

Optimal redistributive tax and education policies in general equilibrium

Bas Jacobs

© The Author(s) 2012. This article is published with open access at Springerlink.com

Abstract This paper studies optimal linear and non-linear income taxes and education subsidies in two-type models with endogenous human capital formation, endogenous labor supply, and endogenous wage rates. Assuming constant human capital elasticities, human capital investment should be efficient under optimal linear policies, whether general equilibrium effects are present or not. Hence, education subsidies should not be used for distributional reasons. Due to general equilibrium effects, optimal linear income taxes may even become negative. Optimal non-linear policies exploit general equilibrium effects for redistribution. The high-skilled type optimally has a negative marginal income tax rate and a positive marginal education subsidy. The low-skilled type optimally faces a positive marginal income tax rate and a marginal tax on education. Simulations demonstrate that general equilibrium effects have only a modest effect on optimal non-linear policies.

Keywords Human capital · General equilibrium · Optimal taxation · Education subsidies

JEL Classification H2 · H5 · I2 · J2

B. Jacobs (✉)

Department of Economics, Erasmus School of Economics, Erasmus University Rotterdam,
PO box 1738, 3000 DR Rotterdam, The Netherlands

e-mail: bjacobs@ese.eur.nl

url: <http://people.few.eur.nl/bjacobs>

B. Jacobs

Tinbergen Institute, Rotterdam, The Netherlands

B. Jacobs

Netspar, Tilburg, The Netherlands

B. Jacobs

CESifo, Munich, Germany

“Educational policies deserve to be programmed not only with a view to improving education in the widest sense, but also in order to influence the income distribution” (Tinbergen 1975, p. 18).

1 Introduction

In his book *Income Distribution*, the Dutch Nobel-prize winner Tinbergen (1975) extensively discusses the merits of increasing the supply of skilled workers relative to unskilled workers to reduce wage inequality. As the relative supply of skilled workers falls, the skill premium is lowered, and wage inequality diminishes. Tinbergen’s concern with growing inequality between skilled and unskilled workers is more relevant today than it was in the 1970s. Many Western countries are currently confronted with sharply increasing skill premiums. Skill-biased technological change causes the demand for skilled workers to increase more rapidly than the supply of skilled workers (Katz and Autor 1999). In Tinbergen’s terminology: the race between education and technological change is currently lost by education. Also, globalization may jeopardize the prospects for low-skilled workers. In light of the deteriorating labor market position of low-skilled workers, it is not surprising that subsidies to foster skill formation have a strong policy appeal. By boosting human capital formation, equality may be served because general equilibrium effects on wages reduce the skill premium. As there is less pre-tax inequality, the need to redistribute incomes through distorting income taxes may diminish at the same time.

The main question of this paper is: Should general equilibrium effects on the wage distribution be exploited in an optimal redistributive tax-cum-education system? To answer this question, this paper analyzes optimal redistributive tax and education policies in a Mirrlees (1971)-framework. Due to imperfect substitution between different skill types in labor demand, the skill premium is determined by both demand and supply conditions in the labor market. Furthermore, skill levels are endogenously determined by human capital investments, and not exogenous as in Mirrlees (1971).

Naturally, governments use the income tax and transfer system to redistribute resources. An important question is, therefore, whether education policies should optimally complement the redistributive tax system. Bovenberg and Jacobs (2005) derive, in a partial equilibrium framework, that human capital formation should neither be taxed nor subsidized (on a net basis) in an optimal redistributive program with linear or nonlinear taxes and subsidies. The intuition is as follows. On the one hand, subsidies on education are implicit subsidies on work effort, since working and learning are complementary in generating income. On the other hand, education subsidies are regressive, since high-ability individuals invest more in human capital (the ‘ability bias’). With the earnings functions used by Bovenberg and Jacobs (2005), both effects cancel out, and the sole role of education subsidies is to offset the distortions of the income tax on skill formation. Indeed, the Diamond and Mirrlees (1971) production efficiency theorem dictates that all investments in human capital (an intermediate input) are efficient under these conditions (Jacobs and Bovenberg 2011).

This paper closely follows the analysis of Bovenberg and Jacobs (2005). However, the main difference is that the wages of workers are now endogenous. This paper

maintains their assumptions to ensure that human capital is optimally not subsidized in the *absence* of general equilibrium effects on wages. In particular, the earnings function is assumed to be weakly separable between ability, labor, and education, both with linear and nonlinear policies. Additionally, the production function for human capital is assumed to have a constant elasticity if linear policies are considered, see also Jacobs and Bovenberg (2011). Individuals also have an iso-elastic utility function, so that labor supply elasticities are constant, and income effects are absent. This simple utility function highlights the crucial role of human capital supply elasticities under linear policy instruments. Finally, the analysis is restricted to two types of agents that differ in their ability to acquire human capital. We refer to the high-ability type as the ‘skilled’ and the low-ability type as the ‘unskilled’ agent.¹

Whether education policy should optimally be employed to provoke redistributive general equilibrium effects on the skill premium is shown to depend crucially on two things. First, can education policy affect the relative supply of human capital in such a way that the skill premium falls? The skill premium will be a downward sloping function of the relative supplies of total ‘effective labor supply’. Effective labor supply is the product of the quantity of labor supply (i.e., hours worked) and the quality of labor supply (i.e., human capital), which is determined by human capital investment. Hence, education subsidies must increase the relative effective labor supply to be helpful in reducing wage inequality. Second, if education subsidies do indeed increase relative supply of effective labor, should education subsidies, when optimally combined with an income tax, also be used? The answers to both questions are not obvious, and they differ fundamentally for linear and nonlinear policy instruments.

The first part of this paper considers optimal *linear* tax and education subsidies. Linear education subsidies tend to increase effective labor supply of both workers at the same time, and it is not clear whether relative supply of human capital is increased at all. The effective labor supply of each agent depends on his human capital and his labor effort. It is demonstrated that general equilibrium effects can *never* be exploited for redistributive reasons when effective labor supply elasticities are equal across agents, and income effects are absent. The reason is that relative effective labor supply remains fixed. Indeed, linear education subsidies will only compress wage differentials if the skilled worker’s supply of effective labor is more elastic with respect to the subsidy than the unskilled worker’s effective labor supply. As agents are assumed to have human capital production functions with a constant elasticity (under linear instruments only), the labor supply elasticity of the skilled worker has to be higher than that of the unskilled worker for education subsidies to have the potential to compress the wage distribution.

The second question is: If the skilled worker has a higher labor supply elasticity, and education subsidies therefore can reduce the skill premium, should education subsidies also be employed in an optimal redistributive program alongside optimal linear taxes? With linear policies, the answer to this question is no. A linear education subsidy is distributionally equivalent to a linear income tax in the absence of

¹Strictly speaking, skills are endogenous, and not an innate trait of individuals. Therefore, referring to the high-(low-)ability type as being ‘skilled’ (‘unskilled’) is slightly abusing language for expositional reasons.

general equilibrium effects. The reason is that gross income is linear in education as a result of the constant elasticity of the production function for human capital. Hence, it is not optimal to exploit general equilibrium effects with the education subsidy because the income tax can do this equally well, while avoiding excessive investment in human capital. By allowing for general equilibrium effects, this paper demonstrates that the distributional equivalence between income taxes and education taxes is preserved. The linear income tax generates the same general equilibrium effects on the wage distribution as the linear education subsidy, while the latter avoids distortions in human capital investments. As a result, the efficiency results of Bovenberg and Jacobs (2005) derived in partial equilibrium carry over to the general equilibrium case.

In contrast, the optimal linear income tax is importantly affected by general equilibrium effects on wages. The assumption that skilled workers have a higher labor supply elasticity than unskilled workers implies that higher income taxes will reduce effective labor supply of the skilled workers relatively more than that of unskilled workers. Linear income taxation thus increases the skill premium, and before-tax income inequality increases. Consequently, general equilibrium effects on wages run against the distributional benefits of a higher income tax, and optimal income taxes are lowered as a result. Theoretically, the optimal linear income tax may even turn *negative* if the indirect general equilibrium effects on the pre-tax wage distribution are strong enough to offset the direct effects of higher marginal tax rates on the post-tax wage distribution. That optimal linear tax rates could be negative is an important new finding in the literature on optimal linear income taxation (Sheshinski 1972; Dixit and Sandmo 1977; Atkinson and Stiglitz 1980).

The second part of the paper considers the simultaneous setting of optimal *non-linear* tax and education policies. Optimal nonlinear policies differ fundamentally from linear policies when general equilibrium effects on wages are present. The key to understanding why they differ is that the relative effective labor supply can, by definition, be directly steered by the nonlinear tax and education schedules—even if all supply elasticities of labor and human capital are equal. By giving a marginal education subsidy on the high type, and a marginal education *tax* on the low type, the government can directly increase the relative supply of skilled human capital to lower the skill premium. Therefore, the first question—can education policy affect the relative supply of education in such a way that the skill premium falls?—can be answered affirmatively.

The second question is: Should education subsidies be optimally employed in an optimal redistributive program? The answer with nonlinear policies is yes. The skilled worker optimally faces a marginal subsidy on education, while the unskilled worker faces a marginal *tax* on education. The optimal marginal income tax on the skilled worker is *negative*, while the optimal marginal tax on the unskilled worker is positive, as in Stern (1982) and Stiglitz (1982). Hence, under nonlinear taxation, the efficiency results break down. We demonstrate that marginal education taxes/subsidies are directly linked to the top rate: marginal subsidies are zero when the top rate is zero. As the skilled worker faces a positive subsidy on both labor effort and education, and the unskilled worker faces a positive tax on both education and labor effort, the skill premium will be reduced. A skilled worker will be less tempted to mimic an unskilled worker, incentive compatibility constraints will be relaxed, and the government can

redistribute more income. In contrast to the linear policies, nonlinear instruments do exploit general equilibrium effects for redistribution.

We analyze the quantitative importance of general equilibrium effects by simulating optimal nonlinear income taxes and education subsidies. We find that the marginal top rate is negative, and rather small for plausible elasticities of substitution between skilled and unskilled labor, which confirms the findings of Stern (1982). Optimal education subsidies are not large either, since there is a direct link between the top rate and education subsidies. However, we demonstrate quantitatively that general equilibrium should be exploited for redistribution when the elasticity of substitution between skilled and unskilled labor is very low.

Tinbergen's suggestion to promote skill formation so as to provoke a decline in the skill premium for redistributive reasons is possible only under nonlinear policies. The case is lost under linear tax instruments. Intuitively, the generic linear education subsidy is an inefficient instrument to reduce the skill premium because it increases human capital supply of all agents simultaneously. Nonlinear education subsidies avoid this simultaneous increase in human capital supplies and can be tailored to increase supply of the high-ability types, while simultaneously lowering the supply of the low-ability types. Numerical simulations of optimal nonlinear policies demonstrate education policies can only be usefully employed for redistribution when the high-skilled and low-skilled workers are poor substitutes in production. However, for empirically plausible values of the elasticity of substitution between skilled and unskilled workers, it appears that education subsidies have only a limited role to play in an optimal tax-cum-education policy.²

This paper is related to a number of other papers. First, by allowing for endogenous wage rates, this paper contributes to a growing literature on optimal income tax and education policies; for earlier contributions, see, for example, Ulph (1977), Hare and Ulph (1979), Tuomala (1990), and Nielsen and Sørensen (1997). More recent contributions include Maldonado (2008), Alstadsaeter et al. (2008), Grossman and Poutvaara (2009), and Schindler (2011). Second, a number of papers analyze optimal taxation if different skill types earn endogenously determined wages, and the government cannot tax or subsidize all production inputs at different rates. Naito (1999, 2004, 2007) and Pirttilä and Tuomala (2001) show that both the Diamond and Mirrlees (1971) production efficiency theorem, and the Atkinson and Stiglitz (1976) theorem break down with endogenous wages. We also demonstrate that when wages are endogenous the zero commodity tax theorem is not applicable to education and production efficiency in human capital formation ceases to be optimal under nonlinear income taxation.

The rest of this paper is structured as follows. The next section presents the model with optimal linear income taxation and education policies in general equilibrium.

²From an empirical point of view, evidence is not particularly abundant either. Heckman et al. (1998) show in a dynamic overlapping generations model of skill formation for the US economy that long-run income inequality can hardly be changed through education policy. They show that after an initial positive impact of education subsidies on the relative supply of skilled labor, and thereby a reduction in the skill premium, the incentives to acquire human capital diminish, and the initial increase in relative supply of skilled workers is almost completely reversed in the long run.

Section 3 studies the same problem using nonlinear instruments. Section 4 simulates optimal nonlinear policies. Section 5 concludes. Appendix contains all the main derivations. Some non-essential derivations and more simulation results are available in an unpublished appendix.³

2 Model

The standard models of optimal income taxation with general equilibrium effects on wages are extended with human capital formation.⁴ This section presents the baseline model with linear instruments. Nonlinear instruments will be discussed later. Individuals differ in their capacity to accumulate human capital and earnings capacities of individuals are endogenous rather than exogenously given. Furthermore, individuals with higher ability have a comparative advantage in skill formation. A ‘one-shot’ model of human capital investments is analyzed. One may view this model as describing life-time investments in human capital, life-time labor supply and life-time consumption, where there are no inter-temporal distortions due to capital taxes or capital market failures, for example.⁵ To fully track down the general equilibrium impact of tax and education policies analytically, the analysis is restricted to two types, as in almost the entire literature.

2.1 Individuals

There is a unit mass of high-ability and low-ability workers who are indexed by $n = 1$ and $n = 2$, respectively. The fraction of high-ability workers is g_1 and the fraction of low-ability workers is g_2 . Each worker has an iso-elastic utility function $u(c_n, l_n)$, which is defined over consumption c_n and work effort l_n according to

$$u_n(c_n, l_n) \equiv c_n - \frac{l_n^{1+1/\varepsilon_n}}{1 + 1/\varepsilon_n}, \quad n = 1, 2, \quad (1)$$

where $\varepsilon_n > 0$ is the (un)compensated wage elasticity of labor supply of individual n . Since income effects are absent, compensated and uncompensated elasticities coincide, and labor supply is always upward sloping. This utility function is also used for its analytical simplicity in Diamond (1998), Saez (2001), and Naito (2004). This specification of utility is sufficiently general to stress the main points at stake, while not introducing additional analytical complexity due to income effects, as in Allen (1982). Moreover, it highlights the crucial role of different labor supply elasticities under linear policy instruments. Indeed, elasticities of labor supply are assumed to differ, and ε_1 is not equal to ε_2 . In the absence of income effects, different elasticities

³This appendix can be downloaded from the author’s website: <http://people.few.eur.nl/bjacobs>.

⁴See, for example, Feldstein (1973), Allen (1982), Stern (1982), and Stiglitz (1982).

⁵The importance of liquidity constraints for human capital formation is increasingly disputed on empirical grounds by, for example, Cameron and Taber (2004) and Carneiro and Heckman (2004).

of labor supply or human capital formation are necessary to obtain general equilibrium effects of policy. As human capital elasticities are assumed to be equal, see below, linear policy instruments would have no general equilibrium effects if labor supply elasticities would be identical.

Gross labor income z_n of each individual is

$$z_n \equiv w_n h_n l_n, \quad n = 1, 2, \tag{2}$$

where w_n denotes the gross wage rate per unit of human capital of an individual of skill type n , h_n is the level of human capital of each agent, i.e., the number of efficiency units of labor, and l_n denotes work effort. Total ‘effective labor supply’ thus equals $h_n l_n$, which encompasses both the quality of labor and the quantity of labor supply.

Human capital is accumulated on the intensive margin. Individuals invest e_n of their resources in education. One can think of e_n as the years enrolled in education or the quality of education where each individual has access to the same educational inputs, but transforms them differently into human capital depending on ability. The production function for human capital is given by

$$h_n \equiv a_n \phi_n(e_n), \quad \phi'(e_n) > 0, \quad \phi''(e_n) < 0, \quad n = 1, 2, \tag{3}$$

where a_n is the exogenous productivity of investment in human capital; $a_1 > a_2$, i.e., high-ability types have a comparative advantage in learning. High-ability types thus generate more human capital with the same amount of educational efforts because $\frac{\partial z_n}{\partial e_n \partial a_n} = w_n \phi'(e_n) l_n > 0$.⁶

To ensure that optimal education subsidies are zero in the absence of general equilibrium effects, the earnings function is weakly separable in ability, education, and labor. The elasticity of the production function is also assumed to be constant under linear policies, and is denoted by $\beta \equiv \frac{\phi'(e_n)e_n}{\phi(e_n)}$ (Jacobs and Bovenberg 2011). In general equilibrium, the high-ability type is assumed to earn a higher gross wage than the low-ability type so as to obtain an economically meaningful redistribution problem, i.e., $w_1 h_1 l_1 > w_2 h_2 l_2$. This assumption guarantees that gross labor earnings for the high-ability type are always larger than gross labor earnings of the low-ability type.

The price of one unit of education is denoted by p and is common for both individuals.⁷ All costs are assumed to be tax deductible since the major costs of education consist of taxed opportunity costs.⁸ Investments in education e_n are subsidized at flat rate s . Gross incomes z_n are taxed at a constant marginal rate t . In addition, every

⁶This form of comparative advantage is sufficiently rich to describe the main results of this paper. We may, however, allow the two types to accumulate both types of human capital, where the high-ability type is more productive in skilled human capital accumulation than the low-ability type. This will not qualitatively affect our main results. Naito (2004) also shows that his results in the more general specification carry over to the restricted specification.

⁷Assuming common costs of investment is without loss of generality, since each individual features a different marginal return on investments in human capital.

⁸One may, additionally, introduce untaxed direct costs, like tuition fees, but this slightly complicates the analysis without affecting our main results as long as all costs of education can be verified, and, therefore, subsidized, see also Bovenberg and Jacobs (2005).

individual may receive a non-individualized lump-sum transfer b . Hence, the income tax is progressive in the sense that average tax rates increase with income. The fundamental informational requirements to levy a linear income tax and to provide linear education subsidies are that aggregate gross incomes and aggregate investment in human capital must be verifiable by the government.

Consumption c_n equals total net labor income minus education expenditures:

$$c_n = (1 - t)(w_n a_n \phi(e_n) l_n - (1 - s) p e_n) + b, \quad n = 1, 2. \tag{4}$$

The first-order conditions for utility maximization yield the following constant elasticity labor supply functions for each individual

$$l_n = \left((1 - t) w_n a_n \phi(e_n) \right)^{\varepsilon_n}, \quad n = 1, 2. \tag{5}$$

Labor supply l_n increases with the net marginal wage rate and taxes depress labor supply. The first-order condition for optimal human capital investment is given by

$$w_n a_n \phi'(e_n) l_n = (1 - s) p, \quad n = 1, 2. \tag{6}$$

Marginal benefits of learning (the left-hand side) should be equal to the marginal costs of learning (the right-hand side). Subsidies increase investment in human capital. Taxes have no direct effect on learning because both marginal costs and marginal benefits are equally affected. Taxation does, however, reduce labor supply and lowers the returns of investments in human capital indirectly.

First-order conditions are necessary, but not sufficient. Second-order conditions are satisfied by imposing the following restriction on the parameters:⁹

$$\mu_n \equiv 1 - \beta(1 + \varepsilon_n) > 0, \quad n = 1, 2. \tag{7}$$

The tax elasticities of labor supply $\varepsilon_n^{lt} \equiv -\frac{\partial l_n}{\partial t} \frac{(1-t)}{l_n} = \frac{\varepsilon_n(1-\beta)}{\mu_n}$, and education $\varepsilon_n^{et} \equiv -\frac{\partial e_n}{\partial t} \frac{(1-t)}{e_n} = \frac{\varepsilon_n}{\mu_n}$ are important determinants of the optimal tax rates. The tax elasticity of labor earnings $z_n = w_n a_n \phi(e_n) l_n$ amounts to $\varepsilon_n^{zt} = \varepsilon_n^{lt} + \beta \varepsilon_n^{et} = \frac{\varepsilon_n}{\mu_n}$. Note that the tax elasticity of total labor earnings z_n equals the tax elasticity of effective labor supply h_n , since both wages and ability are exogenous to the individual. The tax elasticity of gross income $\frac{\varepsilon_n}{\mu_n}$ exceeds the wage elasticity of gross income ε_n . The reason is that the tax rate t reduces the after-tax wage $(1 - t)w_n a_n \phi(e_n)$ both directly (by raising the tax wedge between the before-tax wage and the after-tax wage t) and indirectly (by depressing the before-tax wage rate $w_n a_n \phi(e_n)$ through its negative impact on learning e_n). Learning is harmed indirectly because lower labor supply depresses the utilization rate of human capital. Similarly, the subsidy elasticities are given by: $\varepsilon_n^{ls} \equiv \frac{\partial l_n}{\partial s} \frac{(1-s)}{l_n} = \frac{\beta \varepsilon_n}{\mu_n}$, $\varepsilon_n^{es} \equiv \frac{\partial e_n}{\partial s} \frac{(1-s)}{e_n} = \frac{1}{\mu_n}$, and $\varepsilon_n^{zs} \equiv \frac{\partial z_n}{\partial s} \frac{(1-s)}{z_n} = \frac{\beta(1+\varepsilon_n)}{\mu_n}$.

⁹The derivations of the second-order conditions and the elasticities are available in a separate appendix, which can be downloaded from the author’s website: <http://people.few.eur.nl/bjacobs>.

2.2 Firms

There is one sector of production. A representative firm maximizes profits while taking wage rates w_1 and w_2 for each skill type as given. The firm produces output Y with a neoclassical production function, which features constant returns to scale in labor inputs H_1 and H_2 :

$$Y \equiv F(H_1, H_2), \tag{8}$$

where $F_n(\cdot) > 0$, $F_{nn}(\cdot) < 0$, $F_{12} \geq 0$, $n = 1, 2$, and subscript n refers to the n th argument of differentiation. The income share of the low-income earner is denoted by $\alpha \equiv \frac{w_2 H_2}{w_1 H_1 + w_2 H_2}$. $1 - \alpha \equiv \frac{w_1 H_1}{w_1 H_1 + w_2 H_2}$ is, therefore, the income share of the high-income earner. Further, if $\alpha < 1/2$, the low-ability type earns less than the high-ability type. $\sigma \equiv \frac{F_1(\cdot)F_2(\cdot)}{F_{12}(\cdot)F(\cdot)}$ denotes the partial elasticity of substitution between H_1 and H_2 in the production function $F(\cdot)$.

First-order conditions for profit maximization are necessary and sufficient, and given by

$$w_n = F_n(H_1, H_2), \quad n = 1, 2. \tag{9}$$

The skill premium π is the ratio of wages of skilled and unskilled workers, i.e., $\pi \equiv w_1/w_2$. With constant returns to scale in production, π is only a function of the relative supplies of skilled and unskilled workers, H_1/H_2 (using $f(H_1/H_2) \equiv F(H_1, H_2)/H_2$):

$$\pi(H_1/H_2) \equiv \frac{w_1}{w_2} = \frac{f'(H_1/H_2)}{f(H_1/H_2) - (H_1/H_2)f'(H_1/H_2)}, \tag{10}$$

where $\pi'(H_1/H_2) < 0$. The skill premium decreases if the relative supply of skilled workers, H_1/H_2 , increases.

Note that if both types have the same labor supply elasticity ε_n , all the tax and subsidy elasticities will be equal across individuals (see previous section). Hence, linear policy instruments cannot affect the skill premium π in that case.

2.3 General equilibrium

Labor market clearing requires that supply equals demand for each labor type:

$$H_n = h_n l_n g_n, \quad n = 1, 2. \tag{11}$$

Further, goods market equilibrium requires that total output equals total consumption, plus investments in human capital, plus exogenously given government expenditures Λ :

$$F(H_1, H_2) = (c_1 + pe_1)g_1 + (c_2 + pe_2)g_2 + \Lambda. \tag{12}$$

2.4 Government

The government maximizes a social welfare function over indirect utilities $v_n(b, t, s, w_n)$:

$$\omega_1 v_1(b, t, s, w_1)g_1 + \omega_2 v_2(b, t, s, w_2)g_2, \quad \omega_n \geq 0, \quad n = 1, 2, \quad (13)$$

where ω_n denotes the Pareto weight of type n in social welfare. The Pareto weights sum to one: $\omega_1 + \omega_2 = 1$. If $\omega_1 = \omega_2 = \frac{1}{2}$, the social welfare function is utilitarian, and there is no social preference for redistribution due to the constancy of marginal utility of income at the individual level (no income effects). $\omega_2 > \omega_1$ implies a social preference for redistribution.¹⁰

The government collects taxes to finance the lump-sum transfer, the education subsidies, and the exogenous revenue requirement Λ . The government budget constraint reads as:

$$\begin{aligned} & t(w_1 a_1 \phi(e_1)l_1 - (1 - s)pe_1)g_1 + t(w_2 a_2 \phi(e_2)l_2 - (1 - s)pe_2)g_2 \\ & = spe_1g_1 + spe_2g_2 + b + \Lambda. \end{aligned} \quad (14)$$

3 Optimal linear taxes and education subsidies

The government maximizes social welfare by optimally choosing the lump-sum transfer b , the linear marginal tax t and the linear education subsidy s . Formally, the following Lagrangian is maximized:¹¹

$$\begin{aligned} \max_{\{b,t,s\}} \mathcal{L} = & \omega_1 v_1(b, t, s, w_1)g_1 + \omega_2 v_2(b, t, s, w_2)g_2 \\ & + \eta(t(w_1 a_1 \phi(e_1)l_1 - (1 - s)pe_1)g_1 + t(w_2 a_2 \phi(e_2)l_2 - (1 - s)pe_2)g_2) \\ & - \eta(spe_1g_1 + spe_2g_2 + b + \Lambda), \end{aligned} \quad (15)$$

where η denotes the Lagrange multiplier of the government budget constraint, and the labor market clearing conditions have to be imposed, i.e., $w_n = F_n(H_1, H_2)$, where $H_n = h_n l_n g_n$, $n = 1, 2$.

The first-order condition for the optimal lump-sum transfer is

$$\frac{\omega_1}{\eta} g_1 + \frac{\omega_2}{\eta} g_2 = 1, \quad (16)$$

where we used Roy's lemma ($\frac{\partial v_n}{\partial b} = 1$), and $\frac{\omega_n}{\eta}$ is the social marginal utility of income of type n . The average social marginal benefits of a higher b (i.e., the left-hand side of (16)) should equal the costs in terms of a higher b (i.e., the right-hand side of (16)).

¹⁰Without loss of generality, social welfare weights are kept constant, but these may be derived from a concave function over private utilities.

¹¹Diamond and Mirrlees (1971) show that there is no guarantee that the second-order conditions for the government's maximization problem are automatically satisfied. We assume, however, that the first-order conditions are indeed sufficient to characterize the optimal solution.

With the aid of the first-order condition for b (16), the distributional characteristic ξ^z of labor income is defined as (minus) the normalized covariance between the social marginal utility of income $\frac{\omega_n}{\eta}$, and gross labor income z_n (Atkinson and Stiglitz 1980):

$$\xi^z \equiv - \frac{\sum_n \frac{\omega_n}{\eta} z_n g_n - \sum_n z_n g_n \sum_n \frac{\omega_n}{\eta} g_n}{\sum_n z_n g_n \sum_n \frac{\omega_n}{\eta} g_n} = \frac{\sum_n (1 - \frac{\omega_n}{\eta}) z_n g_n}{\sum_n z_n g_n}, \tag{17}$$

where the second equality follows from (16). With a positive distributional characteristic ξ^z , taxing labor income yields distributional benefits because the high-ability worker has a lower welfare weight than the low-ability worker, i.e., $\frac{\omega_1}{\eta} < \frac{\omega_2}{\eta}$, and earns a higher income, $z_1 > z_2$. Indeed, a zero distributional characteristic implies either that the government is utilitarian ($\omega_2 = \omega_1$), and not interested in redistribution, or that the marginal contribution to the tax base is equal for both ability types (i.e., taxable income z_n is the same for both types).

Similarly, the distributional characteristic of education ξ^e is defined as

$$\xi^e \equiv - \frac{\sum_n \frac{\omega_n}{\eta} e_n g_n - \sum_n e_n g_n \sum_n \frac{\omega_n}{\eta} g_n}{\sum_n \frac{\omega_n}{\eta} g_n \sum_n e_n g_n} = \frac{\sum_n (1 - \frac{\omega_n}{\eta}) e_n g_n}{\sum_n e_n g_n} = \xi^z. \tag{18}$$

A positive distributional characteristic implies that subsidizing (taxing) education results in distributional losses (gains). If education levels are equal for both workers, there is no educational inequality, and subsidizing education yields no distributional losses. The absence of a redistributive motive renders the distributional characteristic zero.

Note that the distributional characteristic of education is equal to the distributional characteristic of income because gross earnings are linear in education due to the constant elasticity of the production function for human capital. This assumption ensures that (negative) education subsidies are distributionally equivalent to income taxes, see also Jacobs and Bovenberg (2011). In the remainder, the superscripts are dropped, and $\xi \equiv \xi^z = \xi^e$.

Simplifying the first-order condition for t yields the optimal linear income tax at optimal education policy (see Appendix)

$$\frac{t}{1-t} = \frac{1}{\bar{\varepsilon}^{zt}} \left(\xi - \left(\frac{\omega_2}{\eta} - \frac{\omega_1}{\eta} \right) \left(\frac{\varepsilon_1}{\mu_1} - \frac{\varepsilon_2}{\mu_2} \right) \frac{\alpha(1-\alpha)}{(1-\beta)\sigma} \right), \tag{19}$$

where $\bar{\varepsilon}^{zt} = ((1-\alpha)\varepsilon_1^{lt} + \alpha\varepsilon_2^{lt})/(1-\beta)$ is the income weighted average of the tax elasticity of total labor earnings (i.e., effective labor supply). This expression generalizes the standard expression for the optimal linear income tax by allowing for human capital formation and general equilibrium effects on wages (Sheshinski 1972; Dixit and Sandmo 1977; Atkinson and Stiglitz 1980).

The optimal income tax formula shows the trade-off between equity (numerator) and efficiency (denominator). The larger the social preference for redistribution, the larger the ξ , and the higher the optimal marginal income tax. If both groups have an equal weight in social welfare ($\omega_2 = \omega_1$), the optimal marginal income tax is zero because $\xi = 0$ and $\frac{\omega_2}{\eta} = \frac{\omega_1}{\eta}$. The larger the income-weighted average elasticity of

labor earnings, the lower the optimal linear income tax because the labor-income tax more heavily distorts labor supply. Both the elasticities of labor supply and human capital formation determine the effective elasticity of earnings, which is due to the feedback between labor supply and human capital formation; see also Bovenberg and Jacobs (2005) for a more elaborate discussion.

The expression for the optimal income tax in general equilibrium differs from its partial-equilibrium counterpart due to the presence of $(\frac{\omega_2}{\eta} - \frac{\omega_1}{\eta})(\frac{\varepsilon_1}{\mu_1} - \frac{\varepsilon_2}{\mu_2})\frac{\alpha(1-\alpha)}{(1-\beta)}\frac{1}{\sigma}$ (Bovenberg and Jacobs 2005). This term measures the distributional losses (or gains) arising from general equilibrium effects on wages, and are subtracted from the direct welfare gains of higher income taxes in reducing inequality (as measured by ξ). The optimal marginal linear income tax is lower with general equilibrium effects than without if: (i) substitution between labor types is finite ($\sigma < \infty$), (ii) the (uncompensated) earnings elasticity of the skilled worker is larger than that of the unskilled type ($\varepsilon_1^{zt} = \frac{\varepsilon_2}{\mu_2} > \varepsilon_2^{zt} = \frac{\varepsilon_2}{\mu_2}$).

The effects of imperfect substitution in labor types have been subject of most of the papers in this field, so these results do not come as a surprise. If labor types are perfect substitutes ($\sigma = \infty$), the last term in the expression of the optimal income tax (19) vanishes. In this case, wage rates per hour worked are not affected by changes in relative factor supplies, and the wage distribution is exogenous.

The second condition on the relative sizes of labor supply elasticities has received no attention so far. The formula demonstrates that general equilibrium effects only contribute to more equality if high-skilled labor earnings respond *less* elastically to an increase in taxes than low-skilled labor earnings ($\varepsilon_1^{zt} = \frac{\varepsilon_1}{\mu_1} < \varepsilon_2^{zt} = \frac{\varepsilon_2}{\mu_2}$). In that case, higher taxes raise relative supply of skilled labor (H_1/H_2), the skill premium $\pi(H_1/H_2)$ declines, see (10), and smaller before-tax wage inequality results. If, however, skilled labor supply responds *more* elastically to taxes than unskilled labor supply ($\varepsilon_1^{zt} = \frac{\varepsilon_1}{\mu_1} > \varepsilon_2^{zt} = \frac{\varepsilon_2}{\mu_2}$), general equilibrium effects work against redistribution of incomes through income taxes by *increasing* before-tax income inequality as the relative supply of skilled labor declines.

If the compensated elasticities of effective total labor supply ε_1^{zt} and ε_2^{zt} are identical ($\varepsilon_1^{zt} = \varepsilon_2^{zt}$ so that $\frac{\varepsilon_1}{\mu_1} = \frac{\varepsilon_2}{\mu_2}$ and $\varepsilon_1 = \varepsilon_2$), the general equilibrium term vanishes as well. The intuition is straightforward. With equal elasticities, linear taxation does not affect relative labor supply. If relative effective labor supply (H_1/H_2) remains constant, relative wages (w_1/w_2) remain constant as well, cf. the skill premium (10). This is the case, for example, if all individuals have identical preferences ($\varepsilon_1 = \varepsilon_2$). Hence, this example illustrates the necessity of heterogeneous preferences (in the absence of income effects) for our model to make sense in a general equilibrium context.¹²

A higher elasticity of skilled labor supply ($\varepsilon_1 > \varepsilon_2$) is a necessary requirement for education subsidies to work in favor of redistribution via general equilibrium effects under linear policy instruments. Intuitively, only if the high-skilled agent

¹²Using identical utility functions, Allen (1982) found that general equilibrium effects may result in both lower and higher optimal linear income tax rates compared to the case with exogenous wages. Higher optimal linear taxes can be found if income effects in labor supply are sufficiently strong. Even when skilled and unskilled workers have identical preferences, income effects can drive the uncompensated wage elasticity of labor supply of skilled workers below the uncompensated labor supply elasticity of unskilled workers, and general equilibrium effects work in favor of redistribution.

has a higher elasticity of human capital investments with respect to subsidies (i.e., $\varepsilon_1^{es} = \frac{1}{1-\beta(1+\varepsilon_1)} > \frac{1}{1-\beta(1+\varepsilon_2)} = \varepsilon_2^{es}$), subsidies on human capital can compress the wage distribution by boosting high-skilled human capital more than low-skilled human capital. Hence, we will assume that $\varepsilon_1 > \varepsilon_2$. Also from an empirical point of view, a larger elasticity of skilled labor earnings relative to unskilled labor earnings is plausible (Gruber and Saez 2003).

The expression for the optimal income tax (19) suggests that optimal marginal income taxes may even become negative if general equilibrium effects are strong enough and elasticities of effective total labor supply are sufficiently diverging. In that case, subsidizing work effort provokes such strong general equilibrium effects that low-ability workers are better off by paying subsidies to the high-ability workers; their before tax wages increase more than is needed to offset the rise in lump-sum taxes to finance the subsidies on work. Jacobs (2007) shows that this can happen even under not very unrealistic values for the elasticities of the model.

The general solution for optimal education subsidies in the presence of general equilibrium effects at the optimal tax system is (see Appendix)

$$s = 0. \quad (20)$$

The government sets the subsidies on education to zero. Therefore, educational investments are efficient in the presence of an optimal income tax. The intuition is similar as in Bovenberg and Jacobs (2005) and Jacobs and Bovenberg (2011).

Education subsidies do not reduce the labor supply distortion more compared to an equally costly reduction in income taxes. Intuitively, the constant elasticity in the human capital production function ensures that there is a linear relation between e_n and z_n . Hence, subsidizing e_n is equivalent to subsidizing z_n . However, besides distorting labor supply, subsidies on education also distort human capital investment. This distortion in education (over-investment) can be avoided by not subsidizing education. Therefore, the government does not want to use education subsidies to reduce the tax wedge on labor supply.

The government is also indifferent between education taxes and income taxes to redistribute income. Again, because education is linear in income, any redistribution that education taxes can achieve, can also be achieved with income taxes. Indeed, the distributional characteristics of income (ξ^z) and education (ξ^e) are equal. While both taxes on income and education reduce labor effort, taxing education e_n additionally causes under-investment in human capital. The income tax does not directly distort human capital investment because costs and benefits of education are equally affected by the marginal tax rate. The government can therefore avoid distortions in human capital accumulation by taxing income instead of taxing education to redistribute resources.

In contrast to Bovenberg and Jacobs (2005), both subsidies and income taxes do cause general equilibrium effects. Nevertheless, their combined effect cancels in the expression for optimal education subsidies, since general equilibrium effects do not upset the linear relationship between earnings and education. Intuitively, (lower) income taxes can, in principle, achieve the same wage compression as (higher) education subsidies without causing inefficiencies in human capital formation. Hence, education subsidies remain distributionally equivalent to income taxes even in the

presence of general equilibrium effects. For the same reasons, education subsidies still cannot be used as implicit taxes on leisure. Optimal subsidies thus ensure efficiency in human capital formation. Thus, from this discussion follows that education subsidies should not be employed to compress the wage distribution in the optimal redistributive program.

Due to the specific form of the earnings function and the linearity of policy instruments, the government can effectively separate production decisions from consumption choices, also in general equilibrium. By ensuring that all costs of education are effectively tax deductible, the government has access to a perfect profit tax to tax the rents from ability in human capital formation. Hence, the Diamond and Mirrlees (1971) production efficiency theorem still applies as in Bovenberg and Jacobs (2005).

The results in this section depend on specific assumptions regarding the utility function and the production function for human capital. Allowing for income effects (Allen 1982) and non-constant elasticities in the human capital production function will generally result in a role for general equilibrium effects determining optimal education policies. However, non-constant elasticities would give a role for education policies even in the absence of general equilibrium effects (Maldonado 2008; Jacobs and Bovenberg 2011). Moreover, income effects in labor supply would generate an additional channel (through the feedback of labor supply with human capital formation) whereby general equilibrium effects affect optimal education policies. We have switched off this (second-order) feedback by assuming that the income effects in labor supply are identical among individuals (i.e., both have a zero income effect). As long as income elasticities in labor supply are identical, not necessarily zero, the results of the current section remain valid. Hence, our findings are a reasonable first-order approximation, given that there is no compelling evidence indicating that income elasticities in labor supply and human capital elasticities differ much among skill types.

4 Optimal nonlinear taxes and subsidies

This section derives the optimal nonlinear tax and education policies. We maintain the assumption that the earnings function is weakly separable in ability, labor, and education, but we can now allow for general utility and production functions for human capital. These conditions still ensure that optimal education subsidies are zero in the absence of general equilibrium effects on wages (Jacobs and Bovenberg 2011). All results presented in this section are qualitatively the same with the specifications for the utility and the human capital production functions that we used to study the linear tax case. We also provide simulations of the optimal nonlinear taxes and subsidies, while maintaining the same structure on preferences and technologies as in the linear case.

The optimal nonlinear tax and subsidy rates are found by deriving the optimal second-best allocation of consumption, gross income and education, as in Stern (1982) and Stiglitz (1982). By using the first-order conditions for individual optimization, we can compute the optimal marginal income taxes and optimal marginal education subsidies that would decentralize the optimal second-best allocation.

Any solution to the optimal second-best problem has to satisfy the incentive compatibility constraints stating that each individual n has a weakly higher utility choosing the bundle (c_n, z_n, e_n) of consumption, gross income and education, which is intended for them, than utility of choosing the alternative bundle of (c_m, z_m, e_m) , which is intended for the other individual m ($n, m = 1, 2, n \neq m$). That is,

$$u_1(c_1, l_1) = u_1\left(c_1, \frac{z_1}{w_1 a_1 \phi(e_1)}\right) \geq u_1\left(c_2, \frac{z_2}{w_1 a_1 \phi(e_2)}\right) = u_1\left(c_2, \frac{w_2 a_2}{w_1 a_1} l_2\right), \quad (21)$$

$$u_2(c_2, l_2) = u_2\left(c_2, \frac{z_2}{w_2 a_2 \phi(e_2)}\right) \geq u_2\left(c_1, \frac{z_1}{w_2 a_2 \phi(e_1)}\right) = u_2\left(c_1, \frac{w_1 a_1}{w_2 a_2} l_1\right). \quad (22)$$

The government maximizes the social welfare function (13) subject to the economy’s resource constraint (12), and incentive compatibility constraint (21). Under normal circumstances, the second incentive constraint (22) is not binding at an optimal solution, and it will be ignored in the remainder (Stern 1982; Stiglitz 1982). Assuming for simplicity that the price of a unit of education is one, and the number of high-ability and low-ability persons are both equal to one, the following Lagrangian for the maximization of social welfare is formulated¹³

$$\begin{aligned} \max_{\{c_1, l_1, e_1, c_2, l_2, e_2\}} \mathcal{L} &= \omega_1 u_1(c_1, l_1) + \omega_2 u_2(c_2, l_2) \\ &+ \eta (F(a_1 \phi(e_1) l_1, a_2 \phi(e_2) l_2) - e_1 - e_2 - c_1 - c_2 - \Lambda) \\ &+ \theta \left(u_1(c_1, l_1) - u_1^* \left(c_2, \pi^{-1} \left(\frac{a_1 \phi(e_1) l_1}{a_2 \phi(e_2) l_2} \right) \frac{a_2}{a_1} l_2 \right) \right), \quad (23) \end{aligned}$$

where the inverse of the skill premium is denoted by π^{-1} , see (10). The conditions for labor market equilibrium (11) are substituted. To avoid confusion in notation, utility of the high-skilled type mimicking the low-skilled type is designated by $u_1^* \equiv u_1(c_2, \frac{w_2 a_2}{w_1 a_1} l_2)$. θ is the Lagrange multiplier on the incentive compatibility constraint.

First-order conditions for an optimal allocation are

$$\frac{\partial \mathcal{L}}{\partial c_1} = \omega_1 \frac{\partial u_1}{\partial c_1} - \eta + \theta \frac{\partial u_1}{\partial c_1} = 0, \quad (24)$$

$$\frac{\partial \mathcal{L}}{\partial l_1} = \omega_1 \frac{\partial u_1}{\partial l_1} + \eta a_1 \phi(e_1) F_1(\cdot) + \theta \frac{\partial u_1}{\partial l_1} - \theta \frac{\partial u_1^*}{\partial l_2} l_2 \frac{a_2}{a_1} \frac{\partial \pi^{-1}(\cdot)}{\partial l_1} = 0, \quad (25)$$

$$\frac{\partial \mathcal{L}}{\partial e_1} = \eta (a_1 \phi'(e_1) l_1 F_1(\cdot) - 1) - \theta \frac{\partial u_1^*}{\partial l_2} l_2 \frac{a_2}{a_1} \frac{\partial \pi^{-1}(\cdot)}{\partial e_1} = 0, \quad (26)$$

$$\frac{\partial \mathcal{L}}{\partial c_2} = \omega_2 \frac{\partial u_2}{\partial c_2} - \eta - \theta \frac{\partial u_1^*}{\partial c_2} = 0, \quad (27)$$

¹³Extensions with p differing from one and different fractions g_n of both types are straightforward and do not provide additional insights.

$$\frac{\partial \mathcal{L}}{\partial l_2} = \omega_2 \frac{\partial u_2}{\partial l_2} + \eta a_2 \phi(e_2) F_2(\cdot) - \theta \frac{\partial u_1^*}{\partial l_2} \frac{a_2}{a_1} \left(\pi^{-1}(\cdot) - l_2 \frac{\partial \pi^{-1}(\cdot)}{\partial l_2} \right) = 0, \tag{28}$$

$$\frac{\partial \mathcal{L}}{\partial e_2} = \eta (a_2 \phi'(e_2) l_2 F_2(\cdot) - 1) - \theta \frac{\partial u_1^*}{\partial l_2} l_2 \frac{a_2}{a_1} \frac{\partial \pi^{-1}(\cdot)}{\partial e_2} = 0. \tag{29}$$

From the first-order conditions for c_1 (24) and l_1 (25) follows the marginal tax rate, $T'_1 \equiv 1 + \frac{1}{w_1 a_1 \phi(e_1)} \frac{\partial u_1 / \partial l_1}{\partial u_1 / \partial c_1}$, on the high-ability type (see Appendix):¹⁴

$$T'_1 = \frac{\theta}{\eta} \frac{1}{\sigma} \frac{\alpha}{(1 - \alpha)} \frac{1}{w_1 a_1 \phi(e_2)} \frac{\partial u_1^*}{\partial l_2} < 0. \tag{30}$$

Consequently, the marginal top rate is exploited for redistribution with general equilibrium effects on wages. Indeed, the general expression of the optimal marginal tax on the skilled type is almost identical to the one without human capital formation (Stern 1982; Stiglitz 1982). The basic intuition of these papers carries over to the present case. A marginal subsidy on work for the high-ability type increases relative supply of skilled human capital, and lowers the skill premium π . Hence, the utility costs of mimicking the low-ability type increase, as it takes the skilled type more labor time to mimic the unskilled type. In the absence of general equilibrium effects on wages ($\sigma = \infty$), there is no distortion at the top, and the marginal tax rate is zero; see also Seade (1977). Further, if the government is not interested in redistribution, the incentive constraint is slack, θ is zero, and optimal marginal taxes are zero as well.

Manipulation of the first-order condition for c_2 (27) and the first-order condition for l_2 (28) yields the marginal tax on labor of the low-ability type $T'_2 \equiv 1 + \frac{1}{w_2 a_2 \phi(e_2)} \frac{\partial u_2 / \partial l_2}{\partial u_2 / \partial c_2}$ (see Appendix):

$$\frac{T'_2}{1 - T'_2} = \frac{\theta}{\eta} \frac{\partial u_1^*}{\partial c_2} + \frac{T'_1}{(1 - T'_2)} \frac{(1 - \alpha)}{\alpha} (\sigma - 1). \tag{31}$$

In the absence of general equilibrium effects, the marginal tax rate on the low-ability type is unambiguously positive, since the top rate is zero ($T'_1 = 0$) in that case (Mirrlees 1971; Stiglitz 1982). The presence of general equilibrium effects (lower σ) increases the marginal tax rate T'_2 on the low type (recall $T'_1 < 0$). This is intuitive, since a higher marginal tax rate on the low type reduces their labor supply, and therefore results in more before-tax wage equality. Hence, the high type is less tempted to mimic the low type, which relaxes the incentive compatibility constraint, and therefore results in more redistribution.

The optimal marginal education subsidy rate for the high type is derived from combining the first-order conditions for e_1 (26) with the optimal marginal income tax for skilled workers (30). Then, we find that the optimal marginal education subsidy for high-ability workers $S'_1 \equiv 1 - w_1 a_1 \phi'(e_1) l_1$ is positive (see Appendix):

$$\frac{S'_1}{1 - S'_1} = -\frac{\theta}{\eta} \frac{1}{\sigma} \frac{\alpha}{(1 - \alpha)} \frac{1}{w_1 a_1 \phi(e_2)} \frac{\partial u_1^*}{\partial l_2} = -T'_1 > 0. \tag{32}$$

¹⁴With only two types, the tax schedule is not continuous and generally not differentiable. As such, one cannot equate $1 + \frac{1}{a_1 \phi(e_1) F_1} \frac{\partial u_1 / \partial l_1}{\partial u_1 / \partial c_1}$ with the derivative of a tax schedule $T(z_1)$ at income point z_1 .

The intuition is that a marginal subsidy on human capital investment of the high type (like the marginal subsidy on work) lowers the skill premium, and makes it more costly for the high-ability type to mimic the low-ability type.

Similarly, we find the optimal marginal subsidy rate on education for the low-ability type $S'_2 \equiv 1 - w_2 a_2 \phi'(e_2) l_2$ from the first-order condition for e_2 (29) and the tax rate (31) (see Appendix):

$$\frac{S'_2}{1 - S'_2} = \frac{\theta}{\eta} \frac{1}{\sigma} \frac{\alpha}{(1 - \alpha)} \frac{1}{w_2 a_1 \phi(e_2)} \frac{\partial u_1^*}{\partial l_2} = \frac{(1 - \alpha)}{\alpha} T'_1 < 0. \tag{33}$$

Therefore, education is taxed on the margin for the low-ability type. Again, the mechanism is that general equilibrium effects relax the incentive compatibility constraint, and the government can redistribute more income.

Note that the expressions for the marginal education subsidies are all directly related to the top rate. Indeed, subsidies or taxes on education are larger if the top rate is lower, i.e., when general equilibrium effects are more important (lower σ), and if the government wishes to redistribute incomes more heavily (larger θ). If labor types are perfect substitutes ($\sigma = \infty$), the optimal marginal education subsidies for both high and low-skilled workers are zero, i.e., $S'_1 = S'_2 = 0$, cf. Bovenberg and Jacobs (2005).

General equilibrium effects on the skill premium should indeed be exploited for redistribution under nonlinear policies, in contrast to optimal linear policies. The intuition is that, by using nonlinear instruments, the government can directly influence the skill premium $\pi(H_1/H_2)$ by setting different marginal tax and subsidy rates for each worker, as long as the policy remains incentive compatible. This holds true irrespective of preferences or technologies. Hence, by optimally employing marginal subsidies on the high-ability type and marginal taxes on the low-ability type, the skill premium falls, the incentive compatibility constraint is relaxed, and the policy achieves more income redistribution.

Indirect taxes/subsidies, such as education subsidies, are not optimally zero under nonlinear income taxation with weakly separable preferences. This bolsters the findings by Naito (1999, 2004, 2007) and Pirttilä and Tuomala (2001) who investigated the desirability of nonzero commodity taxes and deviations from production efficiency in general equilibrium settings with endogenous wages. In the current model, education should optimally be taxed or subsidized under nonlinear income taxation to exploit factor price changes for redistribution, and the Atkinson and Stiglitz (1976) theorem does not apply to education subsidies. Moreover, since education choices are generally not efficient, the Diamond–Mirrlees production efficiency theorem ceases to apply to individual human capital production as well. The intuition is that, due to the nonlinearity of the policy schedules, the optimal policies do not constitute a perfect profit tax on the rents from human capital formation. As a result, consumption and investment choices cannot be (weakly) separated, which is a prerequisite for the production efficiency theorem to apply. In the absence of general equilibrium effects, education would not be taxed or subsidized on a net basis, and would be efficient.

5 Simulations of optimal nonlinear policies

To check whether general equilibrium effects can be quantitatively important for optimal nonlinear tax and education policies, we simulate the model following Stern (1982). We resort to the standard iso-elastic utility function with a constant wage elasticity of labor supply, which is augmented with a preference parameter δ to calibrate the dis-utility of labor supply:

$$u_n(c_n, l_n) \equiv c_n - \delta \left(\frac{l_n^{1+1/\varepsilon_n}}{1+1/\varepsilon_n} \right), \quad n = 1, 2. \quad (34)$$

In the baseline, we set the wage elasticities of labor supply equal for both types at $\varepsilon_1 = \varepsilon_2 = 0.2$, cf. Blundell and MaCurdy (1999). The preference parameter for leisure is calibrated at $\delta = 10$, so as to keep labor effort of both types between 0 and 1. The production function for human capital is Cobb–Douglas, $a_n \phi(e_n) \equiv a_n e_n^\beta$, with an elasticity $\beta = 0.2$. This is probably a lower bound.¹⁵ The ability parameters are set at $a_1 = 10$ and $a_2 = 8$. The aggregate production function features a constant elasticity of substitution between skilled and unskilled labor:

$$Y = (\gamma H_1^\rho + (1 - \gamma) H_2^\rho)^{1/\rho}, \quad \sigma \equiv \frac{1}{1 - \rho}. \quad (35)$$

The elasticity of substitution between high and low-skilled labor is set at $\sigma = 1.5$. The survey provided in Katz and Autor (1999) suggests values for σ between 1 and 2.5.¹⁶

The share parameter is calibrated in the baseline at $\gamma = 0.63$ to get an income share of skilled labor in total output of $1 - \alpha = 0.65$ in the absence of government intervention Stern (1982). The baseline welfare weights in the social welfare function are $\omega_1 = 1 - \omega_2 = 0.25$. Setting $\omega_1 = \omega_2 = 0.5$ corresponds to a utilitarian criterion, which is non-redistributive because the private marginal utility of income is unity. $\omega_1 = 1 - \omega_2 = 0$ corresponds to a Rawlsian social welfare function. At the baseline simulation, the government revenue requirement is set at zero: $\Lambda = 0$. The simulation results are given in Table 1.

Calculations by Stern (1982) in models without endogenous human capital formation reveal that general equilibrium effects have little impact on simulated optimum tax rates. We largely confirm this in all simulations. The top rate is indeed only

¹⁵Standard Mincer wage equations regress the logarithm of gross earnings on years of education and commonly find an estimate of 5–10 % (Card 1999). By log-differentiating gross earnings z with respect to e , we can interpret β/e as a measure for the Mincer return. For example, if high-skilled individuals have 4 years more education ($e = 4$), and the Mincer return is 7.5 %, then $\beta = 0.3$.

¹⁶More recent empirical evidence confirms these estimates. For example, Goldin and Katz (2007) estimate a value of $\sigma = 1.64$. Ciccone and Peri (2005) estimate $\sigma = 1.5$. Blankenau and Cassou (2011) find similar estimates using aggregate data and even (much) higher elasticities using sectoral data. In a separate appendix, all simulations are presented as well for $\sigma = 0.5$ and $\sigma = 2.5$, since the elasticity of substitution is the key elasticity determining general equilibrium effects on the wage structure. The appendix can be downloaded from the author's website: <http://people.few.eur.nl/bjacobs>. The conclusion that general equilibrium effects have only a limited impact on optimal policies is robust to these variations in σ .

Table 1 Optimal non-linear tax and education policies

σ	∞	5	2.5	2	1.5	0.99	0.5
γ	0.54	0.57	0.59	0.6	0.63	0.67	0.8
T_1'	0.0 %	-0.3 %	-0.8 %	-1.0 %	-1.1 %	-1.9 %	-2.7 %
T_2'	31.0 %	31.6 %	31.8 %	32.0 %	32.6 %	33.4 %	35.3 %
S_1'	0.0 %	0.3 %	0.7 %	1.0 %	1.1 %	1.9 %	2.6 %
S_2'	0.0 %	-0.7 %	-1.5 %	-2.0 %	-2.3 %	-4.0 %	-6.2 %
ε_1	0.75	0.5	0.35	0.3	0.25	0.2	0.15
T_1'	-13.2 %	-8.0 %	-4.5 %	-3.2 %	-2.1 %	-1.1 %	-0.4 %
T_2'	24.8 %	28.0 %	30.3 %	31.1 %	31.9 %	32.6 %	33.1 %
S_1'	11.6 %	7.4 %	4.3 %	3.1 %	2.1 %	1.1 %	0.4 %
S_2'	-34.6 %	-19.0 %	-9.9 %	-7.0 %	-4.5 %	-2.3 %	-0.8 %
ε_2	0.75	0.5	0.35	0.3	0.25	0.2	0.15
T_1'	-0.4 %	-0.6 %	-0.8 %	-0.9 %	-1.0 %	-1.1 %	-1.3 %
T_2'	33.0 %	32.9 %	32.8 %	32.7 %	32.6 %	32.6 %	32.5 %
S_1'	0.4 %	0.6 %	0.8 %	0.9 %	1.0 %	1.1 %	1.2 %
S_2'	-1.0 %	-1.4 %	-1.8 %	-1.9 %	-2.1 %	-2.3 %	-2.6 %
β	0.4	0.3	0.25	0.2	0.15	0.1	
T_1'	-2.4 %	-1.6 %	-1.3 %	-1.1 %	-1.0 %	-0.8 %	
T_2'	31.6 %	32.2 %	32.4 %	32.6 %	32.7 %	32.8 %	
S_1'	2.4 %	1.6 %	1.3 %	1.1 %	0.9 %	0.8 %	
S_2'	-5.6 %	-3.5 %	-2.8 %	-2.3 %	-2.0 %	-1.7 %	
ω_1	0.5	0.4	0.3	0.2	0.1	0	
T_1'	0.0 %	-0.5 %	-0.9 %	-1.3 %	-1.7 %	-2.1 %	
T_2'	0.0 %	16.3 %	27.9 %	36.7 %	43.5 %	48.9 %	
S_1'	0.0 %	0.5 %	0.9 %	1.3 %	1.7 %	2.1 %	
S_2'	0.0 %	-1.0 %	-1.9 %	-2.8 %	-3.7 %	-4.6 %	

slightly negative, even with endogenous learning. Since the expressions for optimal education subsidies are all directly related to the top rate, we see that the general equilibrium impact on optimum nonlinear subsidies is rather limited as well.

As expected, the optimal tax expressions are most sensitive to changes in the elasticity of substitution σ . In the top rows of Table 1, we calibrate the parameter γ so as to keep the income share of skilled labor fixed at $1 - \alpha = 0.65$ in the absence of government intervention.¹⁷ We find that marginal education subsidies and taxes are substantially differing from zero when the elasticity of substitution between skilled and unskilled labor falls below unity, which is empirically less plausible.

¹⁷Otherwise, the condition that the skilled wage is always higher than the unskilled wage may be violated, and the redistribution problem will become misspecified.

Further, the elasticity of high-ability labor supply ε_1 is important in explaining the pattern of marginal taxes and subsidies. It is a crucial determinant of the incentive compatibility constraint. The higher this elasticity, the easier it is for the high type to mimic the low type, and the larger are marginal subsidies on high-ability labor supply and education. From the table follows that the main results are not driven by the labor supply elasticity of the low type, nor by the human capital elasticity, or the social welfare function.¹⁸

6 Conclusions

This paper analyzed the simultaneous setting of optimal linear and nonlinear income taxes and education subsidies in two-type models with endogenous labor supply, endogenous human capital formation, and endogenous wage rates. To investigate the potential role of general equilibrium effects in shaping optimal linear tax and education policies, we ensured that optimal subsidies on education are zero in the absence of general equilibrium effects on wages. This required weakly separable earnings functions, and, for linear instruments only, a constant elasticity in the human capital production function. For linear education subsidies to work in favor of redistribution, we further assumed that labor supply of the high-ability ('skilled') type is more elastic than labor supply of the low-ability ('unskilled') type. Linear taxes and subsidies cannot—by assumption—affect the skill premium if all behavioral elasticities are equal.

We showed that optimal linear education subsidies are zero, even if linear tax and education policies have the potential to provoke equilibrium effects on wage rates. The intuition is that linear income taxes are distributionally equivalent to (negative) linear education subsidies, and more efficient because income taxes do not generate distortions in human capital formation, whereas linear subsidies cause overinvestment. This holds true whether general equilibrium effects are present or not. The optimal linear income tax is, however, lowered due to general equilibrium effects on wages if skilled labor supply is more elastic than unskilled labor supply. A higher income tax increases the skill premium because skilled labor supply falls more than unskilled labor supply. These general equilibrium effects work against the direct redistributive gains of a higher income tax rate. The optimal linear income tax rate may even turn negative when general equilibrium effects on wages are sufficiently strong.

The results for optimal nonlinear policies are found to be fundamentally different. With nonlinear instruments, the government can directly affect the relative supply of skilled human capital using specific instruments, such as marginal subsidies on human capital or labor supply of the high type, and marginal taxes on human capital or labor supply of the low type. Consequently, one does not need to impose restrictions on preferences or the production function for human capital to obtain an

¹⁸Due to the absence of income effects, all reported results are completely insensitive to changing the government revenue requirement.

impact of nonlinear policies on the skill premium. The nonlinearity of tax and subsidy schedules is sufficient. Optimal nonlinear policies do exploit general equilibrium effects on wages for redistributive purposes. The skilled worker optimally faces marginal subsidies on both work effort and education, whereas the unskilled worker optimally faces marginal taxes on work and education. As a result, wage differences are reduced, and the incentive compatibility constraint is relaxed because the skilled worker finds it harder to mimic the unskilled worker. However, simulations of optimal nonlinear policies revealed that the impact of general equilibrium effects on optimal policies is modest. Only at low levels of the elasticity of substitution between skilled and unskilled labor, marginal subsidies (taxes) on skilled (unskilled) education, and marginal subsidies on skilled work are found to be substantially positive.

In future research, the current analysis can be cast in a model with two production sectors, each exhibiting different factor intensities, so as to investigate the desirability of aggregate production efficiency, and the optimality of zero commodity taxation. Our conjecture is that deviations from aggregate production efficiency, and nonzero commodity will be optimal, as Naito (1999, 2004, 2007) and Pirttilä and Tuomala (2001) have demonstrated in similar settings, but without education policies and human capital investment. However, it remains unclear how optimal education policies will be affected. The current two-type analysis of nonlinear income taxation may also be extended to a setting with a continuum of skill types in order to further investigate how factor prices should be exploited for redistribution under nonlinear income taxation with more realistic skill distributions, as in, for example, Diamond (1998) and Saez (2001). Nevertheless, the expressions for the incentive compatibility constraints reveal that each of them is dependent on the entire wage distribution, and therefore on all the actions of all other agents. As a consequence, one may need to resort to numerical simulations. Our results under linear income taxation will also change when more general utility or earnings functions are used to analyze the importance of income effects, and to allow for the possibility that education has a varying degree of complementarity with work effort (Jacobs and Bovenberg 2011). Also, extensions with imperfections in labor markets due to for example minimum wages, search frictions, unions, and efficiency wages may be interesting avenues for future research.

Open Access This article is distributed under the terms of the Creative Commons Attribution License which permits any use, distribution, and reproduction in any medium, provided the original author(s) and the source are credited.

Appendix

Optimal linear policies

The first-order condition for the optimal linear income tax is given by

$$\begin{aligned} \frac{\partial \mathcal{L}}{\partial t} = & - \sum_n \omega_n (w_n a_n \phi(e_n) l_n - p(1-s)e_n) g_n \\ & + \eta \sum_n (w_n a_n \phi(e_n) l_n - p(1-s)e_n) g_n \end{aligned}$$

$$\begin{aligned}
 & + \eta t \sum_n \left(w_n a_n \phi(e_n) \frac{\partial l_n}{\partial t} \right) g_n - \eta s \sum_n p \frac{\partial e_n}{\partial t} g_n \\
 & + (\omega_1(1-t)H_1 + \eta t H_1) \left(\frac{\partial w_1}{\partial H_1} \frac{\partial H_1}{\partial t} + \frac{\partial w_1}{\partial H_2} \frac{\partial H_2}{\partial t} \right) \\
 & + (\omega_2(1-t)H_2 + \eta t H_2) \left(\frac{\partial w_2}{\partial H_1} \frac{\partial H_1}{\partial t} + \frac{\partial w_2}{\partial H_2} \frac{\partial H_2}{\partial t} \right) = 0, \tag{36}
 \end{aligned}$$

where Roy’s lemma has been used, i.e., $\frac{\partial v_n}{\partial t} = -(w_n a_n \phi(e_n) l_n - (1-s) p e_n)$, and $\frac{\partial v_n}{\partial w_n} = (1-t) a_n \phi(e_n) l_n$. The first-order condition for the linear income tax can be rewritten by substituting $w_n a_n \phi(e_n) l_n - p(1-s)e_n = (1-\beta)z_n$ and $p e_n = \frac{\beta}{1-s} z_n$ (both are derived from the first-order condition for learning):

$$\begin{aligned}
 \frac{\partial \mathcal{L}}{\partial t} &= - \sum_n \omega_n (1-\beta) z_n g_n + \eta \sum_n (1-\beta) z_n g_n \\
 & + \eta \frac{t}{1-t} \sum_n \left(z_n \frac{\partial l_n}{\partial t} \frac{1-t}{l_n} \right) g_n - \eta \frac{s}{(1-t)(1-s)} \sum_n z_n \beta \frac{\partial e_n}{\partial t} \frac{1-t}{e_n} g_n \\
 & + \left(\omega_1 + \eta \frac{t}{1-t} \right) H_1 \left(H_1 F_{11} \frac{\partial H_1}{\partial t} \frac{1-t}{H_1} + H_2 F_{12} \frac{\partial H_2}{\partial t} \frac{1-t}{H_2} \right) \\
 & + \left(\omega_2 + \eta \frac{t}{1-t} \right) H_2 \left(H_1 F_{12} \frac{\partial H_1}{\partial t} \frac{1-t}{H_1} + H_2 F_{22} \frac{\partial H_2}{\partial t} \frac{1-t}{H_2} \right) = 0. \tag{37}
 \end{aligned}$$

Next, use the distributional characteristic (17), divide by average income $\bar{z} = F(\cdot)$ and η . Use the property that the labor demand functions are homogenous of degree zero:

$$H_1 \frac{\partial w_1}{\partial H_1} = H_1 F_{11}(H_1, H_2) = -H_2 F_{12}(H_1, H_2) = -H_2 \frac{\partial w_2}{\partial H_1}, \tag{38}$$

$$H_2 \frac{\partial w_2}{\partial H_2} = H_2 F_{22}(H_1, H_2) = -L_1 F_{12}(H_1, H_2) = -H_1 \frac{\partial w_1}{\partial H_2}. \tag{39}$$

Substitution of (38) and (39) gives (note that symmetry of the substitution matrix implies $\frac{\partial w_1}{\partial H_2} = \frac{\partial w_2}{\partial H_1} = F_{12}$)

$$\begin{aligned}
 & (1-\beta)\xi - \frac{t}{1-t} \bar{\varepsilon}^{lt} + \frac{s}{(1-t)(1-s)} \beta \bar{\varepsilon}^{et} \\
 & - \left(\frac{\omega_2}{\eta} - \frac{\omega_1}{\eta} \right) (\varepsilon_1^{Ht} - \varepsilon_2^{Ht}) \frac{H_1 F_1}{F} \frac{H_2 F_2}{F} \frac{F_{12} F}{F_1 F_2} = 0, \tag{40}
 \end{aligned}$$

where the income weighted elasticities are defined as $\bar{\varepsilon}^{lt} \equiv \sum_n \varepsilon_n^{lt} z_n g_n / \sum_n z_n g_n = (1-\alpha)\varepsilon_1^{lt} + \alpha\varepsilon_2^{lt}$, and $\bar{\varepsilon}^{et} \equiv \sum_n \varepsilon_n^{et} z_n g_n / \sum_n z_n g_n = (1-\alpha)\varepsilon_1^{et} + \alpha\varepsilon_2^{et}$. Further, $\varepsilon_n^{Ht} \equiv -\frac{\partial H_n}{\partial t} \frac{1-t}{H_n} = \varepsilon_n^{zt} = \varepsilon_n^{lt} + \beta\varepsilon_n^{et} = \frac{\varepsilon_n}{\mu_n}$. Use $1-\alpha \equiv \frac{H_1 F_1}{F}$, $\alpha \equiv \frac{H_2 F_2}{F}$, $\frac{1}{\sigma} \equiv \frac{F_{12} F}{F_1 F_2}$,

and simplify to obtain

$$\xi(1-\beta) - \frac{t}{1-t} \bar{\varepsilon}^{lt} + \frac{s}{(1-t)(1-s)} \beta \bar{\varepsilon}^{et} - \left(\frac{\omega_2}{\eta} - \frac{\omega_1}{\eta}\right) \left(\frac{\varepsilon_1}{\mu_1} - \frac{\varepsilon_2}{\mu_2}\right) \alpha(1-\alpha) \frac{1}{\sigma} = 0. \tag{41}$$

The first-order condition for the linear education subsidy is

$$\begin{aligned} \frac{\partial \mathcal{L}}{\partial s} &= \sum_n \omega_n(1-t)pe_n g_n - \eta \sum_n (1-t)pe_n g_n \\ &+ \eta t \sum_n \left(w_n a_n \phi(e_n) \frac{\partial l_n}{\partial s} g_n \right) - \eta s \sum_n p \frac{\partial e_n}{\partial s} g_n \\ &+ (\omega_1(1-t)H_1 + \eta t H_1) \left(\frac{\partial w_1}{\partial H_1} \frac{\partial H_1}{\partial s} + \frac{\partial w_1}{\partial H_2} \frac{\partial H_2}{\partial s} \right) \\ &+ (\omega_2(1-t)H_2 + \eta t H_2) \left(\frac{\partial w_2}{\partial H_1} \frac{\partial H_1}{\partial s} + \frac{\partial w_2}{\partial H_2} \frac{\partial H_2}{\partial s} \right) = 0, \end{aligned} \tag{42}$$

where Roy’s lemma has been used, i.e., $\frac{\partial v_n}{\partial s} = (1-t)pe_n$, and $\frac{\partial v_n}{\partial w_n} = (1-t)a_n \phi(e_n)l_n$. The first-order condition for s can be rewritten by rearranging, and substituting $pe_n = \frac{\beta}{1-s}z_n$ (derived from the first-order condition for learning):

$$\begin{aligned} \frac{\partial \mathcal{L}}{\partial s} &= \sum_n \omega_n \left(\frac{1-t}{1-s} \right) \beta z_n g_n - \eta \sum_n \left(\frac{1-t}{1-s} \right) \beta z_n g_n \\ &+ \eta \frac{t}{1-s} \sum_n \left(z_n \frac{\partial l_n}{\partial s} \frac{1-s}{l_n} g_n \right) - \eta \frac{s}{1-s} \sum_n \frac{1}{1-s} \beta z_n \frac{\partial e_n}{\partial s} \frac{1-s}{e_n} g_n \\ &+ \left(\omega_1 \left(\frac{1-t}{1-s} \right) + \eta \frac{t}{1-s} \right) H_1 \left(H_1 F_{11} \frac{\partial H_1}{\partial s} \frac{1-s}{H_1} + H_2 F_{12} \frac{\partial H_2}{\partial s} \frac{1-s}{H_2} \right) \\ &+ \left(\omega_2 \left(\frac{1-t}{1-s} \right) + \eta \frac{t}{1-s} H_2 \right) H_2 \left(H_1 F_{12} \frac{\partial H_1}{\partial s} \frac{1-s}{H_1} + H_2 F_{22} \frac{\partial H_2}{\partial s} \frac{1-s}{H_2} \right) \\ &= 0. \end{aligned} \tag{43}$$

Next, use the distributional characteristic (18), divide by $\eta \beta \left(\frac{1-t}{1-s} \right) \bar{z}$, and use $\bar{z} = F(\cdot)$. Apply the property that the labor demand functions are homogenous of degree zero. Substitution of (38) and (39) gives (note that symmetry of the substitution matrix implies $\frac{\partial w_1}{\partial H_2} = \frac{\partial w_2}{\partial H_1} = F_{12}$)

$$-\xi + \frac{t}{1-t} \frac{\bar{\varepsilon}^{ls}}{\beta} - \frac{s}{(1-s)(1-t)} \bar{\varepsilon}^{es} + \left(\frac{\omega_2}{\eta} - \frac{\omega_1}{\eta}\right) \left(\frac{\varepsilon_1^{Hs}}{\beta} - \frac{\varepsilon_2^{Hs}}{\beta}\right) \frac{H_1 F_1}{F} \frac{H_2 F_2}{F} \frac{F_{12} F}{F_1 F_2}, \tag{44}$$

where the income weighted elasticities are defined as $\bar{\varepsilon}^{ls} \equiv \sum_n \varepsilon_n^{ls} z_n g_n / \sum_n z_n g_n = (1-\alpha)\varepsilon_1^{ls} + \alpha\varepsilon_2^{ls}$, and $\bar{\varepsilon}^{es} \equiv \sum_n \varepsilon_n^{es} z_n g_n / \sum_n z_n g_n = (1-\alpha)\varepsilon_1^{es} + \alpha\varepsilon_2^{es}$. Also the elasticity of skilled and unskilled labor is defined as: $\varepsilon_n^{Hs} \equiv \frac{\partial H_n}{\partial s} \frac{(1-s)}{H_n} = \varepsilon_n^{zs} = \varepsilon_n^{ls} +$

$\beta \varepsilon_n^{es} = \frac{\beta(1+\varepsilon_n)}{\mu_n}$. Next, use $1 - \alpha \equiv \frac{H_1 F_1}{F}$, $\alpha \equiv \frac{H_2 F_2}{F}$, $\frac{1}{\sigma} \equiv \frac{F_{12} F}{F_1 F_2}$, and simplify to obtain

$$\xi - \frac{t}{1-t} \frac{\bar{\varepsilon}^{ls}}{\beta} + \frac{s}{(1-s)(1-t)} \bar{\varepsilon}^{es} - \left(\frac{\omega_2}{\eta} - \frac{\omega_1}{\eta} \right) \left(\frac{1 + \varepsilon_1}{\mu_1} - \frac{1 + \varepsilon_2}{\mu_2} \right) \alpha (1 - \alpha) \frac{1}{\sigma} \tag{45}$$

Subtracting the first-order conditions for t and s yields

$$\begin{aligned} \frac{t}{1-t} \left(\frac{\bar{\varepsilon}^{lt}}{1-\beta} - \frac{\bar{\varepsilon}^{ls}}{\beta} \right) &= \frac{s}{(1-s)(1-t)} \left(\frac{\beta \bar{\varepsilon}^{et}}{1-\beta} - \bar{\varepsilon}^{es} \right) - \frac{\alpha(1-\alpha)}{\sigma} \left(\frac{\omega_2}{\eta} - \frac{\omega_1}{\eta} \right) \\ &\times \left(\frac{\varepsilon_1^{Ht} - \varepsilon_2^{Ht}}{1-\beta} - \frac{\varepsilon_1^{Hs} - \varepsilon_2^{Hs}}{\beta} \right) = 0. \end{aligned} \tag{46}$$

Substituting the relevant elasticities $\varepsilon_n^{lt} = \frac{\varepsilon_n(1-\beta)}{\mu_n}$, $\varepsilon_n^{et} = \frac{\varepsilon_n}{\mu_n}$, $\varepsilon_n^{ls} = \frac{\beta \varepsilon_n}{\mu_n}$, $\varepsilon_n^{es} = \frac{1}{\mu_n}$ yields $\frac{\bar{\varepsilon}^{lt}}{1-\beta} - \frac{\bar{\varepsilon}^{ls}}{\beta} = 0$, $\frac{\beta \bar{\varepsilon}^{et}}{1-\beta} - \bar{\varepsilon}^{es} = \frac{-1}{1-\beta}$, $\frac{\varepsilon_1^{Ht} - \varepsilon_2^{Ht}}{1-\beta} - \frac{\varepsilon_1^{Hs} - \varepsilon_2^{Hs}}{\beta} = 0$. Therefore, $\frac{s}{(1-s)(1-t)} = 0$. The optimal income tax follows from substituting $s = 0$ and $\varepsilon_1^{Ht} - \varepsilon_2^{Ht} = \frac{\varepsilon_1}{\mu_1} - \frac{\varepsilon_2}{\mu_2}$ in the expression for the optimal income tax.

Optimal nonlinear policies

The optimal marginal tax rate on the high-ability type follows from multiplying the first-order condition for c_1 (24) by $\frac{\partial u_1 / \partial l_1}{\partial u_1 / \partial c_1}$, and substituting the result into (25). Rearranging while using the definition for $T'_1 \equiv 1 + \frac{1}{w_1 a_1 \phi(e_1)} \frac{\partial u_1 / \partial l_1}{\partial u_1 / \partial c_1}$ yields

$$T'_1 = \frac{\theta}{\eta} \frac{l_2}{w_1 a_1 \phi(e_1)} \frac{a_2}{a_1} \frac{\partial u_1^*}{\partial l_2} \frac{\partial \pi^{-1}}{\partial l_1} = \frac{\theta}{\eta} \frac{l_2}{w_1} \frac{a_2}{a_1} \frac{\partial u_1^*}{\partial l_2} \frac{\partial \pi^{-1}}{\partial H_1} \tag{47}$$

Next, derive that $\frac{\partial \pi^{-1}}{\partial H_1} = \frac{1}{\sigma} \frac{\alpha}{(1-\alpha)} \frac{1}{H_2}$ from differentiating the skill premium:

$$\frac{\partial(\pi^{-1})}{\partial H_1} = \frac{\partial(F_2/F_1)}{\partial H_1} = \frac{F_{21} F_1 - F_{11} F_2}{F_1^2} = \frac{1}{H_2} \frac{1}{\sigma} \frac{\alpha}{(1-\alpha)}, \tag{48}$$

where the property is used that the labor demand function is homogeneous of degree zero (38). Substitution of the last result yields the equation in the text.

Repeating the same procedure with the first-order conditions for c_2 (27) and for l_2 (28) yields the marginal tax on unskilled labor income $T'_2 \equiv 1 + \frac{1}{w_2 a_2 \phi(e_2)} \frac{\partial u_2 / \partial l_2}{\partial u_2 / \partial c_2}$,

$$T'_2 = (1 - T_2) \frac{\theta}{\eta} \frac{\partial u_1^*}{\partial c_2} + \frac{1}{w_2 a_2 \phi(e_2)} \frac{\theta}{\eta} \frac{\partial u_1^*}{\partial l_2} \frac{a_2}{a_1} \left(\pi^{-1}(\cdot) + l_2 \frac{\partial \pi^{-1}(\cdot)}{\partial l_2} \right). \tag{49}$$

$\pi^{-1} + l_2 \frac{\partial \pi^{-1}}{\partial l_2}$ can be written as $\pi^{-1} + H_2 \frac{\partial \pi^{-1}}{\partial H_2} = \pi^{-1} - \frac{1}{\sigma} \frac{\alpha}{1-\alpha} \frac{H_1}{H_2}$ where (39) is used:

$$\frac{\partial(\pi^{-1})}{\partial H_2} = \frac{\partial(F_2/F_1)}{\partial H_2} = \frac{F_{22} F_1 - F_{12} F_2}{F_1^2} = -\frac{1}{H_2} \frac{H_1}{H_2} \frac{1}{\sigma} \frac{\alpha}{(1-\alpha)}. \tag{50}$$

Substitute this result and $\frac{\partial u_1^*}{\partial l_2} = \frac{T_1'}{\frac{\theta}{\eta} \frac{1}{\sigma} \frac{\alpha}{(1-\alpha)} \frac{1}{w_1 a_1 \phi(e_2)}}$ to get

$$T_2' = (1 - T_2') \frac{\theta}{\eta} \frac{\partial u_1^*}{\partial c_2} + T_1' \left(\frac{1 - \alpha}{\alpha} \right) (\sigma - 1). \tag{51}$$

Rearranging yields the expression in the text.

Optimal education subsidies for the skilled worker follow from combining the expression of the optimal income tax (30) with the first-order condition for e_1 (26). First, rewrite the first-order condition for e_1 using the definition of the subsidy on human capital investment $S_1' \equiv 1 - w_1 a_1 \phi'(e_1) l_1$,

$$-S_1' = \frac{\theta}{\eta} \frac{\partial u_1^*}{\partial l_2} l_2 \frac{a_2}{a_1} \frac{\partial \pi^{-1}}{\partial e_1}. \tag{52}$$

Dividing this by the expression for the optimal income tax gives the expression in the text.

Similarly, we find the optimal subsidy on education for the unskilled worker. Rewrite the first-order condition for e_2 using $S_2' \equiv 1 - w_2 a_2 \phi'(e_2) l_2$

$$-S_2' = \frac{\theta}{\eta} \frac{\partial u_1^*}{\partial l_2} l_2 \frac{a_2}{a_1} \frac{\partial \pi^{-1}}{\partial e_2}. \tag{53}$$

Using the properties of the production function and substituting $1 - S_2' = w_2 a_2 \phi'(e_2) l_2$, we obtain

$$S_2' = \frac{\theta}{\eta} \frac{\partial u_1^*}{\partial l_2} l_2 \frac{a_2}{a_1} \frac{1}{w_2 H_2} \frac{H_1}{H_2} \frac{1}{\sigma} \frac{\alpha}{(1 - \alpha)} (1 - S_2'). \tag{54}$$

Dividing the last equation by $(1 - S_2')$ and substituting $T_1' w_1 a_1 \phi(e_2) = \frac{\theta}{\eta} \frac{1}{\sigma} \frac{\alpha}{(1-\alpha)} \frac{\partial u_1^*}{\partial l_2}$ gives the equation in the text.

References

- Allen, F. (1982). Optimal linear income taxation with general equilibrium effects on wages. *Journal of Public Economics*, 17(2), 135–143.
- Alstadsaeter, A., Kolm, A.-S., & Larsen, B. (2008). Money or joy: the choice of educational type. *European Journal of Political Economy*, 24(1), 107–122.
- Atkinson, A. B., & Stiglitz, J. E. (1976). The design of tax structure: direct versus indirect taxation. *Journal of Public Economics*, 6(1–2), 55–75.
- Atkinson, A. B., & Stiglitz, J. E. (1980). *Lectures on public economics*. London/New York: McGraw-Hill.
- Blankenau, W. F., & Cassou, S. P. (2011). Industry estimates of the elasticity of substitution and the rate of biased technological change between skilled and unskilled labour. *Applied Economics*, 43(23), 3129–3142.
- Blundell, R., & MaCurdy, T. (1999). Labor supply: a review of alternative approaches. In O. Ashenfelter & D. Card (Eds.), *Handbook of labor economics* (Vol. 3A, pp. 1559–1695). Amsterdam: Elsevier. Chap. 27.
- Bovenberg, A. L., & Jacobs, B. (2005). Redistribution and education subsidies are Siamese twins. *Journal of Public Economics*, 89(11–12), 2005–2035.

- Cameron, S., & Taber, C. (2004). Estimation of educational borrowing constraints using returns to schooling. *Journal of Political Economy*, 112(1), 132–182.
- Card, D. (1999). The causal effect of education on earnings. In O. Ashenfelter & D. Card (Eds.), *Handbook of labor economics* (Vol. 3A, pp. 1801–1863). Amsterdam: Elsevier.
- Carneiro, P., & Heckman, J. J. (2004). Human capital policy. In J. J. Heckman & A. B. Krueger (Eds.), *Inequality in America: what role for human capital policy?* (pp. 77–240). Cambridge: MIT Press.
- Ciccone, A., & Peri, G. (2005). Long-run substitutability between more and less educated workers: evidence from U.S. States, 1950–1990. *Review of Economics and Statistics*, 87(4), 652–663.
- Diamond, P. A. (1998). Optimal income taxation: an example with a U-shaped pattern of optimal marginal tax rates. *American Economic Review*, 88(1), 83–95.
- Diamond, P. A., & Mirrlees, J. A. (1971). Optimal taxation and public production I: production efficiency. *American Economic Review*, 61(1), 8–27.
- Dixit, A. K., & Sandmo, A. (1977). Some simplified formulae for optimal income taxation. *The Scandinavian Journal of Economics*, 79(4), 417–423.
- Feldstein, M. (1973). On the optimal progressivity of the income tax. *Journal of Public Economics*, 2(4), 357–366.
- Goldin, C., & Katz, L. F. (2007). Long-run changes in the wage structure: narrowing, widening, polarizing. *Brookings Papers on Economic Activity*, 38(2), 135–165.
- Grossman, V., & Poutvaara, P. (2009). Pareto-improving bequest taxation. *International Tax and Public Finance*, 16(5), 647–669.
- Gruber, J., & Saez, E. (2003). The elasticity of taxable income: evidence and implications. *Journal of Public Economics*, 84(1), 1–32.
- Hare, P. G., & Ulph, D. T. (1979). On education and distribution. *Journal of Political Economy*, 87(5), S193–S212.
- Heckman, J. J., Lochner, L., & Taber, C. (1998). A study of tuition policy. *American Economic Review*, 88(2), 381–386.
- Jacobs, B. (2007). *Optimal redistributive tax and education policies in general equilibrium* (CESifo Working Paper No. 2162).
- Jacobs, B., & Bovenberg, A. L. (2011). Optimal taxation of human capital and the earnings function. *Journal of Public Economic Theory*, 13(6), 957–971.
- Katz, L. F., & Autor, D. H. (1999). Changes in the wage structure and wage inequality. In O. Ashenfelter & D. Card (Eds.), *Handbook of labor economics* (Vol. 3A, pp. 1463–1555). Amsterdam: Elsevier.
- Maldonado, D. (2008). Education policies and optimal taxation. *International Tax and Public Finance*, 15(2), 131–143.
- Mirrlees, J. A. (1971). An exploration in the theory of optimum income taxation. *Review of Economic Studies*, 38(2), 175–208.
- Naito, H. (1999). Re-examination of uniform commodity taxes under a non-linear income tax system and its implication for production efficiency. *Journal of Public Economics*, 71(2), 165–188.
- Naito, H. (2004). Endogenous human capital accumulation, comparative advantage and direct vs. indirect redistribution. *Journal of Public Economics*, 88(12), 2685–2710.
- Naito, H. (2007). Atkinson–Stiglitz theorem with endogenous human capital accumulation. *The B.E. Journal in Economic Analysis & Policy*, 7(1), 46.
- Nielsen, S. B., & Sørensen, P. B. (1997). On the optimality of the Nordic system of dual income taxation. *Journal of Public Economics*, 63(3), 311–329.
- Pirttilä, J., & Tuomala, M. (2001). On optimal non-linear taxation and public good provision in an overlapping generations economy. *Journal of Public Economics*, 79(3), 485–501.
- Saez, E. (2001). Using elasticities to derive optimal income tax rates. *Review of Economic Studies*, 68(1), 205–229.
- Schindler, D. (2011). Tuition fees and the dual income tax—the optimality of the Nordic income tax system reconsidered. *German Economic Review*, 12(1), 59–84.
- Seade, J. K. (1977). On the shape of optimal tax schedules. *Journal of Public Economics*, 7(2), 203–235.
- Sheshinski, E. (1972). The optimal linear income tax. *Review of Economic Studies*, 39(3), 297–302.
- Stern, N. H. (1982). Optimum income taxation with errors in administration. *Journal of Public Economics*, 17(2), 181–211.
- Stiglitz, J. E. (1982). Self-selection and Pareto efficient taxation. *Journal of Public Economics*, 17(2), 213–240.
- Tinbergen, J. (1975). *Income distribution*. Amsterdam: North-Holland.
- Tuomala, M. (1990). *Optimal income tax and redistribution*. Oxford: Clarendon.
- Ulph, D. T. (1977). On the optimal distribution of income and educational expenditure. *Journal of Public Economics*, 8(3), 341–356.