Redistribution and Education Subsidies are Siamese Twins

A. Lans Bovenberg

Bas Jacobs

[1] CentER, Tilburg University, Erasmus Universiteit Rotterdam, and CEPR.
[2] University of Amsterdam, CPB Netherlands Bureau for Economic Policy Analysis, and Tinbergen Institute.

Redistribution and Education Subsidies are Siamese Twins[*]

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

Working Paper – September, 2003

Abstract

We develop a model of human capital formation with endogenous labor supply and heterogeneous agents to explore the optimal level of education subsidies along with the optimal schedule of the labor income tax. If the government can observe inputs into the production of human capital, subsidies on education ensure efficiency in human capital accumulation and labor taxes are more progressive than without education policies. Although the able benefit more than proportionally from education subsidies, education subsidies thus play an important role in alleviating the tax distortions in human capital accumulation induced by redistributive policies. If the government can not verify all investments in human capital, education policy offsets some but not all the tax-induced distortions on learning.

Keywords: human capital, education subsidies, progressive taxation, redistribution.
JEL codes: H2, H5, I2, J2.

1 Introduction

Most OECD countries heavily subsidize higher education. These education subsidies are typically justified on the basis of perceived positive external effects of human capital accumulation, capital market imperfections, and redistributional concerns. Positive external effects of higher education, however, are difficult to establish empirically (see, *e.g.*, Heckman and Klenow, 1998; Acemoglu and Angrist, 2000; Krueger and Lindahl, 2002). Also capital market imperfections do not seem to be very important in practice (see, *e.g.*, Shea, 2000; Cameron and Taber, 2000; and Carneiro and Heckman, 2003).[1] Education subsidies, morover, typically transfer resources away from unskilled towards skilled, educated individuals so that their distributional effects appear perverse.[2] Why, then, is education subsidized?

We provide a second-best case for education subsidies on the basis of redistributional considerations rather than externalities and capital market imperfections. Although the able benefit more than proportionally from education subsidies, we show that education subsidies play an important role in alleviating the tax distortions in human capital accumulation induced by redistributive policies.[3] This explains why OECD countries subsidize higher education more heavily if the income tax is more progressive (see Figure 1).

Our paper explores the interaction between public spending and tax policies by viewing education subsidies and tax policies as interdependent instruments aimed at redistribution.[4] We add endogenous human capital formation

---

[1]In any case, loans rather than subsidies are the most direct way to address liquidity constraints (see Jacobs and van Wijnbergen (2002)).

[2]Dur and Teulings (2001), however, argue that education subsidies benefit unskilled individuals by boosting the supply of skilled workers compared to that of unskilled workers, thereby reducing wage differentials between skilled and unskilled workers. For realistic parameter values, they find that this general equilibrium effect broadly offsets the direct regressive impact of educational subsidies.

[3]Boadway *et al.* (1996) and Andersson and Konrad (2001) offer an alternative explanation for education subsidies. They argue that education subsidies are called for if the government cannot commit, and engages in excessive redistribution after investments in human capital have been made. We show that education subsidies are part of an optimal policy package – even if the government can commit to announced policies.

[4]The public finance literature on the effects of taxation on the incentives to accumulate human capital, in contrast, has typically abstracted from both public spending on education and distributional considerations, see *e.g.*, Boskin (1975), Heckman (1976), Kotlikoff and Summers (1979), Eaton and Rosen (1980), Trostel (1993), Nielsen and Sørensen (1997), Milesi-Feretti and Roubini (1998), and Judd (1999).

Figure 1: Marginal taxes and education subsidies in various OECD countries
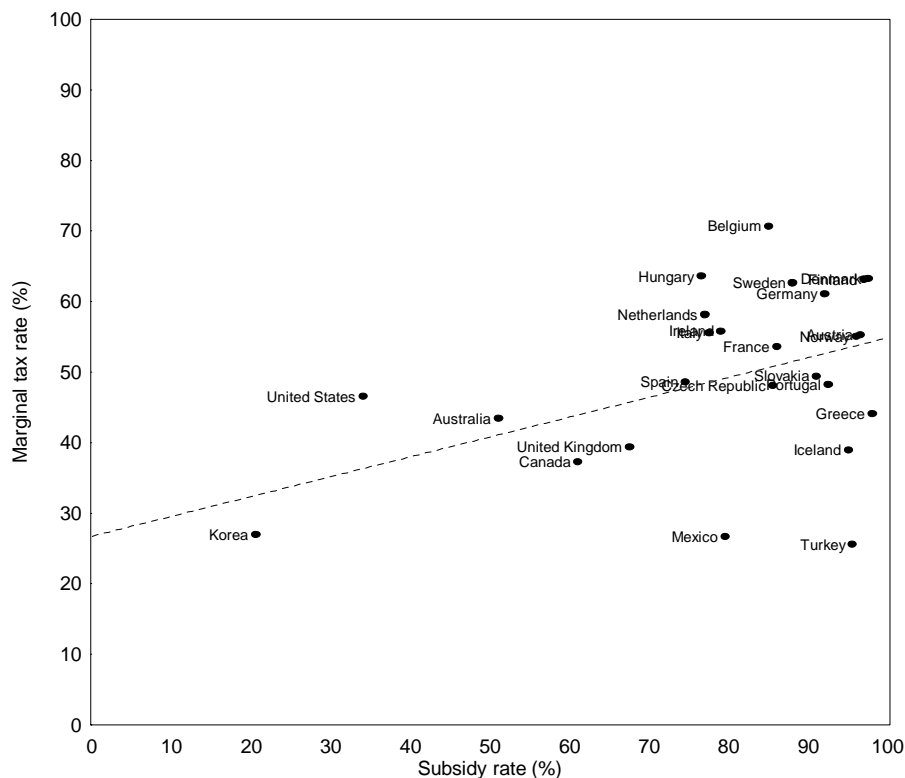
to the standard models of optimal income taxation and investigate how the availability of education subsidies affects the optimal income tax system. Our paper contributes in a number of ways to the optimal tax literature.

First, we derive the optimal income tax in the absence of education policies to show that optimal marginal income taxes are reduced below levels that would be optimal in the absence of endogenous human capital formation. Intuitively, the efficiency costs of redistribution increase, as positive marginal tax rates distort not only labor supply but also human capital accumulation.

Second, if the government has education subsidies at its disposal, the distortions of taxes on learning decisions are eliminated so as to restore efficiency in education choices. Indeed, education subsidies provide the government with sufficient instruments to restore efficiency in the production of human capital (see also Diamond and Mirrlees (1971)). By eliminating

3

the distortions of redistribution on learning, education subsidies make the optimal labor tax more progressive than without education subsidies.

Third, if the income tax is not available, the government would want to tax – rather than subsidize – education. Hence, the availability of an income tax as a redistributional device justifies education subsidies. Intuitively, the income tax is better targeted at income redistribution because all income is taxed – including the rents from ability. An education tax, in contrast, extracts only the returns from human capital formation, which exclude the rents from ability.

We explore also whether these three contributions are robust with respect to the observability of learning and individual incomes. If the government can not verify all investments in human capital, it offsets some but not all the tax-induced distortions on human capital accumulation. Intuitively, whereas education subsidies help to alleviate the adverse impact of taxes on the level of aggregate learning, they also distort the composition of the inputs in human capital formation away from non-verifiable inputs. Furthermore, we show that the main results derived with linear policy instruments also hold if the government has access to a non-linear income tax and non-linear education subsidies, which require the government to observe individual incomes and individual learning efforts.

Our paper extends earlier work by Ulph (1977) and Hare and Ulph (1979). Also these papers explore optimal policy if both the income tax and education expenditures are simultaneously optimized. They assume, however, that the government can directly set the level of education. With agents not deciding on human capital formation themselves, taxation does not distort education decisions. This contrasts with our paper, in which income taxes distort individual decisions on human capital accumulation.

Our paper is related also to Lommerud (1989), and van Ewijk and Tang (2000). In Lommerud (1989), the government taxes labor income in order to internalize the negative externalities from status seeking, but employs education subsidies to restore incentives to undertake education. In van Ewijk and Tang (2000), the government employs progressive taxes to punish wage demands of unions and to raise employment, but this discourages learning efforts. Education subsidies allow the government to set progressive labor taxes without distorting human capital accumulation. Both papers, however, abstract from distributional concerns as the main reason for imposing progressive income taxes.

The rest of this paper is structured as follows. Section 2 describes the

fundamental building blocks of our model: preferences and technologies. Section 3 explores optimal linear tax and linear education policies. Section 4 shows that the main results continue to hold if non-linear policy instruments would be available. Before section 6 concludes, section 5 discusses the policy implications of the analysis. The appendix contains the technical derivations of the results.

2   Preferences and technologies

Our model extends the standard model of optimal income taxation with endogenous earning potential through human capital formation. Individuals are heterogeneous with respect to their exogenous ability n. The cumulative distribution of ability is denoted by $F(n)$. $f(n)$ is the corresponding density function, which is continuously differentiable and strictly positive on the support $[\underline{n}, \overline{n}]$, $\underline{n}, \overline{n} > 0$.[5] The government knows the distribution of abilities, but cannot observe individual ability. Accordingly, it cannot levy individual-specific lump-sum taxes to redistribute incomes, but must rely on distortionary taxes instead.

Individuals invest an amount x_n of their resources on acquiring human capital. These resources are deductible from the income tax. p_x denotes the unit cost of x_n. We can interpret x_n as foregone labor time of an unskilled worker. With this interpretation, p_x represents the wage rate of an unskilled worker. Individuals also invest y_n units of non tax-deductible resources in education. These investments include books, computers and tuition. These costs cannot be deducted from the income tax because individuals typically earn no (or very low) labor incomes when they put these resources into education. The unit cost price of y_n is p_y.

Ability n is the productivity of inputs invested in education, so that more able individuals produce more human capital with the same inputs. The production function for human capital is homothetic in the two inputs x_n and y_n:

$$h_n = n\phi(x_n, y_n) \equiv n(\psi(x_n, y_n))^{\beta}, \tag{1}$$

where h_n denotes human capital, and $\psi(.)$ is a concave constant-returns-to-scale production function, with positive but diminishing returns to the inputs into education: $\psi_x, \psi_y > 0$, $\psi_{xx}, \psi_{yy} < 0$, $\psi_{xy} > 0$. Human capital

---

[5]The upper bound $\overline{n}$ may be infinite.

accumulation exhibits decreasing returns with respect to investment, *i.e.*, $\beta < 1$. Ability and educational investments are complementary inputs in generating human capital, *i.e.*, $\frac{\partial^2 h_n}{\partial x_n \partial n} = \beta(\psi(x_n, y_n))^{\beta-1}\psi_x \geq 0$, $\frac{\partial^2 h_n}{\partial y_n \partial n} = \beta(\psi(x_n, y_n))^{\beta-1}\psi_y \geq 0$. Gross labor income z_n is the product of the number of efficiency units of human capital, h_n, and hours worked l_n, *i.e.*, $z_n \equiv h_n l_n = n\phi(x_n, y_n)l_n$.[6]

Households feature a common, concave, and twice differentiable utility function, which is increasing in consumption c_n, and decreasing in work effort l_n

$$u(c_n, l_n), \tag{2}$$

with $u_c > 0$, $u_l < 0$, $u_{cc} < 0$, $u_{ll} < 0$.

3   Linear tax and education policies

This section derives the optimal linear income tax and education subsidies. Under a linear income tax, the government cannot observe individual gross labor incomes z_n, but only aggregate incomes $\int_{\underline{n}}^{\overline{n}} z_n dF(n)$. The linear income tax consists of an impersonal marginal tax rate, t, and a non-individualized lump-sum transfer or poll subsidy g. If y_n is verifiable, the government can subsidize these resources at rate s_y. In addition, x_n may be subsidized at rate s_x, where *after-subsidy* education expenses on x_n are deductible from the income tax.

---

[6]In this interpretation, the wage rate per efficiency unit of labor is normalized at unity. Alternatively, defining $\phi(x_n, y_n)$ (instead of $n\phi(x_n, y_n)$) as the efficiency units of labor (or human capital) per hour worked, we can interpret ability n as the wage rate per efficiency unit of labor.

3.1 Individual optimization

With this tax and subsidy code, the household budget constraint is[7]

$$c_n = (1-t)(n\phi(x_n, y_n)l_n - (1-s_x)p_x x_n) - (1-s_y)p_y y_n + g. \qquad (3)$$

Taking the policy instruments of the government as given,[8] individuals maximize utility by choosing c_n, l_n, x_n and y_n, subject to the household budget constraint. This yields the first-order conditions for the optimal choices of x_n and y_n

$$nl_n\phi_x = nl_n\beta(\psi(x_n, y_n))^{\beta-1}\psi_x = (1-s_x)p_x, \qquad (4)$$

$$(1-t)nl_n\phi_y = (1-t)nl_n\beta(\psi(x_n, y_n))^{\beta-1}\psi_y = (1-s_y)p_y. \qquad (5)$$

Marginal benefits of investing a unit of resources should equal marginal costs. The marginal tax rate t does not distort the x-margin, whereas it does distort the y-margin. Indeed, since x_n is tax deductible, the tax rate t equally affects the marginal benefits and the marginal costs of investing x_n. With the non-deductible input y_n, in contrast, taxes reduce only marginal benefits and do not affect costs, thereby distorting the optimal investment of y_n.

The first-order conditions for x_n (4) and y_n (5), and the production function of human capital (1) imply that gross labor income z_n is proportional to both x_n and y_n:

$$z_n = nl_n\phi(x_n, y_n) = \frac{(1-s_x)p_x}{\alpha\beta}x_n, \qquad (6)$$

$$z_n = nl_n\phi(x_n, y_n) = \frac{(1-s_y)p_y}{(1-\alpha)\beta(1-t)}y_n. \qquad (7)$$

---

[7]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 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, however, these additional income effects do not affect any of the main results derived below. For notational convenience, we therefore abstract from this additional term in the household budget constraint.

[8]Indeed, the government sets policy before agents determine their behavior. In view of its distributional preferences, the government faces an incentive to renege on its promises after the private sector has accumulated human and financial capital. We thus have to assume that the government has access to a commitment technology (*e.g.*, due to reputational considerations).

With linear policy instruments, the proportionality factors do not depend on ability n, because the shares of x_n and y_n in human capital investment, $\alpha \equiv \frac{x_n \psi_x}{\psi} = \frac{(1-t)(1-s_x)p_x x_n}{(1-t)(1-s_x)p_x x_n + (1-s_y)p_y y_n}$ and $1 - \alpha \equiv \frac{y_n \psi_y}{\psi} = \frac{(1-s_y)p_y y_n}{(1-t)(1-s_x)p_x x_n + (1-s_y)p_y y_n}$, are the same for all agents (see appendix).

The first-order condition for labor supply amounts to

$$\frac{-u_l}{u_c} = (1 - t)n\phi(x_n, y_n) = (1 - t)n(\psi(x_n, y_n))^\beta. \tag{8}$$

The second-order conditions for a maximum imply (see appendix)

$$\mu_n \equiv 1 - \beta(1 + \varepsilon_n) > 0, \tag{9}$$

where $\varepsilon_n \equiv \frac{\partial l_n^*}{\partial w_n} \frac{w_n}{l_n^*}$, denotes the compensated after-tax wage elasticity of labor supply, $w_n \equiv (1 - t)n\phi(x_n, y_n)$, and asterisks stand for compensated changes. The positive feedback between human capital and labor supply implies that decreasing returns in the production of human capital ($\beta < 1$) are not sufficient for the second-order condition to be met. In particular, more learning raises compensated labor supply, which in turn makes learning more attractive. This positive feedback effect, which depends on the compensated wage elasticity of labor supply ε_n, should be offset by sufficiently strong decreasing returns in learning to prevent corner solutions.

Totally differentiating the first-order conditions for x_n and y_n ((4) and (5)) and the compensated labor supply function $l_n^*(w_n)$ yields the following compensated elasticities of the endogenous variables with respect to the policy parameters (see appendix):

$$\varepsilon_{lt} \equiv \frac{\partial l_n^*}{\partial t} \frac{(1-t)}{l_n^*} = -\frac{\varepsilon_n(1 - \alpha\beta)}{\mu_n} \leq 0, \tag{10}$$

$$\varepsilon_{ls_x} \equiv \frac{\partial l_n^*}{\partial s_x} \frac{(1 - s_x)}{l_n^*} = \frac{\varepsilon_n \alpha\beta}{\mu_n} \geq 0, \tag{11}$$

$$\varepsilon_{ls_y} \equiv \frac{\partial l_n^*}{\partial s_y} \frac{(1 - s_y)}{l_n^*} = \frac{\varepsilon_n(1 - \alpha)\beta}{\mu_n} \geq 0, \tag{12}$$

$$\varepsilon_{xt} \equiv \frac{\partial x_n^*}{\partial t} \frac{(1-t)}{x_n^*} = -\frac{\varepsilon_n + (1 - \alpha)\beta(1 + \varepsilon_n)}{\mu_n} + (1 - \alpha)(\sigma - 1) \lessgtr 0, \tag{13}$$

$$\varepsilon_{xs_x} \equiv \frac{\partial x_n^*}{\partial s_x} \frac{(1 - s_x)}{x_n^*} = \frac{1 - (1 - \alpha)\beta(1 + \varepsilon_n)}{\mu_n} + (1 - \alpha)(\sigma - 1) \geq 0, \tag{14}$$

8

$$\varepsilon_{xs_y} \equiv \frac{\partial x_n^*}{\partial s_y} \frac{(1 - s_y)}{x_n^*} = \frac{(1 - \alpha)\beta(1 + \varepsilon_n)}{\mu_n} - (1 - \alpha)(\sigma - 1) \lessgtr 0, \qquad (15)$$

$$\varepsilon_{yt} \equiv \frac{\partial y_n^*}{\partial t} \frac{(1 - t)}{y_n^*} = -\frac{(1 - \alpha\beta)(1 + \varepsilon_n)}{\mu_n} - \alpha(\sigma - 1) \leq 0, \qquad (16)$$

$$\varepsilon_{ys_x} \equiv \frac{\partial y_n^*}{\partial s_x} \frac{(1 - s_x)}{y_n^*} = \frac{\alpha\beta(1 + \varepsilon_n)}{\mu_n} - \alpha(\sigma - 1) \lessgtr 0, \qquad (17)$$

$$\varepsilon_{ys_y} \equiv \frac{\partial y_n^*}{\partial s_y} \frac{(1 - s_y)}{y_n^*} = \frac{1 - \alpha\beta(1 + \varepsilon_n)}{\mu_n} + \alpha(\sigma - 1) \geq 0, \qquad (18)$$

where $\sigma \equiv -\frac{d(x_n/y_n)}{d(\psi_x/\psi_y)}$ denotes the elasticity of substitution between the two inputs in the production function for human capital.[9]

A larger marginal tax rate t affects private decisions through three channels: reducing labor supply (if $\varepsilon_n > 0$), depressing aggregate learning efforts (if $\alpha < 1$, so that some inputs are not tax deductible), and substituting tax-deductible for non tax-deductible resources in the production of human capital. The education subsidies s_x and s_y affect investment decisions in human capital at the aggregate level and through the composition of inputs into human capital formation. In some cases, policy effects cannot be signed because the impact on aggregate learning has the opposite sign of the corresponding effect on the composition of learning. To illustrate, depending on the substitution elasticity σ, the negative aggregate learning effect of a higher tax rate on x_n may offset a positive substitution effect, as tax deductible inputs are substituted for inputs that are not tax deductible.

3.2   Government

The government collects taxes to finance exogenously given expenditures, E, the education subsidies s_x and s_y, and the uniform lump-sum transfer g. The government budget constraint therefore reads as

$$t \int_{\underline{n}}^{\overline{n}} \left(n\phi(x_n, y_n)l_n - (1 - s_x)p_x x_n \right) dF(n) \qquad (19)$$

$$= \int_{\underline{n}}^{\overline{n}} \left(s_x p_x x_n + s_y p_y y_n + g + E \right) dF(n).$$

---

[9]The appendix shows that homotheticity of $\psi(x_n, y_n)$ implies that not only α but also σ is independent of ability.

We can write the indirect utility function as a function of the policy parameters: $v(g, t, s_x, s_y, n)$. The government maximizes a social welfare function defined over individuals' indirect utilities Γ

$$\Gamma = \int_{\underline{n}}^{\overline{n}} \Psi(v(g, t, s_x, s_y, n)) dF(n), \tag{20}$$

where $\Psi' > 0$, and $\Psi'' \leq 0$. With $\Psi' = 1$, the social welfare function is utilitarian (see Atkinson and Stiglitz (1980)).

The Lagrangian for maximizing social welfare is given by

$$\max_{\{g, t, s_x, s_y\}} \mathcal{L} = \int_{\underline{n}}^{\overline{n}} \Psi\left(v(g, t, s_x, s_y, n)\right) dF(n) \tag{21}$$

$$+ \lambda \int_{\underline{n}}^{\overline{n}} \left(t\left(n\phi(x_n, y_n)l_n - (1 - s_x)p_x x_n\right) - s_y p_y y_n - s_x p_x x_n - g - E\right) dF(n),$$

where λ is the Lagrange multiplier of the government budget constraint.

3.3 Optimal lump-sum transfer

The first-order condition for maximizing the Lagrangian (21) with respect to the lump-sum transfer g amounts to

$$\frac{\partial \mathcal{L}}{\partial g} = \int_{\underline{n}}^{\overline{n}} \left(\Psi' \eta_n - \lambda + \lambda t n \phi(.) \frac{\partial l_n}{\partial g} + \lambda \Delta_x \frac{\partial x_n}{\partial g} + \lambda \Delta_y \frac{\partial y_n}{\partial g}\right) dF(n) = 0, \tag{22}$$

where we used Roy's lemma, and η_n denotes the private marginal utility of income of individual with ability n. Δ_x and Δ_y represent the tax wedges on the two inputs into human capital accumulation; *i.e.*, $\Delta_x \equiv t\left(n\phi_x(.)l_n - (1 - s_x)p_x\right) - s_x p_x = -s_x p_x$ (from the first-order condition for x_n (4)) and $\Delta_y \equiv t n \phi_y(.)l_n - s_y p_y = \frac{(t - s_y)p_y}{1 - t}$ (from the first-order condition for y_n (5)).

By defining the net social marginal value of income of an n-man (including the effect on the tax base)

$$b_n \equiv \frac{\Psi' \eta_n}{\lambda} + t n \phi(x_n, y_n) \frac{\partial l_n}{\partial g} + \Delta_x \frac{\partial x_n}{\partial g} + \Delta_y \frac{\partial y_n}{\partial g}, \tag{23}$$

we can write the first-order condition for g (22) as

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

10

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.*, the right-hand side of (24)).

With the aid of the first-order condition for g (24), we define the so-called distributional characteristic ξ of labor income as

$$\xi \equiv \frac{\int_{\underline{n}}^{\overline{n}} (1 - b_n) z_n dF(n)}{\int_{\underline{n}}^{\overline{n}} z_n dF(n)}. \tag{25}$$

ξ is the negative normalized covariance between the welfare weight the government attaches to income of a particular ability b_n and gross labor income z_n (see, *e.g.*, Atkinson and Stiglitz (1980)). This covariance coincides with the normalized covariances of the welfare weights with both inputs in learning (x_n and y_n) and the base of the income tax ($z_n - (1 - s_x) p_x x_n$), because labor income z_n is proportional to investments in education x_n and y_n (see (6) and (7)). A positive distributional characteristic ξ thus implies that the income tax base 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.[10] Indeed, a zero distributional characteristic implies either that the government is not interested in redistribution (so that the welfare weight is the same for all n) or that the marginal contribution to the tax base is equal for all ability types (taxable income is the same for all n).

3.4 Optimal income taxation

This section derives the optimal linear income tax in the absence of education policy ($s_x = s_y = \Delta_x = 0$ and $\Delta_y = \frac{t p_y}{1-t}$). The optimal linear income tax follows from taking the first-order condition for maximizing the Lagrangian

---

[10]The strength of this negative correlation depends both on the concavity of the function Ψ and on inequality in utility levels. In particular, the government attaches a higher priority to redistributing incomes if more unequal utility levels produce larger differences in marginal utilities of income η_n and thus b_n.

(21) with respect to t and substituting the distributional characteristic (25):

$$\frac{t}{1-t} = \frac{\xi(1-\alpha\beta)}{-\bar{\varepsilon}_{lt} - (1-\alpha)\beta\bar{\varepsilon}_{yt}}. \tag{26}$$

Here, and in the rest of the paper, upper bars denote income-weighted elasticities (*i.e.*, $\bar{\epsilon} \equiv \frac{\int_{\underline{n}}^{\bar{n}} \epsilon z_n dF(n)}{\int_{\underline{n}}^{\bar{n}} z_n dF(n)}$, where ϵ represents the elasticity concerned). This optimal tax formula illustrates the fundamental trade-off between equity and efficiency. In particular, the tax rate rises if redistributional concerns become more pressing, as indicated by a larger distributional characteristic ξ. The denominator of (26) captures the distortionary costs of redistributive taxation in terms of the effective elasticity of the tax base. The first income-weighted elasticity $\bar{\varepsilon}_{lt}$ is associated with the distortionary effect of taxes on labor supply. The second term in the denominator $(1-\alpha)\beta\bar{\varepsilon}_{yt}$ represents the human-capital distortion associated with a higher tax rate on account of the non-deductibility of education expenditures. The optimal marginal tax should be lower if larger compensated elasticities (in absolute value) indicate that redistributive taxes distort labor supply and human capital accumulation more substantially. These latter effects conform to standard Ramsey intuition.

We can substitute (10) and (16) to eliminate ε_{lt} and ε_{yt} in (26) to find

$$\frac{t}{1-t} = \frac{\xi}{\left(\frac{\bar{\varepsilon}_n}{\mu_n}\right) + (1-\alpha)\beta\left(\frac{(1+\varepsilon_n)}{\mu_n}\right) + \frac{(1-\alpha)\beta}{1-\alpha\beta}\alpha(\sigma-1)}. \tag{27}$$

In the absence of a learning decision ($\beta = 0$), the effective elasticity of the tax base represented by the denominator at the right-hand side of (27) is simply the income-weighted compensated simple elasticity of labor supply $\bar{\varepsilon}_n$. Hence, our model nests the standard optimal linear income tax model (see Atkinson and Stiglitz (1980)). Even if all investments in human capital are tax deductible ($\alpha = 1$), the efficiency costs of taxation are higher than in the standard model without human capital accumulation ($\beta = 0$) because of the feedback effects of human capital accumulation on labor supply (*i.e.*, $\frac{\varepsilon_n}{\mu_n} = \frac{\varepsilon_n}{1-\beta(1+\varepsilon_n)} > \varepsilon_n$); lower labor supply depresses learning, which harms the after-tax wage rate and thus further reduces labor supply. As a direct consequence, *ceteris paribus* the distributional characteristic ξ and the simple compensated labor supply elasticities ε_n, endogenous learning reduces the optimal tax rate t. Non tax-deductibility of some human capital investments

12

(*i.e.*, $\alpha < 1$) further decreases the optimal tax rate (at given ξ and ε_n), especially if both the aggregate level and the composition of learning are sensitive to the tax rate (*i.e.*, both β and σ are large).

3.5 Optimal education policies

This sub-section investigates optimal education policy if the government cannot simultaneously optimize its tax policy.

3.5.1 Optimal education subsidies

If the government can employ both s_x and s_y to optimize welfare, then the optimal linear education subsidies are found by manipulating the first-order conditions for maximizing the Lagrangian (21) with respect to s_x and s_y. This yields the result that all inputs in learning should be equally taxed or subsidized (see appendix):

$$\frac{-s_x}{1 - s_x} = \frac{t - s_y}{1 - s_y}, \tag{28}$$

so that the composition of human capital investment is not distorted.

The optimal levels of the education subsidies are given by (see appendix)

$$\frac{-s_x}{(1 - s_x)(1 - t)} = \frac{t - s_y}{(1 - s_y)(1 - t)} = \frac{1}{\left(\overline{1/\mu}\right)} \left[\xi - \frac{t}{1 - t} \left(\overline{\frac{\varepsilon_n}{\mu_n}} \right) \right], \tag{29}$$

where $\overline{1/\mu}$ represents the income-weighted elasticity $1/\mu_n$, which captures the strength of the feedback effects between labor supply and human-capital formation (recall that $\mu_n \equiv 1 - \beta(1 + \varepsilon_n)$).

The term in square brackets at the far right-hand side of (29) measures the sub-optimality of the income tax (see expression (33) below). If this term is positive, income taxes are sub-optimally low and thus do not redistribute sufficient resources. Accordingly, the government employs also education taxes to redistribute from the rich, who learn more, to the poor. Education taxes are thus positive (i.e. $-s_x > 0$ and $t - s_y > 0$) if the government cannot tax labor income at all ($t = 0$). These education taxes are relatively large if the effective elasticity of the subsidy base is small; *i.e.*, if the feedback effects between labor supply and learning are only weak and $\overline{1/\mu}$ is small (*ceteris paribus* the distributional characteristic ξ).

13

3.5.2 Optimal education subsidy on x_n only

If y_n is not verifiable, the government cannot employ subsidies on y_n so that $s_y = 0$. In that case, the optimal linear education subsidy on x_n is given by (see appendix)

$$\frac{-s_x}{(1 - s_x)(1 - t)} = \frac{\left(\xi - \frac{t}{1-t}\left[\left(\frac{\overline{\varepsilon_n + 1 - \alpha}}{\mu_n}\right) - (1 - \alpha)\sigma\right]\right)}{\alpha\left(\overline{1/\mu}\right) + (1 - \alpha)\sigma}, \qquad (30)$$

Without any instruments to tax income ($t = 0$), the government employs taxes on education ($s_x < 0$) to reduce inequality. With a Leontief production function for human capital ($\sigma = 0$), the government sets the same net tax on education as in the case in which the government has access to s_y as well. To achieve this allocation, it raises the tax on x_n above the corresponding tax $-s_x$ with verifiable y_n (compare (30) (and $\sigma = t = 0$) with the optimal s_x if both subsidy instruments are used (29)). With non-zero substitution between x_n and y_n ($\sigma > 0$), a tax on the verifiable input in learning distorts the composition of human-capital investments. Hence, the government can no longer costlessly tax non-verifiable investments y_n in an indirect way by levying a heavier tax on the verifiable component x_n. By doing so, the government imposes a smaller tax burden on education compared to the case in which both inputs into education are verifiable. Indeed, *ceteris paribus* the distributional characteristic ξ, more substitution between the two inputs into human capital accumulation reduces the optimal tax on learning, as such a tax distorts not only the aggregate level of human-capital accumulation (and thus labor supply) but also the composition of learning (see (30) with $t = 0$).

With a Leontief production function of human capital, a positive income tax $t > 0$, *ceteris paribus* reduces the optimal education tax. Intuitively, a positive income tax takes over the role of the education tax in redistributing income. With substantial substitution in human-capital accumulation (*i.e.*, σ is large), in contrast, the optimal education tax remains large in the presence of a positive income tax. The reason is that the education tax helps to offset the distortionary impact of income tax deductibility of x_n on the composition of learning. Indeed, if the substitution elasticity σ approaches infinity, the education tax is set so that it implies the same tax on x_n as the income tax imposes on the non-tax deductible input y_n (*i.e.*, $\frac{-s_x}{(1-s_x)(1-t)} = \frac{t}{1-t}$, which corresponds to (28) with $s_y = 0$). In this case, the government gives priority

14

to offsetting the distortionary impact of the income tax on the composition rather than the aggregate level of learning.

3.6 Optimal income taxation and education subsidies

If the government can simultaneously employ all its policy instruments, the optimal subsidies on education are (see appendix)

$$s_x = 0, \tag{31}$$

$$s_y = t. \tag{32}$$

These expressions imply that the tax wedges on both inputs into human capital accumulation ($\Delta_x = -s_x p_x$ and $\Delta_y = \frac{(t-s_y)p_y}{1-t}$) are zero. Optimal education policy thus eliminates learning distortions on account of redistributive taxation. The optimal subsidy s_y ensures that non-deductible education expenses effectively become deductible against the labor income tax rate.

In the presence of this optimal education policy, the optimal linear tax rate is given by (see appendix)

$$\frac{t}{1-t} = \frac{\xi}{\left(\frac{\varepsilon_n}{\mu_n}\right)}. \tag{33}$$

Comparing the optimal income tax with optimal education policy (33) to the optimal income tax without any education subsidies (27), we observe that the additional instrument of the education subsidy s_y (*ceteris paribus* the distributional characteristic ξ) allows for a higher income tax.[11] The reason is that the education subsidy eliminates the tax distortions on the aggregate level and composition of human capital accumulation. With a lower effective elasticity of the tax base implying smaller efficiency costs of redistributive income taxation, the government redistributes income more aggressively.

The deeper intuition for this result is as follows. The tax on labor income falls on income derived from human-capital formation, labor supply

---

[11]The optimal tax rate is still lower than without endogenous human capital accumulation ($\beta = 0$). The reason is that the positive feedback effects between labor supply and learning raise the effective labor-supply elasticity. This increases the deadweight loss associated with the labor-supply distortion, which (in contrast to the human capital distortion) is not affected by the education subsidy.

and ability. With the optimal education subsidy (32) compensating individuals for the taxes imposed on the returns from learning, the labor income tax falls only on additional income from labor supply and the rents from ability. Hence, the education subsidy gives the government a more effective instrument to tax away the infra-marginal rents from ability without distorting the incentives to learn. Indeed, the optimal tax rate would approach 100% if labor supply was inelastic ($\varepsilon_n = 0$). The income tax cum education subsidy then effectively becomes Tinbergen's talent tax (*i.e.*, a tax on rents from ability only).

Education subsidies increase if distributional concerns become more pressing, as indicated by a larger distributional characteristic ξ (substitute (33) into (32) to eliminate t). This result may seem paradoxical because the more able individuals learn more, and thus benefit most from the education subsidies. However, this benefit for the most-skilled agents is more than offset by higher labor taxes imposing a heavier tax burden on the rents from ability. Clearly then, education subsidies and progressive taxation are Siamese twins. Indeed, both optimal education subsidies and marginal taxes are zero in the absence of redistributional considerations ($\xi = 0$).

3.7 Optimal income taxation and imperfect education subsidies

If the government cannot verify y_n, it must resort to subsidies on x_n to offset the learning distortions imposed by the labor tax. The optimal linear taxes and education subsidies are obtained by manipulating the first-order conditions for maximizing the Lagrangian (21) with respect to t and s_x. The optimal combination of subsidies and taxes satisfies (see appendix)

$$\frac{s_x}{1-s_x} = \frac{(1-\sigma)(1-\alpha)}{1-(1-\alpha)(1-\sigma)}t. \tag{34}$$

An education subsidy ($s_x > 0$) alleviates the distortionary effect of the labor tax on aggregate learning. At the same time, however, it exacerbates the distortions of the tax system on the composition of learning. In particular, the 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 causes 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. A large substitution elasticity σ 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 inputs. In contrast, if individuals cannot easily substitute y_n for x_n (*i.e.*, σ is small), x_n is subsidized rather than taxed because subsidizing the observed input x_n is a good substitute for subsidizing the non-observed input y_n, which the government would ideally like to subsidize (compare (32)).

The optimal linear income tax amounts to

$$\frac{t}{1-t} = \frac{\xi}{\left(\frac{\varepsilon_n}{\mu_n}\right) + \left(\frac{\beta(1+\varepsilon_n)}{\mu_n}\right)\frac{\sigma(1-\alpha)}{\alpha+(1-\alpha)\sigma}}. \tag{35}$$

With a Leontief production of human capital ($\sigma = 0$), the optimal income tax is the same as in the case in which all policy instruments are available (see (33)). Without substitution between x_n and y_n, the government can costlessly subsidize non-verifiable inputs y_n in an indirect way by subsidizing verifiable inputs x_n. With non-zero substitution between the two inputs ($\sigma > 0$), 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 the excessive use of x_n. With an imperfect instrument to offset the distortionary impact of the tax system on y_n, the optimal linear income tax is lower compared to the optimal income tax with direct subsidies on y_n (compare (35) and (33)).

With infinite substitution ($\sigma \to \infty$), the optimal income tax with imperfect education subsidies equals the optimal income tax without any subsidies and tax deductible inputs (so that (35) coincides with (27) with $\alpha = 0$). Indeed, in that case, the government finds it optimal to make x_n effectively not tax deductible, so as ensure a level playing field with y_n. The Cobb Douglas case ($\sigma = 1$) is the only case in which the availability of s_x does not raise the optimal tax rate (compare the optimal tax with s_x (35) and the optimal tax at zero subsidies (27)). In that case, education subsidies cannot do anything to offset the distortions of the income tax on the composition and level of learning because the benefits of an educational subsidy in terms of more learning are exactly offset by the costs of such a subsidy in terms of a more distorted composition of learning (excessive use of x_n). Both with easy substitution ($\sigma > 1$) and with difficult substitution ($\sigma < 1$), the availability of s_x allows the government to combat the human-capital distortions

imposed by the labor income tax. With easy substitution, learning is taxed to alleviate the relatively large tax distortions on the composition of human capital. If substitution is difficult, in contrast, learning is subsidized, as the government gives priority to fighting tax distortions on the level rather than the composition of learning.

4 Non-linear tax and education policies

This section explores the robustness of our results by analyzing non-linear policy instruments. In particular, the government can levy a non-linear income tax $T(z_n - (p_x x_n - S_x(p_x x_n)))$ on gross incomes z_n minus net-of-subsidy educational spending on x_n. The marginal income tax rate $T' \equiv dT(z_n - p_x x_n + S_x(p_x x_n))/d(z_n - p_x x_n + S_x(p_x x_n))$ can now differ across individuals with different incomes. Furthermore, the government may employ a non-linear subsidy on resources x_n invested in education. The subsidy is denoted as $S_x(p_x x_n)$, where $S'_x \equiv dS_x(p_x x_n)/dx_n$ represents the marginal subsidy rate on x_n. If non-deductible individual investments $p_y y_n$ are verifiable, the government can subsidize also y_n through a non-linear subsidy $S_y(p_y y_n)$, with $S'_y \equiv dS_y(p_y y_n)/dy_n$ denoting the marginal subsidy on y_n. The non-linear tax and educational subsidies imply stronger information requirements than the linear policy instruments explored in the previous section. In particular, the government must be able to verify gross incomes and inputs into learning of each individual rather than only the economy-wide aggregates.

4.1 Individual optimization

With these non-linear policy instruments, the individual's budget constraint reads as

$$c_n = z_n - p_x x_n + S_x(p_x x_n) - T(z_n - p_x x_n + S_x(p_x x_n)) - p_y y_n + S_y(p_y y_n), \quad (36)$$

where $z_n \equiv n\phi(x_n, y_n)l_n$.

Maximization of utility $u(c_n, l_n)$ subject to the household budget constraint yields

$$\frac{-u_l}{u_c} = (1 - T')n\phi(x_n, y_n), \quad (37)$$

in addition to the first-order conditions for resources invested in education[12]

$$nl_n\phi_x = (1 - S'_x)p_x, \tag{38}$$

$$(1 - T')nl_n\phi_y = (1 - S'_y)p_y. \tag{39}$$

4.2 Government

The government observes $z_n - p_x x_n + S_x(p_x x_n)$ (and x_n and y_n if it subsidizes education) and thereby controls $z_n - p_n x_n$ (and educational investments x_n and y_n if it subsidizes these investments) through the tax (and subsidy) schedule(s). In directly controlling the allocation, the government should respect the incentive-compatibility constraints, which require that each individual n prefers the bundle c_n, z_n, x_n, y_n intended for him over the bundles intended for others

$$U(c_n, z_n, x_n, y_n, n) \geq U(c_m, z_m, x_m, y_m, n), \quad \forall m \in [\underline{n}, \overline{n}], \forall n \in [\underline{n}, \overline{n}], \tag{40}$$

where $U(c_n, z_n, x_n, y_n, n) \equiv u\left(c_n, \frac{z_n}{n\phi(x_n, y_n)}\right) = u(c_n, l_n)$. These global incentive-compatibility constraints can be replaced by the (first-order) incentive-compatibility constraint using the Envelope Theorem (see, *e.g.*, Mirrlees (1971)):[13]

$$\frac{du_n}{dn} = -u_l(c_n, l_n)\frac{l_n}{n}. \tag{41}$$

Utilizing Walras' law, we employ the economy's resource constraint[14]

$$\int_{\underline{n}}^{\overline{n}} (n\phi(x_n, y_n)l_n - p_x x_n - p_y y_n - c_n - E)\, dF(n) = 0 \tag{42}$$

rather than the government budget constraint.

---

[12]Second-order conditions for an interior solution are assumed to be satisfied. In particular, the restriction $\mu_n \equiv 1 - \beta(1 + \varepsilon_n) > 0$ must hold.

[13]In adopting the first-order approach, we assume that the second-order conditions for the second-best allocation are met. This requires single crossing of the utility function with respect to all control variables in addition to all control variables being non-decreasing in ability at the optimum schedules. For an analysis of second-order conditions in optimal non-linear taxation problems, see Mirrlees (1986) and Ebert (1992).

[14]If the government implements an allocation that meets the individuals' budget constraints and the economy's resource constraint, the government budget constraint must be satisfied as well by Walras' law.

We solve for the optimal allocation by applying the maximum principle and setting up a Hamiltonian $\mathcal{H}$, with l_n (and, depending on the availability of non-linear education subsidies, x_n and y_n) as control variable(s), u_n as state variable, and a costate variable θ_n for the incentive-compatibility constraint (41):

$$\max_{\{l,x,y,u\}} \mathcal{H} = \Psi(u_n)f(n) + \theta_n u_l(c_n, l_n)\frac{l_n}{n} \tag{43}$$
$$+\lambda \left(n\phi(x_n, y_n)l_n - p_x x_n - p_y y_n - c_n\right)f(n),$$

where λ is the shadow value of the resource constraint. The transversality conditions are given by

$$\lim_{n\to\overline{n}} \theta_n = 0, \quad \lim_{n\to\underline{n}} \theta_n = 0. \tag{44}$$

Having found the second-best allocation, we determine the optimal marginal taxes T' and subsidy rates S'_x and S'_y by employing the first-order condition(s) for individual optimization with respect to l_n, x_n, and y_n (see (37), (38), and (39), respectively). Optimal consumption follows from the optimal levels of l_n, y_n, x_n, and the household budget constraint (36).

4.3 Optimal income taxation

In analogy of our discussion of linear policy instruments, we first explore the optimal income tax in the absence of education policy. By taking the first-order condition of the Hamiltonian (43) with respect to l_n, we derive the optimal marginal tax rates along the non-linear tax schedule (see appendix)

$$\frac{T'}{1-T'} = \frac{u_c\theta_n/\lambda}{nf(n)}\left(1+\frac{1}{\varepsilon_n^*}\right)\left[\frac{\frac{\varepsilon_n}{\mu_n}}{\frac{\varepsilon_n}{\mu_n} + \frac{\beta(1+\varepsilon_n)(1-\alpha_n)}{\mu_n} + \frac{(1-\alpha_n)\beta}{(1-\alpha_n\beta)}\alpha_n(\sigma_n-1)}\right], \tag{45}$$

where $\varepsilon_n^* \equiv \left(\frac{lu_{ll}}{u_l} - \frac{lu_{cl}}{u_c}\right)^{-1}$. With non-linear instruments both α_n and σ_n are dependent on n, hence we use subscripts. The denominator in the term in square brackets in (45) equals the effective elasticity of the tax base in the corresponding expression of the linear income tax (27) (except that average elasticities feature in the expression of the optimal linear tax rate). The non-linear counterpart of ξ is the term $u_c\theta_n/\lambda$, which measures the benefits of redistribution. In particular, θ_n denotes the welfare gain of redistributing

20

income from individuals above ability n to individuals below n, and depends both on the concavity of the social welfare function and the marginal utility of income.[15] Furthermore, a larger number of individuals at ability n (a larger density $f(n)$) and a higher ability level n exacerbate the distortionary costs of high marginal rates at that ability level.

In the general case with partially deductible costs ($0 \leq \alpha_n < 1$ and $\beta > 0$) the optimal tax rates are reduced compared to the case with exogenous learning (*ceteris paribus* ε_n^*, ε_n, and $u_c\theta_n/\lambda$). The reason is that the tax system distorts not only labor supply, but also the level of learning because not all educational costs are deductible. The additional learning distortions yield lower optimal tax rates.

With substitution between x_n and y_n (if $\sigma_n > 0$), the income tax distorts not only labor supply and the level of learning, but also the composition of learning by favoring deductible inputs x_n over non tax-deductible inputs y_n. Hence, more substitution *ceteris paribus* depresses optimal tax rates. In fact, with easy substitution between deductible and non-deductible investments (*i.e.*, σ_n is large), the term in square brackets may be smaller compared to the case in which none of the costs of investment are deductible ($\alpha_n = 0$). Indeed, easy substitution implies that welfare losses due to a distorted composition of learning dominate those on account of a distorted level of learning. Hence, making some, but not all, investment costs deductible may harm welfare. With a Cobb Douglas learning technology ($\sigma_n = 1$), however, both effects cancel.

4.4   Optimal education policies

We now turn to the case in which the government has no access to the non-linear income tax and can employ only education subsidies. By taking the first-order conditions of the Hamiltonian (43) with respect to x_n and y_n, we derive the optimal marginal subsidy rates (see appendix)

$$\frac{-S'_x}{1 - S'_x} = \frac{-S'_y}{1 - S'_y} = \frac{u_c\theta_n/\lambda}{nf(n)}\left(1 + \frac{1}{\varepsilon_n^*}\right)\varepsilon_n. \qquad (46)$$

As in the linear case, the two education instruments are thus equal (compare (46) to the corresponding optimal linear education taxes (29) with $t = 0$),

---

[15]From the transversality conditions follows that optimal marginal taxes are zero at the top and the bottom of the ability distribution, since $\theta_{\underline{n}} = \theta_{\overline{n}} = 0$.

which avoids distortions in the composition of human capital investments. Moreover, in the absence of the income tax, education is taxed in order to redistribute incomes away from the able agents who learn most.

4.5   Optimal income taxation and education subsidies

If the government simultaneously sets non-linear tax and education subsidy schedules, it determines the optimal allocation by using x_n, y_n, and l_n as control variables. By manipulating the first-order condition of the Hamiltonian (43) with respect to these three controls, we derive optimal marginal tax rates and subsidy rates along the non-linear tax and subsidy schedules. From the first-order conditions for x_n and y_n, we establish production efficiency in learning:

$$n\phi_x l_n = p_x, \tag{47}$$

$$n\phi_y l_n = p_y. \tag{48}$$

Combining (47) with (38), and (48) with (39), we find

$$S'_x(p_x x_n) = 0, \tag{49}$$

$$S'_y(p_y y_n) = T'(z_n - p_x x_n + S(p_x x_n)). \tag{50}$$

Optimal education policy requires that all costs of education should effectively be tax deductible, just as in the linear case (compare (31) and (32)).

Optimal income tax rates along the non-linear tax schedule are (see appendix)

$$\frac{T'}{1 - T'} = \frac{u_c \theta_n / \lambda}{n f(n)} \left(1 + \frac{1}{\varepsilon_n^*}\right). \tag{51}$$

We thus arrive at the same conclusion as with linear taxes: education subsidies allow for higher marginal tax rates by eliminating the distortionary impact of the labor income tax on the aggregate level and composition of human capital accumulation. In particular, comparing the optimal income tax rates with optimal education policy (51) to the optimal income tax without any education subsidies (45), we observe that the additional instrument of the education subsidy s_y (*ceteris paribus* ε_n^*, ε_n, and $u_c \theta_n / \lambda$) allows for higher marginal taxes on labor income. Moreover, stronger redistributive motives, as reflected in higher values for $u_c \theta_n / \lambda$, result in higher education subsidies, even though the rich benefit more from education subsidies (substitute (51) into (50) to eliminate T'). Indeed, whereas the tax system is targeted at

22

redistribution, education subsidies offset the adverse impact of taxes on the incentives to accumulate human capital. Education subsidies and progressive taxation are therefore Siamese twins.

4.6 Optimal income taxation and imperfect education subsidies

If the government cannot verify y_n, it can employ only the non-linear tax system and subsidies on the deductible input x_n to optimize social welfare. By taking the first-order condition of the Hamiltonian (43) with respect to l_n and x_n we derive the optimal marginal tax and subsidy rates. These instruments are related as follows (see appendix):

$$\frac{S'_x}{1 - S'_x} = \frac{(1 - \alpha_n)(1 - \sigma_n)}{1 - (1 - \alpha_n)(1 - \sigma_n)} T', \tag{52}$$

which coincides with the corresponding expression for the linear income tax (see (34) and the explanation given there).

The optimal marginal income tax rates are given by

$$\frac{T'}{1 - T'} = \frac{u_c \theta_n / \lambda}{n f(n)} \left(1 + \frac{1}{\varepsilon^*}\right) \left[\frac{\frac{\varepsilon_n}{\mu_n}}{\frac{\varepsilon_n}{\mu_n} + \frac{\beta(1 + \varepsilon_n)}{\mu_n} \frac{\sigma_n(1 - \alpha_n)}{\alpha_n + (1 - \alpha_n)\sigma_n}} \right]. \tag{53}$$

This formula again resembles the corresponding formula with linear policy instruments (35), with the denominator of the term in squared brackets coinciding with the effective elasticity of the tax base in (35) (except that average elasticities feature there). Optimal marginal tax rates with imperfect subsidies exceed the corresponding marginal tax rates without education subsidies (except if $\sigma_n = 1$, see (45)), but are below the marginal tax rates if all three non-linear policy instruments are available (see (51)) (*ceteris paribus* ε_n^*, ε_n, and $u_c \theta_n / \lambda$). The substitution elasticity σ_n between verifiable and non-verifiable inputs into human capital accumulation determines whether the optimal marginal tax rates are closer to the case with both or without any education subsidies. In particular, without much substitution, a subsidy on verifiable inputs mimics a subsidy on non-verifiable inputs. Therefore, tax distortions on human capital accumulation can be reduced without creating substantial distortions in the composition of human capital investments. Hence, the government can set relatively high tax rates. With easy substitution, in contrast, the government should not use education subsidies to

alleviate tax-induced learning distortions because these education subsidies substantially distort the composition of learning.

5   Policy implications

Our analysis has a number of important policy implications. In particular, if the government can observe educational inputs, it should optimally subsidize education to make all costs of education effectively tax deductible. As students typically do not earn high enough incomes to expense their educational costs from the income tax, governments should subsidize the costs of (higher) education at the marginal tax rate at which the marginal returns of that education are taxed.

To explore whether the current levels of education subsidies in several OECD countries are efficient, we calculate optimal education subsidies and compare them with actual subsidies. We focus on subsidies to higher education because compulsory schooling laws ensure that progressive taxes do not reduce participation in basic education. As a measure for t (or T'), we employ marginal tax rates facing a worker earning 133% of the average production wage (see OECD (2002)). Public subsidy rates on the direct educational costs for tertiary education (see OECD 2003)) are used as a measure of s (or S'). All figures apply to 2000.

Table 1 contains the actual subsidies (as a percentage of direct costs of education) and marginal income taxes. Optimal education subsidies require that the costs of education be made effectively tax deductible from the income tax, so that the subsidy rate equals the marginal tax rate. Although actual subsidies are quite high, a substantial part of education subsidies can be justified by appealing to tax distortions on human capital formation. Indeed, one does not need to rely on capital market imperfections or externalities to argue in favor of substantial education subsidies.

Our analysis suggests also that education should be subsidized more heavily if countries feature more progressive tax systems. Figure 1 shows that this is indeed a stylized fact. As a corollary, countries reducing marginal taxes should cut education subsidies. As many countries have indeed cut marginal taxes in recent years, education subsidies can be reduced as well. Indeed, several countries have done exactly that.

The analysis sheds light on optimal education policies also if educational efforts cannot be adequately monitored. One interpretation of x_n is time

Table 1: Marginal tax rates and education subsidies (in %)

Country	Marginal tax rate	Marginal subsidy rate
Korea	27.0	23.3
United States	46.6	33.9
Japan	28.9	44.9
Australia	43.5	51.0
Canada	37.3	61.0
United Kingdom	39.4	67.7
Spain	48.6	74.4
Hungary	63.6	76.7
Netherlands	58.2	77.4
Italy	55.6	77.5
Ireland	55.8	79.2
Mexico	26.7	79.4
Belgium	70.7	85.2
Czech Republic	48.1	85.5
France	53.6	85.7
Sweden	62.7	88.1
Slovakia	49.4	91.2
Germany	61.1	91.8
Portugal	48.3	92.5
Iceland	39.0	94.9
Turkey	25.6	95.4
Norway	55.1	96.2
Austria	55.3	96.7
Finland	63.2	97.2
Denmark	63.3	97.6
Greece	44.1	99.7

Note: Marginal taxes apply to a worker earning 133% of the average production wage, including social security contributions and local taxes, see OECD (2002). The marginal subsidy rate applies to the total direct costs of tertiary educational institutes, see OECD (2003, table B3.2)

invested in education, while y_n can be interpreted as goods invested in education. Yet, the analysis allows for two alternative interpretations. First, x_n and y_n can be interpreted as basic and higher education, respectively. Since substitution between higher and basic education is very limited, the government may restore the correct incentives to accumulate human capital with substantial subsidies to basic education if it cannot verify students' efforts and resources invested in higher education. Indeed, this case corresponds closely to a Leontief production function of human capital.

As another alternative interpretation, x_n and y_n can be viewed as, respectively, formal and informal (parental) inputs into education. In this case formal, public inputs can be good substitutes for informal, private inputs, as argued by Becker and Tomes (1979). With parental inputs in human capital formation being very hard to observe, subsidizing formal inputs (schools, books, etc) may not be efficient because it merely crowds out informal parental efforts to educate their children.

6   Conclusions

This paper has studied the simultaneous setting of optimal redistributive income taxes and education subsidies. We showed that education subsidies can be a powerful instrument to eliminate distortions in the accumulation of human capital caused by redistributive policies. Moreover, optimal marginal income tax rates can be higher if the government employs education subsidies to reduce the distortions associated with the redistribution of incomes. Education subsidies and redistribution of incomes are thus like Siamese twins – although the ones who benefit from the education subsidies also enjoy relatively high lifetime incomes. The more the government desires to help the poor, the more it should employ education subsidies to offset the learning distortions associated with redistribution. If not all educational efforts can be perfectly observed, however, the government cannot completely eliminate tax-distortions on human capital formation.

In future research, we intend to explore the consequences of non-pecuniary returns of education. If education is a consumption good and substantial returns on education are tacit, the tax wedge on learning is smaller because the returns escape the income tax. This weakens the case for education subsidies to correct for tax distortions (see also Alstadsaeter (2003)). We also would like to extend the current analysis to a more dynamic setting. In

such a dynamic analysis, progressive tax schedules with rising marginal tax rates harm investment in human capital even if all costs of education are tax deductible. The reason is that the returns on learning are taxed at higher marginal rates than the marginal rates at which costs of learning (foregone earnings and direct costs) can be deducted.

References

Acemoglu, Daron J., and Joshua Angrist (2000), "How Large are the Social Returns to Education? Evidence from Compulsory Schooling Laws", *NBER Macroeconomics Annual*, 9-59.

Alstedsaeter, Annette (2003), "Does the Tax System Encourage too Much Education?", *FinanzArchiv*, 59, (1), 27-48.

Andersson, Fredrik, and Kai A. Konrad (2001), "Human Capital Investment and Globalization in Extortionary States", *Journal of Public Economics*, forthcoming.

Atkinson, Anthony B., and Joseph E. Stiglitz (1980), *Lectures on Public Economics*, New York: McGraw-Hill.

Becker, Gary S., and Nigel Tomes (1979), "Human Capital and the Rise and Fall of Families, *Journal of Labor Economics*, 4, (3), S1-S39.

Boadway, Robin, Nicolas Marceau, and Maurice Marchand (1996), "Investment in Education and the Time Inconsistency of Redistributive Tax Policy", *Economica* 63, 171-189.

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

Cameron, Stephen, and Christopher Taber (2000), "Borrowing Constraints and the Returns to Schooling", NBER Working Paper 7761.

Carneiro, Pedro, and James J. Heckman (2003). "Human Capital Policy", NBER Working Paper 9495.

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.

Eaton, Jonathan, and Harvey S. Rosen (1980), "Taxation, Human Capital and Uncertainty", *American Economic Review* 70 (4), 705-715.

Ebert, Udo (1992), "A Reexamination of the Optimal Nonlinear Income Tax", *Journal of Public Economics*, 49, (1), 47-73.

Ewijk, Casper van, and Paul J.G. Tang (2000), "Efficient Progressive Taxes and Education Subsidies", CPB Research Memorandum 170, CPB Netherlands Bureau for Economic Policy Analysis.

Hare, Paul G., and David T. Ulph (1979), "On Education and Distribution", *Journal of Political Economy*, 87, S193-S212.

Heckman, James J., and Peter J. Klenow (1998), "Human Capital Policy", in: Boskin, Michael (ed.). *Policies to Promote Human Capital Formation*, Hoover Institution.

Jacobs, Bas and Sweder J. G. van Wijnbergen (2002), "Optimal Financing of Education with Capital and Insurance Markets Imperfections", mimeo: University of Amsterdam.

Judd, Kenneth L. (1999), "Optimal Taxation and Spending in General Competitive Growth Models", *Journal of Public Economics*, 71, 1-26.

Kotlikoff, Laurence J., and Lawrence H. Summers (1979), "Tax Incidence in a Life Cycle Model with Variable Labor Supply", *Quarterly Journal of Economics*, 93, 705-718.

Krueger, Alan B., and Mikael Lindahl (2002), "Education and Growth: Why and for Whom", *Journal of Economic Literature*, 39, (4), 1101-1136.

Lommerud, Kjell Erik (1989), "Educational Subsidies when Relative Income Matters", *Oxford Economic Papers*, 41, 640-652.

Milesi-Ferretti, Gian Maria, and Nouriel Roubini (1998), "On the Taxation of Human and Physical Capital in Models of Endogenous Growth", *Journal of Public Economics*, 70, 237-254.

Mirrlees, James A. (1971), "An Exploration in the Theory of Optimum Income Taxation", *Review of Economic Studies*, 38, 175-208.

Nielsen, Søren B., and Peter B. Sørensen (1997), "On the Optimality of the Nordic System of Dual Income Taxation", *Journal of Public Economics*, 63, 311-329.

OECD (2002), *Tax Data Base 2002*, Paris: OECD.

OECD (2003), *Education at a Glance 2003*, Paris: OECD.

Shea, John (2000), "Does Parents' Money Matter?", *Journal of Public Economics*, 77, 155-184.

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

Ulph, David T. (1977). "On the Optimal Distribution of Income and Educational Expenditure", *Journal of Public Economics*, 8, 341-356.

Appendix

Second-order conditions

To derive the second-order conditions, we employ a two-step budgeting procedure. In the first stage, individuals select labor supply l_n and aggregate investment in human capital $e_n \equiv \psi(x_n, y_n)$. Writing the household budget constraint as $c_n = (1-t)n\phi(e_n)l_n - p_e e_n + g$, where the price index of educational spending is $p_e \equiv ((1-t)(1-s_x)p_x x_n + (1-s_y)p_y y_n)/e_n$, and substituting this budget constraint into the utility function $u(c_n, l_n)$ to eliminate c_n, we arrive at the following maximization problem

$$\max_{\{l_n, e_n\}} U_n = u\left((1-t)n\phi(.)l_n - p_e e_n + g, l_n\right). \qquad (54)$$

The first-order conditions are

$$\frac{\partial U_n}{\partial l_n} = u_c(1-t)n\phi(.) + u_l = 0, \qquad (55)$$

$$\frac{\partial U_n}{\partial e_n} = u_c[(1-t)nl_n\phi_e(.) - p_e] = 0. \qquad (56)$$

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

$$H \equiv \begin{bmatrix} u_{cl}(1-t)n\phi(.) + u_{ll} & u_c(1-t)n\phi_e \\ u_c(1-t)n\phi_e & u_c(1-t)nl_n\phi_{ee} \end{bmatrix}. \tag{57}$$

For utility to reach a maximum, the Hessian matrix should be negative definite. This is the case if the leading principal minors of H switch sign. The first principal minor is $u_{ll} + u_{cl}(1-t)n\phi(.) = u_{ll} - \frac{u_{cl}u_l}{u_c} < 0$. Therefore, the second leading principal minor must be positive. Given the elasticities of the production function $\phi(e_n) = e_n^\beta$, this implies

$$\beta(1 + \varepsilon_n^*) < 1, \tag{58}$$

where $\varepsilon_n^* \equiv \left(\frac{l u_{ll}}{u_l} - \frac{l u_{cl}}{u_c} \right)^{-1}$.

To relate this inequality to $\mu_n \equiv 1 - \beta(1 + \varepsilon_n)$, we link ε_n^* to the compensated wage elasticity of labor supply ε_n. By totally differentiating the first-order condition $-u_l/u_c = w_n$ (where w_n is the after-tax wage) and substituting the totally differentiated utility function (while holding utility constant, $i.e.$, $u_c dc_n + u_l dl_n = 0$) to eliminate dc_n, we can write the compensated wage elasticity of labor supply ε_n as

$$\varepsilon_n \equiv \frac{dl_n^*}{dw_n} \frac{w_n}{l_n^*} = \left(\left(\frac{l_n u_{ll}}{u_l} - \frac{l_n u_{cl}}{u_c} \right) - \left(\frac{-u_l}{u_c} \right) \left(\frac{l_n u_{cc}}{u_c} - \frac{l_n u_{cl}}{u_l} \right) \right)^{-1}, \tag{59}$$

so that $\varepsilon_n^{-1} = (\varepsilon_n^*)^{-1} + \rho$, where $\rho \equiv - \left(\frac{-u_l}{u_c} \right) \left(\frac{l_n u_{cc}}{u_c} - \frac{l_n u_{cl}}{u_l} \right) \geq 0$. So, we have $\varepsilon_n \leq \varepsilon_n^*$ and we find

$$\mu_n \equiv 1 - \beta(1 + \varepsilon_n) > 1 - \beta(1 + \varepsilon_n^*) > 0. \tag{60}$$

Therefore, the second-order condition implies that $\mu_n > 0$.

In the second stage of the budgeting procedure, individuals allocate their expenditures on education e_n optimally over x_n and y_n. Hence, they maximize $e_n = \psi(x_n, y_n)$ subject to the expenditure constraint: $p_e e_n = (1-t)(1-s_x)p_x x_n + (1-s_y)p_y y_n$. The resulting second-order condition $\psi_{xx}\psi_{yy} - (\psi_{xy})^2 > 0$ is implied by concavity of the sub-production function for human capital $\psi(x_n, y_n)$.

Derivation of compensated elasticities

We find the compensated behavioral elasticities with respect to policy variables by using (4) and (5). We multiply (4) by $(1-t)x_n$ and multiply (5) by y_n and add the results. Using the property that $\psi(x_n, y_n)$ is homogeneous of degree one: $\psi_x x_n + \psi_y y_n = \psi(x_n, y_n)$, we then find

$$\beta(1-t)nl_n\phi(e_n) = p_e e_n, \tag{61}$$

where $e_n \equiv \psi(x_n, y_n)$, $p_e \equiv \left((1-t)(1-s_x)p_x x_n + (1-s_y)p_y y_n\right)/e_n$, and $\phi(e_n) = e_n^\beta$.

Log-linearizing (61) (using $\phi(e_n) = e_n^\beta$) and the definition of p_e (using $e_n \equiv \psi(x_n, y_n)$ and (4) and (5)), we arrive at

$$-\tilde{t} + \tilde{l}_n + \beta\tilde{e}_n = \tilde{p}_e + \tilde{e}_n, \tag{62}$$

$$\tilde{p}_e = -\alpha(\tilde{t} + \tilde{s}_x) - (1-\alpha)\tilde{s}_y, \tag{63}$$

where a tilde stands for a *compensated* relative change (*i.e.*, $\tilde{l}_n \equiv dl_n^*/l_n^*$, $\tilde{x} \equiv dx_n^*/x_n^*$, et cetera), except for the tax rate and the subsidy rates where $\tilde{t} \equiv dt/(1-t)$ and $\tilde{s}_i \equiv ds_i/(1-s_i)$, $i = x, y$.

$\alpha \equiv \frac{x_n\psi_x}{\psi} = \frac{(1-t)(1-s_x)p_x x_n}{p_e e_n}$ does not depend on ability with linear taxes and subsidies. The reason is that (4) and (5) imply that the marginal rate of transformation does not depend on n:

$$\frac{\psi_x}{\psi_y} = \frac{(1-t)(1-s_x)p_x}{(1-s_y)p_y}. \tag{64}$$

ψ_x and ψ_y are functions of x_n/y_n only (since $\psi(.)$ is homogeneous of the first degree). Hence, (64) determines the ratio of the two inputs x_n/y_n as a function of $\frac{(1-t)(1-s_x)p_x}{(1-s_y)p_y}$ only. Hence, if policy instruments are linear, $\alpha = \frac{(1-t)(1-s_x)p_x x_n}{(1-t)(1-s_x)p_x x_n + (1-s_y)p_y y_n} = \frac{(1-t)(1-s_x)p_x(x_n/y_n)}{(1-t)(1-s_x)p_x(x_n/y_n) + (1-s_y)p_y}$ depends only on $(1-t)(1-s_x)p_x$ and $(1-s_y)p_y$, which do not depend on ability. However, if policy instruments are non-linear (and we replace s_x by $S_x'(p_x x_n)$, s_x by $S_x'(p_x x_n)$ and t by $T'(z_n - p_x x_n + S_x(p_x x_n))$ this is no longer true. Therefore, with the non-linear policy instruments, we will use a subscript n for both α and σ.

Expression (8) implies that compensated labor supply depends only on the after-tax wage rate $w \equiv (1-t)n\phi(x_n, y_n) = (1-t)ne_n^\beta$. We can write

$$\tilde{l}_n = \varepsilon_n(-\tilde{t} + \beta\tilde{e}_n), \tag{65}$$

where ε_n is the compensated wage elasticity of labor supply, $i.e.,$ $\varepsilon_n = \frac{\partial l_n^*}{\partial w_n} \frac{w_n}{l_n^*}$.

Substituting (65) and (63) into (62) to eliminate, respectively, $\tilde{l}_n$ and $\tilde{p}_e$, we can solve for $\tilde{\phi}_n = \beta \tilde{e}_n$:

$$\tilde{\phi}_n = \beta \tilde{e}_n = \frac{\beta}{\mu_n} \left((\alpha - (1 + \varepsilon_n)) \, \tilde{t} + \alpha \tilde{s}_x + (1 - \alpha) \tilde{s}_y \right), \qquad (66)$$

where $\mu_n \equiv 1 - (1 + \varepsilon_n) \beta > 0$. Substituting (66) into (65), we solve for $\tilde{l}_n$:

$$\tilde{l}_n = \frac{\beta \varepsilon_n}{\mu_n} \left(\alpha \tilde{s}_x + (1 - \alpha) \tilde{s}_y \right) - \frac{\varepsilon_n (1 - \alpha \beta)}{\mu_n} \tilde{t}. \qquad (67)$$

We find x_n and y_n by differentiating $e_n \equiv \psi(x_n, y_n)$ and using (4) and (5) to eliminate ψ_x and ψ_y to arrive at

$$\tilde{e} = \alpha \tilde{x} + (1 - \alpha) \tilde{y}. \qquad (68)$$

Differentiation of (64) yields

$$\tilde{x} - \tilde{y} = \sigma(\tilde{t} + \tilde{s}_x - \tilde{s}_y), \qquad (69)$$

where $\sigma \equiv -d\left(\frac{x_n}{y_n}\right) / d(\frac{\psi_x}{\psi_y})$ is 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 thus independent of ability (just like α). From (68) and (69), we can express $\tilde{x}$ and $\tilde{y}$ in terms of $\tilde{e}$ and substituting (66) to eliminate $\tilde{e}$, we find:

$$\tilde{x} = \frac{(\alpha - (1 + \varepsilon_n)) \, \tilde{t} + \alpha \tilde{s}_x + (1 - \alpha) \tilde{s}_y}{\mu_n} + (1 - \alpha) \sigma(\tilde{t} + \tilde{s}_x - \tilde{s}_y), \qquad (70)$$

$$\tilde{y} = \frac{(\alpha - (1 + \varepsilon_n)) \, \tilde{t} + \alpha \tilde{s}_x + (1 - \alpha) \tilde{s}_y}{\mu_n} - \alpha \sigma(\tilde{t} + \tilde{s}_x - \tilde{s}_y). \qquad (71)$$

(67), (70) and (71) yield the elasticities in the main text.

First-order conditions linear tax and education policies

We apply Roy's lemma to derive the following properties:

$$\frac{\partial v(g, t, s_x, s_y, n)}{\partial g} = \eta_n, \qquad (72)$$

32

$$\frac{\partial v(g,t,s_x,s_y,n)}{\partial t} = -\eta_n \left(n\phi(x_n,y_n)l_n - (1-s_x)p_x x_n \right), \tag{73}$$

$$\frac{\partial v(g,t,s_x,s_y,n)}{\partial s_x} = \eta_n(1-t)p_x x_n, \tag{74}$$

$$\frac{\partial v(g,t,s_x,s_y,n)}{\partial s_y} = \eta_n p_y y_n, \tag{75}$$

where η_n denotes private marginal utility of income for individual with ability n.

The first-order conditions for maximization of social welfare (21) with respect to t, s_x, and s_y are given by

$$\frac{\partial \mathcal{L}}{\partial t} = \int_{\underline{n}}^{\overline{n}} \left(-\Psi'\eta_n + \lambda \right) \left(nl_n\phi(x_n,y_n) - (1-s_x)p_x x_n \right) dF(n) + \tag{76}$$
$$\int_{\underline{n}}^{\overline{n}} \lambda \left(tn\phi(x_n,y_n)\frac{\partial l_n}{\partial t} + \Delta_x\frac{\partial x_n}{\partial t} + \Delta_y\frac{\partial y_n}{\partial t} \right) dF(n) = 0,$$

$$\frac{\partial \mathcal{L}}{\partial s_x} = \int_{\underline{n}}^{\overline{n}} \left(\Psi'\eta_n - \lambda \right) (1-t)p_x x_n dF(n) + \tag{77}$$
$$\int_{\underline{n}}^{\overline{n}} \lambda \left(tn\phi(x_n,y_n)\frac{\partial l_n}{\partial s_x} + \Delta_x\frac{\partial x_n}{\partial s_x} + \Delta_y\frac{\partial y_n}{\partial s_x} \right) dF(n) = 0,$$

$$\frac{\partial \mathcal{L}}{\partial s_y} = \int_{\underline{n}}^{\overline{n}} \left(\Psi'\eta_n - \lambda \right) p_y y_n dF(n) + \tag{78}$$
$$\int_{\underline{n}}^{\overline{n}} \lambda \left(tn\phi(x_n,y_n)\frac{\partial l_n}{\partial s_y} + \Delta_x\frac{\partial x_n}{\partial s_y} + \Delta_y\frac{\partial y_n}{\partial s_y} \right) dF(n) = 0,$$

where we used (73), (74), and (75). These equations can be simplified using the Slutsky equations for labor supply and investment in human capital (a derivative with an asterisk denotes a compensated change): $\frac{\partial l_n}{\partial t} \equiv \frac{\partial l_n^*}{\partial t} - (n\phi(x_n,y_n)l_n - (1-s_x)p_x x_n)\frac{\partial l_n}{\partial g}$, $\frac{\partial x_n}{\partial t} \equiv \frac{\partial x_n^*}{\partial t} - (n\phi(x_n,y_n)l_n - (1-s_x)p_x x_n)\frac{\partial x_n}{\partial g}$, $\frac{\partial y_n}{\partial t} \equiv \frac{\partial y_n^*}{\partial t} - (n\phi(x_n,y_n)l_n - (1-s_x)p_x x_n)\frac{\partial y_n}{\partial g}$, $\frac{\partial l_n}{\partial s_x} \equiv \frac{\partial l_n^*}{\partial s_x} + (1-t)p_x x_n\frac{\partial l_n}{\partial g}$, $\frac{\partial x_n}{\partial s_x} \equiv \frac{\partial x_n^*}{\partial s_x} + (1-t)p_x x_n\frac{\partial x_n}{\partial g}$, $\frac{\partial y_n}{\partial s_x} \equiv \frac{\partial y_n^*}{\partial s_x} + (1-t)p_x x_n\frac{\partial y_n}{\partial g}$, $\frac{\partial l_n}{\partial s_y} \equiv \frac{\partial l_n^*}{\partial s_y} + p_y y_n\frac{\partial l_n}{\partial g}$, $\frac{\partial x_n}{\partial s_y} \equiv \frac{\partial x_n^*}{\partial s_y} + p_y y_n\frac{\partial x_n}{\partial g}$, and $\frac{\partial y_n}{\partial s_y} \equiv \frac{\partial y_n^*}{\partial s_y} + p_y y_n\frac{\partial y_n}{\partial g}$. Substitution of the last

nine equations, together with the definition of b_n from (23) in the first-order conditions for t, s_x and s_y, yields

$$\int_{\underline{n}}^{\overline{n}} \left((1 - b_n) \left(nl_n\phi(.) - (1 - s_x)p_xx_n \right) + tn\phi(.)\frac{\partial l_n^*}{\partial t} + \Delta_x\frac{\partial x_n^*}{\partial t} + \Delta_y\frac{\partial y_n^*}{\partial t} \right) dF(n) = 0,$$
(79)

$$\int_{\underline{n}}^{\overline{n}} \left((b_n - 1)(1 - t)p_xx_n + tn\phi(.)\frac{\partial l_n^*}{\partial s_x} + \Delta_x\frac{\partial x_n^*}{\partial s_x} + \Delta_y\frac{\partial y_n^*}{\partial s_x} \right) dF(n) = 0,$$
(80)

$$\int_{\underline{n}}^{\overline{n}} \left((b_n - 1)p_yy_n + tn\phi(.)\frac{\partial l_n^*}{\partial s_y} + \Delta_x\frac{\partial x_n^*}{\partial s_y} + \Delta_y\frac{\partial y_n^*}{\partial s_y} \right) dF(n) = 0.$$
(81)

The first-order condition for the tax rate (79) can be rewritten using the tax elasticities ($\varepsilon_{jt} \equiv \frac{\partial j_n}{\partial t}\frac{(1-t)}{j_n}$; $j = l^*, x^*, y^*$), Δ_x, and Δ_y:

$$\int_{\underline{n}}^{\overline{n}} (b_n - 1)\left(nl_n\phi(.) - (1 - s_x)p_xx_n \right) dF(n)$$
(82)

$$= \int_{\underline{n}}^{\overline{n}} \left(\frac{t}{1-t}n\phi(.)l_n\varepsilon_{lt} - \frac{s_xp_xx_n}{1-t}\varepsilon_{xt} + \frac{(t - s_y)}{(1-t)}\frac{p_yy_n}{(1-t)}\varepsilon_{yt} \right) dF(n).$$

We use now (6) and (7) to write p_xx_n and p_yy_n in terms of $nl_n\phi(.)$ to yield (after dividing by $(1 - \alpha\beta)$):

$$\int_{\underline{n}}^{\overline{n}} (b_n - 1) nl_n\phi(.)dF(n) = \frac{t}{1-t}\int_{\underline{n}}^{\overline{n}} \frac{\varepsilon_{lt}}{(1 - \alpha\beta)}nl_n\phi(.)dF(n)$$
(83)

$$- \frac{s_x}{(1 - s_x)(1 - t)}\int_{\underline{n}}^{\overline{n}} \frac{\alpha\beta\varepsilon_{xt}}{(1 - \alpha\beta)}nl_n\phi(.)dF(n)$$

$$+ \frac{t - s_y}{(1 - s_y)(1 - t)}\int_{\underline{n}}^{\overline{n}} \frac{(1 - \alpha)\beta\varepsilon_{yt}}{(1 - \alpha\beta)}nl_n\phi(.)dF(n).$$

Similarly, we can rewrite the first-order condition for the subsidy s_x (80) by using the subsidy elasticities ($\varepsilon_{js_x} \equiv \frac{\partial j_n}{\partial s_x}\frac{(1-s_x)}{j_n}$; $j = l^*, x^*, y^*$), Δ_x, and Δ_y:

$$\int_{\underline{n}}^{\overline{n}} (1 - b_n)(1 - t)p_xx_ndF(n)$$
(84)

$$= \int_{\underline{n}}^{\overline{n}} \left(\frac{t}{1 - s_x}nl_n\phi(.)\varepsilon_{ls_x} - \frac{s_xp_xx_n}{1 - s_x}\varepsilon_{xs_x} + \frac{(t - s_y)p_yy_n}{(1 - t)(1 - s_x)}\varepsilon_{ys_x} \right) dF(n).$$

34

We use now (6) and (7) to write $p_x x_n$ and $p_y y_n$ in terms of $n l_n \phi(.)$ to obtain (after dividing through by $\frac{(1-t)\alpha\beta}{(1-s_x)}$)

$$\int_{\underline{n}}^{\overline{n}} (1 - b_n)\, n l_n \phi(.) dF(n) = \frac{t}{1-t} \int_{\underline{n}}^{\overline{n}} \frac{\varepsilon_{ls_x}}{\alpha\beta} n l_n \phi(.) dF(n) \qquad (85)$$

$$-\frac{s_x}{(1-s_x)(1-t)} \int_{\underline{n}}^{\overline{n}} \varepsilon_{xs_x} n l_n \phi(.) dF(n)$$

$$+\frac{(t-s_y)}{(1-s_y)(1-t)} \int_{\underline{n}}^{\overline{n}} \frac{(1-\alpha)\varepsilon_{ys_x}}{\alpha} n l_n \phi(.) dF(n).$$

Finally, we rewrite the first-order condition for s_y (81) using the subsidy elasticities ($\varepsilon_{js_y} \equiv \frac{\partial j_n}{\partial s_y} \frac{(1-s_y)}{j_n}$; $j = l^*, x^*, y^*$), Δ_x, and Δ_y:

$$\int_{\underline{n}}^{\overline{n}} (1 - b_n)\, p_y y_n dF(n) \qquad (86)$$

$$= \int_{\underline{n}}^{\overline{n}} \left(\frac{t}{1-s_y} n\phi(.) l_n \varepsilon_{ls_y} - \frac{s_x p_x x_n}{1-s_y} \varepsilon_{xs_y} + \frac{t - s_y p_y y_n}{(1-t)(1-s_y)} \varepsilon_{ys_y} \right) dF(n).$$

We use now (6) and (7) to write $p_x x_n$ and $p_y y_n$ in terms of $n l_n \phi(.)$ to obtain (after dividing through by $\frac{(1-t)v}{(1-s_y)}$)

$$\int_{\underline{n}}^{\overline{n}} (1 - b_n)\, n l_n \phi(.) dF(n) = \frac{t}{1-t} \int_{\underline{n}}^{\overline{n}} \frac{\varepsilon_{ls_y}}{(1-\alpha)\beta} n l_n \phi(.) dF(n) \quad (87)$$

$$-\frac{s_x}{(1-s_x)(1-t)} \int_{\underline{n}}^{\overline{n}} \frac{\alpha\varepsilon_{xs_y}}{(1-\alpha)} n l_n \phi(.) dF(n)$$

$$+\frac{t - s_y}{(1-s_y)(1-t)} \int_{\underline{n}}^{\overline{n}} \varepsilon_{ys_y} n l_n \phi(.) dF(n).$$

Optimal linear income taxation

Together with the definition of the distributional characteristic (25), the first-order condition for t (83) yields the optimal linear income tax (26) (with $s_x = s_y = 0$).

Optimal linear education subsidies

If we subtract (87) from (85), we find

$$\frac{t}{1-t}\int_{\underline{n}}^{\overline{n}}\left(\frac{\varepsilon_{ls_x}}{\alpha\beta}-\frac{\varepsilon_{ls_y}}{(1-\alpha)\beta}\right)nl_n\phi(.)dF(n) \tag{88}$$

$$=\frac{s_x}{(1-s_x)(1-t)}\int_{\underline{n}}^{\overline{n}}\left(\varepsilon_{xs_x}-\frac{\alpha\varepsilon_{xs_y}}{(1-\alpha)}\right)nl_n\phi(.)dF(n)$$

$$-\frac{t-s_y}{(1-s_y)(1-t)}\int_{\underline{n}}^{\overline{n}}\left(\frac{(1-\alpha)\varepsilon_{ys_x}}{\alpha}-\varepsilon_{ys_y}\right)nl_n\phi(.)dF(n).$$

Expressing these elasticities in terms of the compensated labor supply elasticity ε_n and the parameters of the production function of human capital (see the expressions at the end of sub-section 3.1), we find that $\frac{\varepsilon_{ls_x}}{\alpha\beta}-\frac{\varepsilon_{ls_y}}{(1-\alpha)\beta}=0$, $\varepsilon_{xs_x}-\frac{\alpha\varepsilon_{xs_y}}{(1-\alpha)}=\sigma$ and $\frac{(1-\alpha)\varepsilon_{ys_x}}{\alpha}-\varepsilon_{ys_y}=-\sigma$ so that

$$\frac{-s_x}{(1-s_x)(1-t)}=\frac{t-s_y}{(1-s_y)(1-t)} \tag{89}$$

if $\sigma>0$. Substituting (89) back into (85) to eliminate $\frac{t-s_y}{(1-s_y)(1-t)}$ and using expressions at the end of sub-section 3.1 to eliminate the elasticities, we arrive at (29). Employing the same expressions to rewrite (85), we establish (30).

Optimal linear income taxation and linear education subsidies

Adding (87) to (83) and using (89) to eliminate $\frac{-s_x}{(1-s_x)(1-t)}$, we find

$$\frac{t}{1-t}\int_{\underline{n}}^{\overline{n}}\left(\frac{\varepsilon_{lt}}{(1-\alpha\beta)}+\frac{\varepsilon_{ls_y}}{(1-\alpha)\beta}\right)nl_n\phi(.)dF(n)= \tag{90}$$

$$-\frac{t-s_y}{(1-s_y)(1-t)}\int_{\underline{n}}^{\overline{n}}\left(\frac{\alpha\beta\varepsilon_{xt}}{(1-\alpha\beta)}+\frac{(1-\alpha)\beta\varepsilon_{yt}}{(1-\alpha\beta)}+\frac{\alpha\varepsilon_{xs_y}}{(1-\alpha)}+\varepsilon_{ys_y}\right)nl_n\phi(.)dF(n)$$

Using the expressions for the elasticities, we derive that $\frac{\varepsilon_{lt}}{(1-\alpha\beta)}+\frac{\varepsilon_{ls_y}}{(1-\alpha)\beta}=0$ and $\frac{\alpha\beta\varepsilon_{xt}}{(1-\alpha\beta)}+\frac{(1-\alpha)\beta\varepsilon_{yt}}{(1-\alpha\beta)}+\frac{\alpha\varepsilon_{xs_y}}{(1-\alpha)}+\varepsilon_{ys_y}=\frac{-1}{1-\alpha\beta}$, so that

$$\frac{s_y-t}{(1-s_y)(1-t)}\left(\int_{\underline{n}}^{\overline{n}}\frac{1}{1-\alpha\beta}n\phi(x_n,y_n)l_ndF(n)\right)=0. \tag{91}$$

Since the term between parentheses is non zero, we must have $s_y = t$. (89) then yields $s_x = 0$. Substituting these results in (83) and employing the expressions at the end of sub-section 3.1 to write the elasticities in terms of ε_n and the parameters of the production function for human capital, we obtain (33).

Optimal linear income taxation and imperfect linear education subsidies

Adding (83) and (85), we establish

$$\frac{t}{1-t} \int_{\underline{n}}^{\overline{n}} \left(\frac{\varepsilon_{lt}}{(1-\alpha\beta)} + \frac{\varepsilon_{ls_x}}{\alpha\beta} \right) nl_n \phi(.) dF(n) \tag{92}$$

$$= \frac{s_x}{(1-s_x)(1-t)} \int_{\underline{n}}^{\overline{n}} \left(\frac{\alpha\beta\varepsilon_{xt}}{1-\alpha\beta} + \varepsilon_{xs_x} \right) nl_n \phi(.) dF(n)$$

$$- \frac{(t-s_y)}{(1-s_y)(1-t)} \int_{\underline{n}}^{\overline{n}} \left(\frac{(1-\alpha)\beta\varepsilon_{yt}}{(1-\alpha\beta)} + \frac{(1-\alpha)\varepsilon_{ys_x}}{\alpha} \right) nl_n \phi(.) dF(n).$$

Expressing the elasticities in terms of the parameters of the production function of human capital and the compensated wage elasticity of labor supply ε_n, we find for the terms in parentheses: $\frac{\varepsilon_{lt}}{(1-\alpha\beta)} + \frac{\varepsilon_{ls_x}}{\alpha\beta} = 0$, $\frac{\alpha\beta\varepsilon_{xt}}{1-\alpha\beta} + \varepsilon_{xs_x} = \frac{\alpha+(1-\alpha)\sigma}{1-\alpha\beta}$ and $\frac{(1-\alpha)\beta\varepsilon_{yt}}{(1-\alpha\beta)} + \frac{(1-\alpha)\varepsilon_{ys_x}}{\alpha} = \frac{(1-\alpha)(1-\sigma)}{1-\alpha\beta}$. Substituting these expressions, which are independent of ability n, into (92), we find

$$\frac{s_x}{1-s_x} = \frac{(1-\alpha)(1-\sigma)}{1-(1-\alpha)(1-\sigma)} t. \tag{93}$$

Substitution of (93) into (85) and using the definitions for the elasticities, we arrive at (35).

Optimal non-linear income taxation

If the government has the non-linear income tax as its only policy instrument, it employs only l_n as a control variable, taking into account the indirect effects of changing that control variable l_n and the state variable u_n on x_n, y_n and c_n. The first-order condition for optimizing the Hamiltonian with respect to

l_n amounts to

$$\frac{\partial \mathcal{H}}{\partial l_n} = \lambda \left(n\phi(x_n, y_n) - \left.\frac{dc_n}{dl_n}\right|_{\bar{u}} \right) f(n) + \tag{94}$$

$$\lambda \left((n\phi_x l_n - p_x) \left.\frac{dx_n}{dl_n}\right|_{\bar{u}} + (n\phi_y l_n - p_y) \left.\frac{dy_n}{dl_n}\right|_{\bar{u}} \right) f(n) +$$

$$\frac{\theta_n}{n} \left(u_l + l_n u_{ll} + l_n u_{lc} \left.\frac{dc_n}{dl_n}\right|_{\bar{u}} \right) = 0.$$

Substituting $\left.\frac{dc_n}{dl_n}\right|_{\bar{u}} = -\frac{u_l}{u_c} = (1 - T')n\phi(.)$ (derived by differentiating utility $u(c_n, l_n)$ and substituting the marginal rate of substitution between work and consumption (37)), the private first-order condition for x_n (38) ($n\phi_x l_n - p_x = 0$ with $S_x = 0$), the private first-order condition for y_n (39) ($n\phi_y l_n - p_y = \frac{T' p_y}{1 - T'}$ and $S_y = 0$), and $u_l = -u_c(1 - T')n\phi(.)$ (derived from (37)), yields the optimal marginal income tax rates:

$$\frac{T'}{1 - T'} = \frac{u_c \theta_n / \lambda}{n f(n)} \left(1 + \frac{1}{\varepsilon_n^*} \right) \left(\frac{1}{1 + v_n \varepsilon_{yl}} \right), \tag{95}$$

where $\varepsilon_n^* \equiv \left(\frac{l u_{ll}}{u_l} - \frac{l u_{cl}}{u_c} \right)^{-1}$, $v_n \equiv \frac{p_y y_n}{(1 - T')n\phi l_n} = (1 - \alpha_n)\beta$, and $\varepsilon_{yl} \equiv \left.\frac{l_n}{y_n}\frac{dy_n}{dl_n}\right|_{\bar{u}}$. The last elasticity can be written as $\varepsilon_{yl} = \frac{\varepsilon_{yt}}{\varepsilon_{lt}}$, where ε_{yt} (ε_{lt}) is the compensated tax elasticity of y_n (l_n), since the government uses the marginal tax rate to change l_n at a given utility level. Upon substitution for the expressions for the elasticities ε_{lt} (10), and ε_{yt} (16), we obtain (45).

The first-order condition for u_n is

$$\frac{\partial \mathcal{H}}{\partial u_n} = \left(\Psi' + \lambda(n\phi_x l_n - p_x) \left.\frac{dx_n}{du_n}\right|_{\bar{l}} + \lambda(n\phi_y l_n - p_y) \left.\frac{dy_n}{du_n}\right|_{\bar{l}} - \lambda \left.\frac{dc_n}{du_n}\right|_{\bar{l}} \right) f(n) \tag{96}$$

$$\theta_n \frac{l_n u_{lc}}{n} \left.\frac{dc_n}{du_n}\right|_{\bar{l}} = \frac{d\theta_n}{dn}.$$

(Note that we defined θ_n negatively.) We can simplify this expression using the Envelope Theorem, which implies that $\left.\frac{dx_n}{du_n}\right|_{\bar{l}} = \left.\frac{dy_n}{du_n}\right|_{\bar{l}} = 0$, so that

$$\frac{\partial \mathcal{H}}{\partial u_n} = \left(\Psi' - \lambda \left.\frac{dc_n}{du_n}\right|_{\bar{l}} \right) f(n) - \theta_n \frac{l_n u_{lc}}{n} \left.\frac{dc_n}{du_n}\right|_{\bar{l}} = -\frac{d\theta_n}{dn}. \tag{97}$$

This is a standard expression (see e.g. Mirrlees (1971)). One cannot solve this differential equation analytically without making additional assumptions on the distribution of abilities or (social) preferences.

Optimal non-linear education subsidies

If the government can only freely set education policy and has no access to the non-linear income tax, the government optimizes both S_x and S_y by controlling x_n and y_n. Further, it must take into account the indirect effects on l_n and c_n. The first-order condition for optimizing social welfare with respect to x_n amounts to

$$
\frac{\partial \mathcal{H}}{\partial x_n} = \lambda \left(n\phi_x l_n - p_x + n\phi(x_n, y_n) \left. \frac{dl_n}{dx_n} \right|_{\bar{y}, \bar{u}} - \left. \frac{dc_n}{dx_n} \right|_{\bar{y}, \bar{u}} \right) f(n) + \quad (98)
$$

$$
\frac{\theta_n}{n} \left(u_l \left. \frac{dl_n}{dx_n} \right|_{\bar{y}, \bar{u}} + l_n u_{ll} \left. \frac{dl_n}{dx_n} \right|_{\bar{y}, \bar{u}} + l_n u_{lc} \left. \frac{dc_n}{dx_n} \right|_{\bar{y}, \bar{u}} \right) = 0
$$

Substituting $\left. \frac{dc_n}{dx_n} \right|_{\bar{y}, \bar{u}} = \frac{dc_n}{dl_n} \left. \frac{dl_n}{dx_n} \right|_{\bar{y}, \bar{u}} = -\frac{u_l}{u_c} \left. \frac{dl_n}{dx_n} \right|_{\bar{y}, \bar{u}} = n\phi(.) \left. \frac{dl_n}{dx_n} \right|_{\bar{y}, \bar{u}}$ (where the last equality follows from the marginal rate of substitution (37) and $T' = 0$), $u_l = -u_c n\phi(.)$, and $n\phi_x l_n - p_x = -S'_x p_x$ (from the first-order condition for x_n (38)), we arrive at

$$
\frac{-S'_x}{1 - S'_x} = \frac{u_c \theta_n / \lambda}{n f(n)} \left(1 + \frac{1}{\varepsilon_n^*} \right) \frac{\varepsilon_{lx}}{\gamma_n}, \qquad (99)
$$

where $\gamma_n \equiv \frac{(1 - S'_x) p_x x_n}{n\phi l_n} = \alpha_n \beta$ and $\varepsilon_{lx} \equiv \frac{x_n}{l_n} \left. \frac{dl_n}{dx_n} \right|_{\bar{y}, \bar{u}}$.

The first-order condition for y_n is

$$
\frac{\partial \mathcal{H}}{\partial y_n} = \lambda \left(n\phi_y l_n - p_y + n\phi(x_n, y_n) \left. \frac{dl_n}{dy_n} \right|_{\bar{x}, \bar{u}} - \left. \frac{dc_n}{dy_n} \right|_{\bar{x}, \bar{u}} \right) f(n) + \quad (100)
$$

$$
\frac{\theta_n}{n} \left(u_l \left. \frac{dl_n}{dy_n} \right|_{\bar{x}, \bar{u}} + l u_{ll} \left. \frac{dl_n}{dx_n} \right|_{\bar{x}, \bar{u}} + l_n u_{lc} \left. \frac{dc_n}{dx_n} \right|_{\bar{x}, \bar{u}} \right) = 0.
$$

In a similar way as with the first-order condition for x_n, we write this expression as

$$
\frac{-S'_y}{1 - S'_y} = \frac{u_c \theta_n / \lambda}{n f(n)} \left(1 + \frac{1}{\varepsilon_n^*} \right) \frac{\varepsilon_{ly}}{\upsilon_n}, \qquad (101)
$$

where $\upsilon_n \equiv \frac{(1 - S'_y) p_y y_n}{n\phi l_n} = (1 - \alpha_n)\beta$ and $\varepsilon_{ly} \equiv \frac{y_n}{l_n} \left. \frac{dl_n}{dy_n} \right|_{\bar{y}, \bar{u}}$.

Using the first-order conditions for x_n and y_n (*i.e.*, (99) and (101)), we find

$$\frac{-S'_x}{1-S'_x}\frac{\gamma_n}{\varepsilon_{lx}} = \frac{-S'_y}{1-S'_y}\frac{\upsilon_n}{\varepsilon_{ly}}. \tag{102}$$

We can rewrite this expression by substituting the corresponding expressions for ε_{lx} and ε_{ly}. We find ε_{lx} by setting $\tilde{s}_x$ and $\tilde{s}_y$ such that $\tilde{y}_n = 0$ (note that in the first-order condition for x_n, y_n is constant, so that $\tilde{y}_n = 0$). This implies that

$$\varepsilon_{ys_x}\tilde{s}_x + \varepsilon_{ys_y}\tilde{s}_y = 0. \tag{103}$$

Solving $\tilde{s}_y$ in terms of $\tilde{s}_x$ and substituting this in the expressions for $\tilde{l}_n = \varepsilon_{ls_x}\tilde{s}_x + \varepsilon_{ls_y}\tilde{s}_y$ and $\tilde{x}_n = \varepsilon_{xs_x}\tilde{s}_x + \varepsilon_{xs_y}\tilde{s}_y$ we find

$$\varepsilon_{lx} = \frac{\tilde{l}_n}{\tilde{x}_n} = \frac{\varepsilon_{ls_x} - \varepsilon_{ls_y}\frac{\varepsilon_{ys_x}}{\varepsilon_{ys_y}}}{\varepsilon_{xs_x} - \varepsilon_{xs_y}\frac{\varepsilon_{ys_x}}{\varepsilon_{ys_y}}} = \frac{\varepsilon_{ls_x}\varepsilon_{ys_y} - \varepsilon_{ls_y}\varepsilon_{ys_x}}{\varepsilon_{xs_x}\varepsilon_{ys_y} - \varepsilon_{xs_y}\varepsilon_{ys_x}}. \tag{104}$$

In a similar way, we find for ε_{ly} (using that $\tilde{x}_n = 0$ in the first-order condition for y_n)

$$\varepsilon_{ly} = \frac{\tilde{l}_n}{\tilde{y}_n} = \frac{\varepsilon_{ls_x} - \varepsilon_{ls_y}\frac{\varepsilon_{xs_x}}{\varepsilon_{xs_y}}}{\varepsilon_{ys_x} - \varepsilon_{ys_y}\frac{\varepsilon_{xs_x}}{\varepsilon_{xs_y}}} = \frac{\varepsilon_{ls_y}\varepsilon_{xs_x} - \varepsilon_{ls_x}\varepsilon_{xs_y}}{\varepsilon_{xs_x}\varepsilon_{ys_y} - \varepsilon_{xs_y}\varepsilon_{ys_x}}. \tag{105}$$

Substituting (104) and (105) into (102), we find the expression

$$\frac{-S'_x}{1-S'_x}\gamma_n\left(\varepsilon_{xs_x} - \frac{\varepsilon_{ls_x}}{\varepsilon_{ls_y}}\varepsilon_{xs_y}\right) = \frac{-S'_y}{1-S'_y}\upsilon_n\left(\frac{\varepsilon_{ls_x}}{\varepsilon_{ls_y}}\varepsilon_{ys_y} - \varepsilon_{ys_x}\right). \tag{106}$$

Moreover, we can use the elasticities at the end of sub-section 3.1 (*i.e.*, (11) and (12)) to eliminate ε_{ls_x} and ε_{ls_y} to arrive at

$$\frac{-S'_x}{1-S'_x}\left(\varepsilon_{xs_x} - \frac{\gamma_n}{\upsilon_n}\varepsilon_{xs_y}\right) = \frac{-S'_y}{1-S'_y}\left(\varepsilon_{ys_y} - \frac{\upsilon_n}{\gamma_n}\varepsilon_{ys_x}\right). \tag{107}$$

This can be simplified using the definitions for the elasticities as

$$S'_x = S'_y. \tag{108}$$

The optimal level of the education tax follows from substitution of the relevant elasticities in the first-order condition for x_n (99). In particular, we can substitute the elasticities at the end of sub-section 3.1 in $\varepsilon_{lx} = \frac{\varepsilon_{ls_x}\varepsilon_{ys_y} - \varepsilon_{ls_y}\varepsilon_{ys_x}}{\varepsilon_{xs_x}\varepsilon_{ys_y} - \varepsilon_{xs_y}\varepsilon_{ys_x}}$ $= \gamma_n\varepsilon_n$ and use (108) in (99), to arrive at (46).

Optimal non-linear income taxation and non-linear education subsidies

If the government optimizes all its instruments, it employs the control variables l_n, x_n, and y_n to optimize social welfare, taking into account the indirect effects on c_n. This yields the following first-order conditions for x_n and y_n

$$\frac{\partial \mathcal{H}}{\partial x_n} = \lambda \left(n\phi_x l_n - p_x - \frac{dc_n}{dx_n}\bigg|_{\bar{u},\bar{l},\bar{y}} \right) f(n) = 0, \tag{109}$$

$$\frac{\partial \mathcal{H}}{\partial y_n} = \lambda \left(n\phi_y l_n - p_y - \frac{dc_n}{dy_n}\bigg|_{\bar{u},\bar{l},\bar{x}} \right) f(n) = 0. \tag{110}$$

For the indirect impacts on consumption, we find $\frac{dc_n}{dx_n}\big|_{\bar{u},\bar{l},\bar{y}} = \frac{dc_n}{dy_n}\big|_{\bar{u},\bar{l},\bar{x}} = 0$ by differentiating the household budget constraint (36) and substituting the individuals' first-order conditions for x_n and y_n (38) and (39). Using these results in (109) and (110), we find (47) and (48).

The first-order condition for l_n is given by

$$\frac{\partial \mathcal{H}}{\partial l_n} = \lambda \left(n\phi(x_n, y_n) - \frac{dc_n}{dl_n}\bigg|_{\bar{u},\bar{x},\bar{y}} \right) f(n) + \frac{\theta_n}{n} \left(u_l + l_n u_{ll} + l_n u_{lc} \frac{dc_n}{dl_n}\bigg|_{\bar{u},\bar{x},\bar{y}} \right) = 0, \tag{111}$$

where we substitute $\frac{dc_n}{dl_n}\big|_{\bar{u},\bar{x},\bar{y}} = (1 - T')n\phi(x_n, y_n)$ (found by taking the total derivative of utility $u(c_n, l_n)$ and substituting the marginal rate of substitution (37) to eliminate u_l), and $u_l = u_c(1 - T'))n\phi(x_n, y_n)$ (found from (37)) to arrive at (51).

Optimal non-linear income taxation and imperfect non-linear education subsidies

If the government can only optimize T and S_x, it uses l_n and x_n as control variables, taking into account the indirect effects of changing l_n and x_n on c_n and y_n. The first-order condition for social-welfare maximization with

41

respect to x_n is given by

$$
\begin{aligned}
\frac{\partial \mathcal{H}}{\partial x_n} = {} & \lambda \left(n\phi_x l_n - p_x - \frac{dc_n}{dx_n}\Big|_{\bar{u},\bar{l}} - (n\phi_y l_n - p_y)\frac{dy_n}{dx_n}\Big|_{\bar{u},\bar{l}} \right) f(n) + \\
& \frac{\theta_n}{n}\left(u_l \frac{dl_n}{dx_n}\Big|_{\bar{u},\bar{l}} + l_n u_{ll}\frac{dl_n}{dx_n}\Big|_{\bar{u},\bar{l}} + l_n u_{lc}\frac{dc_n}{dx_n}\Big|_{\bar{u},\bar{l}} \right) = 0.
\end{aligned}
\tag{112}
$$

Substituting $n\phi_x l_n - p_x = -S'_x p_x$, $\frac{dc_n}{dx_n}\big|_{\bar{u},\bar{l}} = 0$ (found by differentiating the household budget constraint (36) and substituting the individuals' first-order condition for x_n (38)), $(n\phi_y l_n - p_y) = \frac{T' p_y}{1-T'}$, and $\frac{dl_n}{dx_n}\big|_{\bar{u},\bar{l}} = 0$ (by definition), we find

$$
\frac{-S'_x}{(1-S'_x)(1-T')}p_x x_n (1 - S'_x)(1 - T') + \frac{T'}{1-T'}p_y y_n \varepsilon_{yx} = 0,
\tag{113}
$$

where $\varepsilon_{yx} \equiv \frac{x_n}{y_n}\frac{dy_n}{dx_n}\big|_{\bar{u},\bar{l}}$. By using $\upsilon_n \equiv \frac{p_y y_n}{(1-T')n\phi l_n} = (1-\alpha_n)\beta$ and $\gamma_n \equiv \frac{(1-S'_x)p_x x_n}{n\phi l_n} = \alpha_n \beta$, we can write this expression as

$$
\frac{S'_x}{(1-S'_x)(1-T')}\gamma_n = \frac{T'}{1-T'}\upsilon_n \varepsilon_{yx}.
\tag{114}
$$

We find ε_{yx} by defining $\tilde{s}_x$ and $\tilde{t}$ such that $\tilde{l} = 0$:

$$
\varepsilon_{ls_x}\tilde{s}_x + \varepsilon_{lt}\tilde{t} = 0.
\tag{115}
$$

Solving $\tilde{t}$ in terms of $\tilde{s}_x$ and substituting this in the expressions for $\tilde{y}_n = \varepsilon_{ys_x}\tilde{s}_x + \varepsilon_{yt}\tilde{t}$ and $\tilde{x}_n = \varepsilon_{xs_x}\tilde{s}_x + \varepsilon_{xt}\tilde{t}$, we find

$$
\varepsilon_{yx} = \frac{\tilde{y}_n}{\tilde{x}_n} = \frac{\varepsilon_{ys_x} - \varepsilon_{yt}\frac{\varepsilon_{ls_x}}{\varepsilon_{lt}}}{\varepsilon_{xs_x} - \varepsilon_{xt}\frac{\varepsilon_{ls_x}}{\varepsilon_{lt}}}.
\tag{116}
$$

We can substitute the elasticities at the end of sub-section 3.1 to find

$$
\varepsilon_{yx} = \frac{(1-\sigma_n)\alpha_n}{1 - (1-\alpha_n)(1-\sigma_n)}.
\tag{117}
$$

Substituting the last result into (114), and using $\gamma_n = \alpha_n \beta$ and $\upsilon_n = (1-\alpha_n)\beta$, we arrive at (52).

The first-order condition optimizing the Hamiltonian (43) with respect to l_n is

$$\frac{\partial \mathcal{H}}{\partial l_n} = \lambda \left(n\phi(x_n, y_n) - \frac{dc_n}{dl_n}\Big|_{\bar{u},\bar{x}} + (n\phi_y l_n - p_y)\frac{dy_n}{dl_n}\Big|_{\bar{u},\bar{x}} \right) f(n) + (118)$$

$$\frac{\theta_n}{n}\left(u_l + l_n u_{ll} + l_n u_{lc}\frac{dc_n}{dl_n}\Big|_{\bar{u},\bar{x}} \right) = 0.$$

Substituting $\frac{dc_n}{dl_n}\Big|_{\bar{u},\bar{x}} = -\frac{u_l}{u_c} = (1 - T')n\phi(.)$ (derived by differentiating utility $u(c_n, l_n)$ and substituting the marginal rate of substitution (37)), $(n\phi_y l_n - p_y) = \frac{T'p_y}{1-T'}$ (derived from the individuals' first-order conditions for y_n (39) and $S_y = 0$), and $u_l = -u_c(1 - T')n\phi(.)$ (derived from (37)), we find

$$\frac{T'}{1 - T'} = \frac{u_c \theta_n / \lambda}{n f(n)} \left(1 + \frac{1}{\varepsilon_n^*} \right) \left(\frac{1}{1 + \upsilon\varepsilon_{yl}} \right), \tag{119}$$

where $\upsilon_n \equiv \frac{p_y y_n}{(1-T')n\phi l_n}$ and $\varepsilon_{yl} \equiv \frac{l_n}{y_n}\frac{dy_n}{dl_n}\Big|_{\bar{u},\bar{x}}$.

We find ε_{yl} by defining $\tilde{s}_x$ and $\tilde{t}$ such that $\tilde{x} = 0$:

$$\varepsilon_{xs_x}\tilde{s}_x + \varepsilon_{xt}\tilde{t} = 0. \tag{120}$$

Solving $\tilde{t}$ in terms of $\tilde{s}_x$ and substituting this in the expressions for $\tilde{y}_n = \varepsilon_{ys_x}\tilde{s}_x + \varepsilon_{yt}\tilde{t}$ and $\tilde{l}_n = \varepsilon_{ls_x}\tilde{s}_x + \varepsilon_{lt}\tilde{t}$, we obtain

$$\varepsilon_{yl} = \frac{\tilde{y}_n}{\tilde{l}_n} = \frac{\varepsilon_{ys_x} - \varepsilon_{yt}\frac{\varepsilon_{xs_x}}{\varepsilon_{xt}}}{\varepsilon_{ls_x} - \varepsilon_{lt}\frac{\varepsilon_{xs_x}}{\varepsilon_{xt}}}. \tag{121}$$

Substituting the elasticities from the end of sub-section (3.1), we arrive at

$$\varepsilon_{yl} = \frac{1 + \varepsilon_n}{\varepsilon_n}\frac{\sigma_n}{(1 - \alpha_n)\sigma_n + \alpha_n}. \tag{122}$$

Substitution of this last result into (119) yields (53).