

Jacobs, Bas

Working Paper

Optimal redistributive tax and education policies in general equilibrium

CESifo Working Paper, No. 2162

Provided in Cooperation with:

Ifo Institute – Leibniz Institute for Economic Research at the University of Munich

Suggested Citation: Jacobs, Bas (2007) : Optimal redistributive tax and education policies in general equilibrium, CESifo Working Paper, No. 2162, Center for Economic Studies and ifo Institute (CESifo), Munich

This Version is available at:

<https://hdl.handle.net/10419/26206>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

OPTIMAL REDISTRIBUTIVE TAX AND EDUCATION POLICIES IN GENERAL EQUILIBRIUM

BAS JACOBS

CESIFO WORKING PAPER NO. 2162

CATEGORY 1: PUBLIC FINANCE

DECEMBER 2007

PRESENTED AT CESIFO AREA CONFERENCE ON PUBLIC SECTOR ECONOMICS, APRIL 2007

An electronic version of the paper may be downloaded

- *from the SSRN website:* www.SSRN.com
- *from the RePEc website:* www.RePEc.org
- *from the CESifo website:* www.CESifo-group.org/wp

OPTIMAL REDISTRIBUTIVE TAX AND EDUCATION POLICIES IN GENERAL EQUILIBRIUM

Abstract

Should a redistributive government optimally subsidize education to provoke a reduction in the skill premium through general equilibrium effects on wages? To answer this question, this paper studies optimal linear and non-linear redistributive income taxes and education subsidies in two-type models with endogenous human capital formation, endogenous labor supply, and endogenous wage rates. Under optimal linear policies, education should not be subsidized so as to reduce the skill premium. Linear income taxes are distributionally equivalent to (negative) linear education subsidies, but linear taxes do not distort investment in human capital, whether general equilibrium effects are present or not. If skilled labor supply is more elastic than unskilled labor supply, optimal redistributive linear income taxes are lowered as the distributional gains of linear taxes are offset by a rise in the skill premium. Moreover, the optimal linear income tax may even become negative if general equilibrium effects are sufficiently strong. Under non-linear taxation, governments can directly steer the skill premium by exploiting non-linearities in the policy schedules. At the top, the optimal marginal income tax rate is negative, and the optimal marginal education subsidy is positive. At the bottom, the optimal marginal income tax rate is positive, and education is optimally taxed at the margin. Hence, optimal non-linear tax and education policies compress wage differentials, which contributes to redistribution. Simulations show that the top rate and marginal education subsidies are close to zero for a wide range of plausible parameters. Only when high-ability and low-ability workers are rather poor substitutes in production, marginal education subsidies on the high type and marginal education taxes on the low type substantially differ from zero.

JEL Code: H2, H5, I2, J2.

Keywords: human capital, general equilibrium, education subsidies, optimal taxation, direct and indirect redistribution.

Bas Jacobs
Erasmus School of Economics
Erasmus University Rotterdam
PO Box 1738
The Netherlands – 3000 DR Rotterdam
bjacobs@few.eur.nl

November 14, 2007

I am grateful to Thomas Aronsson, Tony Atkinson, Lans Bovenberg, Robert Dur, Thomas Gaube, Mikhail Golosov, Wouter den Haan, Rick van der Ploeg, and Coen Teulings for their helpful discussions and thoughtful comments. I also benefitted from comments and suggestions by seminar participants at the Universities of Utrecht, Amsterdam, and Tilburg, the 62nd IIPF 2006 conference in Cyprus, August 28-31, 2006, and the CESifo Public Sector Economics Area Conference, Munich, April 20-22, 2007.

“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.148).

1 Introduction

In his book *Income Distribution*, the Dutch Nobel-prize winner Jan 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 unskilled workers (Katz and Autor, 1999). In Tinbergen’s (1975) 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).

This analysis closely follows Bovenberg and Jacobs (2005). These authors 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 non-linear 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 exactly, and the sole role of education subsidies is to offset the distortions of the income tax on skill formation (see also Jacobs and Bovenberg, 2007).

This paper maintains the assumptions of Bovenberg and Jacobs (2005) to ensure that human capital is optimally not subsidized in the *absence* of general equilibrium effects on wages. Therefore, the earnings function is assumed to be weakly separable between ability, labor, and education, both with linear and non-linear 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 (2007). 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

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 an abuse of language for expositional reasons.

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 human capital: $\pi(H_1/H_2)$, $\pi' < 0$. Hence, education subsidies must increase the relative supply H_1/H_2 to be helpful in reducing wage inequality. Second, if education subsidies do indeed increase relative supply of human capital, 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 non-linear policy instruments.

The first part of this paper considers optimal *linear* tax and education subsidies. Linear education subsidies tend to increase both H_1 and H_2 at the same time, and it is not clear whether relative supply of human capital is increased at all. The supply of human capital of each agent depends on ability a , education e , and labor effort, i.e., $H_1(a_1, e_1, l_1)$ and $H_2(a_2, e_2, l_2)$, where the first agent has high ability, and the second low ability.² It is demonstrated that general equilibrium effects can *never* be exploited for redistributive reasons when supply elasticities of both labor and education are equal across agents, and income effects are absent. The reason is that relative human capital supply (H_1/H_2) remains fixed. Indeed, linear education subsidies (s) will only compress wage differentials if the skilled worker's supply of human capital is more elastic with respect to the subsidy than the unskilled worker's supply of human capital. 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, even if there are general equilibrium effects on wages. 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. Therefore, 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 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.

The second part of the paper considers the simultaneous setting of optimal *non-linear* tax and education policies. Optimal non-linear 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 supply of human capital (H_1/H_2) can, by definition, be directly steered by the non-linear tax and education schedules – even if all supply elasticities are equal. By

²To avoid confusion, H_n is referred to as the supply of 'broad human capital', encompassing both quality (education) and quantity (labor) dimensions of labor. When we speak of 'investment in human capital', i.e., education, this only applies to the 'narrow' definition of human capital, i.e., the augmentation of labor time.

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 non-linear policies is yes. The skilled worker 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). 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, non-linear instruments do exploit general equilibrium effects for redistribution.

We analyze the quantitative importance of general equilibrium effects by simulating optimal non-linear 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 particularly be exploited for redistribution when the elasticity of substitution between skilled and unskilled labor is 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 non-linear 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. Non-linear 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. However, numerical simulations of non-linear policies give rather weak support to the use of education policies in reducing inequality through general equilibrium effects on wages. Hence, it appears that direct instruments are more efficient than indirect instruments, i.e. education subsidies.³

This paper is related to a number of other papers. First, Dur and Teulings (2001, 2004) pioneered the analysis of optimal tax and education policies in the presence of general equilibrium effects on wages. They developed a log-linear matching model with a continuum of agents. These authors show that education is subsidized in an optimal redistributive tax and education policy as long as the regressive incidence of the education subsidy (ability bias) does not offset the progressive general equilibrium effect on wages. Education policy then allows the government to rely less on distortionary income taxation for redistribution. This paper provides different results, and the reasons for this are twofold. First, Dur and Teulings (2001, 2003) only analyze log-linear policies. Log-linear education policies have the property that high-ability types receive a higher marginal subsidy on education than low-ability types. Thus, the regressive schedule for education subsidies directly reduces the skill premium, as high-ability types will invest relatively more in human capital than low-ability types. Second, the authors assume that increase in the

³From 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.

mean level of human capital always reduces the skill premium, even if all behavioral elasticities are equal and income effects in labor supply are absent, so that relative supplies of human capital remain fixed. This contrasts with the current paper, where an increase in the mean level of human capital, while not changing the relative supplies of human capital, does not affect the wage distribution.

Second, some papers analyzed the question whether direct or indirect instruments should be used for redistribution in general equilibrium settings. If the government cannot observe different skill-types in production, and cannot tax or subsidize all production inputs at different rates, both the Diamond and Mirrlees (1971) production efficiency theorem, and the Atkinson and Stiglitz (1976) zero commodity tax theorem break down. Indirect taxes should therefore optimally complement the non-linear income tax, even when preferences are weakly separable (Naito, 1999). Saez (2004), in contrast, argues against indirect instruments for redistribution, and restores the validity of the production efficiency and zero commodity tax theorems when skill-types are imperfect substitutes in production. Naito (2004, 2007) finds that this holds true only if more able individuals do not have a comparative advantage in high-skilled occupations. By allowing for comparative advantage in skill formation, Naito (2004, 2007) derives that deviations from aggregate production efficiency, and non-zero commodity taxation are optimal.

This paper contributes to this literature and finds that another powerful property of optimal tax structures is violated under linear and non-linear taxation. Diamond and Mirrlees (1971) derive that optimal tax expressions are the same in partial equilibrium (fixed prices), and general equilibrium (variable prices). This is what Saez (2004) calls the ‘tax formula result’. We show that labor demand parameters explicitly enter optimal tax formulae, and factor price changes are exploited for redistribution. Hence, the tax formula result also breaks down with comparative advantage in skill formation, as in Feldstein (1973), Allen (1982), Stern (1982), and Stiglitz (1982). Moreover, indirect instruments, such as education subsidies, should optimally complement the non-linear income tax. Hence, under non-linear income taxation, the zero commodity tax theorem is not applicable to education, which further bolsters the findings by Naito (1999, 2004, 2007).

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. Section 3 studies the same problem using non-linear instruments. Section 4 concludes. An appendix contains all the technical derivations.

2 Model

This section presents the base-line model. The standard models of optimal income taxation with general equilibrium effects on wages are extended with human capital formation.⁴ 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.⁶

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, Carneiro and Heckman (2003) or Cameron and Taber (2004).

⁶Saez (2004) and Dur and Teulings allow (2001, 2004) for an infinite number of skill types.

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), Dur and Teulings (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 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.

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 their ability. Gross labor income z_n of each individual is

$$z_n \equiv w_n a_n \phi(e_n) l_n, \quad n = 1, 2, \quad (2)$$

where $\phi'(e_n) > 0$ and $\phi''(e_n) < 0$. w_n denotes the gross wage rate per unit of human capital of an individual of skill-type n , and, l_n is work effort. $a_n \phi(e_n)$ is the production function for human capital, 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 generate more human capital with the same amount of educational efforts because $\frac{\partial z_n}{\partial e_n \partial a_n} = w_n \phi'(e_n) l_n > 0$.⁷ 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 (Jacobs and Bovenberg, 2007). The elasticity of the production function is also assumed to be constant under linear policies, and is denoted by $\beta \equiv \frac{\phi'(e_n) e_n}{\phi(e_n)}$ (see Jacobs and Bovenberg, 2007). In general equilibrium, type 1 is assumed to earn a higher gross income than type 2, i.e. $w_1 a_1 \phi(e_1) l_1 > w_2 a_1 \phi(e_2) l_2$.

The price of one unit of education is denoted by p_n . Note that these costs might differ between individuals. 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 individual may receive a non-individualized lump-sum transfer b . Hence, the income tax is progressive in the sense that

⁷This 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.

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).

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:

$$c_n = (1 - t)(w_n a_n \phi(e_n) l_n - (1 - s)p_n e_n) + b, \quad n = 1, 2. \quad (3)$$

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

$$l_n = ((1 - t)w_n a_n \phi(e_n))^{\varepsilon_n}, \quad n = 1, 2. \quad (4)$$

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

$$w_n a_n \phi'(e_n) l_n = (1 - s)p_n, \quad n = 1, 2. \quad (5)$$

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. The first-order conditions reveal that investments in education increase if larger labor supply increases the utilization rate of human capital. As investments in human capital increase, net wage rates per hour worked increase, and labor supply expands. Larger labor supply, in turn, results in higher investment in human capital, and so on. Sufficiently strong diminishing returns to human capital accumulation β , or a sufficiently low wage elasticity of labor supply ε_n , should guarantee that this feedback dampens, so that an interior solution in human capital formation and labor supply is attained. Second-order conditions are satisfied by imposing the following restriction on the parameters (see the Appendix for the derivation)

$$\mu_n \equiv 1 - \beta(1 + \varepsilon_n) > 0, \quad n = 1, 2. \quad (6)$$

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, and are derived in the Appendix. The tax elasticity of labor earnings $z_n = w_n a_n \phi(e_n) l_n$ amounts to $\varepsilon_n^{zt} = \varepsilon_n^{lt} + \beta \varepsilon_n^{et} = \frac{\varepsilon_n}{\mu_n}$. 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 $\varepsilon_n^{zs} \equiv \frac{\partial z_n}{\partial s} \frac{(1-s)}{z_n} = \frac{\beta(1+\varepsilon_n)}{\mu_n}$.

2.2 Firms

There is one sector of production.⁹ 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

⁹At least two production sectors producing different commodities, where one sector is intensive in skilled labor and the other in unskilled labor, are necessary to analyze the desirability of aggregate production efficiency and the optimality of zero commodity taxation, see Naito (1999, 2004, 2007). This extension is left for future research.

production function, which features constant returns to scale in labor inputs H_1 and H_2

$$Y \equiv F(H_1, H_2), \quad (7)$$

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

$$w_n = F_n(H_1, H_2), \quad n = 1, 2. \quad (8)$$

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 $f(H_1/H_2) \equiv F(H_1, H_2)/H_2$):

$$\pi(H_1/H_2) \equiv \frac{w_1}{w_2} = \frac{f'(H_1/H_2)}{f(H_1/H_2) - (H_1/H_2)f'(H_1/H_2)}, \quad (9)$$

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:

$$H_n = a_n \phi(e_n) l_n g_n, \quad n = 1, 2. \quad (10)$$

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

$$F(H_1, H_2) = (c_1 + p_1 e_1) g_1 + (c_2 + p_2 e_2) g_2 + \Lambda. \quad (11)$$

2.4 Government

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

$$\omega_1 v_1(b, t, s, w_1) g_1 + \omega_2 v_2(b, t, s, w_2) g_2, \quad \omega_n \geq 0, \quad n = 1, 2, \quad (12)$$

where ω_n denotes the weight of type n in social welfare. The welfare weights sum to one: $\omega_1 + \omega_2 = 1$. If $\omega_1 = \omega_2$, 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

$$\begin{aligned} & t(w_1 a_1 \phi(e_1) l_1 - (1-s)p_1 e_1) g_1 + t(w_2 a_2 \phi(e_2) l_2 - (1-s)p_2 e_2) g_2 \\ & = s p_1 e_1 g_1 + s p_2 e_2 g_2 + b + \Lambda. \end{aligned} \quad (13)$$

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

The government maximizes social welfare by optimally choosing the lump-sum transfer b , the linear marginal tax and the linear education subsidy s . Formally, the following Lagrangian is maximized¹¹

$$\begin{aligned} \max_{\{b,t,s\}} \mathcal{L} = & \omega_1 v_1(b, t, s, w_1) g_1 + \omega_2 v_2(b, t, s, w_2) g_2 \\ & + \eta (t (w_1 a_1 \phi(e_1) l_1 - (1-s) p_1 e_1) g_1 + t (w_2 a_2 \phi(e_2) l_2 - (1-s) p_2 e_2) g_2) \\ & - \eta (s p_1 e_1 g_1 + s p_2 e_2 g_2 + b + \Lambda), \end{aligned} \quad (14)$$

where η denotes the Lagrange multiplier of the government budget constraint, and the labor market clearing conditions have to be imposed $w_1 = F_1(H_1, H_2)$, $w_2 = F_2(H_1, H_2)$, and $H_n = a_n \phi(e_n) l_n$, $n = 1, 2$.

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

$$\frac{\omega_1}{\eta} g_1 + \frac{\omega_2}{\eta} g_2 = 1, \quad (15)$$

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 (15)) should equal the costs in terms of a higher b (i.e., the right-hand side of (15)).

With the aid of the first-order condition for b (15), 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 (see, e.g., Atkinson and Stiglitz, 1980):

$$\xi^z \equiv - \frac{\sum_n \frac{\omega_n}{\eta} z_n g_n - \sum_n z_n g_n \sum_n \frac{\omega_n}{\eta} g_n}{\sum_n z_n g_n \sum_n \frac{\omega_n}{\eta} g_n} = \frac{\sum_n (1 - \frac{\omega_n}{\eta}) z_n g_n}{\sum_n z_n g_n}, \quad (16)$$

where the second equality follows from (15). 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

$$\xi^e \equiv - \frac{\sum_n \frac{\omega_n}{\eta} e_n g_n - \sum_n e_n g_n \sum_n \frac{\omega_n}{\eta} g_n}{\sum_n \frac{\omega_n}{\eta} g_n \sum_n e_n g_n} = \frac{\sum_n (1 - \frac{\omega_n}{\eta}) e_n g_n}{\sum_n e_n g_n} = \xi^z. \quad (17)$$

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 redistributional 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 (2007). In the remainder, the superscripts are dropped, and $\xi \equiv \xi^z = \xi^e$.

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.

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

$$\begin{aligned}
\frac{\partial \mathcal{L}}{\partial t} = & - \sum_n \omega_n (w_n a_n \phi(e_n) l_n - p_n (1-s) e_n) g_n \\
& + \eta \sum_n (w_n a_n \phi(e_n) l_n - p_n (1-s) e_n) g_n \\
& + \eta t \sum_n \left(w_n a_n \phi(e_n) \frac{\partial l_n}{\partial t} \right) g_n - \eta s \sum_n p_n \frac{\partial e_n}{\partial t} g_n \\
& + (\omega_1 (1-t) H_1 + \eta t H_1) \left(\frac{\partial w_1}{\partial H_1} \frac{\partial H_1}{\partial t} + \frac{\partial w_1}{\partial H_2} \frac{\partial H_2}{\partial t} \right) \\
& + (\omega_2 (1-t) H_2 + \eta t H_2) \left(\frac{\partial w_2}{\partial H_1} \frac{\partial H_1}{\partial t} + \frac{\partial w_2}{\partial H_2} \frac{\partial H_2}{\partial t} \right) = 0.
\end{aligned} \tag{18}$$

And the first-order condition for the linear education subsidy is

$$\begin{aligned}
\frac{\partial \mathcal{L}}{\partial s} = & \sum_n \omega_n (1-t) p_n e_n g_n - \eta \sum_n (1-t) p_n e_n g_n \\
& + \eta t \sum_n \left(w_n a_n \phi(e_n) \frac{\partial l_n}{\partial s} g_n \right) - \eta s \sum_n p_n \frac{\partial e_n}{\partial s} g_n \\
& + (\omega_1 (1-t) H_1 + \eta t H_1) \left(\frac{\partial w_1}{\partial H_1} \frac{\partial H_1}{\partial s} + \frac{\partial w_1}{\partial H_2} \frac{\partial H_2}{\partial s} \right) \\
& + (\omega_2 (1-t) H_2 + \eta t H_2) \left(\frac{\partial w_2}{\partial H_1} \frac{\partial H_1}{\partial s} + \frac{\partial w_2}{\partial H_2} \frac{\partial H_2}{\partial s} \right) = 0.
\end{aligned} \tag{19}$$

Roy's lemma has been used in deriving both expressions, i.e., $\frac{\partial v_n}{\partial t} = -(w_n a_n \phi(e_n) l_n - (1-s) p_n e_n)$, and $\frac{\partial v_n}{\partial s} = (1-t) p_n e_n$, and $\frac{\partial v_n}{\partial w_n} = (1-t) a_n \phi(e_n) l_n$.

2.5 Optimal tax policies in general equilibrium

This section first derives the case in which education subsidies are not used ($s = 0$). Furthermore, it starts with deriving the optimal linear income tax in the absence of general equilibrium effects (see, e.g., Sheshinski, 1971; Dixit and Sandmo, 1977; Atkinson and Stiglitz, 1980). In that case, the last two lines in the first-order condition for t (18) are zero, and the optimal linear income tax is (see Appendix):

$$\frac{t}{1-t} = \frac{(1-\beta)\xi}{\bar{\varepsilon}^{lt}} = \frac{\xi}{(1-\alpha)\frac{\varepsilon_1}{\mu_1} + \alpha\frac{\varepsilon_2}{\mu_2}}, \tag{20}$$

where $\bar{\varepsilon}^{lt} \equiv (1-\alpha)\varepsilon_1^{lt} + \alpha\varepsilon_2^{lt}$, and $\bar{\varepsilon}^{lt}/(1-\beta) = \bar{\varepsilon}_{zt}$ is the income weighted average of the tax elasticity of total labor earnings. The denominator is the elasticity of the tax base with respect to the tax rate. The optimal income tax formula (20) shows the trade-off between equity (numerator) and efficiency (denominator). The larger the social preference for redistribution is, the larger is ξ , and, the higher is 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 $\xi = 0$. The larger the income weighted average elasticity of labor earnings is, the lower is 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.

With general equilibrium effects, in contrast, the last two lines in the first-order condition for t (18) are not zero, and need to be taken into account as well. To obtain a general expression

for the optimal linear income tax with general equilibrium effects on wages, the property is exploited that inverse labor demand equations are homogeneous of degree zero as long as there are constant returns to scale in production, i.e.,

$$H_1 \frac{\partial w_1}{\partial H_1} = H_1 F_{11}(H_1, H_2) = -H_2 F_{12}(H_1, H_2) = -H_2 \frac{\partial w_2}{\partial H_1}, \quad (21)$$

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

Simplifying the first-order condition for t then yields (see Appendix)

$$\frac{t}{1-t} = \frac{1}{\varepsilon^{lt}} \left((1-\beta)\xi - \left(\frac{\omega_2}{\eta} - \frac{\omega_1}{\eta} \right) \left(\frac{\varepsilon_1}{\mu_1} - \frac{\varepsilon_2}{\mu_2} \right) \alpha(1-\alpha) \frac{1}{\sigma} \right). \quad (23)$$

The expression for the optimal linear income tax in partial equilibrium differs from the expression for the optimal income tax in general equilibrium. Hence, the ‘tax-formula result’ does not apply (see also Saez, 2004). The difference with formula without general equilibrium effects is the added term in the brackets on the right hand side. $\left(\frac{\omega_2}{\eta} - \frac{\omega_1}{\eta} \right) \left(\frac{\varepsilon_1}{\mu_1} - \frac{\varepsilon_2}{\mu_2} \right) \alpha(1-\alpha) \frac{1}{\sigma}$ 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 (23) 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 (9), and larger 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. Below we derive that a higher elasticity of skilled labor supply is a necessary requirement for education subsidies to work in favor of redistribution via general equilibrium effects under linear policy instruments. Hence, we assume that $\varepsilon_1^{zt} > \varepsilon_2^{zt}$, or, equivalently, $\varepsilon_1 > \varepsilon_2$. Also from an empirical point of view, a larger elasticity of skilled labor relative to unskilled labor is plausible (see Gruber and Saez, 2003). Furthermore, it can be demonstrated that, if both workers have identical CES utility functions, uncompensated wage elasticities of labor supply unambiguously increase with skill type as long as labor supply functions are upward sloping.¹²

¹²Suppose we hold human capital accumulation fixed. If utility has the CES form $u(c_n, l_n) \equiv [\delta c_n^\rho + (1-\delta)(1-l_n)^\rho]^{1/\rho}$, the uncompensated labor supply elasticity with respect to the tax rate is $\varepsilon_n^{lt} \equiv -\frac{\partial l_n}{\partial t} \frac{1-t}{l_n} = \frac{\sigma_{cl}-\zeta}{l_n/(1-l_n)+\zeta}$, where $\sigma_{cl} \equiv 1/(1-\rho)$ is the elasticity of substitution between consumption and leisure, and $\zeta \equiv (1-t)w_n l_n / ((1-t)w_n l_n + b)$ is the share of labor income in total income. The labor supply curve is upward sloping (backward bending) if $\sigma_{cl} > \zeta$ ($\sigma_{cl} < \zeta$). With identical tax rates t and transfers g , the uncompensated tax elasticity of labor supply for high-skilled workers is higher than for low-skilled workers, if the labor supply curve is upward sloping, which can be derived as follows. Differentiation of the elasticity with respect to ability n gives: $\frac{\partial \varepsilon_n^{lt}}{\partial n} = \frac{(\zeta + l_n/(1-l_n)) \frac{\partial \zeta}{\partial n} + (\sigma_{cl} - \zeta) \frac{\partial (l_n/(1-l_n))}{\partial n}}{(\zeta + l_n/(1-l_n))^2} > 0$ if $\sigma_{cl} > \zeta$. $\frac{\partial \zeta}{\partial n} > 0$ as the uniform lump-sum transfer

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 $\frac{\varepsilon_1}{\mu_1} = \frac{\varepsilon_2}{\mu_2}$), the general equilibrium term vanishes as well. The intuition is straightforward. With equal elasticities, linear taxation does not affect relative labor supply. If relative total labor supply (H_1/H_2) remains constant, relative wages (w_1/w_2) remain constant as well, cf. the skill premium (9). 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.¹³

The expression for the optimal income tax (23) suggests that optimal marginal income taxes may even become negative. If the shares of workers are equal, i.e., $g_1 = g_2 = \frac{1}{2}$ we can find an explicit condition under which marginal tax rates are optimally negative ($\xi = (\omega_2 - \omega_1) (\frac{1}{2} - \alpha)$ and $\eta = 1/2$)

$$\frac{\frac{\varepsilon_1}{\mu_1} - \frac{\varepsilon_2}{\mu_2}}{(1 - \beta)} > \frac{\sigma (\frac{1}{2} - \alpha)}{2\alpha(1 - \alpha)}. \quad (24)$$

The last inequality is satisfied if general equilibrium effects are very important (low σ), if there is less inequality (higher α), and if tax elasticities of earnings are higher for high-skilled workers than for low-skilled workers (large $(\varepsilon_1^{zt} - \varepsilon_2^{zt})/(1 - \beta)$). 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.

Optimal negative income taxes may not be a purely theoretical curiosity. As an illustrative example, suppose that the elasticity of substitution between skilled and unskilled workers equals $\sigma = 1.5$, which is not an uncommon empirical value (see for example Katz and Autor, 1999). Assume furthermore that the skilled worker earns 50% more than the unskilled worker so that $\alpha = 0.4$. This can be justified by a wage increase of 10% for every additional year of schooling and 5 years of higher education. A Mincer return of 10% is not an uncommon empirical estimate either (see Card, 1999). $\alpha = 0.4$ may also correspond to the income share of low-skilled workers in labor earnings. Finally, take β to be $1/4$. This can be interpreted as the share of forgone earnings in life-time earnings, or the share of schooling years in total years available for working and education. Then, the difference in human capital supply elasticities must satisfy $\varepsilon_1^{zt} - \varepsilon_2^{zt} = 0.24$. Under this parameterization, a difference of 0.24 in the total elasticity of labor earnings is sufficient to have negative optimal marginal tax rates, when general equilibrium effects on wages are present.

It should be noted that *uncompensated* elasticities determine the general equilibrium impact of taxes. The change in relative supplies determines the effect on wages. This change in relative supply is correctly measured only with uncompensated, and not compensated, changes.¹⁴

constitutes a smaller part of total income for wealthier individuals, and $\frac{\partial(l_n/1-l_n)}{\partial n} > 0$. The latter can be proven by taking the differential of $\frac{\partial(l_n/1-l_n)}{\partial n} = \frac{\partial(w_n l_n / (w_n - w_n l_n))}{\partial n} = \frac{l_n \frac{\partial w_n}{\partial n} + w_n \frac{\partial l_n}{\partial n}}{(w_n - w_n l_n)^2} > 0$. This term is positive under standard monotonicity (or ‘single-crossing’) conditions, which ensure that gross incomes (and consumption) increase in ability, i.e., $\frac{\partial w_n l_n}{\partial n} = l_n \frac{\partial w_n}{\partial n} + w_n \frac{\partial l_n}{\partial n} > 0$, see Mirrlees (1971). Allowing for human capital formation does not qualitatively affect this derivation.

¹³Using 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.

¹⁴The current model ignores income effects, and compensated and uncompensated elasticities coincide. Whether this is an omission remains an open empirical question. Estimates of the income elasticity of labor supply vary substantially, and there appears to be no consensus in the literature on its size, see also Blundell and MaCurdy (1999), and Gruber and Saez (2003).

2.6 Optimal education policies in general equilibrium

This section analyzes the case where the government has only education subsidies at its disposal and does not have access to an income tax ($t = 0$). This is not the relevant case to describe the real world, but it helps to derive the intuition for the results in the next section, where both tax and education policies are simultaneously optimized. Again, first education subsidies in the absence of general equilibrium effects are derived (see Appendix):

$$\frac{-s}{1-s} = \frac{\xi}{\bar{\varepsilon}^{es}} = \frac{\xi}{\frac{1-\alpha}{\mu_n} + \frac{\alpha}{\mu_2}}, \quad (25)$$

where $\bar{\varepsilon}^{es} \equiv (1-\alpha)\varepsilon_1^{es} + \alpha\varepsilon^{es}$ is the income weighted learning elasticity with respect to the subsidy ($\varepsilon_n^{es} \equiv \frac{\partial e_n}{\partial s} \frac{1-s}{e_n} = \frac{1}{\mu_n}$). Since the optimal subsidy is negative, education is *taxed* if the government has no other means to redistribute incomes. The formula again stresses the trade-off between equity (numerator) and efficiency (denominator). The more education taxes yield distributional gains (larger ξ), the larger are taxes on education. If education decisions are more elastic to taxes, $\bar{\varepsilon}^{es}$ increases and optimal taxes on education are lower as a consequence.

In the presence of general equilibrium effects, optimal subsidies on education are given by (see Appendix)

$$\frac{-s}{1-s} = \frac{1}{\bar{\varepsilon}^{es}} \left(\xi - \left(\frac{\omega_2}{\eta} - \frac{\omega_1}{\eta} \right) \left(\frac{1+\varepsilon_1}{\mu_1} - \frac{1+\varepsilon_2}{\mu_2} \right) \alpha(1-\alpha) \frac{1}{\sigma} \right) \quad (26)$$

The difference with the optimal subsidy without general equilibrium effects (25) is again a term containing the general equilibrium impact of the education subsidy $\left(\frac{\omega_2}{\eta} - \frac{\omega_1}{\eta} \right) \left(\frac{1+\varepsilon_1}{\mu_1} - \frac{1+\varepsilon_2}{\mu_2} \right) \alpha(1-\alpha) \frac{1}{\sigma}$. The intuition of the optimal income tax carries over to the present case. If wage elasticities of labor supply are equal ($\varepsilon_1 = \varepsilon_2$, so that $\mu_1 = \mu_2$) education subsidies will not provoke general equilibrium effects since relative supplies of human capital are not affected. Similarly, perfect substitution between labor types ($\sigma = \infty$) gives the same result as the expression without general equilibrium effects.

Taxes on human capital formation are lower if imperfect substitution on the labor market is important (low σ), and if the labor supply elasticity of the high-ability worker is lower than that of the low-ability worker, i.e., $\varepsilon_1 > \varepsilon_2$. The latter condition is necessary to have education subsidies working in favor of redistribution through general equilibrium effects. Indeed, one may get the result that – even in the absence of an income tax – education should be subsidized for distributional reasons if the adverse distributional effects ξ are outweighed by the positive general equilibrium effects on wages. If shares of both types of workers are equal ($g_1 = g_2$), this requires $\frac{1+\varepsilon_1}{\mu_1} - \frac{1+\varepsilon_2}{\mu_2} > \frac{\sigma(\frac{1}{2}-\alpha)}{2\alpha(1-\alpha)}$. We can infer that with $\sigma = 1.5$ and $\alpha = 0.4$, a difference in subsidy elasticities of education equal to 0.31 is sufficient to provoke such strong general equilibrium effects that education should be subsidized rather than taxed.

2.7 Optimal tax and education policies in general equilibrium

When the government simultaneously optimizes tax and education policies, the optimal taxes and subsidies follow from solving the first-order conditions for t and s . The general solution for optimal education subsidies in the presence of general equilibrium effects is (see Appendix)

$$s = 0. \quad (27)$$

The government sets the subsidies on education to zero. Therefore, educational investments are weakly efficient, i.e., conditional upon distorted labor supply, in the presence of an optimal income tax. The intuition is the same as in Bovenberg and Jacobs (2005) and Jacobs and

Bovenberg (2007). The government may use education subsidies as an implicit tax on leisure, because education and work effort are complementary. At the same time, educational investments generate inequality due to the ability bias in education, i.e., the high-ability individuals invest more in education than low-ability individuals (*ceteris paribus*).

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 . Education subsidies do not reduce the labor supply distortion more compared to an equally costly reduction in income taxes. 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 (Jacobs and Bovenberg, 2007).

The government is also indifferent between education taxes and income taxes to redistribute income. Because education is linear in income, any redistribution that education taxes can achieve, can be achieved equally as well 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 using the income tax instead of taxes on education to redistribute incomes (Jacobs and Bovenberg, 2007).

General equilibrium effects do not change the logic of these arguments. Even when general equilibrium effects on wages are present, there is still a linear relationship between earnings and education. Hence, education subsidies remain distributionally equivalent to income taxes, and education subsidies still cannot be used as implicit taxes on leisure. Optimal subsidies thus ensure conditional efficiency in human capital formation. We like to speak of conditional efficiency, because labor supply is still distorted below first-best levels. As the utilization rate of human capital is lower, so are the returns to human capital investments, and education is below first-best levels.

From this discussion follows that education subsidies should not be employed to compress the wage distribution in the optimal redistributive program. This proves the robustness of Bovenberg and Jacobs (2005) in the presence of general equilibrium effects on wages. The optimal linear income tax with optimal education policies remains as in (27). Hence, the reader is referred to the earlier discussion.

3 Non-linear taxes and subsidies

This section derives the optimal non-linear tax and education policies. We can now allow for general utility and production functions. We do nevertheless maintain the assumption that the earnings function is weakly separable in ability, labor, and education. This ensures that optimal education subsidies are zero in the absence of general equilibrium effects on wages (Jacobs and Bovenberg, 2007). We also provide simulations of the optimal non-linear taxes and subsidies, while maintaining the same structure on preferences and technologies as in the linear case.

The optimal non-linear 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 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. That is,

$$U_n(c_n, z_n, e_n, n) \geq U_n(c_m, z_m, e_m, n), \quad n, m = 1, 2, \quad (28)$$

where $U_n(c_n, z_n, e_n, n) \equiv u_n\left(c_n, \frac{z_n}{w_n a_n \phi(e_n)}\right) = u_n(c_n, l_n)$. In the current model, the two incentive compatibility constraints are given by

$$u_1(c_1, l_1) = u_1\left(c_1, \frac{z_1}{w_1 a_1 \phi(e_1)}\right) \geq u_1\left(c_2, \frac{z_2}{w_1 a_1 \phi(e_2)}\right) = u_1\left(c_2, \frac{w_2 a_2}{w_1 a_1} l_2\right), \quad (29)$$

$$u_2(c_2, l_2) = u_2\left(c_2, \frac{z_2}{w_2 a_2 \phi(e_2)}\right) \geq u_2\left(c_1, \frac{z_1}{w_2 a_2 \phi(e_1)}\right) = u_2\left(c_1, \frac{w_1 a_1}{w_2 a_2} l_1\right). \quad (30)$$

The government maximizes the social welfare function (12) subject to the economy's resource constraint (11), and incentive compatibility constraint (29). Under normal circumstances, the second-incentive constraint (30) is not binding at an optimal solution, and it will be ignored in the remainder (see also Stiglitz, 1982, and Stern, 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 formulated¹⁵

$$\begin{aligned} \max_{\{c_1, l_1, e_1, c_2, l_2, e_2\}} \mathcal{L} = & \omega_1 u_1(c_1, l_1) + \omega_2 u_2(c_2, l_2) \\ & + \eta (F(a_1 \phi(e_1) l_1, a_2 \phi(e_2) l_2) - e_1 - e_2 - c_1 - c_2 - \Lambda) \\ & + \theta \left(u_1(c_1, l_1) - u_1^* \left(c_2, \pi^{-1} \left(\frac{a_1 \phi(e_1) l_1}{a_2 \phi(e_2) l_2} \right) \frac{a_2}{a_1} l_2 \right) \right), \end{aligned} \quad (31)$$

where the inverse of the skill premium is denoted by π^{-1} , see (9). The conditions for labor market equilibrium (10) are substituted. To avoid confusion in notation, utility of the high-skilled type mimicking the low-skilled type is designated by $u_1^* \equiv u_1\left(c_2, \frac{w_2 a_2}{w_1 a_1} l_2\right)$. θ is the Lagrange multiplier on the incentive compatibility constraint, and measures the marginal social value of redistributing income from the high-skilled to the low-skilled worker.

First-order conditions for an optimal allocation are

$$\frac{\partial \mathcal{L}}{\partial c_1} = \omega_1 \frac{\partial u_1}{\partial c_1} - \eta + \theta \frac{\partial u_1}{\partial c_1} = 0, \quad (32)$$

$$\frac{\partial \mathcal{L}}{\partial l_1} = \omega_1 \frac{\partial u_1}{\partial l_1} + \eta a_1 \phi(e_1) F_1(\cdot) + \theta \frac{\partial u_1}{\partial l_1} - \theta \frac{\partial u_1^*}{\partial l_2} l_2 \frac{a_2}{a_1} \frac{\partial \pi^{-1}(\cdot)}{\partial l_1} = 0, \quad (33)$$

$$\frac{\partial \mathcal{L}}{\partial e_1} = \eta (a_1 \phi'(e_1) l_1 F_1(\cdot) - 1) - \theta \frac{\partial u_1^*}{\partial l_2} l_2 \frac{a_2}{a_1} \frac{\partial \pi^{-1}(\cdot)}{\partial e_1} = 0, \quad (34)$$

$$\frac{\partial \mathcal{L}}{\partial c_2} = \omega_2 \frac{\partial u_2}{\partial c_2} - \eta - \theta \frac{\partial u_1^*}{\partial c_2} = 0, \quad (35)$$

$$\frac{\partial \mathcal{L}}{\partial l_2} = \omega_2 \frac{\partial u_2}{\partial l_2} + \eta a_2 \phi(e_2) F_2(\cdot) - \theta \frac{\partial u_1^*}{\partial l_2} \frac{a_2}{a_1} \left(\pi^{-1}(\cdot) - l_2 \frac{\partial \pi^{-1}(\cdot)}{\partial l_2} \right) = 0, \quad (36)$$

$$\frac{\partial \mathcal{L}}{\partial e_2} = \eta (a_2 \phi'(e_2) l_2 F_2(\cdot) - 1) - \theta \frac{\partial u_1^*}{\partial l_2} l_2 \frac{a_2}{a_1} \frac{\partial \pi^{-1}(\cdot)}{\partial e_2} = 0. \quad (37)$$

¹⁵ Allowing for p_n to differ between individuals does not affect the results. The extension with different fractions g_n of both types is trivial.

From the first-order conditions for c_1 (32) and l_1 (33) 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)¹⁶

$$T'_1 = \frac{\theta}{\eta} \frac{1}{\sigma} \frac{\alpha}{(1-\alpha)} \frac{1}{w_1 a_1 \phi(e_2)} \frac{\partial u_1^*}{\partial l_2} < 0. \quad (38)$$

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, see Stern (1982) and 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) and Mirrlees (1971). Further, if the government is not interested in redistribution, the marginal social value of transferring incomes from high-ability to low-ability workers, θ , is zero, and optimal marginal taxes are zero as well.

Manipulation of the first-order condition for c_2 (35) and the first-order condition for l_2 (36) 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)

$$\frac{T'_2}{1 - T'_2} = \frac{\theta}{\eta} \frac{\partial u_1^*}{\partial c_2} + \frac{T'_1}{(1 - T'_2)} \frac{(1 - \alpha)}{\alpha} (\sigma - 1). \quad (39)$$

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 (see 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.

Optimal education subsidies for the high type are derived from combining the first-order conditions for e_1 (34) with the optimal marginal income tax for skilled workers (38). 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)

$$\frac{S'_1}{1 - S'_1} = -\frac{\theta}{\eta} \frac{1}{\sigma} \frac{\alpha}{(1-\alpha)} \frac{1}{w_1 a_1 \phi(e_2)} \frac{\partial u_1^*}{\partial l_2} = -T'_1 > 0. \quad (40)$$

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 subsidy 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 (37) and the tax rate (39) (see Appendix)

$$\frac{S'_2}{1 - S'_2} = \frac{\theta}{\eta} \frac{1}{\sigma} \frac{\alpha}{(1-\alpha)} \frac{1}{w_2 a_1 \phi(e_2)} \frac{\partial u_1^*}{\partial l_2} = \frac{(1 - \alpha)}{\alpha} T'_1 < 0. \quad (41)$$

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.

¹⁶With 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 .

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 non-linear policies, in contrast to optimal linear policies. The intuition is that, by using non-linear 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 giving 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.

The ‘tax-formula result’ of Saez (2004) is not applicable either with non-linear policy instruments, as the production elasticity (i.e., σ) enters optimal tax expressions. Further, indirect taxes/subsidies, such as education subsidies, are not optimally zero under non-linear income taxation with weakly separable preferences. This bolsters the findings by Naito (1999, 2004, 2007), who investigated the desirability of non-zero commodity taxes in similar general equilibrium settings with human capital formation, and comparative advantage. In the current model, education should optimally be taxed or subsidized under non-linear income taxation to exploit factor price changes for redistribution, and the Atkinson and Stiglitz (1976) theorem does not apply to education subsidies. In the absence of general equilibrium effects, education would not be taxed or subsidized on a net basis, and would be weakly efficient, i.e., efficient conditional upon distorted labor supply.

To check whether general equilibrium effects can be quantitatively important for optimal non-linear 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

$$u_n(c_n, l_n) \equiv c_n - \delta \left(\frac{l_n^{1+1/\varepsilon_n}}{1 + 1/\varepsilon_n} \right), \quad n = 1, 2. \quad (42)$$

In the baseline, we set the wage elasticities of labor supply equal for both types at $\varepsilon_1 = \varepsilon_2 = 0.25$. 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 $\beta = 0.25$. The ability parameters are set at $a_1 = 10$ and $a_2 = 5$. Hence, the high-ability worker is twice as ‘smart’ as the low-ability worker. The aggregate production function features a constant elasticity of substitution between skilled and unskilled labor:

$$Y = (\gamma H_1^\rho + (1 - \gamma) H_2^\rho)^{1/\rho}, \quad \sigma \equiv \frac{1}{1 - \rho}. \quad (43)$$

The elasticity of substitution between high and low-skilled labor is set at $\sigma = 1.5$, see Katz and Autor (1999). The share parameter is calibrated at $\gamma = 0.59$ to get an income share of skilled labor in total output of $1 - \alpha = 0.67$ in the absence of government intervention (see also 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 an 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. 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

Table 1: Optimal non-linear tax and education policies

σ	∞	5	2	1.5	0.99	0.75	0.6
γ	0.43	0.48	0.55	0.59	0.67	0.74	0.82
T_1'	0.0%	0.0%	-1.5%	-2.3%	-4.6%	-8.4%	-8.8%
T_2'	30.4%	33.0%	31.1%	31.7%	33.4%	36.1%	38.1%
S_1'	0.0%	0.0%	1.5%	2.3%	4.4%	7.8%	8.1%
S_2'	0.0%	-0.1%	-3.4%	-5.2%	-10.3%	-19.5%	-21.9%
ε_1	1	0.75	0.5	0.25	0.1		
T_1'	-22.2%	-16.1%	-9.4%	-2.3%	-0.1%		
T_2'	19.0%	22.7%	27.0%	31.7%	33.3%		
S_1'	18.2%	13.9%	8.6%	2.3%	0.1%		
S_2'	-75.2%	-46.6%	-23.4%	-5.2%	-0.1%		
ε_2	1	0.75	0.5	0.25	0.1		
T_1'	-0.6%	-0.9%	-1.4%	-2.3%	-3.3%		
T_2'	32.7%	32.5%	32.2%	31.7%	31.2%		
S_1'	0.6%	0.9%	1.4%	2.3%	3.2%		
S_2'	-1.9%	-2.5%	-3.5%	-5.2%	-6.8%		
β	0.1	0.2	0.25	0.3	0.4		
T_1'	-1.5%	-2.0%	-2.3%	-2.8%	-4.1%		
T_2'	32.3%	32.0%	31.7%	31.4%	30.3%		
S_1'	1.5%	2.0%	2.3%	2.7%	4.0%		
S_2'	-3.2%	-4.3%	-5.2%	-6.3%	-10.2%		
ω_1	0.5	0.4	0.3	0.2	0.1	0	
T_1'	0.0%	-0.9%	-1.9%	-2.8%	-3.8%	-4.7%	
T_2'	0.0%	15.9%	27.2%	35.7%	42.2%	47.4%	
S_1'	0.0%	0.9%	1.8%	2.7%	3.6%	4.5%	
S_2'	0.0%	-2.0%	-4.1%	-6.3%	-8.7%	-11.4%	
Λ	0.0	0.5	1.0	1.5	2.0	2.5	
T_1'	-2.3%	-2.3%	-2.3%	-2.3%	-2.3%	-2.3%	
T_2'	31.7%	31.7%	31.7%	31.7%	31.7%	31.7%	
S_1'	2.3%	2.3%	2.3%	2.3%	2.3%	2.3%	
S_2'	-5.2%	-5.2%	-5.2%	-5.2%	-5.2%	-5.2%	

Notes: Utility is given by $u_n(c_n, l_n) \equiv c_n - \delta \left(\frac{l_n^{1+1/\varepsilon_n}}{1+1/\varepsilon_n} \right)$. The production function for human capital is $a_n \phi(e_n) \equiv a_n e_n^\beta$. Output is $Y = (\gamma H_1^\rho + (1 - \gamma) H_2^\rho)^{1/\rho}$, $\sigma \equiv \frac{1}{1-\rho}$, $H_n \equiv a_n \phi(e_n) l_n$. And, social welfare is $\omega_1 u_1 + \omega_2 u_2$. The baseline parameters are: $\varepsilon_1 = \varepsilon_2 = 0.25$, $\delta = 10$, $\beta = 0.25$, $a_1 = 10$, $a_2 = 5$, $\sigma = 1.5$, $\gamma = 0.59$, $\omega_1 = 1 - \omega_2 = 0.25$.

confirm this. The the top-rate is indeed only 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 non-linear subsidies is rather limited as well.

As expected, the optimal tax expressions are most sensitive to changes in the elasticity of substitution σ . For different elasticities σ , we calibrate the parameter γ so as to keep the income share of skilled labor fixed at $1 - \alpha = 0.67$ 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.

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 more difficult 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, the social welfare function, or the government revenue requirement.

4 Conclusions

This paper analyzed the simultaneous setting of optimal linear and non-linear 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 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 over-investment. 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 non-linear policies are found to be fundamentally different. With non-linear 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 impact of non-linear policies on the skill premium. The non-linearity of tax and subsidy schedules is sufficient. Optimal non-linear policies do exploit general equilibrium effects on wages for redistributive purposes. The skilled worker optimally faces

¹⁷Otherwise, the condition that skilled labor income is always higher than unskilled labor income ($w_1 H_1 > w_2 H_2$) may be violated, and the maximization problem will become misspecified.

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 an unskilled worker. However, simulations of optimal non-linear 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 non-zero commodity will be optimal, as Naito (1999, 2004, 2007) has demonstrated in similar settings, but without education policies. However, it remains unclear how optimal education policies will be affected. The current two-type analysis of non-linear 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 non-linear 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, the set of incentive compatibility constraints cannot easily be collapsed into a single differential equation on utility, as in Mirrlees (1971), and one needs 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 (see Jacobs and Bovenberg, 2007). 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.

References

- Allen, Franklin (1982). “Optimal Linear Income Taxation with General Equilibrium Effects on Wages”, *Journal of Public Economics*, 17, 135-143.
- Atkinson, Anthony B., and Joseph E. Stiglitz (1976). “The Design of Tax Structure: Direct versus Indirect Taxation”, *Journal of Public Economics*, 6, 55-75.
- Atkinson, Anthony B., and Joseph E. Stiglitz (1980). *Lectures on Public Economics*, New York: McGraw-Hill.
- Blundell, Richard, and Thomas MaCurdy (1999). “Labor Supply”, in: Ashenfelter, Orley, and David Card (eds), *Handbook of Labor Economics – Vol. 3A*, Amsterdam: Elsevier-North-Holland.
- Bovenberg, A. Lans, and Bas Jacobs (2005). “Redistribution and Education Subsidies are Siamese Twins”, *Journal of Public Economics*, 89, 2005-2035.
- Cameron, Stephen, and Christopher Taber (2004). “Estimation of Educational Borrowing Constraints using Returns to Schooling”, *Journal of Political Economy*, 112, (1), 132-182.
- Card, David (1999), “The Causal Effect of Education on Earnings”, in: Ashenfelter, Orley, and David Card (eds) (1999). *Handbook of Labor Economics - Vol. 3A*, Amsterdam: Elsevier-North Holland.

- Carneiro, Pedro, and James J. Heckman (2004). "Human Capital Policy", in: Friedman, Benjamin M. (ed) (2004). *Inequality in America: What Role for Human Capital Policies?*, Cambridge-MA: MIT Press.
- Diamond, Peter A., and James A. Mirrlees (1971). "Optimal Taxation and Public Production I: Production Efficiency", *American Economic Review* 61, 8-27.
- Diamond, Peter A. (1998). "Optimal Income Taxation: An Example with a U-Shaped Pattern of Optimal Marginal Tax Rates", *American Economic Review*, 88, 83-95.
- Dixit, Avinash K. and, Agnar Sandmo (1977). "Some Simplified Formulae for Optimal Income Taxation", *Scandinavian Journal of Economics*, 79, (4), 417-423.
- Dur, Robert, and Coen N. Teulings (2001). "Education and Efficient Redistribution", Tinbergen Institute Discussion Paper 090/3.
- Dur, Robert, and Coen N. Teulings (2004). "Are Education Subsidies an Efficient Redistributive Device", in: Agnell, Jonas et al. (eds) (2004). *Labor Market Institutions and Public Regulation*, Cambridge-MA: MIT-Press.
- Feldstein, Martin (1973). "On the Optimal Progressivity of the Income Tax", *Journal of Public Economics*, 2, 357-366.
- Gruber, Jonathan, and Emmanuel Saez (2003). "The Elasticity of Taxable Income: Evidence and Implications", *Journal of Public Economics*, 84, 1-32.
- Heckman, James J., Lance Lochner, and Christopher Taber (1998). "A Study of Tuition Policy", *American Economic Review*, 88, (2), 381-386.
- Jacobs, Bas, and A. Lans Bovenberg (2007). "Generalized Results on Optimal Taxation and Education Policy and Conditions for Weak Efficiency in Human Capital Production", mimeo: Erasmus Universiteit Rotterdam/Universiteit van Tilburg.
- Katz, Lawrence F. and David H. Autor (1999). "Changes in the Wage Structure and Wage Inequality", in: Ashenfelter, Orley, and David Card (eds) (1999), *Handbook of Labor Economics Vol. 3A*, Elsevier Science BV.
- Mirrlees, James A. (1971). "An Exploration in the Theory of Optimum Income Taxation", *Review of Economic Studies*, 38, 175-208.
- Naito, Hisahiro (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, 165-188.
- Naito, Hisahiro (2004). "Endogenous Human Capital Accumulation, Comparative Advantage and Direct vs. Indirect Redistribution", *Journal of Public Economics*, 88, 2685-2710.
- Naito, Hisahiro (2007). "Atkinson-Stiglitz Theorem with Endogenous Human Capital Accumulation", *B.E. Journal of Economic Analysis & Policy*, 7, (1), article 46.
- Saez, Emmanuel (2001). "Using Elasticities to Derive Optimal Income Tax Rates," *Review of Economic Studies*, 68, 205-229.
- Saez, Emmanuel (2004). "Direct or Indirect Instruments for Redistribution: Short-Run versus Long-Run", *Journal of Public Economics*, 88, 503-518.

- Seade, Jesus K. (1977). “On the Shape of Optimal Tax Schedules”, *Journal of Public Economics*, 7, 203-236.
- Sheshinski, Eytan (1971). “The Optimal Linear Income-Tax”, *Review of Economic Studies*, 39, (3), 297-302.
- Stern, Nicholas H. (1982). “Optimum Income Taxation with Errors in Administration”, *Journal of Public Economics*, 17, 181-211.
- Stiglitz, Joseph E. (1982). “Self-Selection and Pareto Efficient Taxation”, *Journal of Public Economics*, 17, 213-240.
- Tinbergen, Jan (1975). *Income Distribution*, Amsterdam: North-Holland Publishing Company BV.

Appendix

Second-order conditions of individual optimization

By substituting the household budget constraint (3) into the utility function (1) to eliminate c_n , we arrive at the following unconstrained maximization problem

$$\max_{\{l_n, e_n\}} U_n = (1-t)(w_n a_n \phi(e_n) l_n - (1-s)p_n e_n) + b - \frac{l_n^{1+1/\varepsilon_n}}{1+1/\varepsilon_n}. \quad (44)$$

The first-order conditions are

$$\frac{\partial U_n}{\partial l_n} = (1-t)w_n a_n \phi(e_n) - l_n^{1/\varepsilon_n} = 0, \quad (45)$$

$$\frac{\partial U_n}{\partial e_n} = (1-t)(w_n a_n \phi'(e_n) l_n - (1-s)p_n) = 0. \quad (46)$$

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

$$H \equiv \begin{bmatrix} -\frac{1}{\varepsilon_n} l_n^{1/\varepsilon_n - 1} & (1-t)w_n a_n \phi'(e_n) \\ (1-t)w_n a_n \phi'(e_n) & (1-t)w_n a_n l_n \phi''(e_n) \end{bmatrix}. \quad (47)$$

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 signs. The first principal minor is negative. Therefore, the second leading principal minor must be positive, i.e., $-\frac{1}{\varepsilon_n} l_n^{1/\varepsilon_n} (1-t)w_n a_n \phi''(e_n) - ((1-t)w_n a_n \phi'(e_n))^2 > 0$. Using (4) to eliminate l_n and substituting (2), this inequality can be written as

$$\mu_n \equiv 1 - \beta(1 + \varepsilon_n) > 0. \quad (48)$$

Elasticities of individual behavior

Log-linearizing (5) (using $\phi(e_n) = e_n^\beta$) gives

$$\tilde{l}_n + (\beta - 1)\tilde{e}_n = -\tilde{s}. \quad (49)$$

A tilde stands for a relative change (i.e., $\tilde{l}_n \equiv dl_n/l_n$, $\tilde{e}_n \equiv de_n/e_n$, et cetera), except for the tax rate and the subsidy rates, where $\tilde{t} \equiv dt/(1-t)$, and $\tilde{s} \equiv ds/(1-s)$.

Expression (4) implies that labor supply depends only on the after-tax wage rate $(1 - t)w_n a_n \phi(e_n)$ so that

$$\tilde{l}_n = \varepsilon_n(\beta \tilde{e}_n - \tilde{t}). \quad (50)$$

Substituting (50) into (49) to eliminate $\tilde{l}_n$, an expression for $\tilde{e}_n$ is found

$$\tilde{e}_n = \frac{1}{\mu_n} \tilde{s} - \frac{\varepsilon_n}{\mu_n} \tilde{t}, \quad (51)$$

where $\mu_n \equiv 1 - \beta(1 + \varepsilon_n) > 0$. Substitution of (51) into (50), gives a solution for $\tilde{l}_n$

$$\tilde{l}_n = \frac{\beta \varepsilon_n}{\mu_n} \tilde{s} - \frac{\varepsilon_n(1 - \beta)}{\mu} \tilde{t}. \quad (52)$$

Therefore, the following elasticities of l_n and e_n with respect to the policy parameters are obtained

$$\varepsilon_n^{lt} \equiv -\frac{\partial l_n}{\partial t} \frac{(1 - t)}{l_n} = \frac{\varepsilon_n(1 - \beta)}{\mu_n}, \quad (53)$$

$$\varepsilon_n^{et} \equiv -\frac{\partial e_n}{\partial t} \frac{(1 - t)}{e_n} = \frac{\varepsilon_n}{\mu_n}, \quad (54)$$

$$\varepsilon_n^{ls} \equiv \frac{\partial l_n}{\partial s} \frac{(1 - s)}{l_n} = \frac{\beta \varepsilon_n}{\mu_n}, \quad (55)$$

$$\varepsilon_n^{es} \equiv \frac{\partial e_n}{\partial s} \frac{(1 - s)}{e_n} = \frac{1}{\mu_n}. \quad (56)$$

Optimal linear income taxation

The first-order condition for the linear income tax can be rewritten by substituting $w_n a_n \phi(e_n) l_n - p_n(1 - s)e_n = (1 - \beta)z_n$, and $p_n e_n = \frac{\beta}{1 - s} z_n$ (both are derived from the first-order condition for learning):

$$\begin{aligned} \frac{\partial \mathcal{L}}{\partial t} = & -\sum_n \omega_n (1 - \beta) z_n g_n + \eta \sum_n (1 - \beta) z_n g_n \\ & + \eta \frac{t}{1 - t} \sum_n \left(z_n \frac{\partial l_n}{\partial t} \frac{1 - t}{l_n} \right) g_n - \eta \frac{s}{(1 - t)(1 - s)} \sum_n z_n \beta \frac{\partial e_n}{\partial t} \frac{1 - t}{e_n} g_n \\ & + \left(\omega_1 + \eta \frac{t}{1 - t} \right) H_1 \left(H_1 F_{11} \frac{\partial H_1}{\partial t} \frac{1 - t}{H_1} + H_2 F_{12} \frac{\partial H_2}{\partial t} \frac{1 - t}{H_2} \right) \\ & + \left(\omega_2 + \eta \frac{t}{1 - t} \right) H_2 \left(H_1 F_{12} \frac{\partial H_1}{\partial t} \frac{1 - t}{H_1} + H_2 F_{22} \frac{\partial H_2}{\partial t} \frac{1 - t}{H_2} \right) = 0. \end{aligned} \quad (57)$$

Next, use the distributional characteristic (16), divide by average income $\bar{z} = F(\cdot)$, and η . Use the property that the labor demand functions are homogenous of degree zero. Substitution of (21) and (22) gives (note that symmetry of the substitution matrix implies $\frac{\partial w_1}{\partial H_2} = \frac{\partial w_2}{\partial H_1} = F_{12}$)

$$(1 - \beta)\xi - \frac{t}{1 - t} \bar{\varepsilon}^{lt} + \frac{s}{(1 - t)(1 - s)} \beta \bar{\varepsilon}^{et} - \left(\frac{\omega_2}{\eta} - \frac{\omega_1}{\eta} \right) (\varepsilon_1^{Ht} - \varepsilon_2^{Ht}) \frac{H_1 F_1}{F} \frac{H_2 F_2}{F} \frac{F_{12} F}{F_1 F_2} = 0, \quad (58)$$

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 $1 - \alpha \equiv \frac{H_1 F_1}{F}$, $\alpha \equiv \frac{H_2 F_2}{F}$, and $\frac{1}{\sigma} \equiv \frac{F_{12} F}{F_1 F_2}$, and simplify to obtain

$$\xi(1 - \beta) - \frac{t}{1 - t} \bar{\varepsilon}^{lt} + \frac{s}{(1 - t)(1 - s)} \beta \bar{\varepsilon}^{et} - \left(\frac{\omega_2}{\eta} - \frac{\omega_1}{\eta} \right) \left(\frac{\varepsilon_1}{\mu_1} - \frac{\varepsilon_2}{\mu_2} \right) \alpha(1 - \alpha) \frac{1}{\sigma} = 0. \quad (59)$$

To find the optimal linear income without education subsidies, set $s = 0$. Rearranging yields the expression in the main text.

Optimal education subsidies

The first order condition for s can be rewritten by rearranging, and substituting $p_n e_n = \frac{\beta}{1-s} z_n$ (derived from the first-order condition for learning):

$$\begin{aligned} \frac{\partial \mathcal{L}}{\partial s} = & \sum_n \omega_n \left(\frac{1-t}{1-s} \right) \beta z_n g_n - \eta \sum_n \left(\frac{1-t}{1-s} \right) \beta z_n g_n \\ & + \eta \frac{t}{1-s} \sum_n \left(z_n \frac{\partial l_n}{\partial s} \frac{1-s}{l_n} g_n \right) - \eta \frac{s}{1-s} \sum_n \frac{1}{1-s} \beta z_n \frac{\partial e_n}{\partial s} \frac{1-s}{e_n} g_n \\ & + \left(\omega_1 \left(\frac{1-t}{1-s} \right) + \eta \frac{t}{1-s} \right) H_1 \left(H_1 F_{11} \frac{\partial H_1}{\partial s} \frac{1-s}{H_1} + H_2 F_{12} \frac{\partial H_2}{\partial s} \frac{1-s}{H_2} \right) \\ & + \left(\omega_2 \left(\frac{1-t}{1-s} \right) + \eta \frac{t}{1-s} H_2 \right) H_2 \left(H_1 F_{12} \frac{\partial H_1}{\partial s} \frac{1-s}{H_1} + H_2 F_{22} \frac{\partial H_2}{\partial s} \frac{1-s}{H_2} \right) = 0. \end{aligned} \quad (60)$$

Next, use the distributional characteristic (17), 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 (21) and (22) gives (note that symmetry of the substitution matrix implies $\frac{\partial w_1}{\partial H_2} = \frac{\partial w_2}{\partial H_1} = F_{12}$)

$$-\xi + \frac{t}{1-t} \frac{\bar{\varepsilon}^{ls}}{\beta} - \frac{s}{(1-s)(1-t)} \bar{\varepsilon}^{es} + \left(\frac{\omega_2}{\eta} - \frac{\omega_1}{\eta} \right) \left(\frac{\varepsilon_1^{Hs} - \varepsilon_2^{Hs}}{\beta} \right) \frac{H_1 F_1}{F} \frac{H_2 F_2}{F} \frac{F_{12} F}{F_1 F_2}, \quad (61)$$

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

$$\xi - \frac{t}{1-t} \frac{\bar{\varepsilon}^{ls}}{\beta} + \frac{s}{(1-s)(1-t)} \bar{\varepsilon}^{es} - \left(\frac{\omega_2}{\eta} - \frac{\omega_1}{\eta} \right) \left(\frac{1+\varepsilon_1}{\mu_1} - \frac{1+\varepsilon_2}{\mu_2} \right) \alpha (1-\alpha) \frac{1}{\sigma}. \quad (62)$$

By setting $t = 0$, the equation of optimal subsidies in the text is obtained.

Optimal tax and education policies

The first-order conditions for t and s are

$$\xi - \frac{t}{1-t} \frac{\bar{\varepsilon}^{lt}}{(1-\beta)} + \frac{s}{(1-t)(1-s)} \frac{\beta \bar{\varepsilon}^{et}}{(1-\beta)} - \left(\frac{\omega_2}{\eta} - \frac{\omega_1}{\eta} \right) \left(\frac{\varepsilon_1^{Ht} - \varepsilon_2^{Ht}}{1-\beta} \right) \alpha (1-\alpha) \frac{1}{\sigma} = 0, \quad (63)$$

$$\xi - \frac{t}{1-t} \frac{\bar{\varepsilon}^{ls}}{\beta} + \frac{s}{(1-s)(1-t)} \bar{\varepsilon}^{es} - \left(\frac{\omega_2}{\eta} - \frac{\omega_1}{\eta} \right) \left(\frac{\varepsilon_1^{Hs} - \varepsilon_2^{Hs}}{\beta} \right) \alpha (1-\alpha) \frac{1}{\sigma} = 0. \quad (64)$$

Subtracting yields

$$\begin{aligned} \frac{t}{1-t} \left(\frac{\bar{\varepsilon}^{lt}}{1-\beta} - \frac{\bar{\varepsilon}^{ls}}{\beta} \right) = & \frac{s}{(1-s)(1-t)} \left(\frac{\beta \bar{\varepsilon}^{et}}{1-\beta} - \bar{\varepsilon}^{es} \right) - \\ & \frac{\alpha(1-\alpha)}{\sigma} \left(\frac{\omega_2}{\eta} - \frac{\omega_1}{\eta} \right) \left(\frac{\varepsilon_1^{Ht} - \varepsilon_2^{Ht}}{1-\beta} - \frac{\varepsilon_1^{Hs} - \varepsilon_2^{Hs}}{\beta} \right) = 0. \end{aligned} \quad (65)$$

Substituting the relevant elasticities (see earlier in Appendix) yields $\frac{\varepsilon_n^{lt}}{1-\beta} - \frac{\varepsilon_n^{ls}}{\beta} = 0$, $\frac{\beta \bar{\varepsilon}^{et}}{1-\beta} - \bar{\varepsilon}^{es} = \frac{-1}{1-\beta}$, $\frac{\varepsilon_1^{Ht} - \varepsilon_2^{Ht}}{1-\beta} - \frac{\varepsilon_1^{Hs} - \varepsilon_2^{Hs}}{\beta} = 0$. 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 non-linear tax policies

The optimal marginal tax rate on the high-ability type follows from multiplying the first-order condition for c_1 (32) by $\frac{\partial u_1/\partial l_1}{\partial u_1/\partial c_1}$, and substituting the result in (33). 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

$$T'_1 = \frac{\theta}{\eta} \frac{l_2}{w_1 a_1 \phi(e_1)} \frac{a_2}{a_1} \frac{\partial u_1^*}{\partial l_2} \frac{\partial \pi^{-1}}{\partial l_1} = \frac{\theta}{\eta} \frac{l_2}{w_1} \frac{a_2}{a_1} \frac{\partial u_1^*}{\partial l_2} \frac{\partial \pi^{-1}}{\partial H_1}. \quad (66)$$

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:

$$\frac{\partial(\pi^{-1})}{\partial H_1} = \frac{\partial(F_2/F_1)}{\partial H_1} = \frac{F_{21}F_1 - F_{11}F_2}{F_1^2} = \frac{1}{H_2} \frac{1}{\sigma} \frac{\alpha}{(1-\alpha)}, \quad (67)$$

where the property is used that the labor demand function is homogeneous of degree zero (21). Substitution of the last result yields the equation in the text.

Repeating the same procedure with the first-order conditions for c_2 (35) and for l_2 (36) yields the marginal tax on unskilled labor income $T'_2 \equiv 1 + \frac{1}{w_2 a_2 \phi(e_2)} \frac{\partial u_2/\partial l_2}{\partial u_2/\partial c_2}$,

$$T'_2 = (1 - T'_2) \frac{\theta}{\eta} \frac{\partial u_1^*}{\partial c_2} + \frac{1}{w_2 a_2 \phi(e_2)} \frac{\theta}{\eta} \frac{\partial u_1^*}{\partial l_2} \frac{a_2}{a_1} \left(\pi^{-1}(\cdot) + l_2 \frac{\partial \pi^{-1}(\cdot)}{\partial l_2} \right). \quad (68)$$

$\pi^{-1} + l_2 \frac{\partial \pi^{-1}}{\partial l_2}$ 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 (22) is used:

$$\frac{\partial(\pi^{-1})}{\partial H_2} = \frac{\partial(F_2/F_1)}{\partial H_2} = \frac{F_{22}F_1 - F_{12}F_2}{F_1^2} = -\frac{1}{H_2} \frac{H_1}{H_2} \frac{1}{\sigma} \frac{\alpha}{(1-\alpha)}. \quad (69)$$

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

$$T'_2 = (1 - T'_2) \frac{\theta}{\eta} \frac{\partial u_1^*}{\partial c_2} + T'_1 \left(\frac{1-\alpha}{\alpha} \right) (\sigma - 1). \quad (70)$$

Rearranging yields the expression in the text.

Optimal non-linear education policies

Optimal education subsidies for the skilled worker follow from combining the expression of the optimal income tax (38) with the first-order condition for e_1 (34). First, rewrite the first-order condition for e_1 using the definition of the subsidy on human capital investment $S'_1 \equiv 1 - w_1 a_1 \phi'(e_1) l_1$,

$$-S'_1 = \frac{\theta}{\eta} \frac{\partial u_1^*}{\partial l_2} l_2 \frac{a_2}{a_1} \frac{\partial \pi^{-1}}{\partial e_1}. \quad (71)$$

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 $S'_2 \equiv 1 - w_2 a_2 \phi'(e_2) l_2$

$$-S'_2 = \frac{\theta}{\eta} \frac{\partial u_1^*}{\partial l_2} l_2 \frac{a_2}{a_1} \frac{\partial \pi^{-1}}{\partial e_2}. \quad (72)$$

Using the properties of the production function and substituting $1 - S'_2 = w_2 a_2 \phi'(e_2) l_2$ we obtain

$$S'_2 = \frac{\theta}{\eta} \frac{\partial u_1^*}{\partial l_2} l_2 \frac{a_2}{a_1} \frac{1}{w_2 H_2} \frac{H_1}{H_2} \frac{1}{\sigma} \frac{\alpha}{(1-\alpha)} (1 - S'_2). \quad (73)$$

Dividing the last equation by $(1 - S'_2)$ and substitution of $T'_1 w_1 a_1 \phi(e_2) = \frac{\theta}{\eta} \frac{1}{\sigma} \frac{\alpha}{(1-\alpha)} \frac{\partial u_1^*}{\partial l_2}$ gives the equation in the text.

CESifo Working Paper Series

for full list see www.cesifo-group.org/wp

(address: Poschingerstr. 5, 81679 Munich, Germany, office@cesifo.de)

- 2100 Gunther Schnabl and Andreas Hoffmann, Monetary Policy, Vagabonding Liquidity and Bursting Bubbles in New and Emerging Markets – An Overinvestment View, September 2007
- 2101 Panu Poutvaara, The Expansion of Higher Education and Time-Consistent Taxation, September 2007
- 2102 Marko Koethenbueger and Ben Lockwood, Does Tax Competition Really Promote Growth?, September 2007
- 2103 M. Hashem Pesaran and Elisa Tosetti, Large Panels with Common Factors and Spatial Correlations, September 2007
- 2104 Laszlo Goerke and Marco Runkel, Tax Evasion and Competition, September 2007
- 2105 Scott Alan Carson, Slave Prices, Geography and Insolation in 19th Century African-American Stature, September 2007
- 2106 Wolfram F. Richter, Efficient Tax Policy Ranks Education Higher than Saving, October 2007
- 2107 Jarko Fidrmuc and Roman Horváth, Volatility of Exchange Rates in Selected New EU Members: Evidence from Daily Data, October 2007
- 2108 Torben M. Andersen and Michael Svarer, Flexicurity – Labour Market Performance in Denmark, October 2007
- 2109 Jonathan P. Thomas and Tim Worrall, Limited Commitment Models of the Labor Market, October 2007
- 2110 Carlos Pestana Barros, Guglielmo Maria Caporale and Luis A. Gil-Alana, Identification of Segments of European Banks with a Latent Class Frontier Model, October 2007
- 2111 Felicitas Nowak-Lehmann D., Sebastian Vollmer and Immaculada Martínez-Zarzoso, Competitiveness – A Comparison of China and Mexico, October 2007
- 2112 Mark Mink, Jan P.A.M. Jacobs and Jakob de Haan, Measuring Synchronicity and Co-movement of Business Cycles with an Application to the Euro Area, October 2007
- 2113 Ossip Hühnerbein and Tobias Seidel, Intra-regional Tax Competition and Economic Geography, October 2007
- 2114 Christian Keuschnigg, Exports, Foreign Direct Investment and the Costs of Corporate Taxation, October 2007

- 2115 Werner Bönnte, Oliver Falck and Stephan Heblich, Demography and Innovative Entrepreneurship, October 2007
- 2116 Katrin Assenmacher-Wesche and M. Hashem Pesaran, Assessing Forecast Uncertainties in a VECX Model for Switzerland: An Exercise in Forecast Combination across Models and Observation Windows, October 2007
- 2117 Ben Lockwood, Voting, Lobbying, and the Decentralization Theorem, October 2007
- 2118 Andrea Ichino, Guido Schwerdt, Rudolf Winter-Ebmer and Josef Zweimüller, Too Old to Work, too Young to Retire?, October 2007
- 2119 Wolfgang Eggert, Tim Krieger and Volker Meier, Education, Unemployment and Migration, October 2007
- 2120 Stefan Napel and Mika Widgrén, The European Commission – Appointment, Preferences, and Institutional Relations, October 2007
- 2121 Bertil Holmlund and Martin Söderström, Estimating Income Responses to Tax Changes: A Dynamic Panel Data Approach, October 2007
- 2122 Doina Maria Radulescu, From Separate Accounting to Formula Apportionment: Analysis in a Dynamic Framework, October 2007
- 2123 Jelle Brouwer, Richard Paap and Jean-Marie Viaene, The Trade and FDI Effects of EMU Enlargement, October 2007
- 2124 Kurt R. Brekke, Luigi Siciliani and Odd Rune Straume, Competition and Waiting Times in Hospital Markets, October 2007
- 2125 Alexis Direr, Flexible Life Annuities, October 2007
- 2126 Johannes Becker and Clemens Fuest, Quality versus Quantity – The Composition Effect of Corporate Taxation on Foreign Direct Investment, October 2007
- 2127 Balázs Égert, Real Convergence, Price Level Convergence and Inflation Differentials in Europe, October 2007
- 2128 Marko Koethenbuerger, Revisiting the “Decentralization Theorem” – On the Role of Externalities, October 2007
- 2129 Axel Dreher, Silvia Marchesi and James Raymond Vreeland, The Politics of IMF Forecasts, October 2007
- 2130 Andreas Knabe and Ronnie Schöb, Subsidizing Extra Jobs: Promoting Employment by Taming the Unions, October 2007
- 2131 Michel Beine and Bertrand Candelon, Liberalization and Stock Market Co-Movement between Emerging Economies, October 2007

- 2132 Dieter M. Urban, FDI Technology Spillovers and Wages, October 2007
- 2133 Valentina Bosetti, Carlo Carraro, Emanuele Massetti and Massimo Tavoni, Optimal Energy Investment and R&D Strategies to Stabilise Greenhouse Gas Atmospheric Concentrations, October 2007
- 2134 David-Jan Jansen and Jakob de Haan, The Importance of Being Vigilant: Has ECB Communication Influenced Euro Area Inflation Expectations?, October 2007
- 2135 Oliver Falck, Heavyweights – The Impact of Large Businesses on Productivity Growth, October 2007
- 2136 Xavier Freixas and Bruno M. Parigi, Banking Regulation and Prompt Corrective Action, November 2007
- 2137 Jan K. Brueckner, Partial Fiscal Decentralization, November 2007
- 2138 Silvia Console Battilana, Uncovered Power: External Agenda Setting, Sophisticated Voting, and Transnational Lobbying, November 2007
- 2139 Alan J. Auerbach, Michael P. Devereux and Helen Simpson, Taxing Corporate Income, November 2007
- 2140 Lorenzo Cappellari, Paolo Ghinetti and Gilberto Turati, On Time and Money Donations, November 2007
- 2141 Roel Beetsma and Heikki Oksanen, Pension Systems, Ageing and the Stability and Growth Pact, November 2007
- 2142 Hikaru Ogawa and David E. Wildasin, Think Locally, Act Locally: Spillovers, Spillbacks, and Efficient Decentralized Policymaking, November 2007
- 2143 Alessandro Cigno, A Theoretical Analysis of the Effects of Legislation on Marriage, Fertility, Domestic Division of Labour, and the Education of Children, November 2007
- 2144 Kai A. Konrad, Mobile Tax Base as a Global Common, November 2007
- 2145 Ola Kvaløy and Trond E. Olsen, The Rise of Individual Performance Pay, November 2007
- 2146 Guglielmo Maria Caporale, Yannis Georgellis, Nicholas Tsitsianis and Ya Ping Yin, Income and Happiness across Europe: Do Reference Values Matter?, November 2007
- 2147 Dan Anderberg, Tax Credits, Income Support and Partnership Decisions, November 2007
- 2148 Andreas Irmen and Rainer Klump, Factor Substitution, Income Distribution, and Growth in a Generalized Neoclassical Model, November 2007

- 2149 Lorenz Blume, Jens Müller and Stefan Voigt, The Economic Effects of Direct Democracy – A First Global Assessment, November 2007
- 2150 Axel Dreher, Pierre-Guillaume Méon and Friedrich Schneider, The Devil is in the Shadow – Do Institutions Affect Income and Productivity or only Official Income and Official Productivity?, November 2007
- 2151 Valentina Bosetti, Carlo Carraro, Emanuele Massetti and Massimo Tavoni, International Energy R&D Spillovers and the Economics of Greenhouse Gas Atmospheric Stabilization, November 2007
- 2152 Balázs Égert and Dubravko Mihaljek, Determinants of House Prices in Central and Eastern Europe, November 2007
- 2153 Christa Hainz and Hendrik Hakenes, The Politician and his Banker, November 2007
- 2154 Josef Falkinger, Distribution and Use of Knowledge under the “Laws of the Web”, December 2007
- 2155 Thorvaldur Gylfason and Eduard Hochreiter, Growing Apart? A Tale of Two Republics: Estonia and Georgia, December 2007
- 2156 Morris A. Davis and François Ortalo-Magné, Household Expenditures, Wages, Rents, December 2007
- 2157 Andreas Haufler and Christian Schulte, Merger Policy and Tax Competition, December 2007
- 2158 Marko Köthenbürger and Panu Poutvaara, Rent Taxation and its Intertemporal Welfare Effects in a Small Open Economy, December 2007
- 2159 Betsey Stevenson, Title IX and the Evolution of High School Sports, December 2007
- 2160 Stergios Skaperdas and Samarth Vaidya, Persuasion as a Contest, December 2007
- 2161 Morten Bønnedsen and Christian Schultz, Arm’s Length Provision of Public Services, December 2007
- 2162 Bas Jacobs, Optimal Redistributive Tax and Education Policies in General Equilibrium, December 2007