentive compatibility constraints for this non-linear tax problem as a differential equation on utility (see also Mirrlees, 1971):¹³

$$\frac{du_n}{dn} = u_\ell (v(c_n^1, c_n^2), 1 - l_n) \frac{l_n}{n}. \quad (47)$$

The economy's resource constraint amounts to

$$\int_{\underline{n}}^{\bar{n}} (l_n n \phi(e_n) - (1 + r)p_e e_n - (1 + r)c_n^1 - c_n^2 + (1 + r)a_o) dF(n) = \Lambda. \quad (48)$$

We solve the optimal allocation by applying the maximum principle and setting up a Hamiltonian \mathcal{H} , with labor supply l_n and savings a_n as control variables, utility u_n as state variable, and θ_n as the co-state variable for the incentive compatibility constraints (47):

$$\begin{aligned} \max_{\{l_n, a_n, u_n\}} \mathcal{H} &\equiv \Psi(u_n) f(n) + \eta (l_n n \phi(e_n) + (1 + r)a_n - c_n^2) f(n) \\ &\quad - \frac{\theta_n}{n} u_\ell (v(a_o - a_n - p_e e_n, c_n^2), 1 - l_n) l_n, \end{aligned} \quad (49)$$

where we substituted the first-period household-budget constraint (3) to eliminate c_n^1 . η is the shadow value of the resource constraint. Note that θ_n is defined negatively. The transversality conditions are given by

$$\lim_{n \rightarrow \underline{n}} \theta_n = 0, \quad \lim_{n \rightarrow \bar{n}} \theta_n = 0. \quad (50)$$

Having determined the second-best allocation, we implement it as a decentralized market outcome by employing non-linear taxes on labor income $T_z(\cdot)$ and capital income $T_a(\cdot)$. The relationship between optimal marginal capital and labor taxes is identical to the corresponding expression (25) derived under linear policies, except that non-linear tax rates replace the linear ones, and individual elasticities replace aggregate elasticities (see Appendix D):

$$\frac{rT'_a(ra_n)}{1 + (1 - T'_a(ra_n))r} = \frac{T'_z(z_n)}{1 - T'_z(z_n)} (1 - \delta_n), \quad \delta_n \equiv \frac{(1 - \omega)(1 - \beta_n + \gamma_n)\sigma}{\beta_n + (1 - \omega)(1 - \beta_n + \gamma_n)\sigma}, \quad (51)$$

The intuition for the expression is the same as for the case with linear tax instruments. A positive capital tax relaxes incentive compatibility constraints in the presence of a positive marginal labor tax. High-ability types can mimic low-ability types by investing less in human capital. Consequently, mimickers save more in the first period of their life-cycle than non-mimickers do. By taxing saving at the margin, the government reduces the attractiveness for high-ability individuals to mimic

¹³In adopting this approach, we assume that the second-order conditions for the optimal policy problem are met. This requires single crossing of the utility function.

low-ability individuals. Hence, this policy relaxes the incentive compatibility constraints that are associated with the redistribution of income.

As in the case with linear taxes, distributional concerns do not directly affect the optimal capital tax. Indeed, the distributional term θ_n is absent in (51). The capital tax is aimed exclusively at alleviating the learning distortions imposed by the labor tax and is thus directly related to the presence of the labor tax. Accordingly, the shape of the non-linear capital tax closely follows the pattern of marginal tax rates on labor incomes. In particular, marginal capital taxes should be relatively large for those individuals who face relatively large marginal labor taxes, and vice versa.

The optimal non-linear income tax is given by (see Appendix D)

$$\frac{T'_z(z_n)}{1 - T'_z(z_n)} = \frac{\left(1 + \frac{1}{\epsilon_n^*}\right) \frac{\epsilon_n}{\mu_n}}{nf(n)} \frac{u_{c^2} \theta_n / \eta}{[(\omega + (1 - \omega) \delta_n) \epsilon_n / \mu_n + \delta_n \beta_n (1 + \epsilon_n) / \mu_n]}, \quad (52)$$

where $\epsilon_n^* \equiv -(l_n u_{\ell\ell} / u_\ell - l_n u_{\ell v} / u_v)^{-1} > 0$ and $u_{c^2} \equiv u_v v_{c^2}$. The formula is similar to Mirrlees (1971) and consists of four elements. The first is the term in square brackets. This is the elasticity of the non-linear income tax base with respect to the labor tax and is the same as the elasticity of the tax base of the linear income tax (except that individual elasticities replace aggregate elasticities), see equation (27) and the discussion there. In particular, a large optimal non-linear capital tax mitigates the learning distortion and thus reduces δ_n below unity, thereby allowing for larger marginal labor taxes (*ceteris paribus*). The second element is the elasticity $(1 + \epsilon_n^{*-1}) \epsilon_n / \mu_n$. This elasticity transforms the base for learning ability $nf(n)$ into the labor-tax base $z_n q(z_n)$, where $q(z_n)$ is the density of earnings at z_n (see also Saez, 2001; Bovenberg and Jacobs, 2005). A larger elasticity $(1 + \epsilon_n^{*-1}) \epsilon_n / \mu_n$ implies that, for a given ability distribution $f(n)$, the distribution of gross earnings $q(z_n)$ is more unequal. Consequently, marginal tax rates should be set at higher levels. The third element is $nf(n)$. A large number of individuals $f(n)$ with ability n yield relatively large efficiency costs of marginal tax rates $T'_z(z_n)$ and thus reduce the optimal marginal tax rates at ability level n .

The final element in (52) is θ_n / η . This term is the non-linear counterpart of ξ and stands for the benefits of redistribution at skill n . Using the first-order condition for u_n , θ_n / η can be written as (see Appendix D)

$$\frac{\theta_n}{\eta} = \int_n^{\bar{n}} \left(\frac{1}{u_{c^2}} - \frac{\Psi'(\cdot)}{\eta} \right) \exp \left[\int_n^m \frac{l_s u_{\ell v}}{s u_v} ds \right] f(m) dm. \quad (53)$$

This term is well known from the optimal tax literature. It measures the welfare gain from redistributing income from individuals with an ability level higher than n to individuals with a lower ability than n . The transversality conditions imply that marginal benefits of redistribution are zero at the top and the bottom if bunching is absent at the bottom. We refer the reader to Mirrlees (1971), Seade (1977), Atkinson and Stiglitz (1980), Diamond (1998) and Saez (2001) for a more elaborate discussion of this term.

6 Simulations

This section conducts a back-of-the-envelope calculation of optimal capital income taxes by using expression (40) for the optimal dual-income tax structure. We assume that governments optimized the labor tax (t) according to equation (41). We present two simulations. One is based on normal assumptions regarding the parameters ('baseline'). The other simulation is based on conservative parameters that would undermine our case for positive capital taxes ('conservative').

Expression (40) contains three share parameters: α , and ω , and $\bar{\gamma}$. In the baseline, the share of observable costs in total educational expenditure is set at $\alpha = 0.5$. Becker (1964), Boskin (1975) and Trostel (1993) find that the share of goods invested in education is about one-quarter and that the share of (tax deductible) foregone earnings amounts to three-quarters. However, this ignores the effort cost of education (i.e., attending college, studying, etc) and non-verifiable investment in on-the-job training. The share of second-period consumption in total consumption, ω , is set at 0.5 and the average ratio of non-labor income (net transfers and returns from initial wealth) to net labor incomes, $\bar{\gamma}$, is 0.25. The exogenously given revenue requirement Λ balances the budget.

As regards the three relevant behavioral elasticities in (40) (i.e., σ , β , and ρ), the largest empirical literature exists on the intertemporal elasticity of substitution in consumption σ . Whereas older papers found very small elasticities, more recent work (e.g., Hall, 1988; Attanasio and Weber, 1995) suggests that the intertemporal substitution elasticity is around $\sigma = 0.5$. Trostel (1993) contains an extensive discussion on plausible parameter values for the returns to inputs invested in human capital β . Trostel considers $\beta = 0.6$ the most reasonable value and the baseline employs this number. Concerning the elasticity of substitution between inputs in the human capital formation ρ , we follow Trostel (1993) by using $\rho = 1$ as the benchmark value. Hence, the production function of human capital is Cobb Douglas.

In order to compute the optimal capital tax, we adopt a rate of return r on financial investments featuring similar (risk) characteristics as investments in human capital. Since human capital is riskier than governments bonds, a real rate of return of 6% per annum is assumed.¹⁴ In our two-period life-cycle model, each period captures one half of the average overall life span. Since we abstract from a retirement phase, we adopt a cohort length of 25 years.¹⁵

Figure 1 shows the optimal capital income taxes at given labor income taxes for the baseline values of the parameters. Quite surprisingly, a synthetic income tax, which taxes capital and labor incomes at the same marginal rates, appears to be roughly optimal. The case for substantially positive capital taxes also survives in our ‘conservative’ simulation, where we assume that the human capital elasticity β is cut in half to 0.3, the share of verifiable costs α increases to three-quarters, and the intertemporal elasticity σ doubles in value to unity. At a marginal labor tax of 60%, the optimal capital tax would still be around 20%.¹⁶

7 Conclusions

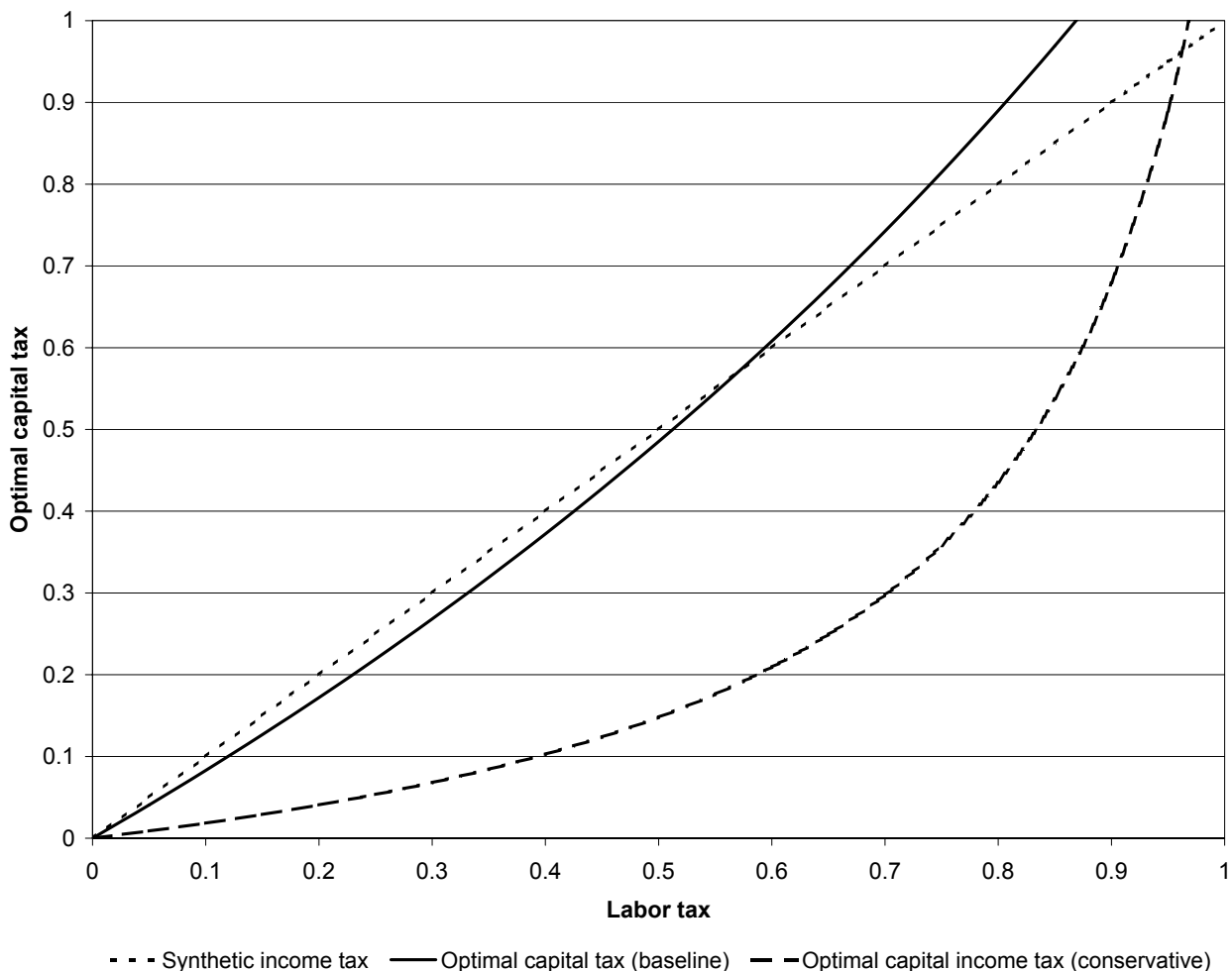
This paper investigated the interactions between labor markets, capital markets and human capital investments in a second-best world in which the government engages in redistribution without being able to verify work and learning efforts. To that end, we developed a two-period life-cycle model of human capital investment, financial saving and labor supply in which individuals feature different abilities to learn and income inequality emerges endogenously. The reason why positive capital-income taxes are optimal is that these taxes are a second-best instrument to alleviate the labor-tax distortions on human capital accumulation. Accordingly, if a government sets large marginal labor

¹⁴Estimated Mincer returns on education typically exceed 6%, see e.g., Ashenfelter et al. (1999), Card (1999), and Harmon et al. (2003). In analogy of the equity premium puzzle, this raises the so-called human capital premium puzzle, see Judd (2000) and Palacios-Huerta (2006).

¹⁵See Jacobs and Bovenberg (2005) for a three-period version of the model with a retirement phase. The quantitative results are hardly affected.

¹⁶Jacobs and Bovenberg (2005) perform more extensive sensitivity analyses. They find that the case for substantial capital taxes is not lost – even for rather extreme parameter values that substantially deviate from empirically estimated values.

Figure 1: Optimal capital income taxes



Key: Optimal capital income taxes as a function of optimal labor income taxes. $r = 0.06$ per year, cohort length is 25 years, $\rho = 1$, $\omega = 0.5$, and $\bar{\gamma} = 0.25$. ‘Baseline’ scenario: $\sigma = 0.5$, $\beta = 0.6$, $\alpha = 0.5$. ‘Conservative’ scenario: $\sigma = 1.0$, $\beta = 0.3$, $\alpha = 0.75$.

tax rates for redistributive reasons, it should also levy large tax rates on capital income in order to combat the adverse impact of high marginal labor taxes on learning. We demonstrate that this applies to both linear and non-linear policies.

The optimal marginal tax on capital income is large compared to the optimal marginal tax on labor income if, compared to the level aggregate saving, the level and composition of learning efforts are relatively elastic with respect to financial incentives. Education subsidies reduce the need for capital taxes but do not eliminate the case for positive capital taxes as long as not all investments in human capital can be verified. The optimal capital tax is zero only in knife-edge cases in which either all investment in human capital can be verified or verifiable and non-verifiable inputs cannot be substituted. All labor-tax distortions on human capital accumulation can then be eliminated through education subsidies. Numerical simulations suggest that optimal capital taxes are substantially positive. Our analysis thus suggests that the welfare gains of replacing income taxes by consumption taxes are overestimated if human capital formation is ignored. Indeed,

consumption taxes are not neutral with respect to human capital formation if not all investments in human capital are tax-deductible (see also Judd, 1999).

This paper shows that policies to enhance human capital investment, private pension savings, labor supply and income redistribution are fundamentally related over the life-cycle and result in important trade-offs. Apart from the standard trade-off between equity and efficiency, our analysis emphasizes a second fundamental trade-off between promoting human capital formation and stimulating financial savings. This is important in light of the aging of the population and the Lisbon agenda. Indeed, many European governments simultaneously want to stimulate human capital investment, private pension savings and labor-force participation while maintaining income equality. This paper demonstrates that reconciling these objectives is a major challenge. To illustrate, promoting (pension) savings by means of subsidies on saving raises the opportunity returns on human capital investments and thus induces households to substitute financial saving for human capital investment.¹⁷ Furthermore, containing inequality by more progressive labor taxation distorts human capital formation. In addition, more progressive labor taxes erode the tax base not only by reducing hours worked, but also by promoting earlier retirement. Policy makers should therefore take into account these life-cycle interactions between saving, working and learning.

References

- Aiyagari, S. Rao (1995), “Optimal Capital Income Taxation with Incomplete Markets, Borrowing Constraints, and Constant Discounting”, *Journal of Political Economy*, 103, (6), 1158-1175.
- Anderberg, Dan (2008), “Optimal Policy and the Risk Properties of Human Capital Reconsidered”, CESifo Working Paper No. 2451, München: CESifo.
- Ashenfelter, Orly, Colm Harmon, and Hessel Oosterbeek (1999), “A Review of Estimates of the Schooling/Earnings Relationship, with Tests for Publication Bias”, *Labour Economics*, 6, 453-470.
- Atkinson, Anthony B., and Joseph E. Stiglitz (1976), “The Design of Tax Structure: Direct versus Indirect Taxation”, *Journal of Public Economics*, 6, 55–75.
- Atkinson, Anthony B., and Joseph E. Stiglitz (1980), *Lectures on Public Economics*, New York: McGraw-Hill.
- Atkinson, Anthony B., and Agnar Sandmo (1980), ”Welfare Implications of the Taxation of Savings”, *Economic Journal* 90, 529-549.
- Attanasio, Orazio P., and Guglielmo Weber (1995), “Is Consumption Growth Consistent with Intertemporal Optimization?”, *Journal of Political Economy*, 103, (6), 1121-1157.
- Becker, Gary S. (1964), *Human Capital: A Theoretical and Empirical Analysis with Special Reference to Education*, Third edition 1993, Chicago: Chicago University Press.
- Bernheim, B. Douglas (2002), “Taxation and Saving”, in: Auerbach, Alan J., and Martin Feldstein (eds.), *Handbook of Public Economics*, Vol 3., Chapter 18, North-Holland: Elsevier.

¹⁷This is reminiscent of Edlin (1991) and Feldstein (1992), who argue that (means-tested) subsidies on education are implicit taxes on financial savings.

- Boskin, Michael (1975), “Notes on the Tax Treatment of Human Capital”, NBER Working Paper 116, Cambridge-MA: NBER.
- Bovenberg, A. Lans, and Bas Jacobs (2003), “Redistribution and Education Subsidies are Siamese Twins”, Tinbergen Institute Discussion Paper 05-037/3, Amsterdam/Rotterdam: Tinbergen Institute.
- Bovenberg, A. Lans, and Bas Jacobs (2005), “Redistribution and Education Subsidies are Siamese Twins”, *Journal of Public Economics*, 89, 2005-2035.
- Cameron, Stephen, and James J. Heckman (2001), “The Dynamics of Educational Attainment for Black, Hispanic, and White Males”, *Journal of Political Economy*, 109, 455-499.
- Card, David (1999), “The Causal Effect of Education on Earnings”, in: Ashenfelter, Orley, and David Card (eds.), *Handbook of Labor Economics – Vol 3A*, Amsterdam: Elsevier-North Holland.
- Carneiro, Pedro, and James J. Heckman (2002), “The Evidence on Credit Constraints in Post-Secondary Schooling”, *Economic Journal*, 112, 705-734.
- Chamley, Christophe (1986), “Optimal Taxation of Capital Income in General Equilibrium with Infinite Lives”, *Econometrica*, 54 (3), 607-622.
- da Costa, Carlos E., and Lucas J. Maestri (2007), “The Risk-Properties of Human Capital and the Design of Government Policies”, *European Economic Review*, 51, 695–713.
- Diamond, Peter A. (1975), “A Many-Person Ramsey Rule”, *Journal of Public Economics*, 4, (4), 335-342.
- Diamond, Peter A. (1998), “Optimal Income Taxation: An Example with a U-Shaped Pattern of Optimal Marginal Tax Rates”, *American Economic Review*, 88, 83-95.
- Diamond, Peter A. (2006), “Comment on New Dynamic Public Finance: A User’s Guide”, *NBER Macroeconomic Annual*, 365-379.
- Diamond, Peter A., and James A. Mirrlees (1971), “Optimal Taxation and Public Production I: Production Efficiency”, *American Economic Review*, 61, 8-27.
- Edlin, Aaron S. (1993), “Is College Financial Aid Equitable and Efficient?”, *Journal of Economic Perspectives*, 7 (2), 143-158.
- Erosa, Andrés, and Martin Gervais (2002), “Optimal Taxation in Life-Cycle Economies”, *Journal of Economic Theory*, 105, 338-369.
- Feldstein, Martin (1995), “College Scholarship Rules and Private Savings”, *American Economic Review*, 85, 552-566.
- Fischer, Stanley (1980), “Dynamic Inconsistency, Cooperation, and the Benevolent Dissembling Government”, *Journal of Economic Dynamics and Control*, 2, 93-107.
- Grochulski, Borys, and Tomasz Piskorski (2005), “Optimal Wealth Taxes with Risky Human Capital”, Federal Reserve Bank of Richmond Working Paper No. 05-13, Richmond.

- Golosov, Mikhail, Nayarana Kocherlakota, and Aleh Tsyvinski (2003), “Optimal Indirect and Capital Taxation”, *Review of Economic Studies*, 70, 569-587.
- Golosov, Mikhail, Aleh Tsyvinski, and Ivan Werning (2006), “New Dynamic Public Finance: A User’s Guide”, *NBER Macroeconomic Annual*, 317–363.
- Hall, Robert E. (1988), “Intertemporal Substitution in Consumption”, *Journal of Political Economy*, 96, (2), 339-357.
- Harmon, Colm, Hessel Oosterbeek, and Ian Walker (2003), “The Returns to Education: Microeconomics”, *Journal of Economic Surveys*, 17, 115-155.
- Heckman, James J. (1976), “A Life-Cycle Model of Earnings, Learning and Consumption”, *Journal of Political Economy*, 4, S11-S44.
- Jacobs, Bas, and A. Lans Bovenberg (2005), “Human Capital and Optimal Positive Taxation of Capital Income”, CEPR Discussion Paper No. 5047, London: CEPR.
- Jacobs, Bas and Lans Bovenberg (2008), “Optimal Taxation of Human Capital and the Earnings Function”, CESifo Working Paper No. 2250, München: CESifo.
- Jones, Larry E., Rodolfo E. Manuelli, and Peter E. Rossi (1993), “Optimal Taxation in Models of Endogenous Technological Change”, *Journal of Political Economy*, 101 (3), 485-517.
- Jones, Larry E., Rodolfo E. Manuelli, and Peter E. Rossi (1997), “On the Optimal Taxation of Capital Income”, *Journal of Economic Theory*, 73, 93-117.
- Judd, Kenneth L. (1985), “Redistributive Taxation in a Simple Perfect Foresight Model“, *Journal of Public Economics*, 28, 59-83.
- Judd, Kenneth L. (1999), “Optimal Taxation and Spending in General Competitive Growth Models”, *Journal of Public Economics*, 71, 1-26.
- Judd, Kenneth L. (2000), “Is Education as Good as Gold?”, mimeo: Hoover Institute.
- Kydland, Finn, and Edward C. Prescott (1977), “Rules Rather than Discretion: The Inconsistency of Optimal Plans”, *Journal of Political Economy*, 85, (3), 473-492.
- Mirrlees, James A. (1971), “An Exploration in the Theory of Optimum Income Taxation”, *Review of Economic Studies*, 38, 175-208.
- Nielsen, Søren B., and Peter B. Sørensen (1997), “On the Optimality of the Nordic System of Dual Income Taxation”, *Journal of Public Economics*, 63, 311-329.
- Ordober, Janush A., and Edmund S. Phelps (1979), “The Concept of Optimal Taxation in the Overlapping-Generations Model of Capital and Wealth”, *Journal of Public Economics* 12, 1-26.
- Palacios-Huerta, Ignacio (2006), “Risk and Market Frictions as Determinants of the Human Capital Premium”, mimeo: Brown University.
- Pigou, Arthur C. (1928, reprint third edition 1949), *A Study in Public Finance*, London: MacMillan & Co. Ltd.

Pirttilä, Jukka, and Matti Tuomala (2001), “On Optimal Non-linear Taxation and Public Good Provision in an Overlapping Generations Economy”, *Journal of Public Economics*, 79, 485-501.

Saez, Emmanuel (2001), “Using Elasticities to Derive Optimal Income Tax Rates,” *Review of Economic Studies*, 68, 205-229.

Saez, Emmanuel (2002), “The Desirability of Commodity Taxation Under Non-Linear Income Taxation and Heterogeneous Tastes”, *Journal of Public Economics*, 83, 217-230.

Seade, Jesus (1977), “On the Shape of Optimal Tax Schedules”, *Journal of Public Economics*, 7, 203-236.

Trostel, Philip A. (1993), “The Effect of Taxation on Human Capital”, *Journal of Political Economy*, 101, 327-350.

Appendix

Appendix A: Second-order conditions

In the first stage of a three-step budgeting procedure, individuals select aggregate consumption v_n , labor supply l_n (leisure demand $l_n \equiv 1 - l_n$), and aggregate investment in human capital $e_n \equiv \psi(x_n, y_n)$. We write the individual budget constraint as $p_c v_n = (1 - t)n\phi(e_n)l_n - Rp_e e_n + Ra_o + g$, where the price indices of aggregate consumption and learning are defined as $p_c \equiv (Rc_n^1 + c_n^2)/v(c_n^1, c_n^2)$ and $p_e \equiv ((1 - t)(1 - s_x)p_x x_n + p_y y_n)/e_n$. p_c and p_e do not depend on n due to the homotheticity of the sub-utility function $v(\cdot)$ and the sub-production function $\psi(\cdot)$, respectively. By employing the intertemporal budget constraint of households (5) to eliminate v_n from the utility function, we arrive at the following maximization problem:

$$\max_{\{l_n, e_n\}} u \left(\frac{(1 - t)n\phi(e_n)l_n}{p_c} - \frac{Rp_e e_n}{p_c} + \frac{g + Ra_o}{p_c}, 1 - l_n \right), \quad (54)$$

The first-order conditions for education and labor supply are

$$(1 - t)nl_n\phi'(e_n) - Rp_e = 0, \quad (55)$$

$$u_v(1 - t)n\phi(e_n) - u_\ell p_c = 0. \quad (56)$$

The second-order partial derivatives are ordered in the Hessian matrix

$$H \equiv \begin{bmatrix} (1 - t)nl_n\phi''(e_n) & (1 - t)n\phi'(e_n) \\ u_v(1 - t)n\phi'(e_n) & -u_{\ell v}(1 - t)n\phi(e_n) + u_{\ell\ell}p_c \end{bmatrix}. \quad (57)$$

The Hessian matrix should be negative definite for utility to reach a maximum. The first principal minor is $(1 - t)nl_n\phi''(e_n) < 0$. Therefore, the second leading principal minor must be positive. Given the production function $\phi(e_n) = e_n^\beta$, this implies

$$\beta(1 + \epsilon_n^*) < 1, \quad (58)$$

where $\epsilon_n^* \equiv -(l_n u_{\ell\ell}/u_\ell - l_n u_{\ell v}/u_v)^{-1}$. To relate this inequality to $\beta(1 + \epsilon_n) < 1$, we link ϵ_n^* to the partial compensated wage elasticity of labor supply ϵ_n . By differentiating the first-order condition

$u_\ell/u_v = w_n/p_c$ (where $w_n \equiv (1-t)n\phi(e_n)$ is the after-tax wage) with respect to l_n and substituting $\frac{dw_n}{dl_n} = \frac{u_\ell}{u_v}$ (derived by taking the total differential of utility, which is held constant), we can write the compensated wage elasticity of labor supply ϵ_n as

$$\epsilon_n \equiv \frac{\partial l_n}{\partial w_n} \frac{w_n}{l_n} = \left(\left(\frac{l_n u_{v\ell}}{u_v} - \frac{l_n u_{\ell\ell}}{u_\ell} \right) - \left(\frac{u_\ell}{u_v} \right) \left(\frac{l_n u_{vv}}{u_v} - \frac{l_n u_{v\ell}}{u_\ell} \right) \right)^{-1}, \quad (59)$$

so that $(\epsilon_n)^{-1} = (\epsilon_n^*)^{-1} + \varrho$, where $\varrho \equiv -(u_\ell/u_v)(l_n u_{vv}/u_v - l_n u_{v\ell}/u_\ell) \geq 0$. We thus have $\epsilon_n \leq \epsilon_n^*$ and

$$\beta(1 + \epsilon_n) \leq \beta(1 + \epsilon_n^*) < 1. \quad (60)$$

Therefore, the second-order condition implies the inequality in the text.

In the second stage of the budgeting procedure, individuals maximize $v(c_n^1, c_n^2)$ subject to the constraint: $p_c v_n = R c_n^1 + c_n^2$. The associated second-order condition, $v_{c^1 c^1} v_{c^2 c^2} - (v_{c^1 c^2})^2 > 0$, is satisfied because $v(c_n^1, c_n^2)$ is strictly concave. In the third stage, individuals allocate their expenditures on education e_n over x_n and y_n by maximizing $e_n = \psi(x_n, y_n)$ subject to the expenditure constraint $p_e e_n = (1-t)(1-s)p_x x_n + p_y y_n$. Concavity of $\psi(x_n, y_n)$ implies the resulting second-order condition $\psi_{xx}\psi_{yy} - (\psi_{xy})^2 > 0$ is met.

Appendix B: Compensated elasticities

To find the various compensated behavior elasticities with respect to policy variables, we multiply the first-order condition for x_n (i.e., (35)) by $(1-t)x_n$, multiply the first-order condition for y_n (i.e., (36)) by y_n , and add the results. Using $\psi_x x_n + \psi_y y_n = \psi(x_n, y_n) = e_n$ (since $\psi(x_n, y_n)$ is homogeneous of degree one), we find

$$\beta(1-t)n l_n \phi(e_n) = R p_e e_n, \quad (61)$$

where $p_e \equiv ((1-t)(1-s)p_x x_n + p_y y_n)/e_n$.

Log-linearizing (61) (using $\phi(e_n) = e_n^\beta$) and the definition of p_e (using $e_n \equiv \psi(x_n, y_n)$ and the first-order conditions for x_n and y_n (i.e., (35) and (36))), we arrive at

$$-\tilde{t} + \tilde{l}_n + \beta \tilde{e}_n = \tilde{p}_e + \tilde{e}_n + \tilde{R}, \quad (62)$$

$$\tilde{p}_e = -\alpha(\tilde{t} + \tilde{s}), \quad (63)$$

where a tilde stands for a *compensated* relative change, e.g., $\tilde{e}_n \equiv de_n/e_n$, except for the tax rate and the subsidy rates where $\tilde{t} \equiv dt/(1-t)$ and $\tilde{s} \equiv ds/(1-s)$. $\alpha \equiv x_n \psi_x / \psi = (1-t)(1-s)p_x x_n / (p_e e_n)$ does not depend on ability. The reason is that the first-order conditions for x_n and y_n ((35) and (36)) imply that the marginal rate of transformation does not depend on n :

$$\frac{\psi_x}{\psi_y} = \frac{(1-t)(1-s)p_x}{p_y}. \quad (64)$$

ψ_x and ψ_y are functions of x_n/y_n only (since $\psi(x_n, y_n)$ is homogeneous of the first degree). (64) determines the ratio of the two inputs x_n/y_n as a function of $(1-t)(1-s)p_x/p_y$ only. $\alpha = (1-t)(1-s)p_x(x_n/y_n) / ((1-t)(1-s)p_x(x_n/y_n) + p_y)$ therefore depends only on $(1-t)(1-s)p_x$ and p_y and not on ability.

The first-order condition for labor supply (56) can be written as

$$\frac{u_\ell}{u_v} = \frac{w_n}{p_c}, \quad (65)$$

where $w_n \equiv (1-t)n\phi(e_n)$ and $p_c \equiv (Rc_n^1 + c_n^2)/v(c_n^1, c_n^2)$. Loglinearizing this expression yields

$$\left(\frac{u_v l_n}{u_v} - \frac{u_{\ell} l_n}{u_{\ell}}\right) \tilde{l}_n + \left(\frac{u_{\ell v} v_n}{u_{\ell}} - \frac{u_{vv} v_n}{u_v}\right) \tilde{v}_n = \tilde{w}_n - \tilde{p}_c. \quad (66)$$

To find the compensated relative changes in labor supply l_n and composite consumption v_n , we differentiate the utility function $u(v_n, 1 - l_n)$ and set the total differential to zero to arrive at:

$$\frac{u_v v_n}{u_{\ell} l_n} \tilde{v}_n = \tilde{l}_n. \quad (67)$$

Substituting (67) into (66) to eliminate \tilde{v}_n , we find

$$\tilde{l}_n = \epsilon_n (\tilde{w}_n - \tilde{p}_c), \quad \epsilon_n \equiv \frac{\partial l_n^*}{\partial w_n} \frac{w_n}{l_n}, \quad (68)$$

where ϵ_n is the compensated wage elasticity of labor supply with respect to the after-tax wage given by (59) in Appendix A.

Substitution of the first-order condition for labor supply (65) to eliminate u_v in (67) gives (by using (11) to write

$$\frac{p_c v_n}{w_n l_n} = \frac{Rc_n^1 + c_n^2}{(1-t)z_n} = (1 - \beta + \gamma_n), \quad \gamma_n \equiv \frac{Ra_o + g}{(1-t)z_n}, \quad (69)$$

Hence, we obtain

$$(1 - \beta + \gamma_n) \tilde{v}_n = \tilde{l}_n. \quad (70)$$

Substitution of (68) in the last equation solves for \tilde{v}_n

$$\tilde{v}_n = \frac{\epsilon_n}{(1 - \beta + \gamma_n)} (\tilde{w}_n - \tilde{p}_c). \quad (71)$$

The log-linearized sub-utility function v_n can be written as

$$\tilde{v}_n = (1 - \omega) \tilde{c}_n^1 + \omega \tilde{c}_n^2. \quad (72)$$

By log-linearizing the definition of p_c , i.e., $p_c v(c_n^1, c_n^2) = Rc_n^1 + c_n^2$, and using (72), we obtain $\tilde{p}_c = (1 - \omega) \tilde{R}$. Substituting this result in (71) and (66), we find

$$\tilde{v}_n = \frac{\epsilon_n}{(1 - \beta + \gamma_n)} (\tilde{w}_n - (1 - \omega) \tilde{R}), \quad (73)$$

$$\tilde{l}_n = \epsilon_n (\tilde{w}_n - (1 - \omega) \tilde{R}). \quad (74)$$

Substituting $\tilde{w}_n = -\tilde{t} + \beta \tilde{e}_n$ (which follows from $w_n \equiv (1-t)n\phi(e_n)$) into (74), and then substituting the result and (63) into (62) to eliminate, respectively, \tilde{l}_n and \tilde{p}_e , we can solve for \tilde{e}_n :

$$\tilde{e}_n = -\frac{(1 + \epsilon_n - \alpha) \tilde{t}}{\mu_n} - \frac{(1 + (1 - \omega) \epsilon_n)}{\mu_n} \tilde{R} + \frac{\alpha}{\mu_n} \tilde{s}, \quad (75)$$

where $\mu_n \equiv 1 - \beta(1 + \epsilon_n) > 0$ from the second-order conditions. Hence, we find for the wage rate

$$\tilde{w}_n = -\left(\frac{1 - \alpha\beta}{\mu_n}\right) \tilde{t} + \frac{\alpha\beta}{\mu_n} \tilde{s} - \frac{\beta(1 + (1 - \omega) \epsilon_n)}{\mu_n} \tilde{R}. \quad (76)$$

Substituting $\tilde{w}_n = -\tilde{t} + \beta\tilde{e}_n$ and (75) into (74), we solve for \tilde{l}_n :

$$\tilde{l}_n = -\frac{\epsilon_n(1-\alpha\beta)}{\mu_n}\tilde{t} + \frac{\epsilon_n\alpha\beta}{\mu_n}\tilde{s} - \epsilon_n\left(\frac{1-\omega(1-\beta)}{\mu_n}\right)\tilde{R}. \quad (77)$$

Consumption in the two periods is found by log-linearizing (9):

$$\tilde{c}_n^1 - \tilde{c}_n^2 = -\sigma\tilde{R}, \quad (78)$$

where $\sigma \equiv d\ln(c_n^1/c_n^2)/d\ln(u_{c^2}/u_{c^1})$ is the intertemporal substitution elasticity in consumption. This elasticity does not depend on ability n . The reason is that the left-hand side of (9) depends only on c_n^1/c_n^2 (as the sub-utility function $v(c_n^1, c_n^2)$ is homothetic), so that one can express c_n^1/c_n^2 as a function of R only.

Substitution of (78) into (72) to eliminate \tilde{c}_n^1 while using (70) to eliminate \tilde{v}_n and substitution of (77) to eliminate \tilde{l}_n yields

$$\tilde{c}_n^1 = \frac{\epsilon_n}{(1-\beta+\gamma_n)}\left(-\frac{(1-\alpha\beta)}{\mu_n}\tilde{t} + \frac{\alpha\beta}{\mu_n}\tilde{s} - \left(\frac{1-\omega(1-\beta)}{\mu_n}\right)\tilde{R}\right) - \sigma\omega\tilde{R}, \quad (79)$$

$$\tilde{c}_n^2 = \frac{\epsilon_n}{(1-\beta+\gamma_n)}\left(-\frac{(1-\alpha\beta)}{\mu_n}\tilde{t} + \frac{\alpha\beta}{\mu_n}\tilde{s} - \left(\frac{1-\omega(1-\beta)}{\mu_n}\right)\tilde{R}\right) + (1-\omega)\sigma\tilde{R}. \quad (80)$$

In the first part of the paper, all learning is non verifiable (i.e., $\alpha = 0$). Hence, the elasticities of learning, labor supply, first- and second-period consumption follow from (use (75), (77), (79) and (80) with $\alpha = 0$).

The second part of the paper analyzes $\alpha > 0$. We find \tilde{x}_n and \tilde{y}_n by differentiating $e_n \equiv \psi(x_n, y_n)$ and using the first-order conditions for x_n and y_n to eliminate ψ_x and ψ_y to arrive at

$$\tilde{e}_n = \alpha\tilde{x}_n + (1-\alpha)\tilde{y}_n. \quad (81)$$

Differentiation of (64) yields

$$\tilde{x}_n - \tilde{y}_n = \rho(\tilde{t} + \tilde{s}), \quad (82)$$

where $\rho \equiv d\ln(x_n/y_n)/d\ln(\psi_y/\psi_x)$ stands for the elasticity of substitution between the two inputs in the production of human capital. This substitution elasticity depends only on x_n/y_n and is independent of ability because x_n/y_n is a function of $(1-t)(1-s)p_x/p_y$ only (see the discussion above on α being independent of ability).

From (81) and (82), we can express \tilde{x}_n and \tilde{y}_n in terms of \tilde{e}_n and then substitute (75) to eliminate learning:

$$\tilde{x}_n = -\frac{(1+\epsilon_n-\alpha)}{\mu_n}\tilde{t} - \frac{(1+(1-\omega)\epsilon_n)}{\mu_n}\tilde{R} + \frac{\alpha}{\mu_n}\tilde{s} + (1-\alpha)\rho(\tilde{t} + \tilde{s}), \quad (83)$$

$$\tilde{y}_n = -\frac{(1+\epsilon_n-\alpha)}{\mu_n}\tilde{t} - \frac{(1+(1-\omega)\epsilon_n)}{\mu_n}\tilde{R} + \frac{\alpha}{\mu_n}\tilde{s} - \alpha\rho(\tilde{t} + \tilde{s}). \quad (84)$$

These expressions yield the elasticities for x_n and y_n with respect to changes in policy.

To find optimal non-linear policies, we need to derive the relative change of savings with respect to all policy parameters in the case that learning is completely non verifiable (i.e., $\alpha = 0$). The change in savings follows from the first-period budget constraint:

$$c_n^1\tilde{c}_n^1 + p_e e_n \tilde{e}_n = -\frac{(1-t)z_n}{R}\tilde{a}_n, \quad (85)$$

where we defined $\tilde{a}_n \equiv Rda_n/((1-t)z_n)$. After substituting the proportionality factors $e_n = (1-t)\beta z_n/(p_e R)$ and $c_n^1 = (1-\omega)(1-\beta+\gamma_n)(1-t)z_n/R$, we find

$$(1-\omega)(1-\beta+\gamma_n)\tilde{c}_n^1 + \beta\tilde{e}_n = -\tilde{a}_n. \quad (86)$$

Substituting (79) to eliminate \tilde{c}_n^1 and (75) to eliminate \tilde{e}_n , we arrive at

$$\begin{aligned} \tilde{a}_n &= \left(\frac{(1-\omega)\epsilon_n + \beta(1+\epsilon_n)}{\mu_n} \right) \tilde{t} \\ &+ \left(\frac{(1-\omega)\epsilon_n(1-\omega(1-\beta)) + \beta(1+(1-\omega)\epsilon_n)}{\mu_n} + (1-\omega)(1-\beta+\gamma_n)\sigma\omega \right) \tilde{R}. \end{aligned} \quad (87)$$

Finally, the elasticities of all variables with respect to the policy variables are found by taking the coefficients of each policy variable in the linearized equations: $\varepsilon_{qj} \equiv \frac{\partial \tilde{q}}{\partial j}$ with $q = e_n, x_n, y_n, c_n^1, c_n^2, l_n, a_n$, and $j = t, R, s$.

Appendix C: Optimal linear taxation

The Lagrangian for maximizing social welfare amounts to

$$\begin{aligned} \max_{\{g,t,R,s\}} \mathcal{L} &= \int_{\underline{n}}^{\bar{n}} \Psi(v(g,t,R,s,n)) dF(n) \\ &+ \eta \int_{\underline{n}}^{\bar{n}} [t(nl_n\phi(e_n) - (1-s)p_x x_n(1+r))] dF(n) \\ &+ \eta \int_{\underline{n}}^{\bar{n}} [(1+r-R)(a_o - c_n^1 - (1-t)(1-s)p_x x_n - p_y y_n)] dF(n) \\ &- \eta \int_{\underline{n}}^{\bar{n}} [(1+r)sp_x x_n + g + \Lambda] dF(n). \end{aligned} \quad (88)$$

The first-order condition for maximization of social welfare with respect to g is given by (using Roy's lemma $\partial v(g,t,R,s,n)/\partial g = \lambda_n$, where λ_n denotes the private marginal utility of income)

$$\frac{\partial \mathcal{L}}{\partial g} = \int_{\underline{n}}^{\bar{n}} \left[\Psi' \lambda_n - \eta + \eta \Delta_x p_x \frac{\partial x_n}{\partial g} + \eta \Delta_y p_y \frac{\partial y_n}{\partial g} + \eta t n \phi(e_n) \frac{\partial l_n}{\partial g} - \eta \tau r \frac{\partial c_n^1}{\partial g} \right] dF(n) = 0, \quad (89)$$

where $\Delta_x \equiv -(Rs + \tau r)$ and $\Delta \equiv tR/(1-t) - \tau r$ denote the tax wedges on x_n and y_n , respectively.

By defining the net marginal social value of income (including the effects on the tax base) as

$$b_n \equiv \frac{\Psi' \lambda_n}{\eta} + \Delta_x p_x \frac{\partial x_n}{\partial g} + \Delta_y p_y \frac{\partial y_n}{\partial g} + t n \phi(e_n) \frac{\partial l_n}{\partial g} - \tau r \frac{\partial c_n^1}{\partial g}, \quad (90)$$

we find

$$\int_{\underline{n}}^{\bar{n}} b_n dF(n) = 1. \quad (91)$$

Using Roy's lemma (i.e., $\partial v(g,t,R,s,n)/\partial t = -\lambda_n(z_n - R(1-s)p_x x_n)$, $\partial v(g,t,R,s,n)/\partial R = \lambda_n a_n = \lambda_n(a_o - c_n^1 - (1-t)(1-s)p_x x_n - p_y y_n)$, and $\partial v(g,t,R,s,n)/\partial s = \lambda_n R(1-t)p_x x_n$), we

write the first-order conditions for t , R , and s as

$$\begin{aligned} \frac{\partial \mathcal{L}}{\partial t} &= \int_{\underline{n}}^{\bar{n}} [-(\Psi' \lambda - \eta) (z_n - R(1-s)p_x x_n)] dF(n) \\ &+ \int_{\underline{n}}^{\bar{n}} \eta \left[\Delta_x p_x \frac{\partial x_n}{\partial t} + \Delta p_y \frac{\partial y_n}{\partial t} + t n \phi(e_n) \frac{\partial l_n}{\partial t} - \tau r \frac{\partial c_n^1}{\partial t} \right] dF(n) = 0, \end{aligned} \quad (92)$$

$$\begin{aligned} \frac{\partial \mathcal{L}}{\partial R} &= \int_{\underline{n}}^{\bar{n}} [(\Psi' \lambda - \eta) a_n] dF(n) \\ &+ \int_{\underline{n}}^{\bar{n}} \eta \left[\Delta_x p_x \frac{\partial x_n}{\partial R} + \Delta p_y \frac{\partial y_n}{\partial R} + t n \phi(e_n) \frac{\partial l_n}{\partial R} - \tau r \frac{\partial c_n^1}{\partial R} \right] dF(n) = 0, \end{aligned} \quad (93)$$

$$\begin{aligned} \frac{\partial \mathcal{L}}{\partial s} &= \int_{\underline{n}}^{\bar{n}} [(\Psi' \lambda - \eta) R(1-t)p_x x_n] dF(n) \\ &+ \int_{\underline{n}}^{\bar{n}} \eta \left[\Delta_x p_x \frac{\partial x_n}{\partial s} + \Delta p_y \frac{\partial y_n}{\partial s} + t n \phi(e_n) \frac{\partial l_n}{\partial s} - \tau r \frac{\partial c_n^1}{\partial s} \right] dF(n) = 0. \end{aligned} \quad (94)$$

The Slutsky equations are given by $\frac{\partial q}{\partial t} \equiv \frac{\partial q^*}{\partial t} - z_n \frac{\partial q}{\partial g}$, $\frac{\partial q}{\partial R} \equiv \frac{\partial q^*}{\partial R} + a_n \frac{\partial q}{\partial g}$, and $\frac{\partial q}{\partial s} \equiv \frac{\partial q^*}{\partial s} + R(1-t)p_x x_n \frac{\partial q}{\partial g}$, where $q = x_n, y_n, l_n, c_n^1$, and the asterisk denotes a compensated change. Using the Slutsky equations, the definition of b_n (i.e., (90)), (91), and the proportionality factors between gross income z_n and the tax and subsidy bases (i.e., $x_n = \alpha \beta z_n / ((1-s)p_x R)$, $y_n = \beta(1-\alpha)(1-t)z_n / (p_y R)$, $Rc_n^1 = (1-\omega)(1-\beta+\gamma_n)(1-t)z_n$ (from (12))) into the first-order conditions for t , R , and s , we establish

$$\begin{aligned} &\int_{\underline{n}}^{\bar{n}} \left[(1-b_n)(1-\alpha\beta)z_n + \frac{\Delta_x}{(1-t)(1-s)R} \alpha \beta \varepsilon_{xt} z_n + \frac{\Delta}{R} \beta(1-\alpha) \varepsilon_{yt} z_n \right] dF(n) \\ &+ \int_{\underline{n}}^{\bar{n}} \left[\frac{t}{1-t} \varepsilon_{lt} z_n - \frac{\tau r}{R} (1-\omega)(1-\beta+\gamma_n) \varepsilon_{c^1 t} z_n \right] dF(n) = 0, \end{aligned} \quad (95)$$

$$\begin{aligned} &\int_{\underline{n}}^{\bar{n}} \left[(1-b_n)(1-\omega(1-\beta))z_n + \frac{\Delta_x}{(1-t)(1-s)R} \alpha \beta \varepsilon_{xR} z_n + \frac{\Delta}{R} \beta(1-\alpha) \varepsilon_{yR} z_n \right] dF(n) \\ &+ \int_{\underline{n}}^{\bar{n}} \left[\frac{t}{1-t} \varepsilon_{lR} z_n - \frac{\tau r}{R} (1-\omega)(1-\beta+\gamma_n) \varepsilon_{c^1 R} z_n \right] dF(n) = 0, \end{aligned} \quad (96)$$

$$\begin{aligned} &\int_{\underline{n}}^{\bar{n}} \left[(b_n-1) \alpha \beta z_n + \frac{\Delta_x}{(1-t)(1-s)R} \alpha \beta \varepsilon_{xs} z_n + \frac{\Delta}{R} \beta(1-\alpha) \varepsilon_{ys} z_n \right] dF(n) \\ &+ \int_{\underline{n}}^{\bar{n}} \left[\frac{t}{1-t} \varepsilon_{ls} z_n - \frac{\tau r}{R} (1-\omega)(1-\beta+\gamma_n) \varepsilon_{c^1 s} z_n \right] dF(n) = 0. \end{aligned} \quad (97)$$

where we defined the compensated elasticities as $\varepsilon_{qt} \equiv \frac{\partial q^*}{\partial t} \frac{(1-t)}{q}$, $\varepsilon_{qR} \equiv \frac{\partial q^*}{\partial R} \frac{R}{q}$, and $\varepsilon_{qs} \equiv \frac{\partial q^*}{\partial s} \frac{(1-s)}{q}$ for $q = x_n, y_n, l_n$ and c_n^1 and used $a_n \equiv a_o - c_n^1 - (1-t)(1-s)p_x x_n - p_y y_n$ to eliminate a_n in (96).

After substitution of the elasticities of labor supply and first-period consumption from Appendix B, we find the following expressions for the first-order conditions for t , R and s

$$\int_{\underline{n}}^{\bar{n}} \left[(1 - b_n) z_n + \frac{\Delta_x}{(1-t)(1-s)R} \frac{\alpha\beta}{(1-\alpha\beta)} \varepsilon_{xt} z_n + \frac{\Delta}{R} \frac{\beta(1-\alpha)}{(1-\alpha\beta)} \varepsilon_{yt} z_n \right] dF(n) + \int_{\underline{n}}^{\bar{n}} \left[- \left(\frac{t}{1-t} - \frac{\tau r}{R} (1-\omega) \right) \frac{\epsilon_n}{\mu_n} z_n \right] dF(n) = 0, \quad (98)$$

$$\int_{\underline{n}}^{\bar{n}} \left[(1 - b_n) z_n + \frac{\Delta_x}{(1-t)(1-s)R} \frac{\alpha\beta}{(1-\omega(1-\beta))} \varepsilon_{xR} z_n + \frac{\Delta}{R} \frac{\beta(1-\alpha)}{(1-\omega(1-\beta))} \varepsilon_{yR} z_n \right] dF(n) + \int_{\underline{n}}^{\bar{n}} \left[- \left(\frac{t}{1-t} - \frac{\tau r}{R} (1-\omega) \right) \frac{\epsilon_n}{\mu_n} z_n + \frac{\tau r}{R} (1-\omega) \sigma \omega \frac{(1-\beta+\gamma_n)}{(1-\omega(1-\beta))} z_n \right] dF(n) = 0, \quad (99)$$

$$\int_{\underline{n}}^{\bar{n}} \left[(b_n - 1) z_n + \frac{\Delta_x}{(1-t)(1-s)R} \varepsilon_{xs} z_n + \frac{\Delta}{R} \frac{\beta(1-\alpha)}{\alpha\beta} \varepsilon_{ys} z_n \right] dF(n) + \int_{\underline{n}}^{\bar{n}} \left[\left(\frac{t}{1-t} - \frac{\tau r}{R} (1-\omega) \right) \frac{\epsilon_n}{\mu_n} z_n \right] dF(n) = 0. \quad (100)$$

To find the optimal structure of education subsidies, add equation (98) to equation (100) to obtain

$$\frac{\Delta_x}{(1-t)(1-s)R} \left(\varepsilon_{xs} + \frac{\beta\alpha}{(1-\alpha\beta)} \varepsilon_{xt} \right) = - \frac{\Delta}{R} \left(\frac{(1-\alpha)}{\alpha} \varepsilon_{ys} + \frac{\beta(1-\alpha)}{(1-\alpha\beta)} \varepsilon_{yt} \right). \quad (101)$$

Use the the relevant elasticities from Appendix B to find

$$\frac{\Delta_x}{(1-t)(1-s)R} = - \frac{\Delta}{R} \left(\frac{(1-\alpha)(1-\rho)}{1-(1-\alpha)(1-\rho)} \right). \quad (102)$$

By substituting $\Delta_x \equiv -(Rs + \tau r)$, we arrive at equation (38) in the main text.

To find equation (39), substitute equation (102) in equations (98) and (99) to eliminate Δ_x and substitute the relevant elasticities from Appendix B:

$$\xi = \frac{\Delta\varphi\beta}{R} \overline{\left[\frac{1+\epsilon_n}{\mu_n} \right]} + \left(\frac{t}{1-t} - \frac{\tau r}{R} (1-\omega) \right) \overline{\left[\frac{\epsilon_n}{\mu_n} \right]}, \quad (103)$$

$$\xi = \frac{\Delta\varphi\beta}{R} \overline{\left[\frac{(1+(1-\omega)\epsilon_n)}{\mu_n(1-\omega(1-\beta))} \right]} + \left(\frac{t}{1-t} - \frac{\tau r}{R} (1-\omega) \right) \overline{\left[\frac{\epsilon_n}{\mu_n} \right]} - \frac{\tau r}{R} (1-\omega) \sigma \omega \frac{(1-\beta+\bar{\gamma})}{(1-\omega(1-\beta))}, \quad (104)$$

where $\varphi \equiv (1 - (1-\alpha)(1-\rho))^{-1} (1-\alpha)\rho$. Subtracting the latter expression from the former, we arrive at (39).

Substituting $\Delta \equiv tR/(1-t) - \tau r$ into (39) to eliminate Δ , we find the dual income tax in (40). The optimal labor tax in equation (41) follows from substituting $\Delta \equiv tR/(1-t) - \tau r$ into equation (103) to eliminate Δ and then eliminating $\tau r/R$ by using (40). The total tax wedge in equation (42) is derived by substituting (40) at the left-hand side of (42) to eliminate $(1-\omega)\tau r/R$ and then using (41) to eliminate $t/(1-t)$.

The corresponding expressions for the case without verifiable learning in section 3 are found by setting $\alpha = 0$.

Appendix D: Optimal non-linear taxation

The first-order conditions for the optimal control problem (49) are given by

$$\begin{aligned} \frac{\partial \mathcal{H}}{\partial l_n} &= \eta \left(n\phi(e_n) + nl_n\phi'(e_n) \frac{de_n}{dl_n} \Big|_{\bar{a}, \bar{u}} - \frac{dc_n^2}{dl_n} \Big|_{\bar{a}, \bar{u}} \right) f(n) \\ &\quad - \frac{\theta_n u_\ell}{n} \left(1 - \frac{l_n u_{\ell\ell}}{u_\ell} - \frac{l_n u_{\ell v} v_{c^1}}{u_\ell} p_e \frac{de_n}{dl_n} \Big|_{\bar{a}, \bar{u}} + \frac{l_n u_{\ell v} v_{c^2}}{u_\ell} \frac{dc_n^2}{dl_n} \Big|_{\bar{a}, \bar{u}} \right) = 0, \end{aligned} \quad (105)$$

$$\begin{aligned} \frac{\partial \mathcal{H}}{\partial a_n} &= \eta \left((1+r) + l_n n\phi'(e_n) \frac{de_n}{da_n} \Big|_{\bar{l}, \bar{u}} - \frac{dc_n^2}{da_n} \Big|_{\bar{l}, \bar{u}} \right) f(n) \\ &\quad - \frac{\theta_n l_n}{n} \left(-u_{\ell v} v_{c^1} (1 + p_e \frac{de_n}{da_n} \Big|_{\bar{l}, \bar{u}}) + u_{\ell v} v_{c^2} \frac{dc_n^2}{da_n} \Big|_{\bar{l}, \bar{u}} \right) = 0, \end{aligned} \quad (106)$$

$$\frac{\partial \mathcal{H}}{\partial u_n} = \Psi'(\cdot) f(n) - \eta \frac{dc_n^2}{du_n} \Big|_{\bar{l}, \bar{a}} f(n) - \frac{\theta_n l_n}{n} u_{\ell v} v_{c^2} \frac{dc_n^2}{du_n} \Big|_{\bar{l}, \bar{a}} = \frac{d\theta_n}{dn}, \quad (107)$$

where we employed $\frac{de_n}{du_n} \Big|_{\bar{l}, \bar{a}} = 0$ by the envelope theorem.

To find the optimal capital tax, totally differentiate utility $u(v(a_0 - a_n - p_e e_n, c_n^2), 1 - l_n)$ at constant l_n and u_n , and substitute the first-order condition for consumption in (45) to find

$$\left(1 + p_e \frac{de_n}{da_n} \Big|_{\bar{l}, \bar{u}} \right) \frac{v_{c^1}}{v_{c^2}} = \left(1 + p_e \frac{de_n}{da_n} \Big|_{\bar{l}, \bar{u}} \right) R_n = \frac{dc_n^2}{da_n} \Big|_{\bar{l}, \bar{u}}. \quad (108)$$

Substitution of these results into (106) implies that the distribution term associated with θ_n drops out. Substituting the first-order condition for learning (46) and simplifying, we arrive at:

$$\frac{rT'_a(ra_n)}{1 + (1 - T'_a(ra_n))r} = -\frac{T'_z(z_n)}{1 - T'_z(z_n)} \beta_n \varepsilon_{ea}, \quad \varepsilon_{ea} \equiv \frac{(1 - T'_z)z_n}{R_n e_n} \frac{de_n}{da_n} \Big|_{\bar{l}, \bar{u}}, \quad (109)$$

and $\beta_n \equiv \phi'(e_n)e_n/\phi(e_n)$.

To find the compensated elasticity ε_{ea} , we use the marginal tax rates on labor and capital income to steer the allocation such that labor supply is fixed and use these changes in marginal tax rates to find $\varepsilon_{ea} = \tilde{e}_n/\tilde{a}_n$ (see also Bovenberg and Jacobs, 2003). We have

$$\tilde{l}_n = \varepsilon_{lt}\tilde{t} + \varepsilon_{lR}\tilde{R} = 0, \quad (110)$$

$$\tilde{e}_n = \varepsilon_{et}\tilde{t} + \varepsilon_{eR}\tilde{R}, \quad (111)$$

$$\tilde{a}_n = \varepsilon_{at}\tilde{t} + \varepsilon_{aR}\tilde{R}. \quad (112)$$

This set of equations can be solved for \tilde{e}_n/\tilde{a}_n by using the first equation to eliminate $\tilde{R} = -(\varepsilon_{lt}/\varepsilon_{lR})\tilde{t}$ and substituting this in the equations for \tilde{e}_n and \tilde{a}_n to find

$$\varepsilon_{ea} = \frac{\tilde{e}_n}{\tilde{a}_n} \Big|_{\bar{l}, \bar{u}} = \frac{\varepsilon_{et}\varepsilon_{lR} - \varepsilon_{eR}\varepsilon_{lt}}{\varepsilon_{at}\varepsilon_{lR} - \varepsilon_{aR}\varepsilon_{lt}}. \quad (113)$$

After substitution of the relevant compensated elasticities from Appendix B, we obtain

$$\varepsilon_{ea} = -\frac{1}{\beta_n + (1 - \omega)(1 - \beta_n + \gamma_n)\sigma}. \quad (114)$$

Using this result in (109) to eliminate ε_{ea} , we arrive at (51).

To find the optimal labor tax, totally differentiate utility $u(v(a_0 - a_n - p_e e_n, c_n^2), 1 - l_n)$ at constant a_n , and u_n , and find

$$\left. \frac{dc_n^2}{dl_n} \right|_{\bar{a}, \bar{u}} = \frac{u_\ell}{u_v v_{c^2}} + p_e \frac{v_{c^1}}{v_{c^2}} \left. \frac{de_n}{dl_n} \right|_{\bar{a}, \bar{u}}. \quad (115)$$

Substitute the first-order condition for consumption (45) and labor supply (44) to eliminate u_ℓ and v_{c^1} , respectively

$$\left. \frac{dc_n^2}{dl_n} \right|_{\bar{a}, \bar{u}} = (1 - T'_z) n \phi(\cdot) + R_n p_e \left. \frac{de_n}{dl_n} \right|_{\bar{a}, \bar{u}} = (1 - T'_z) \left(n \phi(\cdot) + l_n n \phi'(e_n) \left. \frac{de_n}{dl_n} \right|_{\bar{a}, \bar{u}} \right), \quad (116)$$

where the second equality is found by substituting (46) to eliminate $R_n p_e$. Substitute (115) in the second line of (105) and the second equality in (116) in the first line of (105) to eliminate $\left. \frac{dc_n^2}{dl_n} \right|_{\bar{a}, \bar{u}}$ and use the first-order conditions labor supply (44) to eliminate $u_\ell = u_v v_{c^2} (1 - T'_z) n \phi(e_n)$ in the term $\theta_n u_\ell$ in (105) to arrive at:

$$\frac{T'_z(z_n)}{1 - T'_z(z_n)} = \frac{\theta_n u_{c^2} / \eta \left(1 - \frac{l_n u_{\ell\ell}}{u_\ell} + \frac{l_n u_{\ell v}}{u_v} \right)}{n f(n) (1 + \beta_n \varepsilon_{el})}, \quad \varepsilon_{el} \equiv \left. \frac{l_n}{e_n} \frac{de_n}{dl_n} \right|_{\bar{a}, \bar{u}}, \quad (117)$$

where $u_{c^2} \equiv u_v v_{c^2}$.

We proceed by deriving ε_{el} using similar steps as above. At constant levels of saving (controlled by the first-order condition for a_n), we have

$$\tilde{l}_n = \varepsilon_{lt} \tilde{t} + \varepsilon_{lR} \tilde{R}, \quad (118)$$

$$\tilde{e}_n = \varepsilon_{et} \tilde{t} + \varepsilon_{eR} \tilde{R}, \quad (119)$$

$$\tilde{a}_n = \varepsilon_{at} \tilde{t} + \varepsilon_{aR} \tilde{R} = 0. \quad (120)$$

The last equation implies $\tilde{R} = -(\varepsilon_{at}/\varepsilon_{aR}) \tilde{t}$ and substitution of this into the first two equations of the system yields

$$\varepsilon_{el} \equiv \left. \frac{\tilde{e}_n}{\tilde{l}_n} \right|_{\bar{a}, \bar{u}} = \frac{\varepsilon_{et} \varepsilon_{aR} - \varepsilon_{eR} \varepsilon_{at}}{\varepsilon_{lt} \varepsilon_{aR} - \varepsilon_{lR} \varepsilon_{at}}. \quad (121)$$

Substitution of the relevant elasticities from Appendix B and simplifying gives

$$\varepsilon_{el} = -\frac{1}{\varepsilon_n} \frac{(1 - \omega) \varepsilon_n - (1 + \varepsilon_n)(1 - \omega)(1 - \beta_n + \gamma_n)\sigma}{(1 - \omega)(1 - \beta_n + \gamma_n)\sigma + \beta_n}. \quad (122)$$

Consequently, we find

$$\frac{1}{1 + \beta_n \varepsilon_{el}} = \frac{\frac{\varepsilon_n}{\mu_n}}{(\omega + (1 - \omega)\delta_n) \frac{\varepsilon_n}{\mu_n} + \delta_n \frac{(1 + \varepsilon_n)\beta_n}{\mu_n}}, \quad \delta_n \equiv \frac{(1 - \omega)(1 - \beta_n + \gamma_n)\sigma}{\beta_n + (1 - \omega)(1 - \beta_n + \gamma_n)\sigma}, \quad (123)$$

which after substitution into (117) yields (52).

The optimal level of redistribution follows from the first-order condition for u_n (107). At constant levels of l_n and a_n , we find $\left. \frac{dc_n^2}{du_n} \right|_{\bar{l}, \bar{a}} = (u_n v_{c^2})^{-1}$. Substitution this into (107) yields

$$\left(\Psi'(\cdot) - \frac{\eta}{u_{c^2}} \right) f(n) - \theta_n \frac{l_n u_{lv}}{n u_v} = \frac{d\theta_n}{dn}. \quad (124)$$

Integration of this equation and using the transversality condition $\lim_{n \rightarrow \bar{n}} \theta_n = 0$ gives (53).