Optimal Income Taxation with Endogenous Human Capital

BAS JACOBS
University of Amsterdam

Abstract

This paper augments the theory of optimal linear income taxation by taking into account human capital accumulation as a dimension of labor supply. The distribution of earning potentials is endogenous because agents differ in the ability to learn. Taxation affects utilization rates of human capital through labor supply responses. The costs of education that are not deductible from the income tax distort the learning decision as well. We show theoretically that the trade-off between efficiency and equity is worsened. Quantitative analysis shows that the distortionary costs of taxation increase substantially when human capital formation is endogenous.

"It has long been understood that the concept "labor supply" is more general than "hours of work." If one individual is healthier, better educated and more highly motivated than another, then presumably a given number of hours of work will lead to greater effective labor supply for the former than for the latter. Thus, studies on the effect of taxes on other dimensions of labor supply are needed in order to assess the full impact of taxes on work incentives." Rosen (1980, p. 171).

Bas Jacobs, University of Amsterdam and CPB Netherlands Bureau for Economic Policy Analysis. Department of Economics and Econometrics, University of Amsterdam, Roetersstraat 11, 1018 WB Amsterdam, The Netherlands (b.jacobs@uva.nl).

I am grateful to Lans Bovenberg for support and for providing the opportunity to visit Tilburg University and Center. I also want to thank Casper van Ewijk, Hessel Oosterbeek, and seminar participants at Center for comments and suggestions. Valuable help with Gauss by Edwin Leuven is gratefully acknowledged. Finally, I want to thank three anonymous referees whose comments and suggestions greatly improved this paper. This research was supported by the NWO Priority Program "Scholar" funded by the Netherlands Organization for Sciences.

Received April 2002; Accepted July 2003.

Journal of Public Economic Theory, 7 (2), 2005, pp. 295–315.

1. Introduction

The traditional literature on optimal income taxation with endogenous labor supply assumes that labor supply is a one-dimensional variable reflecting the amount of leisure people wish to consume (see, e.g., Mirrlees 1971). Labor supply, however, features many other dimensions as Rosen (1980) suggests.

This paper analyzes optimal income taxation when the learning dimensions of labor supply are taken into account. From human capital theory, we know that earnings per hour are the result of investments aimed at augmenting effective labor supply (see, e.g., Becker 1964). As a result, the distribution of income is endogenously determined by learning the decisions of agents, rather than exogenously given, as in the traditional papers on optimal income taxation.

The first contribution of this paper is to analytically show how optimal linear tax rates are set when human capital accumulation is endogenous. To that end, we formulate a simple two-period life-cycle model of investment in human capital and labor supply. The tax system distorts not only labor supply decisions but also education and training decisions. The more leisure one wishes to consume, the lower are the returns to human capital investments since less time is spent working and the utilization rate of human capital falls. The reverse reasoning also holds: the more human capital one accumulates, the more expensive leisure time becomes as wage rates per hour increase. Education and leisure decisions are thus interdependent and the distortionary effects of taxation increase with the strength of these interaction effects (see, e.g., Kotlikoff and Summers 1979).

Furthermore, tax distortions may arise in human capital investment decisions if the direct costs (besides foregone earnings) are not deductible from the income tax, as is the case in many countries (see, e.g., Boskin 1975, Trostel 1993). Consequently, future earnings are subject to a higher effective rate of tax than total costs of investment because the direct costs remain "untaxed."[1]

We theoretically show that the trade-off between equity and efficiency worsens due to the interaction of working and learning decisions and the presence of nondeductible direct costs of education. Labor supply, defined in a broad sense so that it encompasses both quantity and quality dimensions, becomes more elastic with respect to the tax. Consequently, optimal linear taxes are lower if human capital accumulation is endogenous rather than exogenous.

The second contribution of this paper is to provide numerical calculations on the importance of endogenous learning decisions for the setting of the optimal linear tax schedule. We follow Stern (1976), who analyzed

[1]For reasons of analytical and computational tractability, we do not allow for a nonlinear tax schedule. However, this is an additional channel whereby taxation may harm human capital formation if marginal tax rates on future incomes exceed marginal tax rates on foregone earnings when learning (see, e.g., Bovenberg and Jacobs 2001).

numerically the case of optimal linear taxes without endogenous learning. The traditional literature has found relatively high optimal tax rates in models with solely an endogenous labor supply decision (see, e.g., Stern 1976, Tuomala 1990, Diamond 1998, Saez 2001). We find that optimal tax rates are substantially lower compared to earlier studies in which human capital accumulation is not explicitly taken into account. This confirms our theoretical predictions.

This paper is related to some earlier contributions on optimal taxation when agents differ in their levels of education. Ulph (1977) and Hare and Ulph (1979) study the simultaneous setting of optimal taxation and education expenditures. Ulph (1977) allows for endogenous labor supply. Hare and Ulph (1979) assume that labor supply is fixed and that agents might opt for private education. In both studies, however, the government simply sets the level of education for each agent. Taxation, therefore, does not influence learning decisions. Tuomala (1986) analyzes optimal taxation in a model where learning and labor supply decisions are endogenous. However, Tuomala assumes that leisure is denoted in "effective" leisure time, i.e., effective labor supply increases linearly with the amount of human capital, as in Heckman (1976). Consequently, the separation between working and learning decisions holds and taxes do not affect learning decisions, by assumption.

The rest of this paper is organized as follows. Section 2 describes the model and individual behavior, Section 3 derives optimal fiscal policy, Section 4 discusses the simulations, and Section 5 concludes.

2. Model

We consider a two-period life-cycle model of human capital formation.[2] In the first period, agents choose between working and learning. We assume without loss of generality that there is no consumption-leisure decision in the first period.[3] Additionally, there is a perfect capital market and agents can

[2]Some authors have used multi-period models to analyze the effects of taxation on human capital accumulation (see, e.g., Heckman 1976, Trostel 1993). However, these papers impose strong restrictions on preferences that avoid corner solutions in the choice of leisure. On a balanced growth path, either all available time may be consumed as leisure, or all of it may be devoted to working (see also Weiss 1986). Moreover, the restrictions on preferences that are often made in order to guarantee that a constant fraction of time is spent on leisure eliminate *a priori* the potential distortionary effect of proportional taxation on human capital formation, as the utilization rate of human capital is unaffected by taxation (since these restrictions imply that substitution and income effects in labor supply cancel out). Trostel (1993) uses a unitary elasticity of substitution between consumption and leisure. Heckman (1976) uses preferences defined over consumption and "effective" leisure so that the leisure decision is independent of the level of human capital.

[3]Given a perfect capital market, no important insights are obtained by allowing for first-period consumption. Furthermore, learning decisions, which are the main focus of the current paper, are not affected by the introduction of first-period leisure time.

save or borrow to finance costs of education.[4] The second period is devoted to working only, and agents decide upon the amount of leisure time (or retirement years) they want to consume.

A partial equilibrium model is chosen where the before-tax wage rates and interest rates are taken as given. The model can also be thought of as describing the equilibrium of a small open economy in which perfect capital mobility fixes the real interest rate.

A mass of agents with unit measure lives for two periods. Agents are heterogeneous with respect to the ability to learn α. Agents with a higher ability are more efficient in production of human capital. The distribution of α is denoted by $F(\alpha)$. F has support $[\underline{\alpha}, \infty)$.

In the first period, agents choose to spend their time learning or working. Every agent has one unit of human capital at the beginning of his life. A fraction x_α of total time in the first period is spent on education. The rest, $1 - x_\alpha$, is devoted to working, where the total time endowment is normalized to unity. Education requires, besides time, y_α market goods per year of education.[5] ϕ is the production function for human capital and measures the number of efficiency units of human capital that individuals acquire through learning. Accumulation of human capital is subject to positive, but diminishing, returns with respect to time x_α and goods y_α invested:

$$\phi(\alpha; x_\alpha, y_\alpha) \equiv h(\alpha)\, x_\alpha^\gamma\, y_\alpha^\upsilon, \tag{1}$$

where we assume $h(\alpha) > 0$, and $h'(\alpha) > 0$. Agents with higher ability levels are assumed to be more productive in using time and goods in human capital accumulation, since $\phi_{\alpha x} > 0$ and $\phi_{\alpha y} > 0$. It is further assumed that the production function displays diminishing returns to scale in inputs (x, y) invested in education. This ensures a solution where the individual does not select a corner in which he saves all his income, either in financial or in human form. In the remainder of the analysis, we restrict ourselves to a Cobb–Douglas production function with constant elasticities γ and υ, and $\gamma + \upsilon < 1$. Given the lack of empirical evidence on the precise shape of the production function for human capital, the Cobb–Douglas function is used in almost the entire literature (see, e.g., Weiss 1986, Trostel 1993).

Income is derived from working equals $(1 - t)w(1 - x_\alpha)$, where t is the flat labor income tax rate. We assume that there is perfect substitution in the demand of labor for different skill types and w denotes the common wage rate per efficiency unit of human capital. The tax authority is assumed to be unable to distinguish between income from raw labor (the quantity or hours of work) and human capital (the quality of work). The tax authority cannot observe α, either. The first assumption is equivalent to the commonly used

[4]See Jacobs (2002) for the consequences of imperfect capital markets for the optimal linear income tax.

[5]One may also allow for capital goods in the production function of human capital, rather than consumption goods. This yields qualitatively similar results.

assumption that one cannot observe the wage rate and hours worked. Thus, taxes on income deriving from the quantity of labor and the quality of labor are both equal to t. The second is the standard assumption that excludes individualized lump-sum transfers.

Every agent might receive a uniform nonindividualized lump-sum income transfer g in both periods of his life. The tax system is progressive because the average tax rate increases with income (if income transfers g are positive). Savings, denoted by s_α equal total first-period income minus the direct costs of education py_α. p denotes the unit costs of direct expenditures on education. The first-period budget constraint is therefore given by

$$py_\alpha + s_\alpha = (1 - t)w(1 - x_\alpha) + g. \tag{2}$$

In the second period, human capital is supplied endogenously to the labor market. The total time spent working equals l_α, and the rest is consumed as leisure $1 - l_\alpha$. One may also view leisure as years in retirement (see Kotlikoff and Summers 1979). Income derived from the accumulation of financial assets is $(1 + r)s_\alpha$, where r is the constant real interest rate. In the remainder of the paper, we assume that the real interest rate is zero. All income from human and financial sources is used for consumption c_α. There is no tax on consumption and capital income.[6] The consumption price is chosen as the numéraire. Hence, the second-period budget constraint is

$$c_\alpha = (1 - t)wl_\alpha\phi(\alpha; x_\alpha, y_\alpha) + s_\alpha + g. \tag{3}$$

We assume that the there are no age-dependent taxes. Consequently, the transfer and the tax rate are equal in both periods. This assumption can be justified under the assumption that age discrimination by the government is not possible.

For analytical tractability, we restrict the analysis to an iso-elastic utility function. Utility u is given by

$$u(c_\alpha, l_\alpha) \equiv \ln\left(c_\alpha - \frac{l_\alpha^{1+1/\varepsilon}}{1 + 1/\varepsilon}\right), \tag{4}$$

where $\varepsilon > 0$ is a parameter governing the (un)compensated wage elasticity of labor supply. Since $\varepsilon > 0$, we assume that labor supply is upward sloping. This utility function is used as well by Diamond (1998) and Saez (2001). It is assumed that agents do not derive utility from having human capital.[7]

Agents maximize utility by choosing consumption c_α, labor supply l_α, the amount of time x_α, and the amount of goods invested in education y_α, subject

---

[6] See Bovenberg and Jacobs (2001), who theoretically analyze optimal dual income taxation.

[7] Human capital can also be regarded as a consumption good. Additionally, having more human capital can enhance the effective productivity of leisure in utility. This notion was first applied to human capital theory by Heckman (1976). Both elements can be incorporated. However, this is likely to yield untractable results unless strong restrictions on preferences are imposed.

to their budget constraints, and the production function of human capital. Manipulation of the first-order conditions gives the following labor supply function:

$$l_\alpha = ((1-t)w\phi(\cdot))^\varepsilon. \tag{5}$$

The higher the hourly wage rate, i.e., the product of the net wage rate and the number of efficiency units of human capital, the larger is labor supply. The last equation demonstrates that consumption and investment decisions cannot be separated. Learning increases the hourly wage rate and thereby boosts labor supply. If leisure time is seen as years in retirement, our model can explain the relatively higher participation rates of older workers with more education: a higher level of human capital makes retirement more expensive.

The marginal rate of technical substitution for the optimal choice of time and goods invested in education reads as

$$\frac{\phi_x}{\phi_y} = \frac{\gamma y_\alpha}{\upsilon x_\alpha} = \frac{(1-t)w}{p}. \tag{6}$$

A higher price of time (goods) invested in education should be accompanied by an increase in the marginal product of time (goods) invested in education, and thus implies a lower use of time (goods) relative to goods (time) in the production of human capital.

Finally, there is an arbitrage equation stating that both financial and human savings should yield an equal return:

$$l_\alpha \phi_x = \frac{(1-t)w l_\alpha \phi_y}{p} = 1. \tag{7}$$

This equation determines the amount of time and goods invested in human capital accumulation. Arbitrage between financial and the human capital investments ensures that an optimal plan is characterized by equal returns on both investments. If the rate of return on financial investments is lower, substitution takes place towards human capital investments until rates of return are equalized (as a consequence of diminishing returns in human capital accumulation).

Since the costs of education are not tax deductible, taxes *directly* distort investments in human capital. A higher tax rate reduces the optimal amount of goods invested in human capital, thereby lowering the productivity of time invested. Investments in human capital fall accordingly. Loosely speaking, total marginal costs, consisting of foregone earnings and goods, are reduced less by the tax rate, than the marginal returns, i.e., future marginal earnings. If goods invested in education would be fully deductible, taxes would have no direct effect on investments in human capital, since the marginal costs and the marginal returns of the investment in human capital would be equally reduced by the tax.

Taxes also distort human capital investments *indirectly*, since taxes affect the amount of leisure chosen. Higher taxes on labor income reduce labor supply, and thereby reduce investments in human capital because the effective utilization rate of human capital decreases, which lowers the returns on investments in human capital.

First-order conditions are necessary, but not sufficient. Additionally, we have to guarantee that the second-order conditions are fulfilled. The second-order condition amounts to the following restriction on parameters:[8]

$$(1+\varepsilon)(\gamma+\upsilon) < 1. \tag{8}$$

The second-order condition states that the elasticity of labor supply is not too high, and that the elasticities of time and goods invested in education are also not too high. Intuitively, if more time is spent learning, wage rates per hour increase, thereby inducing substitution toward more labor supply. This, in turn, increases the returns to investments in human capital, so that more time is spent learning, and so on. Due to this interaction between learning and leisure decisions, sufficiently decreasing returns to investments in human capital (low γ or υ) or sufficiently decreasing the marginal utility of leisure (low ε) should guarantee that an interior solution is attained and that the corner solutions with zero leisure time are avoided.

We can analytically solve for the optimal amount of time and goods invested in learning and labor supply. First, use the marginal rate of technical substitution for goods and time invested in education to get y_α as a function of x_α: $y_\alpha = \upsilon w(1-t)x_\alpha/\gamma p$. Substitute the last result in the equation for labor supply to get l_α as a function of x_α only: $l_\alpha = (\upsilon/(p\gamma))^{\varepsilon\upsilon} h(\alpha)^{\varepsilon} (w(1-t))^{\varepsilon(1+\upsilon)} x_\alpha^{\varepsilon(\gamma+\upsilon)}$. Further, we have $\gamma l_\alpha \phi(\cdot) = x_\alpha$, which follows from the arbitrage condition. Solve the last two equations for x_α:

$$x_\alpha^* = \gamma^{\frac{1}{\mu}} h(\alpha)^{\frac{1+\varepsilon}{\mu}} w^{\frac{\varepsilon+\upsilon(1+\varepsilon)}{\mu}} \left(\frac{\upsilon}{\gamma p}\right)^{\frac{\upsilon(1+\varepsilon)}{\mu}} (1-t)^{\frac{\varepsilon+\upsilon(1+\varepsilon)}{\mu}},$$

where $\mu \equiv 1-(1+\varepsilon)(\gamma+\upsilon) > 0$. y_α^* and l_α^* follow from plugging the value for x_α^* into the equations for y_α and l_α.

Since $h'(\alpha) > 0$, agents with higher ability invest more time and goods in human capital accumulation, i.e., $\partial x_\alpha/\partial\alpha > 0$, $\partial y_\alpha/\partial\alpha > 0$ and $\partial l_\alpha/\partial\alpha > 0$ by virtue of the concavity of the production function of human capital; the complementarity between inputs in production of human capital and ability; and, due to the fact that high ability agents supply more labor. Note that the elasticities of x_α, y_α, and l_α with respect to the tax rate t are constant and independent from skill and given by $\varepsilon_{xt} \equiv -\frac{\partial x}{\partial t}\frac{(1-t)}{x} = \frac{\varepsilon+\upsilon(1+\varepsilon)}{1-(1+\varepsilon)(\gamma+\upsilon)} > 0$, $\varepsilon_{yt} \equiv -\frac{\partial y}{\partial t}\frac{(1-t)}{y} = \frac{(1+\varepsilon)(1-\gamma)}{1-(1+\varepsilon)(\gamma+\upsilon)} > 0$, and $\varepsilon_{lt} \equiv -\frac{\partial l}{\partial t}\frac{(1-t)}{l} = \frac{\varepsilon(1-\gamma)}{1-(1+\varepsilon)(\gamma+\upsilon)} > 0$.

[8]The derivation is available upon request from the author.

3. Optimal Linear Income Taxation

The government collects taxes from the households to finance exogenously given expenditures Λ. The government budget constraint therefore reads as

$$t \int_{\underline{\alpha}}^{\infty} H_\alpha \, dF(\alpha) = \Lambda + G, \tag{9}$$

where $G \equiv 2g$ and $H_\alpha \equiv w(1 - x_\alpha) + wl_\alpha \phi(\alpha; x_\alpha, y_\alpha)$ is the gross lifetime value of labor earnings. There are two instruments at the disposal of the government: the linear tax rate on labor income t and the negative income tax G. The tax rate t and the lump-sum transfer G are chosen so as to maximize a social welfare function Γ:

$$\Gamma = \int_{\underline{\alpha}}^{\infty} \Psi(V_\alpha) \, dF(\alpha), \quad \Psi' > 0, \quad \Psi'' \leq 0, \tag{10}$$

where V_α is the indirect utility function of the agents. Different assumptions about Ψ yield, e.g., a Rawlsian objective function or an utilitarian objective function ($\Psi' = 1$) (see also Atkinson and Stiglitz 1980).[9]

The Lagrangian for maximization of social welfare is given by (omitting the indices α)

$$\mathcal{L} = \int_{\underline{\alpha}}^{\infty} (\Psi(V) + \eta(twl\phi(\cdot) + tw(1 - x) - G - \Lambda)) \, dF(\alpha), \tag{11}$$

where η is the Lagrange multiplier associated with the government budget constraint. The first-order condition for G is given by

$$\frac{\partial \mathcal{L}}{\partial G} = \int_{\underline{\alpha}}^{\infty} \left(\Psi'(\cdot)\lambda - \eta + \eta t \frac{\partial H}{\partial G} \right) dF = 0, \tag{12}$$

where we substituted Roy's lemma $\partial V_\alpha / \partial G = \lambda_\alpha$ and λ_α is the private marginal utility of income. Note that there are no income effects on both labor supply and investments in human capital. Consequently, we have $\frac{\partial H}{\partial G} = 0$.

Denote the net social marginal valuation of income of individual α in terms of government revenue by b_α (see Atkinson and Stiglitz 1980):

$$b_\alpha \equiv \frac{\Psi'(\cdot)\lambda_\alpha}{\eta}. \tag{13}$$

The term on the right-hand side denotes the direct social value of redistribution to household α. From the first-order condition for G, we find that the marginal social value of income averaged over all households is given by

$$\bar{b} = 1, \tag{14}$$

---

[9]We assume in the theoretical derivations that lump-sum transfers are never larger than income derived from supplying human capital, i.e., $G < H$, so that agents never decide to be unemployed. This constraint is always nonbinding in the numerical calculations, below.

where $\bar{b} \equiv \int_{\underline{\alpha}}^{\infty} b_\alpha \, dF(\alpha)$. This expression states that social welfare is maximized if a unit increase in the value of the lump-sum transfer given in both the periods is equal to the marginal social utility averaged over all agents.

The first-order condition for the tax rate is given by

$$\frac{\partial \mathcal{L}}{\partial t} = \int_{\underline{\alpha}}^{\infty} (-\Psi'(\cdot)\lambda(wl\phi(\cdot) + w(1-x)) + \eta(wl\phi(\cdot) + w(1-x)))\, dF$$
$$+ \int_{\underline{\alpha}}^{\infty} \left(\eta tw\phi(\cdot)\frac{\partial l}{\partial t} + \eta\left(twl\left(\phi_x \frac{\partial x}{\partial t} + \phi_y \frac{\partial y}{\partial t}\right) - tw\frac{\partial x}{\partial t}\right)\right) dF = 0, \tag{15}$$

where Roy's lemma is used: $\partial V_\alpha / \partial t = -\lambda_\alpha H_\alpha$. (If goods invested in education are deductible we would have an additional $-tp\frac{\partial y}{\partial t}$ in the last term in brackets.)

In order to find an expression for the optimal tax rate, we introduce the distributional characteristic ξ that comprises the distributional impact that human capital has on social welfare (see also Atkinson and Stiglitz 1980):

$$\xi \equiv -\left(\int_{\underline{\alpha}}^{\infty} \left(\frac{H_\alpha}{\bar{H}}\right)\left(\frac{b_\alpha}{\bar{b}}\right) dF(\alpha) - 1\right). \tag{16}$$

The distributional characteristic ξ of the income tax base is given by the negative of the normalized covariance[10] between the welfare weight the government attaches to the income of a particular skill b_α (which is nonincreasing with the skill level α) and the contribution of individual α to the tax base H_α. $\bar{H} \equiv \int_{\underline{\alpha}}^{\infty} H_a dF(\alpha)$ stands for the average value of lifetime earnings. ξ can be interpreted as a "marginal measure of inequality"(see Atkinson and Stiglitz 1980). A positive distributional characteristic ξ thus implies that the tax base is larger for high skills (which feature low welfare weights) than for low skills, so that taxing this base generates positive distributional benefits. The magnitude of the distributional characteristic depends both on the correlation between skills and the tax base and the strength of the redistributive preferences as reflected in the negative correlation between skills and the welfare weights.[11] Indeed, a distributional characteristic of zero may indicate either that the government is not interested in redistribution (so that all skills feature the same welfare weight) or that the marginal contribution to the tax base is the same for all skills (all individuals have equal incomes and there is not need for redistribution).

[10]This can be seen by noting that $\xi = \frac{-1}{\bar{H}\bar{b}}(\int_{\underline{\alpha}}^{\infty} Hb dF - \int_{\underline{\alpha}}^{\infty} H\, dF \int_{\underline{\alpha}}^{\infty} b\, dF) = \frac{-cov(H,b)}{\bar{H}\bar{b}}$.

[11]The strength of this negative correlation depends not only on the concavity of the function Ψ, but also on inequality in life-time incomes. In particular, the government attaches a higher priority to redistributing incomes if life-time incomes become more unequal, since marginal utility of income declines with income.

We obtain an expression for the optimal linear income tax from the first-order condition for t (Equation (15)), after substituting the definition of ξ, the first-order condition for x ($twl\phi_x \frac{\partial x}{\partial t} - tw\frac{\partial x}{\partial t} = 0$), and rewriting the last term in brackets of Equation (15):[12]

$$\frac{t}{1-t} = \frac{\xi}{\omega(\varepsilon_{lt} + \upsilon\varepsilon_{yt})}, \tag{17}$$

where $\omega \equiv \int_{\underline{\alpha}}^{\infty} wl\phi(\cdot)\, dF / \int_{\underline{\alpha}}^{\infty} wl\phi(\cdot) + w(1-x)\, dF$ is the ratio of average second-period income in average total income. The optimum tax formula clearly shows the trade-off between equity (numerator) and efficiency (denominator).

First, the tax rate should be higher if the absolute value of the distributional characteristic ξ is higher, i.e., when the social value of redistributing incomes is higher. This is the case if incomes are more unevenly distributed, or if greater weight is attached to agents at the lower end of the distribution. If all agents have identical abilities, H is identical for all agents, and there is no income inequality. Consequently, $\xi = 0$ and the optimal tax rate is zero.

Second, the denominator of the optimal tax formula shows two elasticities associated with the two tax distortions in our model. The first elasticity (ε_{lt}) is associated with the distortionary effect of taxes on labor supply. The optimal tax rate on labor income should be lower if the elasticity of labor supply is larger. From the definition of the labor supply elasticity we can see that the "true" wage elasticity of broad labor supply, including the learning effects is larger than the "simple" elasticity of labor supply ($\varepsilon_{lt} > \varepsilon$) that would enter in the optimum tax formula in the absence of learning decisions (see, e.g., Atkinson 1995):

$$\frac{\varepsilon_{lt}}{\varepsilon} = \frac{1-(\gamma+\upsilon)}{1-(1+\varepsilon)(\gamma+\upsilon)} + \frac{\upsilon}{1-(1+\varepsilon)(\gamma+\upsilon)} > 1. \tag{18}$$

The first term is larger than the unity and the second term is positive. The first term measures the interaction impact of learning and working decisions, and the second term gives the additional impact of the nondeductibility of the goods invested in education, see also Equation (21) for the optimal tax formula with full tax deductibility of the educational costs. Clearly, the interaction between learning and labor supply decisions makes the labor supply response more elastic and drives the optimal tax rate downward.

Besides the standard elasticity of labor supply, there is a second elasticity in the denominator of the tax formula in Equation (17) ($\upsilon\varepsilon_{yt}$). This is the tax elasticity of investments in human capital. The optimal tax should be lower if the tax elasticity of goods invested in education (ε_{yt}) is larger. It is easily seen

[12]If goods invested in education are deductible, the last term in brackets of Equation (15) is zero by substitution of the first-order condition for y ($twl\phi_y \frac{\partial y}{\partial t} - tp\frac{\partial y}{\partial t} = 0$).

that the tax elasticity of goods invested in education is also enlarged due to the interaction with labor supply (ε). Suppose that labor supply was inelastic ($\varepsilon = 0$), then we can derive for the tax elasticity of learning (ε_{yt}):

$$\varepsilon_{yt}|_{\varepsilon=0} = \frac{1-\gamma}{1-(\gamma+\upsilon)} < \varepsilon_{yt} = \frac{(1+\varepsilon)(1-\gamma)}{1-(1+\varepsilon)(\gamma+\upsilon)}. \tag{19}$$

The elasticities ε_{lt} and ε_{yt} are weighted with the share of second-period income in the total lifetime income ω. The larger second-period income becomes, the more elastic total lifetime income becomes, and the lower optimal linear taxes should be.

If goods invested were fully deductible, the optimum tax would be given by:[13]

$$\frac{t}{1-t} = \frac{\xi}{\omega\varepsilon_{lt}}. \tag{20}$$

Note that only elasticity $\upsilon\varepsilon_{yt}$ drops out compared to Equation (17). Moreover, the elasticity of labor supply is lowered with the elasticity given in Equation (18)

$$\frac{\varepsilon_{lt}}{\varepsilon} = \frac{1-(\gamma+\upsilon)}{1-(1+\varepsilon)(\gamma+\upsilon)} > 1. \tag{21}$$

Clearly, the interaction effect between learning and working remains at work even when the goods invested in education are made tax deductible.

The last formula provides a quantitative idea about the increase in the size of the elasticities when learning is endogenous. Suppose that $\gamma + \upsilon = 0.6$. These are the values suggested by Trostel (1993). Let the elasticity of labor supply be equal to $\varepsilon = 0.25$, which is not an uncommon figure in the literature, see also below. Then we find that the elasticity of broad labor supply, including the interaction effects with learning, is equal to $\varepsilon_{lt} = 0.4$. In other words, the "true" elasticity of the labor supply is about 60% larger than the simple elasticity. Now, suppose that the simple elasticity of the labor supply is $\varepsilon = 0.5$ (an upper bound in the literature) then we find an elasticity of broad labor supply that is four times larger and equal to $\varepsilon_{lt} = 2$. Clearly, the interaction mechanism between the labor supply and the learning decisions has a potentially big impact on the elasticity of broad labor supply, and optimum taxes should be lowered accordingly.

4. Numerical Examples

This section considers some numerical examples of the optimal tax rates. The method employed here stems from Stern (1976). The distribution of ability

[13] This follows from redoing the analysis with $(1-t)p$ as the measure for direct costs on the side of households and adding a cost tpy to the government budget constraint.

is assumed to be normal with mean μ_α, and *SD* σ_α:[14]

$$\alpha \sim N[\mu_\alpha; \sigma_\alpha]. \tag{22}$$

Ability has a mean $\mu_\alpha = -1$, which is a normalization.

The productivity of ability in human capital accumulation is an exponential function:

$$h(\alpha) = A \exp(\alpha)^{\psi}. \tag{23}$$

A is a general efficiency parameter denoting the productivity of learning. If one assumes that ability follows a normal distribution, this specification yields a log-normal wage distribution of second-period incomes, since log second-period income is linear in α. ψ denotes the elasticity of ability in learning and is calibrated to give a realistic spread in the learning distribution.

For the parameterization of the production function of human capital, we refer to Trostel (1993) who provides a very extensive discussion of plausible parameter values. The share of time in the production of human capital is set at $\gamma = 0.3$ and the share of goods in the production of human capital is set at $\upsilon = 0.1$. So, the total returns to private inputs are 0.4. Here, Trostel (1993) uses the values of $\gamma = 0.45$ and $\upsilon = 0.15$. However, these high values turn out to give occasional problems with the second-order conditions, see also Equation (8).

The values $\gamma = 0.3$ and $\upsilon = 0.1$ imply that the direct costs of education are one-quarter of total expenditures in education, so that the foregone earnings make up three-quarters of the total costs of education. Becker (1964) and Boskin (1975) find that the private cost shares of time and goods invested in education are three-quarters and one-quarter, respectively.[15]

In addition, we assume, in contrast to the theoretical derivation, that ability α also affects wage rates w independently of the amount of learning, so that $w(\alpha)$ where $w(\alpha)' > 0$. The reason for making this assumption is that not all income inequality can be attributed to differences in learning behavior. Consequently, the agents with higher ability have a higher wage rate per unit of human capital as well.[16] Wage rates are also assumed to be generated by an exponential wage equation:

$$w(\alpha) = \exp(\alpha), \tag{24}$$

[14]Note that, in contrast to the derivation above, ability does not have a lower bound. However, nothing critical changes in the theoretical derivations if ability runs from $(-\infty, \infty)$ instead. Moreover, we will want to use a specific functional form on h, so that ability always translates into positive learning productivity $h(\alpha)$ for the model to have a useful economic meaning.

[15]The price of direct costs of education is arbitrarily set at $p = 0.5$.

[16]The theoretical derivation earlier is not affected by making this assumption.

so that a log-normal wage distribution results (see also Mirrlees 1971, Stern 1976, Tuomala 1990). The standard deviation of log wages in these papers is set at 0.39.[17]

The social welfare function is a Samuelson–Bergson utility function with a constant elasticity of inequality aversion ν:

$$\Gamma = \int_{\underline{\alpha}}^{\infty} \frac{V^{1-\nu} - 1}{1 - \nu} dF(\alpha). \tag{25}$$

If $\nu = 0$, the social welfare function is utilitarian; if $\nu = \infty$, the social welfare function is Rawlsian (see also Atkinson and Stiglitz 1980). In the base case scenario, the social welfare function is utilitarian, so that $\nu = 0$. Taxes are solely redistributive as the government revenue requirement is set at $\Lambda = 0$.

We use two types of utility functions. First, to make our model comparable with the optimum tax literature, we use the standard CES utility function with a constant elasticity of substitution between consumption and leisure as in Mirrlees (1971), Stern (1976), and Tuomala (1990):

$$u(c, l) = (\beta^{1-\zeta} c^{\zeta} + (1 - \beta)^{1-\zeta} (1 - l)^{\zeta})^{1/\zeta}. \tag{26}$$

The elasticity of substitution between second-period consumption and leisure equals $\sigma \equiv 1/(1 - \zeta)$. We follow common practice by setting $\sigma = 0.5$ in the base line computations. Stern (1976) uses a value of $\sigma = 0.4$ and Tuomala (1990) uses $\sigma = 0.5$ based on reviewing the literature.[18]

Recently, Atkinson (1995), Diamond (1998), and Saez (2001) adopted the iso-elastic utility function that was used in the theoretical derivation. This serves as the basis of our second specification:

$$u(c, l) = \ln\left(c - \theta \frac{l^{1+1/\varepsilon}}{1 + 1/\varepsilon}\right), \tag{27}$$

where we added a parameter θ denoting the preference for leisure. We set the uncompensated wage elasticity of labor supply at $\varepsilon = 0.25$. The uncompensated elasticity of 0.25 is in the middle of elasticities for men and women that are encountered in the microeconometric literature. For men, the elasticity is slightly below 0, whereas significantly higher elasticities, ranging from 0.5 to 1, are reported for women (see Hansson and Stuart 1985, Killingsworth and Heckman 1986, Pencavel 1986). An average value of 0.1 is found in the

[17]We construct a data set with 10 observations representing the deciles according to ability. Within each decile we take the mean value of ability as a data point. We have constructed larger samples, but relatively small increases in precision of the computations were obtained with relatively large increases in computation time. Further, the conclusions of this paper are not sensitive to the sample size.

[18]Again, second-order conditions require that parameters on preferences and production elasticities are restricted, i.e., the elasticity of substitution or the production elasticities are not too high (see also Jacobs 2000).

latter study on the basis of reviewing the literature. In this model, the labor supply decision could also be thought of as the retirement decision, as in Kotlikoff and Summers (1979). A somewhat higher elasticity of labor supply potentially also captures the effects taxes might have on early retirement as these effects are generally ignored in empirical estimates.

The last parameters are jointly calibrated to make the outcomes as realistic as possible. The last parameters include the following: the common learning technology parameter A, the leisure share parameter $\beta(\theta)$, the elasticity of ability in learning ψ, and the *SD* of ability σ_α. We impose four identifying conditions on the model at $t = 0$ and $G = 0$: mean working time is $\bar{1} = 0.67$, mean learning time is $\bar{x} = 0.67$, the *SD* of learning time equals $\sigma_x = 0.12$, the *SD* of the log of total income is 0.40.

The value of mean working time is taken from Stern (1976) and Tuomala (1990). This implies that the average individual would work two-thirds of the day. It could also correspond to a retirement period of 10 years if one regards each period in life as lasting for approximately 30 years.

A mean learning time of 0.67 implies that agents spend on average 20 years on learning in the first period of their lives, if each period in life lasts 30 years. This is high compared with the average time spent on formal education. Harmon and Walker (1999) find that the mean is 11.90 years for the United Kingdom (in the General Household Survey 1974–1994). Ashenfelter and Krueger (1994) report an average of 13.1 schooling years for the United States (from the 1990 Current Population Survey). However, on-the-job training (OJT) is also a part of human capital formation. Mincer (1962) estimates that half of the total human capital formation is on the job. Computations by Heckman et al. (1998) suggest that the contribution by OJT is lower and is in the range of one-quarter of total human capital formation. If we assume that approximately one-third of total human capital formation is OJT, and two-thirds is formal education, then a mean of 13.3 years of formal schooling results. This corresponds with the actual figures.

If we proxy the learning distribution with a normal distribution, then we are able to compute the spread in learning outcomes in the years of formal education from the model. Under the assumption that two-thirds of the human capital is acquired through formal education and each period takes 30 years, a *SD* of 0.12 corresponds to a *SD* in learning time equal to 2.4 years. Harmon and Walker (1999) find that the *SD* equals 2.83 years. Ashenfelter and Krueger (1994) report a *SD* of 2.7 years.

The *SD* of log incomes is calibrated at 0.4, since the distribution of incomes is endogenous. The models of optimum income taxation without learning behavior assume similar inequality. Mirrlees (1971), Stern (1976), and Tuomala (1990) use a *SD* of log wages of 0.39.

The calibration with the CES function yielded a productivity parameter $A = 7.4$, a preference for leisure parameter $\beta = 0.7$, a *SD* of ability of $\sigma_\alpha = 0.31$, and a value of the elasticity of ability $\psi = 0.5$. The calibration with the

Table 1: Outcomes base line parameterization—CES utility

Percentile (%)	c	$1 - l$	x	y	α
0–10	0.43	0.27	0.46	0.07	−1.51
10–20	0.55	0.29	0.53	0.09	−1.32
20–30	0.64	0.30	0.58	0.11	−1.21
30–40	0.73	0.32	0.61	0.13	−1.12
40–50	0.81	0.33	0.65	0.15	−1.04
50–60	0.90	0.34	0.68	0.17	−0.96
60–70	1.01	0.35	0.71	0.20	−0.88
70–80	1.14	0.36	0.75	0.23	−0.79
80–90	1.32	0.38	0.80	0.27	−0.68
90–100	1.70	0.41	0.89	0.36	−0.49
Mean	0.92	0.34	0.67	0.18	−1
SD	0.12	0.04	0.12	0.08	0.31

CELS specification yielded values of $A = 4.4$, $\theta = 5.7$, $\psi = 0$, and $\sigma_\alpha = 0.30$.

We compare the outcomes of the two models in the simulations, see Tables 1 and 2 for the base-line calibration outcomes. A feature of the two utility functions used here is that the labor supply behavior is rather different. In the CES case, we have a backward-bending labor supply curve with $\sigma < 1$. The uncompensated wage elasticity of labor supply is negative at zero nonlabor income (see also Stern 1976). This implies that income taxation induces agents to work more. And, as the utilization rate of human capital increases, learning time increases as well. In the constant elasticity of labor

Table 2: Outcomes base line parameterization—CELS utility

Percentile (%)	c	$1 - l$	x	y	α
0–10	0.43	0.43	0.45	0.07	−1.49
10–20	0.54	0.39	0.52	0.09	−1.31
20–30	0.64	0.37	0.56	0.11	−1.20
30–40	0.72	0.35	0.60	0.13	−1.11
40–50	0.80	0.33	0.64	0.15	−1.04
50–60	0.90	0.32	0.67	0.17	−0.96
60–70	1.00	0.30	0.71	0.20	−0.88
70–80	1.13	0.28	0.76	0.23	−0.80
80–90	1.33	0.25	0.83	0.28	−0.69
90–100	1.74	0.20	0.95	0.38	−0.51
Mean	0.92	0.32	0.67	0.18	−1
SD	0.14	0.07	0.14	0.09	0.30

Table 3: Optimal tax rates t (%) (optimal values for g)

Base: $\boldsymbol{\sigma = 0.5}$	**CES Endogenous**	**Exogenous**	**Base:** $\boldsymbol{\varepsilon = 0.25}$	**CELS Endogenous**	**Exogenous**
$\sigma = 0.2$	24.8 (0.113)	50.9 (0.238)	$\varepsilon = 0.1$	20.5 (0.124)	41.0 (0.239)
$\sigma = 0.3$	21.3 (0.099)	40.5 (0.191)	$\varepsilon = 0.2$	17.1 (0.086)	31.1 (0.157)
$\sigma = 0.4$	19.0 (0.089)	34.1 (0.162)	$\varepsilon = 0.25$	16.1 (0.074)	28.2 (0.133)
$\sigma = 0.5$	17.4 (0.083)	29.8 (0.143)	$\varepsilon = 0.3$	15.3 (0.064)	26.0 (0.114)
$\sigma = 0.6$	16.2 (0.078)	26.7 (0.130)	$\varepsilon = 0.4$	14.2 (0.049)	22.6 (0.087)
$\sigma = 0.7$	15.3 (0.075)	24.5 (0.120)	$\varepsilon = 0.5$	13.6 (0.039)	19.9 (0.068)
$\sigma = 0.8$	14.6 (0.073)	22.6 (0.112)			

supply (CELS) case, labor supply is always upward sloping. Taxation induces agents to work less, on account of a dominant substitution effect, and they also learn less as a consequence.[19]

We derived optimal tax rates for the case in which both learning and leisure are endogenous and for the case in which only labor supply is endogenous, and we fix the investments in human capital at the values that are obtained in the calibration. The latter case provides the natural benchmark to show the effects of endogenous learning decisions.

To compute optimal tax rates, we used a (quasi) Newton algorithm to maximize the social welfare function.[20] This procedure numerically approximates the Hessian matrix of the optimization program. As starting values for (t^o, g^o) we take (0.1, 0.1). We solve numerically the individuals' problem from the first-order conditions and the budget constraints also using a quasi-Newton algorithm. For all simulations, we checked the global behavior of the social welfare function using a grid procedure over the range of all feasible tax rates in order to ensure that our optimization algorithm indeed converges to a global maximum, rather than a possible local one. The grid procedure confirms that there is at most one maximum in all computations. Stern (1976) also finds at most one maximum using a grid method.

Table 3 shows the results for various elasticities of substitution or labor supply elasticities. We find an optimum tax rate of 17.4% in the benchmark case of the CES utility function with $\sigma = 0.5$. The corresponding value of the optimum tax rate when learning is exogenous equals 29.8%. The CELS utility function with a base case value of $\varepsilon = 0.25$ gives values of 16.1% (28.2%) when learning is endogenous (exogenous). Clearly, optimum taxes are much

[19]Moreover, the theoretical models cannot be consistently matched with the empirical literature. A value of the elasticity of substitution between consumption and leisure smaller than 1 cannot be reconciled with an upward-sloping labor supply curve (if nonlabor income is 0). In the remainder, we proceed by analyzing the two cases separately.

[20]The Gauss programs for all computations can be downloaded from: http://www.fee.uva.nl/scholar/mdw/jacobs/optax.zip.

lower when learning is endogenous. In our calculations, optimal taxes are reduced by almost one-half when learning decisions are taken into account. This result is robust to changes in the elasticities of substitution σ or changes in the elasticity of labor supply. A reassuring aspect of our computations is that very similar results are obtained when using the CES and CELS functions.

Regarding the optimal values of the lump-sum transfer g, we note that there is a general positive relation between the optimal tax rates and the value of the transfer. The higher the optimal tax, the larger the transfer.

Our computed tax rates are also lower than the optimal marginal tax rates that are reported in the literature. Our optimal linear taxes are always lower than the ones obtained by Stern (1976) with the CES utility function for various elasticities of substitution. Although, Saez (2001) found marginal rates far above 50% in the model with a CELS utility function, he set the revenue requirement by the government at 0.25 of production. For the sake of comparison, we have computed the optimum rates with this revenue requirement ($\Lambda = 0.23$). For $\varepsilon = 0.25$, we derive an optimum tax $t = 21.0\%$ and for $\varepsilon = 0.50$ we find $t = 22.3\%$. These optimum taxes are substantially lower than the ones from Saez (2001).[21]

Using a CES utility function, Tuomala (1990, p. 98) found optimal nonlinear marginal tax rates ranging from 65% at the first decile of the income distribution to 45% at the ninth decile of the income distribution, with a marginal rate of tax of 59% at the median in the case where $\sigma = 0.5$. The optimal marginal tax rate in our model is 17.4% at $\sigma = 0.5$, which is again a considerably lower marginal tax rate. Using CELS utility functions, Diamond (1998) and Saez (2001) find in nonlinear versions of their models the marginal tax rates on income that are generally higher than 50%, even for the top deciles.

Below, we show robustness checks for various modifications in technology, preferences, or government parameters. The result of lower optimum taxes with endogenous human capital is not sensitive to the parameters used in the model.

In Table 4, we change production elasticities. Here, it must be noted that the range over which the parameters can be varied is limited. Too high values violate second-order conditions, so there are limits on the returns to private inputs (so as to rule out perverse behavior). From Table 4, we can see that changing the elasticities of production yields only small effects on the optimal tax rates. So, the results are robust with respect to the technology parameters of the production function of human capital.[22]

[21] It has to be noted, however, that the amount of pre-tax income inequality is probably different in Saez (2001) compared to other studies, since he uses empirical income distributions rather than artificially generated ones.

[22] We note that optimum taxes increase when the elasticity of time in production of human capital increases for the CES utility function. This can be attributed to the backward-bending labor supply curve. Taxation induces agents to work more, and rates of return to

Table 4: Optimal tax rates t (%) (optimal values for g)—Changing production parameters

Base: $\boldsymbol{\gamma = 0.3}$	**CES** **Endogenous**	**Exogenous**	**CELS** **Endogenous**	**Exogenous**
$\gamma = 0.1$	16.8 (0.086)	27.3 (0.139)	18.6 (0.097)	25.3 (0.129)
$\gamma = 0.2$	16.9 (0.082)	28.6 (0.141)	17.6 (0.083)	26.8 (0.132)
$\gamma = 0.3$	17.4 (0.083)	29.8 (0.143)	16.1 (0.074)	28.2 (0.133)
$\gamma = 0.4$	18.3 (0.090)	32.3 (0.145)	14.2 (0.069)	29.5 (0.134)
$\gamma = 0.5$	20.0 (0.111)	33.5 (0.147)	12.1 (0.074)	30.9 (0.134)
$\gamma = 0.6$	23.2 (0.173)	34.6 (0.149)	10.9 (0.036)	32.0 (0.134)
Base: $\boldsymbol{\upsilon = 0.3}$	**CES** **Endogenous**	**Exogenous**	**CELS** **Endogenous**	**Exogenous**
$\upsilon = 0.05$	19.5 (0.096)	28.2 (0.144)	18.0 (0.089)	26.8 (0.139)
$\upsilon = 0.1$	17.4 (0.083)	29.8 (0.134)	16.1 (0.074)	28.2 (0.134)
$\upsilon = 0.15$	15.9 (0.075)	31.4 (0.142)	14.5 (0.063)	29.4 (0.127)
$\upsilon = 0.2$	–	–	13.0 (0.055)	30.5 (0.120)

Table 5: Optimal tax rates t (%) (optimal values for g)—Changing government parameters

Base: $\boldsymbol{\nu = 0}$	**CES** **Endogenous**	**Exogenous**
$\nu = 0.99$	31.0 (0.139)	43.7 (0.203)
$\nu = 2$	38.3 (0.166)	50.2 (0.229)
$\nu = 3$	42.9 (0.181)	54.2 (0.243)

Base: $\boldsymbol{\Lambda = 0}$	**CES** **Endogenous**	**Exogenous**	**CELS** **Endogenous**	**Exogenous**
$\Lambda = 0.1$	18.1 (0.040)	31.7 (0.108)	17.9 (0.031)	30.3 (0.092)
$\Lambda = 0.3$	19.8 (−0.044)	36.1 (0.041)	23.3 (−0.047)	35.9 (0.015)
$\Lambda = 0.5$	21.6 (−0.125)	41.5 (−0.016)	37.5 (−0.098)	46.6 (−0.044)

Finally, in Table 5, we compute optimum income taxes in the cases where either the revenue requirement of the government or the elasticity of inequality aversion is increased. In both cases, the optimum tax rates increase as expected. Note that the lump-sum transfer g may become negative if the government has large revenue requirements. The earlier conclusion that optimum

investments in human capital to rise as a result. More time is spent learning, the tax base increases, and costs of redistribution fall accordingly. This effect is stronger if the elasticity is larger.

tax rates are lower with endogenous learning decisions is also confirmed here.[23]

5. Conclusion

This paper augments the standard theory of optimal income taxation with human capital. Labor supply has now both a quantity dimension (hours worked) and a quality dimension (acquired human capital). We derived a simple optimal tax formula that disentangled the various tax distortions (labor supply and nondeductible costs of education) that determine the trade-off between efficiency and equity. Quantitative analysis showed that distortions in labor supply may substantially increase when learning decisions are endogenous.

We abstracted from the dynamic consistency of the tax policies and simply assumed that the government can pre-commit. However, in models like the one discussed in this paper, the government has always the incentive to renege on its announcement to set a particular tax schedule after the investments in human capital are made. Accumulated human capital has become a "fixed" factor that can be taxed heavily without high distortionary costs. If the government cannot commit, agents underinvest as a consequence (see, e.g., Boadway et al. 1996). Future research may show that the case for progressive labor income taxes may be reduced even further once dynamic inconsistency issues in policies are taken into account.

In future research one may also allow for imperfect substitution between labor types on the labor market (see, e.g., Dur and Teulings 2001). Again, the case for progressive taxes may be weakened because human capital formation contributes to equality since increases in the supply of skilled workers diminish wage inequality between relatively skilled and unskilled workers. Therefore, the government may have less need to rely on progressive taxes to achieve a certain amount of equality.

Finally, one may allow for the presence of obligatory education and/or education subsidies. Obligatory education reduces the elasticity of broad labor supply with respect to taxes, so that the tax schedule may become more progressive. Similarly, Bovenberg and Jacobs (2001) show that education subsidies allow the government to tax incomes more progressively, since education subsidies reduce the distortions from progressive taxation on human capital formation. Nevertheless, these policies critically hinge on the power of the government to affect human capital formation through education subsidies or law. However, a very large part of human capital formation takes

[23]The optimization routine was not able to compute optimum taxes for the CELS utility function with positive inequality aversion because utility levels were negative for some agents due to the scaling parameter θ. This gave a problem that powers had to be raised to negative numbers. Transformations of the utility function to overcome this problem are not innocuous, since the amount of redistribution depends on the cardinalization of the utility function.

place within the family under circumstances that cannot be easily controlled by governments through education policy (see also Carneiro and Heckman 2003). If education efforts cannot be controlled effectively through government policies, then the results from this paper gain in relevance.

References

ASHENFELTER, O., and A. B. KRUEGER (1994) Estimates of the economic return to schooling from a new sample of twins, *American Economic Review* **84**(5), 1157–1173.

ATKINSON, A. B. (1995) *Public Economics in Action. The Basic Income/Flat Tax Proposal*. Oxford: Oxford University Press.

ATKINSON, A. B., and J. E. STIGLITZ (1980) *Lectures on Public Economics*. New York: McGraw-Hill.

BECKER, G. S. (1964) *Human Capital: A Theoretical and Empirical Analysis with Special Reference to Education*. Third edition 1993, Chicago: Chicago University Press.

BOADWAY, R., N. MARCEAU, and M. MARCHAND (1996) Investment in education and the time inconsistency of redistributive tax policy, *Economica* **63**, 171–189.

BOSKIN, M. (1975) Notes on the tax treatment of human capital. NBER Working Paper 116.

BOVENBERG, A. L., and B. JACOBS (2001) Redistribution and education subsidies are siamese twins. CEPR Discussion Paper No. 3309.

DUR, R., and C. N. TEULINGS (2001) Education and efficient redistribution. Tinbergen Institute Discussion Paper 090/3.

CARNEIRO, P., and J. J. HECKMAN (2003) Human capital policy. Presented at the Alvin Hansen Seminar, Harvard University, April 25, 2002, Revised version: January 21, 2003.

DIAMOND, P. A. (1998) Optimal income taxation: An example with a U-shaped pattern of optimal marginal tax rates, *American Economic Review* **88**(1), 83–95.

HANSSON, I., and C. STUART (1985) Tax revenue and the marginal costs of public funds in sweden, *Journal of Public Economics* **27**, 331–353.

HARE, P. G., and D. T. ULPH (1979) On education and distribution, *Journal of Political Economy* **87**(5), S193–S212.

HARMON, C., and I. WALKER (1999) The marginal and average returns to schooling, *European Economic Review* **43**, 879–887.

HECKMAN, J. J. (1976) A life-cycle model of earnings, learning and consumption, *Journal of Political Economy* **4**, S11–S44.

HECKMAN, J. J., L. LOCHNER, and C. TABER (1998) Explaining rising wage inequality: Explorations with a dynamic general equilibrium model of labor earnings with heterogeneous agents, *Review of Economic Dynamics* **1**, 1–58.

JACOBS, B. (2000) A note on taxation and human capital formation. Mimeo: University of Amsterdam.

JACOBS, B. (2002) Optimal taxation of human capital and credit constraints. Tinbergen Institute Discussion Paper 2002-044/2, March 2002.

KILLINGSWORTH, M. R., and J. J. HECKMAN (1986) Female labor supply: A survey. Chapter 2, in *Handbook of Labor Economics—Vol. I*, O. Ashenfelter and R. Layard, eds. Amsterdam: Elsevier Science, 103–204.

KOTLIKOFF, L. J., and L. H. SUMMERS (1979) Tax incidence in a life cycle model with variable labor supply, *Quarterly Journal of Economics* **93**, 705–718.

MINCER, J. (1962) On the job training: Costs, returns and some implications, *Journal of Political Economy* **70**, 50–79.

MIRRLEES, J. A. (1971) An exploration in the theory of optimum income taxation, *Review of Economic Studies* **38**, 175–208.

PENCAVEL, J. (1986) Labor supply of men: A survey. Chapter 1, in *Handbook of Labor Economics—Vol I*, O. Ashenfelter and R. Layard, eds. Amsterdam: Elsevier Science Publishers BV, 3–102.

ROSEN, H. S. (1980) What is labor supply and do taxes affect it? *American Economic Association—Papers and Proceedings* **70**(2), 171–176.

SAEZ, E. (2001) Using elasticities to derive optimal income tax rates, *Review of Economic Studies* **68**, 205–229.

STERN, N. H. (1976) On the specification of models of optimum income taxation, *Journal of Public Economics* **6**, 123–162.

TROSTEL, P. A. (1993) The effect of taxation on human capital, *Journal of Political Economy* **101**, 327–350.

TUOMALA, M. (1986) On the optimal income taxation and educational expenditures, *Journal of Public Economics* **30**, 183–198.

TUOMALA, M. (1990) *Optimal Income Tax and Redistribution.* Oxford: Clarendon Press.

ULPH, D. T. (1977) On the optimal distribution of income and educational expenditure, *Journal of Public Economics* **8**, 341–356.

WEISS, Y. (1986) The determination of life-cycle earnings: A survey. Chapter 11, in *Handbook of Labor Economics*. O. Ashenfelter and R. Layard, eds. Amsterdam: Elsevier Science Publishers BV, 603–640.