Optimal Taxation of Human Capital and the Earnings Function*

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

February 17, 2008

Abstract

This paper explores how the specification of the earnings function impacts the optimal tax treatment of human capital. If education is complementary to labor effort, education should be subsidized to offset tax distortions on labor supply. However, if most of the education is enjoyed by high ability households, education should be taxed in order to redistribute resources to the poor. The paper identifies the exact conditions under which these two effects cancel and education should be neither taxed nor subsidized. In particular, with non-linear tax instruments, education should be weakly separable from labor and ability in the earnings function. With linear taxes, education should also feature a constant elasticity in a weakly separable earnings function.

Key-words: optimal linear and non-linear taxation, optimal education subsidies, human capital, earnings function
JEL-codes: H2, H5, I2, J2

1 Introduction

Bovenberg and Jacobs (2005) extend optimal tax models with endogenous skill formation. They find that redistributive governments should employ education subsidies to offset tax distortions on skill formation in order to ensure efficiency in human capital investment. Education is neither taxed nor subsidized on a net basis in the optimum. This result holds for both linear and non-linear tax instruments if the government can verify all educational investments.

This paper demonstrates that these efficiency results critically hinge on the presumed specification of the earnings function. In particular, non-linear taxes on education are zero only if the earnings function is weakly separable between ability and labor effort,

*Bas Jacobs gratefully acknowledges financial support from the Netherlands Organization for Sciences (Vidi Grant No. 452-07-013, 'Skill Formation in Distorted Labor Markets'). Part of this research has been done while Jacobs was visiting scholar at the Center of Economic Studies of the Ludwig-Maximillians-Universität München. Comments and suggestions by seminar participants at the University of Konstanz are gratefully acknowledged. The usual disclaimer applies.

[†]Erasmus University Rotterdam, Tinbergen Institute, Netspar and CESifo. Corresponding author: Erasmus School of Economics, Erasmus University Rotterdam, PO box 1738, 3000 DR Rotterdam, The Netherlands. Phone: +31–10–4081481. Fax: +31–10–4089166. E-mail: bjacobs@few.eur.nl. [‡]University of Tilburg, CentER, Netspar, CESifo and CEPR. E-mail: a.l.bovenberg@uvt.nl.

on the one hand, and education, on the other hand. Under this condition, the positive efficiency effect on work effort of subsidizing education exactly offsets the negative equity effect of doing so. Accordingly, education is neither subsidized nor taxed on a net basis in the optimum. Education should optimally be subsidized (on a net basis) for efficiency reasons if education is complementary to labor supply; by acting as an implicit subsidy on work effort, education subsidies alleviate the distortions of a redistributive labor tax on work. However, education should be taxed for equity reasons if education increases with ability. If education is complementary to both work effort and ability, efficiency gains and equity losses of education subsidies work in opposite directions. Education should be subsidized only if education is more complementary with work effort than with ability. In that case, the efficiency gains on account of lower tax distortions on labor supply produced by education subsidies outweigh the distributional losses resulting from the regressive incidence of these education subsidies. If education is weakly separable from ability and labor in the earnings function, both effects exactly cancel so that education is optimally neither subsidized nor taxed.

With linear policy instruments, we show that a separable earnings function should also feature a constant elasticity in education to ensure that optimal net taxes on human capital investments are zero.[1] With this specific earnings function, labor income and educational investments for agents with different skill levels are related in a linear fashion. As a direct consequence, education taxes and income taxes yield the same distributional consequences and produce the same distortions on labor supply. In contrast to labor taxes, however, education taxes distort not only labor supply but also human capital formation. Hence, compared to labor taxes, education taxes produce more distortions in arriving at the same distributional impact. Redistributive governments should thus employ only labor taxes, and should set education taxes to zero.

Whether education should be taxed or subsidized on a net basis in an optimal redistributive program thus crucially depends on two factors, namely, first, the degree of complementarity between education and work effort, and, second, the incidence of education subsidies on various skill levels. Education subsidies become more efficient to fight income-tax distortions on labor supply if education and work are highly complementary. Empirically, education and labor effort are complements because better-educated workers exhibit larger participation rates, retire later and work more hours (OECD, 2006).[2] At the same time, however, education subsidies are regressive, in view of the well-documented ability bias in education (Card, 1999). Which of the two factors is more important for the setting of optimal taxes and education policies thus remains an open empirical question.

The public finance literature on education and taxation typically adopts earnings functions that satisfy the weak separability and constant elasticity requirements.[3] Mal-

[1]Our model considers only one input invested in education. However, if there are multiple inputs in the human capital production function (e.g. time, goods and/or effort), the production function needs to be homothetic in these inputs (see Bovenberg and Jacobs, 2005).

[2]The correlation between labor supply and education may not be causal. A higher ability, for example, may boost both education and labor supply. The empirical literature on education and labor supply (e.g. Card, 1999), however, shows that the returns to education continue to be substantial if one corrects for ability bias while labor supply does respond to exogenous variations in wages (e.g. Blundell and MaCurdy, 1999). Education thus does raise labor supply.

[3]See, for example, Nielsen and Sorensen (1997), Brett and Weymark (2003), Wigger (2004), Jacobs (2005, 2007), Jacobs and Bovenberg (2005), Richter (2006), Blumkin and Sadka (2007), Bohacek and Kapicka (2007), and others. Separability of human capital and labor in labor earnings is also adopted in classical papers on life-cycle models with education; see, for example, Heckman (1976), Kotlikoff and Summers (1979), and Eaton and Rosen (1980); or in modern articles on growth with endogenous human

donado (2007), however, employs an earnings function in which ability and labor effort are not necessarily separable from education. We generalize and interpret his findings by employing more general earnings functions. Ulph (1977) and Hare and Ulph (1979) were not able to obtain clear-cut interpretations of optimal tax and education policies for general earnings functions. This paper, in contrast, provides intuitive characterizations of optimal non-linear tax and education policies for general specifications of earnings functions.

Our findings are also relevant for the 'new public finance' literature. For example, Grochulski and Piskorsi (2006) and Da Costa and Mestri (2007) adopt a similar earnings function as Maldonado (2007), but concentrate mainly on the desirability of capital income and wealth taxes. In particular, Grochulski and Piskorsi (2006) do not optimize education policy, since education is assumed to be non-verifiable, whereas Da Costa and Mestri (2007) do not explicitly elaborate on the role of education policy to alleviate the distortions of the labor income tax. Our paper helps to gain a deeper understanding about the interaction between optimal tax and education policies in the presence of these and more complex earnings functions that are adopted in this literature.

The rest of this paper is structured as follows. Section 2 formalizes the model. Subsequently, sections 3 and 4 explore optimal education taxes for non-linear and linear policy instruments, respectively. Section 5 concludes.

2 The model

This section briefly summarizes the main features of the model, which extends Bovenberg and Jacobs (2005) with a more general earnings function. $n \in [\underline{n}, \overline{n}]$ is individual ability, $f(n)$ is the density of ability, and $F(n)$ is the cumulative distribution of ability. e_n denotes investment in education, l_n represents labor effort, c_n is consumption and z_n stands for gross labor earnings.

Households exhibit identical utility functions and derive utility from consumption c_n and suffer disutility from work effort l_n:

$$u(c_n, l_n), \quad u_c > 0, \quad u_{cc} < 0, \quad u_l < 0, \quad u_{ll} < 0, \tag{1}$$

where subscripts refer to an argument of differentiation (except where it signifies ability n). This specification generalizes the separable utility function in Bovenberg and Jacobs (2005).

In addition, we specify a general earnings function. In particular, gross earnings z_n are a function $\Phi(.)$ of ability n, education e_n, and labor effort l_n:

$$z_n \equiv \Phi(n, l_n, e_n), \quad \Phi_n > 0, \quad \Phi_l > 0, \quad \Phi_{ll} \leq 0, \quad \Phi_e > 0, \quad \Phi_{ee} < 0, \quad \Phi_{ne} \geqslant 0, \quad \Phi_{nl} \geqslant 0. \tag{2}$$

Ability, education and labor effort increase earnings. Marginal returns to education diminish with the level of education, which ensures an interior solution for human capital investment. More hours worked raise earnings at a non-increasing rate, i.e. hours worked may not entail constant returns in earnings as in Mirrlees (1971). Ability is complementary to both education and work effort: more able workers feature a (weakly) higher marginal return to both work and education effort. These latter restrictions ensure single

capital, see for example, Jones et al. (1993, 1997), Milesi-Feretti and Rubini (1998), Judd (1999), and Hendricks (1999).

crossing of the utility functions under non-linear policies.[4] No prior restrictions are imposed on the cross derivative Φ_{el}. Factor prices are given. We thus abstract from general equilibrium effects on returns to work effort and education.

3 Optimal non-linear policies

This section analyzes non-linear policy instruments. The government can verify both gross labor incomes and educational expenditures at the individual level. Accordingly, the government can levy a non-linear income tax $T(z_n)$ on gross incomes $z_n \equiv \Phi(n, l_n, e_n)$. The marginal income tax rate is $T'(z_n) \equiv dT(z_n)/dz_n$. Furthermore, the government employs a non-linear subsidy on resources e_n invested in education. The subsidy is denoted as $S(e_n)$, where $S'(e_n) \equiv dS(e_n)/de_n$ represents the marginal subsidy rate on e_n.

Education requires only resources and the unit cost of education is normalized to one for notational convenience. It does not matter whether education requires only resources or also forgone labor time as long as both time and resources invested in education are verifiable, and can therefore be subsidized (see Bovenberg and Jacobs, 2005). The household budget constraint can thus be written as

$$c_n = \Phi(n, l_n, e_n) - T\left(\Phi(n, l_n, e_n)\right) - e_n + S(e_n). \tag{3}$$

Utility maximization yields the first-order conditions for the optimal choices of educational investment and labor supply

$$\left(1 - T'\left(.\right)\right)\Phi_e(n, l_n, e_n) = 1 - S'\left(.\right), \tag{4}$$

$$\frac{-u_l(c_n, l_n)}{u_c(c_n, l_n)} = (1 - T'\left(.\right))\Phi_l(n, l_n, e_n). \tag{5}$$

Expression (4) reveals that the net marginal returns to education (the left-hand side) should be equal to net marginal costs (the right-hand side); taxes reduce net returns while subsidies reduce costs. Equation (5) indicates that the marginal rate of substitution in utility between leisure and consumption should equal the net real wage, which is reduced by a larger marginal tax rate on earnings.

Incentive compatibility requires that each individual n prefers the bundle c_n, z_n, e_n over the bundles intended for all other individuals m:

$$U(c_n, z_n, e_n, n) \geq U(c_m, z_m, e_m, n), \quad \forall m \in [\underline{n}, \overline{n}], \forall n \in [\underline{n}, \overline{n}], \tag{6}$$

where $U(c_n, z_n, e_n, n) \equiv u\left(c_n, \vartheta(n, z_n, e_n)\right) = u(c_n, l_n)$. The function $l_n \equiv \vartheta(n, z_n, e_n)$ is derived by inverting the gross earnings function $z_n \equiv \Phi(n, l_n, e_n)$, so that its derivatives are given by $\vartheta_n = -\frac{\Phi_n}{\Phi_l} < 0$, $\vartheta_z = \frac{1}{\Phi_l} > 0$, and $\vartheta_e = -\frac{\Phi_e}{\Phi_l} < 0$.

These global incentive-compatibility constraints can be replaced by the (first-order) incentive-compatibility constraint (see, e.g., Mirrlees, 1971)

$$\frac{du_n}{dn} = u_l(c_n, l_n)\vartheta_n(n, l_n, e_n) = -u_l(c_n, l_n)\frac{\Phi_n(n, l_n, e_n)}{\Phi_l(n, l_n, e_n)}. \tag{7}$$

[4] Single crossing and monotonicity (in the skill level) of gross earnings, consumption and education are required for second-order incentive compatibility of the non-linear tax and subsidy schedules.

The government maximizes the following social welfare function, which is concave in individual utilities:

$$\int_{\underline{n}}^{\overline{n}} \Psi(u\left(c_n, l_n\right))dF(n), \quad \Psi' > 0, \quad \Psi'' \leq 0, \tag{8}$$

subject to the economy's resource constraint[5]

$$\int_{\underline{n}}^{\overline{n}} \left(\Phi(n, l_n, e_n) - e_n - c_n\right) dF(n) = E, \tag{9}$$

where E represents the exogenous government revenue requirement.

The Hamiltonian $\mathcal{H}$ for maximization of social welfare is given by[6]

$$\max_{\{l_n, e_n, u_n\}} \mathcal{H} = \Psi(u_n)f(n) + \theta_n u_l(c_n, l_n)\frac{\Phi_n(n, l_n, e_n)}{\Phi_l(n, l_n, e_n)} + \lambda\left(\Phi(n, l_n, e_n) - e_n - c_n - E\right) f(n), \tag{10}$$

where θ_n denotes the costate variable for the incentive-compatibility constraint (7). λ stands for the shadow value of the resource constraint (9).[7]

The optimal net tax on education – when the income tax is optimally set – follows from the first-order condition for e_n and is given by (see Appendix)

$$\frac{(T' - S')}{(1 - T')(1 - S')} = \frac{u_c \theta_n / \lambda}{n f(n)} \frac{\omega_n}{\omega_e} \frac{\partial \ln\left(\Phi_n / \Phi_l\right)}{\partial \ln e_n} = \frac{u_c \theta_n / \lambda}{n f(n)} \omega_n \left(\rho_{ne} - \rho_{le}\right), \tag{11}$$

where $\omega_n \equiv \frac{\Phi_n n}{\Phi}$ and $\omega_e \equiv \frac{\Phi_e e}{\Phi}$ denote the shares in gross earnings of, respectively, ability and education. $\rho_{ne} \equiv \frac{\Phi_{ne}\Phi}{\Phi_n \Phi_e}$ represents Hicks' (1963, 1970) partial elasticity of complementarity between ability and education. The partial elasticity of complementarity measures the extent to which ability and education are gross complements in generating earnings (Bertoletti, 2005). Similarly, $\rho_{le} \equiv \frac{\Phi_{le}\Phi}{\Phi_l \Phi_e}$ stands for Hicks' partial elasticity of complementarity between labor and education.[8] θ_n/λ denotes the marginal value – expressed in monetary units – of redistributing one unit of income from individuals with ability larger than n to individuals with ability smaller than n. The more valuable this redistribution is, the higher will be the net tax (or subsidy) on education (*ceteris paribus*).[9]

If education and labor effort are separable in the earnings function (i.e., $\Phi_{le} = 0$, so that $\rho_{le} = 0$), education should be taxed on a net basis for redistributive reasons as long as education and ability are complementary so that high ability agents exhibit a higher productivity in learning than low ability agents do (i.e., $\Phi_{ne} > 0$, so that $\rho_{ne} > 0$). If Φ_{ne} is larger, then investments in education result in more substantial rents from ability, and optimal net taxes on education are larger in order to combat inequality (*ceteris paribus*).

---

[5]If all individuals respect their budget constraints, and the economy's resource constraint is met, the government budget constraint is automatically satisfied by Walras' law.

[6]We assume that the first-order approach is valid and that no bunching occurs due to either binding non-negativity constraints or the violation of monotonicity conditions.

[7]The transversality conditions are given by $\lim_{n \to \overline{n}} \theta_n = 0$, $\lim_{n \to \underline{n}} \theta_n = 0$.

[8]For classical contributions on the elasticity of complementarity and how it relates to the elasticity of substitution, see Hicks (1963, 1970) and Samuelson (1947, 1973). More recent contributions include Broer (2004), Bertoletti (2005), and Blackorby et al. (2007).

[9]The optimal non-linear income tax at optimal non-linear education subsidies and the expression for the marginal value of redistribution θ_n/λ are derived in the appendix. They are virtually the same as in the optimal tax literature. We refer the reader to Mirrlees (1971), Seade (1977), Atkinson and Stiglitz (1976), and Bovenberg and Jacobs (2005) for the interpretation.

If education is completely separable from ability (i.e., $\Phi_{ne} = 0$, so that $\rho_{ne} = 0$), education should be subsidized on a net basis as long as education and labor effort are complementary (i.e., $\Phi_{le} > 0$ and $\rho_{le} > 0$). A subsidy on education then acts as an implicit tax on leisure because a higher level of learning makes leisure less attractive. Hence, an education subsidy offsets the distortionary impact of a redistributive labor tax on leisure demand. If higher levels of education would result in lower labor effort (i.e., $\Phi_{le} < 0$ and $\rho_{le} < 0$), in contrast, education should be optimally taxed, so as to impose an implicit tax on leisure. Empirical evidence suggests that education and labor effort are complementary (so that $\Phi_{le} > 0$ and $\rho_{le} > 0$) because better skilled workers typically feature higher participation rates, work more, and retire later than low-skilled agents do. Hence, education should be subsidized if education and ability are separable in human-capital formation (i.e., $\Phi_{ne} = 0$). In the general case in which ρ_{ne} and ρ_{nl} take on arbitrary values, net education subsidies can be either positive or negative. Positive net subsidies on education are optimal if the efficiency gains of education subsidies brought about by boosting labor supply dominate the regressive distributional impact of education subsidies, and vice versa.

For education policies not to be employed in an optimal redistributive program, the incentive compatibility constraint (see equation (7)) reveals that $\frac{\Phi_n(n,l_n,e_n)}{\Phi_l(n,l_n,e_n)}$ should not depend on education so that $\frac{\partial \ln(\Phi_n/\Phi_l)}{\partial \ln e_n} = 0$. This condition implies that the earnings function should have the following weakly separable form

$$\Phi(n, l_n, e_n) \equiv \phi(\psi(n, l_n), e_n). \tag{12}$$

With this earnings function, the partial elasticities of complementarity are $\rho_{ne} = \rho_{le} = \frac{\Phi_{\psi e}\Phi}{\Phi_\psi \Phi_e}$ and education policies do not relax the incentive compatibility constraint because the benefits of education subsidies in terms of fewer labor supply distortions exactly offset the distributional losses on account of the regressive incidence of education subsidies. Bovenberg and Jacobs (2005) assume a weakly separable earnings function $z_n \equiv n l_n \phi(e_n)$. With this earnings function, $\rho_{ne} = \rho_{le} = 1$, so that the efficiency gains from education subsidies on labor supply exactly offset the equity losses due to regressive education subsidies.[10]

Maldonado (2007) analyzes a special case of our general earnings function in which education and ability are weakly separable from labor effort: $z_n \equiv \phi(n, e_n) l_n$. With this particular earnings function, we have $\rho_{ne} = \frac{\phi_{ne}\phi}{\phi_n \phi_e}$ and $\rho_{le} = 1$. Consequently, education is taxed (subsidized) on a net basis if $\frac{\phi_{ne}\phi}{\phi_n\phi_e} > (<) 1$.

4 Optimal linear policies

This section derives optimal linear policies. Informational requirements are thus less stringent. In particular, the government needs to be able to verify only aggregate labor

[10] The result of a zero net tax on human capital is in the same spirit as the optimal zero commodity result by Atkinson and Stiglitz (1976). From a purely technical point of view, both results require some form of weak separability: either in the earnings function (between education and the other arguments) or in the utility function (between labor and consumption goods). However, our finding stresses the trade-off between the direct gain of net education taxes in reducing inequality across skills and the loss of these taxes in exacerbating tax distortions on labor. The intuition behind Atkinson and Stiglitz (1976), in contrast, relies primarily on the impact of commodity taxes on the labor-tax distortion rather than the direct effect on inequality across various skill levels.

income and aggregate investment in education. The household problem is the same as in the previous section except that the lump-sum transfer g enters the household budget constraint and the linear tax t and the linear subsidy s replace their non-linear counterparts in the first-order conditions.

The government budget constraint is given by

$$t\int_{\underline{n}}^{\overline{n}}\Phi(n,l_n,e_n)dF(n)=s\int_{\underline{n}}^{\overline{n}}e_n dF(n)+g+E. \tag{13}$$

The government maximizes a concave sum of individuals' indirect utility functions, which are denoted by $\upsilon(g,t,s,n)$. Roy's lemma yields $\frac{\partial \upsilon}{\partial g}=\eta_n$, $\frac{\partial \upsilon}{\partial t}=-\eta_n\Phi(.)$, and $\frac{\partial \upsilon}{\partial s}=\eta_n e_n$, where η_n stands for private marginal utility of income of skill n. We define the social marginal value of income of skill n (including the income effects on the tax base) as

$$b_n\equiv\frac{\Psi'(\upsilon(.))\eta_n}{\lambda}+\left(\frac{t-s}{1-t}\right)\frac{\partial e_n}{\partial g}+t\Phi_l(.)\frac{\partial l_n}{\partial g}, \tag{14}$$

where λ represents again the shadow value of public resources, $\frac{t-s}{1-t}$ stands for the marginal tax wedge on human capital investment (cf. Bovenberg and Jacobs, 2005), and, similarly, $t\Phi_l(.)$ is the marginal tax wedge on labor effort.

The Lagrangian for maximizing social welfare amounts to

$$\max_{\{g,t,s\}}\mathcal{L}=\int_{\underline{n}}^{\overline{n}}\Psi(\upsilon(g,t,s,n))\,dF(n)+\lambda\int_{\underline{n}}^{\overline{n}}(t\Phi(n,l_n,e_n)-se_n-g-E)\,dF(n). \tag{15}$$

Using the definition of b_n and Roy's lemma, we can write the first-order condition for maximizing social welfare with respect to g as

$$\bar{b}\equiv\int_{\underline{n}}^{\overline{n}}b_n dF(n)=1. \tag{16}$$

The marginal social benefits of a higher lump-sum transfer (the left-hand side) should equal the marginal social cost of a higher lump-sum transfer (right-hand side).

In order to facilitate the discussion on the optimal tax and subsidy rates, we introduce the distributional characteristic ξ of labor income, which is given by the normalized covariance between gross earnings z_n and the social welfare weights b_n:

$$\xi\equiv-\frac{\int_{\underline{n}}^{\overline{n}}b_n z_n dF(n)-\int_{\underline{n}}^{\overline{n}}z_n dF(n)\int_{\underline{n}}^{\overline{n}}b_n dF(n)}{\int_{\underline{n}}^{\overline{n}}z_n dF(n)\int_{\underline{n}}^{\overline{n}}b_n dF(n)}=\frac{\int_{\underline{n}}^{\overline{n}}(1-b_n)z_n dF(n)}{\int_{\underline{n}}^{\overline{n}}z_n dF(n)}. \tag{17}$$

Similarly, the distributional characteristic of education ζ is given by

$$\zeta\equiv-\frac{\int_{\underline{n}}^{\overline{n}}b_n e_n dF(n)-\int_{\underline{n}}^{\overline{n}}e_n dF(n)\int_{\underline{n}}^{\overline{n}}b_n dF(n)}{\int_{\underline{n}}^{\overline{n}}e_n dF(n)\int_{\underline{n}}^{\overline{n}}b_n dF(n)}=\frac{\int_{\underline{n}}^{\overline{n}}(1-b_n)e_n dF(n)}{\int_{\underline{n}}^{\overline{n}}e_n dF(n)}. \tag{18}$$

The second equality of both expressions is valid only at the optimized lump-sum transfer so that (16) holds. A larger ξ (ζ) implies that the government wishes to use taxes on earnings (education) for redistributional reasons, due to the negative correlation between welfare weights and incomes (education). ξ (ζ) is zero if z_n (e_n) is the same for all agents so that inequality is absent or if the government does not care about inequality and thus attaches the same social welfare weight b_n to all skills.

If the government sets the optimal education subsidy s alongside the optimal labor-income tax t, we obtain the following expression for the net optimal net tax on education (see appendix)

$$\frac{(t-s)}{(1-s)(1-t)} = \xi\left(\frac{\overline{\omega_e}\ \overline{\varepsilon_{lt}}}{\overline{\varepsilon_{es}}\ \overline{\varepsilon_{lt}} - \overline{\varepsilon_{et}}\ \overline{\varepsilon_{ls}}}\right)\left(\frac{\zeta}{\xi} - \frac{\overline{\varepsilon_{ls}}}{\overline{\omega_e}\ \overline{\varepsilon_{lt}}}\right), \tag{19}$$

The compensated elasticities are defined as $\varepsilon_{et} \equiv -\omega_e \frac{\partial e_n}{\partial t}\frac{1-t}{e_n}$, $\varepsilon_{lt} \equiv -\omega_l \frac{\partial l_n}{\partial t}\frac{1-t}{l_n}$, $\varepsilon_{es} \equiv \omega_e \frac{\partial e_n}{\partial s}\frac{1-s}{e_n}$, $\varepsilon_{ls} \equiv \omega_l \frac{\partial l_n}{\partial s}\frac{1-s}{l_n}$. To save on notation, we have pre-multiplied the standard elasticities with the earnings shares of labor and education (i.e. $\omega_l \equiv \frac{\Phi_l l_n}{\Phi}$, and $\omega_e \equiv \frac{\Phi_e e_n}{\Phi}$). A bar denotes an income-weighted average of a compensated elasticity (i.e. $\overline{\varepsilon_{xi}} \equiv \left(\int_{\underline{n}}^{\overline{n}} \varepsilon_{xi} z_n dF(n)\right)\left(\int_{\underline{n}}^{\overline{n}} z_n dF(n)\right)^{-1}$ for $x = l, e$, $i = t, s$). Similarly, $\overline{\omega_e} \equiv \left(\int_{\underline{n}}^{\overline{n}} \omega_e z_n dF(n)\right)\left(\int_{\underline{n}}^{\overline{n}} z_n dF(n)\right)^{-1}$ is the income-weighted average earnings share of education.

ξ is the analogue of the marginal value of redistribution θ_n in the expression for the non-linear income tax (11). Furthermore, ζ/ξ is closely related to the Hicks partial elasticity of complementarity between education and skill ρ_{ne}. In particular, ζ/ξ measures the distributional advantage of education taxes over income taxes. Similarly, $\frac{\overline{\varepsilon_{ls}}}{\overline{\omega_e}\ \overline{\varepsilon_{lt}}}$ corresponds to the ρ_{le} term in the expression for the optimal non-linear income tax. This term measures the benefits of education subsidies to lower the tax wedge on labor supply.

We can derive the elasticities in terms of the compensated wage elasticity of labor supply and the properties of the earnings function and find that $\varepsilon_{lt} = (\rho_{ee} + \rho_{el})/\Delta > 0$, $\varepsilon_{et} = \left(\frac{1}{\varepsilon\omega_l} + \rho_{ll} + \rho_{el}\right)/\Delta > 0$, $\varepsilon_{ls} = \rho_{el}/\Delta \geq 0$, and $\varepsilon_{es} = \left(\frac{1}{\varepsilon\omega_l} + \rho_{ll}\right)/\Delta > 0$ (see the appendix). $\rho_{ll} \equiv -\frac{\Phi_{ll}\Phi}{\Phi_l^2} > 0$, and $\rho_{ee} \equiv -\frac{\Phi_{ee}\Phi}{\Phi_e^2} > 0$ measure the concavity of the earnings function with respect to labor and education, respectively. The second-order condition of individual optimization implies that the denominator of all elasticities is positive: i.e., $\Delta \equiv \rho_{ee}\left(\frac{1}{\varepsilon\omega_l} + \rho_{ll}\right) - \rho_{el}^2 > 0$. $\varepsilon \equiv \frac{\partial l_n}{\partial w_n}\frac{w_n}{l_n}$ stands for the compensated wage elasticity of labor supply, where $w_n \equiv (1-t)\Phi_l(n, l_n, e_n)$ represents the net marginal wage rate.

Substitution of the elasticities in the expression for the optimal education tax yields

$$\frac{(t-s)}{(1-s)(1-t)} = \xi\overline{\left(\frac{\rho_{ee}+\rho_{el}}{\Delta}\right)}\kappa\overline{\omega_e}\left(\frac{\zeta}{\xi} - \frac{1}{\overline{\omega_e}}\frac{\overline{\left(\frac{\rho_{el}}{\Delta}\right)}}{\overline{\left(\frac{\rho_{ee}+\rho_{el}}{\Delta}\right)}}\right), \tag{20}$$

where $\kappa^{-1} \equiv \overline{\left(\frac{\frac{1}{\varepsilon\omega_l}+\rho_{ll}}{\Delta}\right)}\ \overline{\left(\frac{\rho_{ee}}{\Delta}\right)} - \overline{\left(\frac{\rho_{el}}{\Delta}\right)}^2 > 0$ is also positive. Note that $\kappa \approx 1$, if the covariances of the elasticities with income are small and individual elasticities correspond closely to their aggregate counterparts, since $\Delta \equiv \rho_{ee}\left(\frac{1}{\varepsilon\omega_l} + \rho_{ll}\right) - \rho_{el}^2$.

If education and labor are not complementary ($\rho_{el} = 0$), the cross-elasticity of labor supply with respect to the education subsidy is zero ($\varepsilon_{ls} = 0$). Education should then be taxed for redistributive reasons if an ability bias in education implies that high ability agents learn more (i.e. $\zeta > 0$), so that net taxes on education yield distributional benefits, cf.

$$\left.\frac{(t-s)}{(1-s)(1-t)}\right|_{\rho_{el}=0} = \frac{\zeta\overline{\omega_e}}{\overline{\varepsilon_{es}}} > 0. \tag{21}$$

The government should employ taxes on education more aggressively if education responds less elastically to the tax on education (so that $\bar{\varepsilon}_{es}$ is smaller)– ceteris paribus ζ. This mirrors the findings for the optimal non-linear policies.

Furthermore, if an ability bias in education is absent or the government is not interested in redistribution, the distributional characteristic ζ is zero. In that case, education should be subsidized for efficiency reasons if education and labor effort are complementary (i.e. $\rho_{el} > 0$, so that $\varepsilon_{ls} > 0$):

$$\left.\frac{(t-s)}{(1-s)(1-t)}\right|_{\zeta=0} = -\xi\kappa\overline{\left(\frac{\rho_{el}}{\Delta}\right)} < 0. \tag{22}$$

The government wants to subsidize education in order to boost labor supply, thereby alleviating the tax distortion $t > 0$ associated with redistribution towards those with lower earnings ($\xi > 0$). Hence, also in this case the results are analogous to the case of non-linear policies: the government employs education subsidies to alleviate labor-tax distortions if ability and education are not positively correlated.

If education generates ability rents and is also complementary to labor effort, education taxes yield both efficiency losses and distributional gains. Whether education should be taxed on a net basis at the optimum depends on the attractiveness of education taxes relative to income taxes as a redistributive instrument (as measured by ζ/ξ), versus the effectiveness of education policies to fight labor market distortions as measured by $\frac{1}{\overline{\omega_e}}\overline{\left(\frac{\rho_{el}}{\Delta}\right)}\;\overline{\left(\frac{\rho_{ee}+\rho_{el}}{\Delta}\right)}^{-1}$. At high levels of ζ/ξ, net taxes on education are attractive to combat inequality. If the elasticity of complementarity between learning and working ρ_{el} is large, however, education should be taxed relatively lightly (subsidized heavily). Intuitively, in the presence of a large elasticity of complementarity ρ_{el}, learning indirectly boosts labor supply, thereby alleviating labor tax distortions on labor supply.

The condition $\frac{\zeta}{\xi} \gtrless \frac{1}{\overline{\omega_e}}\overline{\left(\frac{\rho_{el}}{\Delta}\right)}\;\overline{\left(\frac{\rho_{ee}+\rho_{el}}{\Delta}\right)}^{-1}$ determines whether education should be taxed or subsidized on a net basis. If $\frac{\zeta}{\xi} = \frac{1}{\overline{\omega_e}}\overline{\left(\frac{\rho_{el}}{\Delta}\right)}\;\overline{\left(\frac{\rho_{ee}+\rho_{el}}{\Delta}\right)}^{-1}$, we obtain a zero net tax on human capital investments under optimal linear policies. For the class of weakly separable earnings functions ($\rho_{el} = \rho_{en}$), we can derive a sufficient condition for which this requirement is met. First, a constant elasticity of education in the earnings function ensures that $\omega_e = \frac{1}{1+\rho_{ee}}$, where the elasticity ρ_{ee} is constant across agents. Moreover, it implies that $\zeta/\xi = 1$ because education and learning are linearly related. Finally, it yields $\rho_{el} = 1$, since $\omega_e \equiv \frac{\Phi_e e_n}{\Phi}$ is constant (i.e., $\frac{\partial \omega_e}{\partial l_n} = \frac{e_n \Phi_e \Phi_l}{\Phi^2}(\rho_{le} - 1) = 0$). Therefore, the earnings function should be of the following weakly separable form:

$$\Phi(n, l_n, e_n) \equiv \psi(n, l_n) e_n^{\beta}. \tag{23}$$

Bovenberg and Jacobs (2005) assumed that the earnings function was $z_n \equiv n l_n e_n^{\beta}$. This specification exactly meets the requirements for a zero net tax on education. The intuition for their result is that labor earnings and education are related in a linear fashion across different ability levels if the earnings function is weakly separable and features a constant elasticity in education (cf. the first-order condition for education (4)). Compared to labor income taxes, education taxes therefore imply both the same distortions on labor supply (i.e. $\overline{\varepsilon_{lt}} = \frac{\overline{\varepsilon_{ls}}}{\overline{\omega_e}}$) and the same effects on the income distribution (i.e. $\zeta = \xi$). In contrast to labor taxes, however, taxes on education distort the education decision. Consequently, the government does not employ net taxes on education and only adopts a positive labor tax. The labor tax yields the same distributional benefits and imposes the same tax-distortions on labor as the education tax, but avoids distortions in skill formation.[11]

5 Conclusions

This paper has generalized Bovenberg and Jacobs' (2005) models of optimal linear and non-linear taxes and education subsidies in models of labor supply and human capital formation. Using general earnings functions, we show that education decisions are generally not efficient. Efficiency in human capital formation is obtained only under restrictive conditions. In particular, with non-linear policy instruments, the earnings function should be weakly separable in ability and labor, on the one hand, and in education, on the other hand. With linear policy instruments, a weakly separable earnings function should in addition feature a constant elasticity in education to arrive at a zero net tax on human capital. The analysis of linear and non-linear tax and education policies reveals that (net) subsidies on education are optimal only if sufficiently large efficiency gains of lower labor supply distortions due to complementarities between learning and working offset the regressive incidence of education subsidies.

Bovenberg and Jacobs (2005) argued that their efficiency results are due to the Diamond and Mirrlees (1971) production efficiency theorem applied to individual production functions for human capital. This interpretation can be amended if more general earnings functions are employed. Indeed, the applicability of the production efficiency theorem to human capital production crucially depends on the presence of a non-distorting profit tax to skim off the quasi-rents from ability in human capital returns. These rents arise due to diminishing returns in human capital formation. A perfect profit tax is available with the earnings functions used by Bovenberg and Jacobs (2005). By combining a labor income tax with an education subsidy, the government can perfectly extract the rents from ability in human capital production, without directly interfering with the consumption choices of households. However, this optimal education-tax policy package fails to constitute a perfect profit tax on education rents from ability in the presence of more general earnings functions. In fact, in the absence of weak separability, the government cannot tax away the inframarginal ability rents in human capital production at zero efficiency costs because consumption and investment choices are not separable. Consequently, the production efficiency theorem breaks down, and education may be taxed or subsidized for redistributive reasons, depending on the degree of complementarity of education and labor effort and the incidence of education subsidies.

References

Atkinson, Anthony B., and Joseph E. Stiglitz (1976), "The Design of Tax Structure: Direct versus Indirect Taxation", *Journal of Public Economics,* 6, 55–75.

[11]Again, one could relate these findings to Corlett and Hague (1953) on the optimality of differentiated linear commodity taxes. Just as in the discussion of non-linear policies, the analogy is merely technical; homotheticity is required in the earnings function to make linear net education taxes zero, just like homothetic (and separable) utility would be required to find zero linear commodity taxes. However, the economic mechanism for our result and that of Corlett and Hague (1953) is different; see previous footnote.

Bertoletti, Paolo (2005), "Elasticities of Substitution and Complementarity: A Synthesis", *Journal of Productivity Analysis*, 24, (2), 183-196.

Blackorby, Charles, Daniel Primont, and R. Robert Russell (2007), The Morishima Gross Elasticity of Substitution, *Journal of Productivity Analysis*, 28, (3), 203-208.

Blumkin, Tomer, and Efraim Sadka (2008), "A Case for Taxing Education", *International Tax and Public Finance*, 15 (2), 145-163.

Blundell, Richard and Thomas MaCurdy (1999), "Labor Supply: A Review of Alternative Approaches", in: Orley Ashenfelter and David Card (eds.), *Handbook of Labor Economics – Vol 3A*, Amsterdam: Elsevier-North Holland, 1559-1695.

Bohacek, Radim, and Marek Kapicka (2007), "Optimal Human Capital Policies", mimeo.

Bovenberg, A. Lans, and Bas Jacobs (2005), "Redistribution and Education Subsidies are Siamese Twins", *Journal of Public Economics*, 89, 2005–2035.

Brett, Craig, and John A. Weymark (2003), "Financing Education using Optimal Redistributive Taxation", *Journal of Public Economics*, 87, (11), 2549-2569.

Broer, Peter (2004), "The Elasticities of Complementarity and Substitution", Tinbergen Institute Discussion Paper 101/1.

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.

Corlett Wilfred J., and Douglas C. Hague (1953),"Complementarity and the Excess Burden of Taxation", *Review of Economic Studies*, 21, (1), 21-30.

Da Costa, Carlos E., and Lucas J. Maestri (2007), "The Risk-Properties of Human Capital and the Design of Government Policies", *European Economic Review*, 51, (3), 695-713.

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

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

Grochulski, Borys, and Tomasz Piskorski, (2006), "Risky Human Capital and Deferred Capital Income Taxation", Federal Reserve Bank of Richmond Working Paper No. 06-13.

Hare, Paul G., and David T. Ulph (1979), "On Education and Distribution", *Journal of Political Economy*, 87 (5), S193-S212.

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

Hendricks, Lutz (1999), "Taxation and Long-Run Growth", *Journal of Monetary Economics*, 43, 411-434.

Hicks, John R. (1963), *The Theory of Wages*, London: Macmillan, 2nd edition.

Hicks, John R. (1970), "Elasticity of Substitution Again: Substitutes and Complements", *Oxford Economic Papers*, 25, 289-296.

Jacobs, Bas (2005), "Optimal Income Taxation with Endogenous Human Capital", *Journal of Public Economic Theory*, 7, (2), 295-315.

Jacobs, Bas (2007), "Optimal Redistributive Tax and Education Policies in General Equilibrium", CESifo Working Paper No. 2162.

Jacobs, Bas, and A. Lans Bovenberg (2005), "Human Capital and Optimal Positive Taxation of Capital Income", CEPR Working Paper No. 5047.

Jones, Larry E., Rodolfo E. Manuelli, and Peter E. Rossi (1993), "Optimal Taxation in Models of Endogenous Technological Change", *Journal of Political Economy*, 101 (3), 485-517.

Jones, Larry E., Rodolfo E. Manuelli, and Peter E. Rossi (1997), "On the Optimal Taxation of Capital Income", *Journal of Economic Theory*, 73, 93-117.

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

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

Maldonado, Darío (2007), "Education Policies and Optimal Taxation", *International Tax and Public Finance*, 15 (2), 131-143.

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

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

OECD (2006), *OECD Labor Force Statistics Database*, Paris: OECD.

Richter, Wolfram F. (2006), "Taxing Human Capital Efficiently: The Double Dividend of Taxing Non-qualified Labour more Heavily than Qualified Labour", CESifo Working Paper No. 1832.

Samuelson, Paul A. (1947), *Foundations of Economic Analysis*, Cambridge-MA: Harvard University Press.

Samuelson, Paul A. (1973), "Complementarity – An Essay on the 40th Anniversary of the Hicks-Allen Revolution in Demand Theory", *Journal of Economic Literature*, 12, (4), 1255-1289.

Seade, Jesus K. (1977), "On the Shape of Optimal Tax Schedules", *Journal of Public Economics*, 7, 203-235.

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.

Wigger, Berthold (2004), "Are Higher Education Subsidies Second Best?", *Scandinavian Journal of Economics*, 106, (1), 65-82.

Appendix

Optimal non-linear policies

We solve for the optimal allocation by applying the maximum principle and setting up a Hamiltonian $\mathcal{H}$, with l_n and e_n as control variables, u_n as state variable, and θ_n as costate variable for the incentive-compatibility constraint (7):

$$\max_{\{l_n,e_n,u_n\}} \mathcal{H} = \Psi(u_n)f(n) + \theta_n u_l(c_n,l_n)\Xi(n,l_n,e_n) + \lambda\left(\Phi(n,l_n,e_n) - e_n - c_n - E\right)f(n), \tag{24}$$

where $\Xi(n,l_n,e_n) \equiv \frac{\Phi_n(n,l_n,e_n)}{\Phi_l(n,l_n,e_n)}$ and λ represents the shadow value of the resource constraint. The transversality conditions are given by

$$\lim_{n\to\bar{n}} \theta_n = 0, \quad \lim_{n\to\underline{n}} \theta_n = 0. \tag{25}$$

The first-order condition for e_n is given by

$$\frac{\partial\mathcal{H}}{\partial e_n} = \lambda\left(\Phi_e - 1 - \left.\frac{dc_n}{de_n}\right|_{\bar{u},\bar{l}}\right)f(n) + \theta_n\Xi u_{lc}\left.\frac{dc_n}{de_n}\right|_{\bar{u},\bar{l}} + \theta_n u_l\Xi_e = 0. \tag{26}$$

For the indirect impacts on consumption, we find $\left.\frac{dc_n}{de_n}\right|_{\bar{u},\bar{l}} = 0$ by differentiating the household budget constraint and substituting the individuals' first-order condition for e_n. Substitution of $\left.\frac{dc_n}{de_n}\right|_{\bar{u},\bar{l}} = 0$, the first-order condition for learning $1 - S' = (1-T')\Phi_e$, the first-order condition for labor supply $-u_l = (1-T')\Phi_l u_c$ and $\Xi_e = \frac{\Phi_{ne}\Phi_l - \Phi_{le}\Phi_n}{(\Phi_l)^2}$ gives

$$\frac{(T'-S')}{(1-T')(1-S')} = \frac{\theta_n u_c/\lambda}{nf(n)}\frac{n\Phi_l}{\Phi_e}\left(\frac{\Phi_{ne}\Phi_l - \Phi_{le}\Phi_n}{(\Phi_l)^2}\right), \tag{27}$$

which yields (11).

The first-order condition for l_n is given by

$$\frac{\partial\mathcal{H}}{\partial l_n} = \lambda\left(\Phi_l - \left.\frac{dc_n}{dl_n}\right|_{\bar{u},\bar{e}}\right)f(n) + \theta_n\Xi\left(u_{ll} + u_{lc}\left.\frac{dc_n}{dl_n}\right|_{\bar{u},\bar{e}}\right) + \theta_n u_l\Xi_l = 0, \tag{28}$$

next we substitute $\left.\frac{dc_n}{dl_n}\right|_{\bar{u},\bar{e}} = (1-T')\Phi_l(n,l_n,e_n)$ (found by taking the total derivative of utility $u(c_n,l_n)$ and substituting the first-order condition for labor supply to eliminate u_l), and $\Xi_l = \frac{\Phi_{nl}\Phi_l - \Phi_{ll}\Phi_n}{(\Phi_l)^2}$ to arrive at

$$\frac{T'}{1-T'} = \frac{\theta_n u_c/\lambda}{f(n)}\left(\frac{\Phi_{nl}\Phi_l - \Phi_{ll}\Phi_n}{\Phi_l^2} + \frac{\Phi_n}{\Phi_l}\left(\frac{u_{ll}}{u_l} - \frac{u_{lc}}{u_c}\right)\right). \tag{29}$$

Rearranging gives

$$\frac{T'}{1-T'} = \frac{u_c\theta_n/\lambda}{nf(n)}\omega_n\left(\rho_{nl}+\frac{1}{\omega_l}\frac{1}{\varepsilon^*}\right), \tag{30}$$

where $\rho_{nl} \equiv \frac{\Phi_{nl}\Phi}{\Phi_n\Phi_l}$ is Hicks' partial elasticity of complementarity between ability and work effort in earnings. $\varepsilon^* \equiv \left(\frac{u_{ll}l_n}{u_l} - \frac{u_{lc}l_n}{u_c} - \frac{\Phi_{ll}l_n}{\Phi_l}\right)^{-1} > 0$ is a measure for the compensated wage elasticity of labor supply, which depends on the the curvature of both the utility function and the earnings function. As before, we also find here that marginal taxes increase if ability rents increase with labor effort (ρ_{nl} is higher). If the earnings function is linear in ability and labor ($\omega_n = \omega_l = \rho_{nl} = 1$), the expression found by Mirrlees (1971) results.[12]

The first-order condition for u_n is

$$\frac{\partial\mathcal{H}}{\partial u_n} = \left(\Psi'(u_n) - \lambda\left.\frac{dc_n}{du_n}\right|_{\bar{l},\bar{e}}\right)f(n) + \theta_n\Xi u_{lc}\left.\frac{dc_n}{du_n}\right|_{\bar{l},\bar{e}} = \frac{d\theta_n}{dn}. \tag{31}$$

We defined θ_n negatively; hence, there is no minus sign on the right-hand side. Substitution of $\left.\frac{dc_n}{du_n}\right|_{\bar{l},\bar{e}} = \frac{1}{u_c}$ and some rearranging yield a first-order differential equation in θ_n

$$\frac{d\theta_n}{dn} + \delta_n\theta_n = \kappa_n, \tag{32}$$

where $\delta_n \equiv -\frac{\Phi_n}{\Phi_l}\frac{u_{lc}}{u_c}$ and $\kappa_n \equiv \left(\Psi'(u_n) - \frac{\lambda}{u_c}\right)f(n)$. This equation can be solved analytically to find

$$\frac{\theta_n}{\lambda} = \int_n^{\overline{n}}\left(\frac{1}{u_c(.)} - \frac{\Psi'(u_m)}{\lambda}\right)\exp\left(\int_m^n -\frac{\Phi_s(.)}{\Phi_l(.)}\frac{u_{lc}(.)}{u_c(.)}ds\right)f(m)dm. \tag{33}$$

See Mirrlees (1971), Seade (1977), Atkinson and Stiglitz (1976), and Bovenberg and Jacobs (2005) for the interpretation.

Optimal linear policies

The Lagrangian for maximizing social welfare is designated by

$$\max_{\{g,t,s\}}\mathcal{L} = \int_{\underline{n}}^{\overline{n}}\Psi(v(g,t,s,n))\,dF(n) + \lambda\int_{\underline{n}}^{\overline{n}}(t\Phi(n,l_n,e_n) - se_n - g - E)\,dF(n). \tag{34}$$

The first-order conditions for maximization of social welfare (34) with respect to t and s are

$$\frac{\partial\mathcal{L}}{\partial t} = \int_{\underline{n}}^{\overline{n}}\left[(-\Psi'(v_n)\eta_n + \lambda)\Phi(.) + \lambda\left(t\Phi_l(.)\frac{\partial l_n}{\partial t} + \frac{(t-s)}{(1-t)}\frac{\partial e_n}{\partial t}\right)\right]dF(n) = 0, \tag{35}$$

$$\frac{\partial\mathcal{L}}{\partial s} = \int_{\underline{n}}^{\overline{n}}\left[(\Psi'(v_n)\eta_n - \lambda)e_n + \lambda\left(t\Phi_l(.)\frac{\partial l_n}{\partial s} + \frac{(t-s)}{(1-t)}\frac{\partial e_n}{\partial s}\right)\right]dF(n) = 0, \tag{36}$$

[12] We note here that the elasticities of gross income with respect to the marginal tax rates are higher than in the case where human capital formation is exogenous. Optimal marginal income taxes are consequently lower. In order to show this, one needs to write the optimal tax formula in terms of the density of gross earnings. See Bovenberg and Jacobs (2005).

where we used the first-order condition for learning and Roy's lemma.

The first-order condition for t can be simplified upon substitution of the Slutsky equations $\frac{\partial l_n}{\partial t} = \frac{\partial l_n^*}{\partial t} - \Phi(.) \frac{\partial l_n}{\partial g}$ and $\frac{\partial e_n}{\partial t} = \frac{\partial e_n^*}{\partial t} - \Phi(.) \frac{\partial e_n}{\partial g}$ (asterisks denote compensated demands) and the definition of b_n from (14)

$$\int_{\underline{n}}^{\overline{n}} \left[(1 - b_n) \Phi(.) + \frac{t}{1-t} \Phi_l(.) l_n \frac{\partial l_n^*}{\partial t} \frac{1-t}{l_n} + \frac{(t-s)e_n}{(1-t)^2} \frac{\partial e_n^*}{\partial t} \frac{1-t}{e_n} \right] dF(n) = 0. \tag{37}$$

Next, use the first-order condition for learning to establish $\omega_e (1-t) \Phi(n, l_n, e_n) = (1-s)e_n$, where $\omega_e \equiv \frac{\Phi_e e_n}{\Phi}$, and simplify

$$\int_{\underline{n}}^{\overline{n}} \left[(1 - b_n) \Phi(.) - \frac{t}{1-t} \varepsilon_{lt} \Phi(.) + \frac{(t-s)}{(1-t)(1-s)} \varepsilon_{et} \Phi(.) \right] dF(n) = 0, \tag{38}$$

$\varepsilon_{lt} \equiv -\omega_l \frac{\partial l_n^*}{\partial t} \frac{1-t}{l_n} > 0$, and $\varepsilon_{et} \equiv -\omega_e \frac{\partial e_n^*}{\partial t} \frac{1-t}{e_n} > 0$. Next divide by $\int_{\underline{n}}^{\overline{n}} \Phi(n, l_n, e_n) dF(n)$ to get

$$\xi - \frac{t}{1-t} \overline{\varepsilon_{lt}} - \frac{(t-s)}{(1-s)(1-t)} \overline{\varepsilon_{et}} = 0, \tag{39}$$

where a bar denotes a weighted average elasticity with income as weights, i.e.,
$\overline{\varepsilon_{xt}} \equiv \left(\int_{\underline{n}}^{\overline{n}} \varepsilon_{xt} z_n dF(n) \right) \left(\int_{\underline{n}}^{\overline{n}} z_n dF(n) \right)^{-1}$ for $x = l, e$.

Similarly, we can simplify the first-order condition for s upon substitution of the Slutsky equations $\frac{\partial l_n}{\partial s} = \frac{\partial l_n^*}{\partial s} + e_n \frac{\partial l_n}{\partial g}$ and $\frac{\partial e_n}{\partial t} = \frac{\partial e_n^*}{\partial s} + e_n \frac{\partial e_n}{\partial g}$ and the definition of b_n from (14)

$$\int_{\underline{n}}^{\overline{n}} \left[(b_n - 1) e_n + \frac{t}{1-s} \Phi_l(.) l_n \frac{\partial l_n^*}{\partial s} \frac{1-s}{l_n} + \frac{(t-s)e_n}{(1-t)(1-s)} \frac{\partial e_n^*}{\partial s} \frac{1-s}{e_n} \right] dF(n) = 0. \tag{40}$$

Next, use $\omega_e (1-t) \Phi(n, l_n, e_n) = (1-s)e_n$, and simplify

$$\int_{\underline{n}}^{\overline{n}} \left[(b_n - 1) e_n + \frac{t}{1-s} \varepsilon_{ls} \Phi(.) + \frac{(t-s)}{(1-s)^2} \varepsilon_{es} \Phi(.) \right] dF(n) = 0, \tag{41}$$

$\varepsilon_{ls} \equiv \omega_l \frac{\partial l_n^*}{\partial s} \frac{1-s}{l_n} > 0$, and $\varepsilon_{es} \equiv \omega_e \frac{\partial e_n^*}{\partial s} \frac{1-s}{e_n} > 0$. Next divide by $\int_{\underline{n}}^{\overline{n}} \Phi(n, l_n, e_n) dF(n)$ to get

$$-\zeta \frac{\int_{\underline{n}}^{\overline{n}} e_n dF(n)}{\int_{\underline{n}}^{\overline{n}} z_n dF(n)} + \frac{t}{1-s} \overline{\varepsilon_{ls}} + \frac{(t-s)}{(1-s)^2} \overline{\varepsilon_{es}} = 0. \tag{42}$$

Again, use the first-order condition for learning to find

$$\zeta - \frac{t}{1-t} \frac{\overline{\varepsilon_{ls}}}{\overline{\omega_e}} - \frac{(t-s)}{(1-s)(1-t)} \frac{\overline{\varepsilon_{es}}}{\overline{\omega_e}} = 0, \tag{43}$$

where $\overline{\varepsilon_{xs}} \equiv \left(\int_{\underline{n}}^{\overline{n}} \varepsilon_{xs} z_n dF(n) \right) \left(\int_{\underline{n}}^{\overline{n}} z_n dF(n) \right)^{-1}$ $x = l, e$ and
$\overline{\omega_e} \equiv \left(\int_{\underline{n}}^{\overline{n}} \omega_e z_n dF(n) \right) \left(\int_{\underline{n}}^{\overline{n}} z_n dF(n) \right)^{-1}$.

The optimal net tax on education follows by solving the first-order condition for t (39) and the first-order condition for s (43) for $\frac{(t-s)}{(1-s)(1-t)}$ and $\frac{t}{1-t}$. The optimal expression for $\frac{(t-s)}{(1-s)(1-t)}$ is given in the text. The optimal linear income tax is

$$\frac{t}{1-t} = \xi \left(\frac{\overline{\omega_e}\, \overline{\varepsilon_{et}}}{\overline{\varepsilon_{es}}\, \overline{\varepsilon_{lt}} - \overline{\varepsilon_{et}}\, \overline{\varepsilon_{ls}}} \right) \left(\frac{\overline{\varepsilon_{es}}}{\overline{\omega_e}\, \overline{\varepsilon_{et}}} - \frac{\zeta}{\xi} \right). \tag{44}$$

Compensated elasticities

To derive compensated elasticities, we keep utility fixed. Hence, in totally differentiated form, utility can be written as

$$\tilde{c}_n = \frac{\gamma_l}{\gamma_c}\tilde{l}_n, \tag{45}$$

where $\gamma_l \equiv -\frac{u_l l_n}{u} > 0$ and $\gamma_c \equiv \frac{u_c c_n}{u} > 0$ are the shares of labor and labor in utility, respectively. A tilde denotes a log-linear deviation, e.g. $\tilde{c}_n \equiv dc_n/c_n$.The first-order condition for labor supply (see equation (5)) can also be totally differentiated to find

$$(\mu_{ll} - \mu_{cl})\,\tilde{l}_n + (\mu_{cc} - \mu_{lc})\,\tilde{c}_n = -\tilde{t} - \alpha_{ll}\tilde{l}_n + \alpha_{le}\tilde{e}_n, \tag{46}$$

where $\mu_{cc} \equiv -\frac{u_{cc}c_n}{u_c} > 0$, $\mu_{ll} \equiv -\frac{u_{ll}l_n}{u_l} > 0$, $\mu_{cl} \equiv \frac{u_{cl}l_n}{u_c} \gtrless 0$, $\mu_{lc} \equiv \frac{u_l l_n}{u_l} \gtrless 0$, $\alpha_{ll} \equiv -\frac{\Phi_{ll}l_n}{\Phi_l} \geq 0$, $\alpha_{le} \equiv \frac{\Phi_{le}e_n}{\Phi_l} \gtrless 0$, and the linearized tax and subsidy rates are defined as $\tilde{t} \equiv dt/(1-t)$ and $\tilde{s} \equiv ds/(1-s)$.

Using the linearized utility function to substitute out $\tilde{c}$ gives

$$\tilde{l}_n = \varepsilon\left(-\tilde{t} - \alpha_{ll}\tilde{l}_n + \alpha_{le}\tilde{e}_n\right), \tag{47}$$

where $\varepsilon^{-1} \equiv \mu_{ll} - \mu_{cl} + (\mu_{cc} - \mu_{lc})\frac{\gamma_l}{\gamma_c} > 0$ from the second-order conditions. $\varepsilon = \frac{\partial l_n}{\partial w_n}\frac{w_n}{l_n}$ can be interpreted as the compensated *wage* elasticity of labor supply, where $w_n \equiv (1-t)\Phi_l(n, l_n, e_n)$. In the special case that $u(.)$ is homothetic, ε is related to the elasticity of substitution between leisure and consumption (up to a share parameter).

Linearizing the first-order condition for education (see equation (4)) yields

$$-\tilde{t} - \alpha_{ee}\tilde{e}_n + \alpha_{el}\tilde{l}_n = -\tilde{s}, \tag{48}$$

where $\alpha_{ee} \equiv -\frac{\Phi_{ee}e_n}{\Phi_e} > 0$ and $\alpha_{el} \equiv \frac{\Phi_{el}l_n}{\Phi_e} \gtrless 0$.

The last two equations can be solved for $\tilde{l}_n$ and $\tilde{e}_n$ so as to find

$$\tilde{l}_n = -\frac{\varepsilon(\alpha_{ee} + \alpha_{le})}{\Omega}\tilde{t} + \frac{\varepsilon\alpha_{le}}{\Omega}\tilde{s}, \tag{49}$$

$$\tilde{e}_n = -\frac{1 + \varepsilon(\alpha_{ll} + \alpha_{el})}{\Omega}\tilde{t} + \frac{(1 + \varepsilon\alpha_{ll})}{\Omega}\tilde{s}, \tag{50}$$

where $\Omega \equiv \alpha_{ee}(1 + \varepsilon\alpha_{ll}) - \varepsilon\alpha_{el}\alpha_{le} > 0$ from the second-order conditions for utility maximization.

We can rewrite the last expressions, by defining $\rho_{ee} \equiv -\frac{\Phi_{ee}\Phi}{\Phi_e^2} > 0$, and $\rho_{ll} \equiv -\frac{\Phi_{ll}\Phi}{\Phi_l^2} > 0$, as measures for the concavity of the earnings function with respect to learning and labor effort, and using the elasticity of complementarity between learning and working $\rho_{el} \equiv \frac{\Phi_{el}\Phi}{\Phi_e\Phi_l}$ to obtain $\alpha_{ee} \equiv -\frac{\Phi_{ee}e_n}{\Phi_e} = \omega_e\rho_{ee}$, $\alpha_{el} \equiv \frac{\Phi_{el}l_n}{\Phi_e} = \omega_l\rho_{el}$, $\alpha_{le} \equiv \frac{\Phi_{le}e_n}{\Phi_l} = \omega_e\rho_{el}$, and $\alpha_{ll} \equiv -\frac{\Phi_{ll}l_n}{\Phi_l} = \omega_l\rho_{ll}$, where $\omega_l \equiv \frac{\Phi_l l_n}{\Phi}$ and $\omega_e \equiv \frac{\Phi_e e_n}{\Phi}$.

Similarly, Ω can be rewritten to find

$$\Omega = \omega_l\omega_e\varepsilon\left(\rho_{ee}\left(\frac{1}{\varepsilon\omega_l} + \rho_{ll}\right) - \rho_{el}^2\right) > 0. \tag{51}$$

Hence, the second-order condition can be written as $\rho_{ee}\left(\frac{1}{\varepsilon\omega_l} + \rho_{ll}\right) > \rho_{el}^2$. Consequently, the feedback between learning and working in the earnings function, as measured by

ρ_{el}, should be sufficiently small so as to guarantee that second-order conditions are met. Indeed, when there is no feedback ($\rho_{el} = 0$) second-order conditions are always satisfied. Substitution of the α terms and Δ in the expressions for $\tilde{l}_n$ and $\tilde{e}_n$ gives

$$\omega_l \tilde{l}_n = -\frac{\rho_{ee} + \rho_{el}}{\rho_{ee}\left(\frac{1}{\varepsilon\omega_l} + \rho_{ll}\right) - \rho_{el}^2}\tilde{t} + \frac{\rho_{el}}{\rho_{ee}\left(\frac{1}{\varepsilon\omega_l} + \rho_{ll}\right) - \rho_{el}^2}\tilde{s}, \tag{52}$$

$$\omega_e \tilde{e}_n = -\frac{\frac{1}{\varepsilon\omega_l} + \rho_{ll} + \rho_{el}}{\rho_{ee}\left(\frac{1}{\varepsilon\omega_l} + \rho_{ll}\right) - \rho_{el}^2}\tilde{t} + \frac{\frac{1}{\varepsilon\omega_l} + \rho_{ll}}{\rho_{ee}\left(\frac{1}{\varepsilon\omega_l} + \rho_{ll}\right) - \rho_{el}^2}\tilde{s}. \tag{53}$$

The elasticities of labor supply and human capital investment with respect to the policy parameters t and s can, therefore, be written as

$$\varepsilon_{lt} \equiv -\omega_l \frac{\partial l_n}{\partial t}\frac{1-t}{l_n} = \frac{\rho_{ee} + \rho_{el}}{\rho_{ee}\left(\frac{1}{\varepsilon\omega_l} + \rho_{ll}\right) - \rho_{el}^2}, \tag{54}$$

$$\varepsilon_{et} \equiv -\omega_e \frac{\partial e_n}{\partial t}\frac{1-t}{e_n} = \frac{\frac{1}{\varepsilon\omega_l} + \rho_{ll} + \rho_{el}}{\rho_{ee}\left(\frac{1}{\varepsilon\omega_l} + \rho_{ll}\right) - \rho_{el}^2}, \tag{55}$$

$$\varepsilon_{ls} \equiv \omega_l \frac{\partial l_n}{\partial s}\frac{1-s}{l_n} = \frac{\rho_{el}}{\rho_{ee}\left(\frac{1}{\varepsilon\omega_l} + \rho_{ll}\right) - \rho_{el}^2}, \tag{56}$$

$$\varepsilon_{es} \equiv \omega_e \frac{\partial e_n}{\partial s}\frac{1-s}{e_n} = \frac{\frac{1}{\varepsilon\omega_l} + \rho_{ll}}{\rho_{ee}\left(\frac{1}{\varepsilon\omega_l} + \rho_{ll}\right) - \rho_{el}^2}. \tag{57}$$

Note that all the elasticities increase with the elasticity of complementarity ρ_{el}.