



ELSEVIER

Journal of Public Economics 89 (2005) 2005–2035

JOURNAL OF
PUBLIC
ECONOMICS

www.elsevier.com/locate/econbase

Redistribution and education subsidies are Siamese twins

A. Lans Bovenberg^{a,*}, Bas Jacobs^b

^aUniversity of Tilburg, Netspar, and CEPR, The Netherlands

^bUniversity of Tilburg, Netspar, University of Amsterdam, and Tinbergen Institute, The Netherlands

Received 17 October 2002; received in revised form 10 November 2004; accepted 13 December 2004

Available online 13 May 2005

Abstract

We develop models of optimal linear and non-linear income taxation with endogenous human capital formation to explore optimal education subsidies. Optimal subsidies on education ensure efficiency in human capital accumulation and thus play an important role in alleviating the tax distortions on learning induced by redistributive policies. If the government cannot verify all investments in human capital, education policy offsets some but not all tax-induced distortions on learning. Non-pecuniary educational costs (benefits) may increase (decrease) subsidies on education, especially if they are complementary to work effort.

© 2005 Elsevier B.V. All rights reserved.

JEL classification: H2; H5; I2; J2

Keywords: Human capital; Education subsidies; Progressive taxation; Redistribution

1. Introduction

Most OECD countries heavily subsidize higher education. These education subsidies are typically justified on the basis of perceived positive external effects of human capital

* Corresponding author. CentER, Tilburg University, P.O. Box 90153, 5000 LE Tilburg, The Netherlands. Tel.: +31 13 466 2912; fax: +31 13 466 3066.

E-mail addresses: a.l.bovenberg@uvt.nl (A. Lans Bovenberg), bas.jacobs@iue.it (B. Jacobs).

accumulation, capital market imperfections, and redistributive concerns. Positive external effects of higher education, however, are difficult to establish empirically (see, e.g., Heckman and Klenow, 1998; Acemoglu and Angrist, 2000; Krueger and Lindahl, 2002). Also capital market imperfections do not seem to be very important in practice (see, e.g., Shea, 2000; Cameron and Taber, 2000; Carneiro and Heckman, 2004).¹ Education subsidies, moreover, typically transfer resources away from unskilled towards skilled, educated individuals so that their distributional effects appear perverse.² Why, then, is education subsidized?

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

Our paper explores the interaction between public spending and tax policies by viewing education subsidies and tax policies as interdependent instruments aimed at redistribution.⁴ We add endogenous human capital formation to the standard models of optimal income taxation and investigate how the availability of education subsidies affects the optimal income tax system.⁵ Our paper contributes in a number of ways to the optimal tax literature.

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

Second, if the government has education subsidies at its disposal, the distortions of redistributive taxes on learning decisions are eliminated so as to restore efficiency in

¹ In any case, loans rather than subsidies are the most direct way to address liquidity constraints (see Jacobs and van Wijnbergen, 2005).

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

³ Boadway et al. (1996) and Andersson and Konrad (2003) offer an alternative explanation for education subsidies. They argue that education subsidies are called for if the government cannot commit, and engages in excessive redistribution after investments in human capital have been made. We show that education subsidies are optimal even if the government can commit to announced policies.

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

⁵ Our paper is related to van Ewijk and Tang (2000). In this study, the government adopts progressive taxes on labor income to raise employment by reducing wage demands of unions. Education subsidies allow the government to employ progressive taxes without distorting human capital accumulation. van Ewijk and Tang (2000) abstract from distributional concerns as the main reason for imposing progressive income taxes.

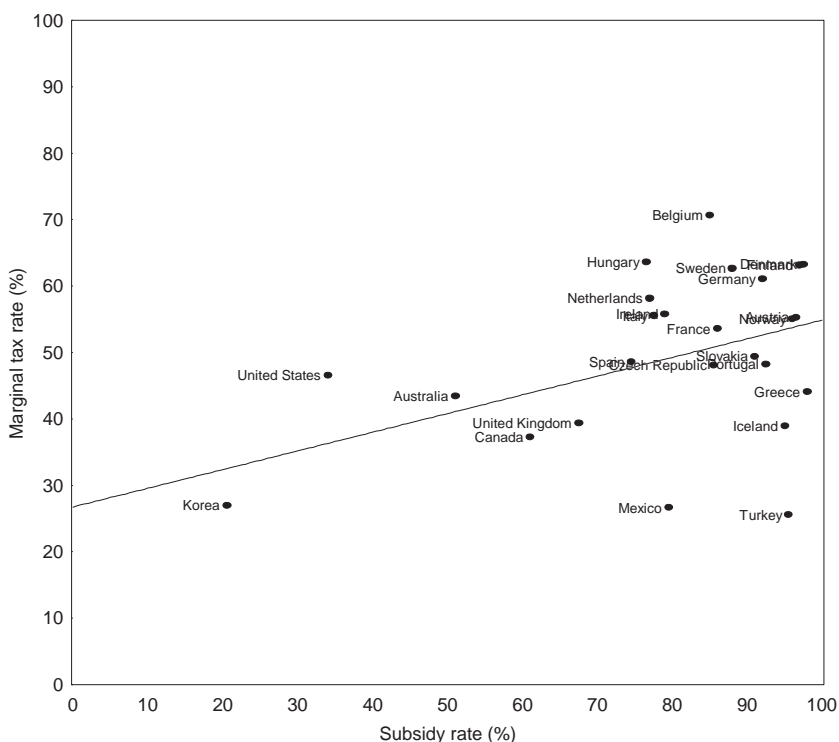


Fig. 1. Marginal taxes and education subsidies in various OECD countries.

education choices. This result can be viewed as an application of the celebrated production efficiency theorem of [Diamond and Mirrlees \(1971\)](#) to a model with individual production functions for human capital. By eliminating the distortions of redistribution on learning, education subsidies make the optimal labor tax more progressive than it would be in the absence of education subsidies. However, compared to the case with exogenous learning, optimal marginal tax rates on labor income are lower. With endogenous human capital, positive marginal tax rates depress after-tax wages not only directly but also indirectly by reducing human capital accumulation through a lower utilization rate of human capital. Hence, the adverse substitution effects associated with redistribution become larger, and optimal marginal taxes decline.

Third, we show that the government optimally subsidizes education if it can freely employ a redistributive income tax—even though the rich benefit more than proportionally from these subsidies. The reason is that education subsidies are aimed at eliminating the adverse impact of redistributive taxes on the incentives to accumulate human capital, whereas the income tax is targeted at redistributing incomes. The combination of taxes and subsidies allows the government to extract rents from ability with smaller distortions on human capital formation. Hence, higher labor taxes more than offset the benefits from education subsidies to the most skilled agents. In this way, progressive taxation and education benefits are Siamese twins.

We investigate whether these three contributions are robust with respect to the observability of learning and individual incomes and the pecuniary nature of educational costs and benefits. If the government is not able to verify some investments in human capital, it cannot subsidize all investments. Hence, the government offsets some, but not all, tax-induced distortions on human capital accumulation. Intuitively, education subsidies on verifiable investments help to alleviate the adverse impact of taxes on the *level* of investments in human capital, but at the same time distort the *composition* of the inputs invested in human capital away from the non-verifiable inputs. We also establish that the main results derived with linear policy instruments continue to hold if the government observes individual incomes and individual learning efforts so that it can levy non-linear income tax and non-linear education subsidies. Finally, by allowing for non-verifiable non-pecuniary inputs (which are complements to verifiable pecuniary inputs) and non-pecuniary outputs of education, we show that the case for subsidies on verifiable educational inputs is strengthened (weakened) with non-pecuniary costs (benefits). The intuition is that distortions of redistributive tax systems increase (decrease) if a larger part of the costs (benefits) cannot be subsidized (are not taxed). Furthermore, the government should optimally subsidize education more heavily if non-pecuniary costs and benefits of education are complementary to work effort. The reason is that additional education then helps to offset the distortionary impact of redistributive taxes on work effort.

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

The rest of this paper is structured as follows. Section 2 explores optimal linear tax and education policies. After Section 3 extends the analysis to non-verifiable investments in human capital, Section 4 shows that the main results continue to hold if non-linear policy instruments are available. Section 5 considers non-monetary effort costs and consumption benefits. Before Section 7 concludes, Section 6 discusses the policy implications of the analysis. Four appendices contain the technical derivations of the results.

2. Optimal linear tax and education policies

2.1. Preferences and technologies

Our model extends the standard model of optimal income taxation by endogenizing individuals' earning potentials through human capital formation. Individuals are heterogeneous with respect to their exogenous ability n . The mass of individuals has unit measure, and the cumulative distribution of ability is denoted by $F(n)$. $f(n)$ is the corresponding density function, which is continuously differentiable and strictly positive on the support $[\underline{n}, \bar{n}]$, $\underline{n}, \bar{n} > 0$. The upper bound \bar{n} may be infinite. The government knows the distribution of abilities, but cannot observe individual ability. Accordingly, it cannot

levy individual-specific lump-sum taxes to redistribute incomes, but must rely on distortionary taxes instead.

Investment in human capital is denoted e_n (for education). The production function for human capital is homothetic and given by⁶

$$h_n = n\phi(e_n) \equiv ne_n^\beta, \tag{1}$$

where h_n denotes human capital. Human capital accumulation features decreasing returns with respect to investment, i.e., $\beta < 1$. Ability n is the productivity of human capital investments. Ability and educational investments are complementary inputs in generating human capital, i.e., $(\partial^2 h_n)/(\partial e_n \partial n) = \beta ne_n^{\beta-1} > 0$. Hence, more able individuals produce more human capital with the same educational investment. Gross labor income z_n is the product of the number of efficiency units of human capital, h_n , and hours worked l_n , i.e., $z_n \equiv h_n l_n = n\phi(e_n)l_n$.⁷

Households derive utility from consumption c_n and suffer disutility from work effort l_n . The utility function is separable without income effects and features a constant elasticity of labor supply (as in Diamond, 1998):

$$u(c_n, l_n) \equiv c_n - \frac{l_n^{1+1/\varepsilon}}{1 + 1/\varepsilon}, \tag{2}$$

where $\varepsilon > 0$ is a parameter governing the (un)compensated after-tax wage elasticity of labor supply. We adopt this simple utility function for expositional reasons. Our main results can be generalized to a general, non-separable utility function yielding income effects and non-constant wage elasticities of labor supply (see Bovenberg and Jacobs, 2003).

2.2. Individual optimization

The income tax schedule features an impersonal marginal tax rate on gross labor income, t , and a non-individualized lump-sum transfer or poll subsidy g . Education is subsidized at a flat rate s . With these linear policy instruments, the household budget constraint can be written as

$$c_n = (1 - t)n\phi(e_n)l_n - (1 - s)p_e e_n + g, \tag{3}$$

where p_e denotes the unit cost of e_n . Taking the policy instruments of the government as given,⁸ individuals maximize utility by choosing c_n , l_n , and e_n , subject to the household

⁶ A homothetic production function for human capital is assumed for aggregation reasons in the cases for linear income taxes and subsidies. Homotheticity is in principle not needed in the non-linear cases.

⁷ In this interpretation, the wage rate per efficiency unit of labor is normalized at unity. Alternatively, defining $\phi(e_n)$ (instead of $n\phi(e_n)$) as the efficiency unit of labor (or human capital) per hour worked, we can interpret ability n as the wage rate per efficiency unit of labor.

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

budget constraint. This yields the following first-order condition for the optimal choice of educational investment

$$(1-t)n\phi'(e_n)l_n = (1-s)p_e. \quad (4)$$

Marginal benefits of investing a unit of resources in education should equal marginal costs. The marginal tax rate t distorts the educational decision. In particular, this tax rate depresses the marginal benefits of education (i.e., the left-hand side of Eq. (4)). The education subsidy s , in contrast, reduces the marginal costs of learning (i.e., the right-hand side of Eq. (4)). Larger labor supply l_n raises the utilization rate of human capital and thus boosts the marginal benefits from learning on the left-hand side of Eq. (4).

The first-order condition for labor supply yields labor supply as a function of the after-tax wage rate $(1-t)n\phi(e_n)$

$$l_n = ((1-t)n\phi(e_n))^\varepsilon. \quad (5)$$

Compensated and uncompensated wage elasticities of labor supply coincide due to the absence of income effects. More education raises labor supply by increasing before-tax wages.

The positive feedback between human capital and labor supply implies that decreasing returns in the production of human capital ($\beta < 1$) are not sufficient for the second-order conditions for utility maximization to be met. In particular, by raising after-tax wages, more learning boosts labor supply, which in turn makes learning more attractive. This positive feedback effect, which depends on the wage elasticity of labor supply ε , should be offset by sufficiently strong decreasing returns in learning (lower β) in order to prevent corner solutions (see Appendix A). Therefore, second-order conditions require

$$\mu \equiv 1 - \beta(1 + \varepsilon) > 0, \quad (6)$$

where μ is a measure of the strength of feedback effects.

The tax elasticities of labor supply $\varepsilon_{lt} \equiv (\partial l_n / \partial (1-t))((1-t)/l_n) = \varepsilon/\mu$, education $\varepsilon_{et} \equiv (\partial e_n / \partial (1-t))((1-t)/e_n) = (1+\varepsilon)/\mu$, and gross earnings $\varepsilon_{zt} \equiv (\partial z_n / \partial (1-t))((1-t)/z_n) = (\varepsilon + \beta(1+\varepsilon))/\mu$ are important determinants of the optimal tax rates and are derived in Appendix B. The tax elasticity of labor supply ε/μ exceeds the wage elasticity of labor supply ε . The reason is that the tax rate t reduces the after-tax wage $(1-t)n\phi(e_n)$ 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 $n\phi(e_n)$ through its negative impact on learning e_n). Learning is harmed directly because the tax rate reduces the returns from investment in human capital, and indirectly because lower labor supply depresses the utilization rate of human capital. Indeed, the tax elasticity of educational investment is given by $(1+\varepsilon)/\mu$, where the two terms in the denominator represent these direct and indirect effects on learning. Given the tax elasticities of labor supply and learning, the tax elasticity of labor earnings $z_n = n\phi(e_n)l_n$ amounts to $\varepsilon_{zt} = \varepsilon_{lt} + \beta\varepsilon_{et} = (\varepsilon + \beta(1+\varepsilon))/\mu$.

2.3. Government

The government collects taxes on aggregate labor incomes to finance exogenously given expenditures, E , the impersonal education subsidy on aggregate learning, s , and the uniform lump-sum transfer g . The government budget constraint therefore reads as

$$t \int_{\underline{n}}^{\bar{n}} n\phi(e_n)l_n dF(n) = \int_{\underline{n}}^{\bar{n}} (sp_e e_n + g + E) dF(n). \tag{7}$$

The government must be able to observe aggregate labor incomes $\int_{\underline{n}}^{\bar{n}} z_n dF(n)$ and aggregate learning $\int_{\underline{n}}^{\bar{n}} e_n dF(n)$ to employ t and s .

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

$$\Gamma = \int_{\underline{n}}^{\bar{n}} \Psi(v(g, t, s, n)) dF(n), \tag{8}$$

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

The Lagrangian for maximizing social welfare is therefore given by

$$\max_{\{g, t, s\}} \mathcal{L} = \int_{\underline{n}}^{\bar{n}} \Psi(v(g, t, s, n)) dF(n) + \lambda \int_{\underline{n}}^{\bar{n}} (tn\phi(e_n)l_n - sp_e e_n - g - E) dF(n), \tag{9}$$

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

2.4. Optimal lump-sum transfer

By defining the social marginal value of income of an individual with ability n

$$b_n \equiv \frac{\Psi'(v(g, t, s, n))}{\lambda}, \tag{10}$$

we can write the first-order condition for maximizing the Lagrangian (Eq. (9)) with respect to g as

$$\bar{b} \equiv \int_{\underline{n}}^{\bar{n}} b_n dF(n) = 1. \tag{11}$$

The average social marginal benefits of a higher g (i.e., the left-hand side of Eq. (11)) should equal the costs in terms of a higher g (i.e., the right-hand side of Eq. (11)).

With the aid of the first-order condition for g (Eq. (11)), we define the so-called distributional characteristic ξ of labor income as the negative normalized covariance

between the welfare weight the government attaches to income of a particular ability b_n and gross labor income z_n (see, e.g., Atkinson and Stiglitz, 1980)⁹:

$$\xi \equiv - \frac{\int_{\underline{n}}^{\bar{n}} b_n z_n dF(n) - \int_{\underline{n}}^{\bar{n}} z_n dF(n) \int_{\underline{n}}^{\bar{n}} b_n dF(n)}{\int_{\underline{n}}^{\bar{n}} z_n dF(n) \int_{\underline{n}}^{\bar{n}} b_n dF(n)} = \frac{\int_{\underline{n}}^{\bar{n}} (1 - b_n) z_n dF(n)}{\int_{\underline{n}}^{\bar{n}} z_n dF(n)}, \quad (12)$$

where the second equality follows from Eq. (11). A positive distributional characteristic ξ implies that the income tax base is larger for high-ability agents (who feature relatively low welfare weights b_n) than for low-ability agents (who feature relatively high welfare weights), so that taxing labor income yields distributional benefits. The magnitude of the distributional characteristic depends not only on the correlation between ability and the tax base, but also on the correlation between ability and the welfare weights.¹⁰ Indeed, a zero distributional characteristic implies either that the government is utilitarian and not interested in redistribution (so that the welfare weight is the same for all n) or that the marginal contribution to the tax base is equal for all ability types (i.e., taxable income is the same for all n , so that there is no inequality).

2.5. Optimal income taxation

The optimal linear income tax in the absence of educational subsidies (i.e., $s = 0$) can be written as (see Appendix C)

$$\frac{t}{1-t} = \frac{\xi}{\varepsilon_{zt}} = \frac{\xi}{\frac{\varepsilon}{\mu} + \beta \frac{1+\varepsilon}{\mu}}. \quad (13)$$

The optimal tax formula (13) illustrates the fundamental trade-off between equity and efficiency. In particular, the tax rate rises if redistributive concerns become more pressing, as indicated by a larger distributional characteristic ξ . The denominator of Eq. (13) captures the distortionary costs of redistributive taxation in terms of the effective elasticity of the total tax base with respect to the tax rate. The optimal marginal tax rate t should be lower if a larger tax elasticity of earnings $\varepsilon_{zt} = \varepsilon/\mu + \beta(1+\varepsilon)/\mu$ indicates that redistributive taxes distort labor supply and human capital accumulation more substantially. This conforms to standard Ramsey intuition.

In the absence of a learning decision ($\beta = 0$ so that $\mu = 1$), the effective elasticity of the tax base represented by the denominator at the right-hand side of Eq. (13) boils down to the wage elasticity of labor supply ε . Accordingly, the optimal tax is given by

$$\frac{t}{1-t} = \frac{\xi}{\varepsilon}. \quad (14)$$

⁹ The distributional characteristic for labor income z_n coincides with that of education e_n because labor income z_n is proportional to investments in education e_n due to the homothetic production function.

¹⁰ The negative correlation increases if more unequal private utility levels u_n produce larger differences in marginal social utilities $U'(u_n)$ (and thus in b_n).

This is the familiar optimal tax formula from the standard optimal linear income tax model, which abstracts from endogenous learning (see Atkinson and Stiglitz, 1980).

With endogenous learning, the wage elasticity of labor supply $l_n(\varepsilon)$ in the usual inverse elasticity formula (14) is replaced by the larger tax elasticity of total earnings $z_n(\varepsilon/\mu + \beta(1+\varepsilon)/\mu)$ in Eq. (13). This result resembles Diamond and Mirrlees (2002), who analyze optimal taxation if a single tax rate distorts multiple decision margins. They show that the optimal tax rate should typically be lower if the tax rate distorts more decision margins. In our model, the income tax rate t distorts not only labor supply but also learning. Hence, optimal taxes are lower compared to the case in which learning is exogenous and the tax rate distorts only the labor supply margin.

2.6. Optimal income taxation and education subsidies

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

$$s = t. \quad (15)$$

The optimal subsidy s ensures that education expenses effectively become deductible against the labor income tax rate. In practice, many educational costs, such as books, computers, and tuition (corresponding to, among other things, the wages of the teachers), cannot be deducted from the income tax because individuals typically earn no (or very low) labor incomes when they put these resources into education. Hence, these costs should optimally be subsidized. Other costs of education, however, are already effectively tax deductible. The main example of tax deductible costs are earnings foregone while enrolled in education; these costs are in fact deductible against the income tax rate t . Expression (15) shows that this treatment is in fact optimal.

Expression (15) implies that the total net tax wedge Δ on learning, which measures the extent to which the tax and subsidy instruments reduce the marginal returns to learning, is zero, i.e.; $\Delta \equiv t\phi(e_n)l_n - sp_e = (t-s)p_e/(1-t) = 0$ (where the second equality follows from Eq. (4)). Accordingly, optimal education subsidies ensure efficiency in human capital investments. This result can be viewed as an extension of the production efficiency theorem of Diamond and Mirrlees (1971) to a model with individual production functions for human capital.¹¹ In our set-up, investments in human capital e_n are intermediate goods entering individual production functions of human capital. Expression (15) states that, in accordance with Diamond and Mirrlees (1971), intermediate inputs into human capital formation should not be taxed on a net basis.

Even with decreasing returns to human capital accumulation, the Diamond–Mirrlees result applies because the government in fact has access to an implicit profit tax on the ability rents in the returns from human capital accumulation. In particular, by combining the income tax and an educational subsidy (in a manner given by Eq. (15)), the government can tax away the infra-marginal rents from human capital formation without

¹¹ Diamond and Mirrlees (1971) focussed on an aggregate production function, but recognized that their theorem also applies in the presence of several production sectors featuring their own production function.

distorting the incentives to learn. Indeed, the optimal education subsidy (Eq. (15)) exactly compensates individuals for taxes imposed on the returns from learning. Hence, the labor income tax falls only on the ability rents generated through human capital formation and on additional income from labor supply. Indeed, if labor supply is inelastic ($\varepsilon = 0$), the optimal income tax cum education subsidy would effectively become Tinbergen's (1970) ability tax (i.e., a tax on rents from ability only).

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

$$\frac{t}{1-t} = \frac{\xi}{\frac{\varepsilon}{\mu}}. \quad (16)$$

Comparing the optimal income tax with optimal education policy (Eq. (16)) to the optimal income tax without any education subsidies (Eq. (13)), we observe that the additional instrument of the education subsidy s allows for a higher income tax (*ceteris paribus* the distributional characteristic ξ). The reason is that the education subsidy eliminates the distortions on human capital accumulation. With a lower effective elasticity of the tax base, implying smaller efficiency costs of redistributive income taxation, the government redistributes income more aggressively.

Despite the availability of education subsidies, the optimal tax rate (Eq. (16)) is lower than with exogenous learning; see Eq. (14) (*ceteris paribus* the distributional characteristic ξ). The reason is that the tax elasticity of labor supply ε/μ replaces the wage elasticity of labor supply ε in the standard optimal tax formula (14). Intuitively, the complementarity between labor supply and (endogenous) learning makes labor supply more sensitive to the tax rate than is the case with exogenous learning. This is because the tax rate reduces the after-tax wage not only directly, but also indirectly by depressing learning and thereby the before-tax wage rate. Although the tax wedge on learning is zero if the government can optimally set education subsidies (i.e., $\Delta = 0$), learning is reduced compared to the first-best because the labor tax reduces learning by decreasing the utilization rate of human capital through lower labor supply.

Education subsidies rise if distributional concerns become more pressing, as indicated by a larger distributional characteristic ξ (substitute Eq. (16) into Eq. (15) to eliminate t). This result may seem counterintuitive because the more able individuals learn more, and thus benefit the most from the education subsidies. However, this benefit for the most-skilled agents is more than offset by higher labor taxes. Indeed, the combination of labor taxes and education subsidies implies a positive tax on rents from ability. Clearly then, education subsidies and progressive taxation are Siamese twins. Both optimal education subsidies and marginal taxes are zero in the absence of redistributive considerations ($\xi = 0$).

If the government can optimize only over the education subsidy at an exogenous tax rate t , the optimal education subsidy is given by (see Appendix C)

$$\frac{t-s}{(1-s)(1-t)} = \mu \left[\xi - \frac{t}{(1-t)} \frac{\varepsilon}{\mu} \right]. \quad (17)$$

The term in square brackets at the right-hand side of Eq. (17) measures the sub-optimality of the income tax (see expression (16)). If this term is positive, income taxes are sub-optimally low and thus do not redistribute sufficient resources from higher to lower abilities. Accordingly, the government levies a positive tax on education ($\Delta = (t-s)p_e/(1-t) > 0$) to redistribute resources from the rich, who learn more, to the poor. Indeed, if the government cannot tax labor income at all ($t = 0$), the government taxes rather than subsidizes education (i.e., $s < 0$). These education taxes are relatively large if the effective elasticity of the tax base is small: i.e., if the feedback effects between labor supply and learning are only weak so that μ is large (*ceteris paribus* the distributional characteristic ξ).

3. Non-verifiable learning

In order to establish overall efficiency in the production of human capital, the government must have access to sufficient policy instruments. In particular, it must be able to verify all investments in human capital so that it can subsidize these educational inputs in order to eliminate the tax wedge on learning. This section explores optimal tax and education policies if the government cannot observe all educational inputs.

To that end, we distinguish between verifiable investments in human capital, which the government can subsidize at rate s , and non-verifiable investments, which cannot be subsidized (see also van Ewijk and Tang, 2000). Examples of verifiable investments are tuition fees and the number of years enrolled in education. In practice, governments do indeed subsidize tuition costs and school enrollment. Books, computers and travelling costs, however, are difficult to verify because individuals may misrepresent expenditures for private consumption purposes as investments in education. Non-verifiable investments include also effort costs (preparing for exams, attending college, etc.). Non-verifiable learning effort is analogous to non-verifiable work effort, which the optimal tax literature typically assumes.

Verifiable educational investments x_n and non-verifiable investments y_n are combined into aggregate investments in human capital e_n through a homothetic concave constant-returns-to-scale sub-production function ψ :

$$e_n \equiv \psi(x_n, y_n), \tag{18}$$

where $\psi_x(\cdot), \psi_y(\cdot) > 0, \psi_{xx}(\cdot), \psi_{yy}(\cdot) < 0, \psi_{xy}(\cdot) > 0$.

The household budget constraint is now given by

$$c_n = (1-t)n\phi(\psi(x_n, y_n))l_n - (1-s)p_x x_n - p_y y_n + g, \tag{19}$$

where p_x and p_y denote the unit cost of x_n and y_n , respectively. The first-order conditions for x_n and y_n are

$$(1-t)n\phi'(e_n)l_n\psi_x(x_n, y_n) = (1-s)p_x, \tag{20}$$

$$(1-t)n\phi'(e_n)l_n\psi_y(x_n, y_n) = p_y. \tag{21}$$

Whereas the marginal tax rate t reduces the marginal benefits of both types of investment, subsidies on education s decrease the marginal costs of only verifiable inputs. Labor supply (Eq. (5)) is not affected.

The government budget constraint is now given by

$$t \int_{\underline{n}}^{\bar{n}} n\phi(\psi(x_n, y_n))I_n dF(n) = \int_{\underline{n}}^{\bar{n}} (sp_x x_n + g + E) dF(n). \tag{22}$$

If the government optimizes over all its instruments (i.e., t , s , and g), the optimal education subsidy can be expressed in terms of the optimal tax rate as (see Appendix C)

$$s = \frac{t}{1 - (1 - \sigma)(1 - \alpha)(1 - t)}, \tag{23}$$

where $\alpha \equiv x_n \psi_x(\cdot) / \psi(\cdot)$ denotes the share of verifiable inputs in human capital formation, and $\sigma \equiv -(\text{dlog}(x_n / y_n)) / (\text{dlog}(\psi_x(\cdot) / \psi_y(\cdot)))$ stands for the elasticity of substitution between the two inputs in the production function for human capital.¹²

In the case of a Leontief production of human capital (i.e., $\sigma = 0$), the government offsets the entire tax wedge on learning (if $t > 0$) through the provision of an implicit subsidy on non-verifiable investments by subsidizing complementary verifiable investments. Consequently, the government subsidizes verifiable inputs beyond tax deductibility of these inputs (i.e., $s > t$, if $t > 0$ and $\alpha < 1$). With non-zero substitution between the two inputs into learning (i.e., $\sigma > 0$), however, the government cannot costlessly mimic a subsidy on y_n by subsidizing x_n , because subsidizing x_n distorts the composition of human capital accumulation towards the excessive use of verifiable inputs. Nevertheless, as long as substitution is imperfect (i.e., $\sigma < \infty$), the optimal subsidy on verifiable inputs is positive if taxation distorts learning ($t > 0$). Intuitively, starting from an initial situation without subsidies on education (and $t > 0$), introducing a small subsidy produces a first-order welfare gain by offsetting some of the tax distortions on human capital formation, while the distortions on the composition of investments in human capital are only second order (if $\sigma < \infty$). In the optimum, the government optimally balances the distortions on the composition and the level of human capital investment so that both distortions are first order. With a Cobb–Douglas sub-production function of aggregate human capital investment (i.e., $\sigma = 1$), the optimal education subsidy makes verifiable investments effectively tax deductible (i.e., $s = t$). In this case, the positive welfare impact of additional education subsidies on the level of learning exactly balances the distortionary effect of additional education subsidies on the composition of learning. Hence, the presence of non-verifiable inputs does not affect the case for making verifiable inputs effectively tax deductible. Only with infinite substitution between the two inputs ($\sigma \rightarrow \infty$), optimal subsidies are zero ($s = 0$). In that case, the government finds it optimal not to subsidize x_n at all in order to ensure a level playing field with y_n .

The optimal linear income tax amounts to (see Appendix C)

$$\frac{t}{1 - t} = \frac{\xi}{\frac{\varepsilon}{\mu} + \rho\beta \frac{1 + \varepsilon}{\mu}}, \tag{24}$$

¹² Appendix B shows that homotheticity of $\psi(x_n, y_n)$ implies that α and σ are independent of ability.

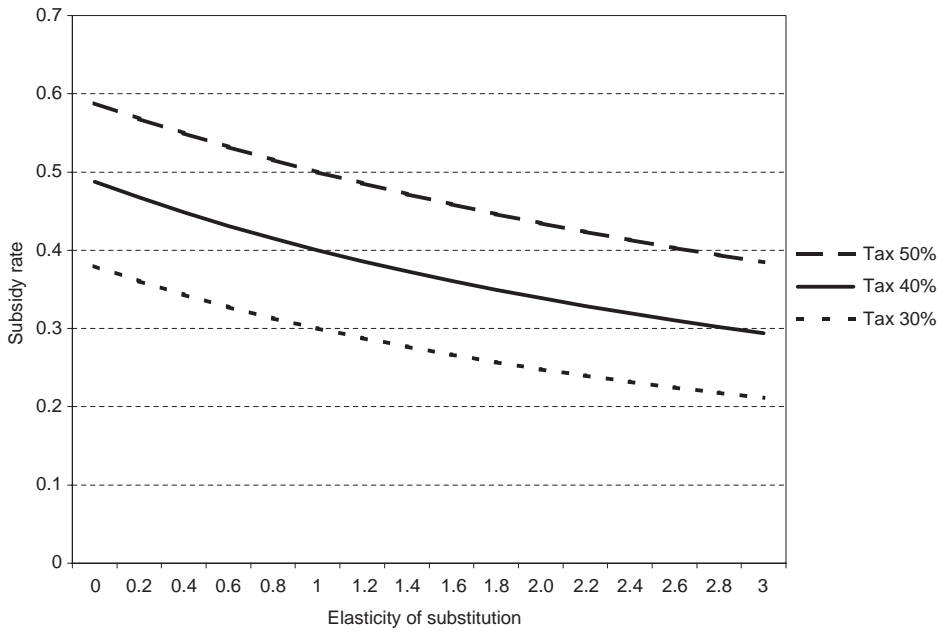


Fig. 2. Optimal education subsidies ($\alpha = 0.8$).

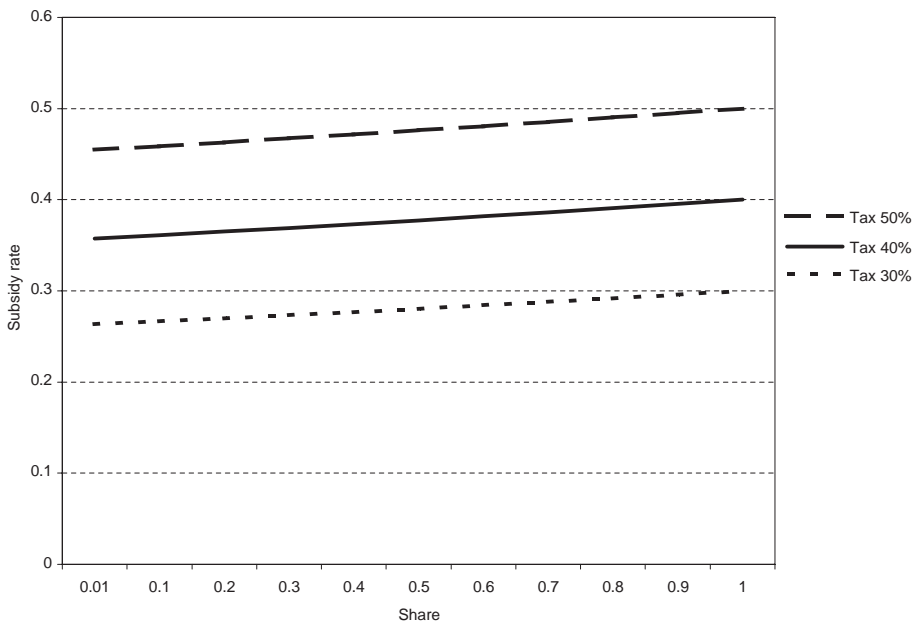


Fig. 3. Optimal education subsidies ($\sigma = 1.2$).

where $0 \leq \rho \equiv (1 - \alpha)\sigma / (\alpha + (1 - \alpha)\sigma) \leq 1$. The parameter ρ indicates which share of the learning distortion of redistributive taxation remains after optimally employing education subsidies. It thus measures the extent to which the production efficiency is violated. Depending on the share of non-verifiable investments α and the substitution elasticity between the two types of investment σ , the optimal tax formula is closer to the corresponding formula without any educational subsidies (Eq. (13)) or to that with only verifiable investments (Eq. (16)). The tax wedge on aggregate learning ρ optimally remains substantial if the share of non-verifiable learning in aggregate learning $(1 - \alpha)$ and the substitution elasticity between verifiable and non-verifiable learning σ are large. In that case, education subsidies are ineffective in offsetting the learning distortion imposed by labor taxes. With infinite substitution ($\sigma \rightarrow \infty$), the optimal tax formula approaches the corresponding formula without any educational subsidies (Eq. (13)). Without substitution ($\sigma = 0$), in contrast, the optimal tax formula equals the optimal tax formula with only verifiable investments (Eq. (16)).

In order to check how sensitive optimal education subsidies on verifiable inputs are with respect to substitution effects between verifiable and non-verifiable inputs, Figs. 2 and 3 plot the optimal subsidy rates s at various elasticities of substitution (σ between 0 and 3) and various shares of verifiable inputs (α between 0 and 1), respectively. These figures reveal that the optimal subsidy rate on verifiable inputs remains close to the tax rate for a wide range of production functions, even though production efficiency may be violated substantially (as a result of large values for σ and $(1 - \alpha)$).

4. Optimal non-linear tax and education policies

This section explores the robustness of our results by analyzing non-linear policy instruments in our benchmark model with a single educational input. In particular, the government can levy a non-linear income tax $T(z_n)$ on gross incomes $z_n = n\phi(e_n)l_n$. The marginal income tax rate $T'(z_n) \equiv dT(z_n)/dz_n$ can thus vary across individuals earning different incomes. Furthermore, the government may employ a non-linear subsidy on resources e_n invested in education. The subsidy is denoted as $S(p_e e_n)$, where $S'(p_e e_n) \equiv dS(p_e e_n)/de_n$ represents the marginal subsidy rate on e_n . The non-linear tax and educational subsidies imply substantial informational requirements because the government must be able to verify gross income and educational investment of each individual. With linear policy instruments, in contrast, the government needs to observe only economy-wide aggregates.

In the presence of non-linear policy instruments, the individual's budget constraint reads as

$$c_n = z_n - T(z_n) - p_e e_n + S(p_e e_n). \quad (25)$$

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

$$l_n = ((1 - T'(z_n))n\phi(e_n))^{\varepsilon}, \quad (26)$$

in addition to the first-order condition for resources invested in education¹³

$$(1 - T'(z_n))n\phi'(e_n)l_n = (1 - S'(p_e e_n))p_e. \tag{27}$$

The government observes z_n (and e_n if it subsidizes education) and thereby controls z_n (and educational investments e_n if it subsidizes these investments) through the tax (and subsidy) schedule(s). The government should respect the incentive-compatibility constraints, which require that each individual n prefers the bundle c_n, z_n, e_n intended for him over the bundles intended for others

$$U(c_n, z_n, e_n, n) \geq U(c_m, z_m, e_m, n), \forall m \in [\underline{n}, \bar{n}], \forall n \in [\underline{n}, \bar{n}], \tag{28}$$

where $U(c_n, z_n, e_n, n) \equiv u(c_n, z_n / (n\phi(e_n))) = u(c_n, l_n)$. These global incentive-compatibility constraints can be replaced by the (first-order) incentive-compatibility constraint using the envelope theorem (see, e.g., Mirrlees, 1971)¹⁴:

$$\frac{du_n}{dn} = \frac{l_n^{1+1/\varepsilon}}{n}. \tag{29}$$

The economy's resource constraint is given by¹⁵

$$\int_{\underline{n}}^{\bar{n}} (n\phi(e_n)l_n - p_e e_n - c_n - E) dF(n) = 0. \tag{30}$$

We solve for the optimal allocation by applying the maximum principle and setting up a Hamiltonian \mathcal{H} , with l_n (and, depending on the availability of non-linear education subsidies, e_n) as control variable(s), u_n as state variable, and θ_n as costate variable for the incentive-compatibility constraint (Eq. (29)):

$$\max_{\{l_n, e_n, u_n\}} \mathcal{H} = \Psi(u_n)f(n) - \theta_n \frac{l_n^{1+1/\varepsilon}}{n} + \lambda(n\phi(e_n)l_n - p_e e_n - c_n - E)f(n), \tag{31}$$

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

$$\lim_{n \rightarrow \bar{n}} \theta_n = 0, \lim_{n \rightarrow \underline{n}} \theta_n = 0. \tag{32}$$

Having found the second-best allocation, we determine the optimal marginal taxes $T'(z_n)$ and subsidy rates $S'(p_e e_n)$ by employing the first-order conditions for individual optimization with respect to l_n and e_n (see Eqs. (26) and (27), respectively). Optimal

¹³ The restriction $\mu \equiv 1 - \beta(1 + \varepsilon) > 0$ holds so that the second-order conditions for an interior solution to the individual optimization problem are satisfied.

¹⁴ In adopting the first-order approach, we assume that the second-order conditions for the optimal policy problem are met. This requires single crossing of the utility function with respect to all control variables in addition to all control variables being non-decreasing in ability at the optimum schedules. For an analysis of second-order conditions in optimal non-linear taxation problems, see Mirrlees (1986) and Ebert (1992).

¹⁵ Walras' law implies that the government budget constraint is met if Eq.(30) holds.

consumption follows from the optimal levels of l_n , e_n , and the household budget constraint (Eq. (25)).

4.1. Optimal income taxation

In analogy of our discussion of linear policy instruments, we first explore the optimal income tax in the absence of education policy. By taking the first-order condition of the Hamiltonian (Eq. (31)) with respect to l_n , we derive the optimal marginal tax rates along the non-linear tax schedule (see Appendix D)

$$\frac{T'(z_n)}{1 - T'(z_n)} = \frac{\theta_n/\lambda}{nf'(n)} \left(\frac{1 + \varepsilon}{\varepsilon + \beta(1 + \varepsilon)} \right). \quad (33)$$

Following Saez (2001), we can write the optimal tax rate in terms of the earnings density $q(z_n)$ rather than the ability density $f(n)$ because we want to explore the impact of endogenous learning at a given distribution of earnings (see Appendix D):

$$\frac{T'(z_n)}{1 - T'(z_n) + \frac{\varepsilon + \beta(1 + \varepsilon)}{\mu} z_n T''(z_n)} = \frac{\theta_n/\lambda}{z_n q(z_n)} \left(\frac{\varepsilon}{\mu} + \beta \frac{1 + \varepsilon}{\mu} \right)^{-1}. \quad (34)$$

The term in brackets at the right-hand side of Eq. (34) is the tax elasticity of labor earnings z_n , which is equivalent to the corresponding tax elasticity of labor earnings in the expression for the optimal linear income tax (Eq. (13)). The non-linear counterpart of ξ is the term $\theta_n/\lambda = \int_n^{\bar{n}} (1 - b_m) f(m) dm > 0$, which corresponds to the benefits of redistribution (see Appendix C). In particular, it measures the welfare gain from redistributing income from individuals featuring abilities above n (or earnings exceeding z_n) to individuals with abilities below n (or earnings below z_n).¹⁶ Expression (34) shows that a larger number of individuals with earnings z_n (larger density $q(z_n)$) and higher earnings z_n exacerbate the distortionary costs of large marginal rates at earnings level z_n and therefore reduce the optimal marginal tax rate at that earnings level (*ceteris paribus* the distributional term θ_n/λ).¹⁷

¹⁶ The transversality conditions $\theta_{\bar{n}} = \theta_n = 0$ imply the standard results of zero optimal marginal tax rates at the top and the bottom of the earnings distribution.

¹⁷ Diamond (1998) shows that asymptotic marginal tax rates (i.e., $T'(z_n)$ for $n \rightarrow \infty$) converge to a positive number if the ability distribution converges to a Pareto distribution. With $T''(z_n)$ thus converging to zero, the asymptotic marginal rate $T'(z_n)$ is then given by

$$\frac{T'(z_n)}{1 - T'(z_n)} = \frac{(1 - b_\infty)}{\pi \left(\frac{\varepsilon}{\mu} + \beta \frac{1 + \varepsilon}{\mu} \right)},$$

where $\pi = z_n q(z_n)/(1 - Q(z_n))$ is the parameter of the Pareto distribution of earnings and $b_\infty \equiv \lim_{n \rightarrow \infty} \Psi'(u_n)/\lambda$.

4.2. Optimal income taxation and education subsidies

If the government simultaneously sets non-linear tax and education subsidy schedules, it determines the optimal allocation by using e_n and l_n as control variables. By manipulating the first-order condition of the Hamiltonian (Eq. (31)) with respect to these two controls, we derive optimal marginal tax rates and subsidy rates along the non-linear tax and subsidy schedules. From the first-order conditions for e_n , we again establish production efficiency in learning (see Appendix D):

$$n\phi'(e_n)l_n = p_e. \quad (35)$$

Combining Eq. (35) with Eq. (27), we find

$$S'(p_e e_n) = T'(z_n). \quad (36)$$

Optimal education policy requires that the costs of education should effectively be tax deductible, just as in the linear case (compare Eq. (15)).

Optimal marginal tax rates along the non-linear tax schedule are (see Appendix D)¹⁸

$$\frac{T'(z_n)}{1 - T'(z_n) + \frac{\varepsilon}{\mu} z_n T''(z_n)} = \frac{\theta_n/\lambda}{z_n q(z_n)} \left(\frac{\varepsilon}{\mu}\right)^{-1}. \quad (37)$$

We arrive at the same conclusion as with linear taxes: education subsidies allow for higher marginal tax rates by eliminating the distortionary impact of the labor income tax on human capital accumulation. In particular, comparing the optimal income tax rates with optimal education policy (Eq. (37)) to the optimal income tax without any education subsidies (Eq. (34)), we observe that the additional instrument of the education subsidy allows for higher marginal taxes on labor income (*ceteris paribus* θ_n/λ). Moreover, stronger redistributive motives, as reflected in larger values for $\theta_n/\lambda = \int_n^{\bar{n}} (1 - b_m)f(m)dm$, increase education subsidies, even though the rich benefit the most from education subsidies (combine Eqs. (36) and (37)). Indeed, whereas the tax system is targeted at redistribution, education subsidies offset the adverse impact of taxes on the incentives to accumulate human capital. Also in the case of non-linear policy instruments, therefore, education subsidies and progressive taxation are Siamese twins.

5. Non-monetary costs and benefits of education

As a final check on the robustness of optimal education subsidies, this section introduces non-pecuniary costs and benefits of learning in the model with non-linear taxes

¹⁸ Bovenberg and Jacobs (2003) show that with non-verifiable investments in human capital along the lines of Section 3, the term ε/μ is replaced by $\varepsilon/\mu + \beta\rho(1+\varepsilon)/\mu$, where $\rho \equiv (1-\alpha)\sigma/(\alpha+(1-\alpha)\sigma)$ and σ denotes the elasticity of substitution between verifiable and non-verifiable inputs in the production function for human capital. Hence, the optimal linear taxes (Eq. (24)) have a straightforward parallel to non-linear instruments. Similarly, the optimal marginal subsidy is given by expression (23), in which s and t are replaced by $S'(p_x x_n)$ and $T'(z_n)$, respectively.

and subsidies. To that end, e_n enters the utility function.¹⁹ We stick to the Diamond (1998) type of preferences without income effects, but adopt a sub-utility function $v(\cdot)$ over work effort l_n and educational investments e_n :

$$u(c_n, l_n, e_n) \equiv c_n - v(l_n, e_n), \quad (38)$$

where $v_l(\cdot) > 0$, $v_{ll}(\cdot) < 0$, $v_{ee}(\cdot) > 0$, and $v_{le}(\cdot) \geq 0$. $v_e(\cdot) > 0$ if education requires non-material effort in addition to material resources, while $v_e(\cdot) < 0$ if education yields non-material benefits in addition to material benefits. The first-order conditions for labor supply l_n and investments in human capital e_n are now given by

$$v_l(l_n, e_n) = (1 - T'(z_n))n\phi(e_n), \quad (39)$$

$$(1 - T'(z_n))n\phi'(e_n)l_n = (1 - S'(p_e e_n))p_e + v_e(l_n, e_n). \quad (40)$$

Appendix D derives the following expression characterizing optimal non-linear education subsidies in the presence of non-pecuniary costs of education ($v_e(\cdot) > 0$)

$$\frac{S'(p_e e_n) - T'(z_n)}{(1 - T'(z_n))(1 - S'(p_e e_n))} = \omega_n \left(\frac{T'(z_n)}{1 - T'(z_n)} - \frac{\theta_n/\lambda}{nf(n)} \frac{l_n v_{le}(\cdot)}{v_e(\cdot)} \right), \quad (41)$$

where $\omega_n \equiv v_e(l_n, e_n)/((1 - S'(p_e e_n))p_e) > 0$ denotes the ratio between non-material and material costs and therefore measures the importance of non-material costs.

If utility is separable (i.e., $v_{le}(\cdot) = 0$), we find that

$$S'(p_e e_n) = \left(\frac{1 + \omega_n}{1 + \omega_n T'(z_n)} \right) T'(z_n) > T'(z_n). \quad (42)$$

Hence, *ceteris paribus* the marginal tax rate $T'(z_n)$, optimal subsidies on material inputs are larger in the presence of non-material effort costs (i.e., $S'(p_e e_n) > T'(z_n)$ if $\omega_n > 0$, while $S'(p_e e_n) = T'(z_n)$ if $\omega_n = 0$). The intuition is that the government employs additional subsidies on the verifiable material inputs to offset the net tax on the non-verifiable non-pecuniary inputs.²⁰

With non-separable utility, we can distinguish two cases: *learning-or-doing* and *learning-by-doing*. In particular, if additional work effort raises the effort costs of learning (i.e., $v_{le}(\cdot) > 0$), we have *learning-or-doing*, since work effort and learning efforts are substitutes. In that case, optimal subsidies on verifiable materials inputs are lower than with separable utility (*ceteris paribus* $T'(z_n)$ and ω_n). Intuitively, education subsidies, which boost learning, worsen the tax distortions on work effort by further raising the

¹⁹ We assume that effort and material inputs invested in education cannot be substituted. With substitution between non-verifiable non-pecuniary effort and verifiable material inputs, optimal policies are straightforward combinations of the results in this section and Section 3. The formal results are available at www.fee.uva.nl/scholar/mdw/jacobs. Section 3 reveals that optimal subsidies on verifiable inputs are not very sensitive to the elasticity of substitution between verifiable and non-verifiable inputs.

²⁰ $S'(p_e e_n) > T'(z_n)$ continues to hold with positive substitution between effort and material inputs into learning if the substitution elasticity between the inputs is smaller than 1. In that case, the impact of additional subsidies for material inputs (starting from $S'(p_e e_n) = T'(z_n)$) on the aggregate learning distortion dominates that on the composition of learning. Indeed, the intuition is the same as in Section 3.

marginal cost of work effort. If additional work effort reduces the effort costs of learning (i.e., $v_{le}(\cdot) < 0$), in contrast, we have a learning-by-doing model and education should be subsidized even more than with separable utility (*ceteris paribus* $T'(z_n)$). The reason is that additional education subsidies alleviate tax distortions on not only learning but also work effort.

In the presence of consumption benefits of education ($v_e(\cdot) < 0$), optimal education subsidies are characterized by (see Appendix D)

$$\frac{T'(z_n) - S'(p_e e_n)}{(1 - T'(z_n))(1 - S'(p_e e_n))} = -\omega_n \left(\frac{T'(z_n)}{1 - T'(z_n)} + \frac{\theta_n/\lambda}{nf(n)} \frac{I_n v_{le}(\cdot)}{-v_e(\cdot)} \right), \tag{43}$$

where $-\omega_n \equiv \frac{-v_e(I_n, e_n)}{(1 - T'(z_n))n\phi'(e_n)I_n - v_e(I_n, e_n)} = \frac{-v_e(I_n, e_n)}{(1 - S'(p_e e_n))p_e} > 0$ measures the share of non-material benefits in the overall benefits from learning.

If utility is separable between work effort and consumption benefits of education (i.e., $v_{le}(\cdot) = 0$), the optimal subsidy satisfies (note that $-1 < \omega_n < 0$)

$$S'(p_e e_n) = \left(\frac{1 + \omega_n}{1 + \omega_n T'(z_n)} \right) T'(z_n) < T'(z_n). \tag{44}$$

Optimal education subsidies fall if more benefits of education become non-material and thus escape the income tax (*ceteris paribus* $T'(z_n)$). Hence, education should be subsidized to the extent that benefits are taxed.²¹ In the general case with both consumption and investment benefits, optimal education subsidies are between zero (if all benefits escape tax and $\omega_n = -1$) and $T'(z_n)$ (if none of the benefits escape tax and $\omega_n = 0$).

If utility is not separable, just as in the case with non-pecuniary effort costs, education subsidies may exacerbate or alleviate the tax distortions on work effort, depending on the sign of the cross derivative $v_{le}(\cdot)$. In particular, non-separabilities raise (reduce) education subsidies if additional education decreases (increases) the utility cost of work effort (*ceteris paribus* $T'(z_n)$ and note that $v_e(\cdot) < 0$ in this case). According to Becker (1965), individuals featuring higher levels of human capital are also more productive in producing leisure time.²² Casual observation may indeed suggest that skilled people distill more satisfaction from leisure activities such as travelling, visiting museums, and watching opera. In that case, more education may raise the utility cost of work effort (i.e., $v_{le}(\cdot) > 0$), thereby weakening the case for education subsidies. However, highly skilled people typically enjoy more attractive work that yields various non-pecuniary benefits, including more flexibility in working hours, more challenging and exciting jobs, and more societal status and power. Indeed, consumption benefits of education are typically related more to work than to leisure so that $v_{le}(\cdot) < 0$. This strengthens our case for education subsidies because these subsidies then reduce tax distortions on both learning and work effort.

We conclude that our main result that education should be subsidized to alleviate the distortions due to a redistributive tax schedule is quite robust. Whereas non-material costs strengthen the case for education subsidies as long as these non-material effort costs are

²¹ Alstedsaeter (2003) shows that education should optimally be taxed rather than subsidized if all costs are effectively deductible from the income tax and education yields consumption benefits.

²² See also Heckman (1976) for an application of Becker (1965).

complements to verifiable education inputs, non-material benefits may weaken that case unless consumption benefits are more related to work than to leisure. In any case, non-material costs are likely to be more important than non-material benefits because returns on education generally exceed the returns on alternative investments in, for example, government bonds.²³ These higher returns compensate for non-material costs of learning. If non-material benefits were more important than non-material costs, the returns on education would be lower than the real return on government bonds. Hence, large non-material benefits do not fit the facts.

6. Policy implications

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

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

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

Our analysis suggests also that education should be subsidized more heavily if countries feature more progressive tax systems. Fig. 1 shows that this is indeed a stylized fact. As a corollary, countries reducing marginal taxes should cut education subsidies. Indeed, several countries have cut both marginal taxes and education subsidies in recent years.

7. Conclusions

This paper has studied the simultaneous setting of optimal redistributive income taxes and education subsidies in models of optimal linear and non-linear income taxation with

²³ See recent surveys on the returns to education by, for example, Card (1999), and Harmon et al. (2003).

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

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

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

endogenous learning determining individuals' earning potentials. We showed that education subsidies can be a powerful instrument to eliminate distortions in the accumulation of human capital caused by redistributive policies. In particular, education subsidies should effectively make costs of investments in human capital deductible against the rate at which the returns of education are taxed. Moreover, optimal marginal income tax rates can be larger if education subsidies alleviate the tax distortions associated with redistribution. The more the government desires to help the poor, the more it should employ education subsidies to offset the tax distortions associated with redistribution. Education subsidies and redistribution of incomes are thus like Siamese twins. Intuitively, the more redistributive tax system more than compensates for the regressive incidence of education subsidies.

If the government cannot verify all educational efforts, it is not able to completely eliminate the tax distortions on human capital formation. The reason is that the government can subsidize only verifiable investments. Optimal subsidies therefore balance distortions in the level of human capital accumulation with distortions in the composition of learning. Nevertheless, under plausible parameter combinations, the

government still optimally subsidizes observable investments at roughly the marginal rate at which the returns to the investments in human capital are taxed. Non-pecuniary costs and benefits of education do not affect this result substantially unless non-pecuniary benefits are more important than non-pecuniary costs and are more related to leisure rather than work. These conditions, however, are unlikely to hold in practice. From a policy perspective, a substantial part of the observed education subsidies can be explained as an instrument to alleviate the distortionary impact of the redistributive tax system on investments in human capital.

Acknowledgements

We thank Thomas Piketty and two anonymous referees for their useful comments and suggestions. We benefited also from comments by participants of seminars held at CPB Netherlands Bureau for Economic Policy Analysis, the Dutch Central Bank and the European University Institute - Florence. Bas Jacobs gratefully acknowledges financial support from the NWO Priority Program ‘Scholar’ funded by the Netherlands Organization for Sciences. Earlier versions of this paper (CEPR Discussion Paper No. 3309 and a working paper) circulated under the same title but with quite different contents.

Appendix A. Second-order conditions of individual optimization

Substituting the household budget constraint (Eq. (3)) into the utility function (Eq. (2)) to eliminate c_n , we arrive at the following maximization problem

$$\max_{\{l_n, e_n\}} U_n = (1-t)n\phi(e_n)l_n - (1-s)p_e e_n + g - \frac{l_n^{1+1/\varepsilon}}{1+1/\varepsilon}. \quad (45)$$

The first-order conditions are

$$\frac{\partial U_n}{\partial l_n} = (1-t)n\phi(e_n) - l_n^{1/\varepsilon} = 0, \quad (46)$$

$$\frac{\partial U_n}{\partial e_n} = (1-t)n\phi'(e_n)l_n - (1-s)p_e = 0. \quad (47)$$

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

$$H \equiv \begin{bmatrix} -\frac{1}{\varepsilon} l_n^{1/\varepsilon-1} & (1-t)n\phi'(e_n) \\ (1-t)n\phi'(e_n) & (1-t)n l_n \phi''(e_n) \end{bmatrix}. \quad (48)$$

For utility to reach a maximum, the Hessian matrix should be negative definite. This is the case if the leading principal minors of H switch sign. The first principal minor is negative. Therefore, the second leading principal minor must be positive, i.e.,

$-(1/\varepsilon) l_n^{1/\varepsilon} (1-t)n\phi''(e_n) - ((1-t)n\phi'(e_n))^2 > 0$. Using Eq. (26) to eliminate l_n and substituting Eq. (1), this inequality can be written as

$$\mu \equiv 1 - \beta(1 + \varepsilon) > 0. \tag{49}$$

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

Appendix B. Elasticities of individual behavior

We derive the behavioral elasticities from the model with two educational inputs. The model with one educational input is a special case of that more general model. The first-order conditions for utility maximization with respect to x_n and y_n are

$$(1-t)n\phi'(e_n)l_n\psi_x(x_n, y_n) = (1-s)p_x, \tag{50}$$

$$(1-t)n\phi'(e_n)l_n\psi_y(x_n, y_n) = p_y. \tag{51}$$

We multiply Eq. (50) by $(1-t)x_n$ and multiply Eq. (51) by y_n , and add the results. Using the property that $\psi(x_n, y_n)$ is homogeneous of degree one, $\psi_x(\cdot)x_n + \psi_y(\cdot)y_n = \psi(x_n, y_n)$, we then find (using $\phi'(e_n) = \beta\phi(e_n)/e_n$)

$$\beta(1-t)n\phi(e_n)l_n = p_e e_n, \tag{52}$$

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

Log-linearizing Eq. (52) (using $\phi(e_n) = e_n^\beta$) and the definition of p_e (using $e_n \equiv \psi(x_n, y_n)$ and Eqs. (50) and (51)), we arrive at

$$-\tilde{t} + \tilde{n} + \tilde{l}_n + \beta\tilde{e}_n = \tilde{p}_e + \tilde{e}_n, \tag{53}$$

$$\tilde{p}_e = -\alpha\tilde{s}, \tag{54}$$

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

$\alpha_n \equiv x_n\psi_x/\psi = (1-s)p_x x_n/(p_e e_n)$ does not depend on ability, except where we have non-linear education subsidies. The reason is that Eqs. (50) and (51) imply that the marginal rate of transformation does not depend on n :

$$\frac{\psi_x(x_n, y_n)}{\psi_y(x_n, y_n)} = \frac{(1-s)p_x}{p_y}. \tag{55}$$

$\psi_x(x_n, y_n)$ and $\psi_y(x_n, y_n)$ are functions of x_n/y_n only (since $\psi(x_n, y_n)$ is homogeneous of the first degree). Hence, Eq.(55) determines the ratio of the two inputs x_n/y_n as a function

of $(1-s)p_x/p_y$ only. $\alpha = \frac{(1-s)p_x x_n}{(1-s)p_x x_n + p_y y_n} = \frac{(1-s)p_x(x_n/y_n)}{(1-s)p_x(x_n/y_n) + p_y}$ depends therefore only on $(1-s)p_x$ and p_y , which do not depend on ability.

Expression (5) implies that compensated labor supply depends only on the after-tax wage rate $w \equiv (1-t)n\phi(e_n)$, so that

$$\tilde{l}_n = \varepsilon(-\tilde{t} + \tilde{n} + \beta\tilde{e}_n). \tag{56}$$

Substituting Eqs. (54) and (56) into Eq. (53) to eliminate, respectively, \tilde{l}_n and \tilde{p}_e , we can solve for $\tilde{\phi}_n = \beta\tilde{e}_n$:

$$\tilde{\phi}_n = \beta\tilde{e}_n = \frac{\beta}{\mu}((1 + \varepsilon)(\tilde{n} - \tilde{t}) + \alpha\tilde{s}), \tag{57}$$

where $\mu \equiv 1 - \beta(1 + \varepsilon) > 0$. Substituting Eq. (57) into Eq. (56), we solve for \tilde{l}_n :

$$\tilde{l}_n = \frac{\alpha\beta\varepsilon}{\mu}\tilde{s} + \frac{\varepsilon}{\mu}(\tilde{n} - \tilde{t}). \tag{58}$$

We find \tilde{x}_n and \tilde{y}_n by differentiating $e_n \equiv \psi(x_n, y_n)$ and using Eqs. (50) and (51) to eliminate ψ_x and ψ_y to arrive at

$$\tilde{e}_n = \alpha\tilde{x}_n + (1 - \alpha)\tilde{y}_n. \tag{59}$$

Differentiation of Eq. (55) yields

$$\tilde{x}_n - \tilde{y}_n = \sigma\tilde{s}, \tag{60}$$

where $\sigma \equiv -(\text{dlog}(x_n)/(y_n))/(\text{dlog}(\psi_x(\cdot)/\psi_y(\cdot)))$ is the elasticity of substitution between the two inputs in the production of human capital. This substitution elasticity depends only on x_n/y_n and is thus independent of ability in all cases (except where we have non-linear education subsidies) just like α . Using Eqs. (59) and (60), we can express \tilde{x}_n and \tilde{y}_n in terms of \tilde{e}_n and substituting Eq. (57) to eliminate \tilde{e}_n , we find:

$$\tilde{x}_n = \frac{(1 + \varepsilon)(\tilde{n} - \tilde{t}) + \alpha\tilde{s}}{\mu} + (1 - \alpha)\sigma\tilde{s}, \tag{61}$$

$$\tilde{y}_n = \frac{(1 + \varepsilon)(\tilde{n} - \tilde{t}) + \alpha\tilde{s}}{\mu} - \alpha\sigma\tilde{s}. \tag{62}$$

Eqs. (58), (61), (62), and $\tilde{e}_n = \alpha\tilde{x}_n + (1 - \alpha)\tilde{y}_n$ yield the following elasticities:

$$\varepsilon_{lt} \equiv \frac{\partial l_n}{\partial(1-t)} \frac{(1-t)}{l_n} = \frac{\varepsilon}{\mu}, \tag{63}$$

$$\varepsilon_{et} \equiv \frac{\partial e_n}{\partial(1-t)} \frac{(1-t)}{e_n} = \frac{1 + \varepsilon}{\mu}, \tag{64}$$

$$\varepsilon_{xt} \equiv \frac{\partial x_n}{\partial(1-t)} \frac{(1-t)}{x_n} = \frac{1+\varepsilon}{\mu}, \tag{65}$$

$$\varepsilon_{yt} \equiv \frac{\partial y_n}{\partial(1-t)} \frac{(1-t)}{y_n} = \frac{1+\varepsilon}{\mu}, \tag{66}$$

$$\varepsilon_{ls} \equiv \frac{\partial l_n}{\partial(1-s)} \frac{(1-s)}{l_n} = -\frac{\alpha\beta\varepsilon}{\mu}, \tag{67}$$

$$\varepsilon_{es} \equiv \frac{\partial e_n}{\partial(1-s)} \frac{(1-s)}{e_n} = -\frac{\alpha}{\mu}, \tag{68}$$

$$\varepsilon_{xs} \equiv \frac{\partial x_n}{\partial(1-s)} \frac{(1-s)}{x_n} = -\frac{\alpha}{\mu} - (1-\alpha)\sigma, \tag{69}$$

$$\varepsilon_{ys} \equiv \frac{\partial y_n}{\partial(1-s)} \frac{(1-s)}{y_n} = -\frac{\alpha}{\mu} + \alpha\sigma. \tag{70}$$

We replace t and s by T' and S' when we consider non-linear tax instruments.

Appendix C. Optimal linear policies

We derive optimal policy in the model with two educational inputs. The model with one educational input is a special case. The Lagrangian for maximizing social welfare is designated by

$$\max_{\{g,t,s\}} \mathcal{L} = \int_{\underline{n}}^{\bar{n}} \Psi(v(g,t,s,n))dF(n) + \lambda \int_{\underline{n}}^{\bar{n}} (tn\phi(\psi(x_n,y_n))l_n - sp_x x_n - g - E)dF(n). \tag{71}$$

The first-order conditions for maximization of social welfare (Eq. (71)) with respect to t and s are

$$\begin{aligned} \frac{\partial \mathcal{L}}{\partial t} = \int_{\underline{n}}^{\bar{n}} & \left[(-\Psi(u_n) + \lambda)n\phi(\cdot)l_n \right. \\ & \left. + \lambda \left(tn\phi(\cdot) \frac{\partial l_n}{\partial t} + \frac{(t-s)p_x}{(1-t)} \frac{\partial x_n}{\partial t} + \frac{tp_y}{(1-t)} \frac{\partial y_n}{\partial t} \right) \right] dF(n) = 0, \end{aligned} \tag{72}$$

$$\begin{aligned} \frac{\partial \mathcal{L}}{\partial s} = \int_{\underline{n}}^{\bar{n}} & \left[(\Psi(u_n) - \lambda)p_x x_n \right. \\ & \left. + \lambda \left(tn\phi(\cdot) \frac{\partial l_n}{\partial s} + \frac{(t-s)p_x}{(1-t)} \frac{\partial x_n}{\partial s} + \frac{tp_y}{(1-t)} \frac{\partial y_n}{\partial s} \right) \right] dF(n) = 0, \end{aligned} \tag{73}$$

where we used Eqs. (50), (51), and Roy’s lemma

$$\frac{\partial v(g, t, s, n)}{\partial t} = -n\phi(e_n)l_n, \tag{74}$$

$$\frac{\partial v(g, t, s, n)}{\partial s} = p_x x_n. \tag{75}$$

The first-order condition for t can be simplified upon substitution of the definition of b_n from Eq. (10) and the tax elasticities ($\varepsilon_{jt} \equiv (\partial j_n / \partial (1-t))((1-t)/j_n)$, $j = l_n, x_n, y_n$):

$$\int_{\underline{n}}^{\bar{n}} \left[(1 - b_n)n\phi(\cdot)l_n - \frac{tn\phi(\cdot)l_n\varepsilon_{lt}}{1-t} - \frac{(t-s)}{(1-t)} \frac{p_x x_n \varepsilon_{xt}}{(1-t)} - \frac{t}{(1-t)} \frac{p_y y_n \varepsilon_{yt}}{(1-t)} \right] dF(n) = 0. \tag{76}$$

We use now Eqs. (20) and (21) to write $p_x x_n$ and $p_y y_n$ in terms of $n\phi(\cdot)l_n$ to yield (after dividing by $\int_{\underline{n}}^{\bar{n}} n\phi(\cdot)l_n dF(n)$):

$$\xi - \frac{t}{(1-t)} \varepsilon_{lt} - \frac{(t-s)}{(1-s)(1-t)} \alpha \beta \varepsilon_{xt} - \frac{t}{(1-t)} (1-\alpha) \beta \varepsilon_{yt} = 0. \tag{77}$$

Similarly, we can rewrite the first-order condition for the subsidy s (Eq. (73)) by using the subsidy elasticities ($\varepsilon_{js} \equiv (\partial j_n / \partial (1-s))((1-s)/j_n)$, $j = l_n, x_n, y_n$):

$$\int_{\underline{n}}^{\bar{n}} \left[(b_n - 1)p_x x_n - \frac{tn\phi(\cdot)l_n}{1-s} \varepsilon_{ls} - \frac{(t-s)p_x x_n \varepsilon_{xs}}{(1-t)(1-s)} - \frac{tp_y y_n \varepsilon_{ys}}{(1-t)(1-s)} \right] dF(n) = 0. \tag{78}$$

We employ Eqs. (20) and (21) to write $p_x x_n$ and $p_y y_n$ in terms of $n\phi(\cdot)l_n$ to obtain (after dividing through by $-\frac{(1-t)\alpha\beta}{(1-s)} \int_{\underline{n}}^{\bar{n}} n\phi(\cdot)l_n dF(n)$)

$$\xi + \frac{t}{(1-t)} \frac{\varepsilon_{ls}}{\alpha\beta} + \frac{(t-s)}{(1-s)(1-t)} \varepsilon_{xs} + \frac{t}{(1-t)} \frac{(1-\alpha)}{\alpha} \varepsilon_{ys} = 0. \tag{79}$$

C.1. Optimal linear income taxation without education subsidies

With $s = 0$, expression (77) yields Eq. (13) by setting $\alpha = 1$ and $\varepsilon_{xt} = \varepsilon_{et}$ (since e_n consists only of verifiable investment x_n in this case), and using $\varepsilon_{zt} = \varepsilon_{lt} + \varepsilon_{et}$.

C.2. Optimal linear education subsidies with exogenous tax rate

Optimal education subsidies with only one observable input (Eq. (17)) follow from Eq. (79) by substituting the elasticities (67) and (69) and setting $\alpha = 1$ (since e_n consists of verifiable investment x_n only).

C.3. Optimal linear income taxation and linear education subsidies

Subtracting Eq. (77) from (79), we find

$$\frac{t}{1-t} \left(\varepsilon_{lt} + \frac{\varepsilon_{ls}}{\alpha\beta} \right) + \frac{t-s}{(1-t)(1-s)} (\alpha\beta\varepsilon_{xt} + \varepsilon_{xs}) + \frac{t}{1-t} \left((1-\alpha)\beta\varepsilon_{yt} + \frac{1-\alpha}{\alpha}\varepsilon_{ys} \right) = 0. \tag{80}$$

Using the expressions for the elasticities (63), (65–67), (69), and (70), we derive that $\varepsilon_{lt} + \varepsilon_{ls}/(\alpha\beta) = 0$ and $\alpha\beta\varepsilon_{xt} + \varepsilon_{xs} = -1 + (1-\alpha)(1-\sigma)$ and $(1-\alpha)\beta\varepsilon_{yt} + (1-\alpha)/\alpha\varepsilon_{ys} = -(1-\alpha)(1-\sigma)$, so that

$$\frac{s-t}{1-s} = \frac{(1-\alpha)(1-\sigma)}{1-(1-\alpha)(1-\sigma)} t, \tag{81}$$

which can be written as Eq. (23). Substitution of Eq. (81) into (79) to eliminate the term with $(t-s)$ and using the definitions for the various elasticities, we arrive at Eq. (24).

Setting $\alpha = 1$ in Eqs. (81) and (24) yield the corresponding expressions with only verifiable inputs in human capital (i.e., Eqs. (15) and (16)).

Appendix D. Optimal non-linear policies

D.1. Optimal non-linear income taxation without education subsidies

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

$$\frac{\partial \mathcal{H}}{\partial l_n} = \lambda \left(n\phi(e_n) - \frac{dc_n}{dl_n} \Big|_{\bar{u}} + (n\phi'(e_n)l_n - p_e) \frac{de_n}{dl_n} \Big|_{\bar{u}} \right) f(n) - \frac{\theta_n}{n} \left(1 + \frac{1}{\varepsilon} \right) l_n^{1/\varepsilon} = 0. \tag{82}$$

Substituting $\frac{dc_n}{dl_n} \Big|_{\bar{u}} = -\frac{u_l}{u_c} = l_n^{1/\varepsilon} = (1-T')n\phi(e_n)$ (derived by totally differentiating utility $u(c_n, l_n)$ and substituting labor supply (Eq. (26))), the private first-order condition for e_n (Eq. (27)), (i.e., $(1-T')n\phi'(e_n)l_n - p_e = 0$ with $S(\cdot) = 0$), and Eq. (26) (to eliminate l_n), we establish the optimal marginal income tax rates:

$$\frac{T'}{1-T'} = \frac{\theta_n/\lambda}{nf(n)} \left(1 + \frac{1}{\varepsilon} \right) \left(\frac{1}{1 + \beta\varepsilon_{el}} \right), \tag{83}$$

where $\varepsilon_{el} \equiv \frac{l_n}{e_n} \frac{de_n}{dl_n} \Big|_{\bar{u}} = \frac{\varepsilon_{el}}{\varepsilon_{lt}} = \frac{1+\varepsilon}{\varepsilon}$, since the government employs the marginal tax rate to change l_n at a given utility level. $\varepsilon_{el}/\varepsilon_{lt} = (1+\varepsilon)/\varepsilon$ from Eqs. (63) and (64). Upon

substitution of the expressions for the elasticities and the elasticity of earnings with respect to ability (Eq. (86)), we establish Eq. (33). To rewrite the last expression in terms of the earnings distribution rather than the ability distribution, we employ the identity $Q(z_n) \equiv F(n)$, which states that the cumulative distribution of abilities must by definition equal the cumulative distribution of gross earnings. Differentiating this identity with respect to n , we arrive at

$$nf(n) = z_n q(z_n) \varepsilon_{zn}, \quad (84)$$

where $\varepsilon_{zn} \equiv (dz_n/dn)(n/z_n)$ denotes the elasticity of earnings with respect to ability. In order to derive the total differential of z_n with respect to ability n , we observe that a change in the ability distribution modifies T' by affecting z_n . The partial elasticities of earnings with respect to abilities and marginal taxes are $\varepsilon_{zn} \equiv (\partial z_n/\partial n)(n/z_n) = (1+\varepsilon)/\mu$ and $\varepsilon_{zt} \equiv (\partial z_n/\partial(1-T'(z_n)))(1-T'(z_n))/z_n = (\varepsilon + \beta(1+\varepsilon))/\mu$, respectively. These elasticities are derived using $(\partial l_n/\partial n)n/l_n = \varepsilon_{lt} = \varepsilon/\mu$, $(\partial e_n/\partial n)n/e_n = \varepsilon_{et} = (1+\varepsilon)/\mu$, and $\tilde{z}_n = \tilde{l}_n + \beta\tilde{e}_n + \tilde{n}$, since $z_n = l_n n \phi(e_n)$. Hence, for the overall change in earnings, which depends on changes in ability and marginal tax rates, we can write,

$$\tilde{z}_n = \frac{(1+\varepsilon)}{\mu} \tilde{n} - \frac{\varepsilon + \beta(1+\varepsilon)}{\mu} \tilde{T}', \quad (85)$$

where $\tilde{T}' \equiv dT'(z_n)/(1-T') = (z_n T''/(1-T')) \tilde{z}_n$. Consequently, we have

$$\varepsilon_{zn} \equiv \frac{dz_n}{dn} \frac{n}{z_n} = \frac{\tilde{z}_n}{\tilde{n}} = \frac{\frac{1+\varepsilon}{\mu}}{1 + \frac{\varepsilon + \beta(1+\varepsilon)}{\mu} \frac{z_n T''(z_n)}{1-T'(z_n)}}. \quad (86)$$

Using this elasticity in $nf(n) = z_n q(z_n) \varepsilon_{zn}$ and substituting of the result in Eq. (33), we arrive at Eq. (34).

The first-order condition for u_n is

$$\frac{\partial \mathcal{H}}{\partial u_n} = \left(\Psi'(u_n) + \lambda(n\phi'(e_n)l_n - p_e) \frac{de_n}{du_n} \Big|_{\bar{l}} - \lambda \frac{dc_n}{du_n} \Big|_{\bar{l}} \right) f(n) = \frac{d\theta_n}{dn}. \quad (87)$$

Recall we defined θ_n negatively; hence there is no minus sign on the right-hand side. We can simplify this expression using the envelope theorem, which implies that $\frac{de_n}{du_n} \Big|_{\bar{l}} = 0$, and $\frac{dc_n}{du_n} \Big|_{\bar{l}} = 1$ (due to constant unit marginal utility of income), so that

$$(\Psi'(u_n) - \lambda)f(n) = \frac{d\theta_n}{dn}. \quad (88)$$

Integrating this expression and employing the transversality conditions (32), we find

$$\frac{\int_n^{\bar{n}} \frac{d\theta_n}{dm} dm}{\lambda} = -\frac{\theta_n}{\lambda} = \int_n^{\bar{n}} \left(\frac{\Psi'(u_n)}{\lambda} - 1 \right) f(m) dm = \int_n^{\bar{n}} (b_m - 1) f(m) dm. \quad (89)$$

D.2. Optimal non-linear income taxation and non-linear education subsidies

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

$$\frac{\partial \mathcal{H}}{\partial e_n} = \lambda \left(n\phi'(e_n)l_n - p_e - \frac{dc_n}{de_n} \Big|_{\bar{u}, \bar{l}} \right) f(n) = 0. \tag{90}$$

For the indirect impacts on consumption, we find $\frac{dc_n}{de_n} \Big|_{\bar{u}, \bar{l}} = 0$ by differentiating the household budget constraint (Eq. (25)) and substituting the individuals' first-order condition for e_n (Eq. (27)). Using these results in Eq. (90), we find Eq. (35).

The first-order condition for l_n is given by

$$\frac{\partial \mathcal{H}}{\partial l_n} = \lambda \left(n\phi(e_n) - \frac{dc_n}{dl_n} \Big|_{\bar{u}, \bar{e}} \right) f(n) - \frac{\theta_n}{n} \left(1 + \frac{1}{\varepsilon} \right) l_n^{1/\varepsilon} = 0, \tag{91}$$

where we substitute $\frac{dc_n}{dl_n} \Big|_{\bar{u}, \bar{e}} = (1 - T')n\phi(e_n)$ (found by taking the total derivative of utility $u(c_n, l_n)$ and substituting Eq. (26) to eliminate $u_l = -l_n^{1/\varepsilon}$), and use the first-order condition for labor supply (Eq. (26)) to arrive at

$$\frac{T'}{1 - T'} = \frac{\theta_n/\lambda}{nf(n)} \left(1 + \frac{1}{\varepsilon} \right). \tag{92}$$

In order to write this in terms of the earnings distribution, we need to find the ability elasticity of earnings. A change in the ability distribution changes T' and S' through changes in z_n and e_n . In addition to the partial earnings elasticities of abilities and marginal taxes (see the derivations above for the optimal non-linear tax in the absence of education subsidies), we use the elasticities of marginal subsidies ($\varepsilon_{ls} = -\beta\varepsilon/\mu$, and $\varepsilon_{es} = -1/\mu$) so that $\varepsilon_{zs} \equiv (\partial z_n / \partial (1 - S'))((1 - S')/z_n) = -\beta(1 + \varepsilon)/\mu$. The overall change in earnings, which depends on changes in abilities, marginal tax rates, and marginal subsidy rates, given by

$$\tilde{z} = \frac{(1 + \varepsilon)}{\mu} \tilde{n} - \frac{\varepsilon + \beta(1 + \varepsilon)}{\mu} \tilde{T}' + \frac{\beta(1 + \varepsilon)}{\mu} \tilde{S}'. \tag{93}$$

$S' = T'$ if educational subsidies are optimal (see Eq. (36)), so that we have (using $\tilde{S}' = \tilde{T}' \equiv dT'(z_n)/(1 - T') = (z_n T''/(1 - T'))\tilde{z}_n$

$$\varepsilon_{zn} \equiv \frac{dz_n}{dn} \frac{n}{z_n} = \frac{\tilde{z}_n}{\tilde{n}} = \frac{\frac{1 + \varepsilon}{\mu}}{1 + \frac{\varepsilon}{\mu} \frac{z_n T''(z_n)}{1 - T'(z_n)}}. \tag{94}$$

Substitution of this expression into Eq. (84), and the result into Eq. (92), yields Eq. (37).

D.3. Optimal non-linear policies and immaterial costs and benefits of education

With the utility function $U(c_n, z_n, e_n, n) \equiv c_n - v((z_n)/(n\phi(e_n)), e_n)$, the incentive compatibility constraints are given by (after applying the envelope theorem)

$$\frac{dU_n}{dn} = \frac{l_n}{n} v_l(l_n, e_n). \quad (95)$$

The Hamiltonian for maximizing social welfare is thus given by

$$\max_{\{e_n, l_n, u_n\}} \mathcal{H} = \Psi(u_n)f(n) - \theta_n \frac{l_n}{n} v_l(l_n, e_n) + \lambda(n\phi(e_n)l_n - p_e e_n - c_n - E)f(n). \quad (96)$$

Accordingly, the first-order condition for e_n amounts to

$$\frac{\partial \mathcal{H}}{\partial e_n} = \lambda \left(n\phi'(e_n)l_n - p_e - \frac{dc_n}{de_n} \Big|_{\bar{u}, \bar{l}} \right) f(n) - \frac{\theta_n}{n} l_n v_{le}(\cdot) = 0. \quad (97)$$

Totally differentiating the household budget constraint (at $dl_n = 0$) and substituting the first-order condition for e_n (Eq. (40)) gives $\frac{dc_n}{de_n} \Big|_{\bar{l}} = v_e$. Furthermore, the first-order condition for e_n (Eq. (40)) can be written as $p_e/v_e = 1/(\omega_n(1-S'))$, where $\omega_n \equiv v_e/((1-S')p_e) = v_e/((1-T')nl_n\phi' - v_e)$. Substituting Eq. (40) into Eq. (97) to eliminate $n\phi'(e_n)l_n$, and substituting $\frac{dc_n}{de_n} \Big|_{\bar{l}} = v_e$ and $p_e/v_e = 1/(\omega_n(1-S'))$, we arrive at Eqs. (41) and (43).

References

- Acemoglu, Daron J., Angrist, Joshua, 2000. How large are the social returns to education? Evidence from compulsory schooling laws. NBER Macroeconomics Annual, 9–59.
- Alstedsaeter, Annette, 2003. Does the tax system encourage too much education? Finanzarchiv 59 (1), 27–48.
- Andersson, Fredrik, Konrad, Kai A., 2003. Human capital investment and globalization in extortionary states. Journal of Public Economics 87 (7–8), 1539–1555.
- Atkinson, Anthony B., Stiglitz, Joseph E., 1980. Lectures on Public Economics. McGraw-Hill, New York.
- Becker, Gary S., 1965. A theory of the allocation of time. Economic Journal 75, 493–517.
- Boadway, Robin, Marceau, Nicolas, Marchand, Maurice, 1996. Investment in education and the time inconsistency of redistributive tax policy. Economica 63, 171–189.
- Boskin, Michael, 1975. Notes on the tax treatment of human capital. NBER Working Paper, 116.
- Bovenberg, Lans A., Jacobs, Bas, 2003. Redistribution and Education Subsidies are Siamese Twins, Tinbergen Institute Discussion Paper 05-036/3.
- Cameron, Stephen, Taber, Christopher, 2000. Borrowing constraints and the returns to schooling. NBER Working Paper, 7761.
- Card, David, 1999. The causal effect of education on earnings. In: Ashenfelter, Orley, Card, David (Eds.), Handbook of Labor Economics, vol. 3A. Elsevier-North Holland, Amsterdam.
- Cameiro, Pedro, Heckman, James J., 2004. Human capital policy. In: Friedman, Benjamin M. (Ed.), Inequality in America: What Role for Human Capital Policies? MIT Press, Cambridge-MA.
- Diamond, Peter A., 1998. Optimal income taxation: an example with a u-shaped pattern of optimal marginal tax rates. American Economic Review 88, 83–95.

- Diamond, Peter A., Mirrlees, James A., 1971. Optimal taxation and public production: I. Production efficiency. *American Economic Review* 61, 8–27.
- Diamond Peter A., Mirrlees, James A., 2002. Optimal Taxation and the Le Chatelier Principle, MIT Department of Economics Working Paper Series 02–30.
- Dur, Robert, Teulings, Coen N., 2001. Education and Efficient Redistribution, Tinbergen Institute Discussion Paper 090/3.
- Eaton, Jonathan, Rosen, Harvey S., 1980. Taxation, human capital and uncertainty. *American Economic Review* 70 (4), 705–715.
- Ebert, Udo, 1992. A reexamination of the optimal nonlinear income tax. *Journal of Public Economics* 49 (1), 47–73.
- Ewijk, Casper van, Tang, Paul J.G., 2000. Efficient Progressive Taxes and Education Subsidies. CPB Research Memorandum vol. 170. CPB Netherlands Bureau for Economic Policy Analysis.
- Hare, Paul G., Ulph, David T., 1979. On education and distribution. *Journal of Political Economy* 87, S193–S212.
- Harmon, Colm, Oosterbeek, Hessel, Walker, Ian, 2003. The returns to education: microeconomics. *Journal of Economic Surveys* 17, 115–155.
- Heckman, James J., 1976. A life-cycle model of earnings, learning and consumption. *Journal of Political Economy* 4, S11–S44.
- Heckman, James J., Klenow, Peter J., 1998. Human capital policy. In: Boskin, M. (Ed.), *Policies to Promote Human Capital Formation*. Hoover Institution.
- Jacobs, Bas, Wijnbergen, Sweder J.G. van, 2005. Capital market failure, adverse selection and equity financing of higher education, Tinbergen Institute Discussion Paper 05-037/3.
- Judd, Kenneth L., 1999. Optimal taxation and spending in general competitive growth models. *Journal of Public Economics* 71, 1–26.
- Kotlikoff, Laurence J., Summers, Lawrence H., 1979. Tax incidence in a life cycle model with variable labor supply. *Quarterly Journal of Economics* 93, 705–718.
- Krueger, Alan B., Lindahl, Mikael, 2002. Education and growth: why and for whom. *Journal of Economic Literature* 39 (4), 1101–1136.
- Milesi-Ferretti, Gian Maria, Roubini, Nouriel, 1998. On the taxation of human and physical capital in models of endogenous growth. *Journal of Public Economics* 70, 237–254.
- Mirrlees, James A., 1971. An exploration in the theory of optimum income taxation. *Review of Economic Studies* 38, 175–208.
- Mirrlees, James A., 1986. The theory of optimal income taxation. In: Arrow, Kenneth J., Intriligator, Michael D. (Eds.), *Handbook of Mathematical Economics*, vol. 3. Elsevier-North Holland, Amsterdam.
- Nielsen, Søren B., Sørensen, Peter B., 1997. On the optimality of the nordic system of dual income taxation. *Journal of Public Economics* 63, 311–329.
- OECD, 2002. *Tax Data Base 2002*. OECD, Paris.
- OECD, 2003. *Education at a Glance 2003*. OECD, Paris.
- Saez, Emmanuel, 2001. Using elasticities to derive optimal income tax rates. *Review of Economic Studies* 68, 205–229.
- Shea, John, 2000. Does parents' money matter? *Journal of Public Economics* 77, 155–184.
- Tinbergen, Jan, 1970. Belasting op Bekwaamheid. *Intermediair* 30 (6), 1–3 (“Tax on Ability” in Dutch).
- Trostel, Philip A., 1993. The effect of taxation on human capital. *Journal of Political Economy* 101, 327–350.
- Ulph, David T., 1977. On the optimal distribution of income and educational expenditure. *Journal of Public Economics* 8, 341–356.