Human Capital and Optimal Positive Taxation of Capital Income*

Bas Jacobs[†] and A. Lans Bovenberg[‡◇]

April 5, 2005

Abstract

This paper analyzes optimal linear taxes on capital and labor incomes in a life-cycle model of human capital investment, financial savings, and labor supply with heterogenous 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 inputs in human capital formation is difficult; and most investments in human capital are verifiable. Numerical calculations suggest that the optimal marginal tax rate on capital income is close to the tax rate on labor income.

Keywords: human capital, labor income taxation, capital income taxation, life cycle, education subsidies.
JEL codes: H2, H5, I2, J2.

1 Introduction

The case for zero taxation of capital income is quite strong in models with infinitely lived individuals. In particular, this zero-tax result holds for quite general specifications of preferences (see Chamley, 1986; and Judd, 1985, 1999). Erosa and Gervais (2002) analyze optimal capital income taxes in life-cycle models rather than models with infinitely lived individuals.

*The authors thank Rick van der Ploeg, Søren Bo Nielsen, seminar participants at the University of Amsterdam, the London School of Economics, and members of the CES-Ifo 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' funded by the Netherlands Organization for Sciences. This paper previously circulated under the title "Optimal Capital Income Taxation with Endogenous Human Capital Formation".

[†]European University Insitute, University of Amsterdam, and Tinbergen Institute. E-mail: bas.jacobs@iue.it. [‡]University of Tilburg, CentER, Erasmus University, OCfEB, and CEPR. E-mail: a.l.bovenberg@uvt.nl. [◇]Corresponding author: CentER, University of Tilburg, P.O. Box 90153, 5000 LE Tilburg, The Netherlands. Phone: (+31) - 13 - 466 2912. Fax: (+31) - 13 - 466 3066.

They show that optimal capital taxes are not necessarily zero if utility is not additively separable in consumption and leisure, or, if utility does not feature weak separability between leisure and homothetic sub-utility over consumption. Building upon Ramsey-Corlett-Hague intuition, Erosa and Gervais (2002) demonstrate that optimal capital taxes are positive if leisure and consumption are more complementary later in life than they are earlier in life. See also Bernheim (2002) for a more elaborate discussion.[1]

Another strand of literature stresses the role of incomplete markets rather than life-cycle issues. Aiyagari (1995) and Golosov, Kocherlakota and Tsyvinski (2004) derive optimal taxes on capital incomes in infinitely lived agent models in the presence of incomplete capital or insurance markets. Both papers find that capital income is optimally taxed at positive rates. Intuitively, positive capital income taxes 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.

We establish another case for positive capital taxes by allowing for endogenous human capital formation in a life-cycle model.[2] In contrast to the papers mentioned above, our case for positive capital income taxes relies neither on non-separable preferences nor market failures.[3] Indeed, in our setting with weakly separable preferences and perfect financial markets, optimal capital taxes would be zero in the absence of human capital formation. To develop our case, we formulate a simple three period life-cycle model of labor supply, human capital and financial savings 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 and consume leisure. They retire and consume all their assets in the last, third period.

We demonstrate that positive capital taxes are optimal to alleviate 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.

Optimal taxes on capital are positive because the government cannot employ subsidies or tax deductions for human capital investments to offset all the distortions of the labor

[1] In life-cycle models with heterogenous households, Ordover and Phelps (1979) and, Atkinson and Sandmo (1980) showed earlier that the optimal capital tax is zero if leisure is weakly separable from consumption.

[2] The literature on the optimal taxation of capital income typically abstracts from human capital accumulation. See Boskin (1975) and Heckman (1976) for earlier treatments of the effects of capital income taxes on human capital formation. These treatments are not put in an optimal tax setting.

[3] All 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.

income tax on human capital formation. Non-verifiability of human capital investments excludes these more direct instruments. Non-verifiability of learning is analogous to the non-verifiable nature of work effort in the optimal tax literature.[4] Direct educational expenditures on books, computers and traveling are important examples of non-verifiable investments in human capital. Moreover, costs of effort while enrolled in education, such as studying hard, sacrificing leisure activities, preparing exams, are important immaterial costs that the government cannot verify easily.[5] In contrast, the indirect costs of education, forgone labor earnings while enrolled in education, are in effect deductible from labor taxation as lower labor earnings reduce the labor tax bill. Tuition costs, however, cannot be deducted for income-tax purposes in many countries, so that these costs are effectively non-verifiable.

We demonstrate that the government optimally taxes capital if at least some investments in human capital are non-verifiable. Only if it can verify all investments in education (so that it can subsidize these costs), the government does not have to rely on capital income taxes to alleviate the labor tax distortions on learning. Distortions in the aggregate level of saving can thus be avoided. This rather special 'knife-edge' case explains the findings of some recent papers that the optimal capital tax is zero in the presence of human capital accumulation (see Judd 1999; and Bovenberg and Jacobs, 2001).[6] Trostel (1993) estimates that the share of non-verifiable costs of education approaches as much as one quarter of total costs, even though he does not include the effort costs of education. Our numerical simulations show that optimal tax rate on capital income approaches that on labor incomes, even if non-verifiable investments account for only 20% of total investments in human capital.

The case for substantially positive capital income taxes does not directly depend on the redistributive preferences of the government. It relies only on the presence of positive labor income taxes. In our model, distortionary labor taxation increases with a stronger distributional preference of the government. Hence, in contrast to representative agent models (see, e.g. Atkinson and Sandmo, 1980; Nielsen and Sørensen, 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.

The rest of this paper is organized as follows. Section 2 describes individual behavior. Section 3 derives optimal tax policy if all educational efforts are non-verifiable. Subsequently, section 4 introduces verifiable educational efforts, which the government can subsidize. Section 5 performs some numerical simulations and demonstrates that a synthetic income tax is roughly optimal under a wide range of parameter values. Section 6 concludes and discusses the policy implications of the analysis. Four appendices contain the technical details of our analysis.

[4]The literature on optimal labor taxation follows Mirrlees' (1971) pioneering analysis by assuming that the government cannot verify work effort and individual ability, thereby excluding individualized lump-sum taxes.

[5]Education may also generate immaterial benefits, such as the fun of studying, nicer jobs, additional status, more freedom of occupational choice, et cetera. Immaterial 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 probably also compensate investors for non-pecuniary costs that exceed non-pecuniary benefits (see also Judd, 2000; and, Palacios-Huerta, 2003).

[6]Nielsen and Sørensen (1997) assume an exogenously given positive capital tax. If the government would optimize the capital tax, it would set this tax equal to zero.

2 The model

2.1 Preferences and technologies

We consider a partial equilibrium three-period life-cycle model without uncertainty. Before-tax wage rates and interest rates are exogenously fixed.[7] A mass of agents with unit measure lives for three periods. 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. Individuals are retired in the third period and consume all their assets. Perfect capital markets allow individuals to freely transfer resources across the three periods.[8]

Individuals are heterogeneous in exogenous ability n. The cumulative distribution of ability is $F(n)$. $f(n)$ represents the corresponding density function, which is continuously differentiable and strictly positive on the support $[\underline{n}, \overline{n}]$, $\underline{n}, \overline{n} > 0$. 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 e^n in education. Initially, we assume that the government cannot observe any of these educational inputs so that it cannot subsidize these investments. Educational investment therefore consists only of direct expenditures and effort costs. Section 4 shows that our main results continue to hold if verifiable costs of education are allowed for as long as some non-verifiable costs remain.[9] For analytical convenience, we abstract from first-period consumption and leisure demands.[10]

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 is homothetic:

$$h^n = n\phi(e^n) \equiv n\,(e^n)^{\beta}\,, \tag{1}$$

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 second-period 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., $z^n \equiv h^n l^n = n\phi(e^n)l^n$.[11] In the third period,

[7] The model can thus be viewed as a model of a general equilibrium small open economy in which the international capital market fixes the real interest rate.

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

[9] Bovenberg and Jacobs (2005) explore optimal labor tax and education policies in the presence of both verifiable and non-verifiable inputs into human capital formation. They abstract, however, from capital income taxes.

[10] First-period consumption and leisure do not affect our main result that optimal capital taxes are positive. The presence of first-period consumption, however, would exacerbate the welfare costs of positive capital income taxes. Endogenous first-period leisure demands would increase the distortionary costs of the labor tax. Hence, the effect of first-period consumption and leisure on the optimal structure of labor and capital taxes is ambiguous.

[11] With ability defined as the productivity of learning, the wage rate per efficiency unit of labor is normalized

individuals are retired and finance their consumption by selling the financial assets they have accumulated in the second period.[12]

Individuals feature a common, concave, and twice differentiable utility function defined over consumption in the second period c_1^n , leisure $\ell^n \equiv 1 - l^n$, and consumption when retired c_2^n:

$$u(v(c_1^n, c_2^n), \ell^n), \tag{2}$$

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 $v(c_1^n, c_2^n)$ is homothetic and weakly separable from leisure ℓ^n. We employ this particular utility function because it most clearly shows how endogenous human capital formation affects the optimal capital income tax. Indeed, with this utility function, the optimal capital income tax would be zero in the absence of human capital formation (see Bernheim, 2002; and, Erosa and Gervais, 2002).

2.2 Budget constraints

The first-period budget constraint is

$$-a_1^n = e^n. \tag{3}$$

First-period saving a_1^n is negative because individuals borrow to finance their education. In the second period, individuals consume, work, pay off their first-period debts and save for retirement. Hence, the second-period budget constraint amounts to

$$c_1^n = (1-t) l^n n \phi(e^n) + R a_1^n - a_2^n + g, \tag{4}$$

where $R \equiv 1 + (1-\tau) r$, 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.

The individual consumes all assets in the third period:

$$c_2^n = R a_2^n. \tag{5}$$

We arrive at the life-time budget constraint by substituting the first-period and third-period budget constraints (i.e. (3) and (5), respectively) into the second-period budget constraint (4) to eliminate a_1^n and a_2^n:

$$c_1^n + \frac{c_2^n}{R} = (1-t) l^n n \phi(e^n) - R e^n + g. \tag{6}$$

at unity. Alternatively, ability n can be interpreted as the skill in second-period work corresponding to the wage rate per efficiency unit of labor. In that case, $\phi(e^n)$ (instead of $n\phi(e^n)$) defines the efficiency units of labor (or human capital) per hour worked. Dur and Teulings (2001) investigate optimal education policies in a general equilibrium model in which different labor types are imperfect substitutes in labor demand.

[12] We include a third period to allow high ability agents to save more than low ability agents do. In the absence of a third period, high ability agents would feature the lowest life-time savings, because they borrow more than low ability agents to finance their education in the first period. Our main results, however, do not rely on any particular relationship of financial savings with ability.

2.3 Individual optimization

Individuals maximize their utilities by choosing c_1^n, c_2^n, ℓ^n, and e^n, subject to their life-time budget constraints, and taking the policy instruments of the government as given.[13] The resulting first-order condition for the optimal choice of e^n amounts to

$$(1-t)l^n n\phi'(e^n) = 1 + (1-\tau)r = R. \tag{7}$$

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 (7) and the production function of human capital (1) imply that second-period gross labor income z^n is proportional to e^n:

$$z^n = l^n n\phi(e^n) = \frac{R}{(1-t)\beta} e^n, \tag{8}$$

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

The first-order condition for labor supply amounts to

$$\frac{u_\ell}{u_{c_1}} = w^n \equiv (1-t)n\phi(e^n), \tag{9}$$

while the Euler equation for savings is

$$\frac{u_{c_1}}{u_{c_2}} = 1 + (1-\tau)r = R. \tag{10}$$

Substituting the link between e^n and z^n (8) into the life-time budget constraint (6), we find that the discounted value of life-time consumption $c_1^n + \frac{c_2^n}{R}$ is linear in gross income (and in view of (8) also linear in learning):

$$c_1^n + \frac{c_2^n}{R} = (1-\beta)(1-t)z^n + g. \tag{11}$$

Weak separability between leisure and consumption in utility and homotheticity of $v(c_1^n, c_2^n)$ imply that the shares of third- and second-period consumption in after-tax labor income do not depend on ability n, i.e.,

$$\omega \equiv \frac{c_2^n/R}{(1-\beta)(1-t)l^n n\phi(e^n) + g}, \tag{12}$$

[13] The government sets policy before agents determine their behavior. In view of its distributional preferences, the government faces an incentive to change its policy after the private sector has accumulated human and financial capital. The government thus must have access to a commitment technology (e.g., due to reputational considerations).

$$1-\omega \equiv \frac{c_1^n}{(1-\beta)(1-t)l^n n\phi(e^n)+g}. \tag{13}$$

The second-order conditions for a utility maximization imply (see Appendix A)

$$\mu^n \equiv 1-\beta(1+\epsilon^n) > 0, \tag{14}$$

where $\epsilon^n \equiv \frac{\partial l^n}{\partial w^n}\frac{w^n}{l^n}$, denotes the compensated wage elasticity of labor supply with respect to the after-tax wage $w^n \equiv (1-t)n\phi(e^n)$.[14] Decreasing returns in human capital accumulation ($\beta < 1$) are not sufficient for the second-order condition to be met because of the positive feedback between human capital and labor supply. 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, which rises with the compensated wage elasticity of labor supply ϵ^n.

2.4 Government

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 $\int_{\underline{n}}^{\overline{n}} l^n n\phi(e^n)dF(n)$ and aggregate capital incomes $\int_{\underline{n}}^{\overline{n}} r\left(a_1^n + \frac{a_2^n}{1+r}\right)dF(n)$.

The government budget constraint reads as

$$\int_{\underline{n}}^{\overline{n}} \left[tl^n n\phi(e^n) + \tau r\left(a_1^n + \frac{a_2^n}{1+r}\right) - g - \Lambda\right] dF(n) = 0. \tag{15}$$

The government budget constraint can be written in terms of t and R by using the first and third-period households budget constraints (3) and (5) to eliminate a_1^n and a_2^n:

$$\int_{\underline{n}}^{\overline{n}} \left[tl^n n\phi(e^n) + Re^n + \frac{c_2^n}{R} - (1+r)e^n - \frac{c_2^n}{1+r} - g - \Lambda\right] dF(n) = 0. \tag{16}$$

The government's budget is fully funded and the government can freely borrow and lend at the capital market at rate r.[15]

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

$$\Gamma \equiv \int_{\underline{n}}^{\overline{n}} \Psi(\upsilon(g,t,R,n))dF(n), \tag{17}$$

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

[14] The compensated tax elasticity of labor supply (ϵ^n/μ^n) exceeds the compensated wage elasticity of labor supply ϵ^n. The reason is that the tax rate depresses the after-tax wage rate not only directly by raising the tax wedge, but also indirectly by reducing human capital investments and thus the before-tax reward for each additional unit of labor effort l^n. We can express the compensated elasticities of the endogenous variables with respect to the policy parameters in terms of ϵ^n by totally differentiating the first-order conditions (7), (9), and (10) (see Appendix B for the derivations).

[15] As in Nielsen and Sørensen (1997), one can interpret our model as the steady state of an overlapping generation's economy. In that case, our 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.

3 Optimal taxation

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

$$\max_{\{g,t,R\}} \mathcal{L} = \int_{\underline{n}}^{\overline{n}} \Psi\left(\upsilon(g,t,R,n)\right) dF(n) + \eta \int_{\underline{n}}^{\overline{n}} \left[tnl^n\phi(e^n) + Re^n + \frac{c_2^n}{R} - (1+r)e^n - \frac{c_2^n}{1+r} - g - \Lambda \right] dF(n), \tag{18}$$

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

We apply Roy's lemma (using (6)) to derive the following properties:

$$\frac{\partial \upsilon(g,t,R,n)}{\partial g} = \lambda^n, \tag{19}$$

$$\frac{\partial \upsilon(g,t,R,n)}{\partial t} = -\lambda^n l^n n\phi(e^n), \tag{20}$$

$$\frac{\partial \upsilon(g,t,R,n)}{\partial R} = \lambda^n \left(-e^n + \frac{c_2^n}{R^2} \right), \tag{21}$$

where λ^n denotes private marginal utility of second-period income for an individual with ability n.

3.1 Optimal lump-sum transfer

The first-order condition for maximizing the Lagrangian (18) with respect to the lump-sum transfer g amounts to

$$\frac{\partial \mathcal{L}}{\partial g} = \int_{\underline{n}}^{\overline{n}} \left[\Psi'\lambda^n - \eta + \eta\Delta\frac{\partial e^n}{\partial g} + \eta tn\phi(e^n)\frac{\partial l^n}{\partial g} + \eta\frac{\tau r}{(1+r)R}\frac{\partial c_2^n}{\partial g} \right] dF(n) = 0, \tag{22}$$

where we used Roy's lemma (19). $\Delta \equiv tnl^n\phi'(e^n) - \tau r = \left(\frac{t}{1-t}\right) R - \tau r$ represents the total tax wedge on investments in human capital (where the second equality is derived by using the first-order condition for learning (7)).

By defining the net social marginal value of income of an individual with ability n (including the effect on the tax base)

$$b^n \equiv \frac{\Psi'\lambda^n}{\eta} + \Delta\frac{\partial e^n}{\partial g} + tn\phi(e^n)\frac{\partial l^n}{\partial g} + \frac{\tau r}{(1+r)R}\frac{\partial c_2^n}{\partial g}, \tag{23}$$

we can write first-order condition (22) as

$$\bar{b} \equiv \int_{\underline{n}}^{\overline{n}} b^n dF(n) = 1. \tag{24}$$

The average social marginal benefits of a higher g (i.e., the left-hand side of (24)) should equal the costs in terms of a higher g (i.e., last the right-hand side of (24)).

In order to facilitate the discussion of the optimal tax schedules below, we define the so-called distributional characteristic ξ of labor income

$$\xi \equiv \frac{\int_{\underline{n}}^{\overline{n}}(1-b^n)z^n dF(n)}{\int_{\underline{n}}^{\overline{n}} z^n dF(n) \int_{\underline{n}}^{\overline{n}} b^n dF(n)}, \tag{25}$$

as the negative normalized covariance between the social value the government attaches to income of a particular ability b^n and gross second-period labor income z^n (see also Atkinson and Stiglitz (1980)). This covariance coincides with the normalized covariances of the welfare weights b^n with learning e^n because labor income z^n is proportional to education e^n (see (8)). A positive distributional characteristic ξ implies that 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).

3.2 Optimal labor income tax

The first-order condition for maximizing the Lagrangian for social welfare (18) with respect to t is given by

$$\begin{aligned} \frac{\partial \mathcal{L}}{\partial t} &= \int_{\underline{n}}^{\overline{n}} \left[-\Psi'\lambda^n n l^n \phi(e^n) + \eta n l^n \phi(e^n)\right] dF(n) + \\ & \int_{\underline{n}}^{\overline{n}} \eta \left[\Delta \frac{\partial e^n}{\partial t} + tn\phi(e^n)\frac{\partial l^n}{\partial t} + \frac{\tau r}{(1+r)R}\frac{\partial c_2^n}{\partial t}\right] dF(n) = 0, \end{aligned} \tag{26}$$

where we used Roy's lemma (20). We substitute the Slutsky equations, and use the definition of b^n (see (23)) to obtain (see Appendix C)

$$\begin{aligned} \xi &= -\frac{\Delta}{R}\beta\bar{\varepsilon}_{et} + \left(\frac{t}{1-t} + \frac{\tau r}{(1+r)}\frac{\omega}{\kappa}\right)\bar{\varepsilon}_{lt} \\ &= \frac{\Delta\beta}{R}\left(\overline{\frac{1+\epsilon}{\mu}}\right) + \left(\frac{t}{1-t} + \frac{\tau r}{(1+r)}\frac{\omega}{\kappa}\right)\left(\overline{\frac{\epsilon}{\mu}}\right), \end{aligned} \tag{27}$$

where κ is the degree of homogeneity of the sub-utility function $v(c_1^n, c_2^n)$, and $\varepsilon_{qt}^n \equiv \frac{\partial q^n}{\partial t}\frac{1-t}{q}$ (for $q^n = e^n$, l^n, c_2^n) are the compensated elasticities with respect to the labor tax. A bar denotes an income-weighted average of a skill-specific variable (i.e., $\bar{\varepsilon} \equiv \int_{\underline{n}}^{\overline{n}} \varepsilon^n z^n dF(n) / \int_{\underline{n}}^{\overline{n}} z^n dF(n)$, where ε^n is a variable that depends on ability).

The interpretation of the optimal labor tax (27) is as follows. The distributional benefits of a higher tax rate (ξ) (i.e. the left-hand side of (27)) should correspond to the additional first-order welfare losses as a result of the higher tax rate (i.e. the right-hand side of (27)).

These welfare losses arise because of first-order impacts on the learning distortion Δ and the labor-supply distortion $\frac{t}{1-t} + \frac{\tau r}{(1+r)}\frac{\omega}{\kappa}$. The capital tax τ features in the labor-supply distortion because the capital tax acts as an implicit tax on labor income by taxing retirement consumption as one of the uses of second-period labor income. In this way, a capital tax induces cross-substitution away from second-period consumption towards leisure.

In the absence of capital income taxes (i.e., $\tau = 0$ so that $\Delta = \frac{t}{1-t}R$), the optimal marginal tax rate on labor income is

$$\frac{t}{1-t} = \frac{\xi}{\left(\frac{\overline{\epsilon}}{\mu}\right) + \beta\left(\frac{\overline{1+\epsilon}}{\mu}\right)}. \tag{28}$$

This expression illustrates the fundamental trade-off between equity and efficiency. If redistributional 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 denominator of (28) represents the distortionary costs of redistributive taxation in terms of the tax elasticity of total labor income. The first income-weighted elasticity $\left(\frac{\overline{\epsilon}}{\mu}\right) = -\bar{\varepsilon}_{lt}$ captures the distortionary effect of the marginal labor tax rate on labor supply and thus the base of the labor tax. The second term in the denominator $\beta\left(\frac{\overline{1+\epsilon}}{\mu}\right) = -\beta\bar{\varepsilon}_{et}$ stands for the distortion of the labor tax on human-capital accumulation. 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).

Expression (28) indicates that endogenous human capital formation ($\beta > 0$) raises the effective elasticity of the tax base in two distinct ways.[16] First, due to positive feedback effects with labor supply, endogenous learning raises the absolute value of the effective elasticity of labor supply ($-\bar{\varepsilon}_{lt} \equiv \frac{\overline{\epsilon}}{\mu}$) by lowering $\mu^n \equiv 1 - \beta(1+\epsilon^n)$. Second, labor taxes distort not only labor supply but also investments in human capital. The term $\beta\left(\frac{\overline{1+\epsilon}}{\mu}\right)$ captures this latter distortionary effect.

3.3 Optimal capital income tax

Maximization of social welfare (18) with respect to the capital income tax yields the following first-order condition:

$$\begin{aligned}\frac{\partial \mathcal{L}}{\partial R} &= \int_{\underline{n}}^{\overline{n}} \left[\Psi'\lambda^n\left(\frac{c_2^n}{R^2} - e^n\right) - \eta\left(\frac{c_2^n}{R^2} - e^n\right)\right] dF(n) + \\ &\int_{\underline{n}}^{\overline{n}} \eta\left[\Delta\frac{\partial e^n}{\partial R} + tn\phi(e^n)\frac{\partial l^n}{\partial R} + \frac{\tau r}{(1+r)R}\frac{\partial c_2^n}{\partial R}\right] dF(n) = 0,\end{aligned} \tag{29}$$

[16] With exogenous learning (i.e., $\beta = 0$), the model collapses to the standard model of optimal linear income taxation with endogenous labor supply in which the optimal marginal tax t is given by $\frac{t}{1-t} = \frac{\xi}{\overline{\epsilon}}$, see Atkinson and Stiglitz (1980).

where we used Roy's lemma (see (21)). Substitution of the Slutsky equations and the definition of b_n (23) yields (see Appendix C)

$$\begin{aligned} & (\omega(1-\beta)-\beta)\xi + \frac{\Delta\beta}{R}\left(\frac{\overline{1-\omega\epsilon}}{\mu}\right) \\ = & \left(\frac{t}{1-t} + \frac{\tau r}{(1+r)}\frac{\omega}{\kappa}\right)(\omega(1-\beta)-\beta)\left(\frac{\overline{\epsilon}}{\mu}\right) + \frac{\tau r}{(1+r)}\left(\omega(1-\omega)\sigma\bar{\gamma}\right), \end{aligned} \tag{30}$$

where $\sigma \equiv d\log\left(c_2^n/c_1^n\right)/d\log(u_{c_1}/u_{c_2})$ denotes the intertemporal elasticity of substitution in consumption and $\gamma^n \equiv 1-\beta+\frac{g}{(1-t)z^n}$. The factor $\omega(1-\beta)-\beta$ is the marginal propensity to save out of life-time labor income. In particular, a large share of retirement consumption ω implies that a considerable part of second-period labor income is saved for retirement, while a small elasticity of learning (β) reduces borrowing in the first period to finance education. If $\omega(1-\beta) > \beta$, the marginal saving quote is positive and individuals with higher ability thus save most because they feature the highest labor income. Hence, by reducing the return on saving, the capital income tax yields positive distributional benefits as captured by the first term at the left-hand side of (30). The second term at the left-hand side stands for the welfare effect of a higher capital income tax on the tax wedge on human capital Δ (note that $\beta\left(\frac{\overline{1-\omega\epsilon}}{\mu}\right) = -\beta\bar{\varepsilon}_{eR} > 0$ (see Appendix B) because a higher capital tax stimulates learning). The first term at the right-hand side of (30) represents the welfare effects of capital income taxation in terms of exacerbating the labor-supply distortion $\frac{t}{1-t}+\frac{\tau r}{(1+r)}\frac{\omega}{\kappa}$, where the compensated labor supply elasticity with respect to the after-tax return on financial savings is $(\omega(1-\beta)-\beta)\left(\frac{\overline{\epsilon}}{\mu}\right) = \bar{\varepsilon}_{lR}$, see Appendix B). The sign of this compensated elasticity is ambiguous. While a higher capital tax harms labor supply by taxing the use of labor income for retirement consumption, it boosts labor supply by stimulating human capital accumulation. The second term at right-hand side of (30) stands for the welfare losses of a higher capital tax in terms of exacerbating distortions in the intertemporal allocation of consumption (given a positive tax wedge $\frac{\tau r}{(1+r)} > 0$ and a positive intertemporal substitution elasticity $\sigma > 0$).

In the absence of a labor income tax (i.e., $t = 0$ so that $\Delta = \tau r$), the optimal capital income tax can be written as

$$\frac{\tau r}{(1+r)} = \frac{(\omega(1-\beta)-\beta)\xi}{\frac{\omega}{\kappa}(\omega(1-\beta)-\beta)\left(\frac{\overline{\epsilon}}{\mu}\right) + \frac{\beta(1+r)}{R}\left(\frac{\overline{1-\omega\epsilon}}{\mu}\right) + \omega(1-\omega)\bar{\gamma}\sigma}. \tag{31}$$

Capital must thus be taxed for redistributive reasons if the rich save more than the poor (i.e., $\omega(1-\beta) > \beta$). The optimal capital income tax rises with the distributional gains as captured by the numerator at the right-hand side of (31), and declines with the welfare costs of capital income taxation. These welfare costs consist of three components represented by the three terms in the denominator at the right-hand side of (31). These terms represent the effects of capital income taxation on, respectively, labor supply, human capital accumulation, and the intertemporal allocation of consumption. As regards the labor-supply effect, if the rich save more (i.e., $\omega(1-\beta) > \beta$), a higher capital tax depresses labor supply because the additional tax burden on third-period consumption imposed by the capital tax (represented

by the term $\omega(1-\beta))$ dominates lower educational costs as a result of the capital tax increasing the net present value of investments in human capital (represented by the term β). This labor supply effect is substantial if the labor supply elasticity $\left(\frac{\bar{\epsilon}}{\mu}\right)$ is large. As far as learning is concerned, if the learning elasticity $\beta\left(\frac{\overline{1-\omega\epsilon}}{\mu}\right) = -\beta\bar{\varepsilon}_{eR}$ is large, capital income taxes induce substantial excessive learning by reducing the opportunity costs of education. Finally, concerning the third component of the welfare costs of capital income taxation, a large intertemporal substitution elasticity σ raises the intertemporal distortions of capital taxes on the intertemporal allocation of consumption.

3.4 Optimal dual income tax

If the government can optimally set both the labor income tax and the capital tax, we can combine the two first-order conditions (27) and (30) to obtain (see Appendix C)

$$\frac{\Delta\beta}{R} = \frac{\tau r}{(1+r)}\omega^*\bar{\gamma}\sigma, \tag{32}$$

where $\omega^* \equiv \frac{\omega(1-\omega)}{1+\omega} > 0$. This expression clearly shows the role of the optimal capital income tax in alleviating learning distortions imposed 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 positive labor taxes distort learning (i.e., if $t > 0$ and thus $\Delta > 0$). In that case, 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. At the same time, however, a capital income tax distorts the intertemporal allocation of consumption. At small capital income taxes, however, these welfare costs associated with distorted saving behavior are only second order, while the welfare benefits of alleviating the learning distortion are first order. Hence, the introduction of a small capital income tax enhances welfare. At the optimal capital tax, the additional welfare benefits of a higher capital income tax in terms of stimulating distorted learning (i.e. the left-hand side of (32)) balance the additional welfare costs in terms of a more distorted intertemporal allocation of consumption (i.e. the right-hand side of (32)). The optimal capital tax thus trades off production efficiency in the composition of saving (i.e., a level playing field between financial and human capital) against allocative efficiency in the level of saving.

Using $\Delta = \frac{t}{1-t}R - \tau r$, we find the optimal structure of capital and labor taxes from (32):

$$\frac{\tau r}{(1+r)} = \frac{t}{1-t}\left(\frac{(1+r)}{R} + \frac{\omega^*\bar{\gamma}\sigma}{\beta}\right)^{-1}. \tag{33}$$

The capital tax is thus relatively low compared to the labor tax if saving is elastic (large σ) compared to learning (small β). In that case, labor taxes impose only small distortions on learning while capital taxes substantially distort saving.

Substituting the learning distortion (32) to eliminate Δ from the first-order condition for t (27) and subsequently substituting the optimal tax structure (33) into the result to

eliminate $\frac{t}{1-t}$, we derive the optimal capital income tax

$$\frac{\tau r}{1+r} = \frac{\xi}{\left(\frac{\overline{\epsilon}}{\mu}\right)\left(\frac{1+r}{R} + \frac{\omega^*\bar{\gamma}\sigma}{\beta} + \frac{\omega}{\kappa}\right) + \beta\left(\frac{\overline{1+\epsilon}}{\mu}\right)\frac{\omega^*\bar{\gamma}\sigma}{\beta}}. \tag{34}$$

Capital income is thus taxed at positive rates even if the rich do not save more than the poor. In the absence of a labor income tax, the factor $\omega(1-\beta)-\beta$, which measures the correlation between skill and saving, determines the sign of the optimal capital income tax (see (31)). However, this factor does not enter the expression for the optimal capital income tax if the government can also optimally set the labor income tax (see (34)). In the latter case, the distributional characteristic of labor rather than capital incomes determines the optimal capital tax. Thus, even if the capital tax is regressive (i.e. $\omega(1-\beta)-\beta<0$), the government relies more heavily on the capital tax if distributional considerations become more pressing as indicated by a larger distributional characteristic ξ.

The reason for this seemingly paradoxical result is that the capital tax is not an efficient instrument for redistributing incomes if the government can also employ a labor tax. With weakly separable and homothetic preferences, saving is proportional to labor income so that a tax on saving in fact acts as a tax on additional labor earnings. A capital tax thus does not redistribute more effectively than the labor tax, while it causes an additional distortion, namely on the intertemporal allocation of consumption. Furthermore, since consumption and leisure are (weakly) separable, the capital tax does not reduce overall labor supply distortions by serving as an implicit tax on leisure. Hence, the labor income tax is aimed at redistributing income, while the capital tax is targeted solely at alleviating the distortions of the labor income tax on human capital formation.

Whereas by itself a capital tax may be regressive, together with the labor tax it enhances redistribution by alleviating the learning distortions of the labor tax. The intuition is that the labor tax, combined with the capital tax, becomes a more effective instrument to tax the rents from ability. The combination of capital taxes and labor taxes therefore results in larger redistribution of incomes and contributes to more equality, even if the capital tax is regressive. Indeed, the direct regressive impact of the capital tax is more than offset by a more progressive labor tax.

Substitution of (34) into (33) to eliminate τ yields the following expression for the optimal labor tax

$$\frac{t}{1-t} = \frac{\xi}{\left(\frac{\overline{\epsilon}}{\mu}\right)\left(1 + \frac{\omega/\kappa}{\frac{1+r}{R} + \frac{\omega^*\bar{\gamma}\sigma}{\beta}}\right) + \beta\left(\frac{\overline{1+\epsilon}}{\mu}\right)\left(\frac{\frac{\omega^*\bar{\gamma}\sigma}{\beta}}{\frac{(1+r)}{R} + \frac{\omega^*\bar{\gamma}\sigma}{\beta}}\right)}. \tag{35}$$

Both labor and capital income are thus optimally taxed at positive marginal tax rates, which rise with the distributional characteristic ξ.

The overall tax wedge on labor supply $\frac{t}{1-t} + \frac{\tau r}{(1+r)}\frac{\omega}{\kappa}$ amounts to (use (34) and (35))

$$\frac{t}{1-t} + \frac{\tau r}{(1+r)}\frac{\omega}{\kappa} = \frac{\xi}{\left(\frac{\overline{\epsilon}}{\mu}\right) + \beta\chi\left(\frac{\overline{1+\epsilon}}{\mu}\right)}, \tag{36}$$

where $0 < \chi \equiv \frac{\frac{\omega^*\bar{\gamma}\sigma}{\beta}}{\frac{\omega^*\bar{\gamma}\sigma}{\beta} + \left(\frac{(1+r)}{R} + \frac{\omega}{\kappa}\right)} < 1$. Comparison of (36) with the overall tax distortion on labor supply in the absence of a capital income tax (28) reveals how the availability of the

capital tax raises the optimal overall labor tax $\frac{t}{1-t} + \frac{\tau r}{(1+r)} \frac{\omega}{\kappa}$ by alleviating the associated distortionary costs on learning through the factor χ. In fact, without intertemporal distortions ($\sigma = 0$), the capital tax completely eliminates the learning distortions (i.e., $\chi = 0$). With large intertemporal distortions compared to learning distortions, in contrast, the capital tax does not help much in reducing distortions in human capital formation. Indeed, the efficacy of the capital tax in alleviating learning distortions depends on the intertemporal substitution elasticity σ, which determines the sensitivity of the *level* of saving, compared to the learning elasticity β, which determines the sensitivity of the *composition* of saving.

The optimal overall tax wedge (36) is lower if behavior becomes more sensitive to taxes. In particular, *ceteris paribus* the distributional characteristic, optimal taxes decline with the wage elasticity of labor supply ϵ^n, the learning elasticity β (both reduce μ^n) and the intertemporal substitution elasticity σ. By raising the labor-supply distortions associated with the labor tax, a higher wage elasticity of labor supply ϵ^n reduces the optimal labor tax and therefore the need for capital taxes to alleviate the distortions of the labor tax on learning. By boosting the costs of employing the capital tax to alleviate the learning distortions imposed by the labor tax, a higher intertemporal substitution elasticity σ reduces the optimal labor supply wedge $\frac{t}{1-t} + \frac{\tau r}{(1+r)} \frac{\omega}{\kappa}$, as the optimal capital tax can correct for only a small part of the labor tax distortions on learning. Redistribution thus becomes more costly and the optimal tax wedge declines, *ceteris paribus*.

Whereas all behavioral margins reduce the overall tax wedge (36), only the learning and saving margin affect the composition of the tax burden (33). The wage elasticities of labor supply ϵ^n do not impact the composition of the tax burden over labor and capital taxes, since shifting the tax burden from labor to capital income taxes will not reduce the distortions in leisure demand (due to (weakly) separable preferences and a constant marginal saving quote).

The impact of the behavioral margins can be illustrated with three special cases. In each of these cases, one of the three behavioral margins (labor supply, learning, or the intertemporal allocation of consumption) does not operate. 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

$$\frac{t}{1-t} = \frac{\xi}{\bar{\epsilon}}. \tag{37}$$

In the absence of intertemporal substitution in consumption (i.e., $\sigma = 0$), the capital income tax can costlessly accomplish production efficiency in learning (i.e., $\Delta = 0$ and $\frac{t}{1-t} = \frac{\tau r}{R}$). The optimal overall tax wedge is given by

$$\frac{t}{1-t} + \frac{\tau r}{(1+r)} \frac{\omega}{\kappa} = \frac{\xi}{\left(\frac{\bar{\epsilon}}{\mu}\right)}. \tag{38}$$

In this case, the capital income tax allows the government to tax the infra-marginal rents of learning at zero costs, even though learning efforts are non-verifiable. Tax rates, however, 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 $\left(\frac{\epsilon}{\mu}\right)$, because the learning elasticity $\beta\left(\frac{1+\epsilon}{\mu}\right)$ drops out of the denominator of (38) (compare (38) with (36)).

Exogenous labor supply ($\epsilon^n = 0$) does not directly affect the optimal tax structure (33) but only raises overall tax levels. The case for taxing capital income thus depends on endogenous learning ($\beta > 0$) rather than endogenous labor supply. The overall tax level

$$\frac{t}{1-t} + \frac{\tau r}{(1+r)}\frac{\omega}{\kappa} = \frac{\xi(1-\beta)}{\beta\chi}, \tag{39}$$

remains finite because the labor tax causes learning distortions, which can be offset by the capital tax only at the cost of distortions in savings (with $\sigma > 0$ and thus $\chi > 0$).

4 Verifiable investment in human capital

This section explores how verifiable educational efforts, 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 forgone earnings are tax deductible and enrollment in (higher) education is widely subsidized across the Western world. In the presence of subsidies on education, the government has access to a direct instrument to offset the learning distortions caused by the labor tax. This raises the question whether the government still wants to rely on the indirect instrument of the capital income tax to alleviate the labor-tax distortions on human capital.

To anwer 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 y^n as direct costs and monetized effort costs. The verifiable input x^n is 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:

$$e^n \equiv \psi(x^n, y^n), \tag{40}$$

where $\psi_x, \psi_y > 0$; $\psi_{xx}, \psi_{yy} \leq 0$ and $\psi_{xy} \geq 0$. Hence, the production function for human capital (1) becomes

$$h^n = n\phi(e^n) \equiv n\left(\psi(x^n, y^n)\right)^{\beta}. \tag{41}$$

The household budget constraint is now given by

$$c_1^n + \frac{c_2^n}{R} = (1-t)l^n n\phi(\psi(x^n, y^n)) - R[(1-t)(1-s)p_x x^n + p_y y^n] + g, \tag{42}$$

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.[17]

[17] If we interpret $(1-t)p_x x^n$ as the taxed foregone earnings of an unskilled worker, we may add a (net) endowment of unskilled labor, $(1-t)p_x$, to the household budget constraint – where we normalized the endowment of unskilled labor time to unity. x^n then denotes the fraction of unskilled labor time invested

The first-order conditions for maximizing utility with respect to the two educational inputs x^n and y^n amount to

$$l^n n\phi'(.)\psi_x(x^n, y^n) = R(1-s)p_x, \tag{43}$$

$$(1-t)l^n n\phi'(.)\psi_y(x^n, y^n) = Rp_y. \tag{44}$$

The tax rate t does not enter (43) because tax deductibility of x^n implies that the tax rate equally reduces the marginal benefits and the marginal costs of verifiable learning x^n. The subsidy s boosts investments of verifiable inputs x^n. In the case of non-verifiable learning, in contrast, the tax rate directly reduces only the benefits and leaves the costs unaffected (see (44)).

The first-order conditions for x^n (43) and y^n (44), and the production function of human capital (41) imply that gross labor income z^n is proportional to both x^n and y^n:

$$z^n = l^n n\phi(e^n) = \frac{(1-s)p_x}{\alpha\beta}x^n, \tag{45}$$

$$z^n = l^n n\phi(e^n) = \frac{p_y}{(1-\alpha)\beta(1-t)}y^n. \tag{46}$$

The proportionality factors do not depend on ability n, because the shares of x^n and y^n in human capital investment, $\alpha \equiv \frac{x^n\psi_x}{\psi} = \frac{(1-t)(1-s)p_x x^n}{(1-t)(1-s)p_x x^n + p_y y^n}$ and $1-\alpha \equiv \frac{y^n\psi_y}{\psi} = \frac{p_y y^n}{(1-t)(1-s)p_x x^n + p_y y^n}$, are the same for all agents (see Appendix B).

The government budget constraint is now given by

$$\begin{aligned} &\int_{\underline{n}}^{\overline{n}} \left[t\left(nl^n\phi(e^n) - (1-s)p_x x^n(1+r)\right)\right] dF(n) + \\ &\int_{\underline{n}}^{\overline{n}} \left[\tau r\left(\frac{c_2^n}{R(1+r)} - (1-s)(1-t)p_x x^n - p_y y^n\right)\right] dF(n) \\ = &\int_{\underline{n}}^{\overline{n}} \left[(1+r)sp_x x^n + g + \Lambda\right] dF(n). \end{aligned} \tag{47}$$

4.1 Optimal education subsidies

If the government simultaneously optimizes over the labor tax and the educational subsidy, the optimal educational subsidy satisfies (see Appendix D)

$$\frac{s + \frac{\tau r}{R}}{(1-s)(1-t)} = \left(\frac{(1-\alpha)(1-\rho)}{1-(1-\alpha)(1-\rho)}\right)\frac{\Delta}{R}, \tag{48}$$

where $\rho \equiv d\log\left(\frac{x^n}{y^n}\right)/d\log(\frac{\psi_y}{\psi_x})$ stands for the elasticity of substitution between x^n and y^n in the composite of aggregate investment in human capital e^n (40). This expression shows the

in education. The additional term in the budget constraint, $(1-t)p_x$, implies that changes in t give rise to additional income effects. As long as the government can optimally set g, these additional income effects do not affect any of the main results derived below. For notational convenience, we abstract from this additional term in the household budget constraint.

optimal relationship between the subsidy wedge on verifiable investments in human capital, $s + \frac{\tau r}{R}$, and the tax wedge on non-verifiable investments in human capital, $\frac{\Delta}{R} \equiv \frac{t}{1-t} - \frac{\tau r}{R}$. A capital income tax $\tau > 0$ reduces the tax wedges on verifiable and non-verifiable learning alike. An education subsidy $s > 0$, in contrast, decreases only the tax wedge on verifiable learning, while a labor tax $t > 0$ increases only the tax wedge on non-verifiable learning.

To interpret (48), we first consider the case in which the capital income tax is zero and the labor income tax is positive (i.e., $\tau = 0$, $t > 0$). In that case, an education subsidy $s > 0$ alleviates the distortionary effect of the labor tax ($t > 0$) on aggregate learning. At the same time, however, a subsidy exacerbates the distortions of the labor tax 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. By further reducing the effective cost of tax deductible inputs, an education subsidy results in even more substitution away from y^n to the tax deductible inputs x^n. The sign of the optimal education subsidy depends on the relative strengths of these two effects on the aggregate level and composition of learning. A large substitution elasticity ($\rho > 1$) 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 contrast, if individuals cannot easily substitute y^n for x^n (i.e., $\rho < 1$), x^n is subsidized rather than taxed because subsidizing the observed input x^n helps to offset the distortionary effect of the labor tax on aggregate learning.

In the presence of a capital tax, (48) shows that the composite tax wedge on verifiable learning $-(s + \frac{\tau r}{R})$ and that on non-verifiable learning $\frac{t}{1-t} - \frac{\tau r}{R}$ have the same sign if and only if the substitution elasticity between the two educational inputs exceeds unity. In the extreme case that inputs in education are perfectly substitutable (i.e., $\rho \to \infty$), taxes on verifiable education completely offset the distortionary effects of the tax-deductibility of only verifiable learning on the composition of learning, so that the tax wedges on both types of learning are the same, i.e., $-s/(1-s) = t$. In this case, the presence of verifiable inputs does not reduce the learning distortion compared to the case in which all investments in human capital are non-verifiable.

With a substitution elasticity ρ smaller than one, in contrast, a subsidy on verifiable learning should optimally offset the adverse impact on aggregate learning of the tax wedge on non-verifiable learning $\frac{t}{1-t} - \frac{\tau r}{R}$. If a positive tax wedge on non-verifiable learning $\frac{t}{1-t} - \frac{\tau r}{R} > 0$ depresses aggregate investment in human capital, a subsidy on verifiable learning is called for to boost aggregate investment in human capital. Without any substitution between the two inputs (i.e., $\rho = 0$), the government can costlessly offset all distortions of non-zero taxes on non-verifiable learning by offsetting subsidies on verifiable learning.

If the production function for aggregate learning is Cobb-Douglas ($\rho = 1$), verifiable learning is taxed in the presence of a positive capital tax τ. Intuitively, the education tax $s = -\frac{\tau r}{R} < 0$ offsets the implicit education subsidy implied by the capital income tax reducing the opportunity costs of learning. This case for a tax on learning is similar to Nielsen and Sørensen (1997), who argue that the government should tax human capital formation in the presence of a capital income tax.

4.2 Optimal labor taxation

If the government can freely set the educational subsidy, the optimal labor tax is determined by the following first-order condition (see Appendix D)

$$\xi = \frac{\Delta\beta\theta}{R}\left(\frac{\overline{1+\epsilon}}{\mu}\right) + \left(\frac{t}{1-t} + \frac{\tau r}{(1+r)}\frac{\omega}{\kappa}\right)\left(\frac{\overline{\epsilon}}{\mu}\right), \tag{49}$$

where $0 \leq \theta \equiv \frac{\rho(1-\alpha)}{\alpha+\rho(1-\alpha)} \leq 1$. θ measures the extent to which education subsidies eliminate the tax wedge on learning. If θ is small (large), education subsidies are a powerful (weak) instrument to alleviate the tax distortions on learning. Compared to the corresponding expression without verifiable learning (i.e., $\alpha = 0$ (and $\theta = 1$), see (27)), the learning distortion is reduced, since $\theta < 1$. The intuition is that the additional subsidy instrument allows the government to alleviate the distortions of the labor tax on human capital formation. 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.

With a Leontief production function of human capital ($\rho = 0$ and $\theta = 0$), the government can completely offset the tax distortions on non-verifiable learning y^n by subsidizing verifiable inputs x^n. The reason is that the government can indirectly subsidize non-verifiable inputs y^n by subsidizing verifiable inputs x^n without inducing substitution away from y^n towards x^n. With non-zero substitution between the two inputs ($\rho > 0$), however, the government cannot costlessly mimic a subsidy on y^n by subsidizing x^n, because subsidizing x^n distorts the composition of human capital accumulation towards excessive use of x^n. Hence, the learning distortion of the labor tax is not reduced to zero. In the extreme case of infinite substitution between the two inputs into human capital formation (i.e., $\rho \to \infty$), the presence of verifiable inputs does not reduce the learning distortion compared to the case in which all inputs are non verifiable. Indeed, in this case, the government finds it optimal to tax x^n (i.e., $s < 0$) so as make x^n effectively non-tax deductible. This ensures a level playing field with y^n. With Cobb Douglas production of aggregate learning (i.e., $\rho = 1$), θ corresponds to the share of non-verifiable inputs in aggregate investment in human capital $(1-\alpha)$. θ is larger (smaller) than $(1-\alpha)$ if the substitution elasticity between the inputs ρ is larger (smaller) than unity.

The formula for the optimal marginal tax on labor income in the absence of capital income taxes (i.e., $\tau = 0$ so that $\Delta = \frac{t}{1-t}R$) becomes

$$\frac{t}{1-t} = \frac{\xi}{\left(\frac{\overline{\epsilon}}{\mu}\right) + \beta\theta\left(\frac{\overline{1+\epsilon}}{\mu}\right)}. \tag{50}$$

Ceteris paribus the distributional characteristic ξ and the labor supply elasticities ϵ^n, the presence of educational subsidies for verifiable learning raises the optimal labor tax since $\theta < 1$.

4.3 Optimal capital taxation

If the government has free access to not only the labor tax t and the subsidy on investments in human capital s, but also the capital income tax τ, the first-order condition for the optimal

capital tax can be written as (see Appendix D)

$$
\begin{aligned}
& (\omega(1-\beta)-\beta)\xi + \frac{\Delta\beta\theta}{R}\left(\frac{1-\omega\epsilon}{\mu}\right) \qquad (51)\\
= & \left(\frac{t}{1-t} + \frac{\tau r}{(1+r)}\frac{\omega}{\kappa}\right)(\omega(1-\beta)-\beta)\left(\frac{\overline{\epsilon}}{\mu}\right) + \frac{\tau r}{(1+r)}\left(\omega(1-\omega)\sigma\bar{\gamma}\right).
\end{aligned}
$$

Also in this first-order condition, the availability of education subsidies reduces effective learning distortion through the additional multiplicative factor $0 \leq \theta \equiv \frac{\rho(1-\alpha)}{\alpha+\rho(1-\alpha)} \leq 1$. Otherwise, the expression for the optimal capital tax is the same as derived before with only non-verifiable investments in human capital, see equation (30).

4.4 Optimal dual income tax

Combining the two first-order conditions (49) and (51), we obtain the (see Appendix D)

$$\frac{\Delta\beta\theta}{R} = \frac{\tau r}{(1+r)}\omega^*\bar{\gamma}\sigma, \qquad (52)$$

which yields the optimal dual tax structure (using $\Delta = \frac{t}{1-t}R - \tau r$)

$$\frac{\tau r}{(1+r)} = \frac{t}{1-t}\left(\frac{(1+r)}{R} + \frac{\omega^*\sigma\bar{\gamma}}{\beta\theta}\right)^{-1}, \qquad (53)$$

and the optimal overall tax wedge on labor supply

$$\frac{t}{1-t} + \frac{\tau r}{(1+r)}\frac{\omega}{\kappa} = \frac{\xi}{\left(\frac{\overline{\epsilon}}{\mu}\right) + \beta\chi\theta\left(\frac{\overline{1+\epsilon}}{\mu}\right)}. \qquad (54)$$

Compared to the case without verifiable learning, subsidized verifiable learning in effect reduces the elasticity of human capital investment β to $\theta\beta$ (compare (52) with (32), (53) with (33), and (54) with (36)). Optimal capital taxes thus remain positive ($\tau > 0$). The only exception is the case in which education subsidies do not distort the composition of learning and can thus costlessly eliminate the entire labor-tax distortion on human capital accumulation ($\theta = 0$). This is the case only if either substitution between verifiable and non-verifiable learning is completely absent ($\rho = 0$) or all learning is verifiable ($\alpha = 1$).

Indeed, the analysis in section 3 remains valid, except that verifiable learning introduces an additional behavioral margin: substitution between verifiable and non-verifiable learning. Just as the other behavioral margins, more elastic behavior on account of a higher substitution elasticity ρ reduces the optimal overall tax wedge (54). Moreover, just as the aggregate learning and saving margin, the learning composition margin affects the composition of the tax burden (53). In particular, the capital tax becomes a more important instrument for alleviating the labor-tax distortions on learning if a high elasticity ρ renders the educational subsidy a relatively inefficient instrument to offset these learning distortions.

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. Both these instruments are imperfect, however, as they distort either the composition of learning (in the case of education subsidies) or aggregate saving (in the case of the capital income tax). At the optimum, the government balances distortions on aggregate learning with those on the composition of learning and the intertemporal allocation of saving. By optimally balancing these three distortions, the government contains the costs of redistribution.

5 Numerical simulations

To further check the robustness of our results to the presence of verifiable investments in human capital, this section quantifies optimal capital income taxes by using expression (53) for the optimal dual income tax structure. We can employ this expression assuming that governments optimized the labor tax (t) and the amount of income redistribution (g).[18] One should interpret our findings with some caution as our simple three-period life-cycle structure abstracts from several real-world complications, such as bequests and risk.

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.[19] In our three-period life-cycle model, each period captures a third of the average overall life-span. Hence, we adopt a cohort length of 25 years.

Expression (53) contains three share parameters: α, ω, and $\bar{\gamma}$. The share of observable costs in total educational expenditure is set at $\alpha = 0.5$. Becker (1964) and Boskin (1975) find that the share of goods invested in education is about one-quarter and the share of (tax deductible) forgone earnings amounts to three-quarters. We do not set α to 0.75, however, because this would ignore the effort cost of education (i.e., attending college, studying, etc). The share of retirement consumption in total consumption is set at $\omega = 0.33$. Finally, to compute $\bar{\gamma}$, we adopt an average ratio of lump-sum transfers to gross labor incomes of $\overline{g/z} = 0.25$.

As regards the three relevant behavioral elasticities in (53) (i.e., σ, β, and ρ), the largest empirical literature exists on the intertemporal elasticity of substitution in consumption σ. Whereas older papers found extremely small elasticities, more recent work (e.g., Hall (1988) and Attanasio and Weber (1995)) suggests that the intertemporal substitution elasticity is substantially positive at around $\sigma = 0.5$.[20] Trostel (1993) contains an extensive discussion on plausible parameter values for the returns to inputs invested in human capital β. Based on this, we set $\beta = 0.5$.[21] Concerning the elasticity of substitution between inputs in the

[18] The exogenously given revenue requirement Λ balances the budget.

[19] Estimated Mincer returns on education typically exceed 6%, see e.g. Ashenfelter et al. (1999). In analogy of the equity premium puzzle, this raises the so-called human capital risk premium puzzle, see Judd (2000).

[20] This value may be on the high side for a three-period model with periods of about 20 years in which intraperiod intertemporal substitution is in effect infinite. A lower value for σ would strengthen the case for large capital taxes further.

[21] The second-order conditions for individual utility maximization imply $\beta(1+\epsilon^n) < 1$. With the compensated wage elasticity of labor supply ϵ^n taking an empirically plausible value of 0.5, β should be smaller than 0.67 for the second-order conditions to be met.

human capital formation ρ, we follow Trostel (1993) by using $\rho = 1$ as the benchmark value. Hence, the production function if aggregate learning is Cobb-Douglas.

Figure 1 shows the optimal capital income taxes at given labor income taxes for the benchmark 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. In order to check whether this result is robust, Figures 2 – 9 perform extensive sensitivity analyses. The results are most sensitive to the real return on financial savings r. Capital income taxes are substantially larger than labor income taxes if the real rate of return r and the cohort length are lower than 6% and 25 years (see Figures 2 and 3). Intuitively, the aggregate capital income from financial saving is lower in this case so that the tax rate has to be higher to yield the same net implicit subsidy on learning.

As regards the behavioral elasticities, the optimality of a synthetic income tax is only violated if the elasticity of substitution in consumption (σ) substantially exceeds unity or the elasticity of human capital formation β and the elasticity of substitution between inputs in education ρ are smaller than 0.2 (see Figures 4, 5 and 6). These are rather extreme parameter values that substantially deviate from empirically estimated values. Hence, the result that optimal capital income taxes substantially exceed zero is quite robust.

The results are also not very sensitive to the shares parameters (see figures 7, 8 and 9). The case for a positive capital tax becomes substantially weaker only if the share of verifiable inputs in learning α exceeds 0.9. The shares of retirement consumption ω and lump-sum government transfers hardly affect the results. Overall, the numerical calculations suggest that optimal marginal capital income taxes are positive and quite close to the marginal tax rate on labor.

6 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. In order to investigate optimal linear taxes on labor and capital incomes, we developed a three-period life-cycle model of human capital investment, financial saving and labor supply in which individuals exhibit different abilities to learn. We demonstrated that a dual income tax with positive marginal tax rates on both labor and capital income is optimal, even if preferences are homothetic and weakly separable in leisure and consumption (these latter conditions imply an optimal zero capital tax in the absence of human capital formation). The reason why positive capital income taxes are nevertheless 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 tax rates for redistributional 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.

The optimal marginal tax on capital income is large compared to the optimal marginal tax on labor income if the level and composition of learning efforts are relatively elastic compared to aggregate saving. Education subsidies on verifiable educational inputs reduce the need for capital taxes to boost learning but do not eliminate the case for positive capital taxes. The

optimal capital tax is zero only in knife-edge cases in which non-verifiable learning is either completely absent or is directly tied to verifiable learning. In these limiting cases, education subsidies are a perfect instrument to eliminate all labor-tax distortions on human capital accumulation so that positive capital taxes are no longer needed. Numerical simulations reveal that optimal capital taxes are substantially positive for a wide variety of plausible parameterizations. In fact, a synthetic income tax, which taxes capital and labor incomes at the same marginal rates, appears to be roughly optimal. Hence, the welfare gains of replacing income taxes by consumption taxes are over-estimated if human capital formation is ignored.[22]

The case for positive capital taxes rests on efficiency concerns rather than its direct redistributional effects. In fact, the optimal capital tax remains positive even if the capital tax itself is regressive. Intuitively, with separable preferences, the labor tax is a more efficient instrument to redistribute resources than the capital tax is. The labor tax is therefore targeted at redistribution, while the capital tax is aimed solely at alleviating the learning distortions imposed by the labor income tax. As a direct consequence, the marginal tax rate on capital income is directly tied to the tax rate on labor income because a larger labor tax rate calls for a larger capital tax to alleviate the additional distortions on human capital accumulation. The capital tax therefore rises with the inequality in labor incomes rather than that in capital incomes.

Our paper is not only of theoretical interest but has policy relevance as well. In modern economies, human capital is viewed as one of the main engines of growth. At the same time, governments seek to promote private saving in order to alleviate future pressures on government budgets in aging societies. Moreover, governments increasingly rely on in-work tax credits targeted at low-skilled workers in order to strengthen the labor-market position of these workers. This paper stresses the fundamental trade-offs between the three objectives of boosting human capital formation, increasing financial saving, and redistributing resources towards low-skilled labor. Given the redistributional objectives of the government, more incentives for financial savings harm human capital accumulation. Moreover, if the government improves the employment prospects of the low skilled through more redistributive labor taxation, the government should move part of the tax burden from the labor market to the capital market in order to alleviate the distortionary effects of higher marginal labor taxes on human capital formation, even though higher capital taxes discourage private saving.

In future research, we want to explore the consequences of general equilibrium effects on wages and interest rates in a multi-period setting, possibly with a representative infinitely lived agent with capital and human capital stocks evolving dynamically over time. We also intend to explore the optimal non-linear taxes on labor incomes and perhaps also on capital incomes for the case in which the government is able to observe individual incomes rather than just the aggregates. This requires the analysis of incentive compatibility constraints. Another important is extension involves wage rigidities and union power, which can be viewed as implicit labor taxes excerbating labor-market distortions.

[22] This point is also stressed by Judd (1999).

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.

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 (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*, 130, (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.

Boskin, Michael (1975), "Notes on the Tax Treatment of Human Capital", NBER Workingpaper 116.

Bovenberg, A. Lans, and Bas Jacobs (2001), "Redistribution and Education Subsidies are Siamese Twins", CEPR Discussion Paper No. 3309.

Bovenberg, A. Lans, and Bas Jacobs (2005), "Redistribution and Education Subsidies are Siamese Twins", *Journal of Public Economics*, forthcoming.

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.

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.

Diamond, Peter A., and James A. Mirrlees (1971), "Optimal Taxation and Public Production I: Production Efficiency", *American Economic Review* 61, 8-27.

Dur, Robert, and Coen N. Teulings (2001), "Education and Efficient Redistribution", Tinbergen Institute Discussion Paper 090/3.

Erosa, Andrés, and Martin Gervais (2002), "Optimal Taxation in Life-Cycle Economies", *Journal of Economic Theory*, 105, 338-369.

Fischer, Stanley (1980), "Dynamic Inconsistency, Cooperation, and the Benevolent Dissembling Government", *Journal of Economic Dynamics and Control*, 2, 93-107.

Golosov, Mikhail, Nayarana Kocherlakota, and Aleh Tsyvinski (2003), "Optimal Indirect and Capital Taxation", *Review of Economic Studies*, 70, 569-587.

Hall, Robert E. (1988), "Intertemporal Substitution in Consumption", *Journal of Political Economy*, 96, (2), 339-357.

Heckman, James J. (1976), "A Life-Cycle Model of Earnings, Learning and Consumption", *Journal of Political Economy*, 4, S11-S44.

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.

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.

Ordover, J.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 (2003), "Risk and Market Frictions as Determinants of the Human Capital Premium", mimeo: Brown University.

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

Appendix

Appendix A: Second-order conditions

We employ a three-step budgeting procedure to derive the second-order conditions. In the first stage, individuals select aggregate consumption v^n, labor supply l^n (leisure demand $\ell^n = 1 - l^n$), and aggregate investment in human capital $e^n (\equiv \psi(x^n, y^n))$. We write the individual budget constraint as $p_c^n v^n = (1-t)n\phi(e^n)l^n - Rp_e e^n + g$, where the price indices of aggregate consumption and learning are defined as $p_c^n \equiv (c_1^n + c_2^n/R)/v(c_1^n, c_2^n)$ and

$p_e \equiv ((1-t)(1-s_x)p_x x^n + p_y y^n)/e_n$, respectively. In the first part of the paper where all educational inputs are non-observable, p_e is normalized to unity. By employing the individual budget constraint to eliminate v^n from the utility function, we arrive at the following maximization problem:

$$\max_{\{l^n, e^n\}} u\left(\frac{(1-t)n\phi(e^n)l^n}{p_c^n} - \frac{Rp_e e^n}{p_c^n} + \frac{g}{p_c^n}, 1-l^n\right), \tag{55}$$

The first-order conditions for education and labor supply are

$$(1-t)nl^n\phi'(e^n) - Rp_e = 0, \tag{56}$$

$$u_v(1-t)n\phi(e^n) - u_\ell p_c^n = 0. \tag{57}$$

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

$$H \equiv \begin{bmatrix} (1-t)nl^n\phi''(e^n) & (1-t)n\phi'(e^n) \\ u_v(1-t)n\phi'(e^n) & -u_{\ell v}(1-t)n\phi(e^n) + u_{\ell\ell}p_c \end{bmatrix}. \tag{58}$$

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

$$\beta(1+\epsilon^{n*}) < 1, \tag{59}$$

where $\epsilon^{n*} \equiv \left(\frac{lu_{ll}}{u_l} - \frac{lu_{vl}}{u_v}\right)^{-1}$. 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 $\frac{u_\ell}{u_v} = \frac{w^n}{p_c^n}$ (where $w^n \equiv (1-t)n\phi(e^n)$ is the after-tax wage) with respect to l^n and substituting $\frac{dv^n}{dl^n} = -\frac{u_l}{u_v}$ (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

$$\epsilon^n \equiv \frac{\partial l^{n*}}{\partial w^n}\frac{w^n}{l^{n*}} = \left(\left(\frac{l^n u_{v\ell}}{u_v} - \frac{l^n u_{\ell\ell}}{u_\ell}\right) - \left(\frac{u_\ell}{u_v}\right)\left(\frac{l^n u_{vv}}{u_v} - \frac{l^n u_{v\ell}}{u_\ell}\right)\right)^{-1}, \tag{60}$$

so that $(\epsilon^n)^{-1} = (\epsilon^{n*})^{-1} + \delta$, where $\delta \equiv -\left(\frac{u_\ell}{u_v}\right)\left(\frac{l^n u_{vv}}{u_v} - \frac{l^n u_{v\ell}}{u_\ell}\right) \geq 0$. We thus have $\epsilon^n \leq \epsilon^{n*}$ and

$$\beta(1+\epsilon^n) \leq \beta(1+\epsilon^{n*}) < 1. \tag{61}$$

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

In the second stage of the budgeting procedure, individuals maximize $v(c_1^n, c_2^n)$ subject to the constraint: $p_c^n v^n = c_1^n + \frac{c_2^n}{R}$. The associated second-order condition, $v_{c_1c_1}v_{c_2c_2} - (v_{c_1c_2})^2 > 0$, is satisfied because $v(c_1^n, c_2^n)$ 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 $p_e e_n = (1-t)(1-s_x)p_x x^n + p_y y^n$. 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. (43)) by $(1-t)x^n$, multiply the first-order condition for y^n (i.e. (44)) 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

$$\beta(1-t)nl^n\phi(e^n) = Rp_e e^n, \tag{62}$$

where $p_e \equiv ((1-t)(1-s)p_x x^n + p_y y^n)/e^n$.

Log-linearizing (62) (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. (43) and (44)), we arrive at

$$-\tilde{t} + l^n + \beta\tilde{e}^n = \tilde{p}_e + \tilde{e}^n + \tilde{R}, \tag{63}$$

$$\tilde{p}_e = -\alpha(\tilde{t} + \tilde{s}), \tag{64}$$

where a tilde stands for a *compensated* relative change, except for the tax rate and the subsidy rates where $\tilde{t} \equiv dt/(1-t)$ and $\tilde{s} \equiv ds/(1-s)$. $\alpha \equiv \frac{x^n\psi_x}{\psi} = \frac{(1-t)(1-s)p_x x^n}{p_e e^n}$ does not depend on ability. The reason is that the first-order conditions for x^n and y^n ((43) and (44)) imply that the marginal rate of transformation does not depend on n:

$$\frac{\psi_x}{\psi_y} = \frac{(1-t)(1-s)p_x}{p_y}. \tag{65}$$

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). Hence, (65) determines the ratio of the two inputs x^n/y^n as a function of $\frac{(1-t)(1-s)p_x}{p_y}$ only. $\alpha = \frac{(1-t)(1-s)p_x x^n}{(1-t)(1-s)p_x x^n + p_y y^n} = \frac{(1-t)(1-s)p_x(x^n/y^n)}{(1-t)(1-s)p_x(x^n/y^n)+p_y}$ depends therefore only on $(1-t)(1-s)p_x$ and p_y – and not on ability.

The first-order condition for labor supply (57) can be written as

$$\frac{w^n}{p_c^n} = \frac{u_\ell}{u_v}, \tag{66}$$

where $w^n \equiv (1-t)n\phi(e^n)$ and $p_c^n \equiv \left(c_1^n + \frac{c_2^n}{R}\right)/v(c_1^n, c_2^n)$. The compensated relative changes in labor supply l^n and composite consumption v^n are related as follows (differentiate the utility function $u(v^n, 1-l^n)$ and set the total differential to zero): $(u_v v^n/u_\ell l^n)\tilde{v}^n = \tilde{l}^n$. Substitution of the first-order condition for labor supply to eliminate u_v yields

$$\gamma^n\tilde{v}^n = \tilde{l}^n, \tag{67}$$

where $\gamma^n \equiv \frac{p_c^n v^n}{(1-t)l^n n\phi(e^n)} = \frac{(1-\beta)(1-t)l^n n\phi(e^n)+g}{(1-t)l^n n\phi(e^n)} = 1 - \beta + \frac{g}{(1-t)z^n}$ (where the second equality is derived from the definition of $p_c^n \equiv \left(c_1^n + \frac{c_2^n}{R}\right)/v(c_1^n, c_2^n)$ and (11)). Log-linearizing the definition of p_c^n (i.e. $p_c^n v(c_1^n; c_2^n) = c_1^n + \frac{c_2^n}{R}$), we arrive at

$$\tilde{p}_c^n = -\omega\tilde{R} + \left(\frac{1-\kappa}{\kappa}\right)\tilde{v}^n, \tag{68}$$

where κ is the degree of homogeneity of $v(c_1^n, c_2^n)$ so that

$$\tilde{v}^n = \kappa\left((1-\omega)\tilde{c}_1^n + \omega\tilde{c}_2^n\right). \tag{69}$$

The right-hand side of (66) depends only on l^n and v^n. Substituting (67) and (68) to eliminate $\tilde{v}^n$ and $\tilde{p}_c^n$ from the log-linearized version of (66), we arrive at the following labor-supply function:

$$\tilde{l}^n = \epsilon^n(\tilde{w}^n + \omega\tilde{R}), \tag{70}$$

where $\epsilon^n \equiv \frac{\partial l^{n*}}{\partial w^n}\frac{w^n}{l^{n*}}$ is a parameter governing the compensated wage elasticity of second-period labor supply with respect to the after-tax second-period wage.

Consumption in the two periods is found by log-linearizing (10):

$$\tilde{c}_1^n - \tilde{c}_2^n = -\sigma\tilde{R}, \tag{71}$$

where $\sigma \equiv d\log\left(\frac{c_1^n}{c_2^n}\right)/d\log\left(\frac{u_{c_2}}{u_{c_1}}\right)$ is the intertemporal substitution elasticity in consumption. This elasticity does not depend on ability n. The reason is that the left-hand side of (10) depends only on c_1^n/c_2^n (as the sub-utility function $v(c_1^n, c_2^n)$ is homothetic), so that one can express c_1^n/c_2^n as a function of R only (independent of ability n). Accordingly, the elasticity σ, which depends on c_1^n/c_2^n only, does not depend on ability n.

Substitution of (71) into (69) to eliminate $\tilde{c}_1^n$ yields (where the second equality follows from (67)):

$$\tilde{c}_2^n = \frac{\tilde{v}^n}{\kappa} + (1-\omega)\sigma\tilde{R} = \frac{\tilde{l}^n}{\kappa\gamma^n} + (1-\omega)\sigma\tilde{R}. \tag{72}$$

Substituting $\tilde{w}^n = -\tilde{t} + \beta\tilde{e}^n$ (which follows from $w^n \equiv (1-t)n\phi(e^n)$) into (70), and then substituting the result and (64) into (63) to eliminate, respectively, l^n and $\tilde{p}_e$, we can solve for $\tilde{e}_n$ ::

$$\tilde{e}^n = \frac{1}{\mu^n}\left((\alpha - (1+\epsilon^n))\tilde{t} - (1-\omega\epsilon^n)\tilde{R} + \alpha\tilde{s}\right), \tag{73}$$

where $\mu^n \equiv 1 - \beta(1+\epsilon^n) > 0$ from the second-order conditions.

Substituting $\tilde{w}^n = -\tilde{t} + \beta\tilde{e}^n$ and (73) into (70), we solve for $\tilde{l}^n$:

$$\tilde{l}^n = \left(-\frac{\epsilon^n(1-\beta\alpha)}{\mu_n}\right)\tilde{t} + \frac{\epsilon^n}{\mu_n}\left(\alpha\beta\tilde{s} + (\omega(1-\beta)-\beta)\tilde{R}\right). \tag{74}$$

Substitution of this reduced form into (72) yields

$$\tilde{c}_2^n = \frac{\epsilon^n}{\mu^n\kappa\gamma^n}\left(-(1-\beta\alpha)\tilde{t} + (\omega(1-\beta)-\beta)\tilde{R} + \alpha\beta\tilde{s}\right) + (1-\omega)\sigma\tilde{R}. \tag{75}$$

From the last three expressions follow the compensated elasticities of aggregate learning, labor supply and third-period consumption with respect to the policy variables.

In the first part of the paper, all learning is non verifiable (i.e., $\alpha = 0$). Hence, the elasticities of learning, labor supply and third-period consumption follow from (use (63), (70), and (72) with $\alpha = 0$):

$$\tilde{e}^n = -\left(\frac{(1+\epsilon^n)\tilde{t} + (1-\omega\epsilon^n)\tilde{R}}{\mu^n}\right), \tag{76}$$

$$\tilde{l}^n = \frac{\epsilon^n}{\mu_n}\left(-\tilde{t} + (\omega(1-\beta)-\beta)\tilde{R}\right), \tag{77}$$

$$\tilde{c}_2^n = \frac{\epsilon^n}{\mu^n \kappa \gamma^n}\left(-\tilde{t} + (\omega(1-\beta)-\beta)\tilde{R}\right) + (1-\omega)\sigma\tilde{R}. \tag{78}$$

The second part of the paper analyzes $\alpha > 0$. We find x^n and 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

$$\tilde{e}^n = \alpha\tilde{x}^n + (1-\alpha)\tilde{y}^n. \tag{79}$$

Differentiation of (65) yields

$$\tilde{x}^n - \tilde{y}^n = \rho(\tilde{t} + \tilde{s}), \tag{80}$$

where $\rho \equiv d\log\left(\frac{x^n}{y^n}\right)/d\log(\frac{\psi_y}{\psi_x})$ stands for the elasticity of substitution between the two inputs in the production of human capital. This substitution elasticity depends only on $\frac{x^n}{y^n}$ and is independent of ability because $\frac{x^n}{y^n}$ is a function of $\frac{(1-t)(1-s)p_x}{p_y}$ only (see the discussion above on α being independent of ability). From (79) and (80), we can express $\tilde{x}$ and $\tilde{y}$ in terms of $\tilde{e}$; substituting (73) to then eliminate $\tilde{e}$, we find:

$$\tilde{x}^n = \frac{(\alpha - (1+\epsilon^n))\tilde{t} - (1-\omega\epsilon^n)\tilde{R} + \alpha\tilde{s}}{\mu^n} + (1-\alpha)\rho(\tilde{t}+\tilde{s}), \tag{81}$$

$$\tilde{y}^n = \frac{(\alpha - (1+\epsilon^n))\tilde{t} - (1-\omega\epsilon^n)\tilde{R} + \alpha\tilde{s}}{\mu^n} - \alpha\rho(\tilde{t}+\tilde{s}). \tag{82}$$

These expressions yield the elasticities for x^n and y^n with respect to changes in policy.

Appendix C: Optimal taxation with non-verifiable learning

We can rewrite the first-order condition for t (26) by using the Slutsky equations

$$\frac{\partial q}{\partial t} \equiv \frac{\partial q^*}{\partial t} - nl^n\phi(e^n)\frac{\partial q}{\partial g}, \tag{83}$$

where $q = e^n, l^n, c_2^n$. A star denotes a compensated change. Substitution of the Slutsky terms into (26) and use of the definition of b^n (see 23)) results in

$$\begin{aligned}\frac{\partial \mathcal{L}}{\partial t} &= \int_{\underline{n}}^{\overline{n}} \left[(-b^n+1)\, z^n\right] dF(n) + \\ &\int_{\underline{n}}^{\overline{n}} \left[\frac{\Delta}{1-t}\frac{e^n}{z^n}z^n\varepsilon_{et}^n + \frac{t}{1-t}z^n\varepsilon_{lt}^n + \frac{\tau r}{(1+r)}\frac{c_2^n/R}{(1-t)z^n}z^n\varepsilon_{c_2t}^n\right] dF(n) = 0,\end{aligned} \tag{84}$$

where we defined the elasticities $\varepsilon_{qt} \equiv \frac{\partial q^*}{\partial t}\frac{1-t}{q}$ for $q = e^n$, l^n and c_2^n. Substituting the elasticities from (76), (77), and (78) into (84), we arrive at (using (8) to eliminate $\frac{e^n}{z^n}$, $\frac{c_2^n/R}{(1-t)z^n} = \omega\gamma^n$, and (24)) the expression in the main text (27).

The first-order condition for the optimal capital income tax (29) is rewritten by employing the definition of b_n (see (23)) and the Slutsky equations

$$\frac{\partial q}{\partial R} \equiv \frac{\partial q^*}{\partial R} + \left(\frac{c_2^n}{R^2} - e^n\right)\frac{\partial x}{\partial g}, \tag{85}$$

where $q = e^n, l^n, c_2^n$. Using $\frac{c_2^n}{R} = \omega\left((1-\beta)(1-t)z_n + g\right)$ (from (12)) and $e^n = \frac{(1-t)\beta}{R}z^n$ (from (8)) to rewrite the first term at the first right-hand side, we find

$$\begin{aligned}\frac{\partial \mathcal{L}}{\partial R} &= -\frac{(\omega(1-\beta)-\beta)(1-t)}{R}\int_{\underline{n}}^{\overline{n}}\left[(-b^n+1)\,z^n\right]dF(n) + \\ &\int_{\underline{n}}^{\overline{n}}\left[\frac{\Delta}{R}\frac{e^n}{z^n}z^n\varepsilon_{eR}^n + \frac{t}{R}z^n\varepsilon_{lR}^n + \frac{\tau r}{(1+r)R}\frac{c_2^n/R}{z^n}z^n\varepsilon_{c_2R}^n\right]dF(n) = 0,\end{aligned} \tag{86}$$

where $\varepsilon_{qR} \equiv \frac{\partial q^*}{\partial R}\frac{R}{q}$ for $q = e^n$, l^n and c_2^n. Substituting the elasticities from (76), (77), and (78) into (84) and dividing both sides by $\frac{(1-t)}{R}$, we arrive at (30) (using (8) to eliminate $\frac{e^n}{z^n}$, $\frac{c_2^n/R}{(1-t)z^n} = \omega\gamma^n$ with $\gamma^n \equiv \frac{(1-\beta)(1-t)nl^n\phi(e^n)+g}{(1-t)nl^n\phi(e^n)}$, and (24)).

Multiplying (27) by $(\omega(1-\beta)-\beta)$ and subtracting (30), we find

$$\frac{\Delta\beta}{R}\left[\left(\overline{\frac{1-\omega\epsilon}{\mu}}\right) + (\omega(1-\beta)-\beta)\left(\overline{\frac{1+\epsilon}{\mu}}\right)\right] = \frac{\tau r}{1+r}\omega(1-\omega)\bar{\gamma}\sigma. \tag{87}$$

Using $\left(\overline{\frac{1-\omega\epsilon}{\mu}}\right) + (\omega(1-\beta)-\beta)\left(\overline{\frac{1+\epsilon}{\mu}}\right) = 1+\omega$, we arrive at (32).

Appendix D: Optimal taxation with verifiable learning

Roy's lemma yields

$$\frac{\partial \upsilon(g,t,R,s,n)}{\partial g} = \lambda^n, \tag{88}$$

$$\frac{\partial \upsilon(g,t,R,s,n)}{\partial t} = -\lambda^n\left(nl^n\phi(e^n) - R(1-s)p_x x^n\right), \tag{89}$$

$$\frac{\partial \upsilon(g,t,R,s,n)}{\partial R} = -\lambda^n\left(-p_e e^n + \frac{c_2^n}{R^2}\right), \tag{90}$$

$$\frac{\partial \upsilon(g,t,R,s,n)}{\partial s} = \lambda^n R(1-t)p_x x^n, \tag{91}$$

where $p_e \equiv \left((1-t)(1-s)p_x x^n + p_y y^n\right)/e^n$ and λ^n is marginal utility of second-period income (see Appendix B).

The government budget constraint (47) can be written in terms of t and R as

$$\begin{aligned}&\int_{\underline{n}}^{\overline{n}}\left[t\left(nl^n\phi(e^n) - (1-s)p_x x^n R\right) + R\left(p_x x^n + p_y y^n\right) + \frac{c_2^n}{R}\right]dF(n) \\ =\ &\int_{\underline{n}}^{\overline{n}}\left[(1+r)\left(p_x x^n + p_y y^n\right) + \frac{c_2^n}{1+r} + Rsp_x x^n + g + \Lambda\right]dF(n).\end{aligned} \tag{92}$$

The Lagrangian for maximizing social welfare amounts to

$$\begin{aligned}\max_{\{g,t,R,s\}} \mathcal{L} &= \int_{\underline{n}}^{\overline{n}} \Psi\left(v(g,t,R,s,n)\right) dF(n) \\ &+\eta \int_{\underline{n}}^{\overline{n}} \left[t\left(nl^n\phi(e^n) - R(1-s)p_x x^n\right) + R\left(p_x x^n + p_y y^n\right) + \frac{c_2^n}{R}\right] dF(n) \\ &-\eta \int_{\underline{n}}^{\overline{n}} \left[(1+r)\left(p_x x^n + p_y y^n\right) + \frac{c_2^n}{1+r} + Rsp_x x^n + g + \Lambda\right] dF(n).\end{aligned} \tag{93}$$

The first-order condition for maximization of social welfare with respect to g is given by

$$\begin{aligned}\frac{\partial \mathcal{L}}{\partial g} = \int_{\underline{n}}^{\overline{n}} \left[\Psi'\lambda^n - \eta + \eta\Delta_x p_x \frac{\partial x^n}{\partial g} + \eta\Delta p_y \frac{\partial y^n}{\partial g}\right] dF(n) + \\ \int_{\underline{n}}^{\overline{n}} \left[\eta tn\phi(e^n)\frac{\partial l^n}{\partial g} + \eta\frac{\tau r}{(1+r)R}\frac{\partial c_2^n}{\partial g}\right] dF(n) = 0,\end{aligned} \tag{94}$$

where $\Delta_x \equiv -\left(Rs + \tau r\right)$ and $\Delta \equiv \frac{t}{1-t}R - \tau r$ 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

$$b^n \equiv \frac{\Psi'\lambda^n}{\eta} + \Delta_x p_x \frac{\partial x^n}{\partial g} + \Delta p_y \frac{\partial y^n}{\partial g} + tn\phi(e^n)\frac{\partial l^n}{\partial g} + \frac{\tau r}{(1+r)R}\frac{\partial c_2^n}{\partial g}, \tag{95}$$

we find

$$\int_{\underline{n}}^{\overline{n}} b^n dF(n) = 1. \tag{96}$$

Using (89), (90), and (91), we write the first-order conditions for t, R, and s as

$$\begin{aligned}\frac{\partial \mathcal{L}}{\partial t} &= \int_{\underline{n}}^{\overline{n}} \left[-\Psi'\lambda\left(nl^n\phi(e^n) - R(1-s)p_x x^n\right)\right] dF(n) + \\ &\int_{\underline{n}}^{\overline{n}} \eta n\left(nl^n\phi(e^n) - R(1-s)p_x x^n\right) dF(n) + \\ &\int_{\underline{n}}^{\overline{n}} \eta\left[\Delta_x p_x \frac{\partial x^n}{\partial t} + \Delta p_y \frac{\partial y^n}{\partial t} + tn\phi(e^n)\frac{\partial l^n}{\partial t} + \frac{\tau r}{(1+r)R}\frac{\partial c_2^n}{\partial t}\right] dF(n) = 0,\end{aligned} \tag{97}$$

$$\begin{aligned}\frac{\partial \mathcal{L}}{\partial R} &= \int_{\underline{n}}^{\overline{n}} \left[\Psi'\lambda^n\left(\frac{c_2^n}{R^2} - p_e e^n\right) - \eta\left(\frac{c_2^n}{R^2} - p_e e^n\right)\right] dF(n) + \\ &\int_{\underline{n}}^{\overline{n}} \eta\left[\Delta_x p_x \frac{\partial x^n}{\partial R} + \Delta p_y \frac{\partial y^n}{\partial R} + tn\phi(e^n)\frac{\partial l^n}{\partial R} + \frac{\tau r}{(1+r)R}\frac{\partial c_2^n}{\partial R}\right] dF(n) = 0,\end{aligned} \tag{98}$$

$$\begin{aligned}\frac{\partial \mathcal{L}}{\partial s} &= \int_{\underline{n}}^{\overline{n}} \left[\Psi'\lambda^n R(1-t)p_x x^n - \eta R(1-t)p_x x^n\right] dF(n) + \\ &\int_{\underline{n}}^{\overline{n}} \eta\left[\Delta_x p_x \frac{\partial x^n}{\partial s} + \Delta p_y \frac{\partial y^n}{\partial s} + tn\phi(e^n)\frac{\partial l^n}{\partial s} + \frac{\tau r}{(1+r)R}\frac{\partial c_2^n}{\partial s}\right] dF(n) = 0.\end{aligned} \tag{99}$$

Substituting the Slutsky equations, the definition of b_n (95), and the proportionality factors between gross income z^n and the tax and subsidy bases (i.e., $x^n = \frac{\alpha\beta}{(1-s)p_x R} z^n$ (from (45)), $y^n = \frac{\beta(1-\alpha)(1-t)}{p_y R} z^n$ (from (46)), $p_e e^n = \frac{\beta(1-t)}{R} z_n$ (from (62)), $\frac{c_2^n}{R} = \omega[(1-\beta)(1-t)z^n + g]$ (from (12)) into these first-order conditions, we establish

$$\begin{aligned}\frac{\partial \mathcal{L}}{\partial t} &= \int_{\underline{n}}^{\overline{n}} \left[(-b+1)(1-\alpha\beta) z^n dF(n) + \frac{\Delta_x}{1-t} \frac{p_x x^n}{z^n} z^n \varepsilon_{xt} + \frac{\Delta}{1-t} \frac{p_y y^n}{z^n} z^n \varepsilon_{yt} \right] dF(n) + \\ &\int_{\underline{n}}^{\overline{n}} \left[\frac{t}{1-t} z^n \varepsilon_{lt}^n + \frac{\tau r}{(1+r)} \frac{c_2^n/R}{(1-t)z^n} z^n \varepsilon_{c_2 t}^n \right] dF(n) = 0, \end{aligned} \tag{100}$$

$$\begin{aligned}\frac{\partial \mathcal{L}}{\partial R} &= \int_{\underline{n}}^{\overline{n}} \left[(b-1) \frac{(\omega(1-\beta)-\beta)(1-t)}{R} z^n + \frac{\Delta_x}{R} \frac{p_x x^n}{z^n} z^n \varepsilon_{xR} + \frac{\Delta}{R} \frac{p_y y^n}{z^n} z^n \varepsilon_{yR} \right] dF(n) + \\ &\int_{\underline{n}}^{\overline{n}} \left[\frac{t}{R} z^n \varepsilon_{lR}^n + \frac{\tau r}{(1+r)} \frac{c_2^n/R}{R z^n} z^n \varepsilon_{c_2 R}^n \right] dF(n) = 0, \end{aligned} \tag{101}$$

$$\begin{aligned}\frac{\partial \mathcal{L}}{\partial s} &= \int_{\underline{n}}^{\overline{n}} \left[(b-1) \frac{\alpha\beta(1-t)}{(1-s)} z^n + \frac{\Delta_x}{(1-s)} \frac{p_x x^n}{z^n} z^n \varepsilon_{xs} + \frac{\Delta}{(1-s)} \frac{p_y y^n}{z^n} z^n \varepsilon_{ys} \right] dF(n) + \\ &\int_{\underline{n}}^{\overline{n}} \left[\frac{t}{(1-s)} z^n \varepsilon_{ls}^n + \frac{\tau r}{(1+r)} \frac{c_2^n/R}{(1-s)z^n} z^n \varepsilon_{c_2 s}^n \right] dF(n) = 0, \end{aligned} \tag{102}$$

where we defined the compensated elasticities $\varepsilon_{qt} \equiv \frac{\partial q^*}{\partial t} \frac{(1-t)}{q}$; $\varepsilon_{qR} \equiv \frac{\partial q^*}{\partial R} \frac{R}{q}$; $\varepsilon_{qs} \equiv \frac{\partial q^*}{\partial s} \frac{(1-s)}{q}$ for $q = x^n$, y^n, l^n and c_2^n. We substitute $\frac{p_x x^n}{z^n} = \frac{\alpha\beta}{(1-s)R}$, $\frac{p_y y^n}{z^n} = \frac{\beta(1-\alpha)(1-t)}{R}$, $\frac{c_2^n/R}{z^n} = \omega\gamma^n(1-t)$ (from the definition of $\gamma^n \equiv \frac{p_c^n v^n}{(1-t)l^n n\phi(e^n)}$), the elasticities from (74) and (75), the first-order condition for g (96), and the definition of the distributional characteristic (25) to arrive at

$$\xi - \left(\frac{t}{1-t} + \frac{\tau r}{(1+r)} \frac{\omega}{\kappa} \right) \left(\frac{\overline{\epsilon}}{\mu} \right) + \frac{\Delta_x/R}{(1-s)(1-t)} \frac{\alpha\beta\bar{\varepsilon}_{xt}}{(1-\alpha\beta)} + \frac{\Delta}{R} \frac{\beta(1-\alpha)\bar{\varepsilon}_{yt}}{(1-\alpha\beta)} = 0, \tag{103}$$

$$\begin{aligned}&\left[\xi - \left(\frac{t}{1-t} + \frac{\tau r}{(1+r)} \frac{\omega}{\kappa} \right) \left(\frac{\overline{\epsilon}}{\mu} \right) \right] (\omega(1-\beta)-\beta) \\ &- \frac{\tau r}{(1+r)} (\omega(1-\omega)\sigma\bar{\gamma}) - \frac{\Delta_x/R}{(1-s)(1-t)} \alpha\beta\bar{\varepsilon}_{xR} - \frac{\Delta}{R} \beta(1-\alpha)\bar{\varepsilon}_{yR} = 0 \end{aligned} \tag{104}$$

$$\xi - \left(\frac{t}{1-t} + \frac{\tau r}{(1+r)} \frac{\omega}{\kappa} \right) \left(\frac{\overline{\epsilon}}{\mu} \right) - \frac{\Delta_x/R}{(1-s)(1-t)} \bar{\varepsilon}_{xs} - \frac{\Delta}{R} \frac{(1-\alpha)\bar{\varepsilon}_{ys}}{\alpha} = 0. \tag{105}$$

To find the optimal labor tax if we are also able to optimize the educational subsidies s, we subtract (105) from (103):

$$\frac{\Delta_x/R}{(1-t)(1-s)} \left(\frac{\alpha\beta\bar{\varepsilon}_{xt}}{(1-\alpha\beta)} + \bar{\varepsilon}_{xs} \right) = -\frac{\Delta}{R} \left(\frac{(1-\alpha)\beta\bar{\varepsilon}_{yt}}{(1-\alpha\beta)} + \frac{(1-\alpha)\bar{\varepsilon}_{ys}}{\alpha} \right). \tag{106}$$

Substitution of the corresponding elasticities from (81) and (82) yields

$$\frac{\Delta_x/R}{(1-t)(1-s)} = -\frac{\Delta}{R}\left(\frac{(1-\alpha)(1-\rho)}{1-(1-\alpha)(1-\rho)}\right). \tag{107}$$

Substituting $\Delta_x \equiv -(Rs+\tau r)$ and $\Delta \equiv \frac{t}{1-t}R - \tau r$, we arrive at equation (48) in the main text.

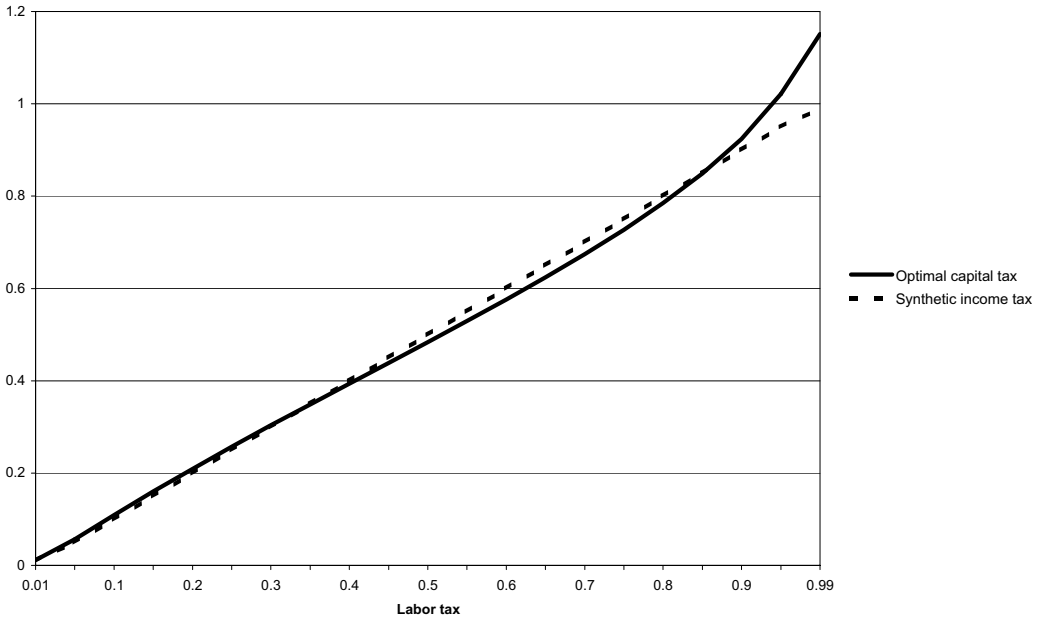
We establish equation (49) in the main text by substituting (107) into (103) and using (81) and (82) to eliminate the elasticities $\bar{\varepsilon}_{xt}$ and $\bar{\varepsilon}_{yt}$ from the resulting expression. Similarly, we derive (51) by substituting (107) into (104) and using (81) and (82) to eliminate the elasticities $\bar{\varepsilon}_{xR}$ and $\bar{\varepsilon}_{yR}$ from the resulting expression.

We arrive at (52) by multiplying (49) with $(\omega(1-\beta)-\beta)$ and subtracting (49):

$$\frac{\Delta\beta\theta}{R}\left[\left(\overline{\frac{1-\omega\epsilon}{\mu}}\right) + (\omega(1-\beta)-\beta)\left(\overline{\frac{1+\epsilon}{\mu}}\right)\right] = \frac{\tau r}{1+r}\omega(1-\omega)\bar{\gamma}\sigma, \tag{108}$$

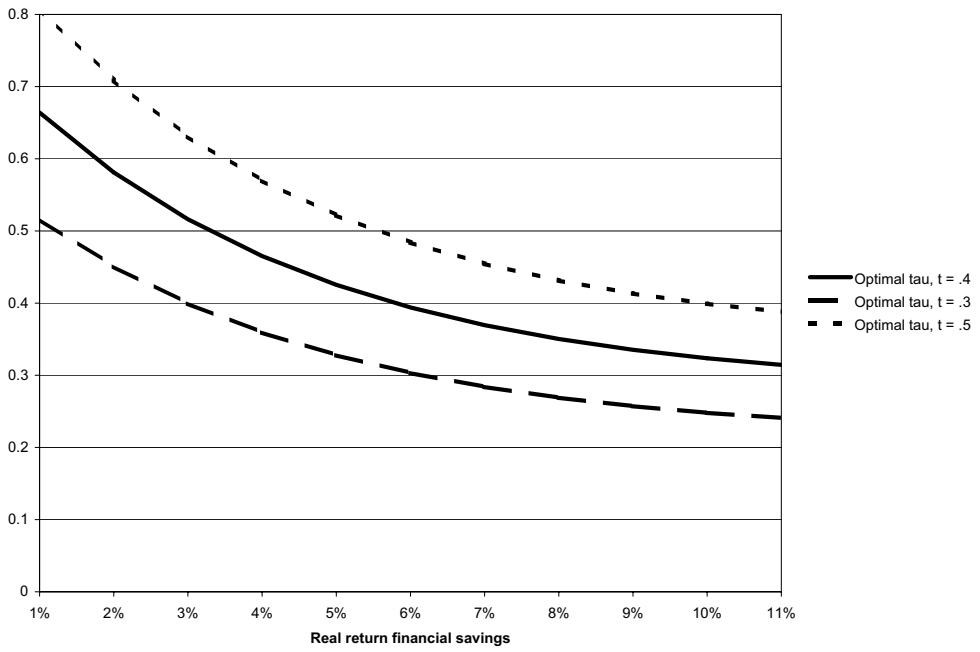
where $\left(\overline{\frac{1-\omega\epsilon}{\mu}}\right) + (\omega(1-\beta)-\beta)\left(\overline{\frac{1+\epsilon}{\mu}}\right) = 1+\omega$.

Figure 1: Optimal capital income taxes



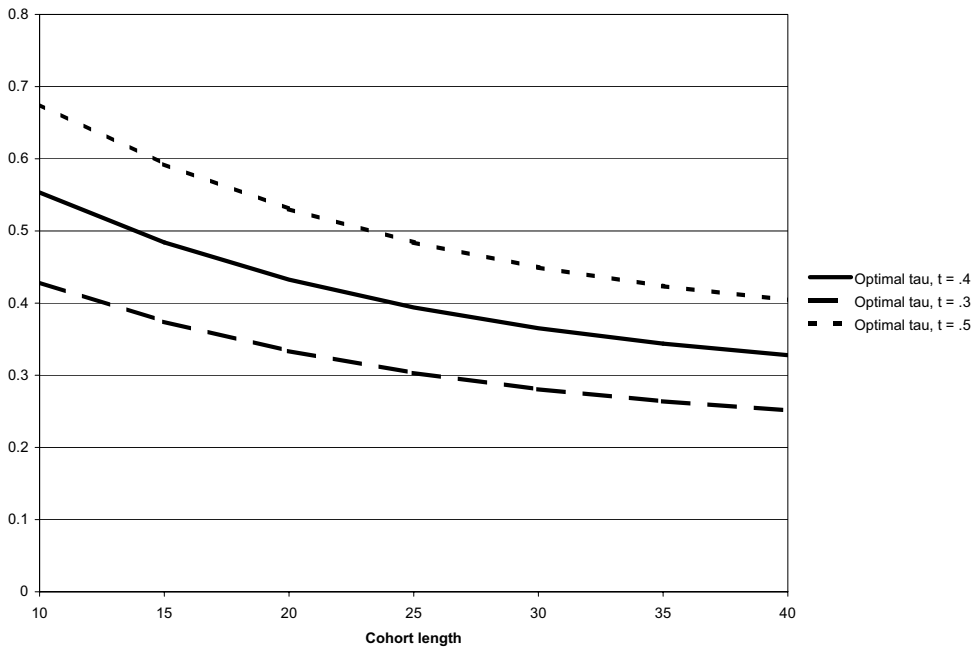
Key: Optimal capital income taxes as a function of optimal labor income taxes at baseline values: $r = 0.06$ per year, cohort length is 25 years, $\sigma = 0.5$, $\beta = 0.6$, $\rho = 1$, $\alpha = 0.5$, $\omega = 0.33$, and $\overline{g/z} = 0.25$.

Figure 2: Optimal capital income taxes and returns to financial investments



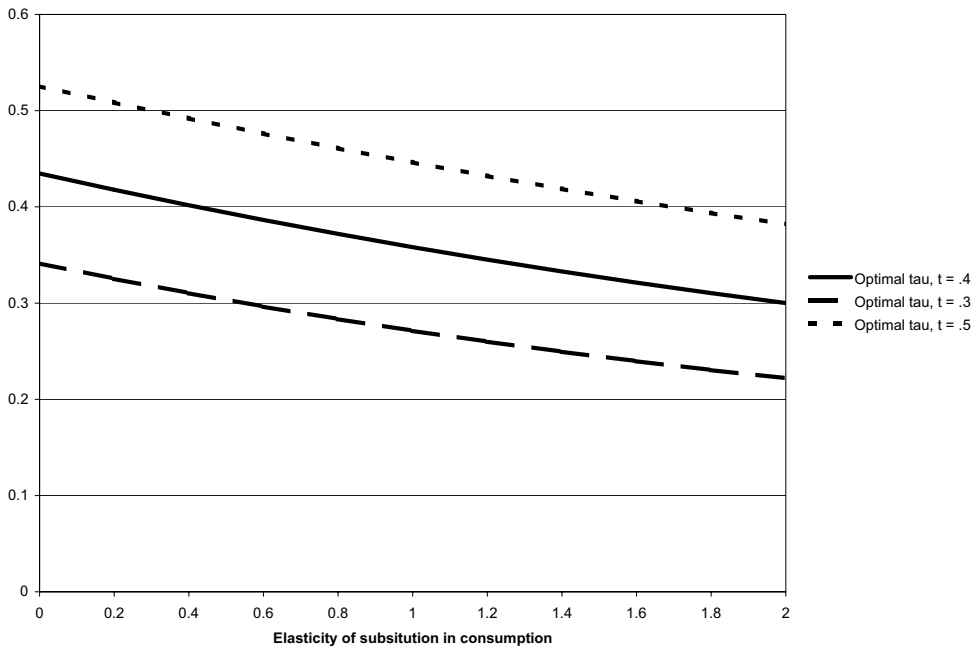
Key: Optimal capital income taxes at given optimal labor income taxes when the real return to financial investments r varies. Other parameters take baseline values: cohort length is 25 years, $\sigma = 0.5$, $\beta = 0.5$, $\rho = 1$, $\alpha = 0.5$, $\omega = 0.33$, and $\overline{g/z} = 0.25$.

Figure 3: Optimal capital income taxes and cohort length



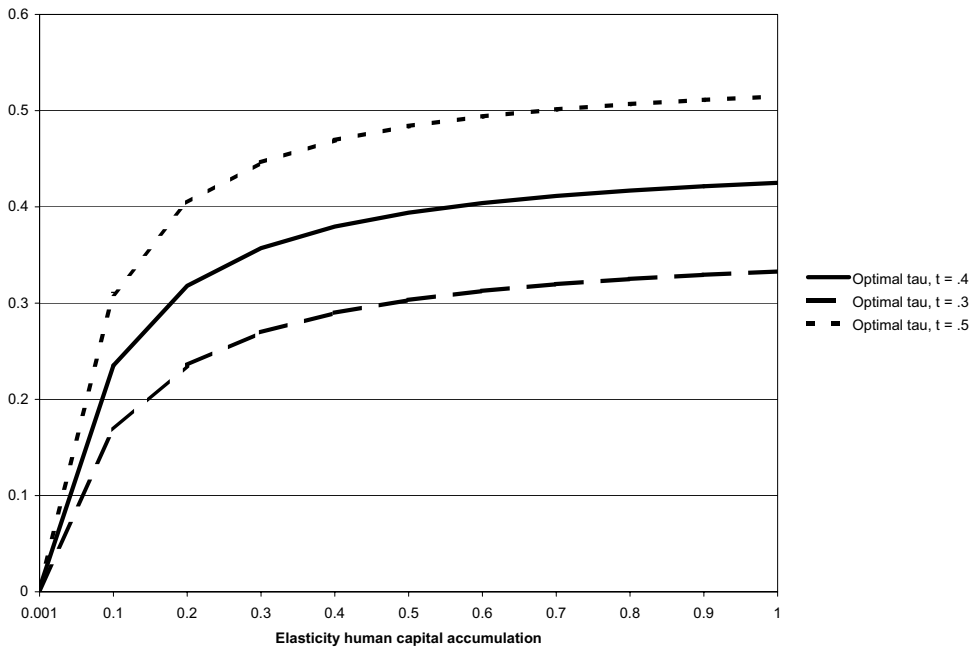
Key: Optimal capital income taxes at given optimal labor income taxes when the cohort length varies. Other parameters take baseline values: $r = 0.06$, $\sigma = 0.5$, $\beta = 0.5$, $\rho = 1$, $\alpha = 0.5$, $\omega = 0.33$, and $\overline{g/z} = 0.25$.

Figure 4: Optimal capital income taxes and the intertemporal elasticity of substitution



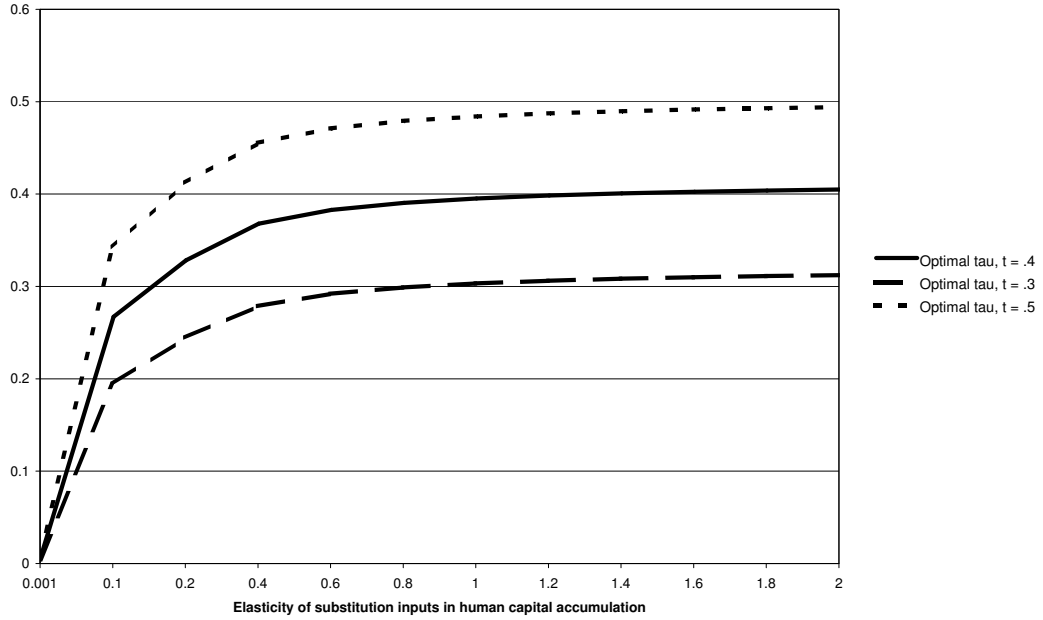
Key: Optimal capital income taxes at given optimal labor income taxes when the intertemporal elasticity of substitution in consumption σ varies. Other parameters take baseline values: $r = 0.06$, cohort length is 25 years, $\beta = 0.5$, $\rho = 1$, $\alpha = 0.5$, $\omega = 0.33$, and $\overline{g/z} = 0.25$.

Figure 5: Optimal capital income taxes and elasticity of human capital formation



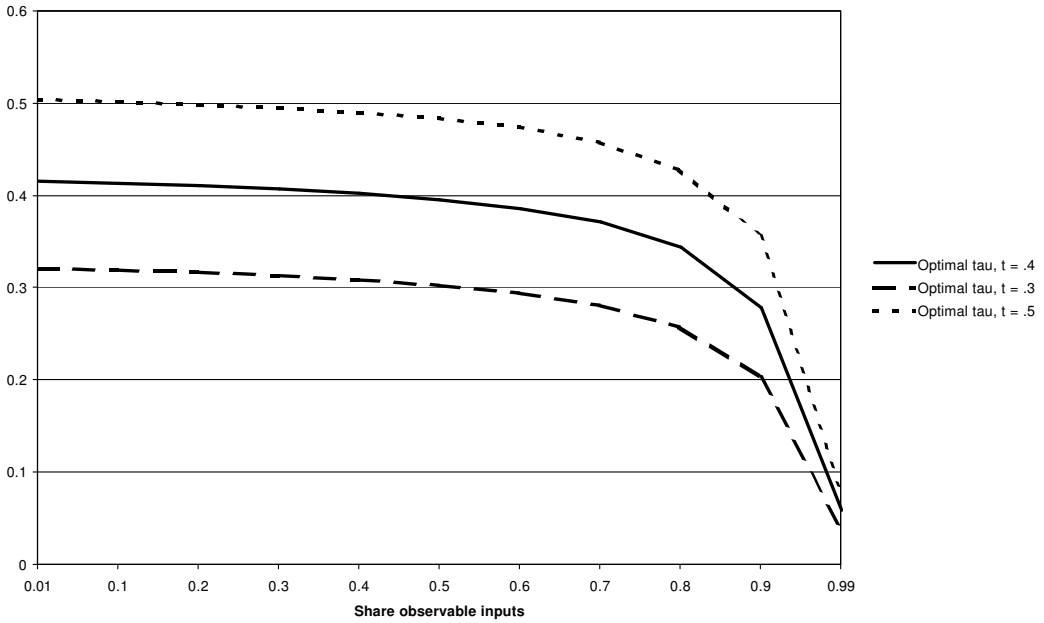
Key: Optimal capital income taxes at given optimal labor income taxes when the elasticity of human capital formation β varies. Other parameters take baseline values: $r = 0.06$, cohort length is 25 years, $\sigma = 0.5$, $\rho = 1$, $\alpha = 0.5$, $\omega = 0.33$, and $\overline{g/z} = 0.25$.

Figure 6: Optimal capital income taxes and the elasticity of substitution between inputs in human capital formation



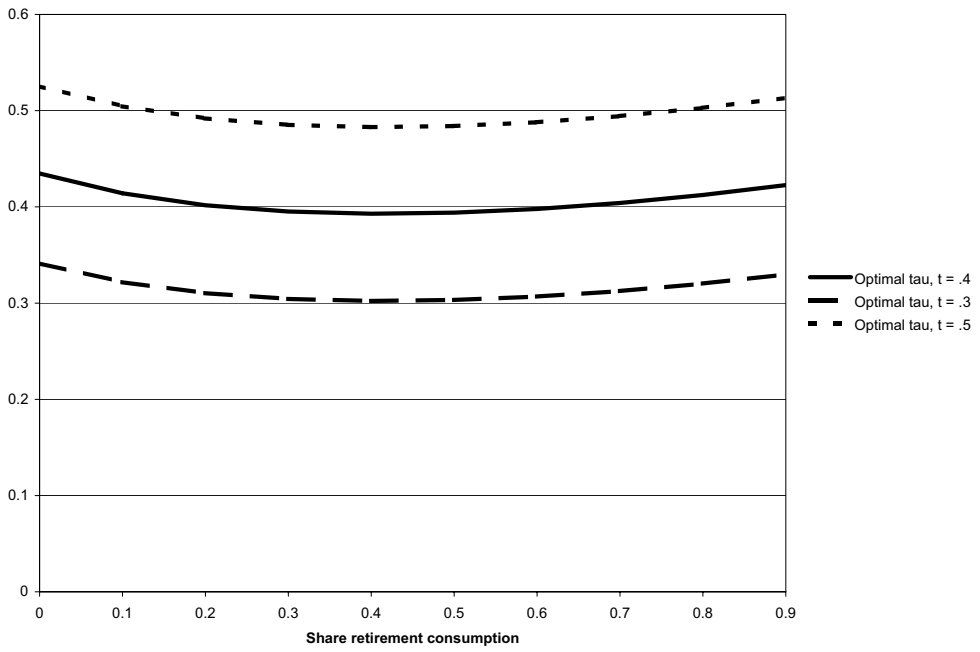
Key: Optimal capital income taxes at given optimal labor income taxes when the elasticity of substitution in human capital formation ρ varies. Other parameters take baseline values: $r = 0.06$, cohort length is 25 years, $\sigma = 0.5$, $\beta = 0.5$, $\alpha = 0.5$, $\omega = 0.33$, and $\overline{g/z} = 0.25$.

Figure 7: Optimal capital income taxes and the share of verifiable inputs in human capital formation



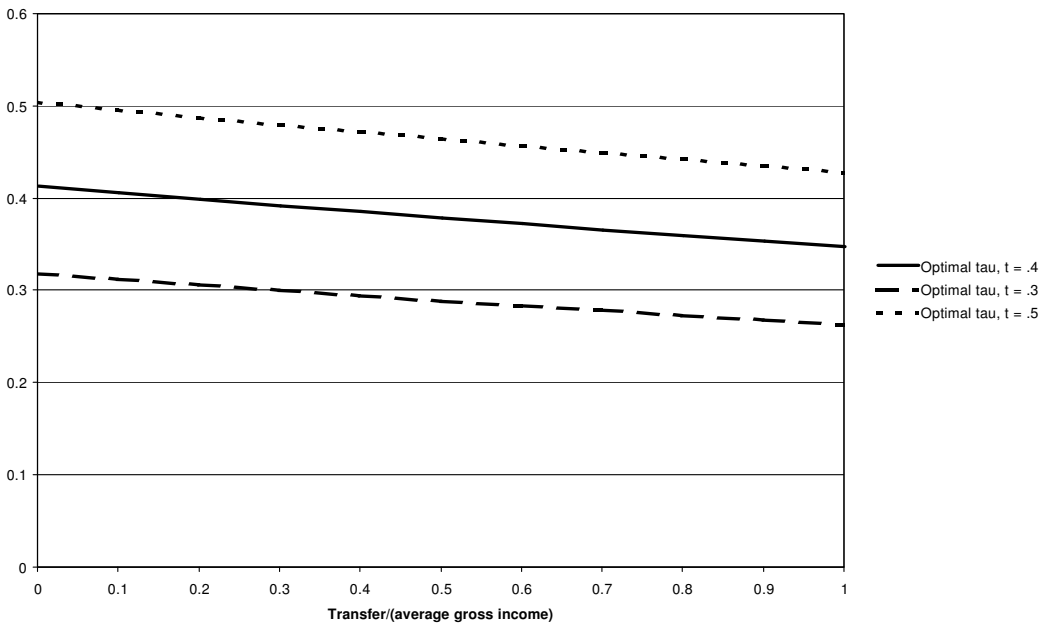
Key: Optimal capital income taxes at given optimal labor income taxes when the share of verifiable inputs in human capital formation α varies. Other parameters take baseline values: $r = 0.06$, cohort length is 25 years, $\sigma = 0.5$, $\beta = 0.5$, $\rho = 1$, $\omega = 0.33$, and $\overline{g/z} = 0.25$.

Figure 8: Optimal capital income taxes and the share of retirement consumption



Key: Optimal capital income taxes at given optimal labor income taxes when the share of retirement consumption ω varies. Other parameters take baseline values: $r = 0.06$, cohort length is 25 years, $\sigma = 0.5$, $\beta = 0.5$, $\rho = 1$, $\alpha = 0.5$, and $\overline{g/z} = 0.25$.

Figure 9: Optimal capital income taxes and the share of transfers



Key: Optimal capital income taxes at given optimal labor income taxes when the share of transfers relative to gross incomes $\overline{g/z}$ varies. Other parameters take baseline values: $r = 0.06$, cohort length is 25 years, $\sigma = 0.5$, $\beta = 0.5$, $\rho = 1$, $\alpha = 0.5$, and $\omega = 0.33$.