



TI 2002-045/2

Tinbergen Institute Discussion Paper

Optimal Income Taxation with Endogenous Human Capital

Bas Jacobs

*Department of General Economics, Faculty of Economics and Econometrics, University of
Amsterdam, and CPB Netherlands Bureau for Economic Policy Analysis*

Tinbergen Institute

The Tinbergen Institute is the institute for economic research of the Erasmus Universiteit Rotterdam, Universiteit van Amsterdam and Vrije Universiteit Amsterdam.

Tinbergen Institute Amsterdam

Keizersgracht 482
1017 EG Amsterdam
The Netherlands
Tel.: +31.(0)20.5513500
Fax: +31.(0)20.5513555

Tinbergen Institute Rotterdam

Burg. Oudlaan 50
3062 PA Rotterdam
The Netherlands
Tel.: +31.(0)10.4088900
Fax: +31.(0)10.4089031

Most TI discussion papers can be downloaded at
<http://www.tinbergen.nl>

Optimal Income Taxation with Endogenous Human Capital*

Bas Jacobs[†]

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 earnings potentials is endogenous, because agents differ in ability to learn. Taxation affects utilization rates of human capital as labor supply is affected. Non tax-deductible costs distort the learning decision as well. We show theoretically that the trade-off between efficiency costs of taxation and equity aspects of redistribution is worsened. Quantitative analysis indicates that distortionary costs of taxation increase substantially when learning is endogenous.

Keywords: optimal linear taxation, labor supply, and human capital accumulation.

JEL codes: H21, J22, J24.

*The author is grateful to Lans Bovenberg for support and for providing the opportunity to visit Tilburg University and CentER. The author further thanks Casper van Ewijk, Hessel Oosterbeek, and seminar participants at CentER for comments and suggestions. I also want to thank Edwin Leuven for his valuable help with Gauss. Financial support from the NWO Priority Program 'Scholar' funded by the Netherlands Organization for Sciences is gratefully acknowledged.

[†]First draft: February, 2000. This draft: March, 2002. University of Amsterdam, Tinbergen Institute, NWO Priority Program 'Scholar', and CPB Netherlands Bureau for Economic Policy Analysis. Address: Faculty of Economics and Econometrics, University of Amsterdam, Roetersstraat 11, 1018 WB Amsterdam, The Netherlands. Phone: (+31) - 20 - 525 5088. Fax: (+31) - 20 - 525 4310. E-mail: jacobs@fee.uva.nl.

“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).

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), Sheshinski (1972), and Atkinson and Stiglitz (1980). However, as Rosen (1980) points out, labor supply features many other dimensions. First, labor supply reflects also participation decisions and not only work effort. This implies that, for example, the amount of time people wish to spend in retirement is also a dimension of labor supply. Second, the intensity of work effort may vary from one individual to another. Hours of work may not fully reflect the intensity of work effort, although the latter variable is hard to measure. And, third, effective labor supply increases not only by more hours of work, i.e. the quantity of labor, but also by a higher quality of labor. Differences in levels of education and health of workers are therefore also dimensions of labor supply.

In this paper we analyze optimal income taxation when the learning dimensions of labor supply are taken into account. Human capital theory points out that earnings per hour are the result of investments aimed at augmenting effective labor supply, see e.g. Becker (1964). If human beings have different abilities to accumulate human capital, the distribution of income is endogenously determined by learning decisions of agents, rather than exogenously given.

The first contribution of this paper is to show analytically how optimal linear tax rates are set when human capital accumulation is endogenous. The tax system distorts not only labor supply decisions but also learning decisions. Tax distortions arise in learning decisions due to the direct costs, besides foregone earnings, that are associated with investments in human

capital. These costs originate from the use of market (or capital) goods in the production of human capital, see e.g. Lord (1989), Rebelo (1991), Trostel (1993), Pecorino (1993), Nerlove et al. (1993), Jones et al. (1993, 1997), Stokey and Rebelo (1995), Milesi-Feretti and Roubini (1998), and Judd (1999). Consequently, future earnings are subject to a higher effective rate of tax than total costs of investment because the direct costs remain ‘untaxed’.¹

Furthermore, labor supply, which now encompasses both quantity and quality dimensions, becomes more elastic if learning is endogenous. The more leisure one wishes to consume, the lower are returns on human capital since less time is spent working so that the utilization of human capital falls. The reverse reasoning also holds: the more one learns, the more expensive leisure time becomes as wage rates per hour increase. Learning and leisure decisions are thus interdependent and the distortionary effects of taxation increase, see also Kotlikoff and Summers (1979), Eaton and Rosen (1980), and Driffil and Rosen (1983).²

We show theoretically that the trade-off between equity and efficiency is worsened due to the direct tax distortion on learning decisions and due to the interaction of working and learning decisions. Labor supply - now defined in a broad sense - becomes more elastic with respect to the tax. Consequently, optimal linear taxes are lowered if human capital accumulation is endogenous.

The second contribution of this paper is to provide quantitative evidence on the importance of endogenous learning decisions for the setting of the optimal linear tax schedule. We compute optimal income taxes, along the lines of Stern (1976). 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) and Saez (2001). We derive that optimal tax rates are substantially lower compared to earlier studies when taking into account human capital accumulation. This confirms our theoretical predictions.

This paper is related to some earlier contributions on optimal taxation

¹Even though costs of education are generally highly subsidized, tuition fees and other direct costs are generally not deductible from the income tax.

²We do not allow for a non-linear tax schedule for reasons of analytical and computational tractability. 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 Van Ewijk (1997) and Nielsen and Sørensen (1997).

when agents differ in their levels of education. Atkinson (1973) analyzes optimal linear income taxation in a model where agents decide upon their levels of education. The distortion on learning stems from the fact that foregone income while learning is not subject to the income tax, whereas income derived from working after learning is. In contrast with Atkinson we allow for taxed foregone earnings as one may argue that non-taxed foregone earnings are of limited relevance. Furthermore, Atkinson (1973) does not pay attention to the labor supply decision. Ulph (1977) and Hare and Ulph (1979) study the problem of optimal taxation and education expenditures where redistributive and educational targets are simultaneously optimized. Ulph (1977) allows for endogenous labor supply whereas Hare and Ulph (1979) assume that labor supply is fixed and agents might opt for private education. In both studies, however, the government simply sets the level of education for each agent, so agents do not choose their levels of learning. Taxation does therefore not influence learning decisions.³ 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, 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 numerical examples, Section 5 presents some sensitivity analyzes, and Section 6 concludes.

³Here, the important assumption is made that ability to earn income is not observable by the government, whereas ability to learn is. We assume that both the quantity (hours worked) and the quality (education) of labor supply are not observable by the government, only earned income.

2 Model

We consider a two-period life-cycle model of human capital formation.⁴ 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.⁵ Additionally, there is a perfect capital market and agents can save or borrow to finance costs of education.⁶ 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 ability to learn α . Agents with a higher ability are more efficient in the production of human capital. In addition we assume that ability α also affects wage rates w independently of the amount of learning, so that $w(\alpha)$ where $w' > 0$. The reason for making this assumption is that not all income inequality can be attributed to differences in learning behavior. Consequently, agents with higher ability have a higher wage rate per unit of human capital as well. The distribution of α is denoted by $F(\alpha)$. F has support $[\underline{\alpha}, \infty)$.

⁴Some authors have used multi-period models for analyzing 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 time may be consumed as leisure, or all time 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 due to a change in the level of human capital cancel out. King, Plosser and Rebelo (1988) use 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 human capital.

⁵Given the perfect capital market no important insights are obtained by allowing for first period consumption. Furthermore, allowing for first-period leisure, yields only the standard labor supply distortion and learning decisions are not affected by introducing first period leisure time. Therefore, inter-action effects with learning are absent.

⁶See Jacobs (2001) for the consequences of imperfect capital markets for the optimal taxation of income.

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 its life. A fraction x of total time in the first period is spend on education. The rest, $1 - x$, is devoted to working, where the total time endowment is normalized at unity. Education requires, besides time, y market goods per year of education.⁷ ϕ is the production function for human capital with positive but diminishing returns time x , and goods y invested in education:

$$\phi(\alpha; x_\alpha, y_\alpha) \equiv h(\alpha)x_\alpha^\gamma y_\alpha^v, \quad (1)$$

where we have $h'(\alpha) > 0$. 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 to ensure an interior solution with a fixed real interest rate. In the remainder we restrict the analysis to a Cobb-Douglas production function with constant elasticities γ and v , and $\gamma + v < 1$. 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. Ben-Porath (1967), Weiss (1986), Trostel (1993).

Income derived from working equals $(1 - t)w(\alpha)(1 - x)$, where t is the flat labor income tax rate. 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 assumption that one cannot observe the wage rate and hours worked. So 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 an uniform non-individualized lump-sum income transfer g in both periods of its life. With positive transfers g the tax system is progressive since the average tax rate increases with income. Savings s can be made to smooth consumption over time. Savings equal total first period income minus the direct costs of education py . p denotes the unit costs of direct expenditures on education. The first period budget constraint

⁷Rather than using market goods in the production of human capital one may also use capital goods in the production function. This yields qualitatively similar results.

is therefore given by:

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

In the second period, human capital is supplied endogenously to the labor market. Total time spend 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 accumulation of financial assets is $(1 + r)s_\alpha$ where r is the constant real interest rate. In the remainder 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.⁸ The consumption price is chosen as the numéraire. Hence, the second period budget constraint is:

$$c_\alpha = (1 - t)w(\alpha)l_\alpha\phi(\alpha; x_\alpha, y_\alpha) + s_\alpha + g. \quad (3)$$

We restrict the analysis to an iso-elastic utility function.⁹ Utility u is given by:

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

where $\varepsilon > 0$ is 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). The analytical convenience is twofold. First, there are no income effects on labor supply. Second, the elasticity of labor supply is constant and equal for all households. This allows us to aggregate easily over all households. If labor supply is exogenous ($\varepsilon = 0$), then the pure human capital model applies, and agents simply maximize life-time earnings. It is assumed that agents do not derive utility from having human capital.¹⁰

⁸See Nielsen and Sørensen and (1997) for an analysis on optimal dual income taxation with endogenous learning with homogeneous agents. Bovenberg and Jacobs (2001) analyze optimal dual income taxation in a similar model as presented in this paper.

⁹We were able to derive a general characterization of the solution with a general utility function. This turns out to give very similar results, although the different roles played by the tax distortions cannot be easily traced analytically.

¹⁰Human capital can also be regarded as a consumption good, see Lazear (1977). Additionally, having more human capital can enhance the effective productivity of leisure in utility. This notion stems from Becker (1965) and has been applied to human capital theory first by Heckman (1976). Both elements can be incorporated. However, this is likely to yield untractable results unless we impose strong restrictions on preferences.

Agents maximize utility by choosing consumption c , labor supply l , the optimal amount of learning x and the goods invested in education y , subject to their budget constraints, and the production function of human capital. Manipulation of the first-order conditions gives the following labor supply function - omitting the agent's indices α :¹¹

$$l = [(1 - t)w\phi(.)]^\varepsilon. \quad (5)$$

The higher the hourly wage rate, the larger is labor supply. From the last equation can be seen that consumption and investment decisions cannot be separated. Learning increases the hourly wage rate and thereby increases labor supply. If one views leisure time as years in retirement, our model can explain the relatively higher participation rates of older workers with more education. The reason is that 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:

$$\frac{\phi_x}{\phi_y} = \frac{\gamma y}{vx} = \frac{(1 - t)w}{p}. \quad (6)$$

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.

$$wl\phi_x = \frac{(1 - t)wl\phi_y}{p} = 1. \quad (7)$$

This equation pegs the amount of time and goods invested in human capital accumulation. Arbitrage between financial and 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 to human capital investments, until rates of return are equalized as a consequence of diminishing returns in human capital accumulation.

Since costs of education are not tax deductible, taxes *directly* distort investments in human capital. A higher tax rate reduces the optimal amount

¹¹The appendix to this paper contains all derivations.

of goods invested in human capital, thereby lowering the productivity of time invested. Investments in human capital fall accordingly. Loosely speaking, foregone earnings and goods are less affected by changes in the tax rate, than the returns, i.e. future earnings. If goods were fully tax deductible, the tax rate would have no direct effect on investments in human capital since costs and returns are equally affected 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. The reason is that the effective utilization rate of human capital decreases so that returns on investments in human capital are lowered.

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:

$$\mu \equiv (1 + \varepsilon)(\gamma + v) < 1. \quad (8)$$

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 not too high. Intuitively, if more time is spend learning, wage rates per hour increase and substitution towards more labor supply is induced. This, in turn, increases returns to investments in human capital so that more time is spend learning, and so on. Due to this interaction between learning and leisure decisions sufficiently decreasing returns to investments in human capital (low γ and v) or sufficiently decreasing marginal utility of leisure (low ε) should guarantee that an interior solution is attained and 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:

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

$$y^* = \frac{v w (1-t)}{\gamma p} x^*,$$

$$l^* = h(\alpha)^\varepsilon (1-t)^{\varepsilon(1+v)} \left(\frac{v}{\gamma} \right)^{\varepsilon v} x^{*\varepsilon(\gamma+v)}.$$

Since $w'(\alpha) > 0$ and $h'(\alpha) > 0$ we derive that agents with higher ability invest more time and goods in human capital accumulation, i.e. $\partial x/\partial\alpha > 0$, $\partial y/\partial\alpha > 0$ and $\partial l/\partial\alpha > 0$ 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 , y and l w.r.t. t are constant and given by:

$$\varepsilon_{xt} \equiv -\frac{\partial x}{\partial t} \frac{(1-t)}{x} = \frac{\varepsilon + v(1+\varepsilon)}{1 - (1+\varepsilon)(\gamma+v)} > 0, \quad (9)$$

$$\varepsilon_{yt} \equiv -\frac{\partial y}{\partial t} \frac{(1-t)}{y} = \frac{(1+\varepsilon)(1-\gamma)}{1 - (1+\varepsilon)(\gamma+v)} > 0, \quad (10)$$

$$\varepsilon_{lt} \equiv -\frac{\partial l}{\partial t} \frac{(1-t)}{l} = \frac{\varepsilon(1-\gamma)}{1 - (1+\varepsilon)(\gamma+v)} > 0. \quad (11)$$

These are useful properties later on.

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:

$$t \int_{\alpha}^{\infty} H_{\alpha} dF(\alpha) = \Lambda + G, \quad (12)$$

where $G \equiv 2g$ and $H \equiv w(\alpha)(1-x_{\alpha}) + w(\alpha)l_{\alpha}\phi(\alpha; x_{\alpha}, y_{\alpha})$ is the gross life-time value of human capital. 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 rates t and lump-sum transfers G are chosen so as to maximize a social welfare function Γ :¹²

$$\Gamma = \int_{\alpha}^{\infty} \Psi(V_{\alpha}) dF(\alpha), \quad \Psi' > 0, \quad \Psi'' \leq 0, \quad (13)$$

¹²We abstract from issues dealing with the dynamic consistency of the tax policies and simply assume that the government can pre-commit. However, in models like the one discussed here the government has always the incentive to renege on its announcement to set a particular tax rate after the investments in human capital are made. The reason is that human capital is accumulated and 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 for example Fischer (1980) and Boadway, Marceau, and Marchand (1996).

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).¹³

Form the first-order conditions for the lump-sum element G we derive the net social marginal valuation of income in terms of government revenue, b , see Atkinson and Stiglitz (1980):¹⁴

$$b_\alpha \equiv \frac{\Psi' \lambda_\alpha}{\eta}, \quad (14)$$

where λ is the private marginal utility of life-time income and η is the Lagrange multiplier associated with the government budget constraint. The term on the right-hand side denotes the direct social value of redistribution to household α . We find that the average social value of income averaged over all households is given by:

$$\bar{b} = 1, \quad (15)$$

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

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 (1976):

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

The term in brackets is the normalized covariance of human capital¹⁵ and the marginal social valuation of income. $\bar{H} \equiv \int_{\underline{\alpha}}^{\infty} H_\alpha dF(\alpha)$ stands for the average supply of human capital. ξ can be interpreted as a ‘marginal measure of inequality’, see Atkinson and Stiglitz (1980).

¹³We assume in the theoretical derivations that the lump-sum transfer is never larger than income derived from supply of human capital, i.e. $G < H$. This constraint precludes that agents voluntarily decide to be unemployed. This constraint is always non-binding in the numerical calculations.

¹⁴We used Roy’s lemma in the derivations: $\partial V / \partial G = \lambda$, $\partial V / \partial t = -\lambda H$.

¹⁵This can be seen by noting that $\xi = \frac{-1}{\bar{H}\bar{b}} \left(\int_{\underline{\alpha}}^{\infty} H b dF - \int_{\underline{\alpha}}^{\infty} H dF \int_{\underline{\alpha}}^{\infty} b dF \right) = \frac{-cov(H,b)}{\bar{H}\bar{b}}$.

The distributional characteristic is positive ($\xi > 0$), because the normalized covariance is negative. If human capital H increases, the marginal social value of income b decreases as a consequence of the diminishing marginal utility of income, and due to the fact that a larger weight is attached to poorer agents ($\Psi'' < 0$). i.e. it becomes socially less efficient to redistribute resources to richer agents, given that the social welfare function features diminishing marginal social welfare in utility of the agents. As the marginal social valuation of income decreases as earnings increase, the term in brackets is negative.

If the government was not interested in redistribution, every agent has the same social value of income: $b = \bar{b}$, which yields $\xi = 0$ in that case. The distributional characteristic increases if incomes become more unevenly distributed, or if the government has a larger preference for income equality.

Straightforward manipulation gives the optimal tax on labor income:

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

where $\omega \equiv \int_{\alpha}^{\infty} wl\phi(\cdot)dF / \int_{\alpha}^{\infty} wl\phi(\cdot) + w(1-x)dF$ is the ratio of average second period income in average total income. The optimum tax formula clearly shows the trade-off between equity and efficiency considerations.¹⁶

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, the optimal tax rate on human capital is zero. The reason is that every agent invests the same amount of resources in human capital accumulation. H is therefore identical for all agents and there is no income inequality. Consequently, the distributional characteristic $\xi = 0$ and the optimal tax rate is zero.

Second, the numerator 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

¹⁶This last formula is a similar way of expressing the optimal linear income tax as in Dixit and Sandmo (1977). They use the non-normalized covariance to obtain expressions that have distributional concerns in the numerator and efficiency costs of taxation in the denominator. Tuomala (1985) uses the government budget constraint to obtain an optimal tax formula where equity considerations enter in numerator the and efficiency costs are captured by the denominator.

optimal tax rate on labor income should be lower if the elasticity of labor supply is larger.

However, from the definition of the labor supply elasticity we can see that the ‘true’ wage elasticity of labor supply - including the learning effects - is larger than the ‘simple’ elasticity of labor supply ($\varepsilon_{tt} > \varepsilon$) that would enter in the optimum tax formula in the absence of learning decisions, see for example Atkinson (1995):

$$\frac{\varepsilon_{tt}}{\varepsilon} = \frac{1 - (\gamma + v)}{1 - (1 + \varepsilon)(\gamma + v)} + \frac{v}{1 - (1 + \varepsilon)(\gamma + v)} > 1. \quad (18)$$

The first term in brackets is larger than 1 and the second term is positive. The first term measures the interaction impact of learning and working decisions and the second term measures the additional impact of the non-deductibility of goods invested in education. Clearly, the interaction between learning and labor supply decisions makes the labor supply response more elastic and is driving the optimal tax rate downwards.

The second term in the denominator of the tax formula $v\varepsilon_{yt}$ captures the tax distortion associated with the non-deductibility of education expenditures. The optimal tax should be lower if the tax-elasticity of goods invested in education is larger ε_{yt} . The more elastic learning behavior responds to the tax the lower should be the optimal tax.

It is easily seen that the tax elasticity of goods invested in education is magnified by the elasticity of labor supply ε . Suppose that labor supply was inelastic ($\varepsilon = 0$) then we have from our definition of ε_{yt} :

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

Due to the inter action effect we see that the size of the tax elasticity of learning is higher and the optimum tax should be lower accordingly.

From both definitions of the elasticities it is easily established that the optimum tax should decrease if either one of the elasticities γ , v , or ε increases. These effects are in conformity with standard Ramsey intuition.

The elasticities are weighed with the share of second period income in total life-time income ω . The larger is second period income, the more elastic total life-time income gets and the lower optimal linear taxes should be.

If goods invested were fully deductible, the optimum tax is given by:¹⁷

$$\frac{t}{1-t} = \frac{\xi}{\omega\varepsilon_{lt}}. \quad (20)$$

The difference between the denominators in the optimum tax formulae with and without deductible goods is the term associated with tax distortion of non-deductible goods $v\varepsilon_{yt}$. This increases the optimum income tax as tax distortions associated with redistribution are smaller. Moreover, the elasticity of labor supply is lowered:

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

Still the inter-action effect between learning and working remains. If the elasticity of labor supply is zero as well, the optimum tax goes to infinity (the denominator goes to zero) since all distortions are eliminated then.

Based on the last formulae one can get a quantitative idea on the increase in the size of the elasticities when learning is endogenous. Suppose that $\gamma + v = .6$. These are the values suggested by Trostel (1993). Let the elasticity of labor supply be equal to $\varepsilon = .25$ which is not an uncommon figure in the literature, see also below. Then we find that the elasticity of broad labor supply is equal to $\varepsilon_{lt} = .4$. In other words, the ‘true’ elasticity is about 60% larger than the simple elasticity of labor supply. Now, suppose that the simple elasticity of labor supply is $\varepsilon = .5$, 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 labor supply and 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 is assumed to be normal with mean μ_α , and standard deviation σ_α :

$$\alpha \sim N[\mu_\alpha; \sigma_\alpha]. \quad (22)$$

¹⁷This 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.

Ability has a mean $\mu_\alpha = -1$ which is a normalization. Wage rates are assumed to be generated by following exponential wage equation:

$$w(\alpha) = \exp(\alpha), \quad (23)$$

so that a log-normal wage distribution results see also Mirrlees (1971), Stern (1976) and Tuomala (1990). The standard deviation of log wages in these papers is set at .39.¹⁸

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

$$h(\alpha) = A \exp(\alpha)^\psi. \quad (24)$$

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-normally distributed 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) for a very extensive discussion of plausible parameter values. The share of time in production of human capital is set at $\gamma = .3$ and the share of goods in production of human capital is set at $v = .1$. So total returns to private inputs are .4.¹⁹ Here, Trostel (1993) uses the values of $\gamma = .45$ and $v = .15$. However, these high values turn out to give occasional problems with the second-order conditions, see also equation (8).

The values $\gamma = .3$ and $v = .1$ imply that direct costs of education are 1/4 of total expenditures in education, so that foregone earnings make up 3/4 of total costs of education. Becker (1964) and Boskin (1975) find that the private costs shares of time and goods invested in education are 3/4 and 1/4.²⁰

¹⁸We 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 a relatively large cost in terms of computation time.

¹⁹Davies and Whalley (1989) use a model with only time as an input in human capital where returns to private inputs ($\gamma + v$) are .5. Heckman (1976) and Haley (1976) find that returns to private inputs are approximately .55. Rosen (1976) finds results that yield private returns of .65, see also the references in Lucas (1990). Lucas (1990) uses private returns to human capital accumulation equal to .8.

²⁰The price of direct costs of education is arbitrarily set at $p = .5$.

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

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

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):

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

The elasticity of substitution between second period consumption and leisure equals $\sigma \equiv 1/(1-\zeta)$. We follow common practice by setting $\sigma = .5$ in the base case scenario. Stern (1976) uses a value of $\sigma = .4$ and Tuomala (1990) uses $\sigma = .5$ based on reviewing the literature, see e.g. Ashenfelter and Heckman (1973).²¹

Recently, Atkinson (1990, 1995), Diamond (1998) and Saez (2001) have adopted the iso-elastic utility function that was used in the theoretical derivation. This serves as the basis of our second specification:

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

where we added a parameter θ denoting the preference for leisure. We set the uncompensated wage elasticity of labor supply at $\varepsilon = .25$. The uncompensated elasticity of .25 is in the middle of elasticities for men and women that are encountered in the micro-econometric literature. For men, the elasticity is slightly below 0, whereas significantly higher elasticities, ranging from .5 to 1, are reported for women, see Pencavel (1986), Killingsworth and Heckman (1986) or Hansson and Stuart (1985). An average value of .1 is found

²¹Again, 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. For the CES utility function this amounts to: $[\psi(1-\gamma-v) - \sigma(\gamma+v)]^{-1} > 0$, and $\psi \equiv \frac{wl\phi(\cdot)}{c} + \frac{l}{1-l} > 0$, see also Jacobs (2000).

in the latter study on the basis of reviewing the literature. In exercises with dynamic growth models an uncompensated elasticity of .2 is commonly assumed, see Lucas (1990), Stokey and Rebelo (1995), and Hendricks (1999). 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 are: the common learning technology parameter A , the leisure share parameter β (θ), the elasticity of ability in learning ψ and the standard deviation of ability σ_α . We impose four identifying conditions on the model at $t = 0$ and $G = 0$: mean working time is $1 - l = .67$, mean learning time is $x = .67$, the standard deviation of learning time equals $\sigma_x = .12$, the standard deviation of the log of total income is .40.

The value of mean working time is taken from Stern (1976) and Tuomala (1990). This implies that the average individual would work 2/3 of the day. It could also correspond to a retirement period of 10 years if one regards each period in life as during approximately 30 years.

A mean learning time of .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 if one compares this with average time spend on formal education. Harmon and Walker (1999) find that the mean is 11.90 years for the UK in the General Household Survey 1974-1994. Ashenfelter and Krueger (1994) report an average of 13.1 schooling years of the US 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 total human capital formation is on the job. Computations by Heckman et al. (1998) suggest that the contribution by OJT is lower and in the range of one quarter of total human capital formation. If we assume that approximately 1/3 of human capital formation is OJT, and 2/3 is formal education, then a mean of $(2/3)^2 * 30 = 13.3$ years of formal schooling results. This corresponds with figures from reality.

If we proxy the learning distribution with a normal distribution, then we are able to compute the spread in learning outcomes in years of formal education from the model. Under the assumption that two thirds of human capital is acquired through formal education and each period takes 30 years, a standard deviation of .12 corresponds to a standard deviation in

learning time equal to 2.4 years. Harmon and Walker (1999) find that the standard deviation equals 2.83 years. Ashenfelter and Krueger (1994) report a standard deviation of 2.7 years.

The standard deviation of log incomes is calibrated at .4, since the distribution of incomes is endogenous. In the models of optimum income taxation without learning behavior similar inequality is assumed. Mirrlees (1971), Stern (1976) and Tuomala (1990) use a standard deviation of log wages of .39.

The calibration with the CES function yielded a productivity parameter $A = 7.4$, a preference for leisure parameter $\beta = .7$, a standard deviation of ability of $\sigma_\alpha = .31$, and a value of the elasticity of ability $\psi = .5$. The calibration with the alternative specification yielded values of $A = 4.4$, $\theta = 5.7$, $\psi = 0$, and $\sigma_\alpha = .30$.

We compare the outcomes of the two models in the simulations. A feature of the two utility functions used here is that 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 non-labor 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 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.²²

We derived optimal tax rates in the case where both learning and leisure are endogenous and for the case where 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. Table 1 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 $\sigma = .5$. The corresponding value of the optimum tax rate when learning is exogenous equals 29.8%. The CELS utility function with base case value of $\varepsilon = .25$ gives values of 16.1% and 28.2% when learning is endogenous and when learning is exogenous respectively. Clearly

²²Moreover, the theoretical models cannot be consistently matched with the empirical literature. A value of the elasticity of substitution between consumption and leisure smaller than one cannot be reconciled with an upward sloping labor supply curve (if non-labor income is zero). In the remainder we proceed by analyzing the two cases separately.

optimum taxes are much 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. An assuring aspect of our computations is that very similar results are obtained when using the CES and CELS functions.

These tax rates are also lower than the optimal marginal tax rates that are reported in the literature. Our findings of optimal linear taxes are always lower than the ones obtained by Stern (1976) with the CES utility function for various elasticities of substitution. Saez (2001) found marginal rates far above 50% in the model with a CELS utility function. However, set the revenue requirement by the government at .25 of production. For the sake of comparison we have computed the optimum rates with this revenue requirement ($\Lambda = .23$). For $\varepsilon = .25$ we derive an optimum tax $t = 21.0\%$ and for $\varepsilon = .50$ we find $t = 22.3\%$. These optimum taxes are substantially lower than the ones from Saez (2001).

Using a CES utility function, Tuomala (1990, p.98) found optimal non-linear marginal tax rates ranging from 65% in 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 $\sigma = .5$. The optimal marginal tax rate in our model is 17.4% at $\sigma = .5$ which is again a considerably lower marginal tax rate. Using CELS utility functions, Diamond (1998) and Saez (2001) find in non-linear versions of their models marginal tax rates on income that are generally higher than 50%, even for the top deciles. The reduction in the optimal marginal tax-rates in our model where earnings potentials are endogenous is quite striking.

In the appendix we show robustness checks for various modifications to technology, preferences or government parameters. The result that optimum taxes are lower with endogenous human capital are *not* sensitive to the parameters used in the model.

5 Conclusion

This paper augmented the theory of optimal income taxation with the analysis of optimal taxation of human capital. To that end, a two period life-cycle model of human capital accumulation, consumption and saving is analyzed. Agents differ in their ability to earn income and to learn. This makes the dis-

Table 1: Optimal tax rates (%)

	CES			CELS	
	Endogenous	Exogenous		Endogenous	Exogenous
$\sigma = .2$	24.8	50.9	$\varepsilon = .1$	20.5	41.0
$\sigma = .3$	21.3	40.5	$\varepsilon = .2$	17.1	31.1
$\sigma = .4$	19.0	34.1	$\varepsilon = .25$	16.1	28.2
$\sigma = .5$	17.4	29.8	$\varepsilon = .3$	15.3	26.0
$\sigma = .6$	16.2	26.7	$\varepsilon = .4$	14.2	22.6
$\sigma = .7$	15.3	24.5	$\varepsilon = .5$	13.6	19.9
$\sigma = .8$	14.6	22.6			

tribution of earnings potentials endogenous. Taxation does not only distort the decision to supply hours of work, but also learning decisions. As such, labor supply has both a quantity dimension - hours worked - and a quality dimension - years spend on education.

Two tax distortions on learning are present. First, taxation affects demand for leisure. Therefore, agents invest less in their human capital if they consume more leisure since the returns of their investments have fallen. Second, costs of education are not tax deductible, so that costs of learning (foregone wages and direct expenditures) are more affected than benefits of learning (future wages). We derived a simple optimal tax formula that shows the trade-off between efficiency costs of taxation and equity aspects of redistribution.

Quantitative analysis showed that distortions in labor supply may substantially increase when learning decisions are endogenous. Our results suggest that optimum marginal tax rates are significantly reduced when the learning dimensions of labor supply are taken into account. As such, these results vindicate Rosen's (1980) idea that other dimensions of labor supply matter. The analysis therefore underpins concerns with adverse incentives that are associated with redistributive policies (e.g. the EITC) not only on labor supply but also on learning efforts. In addition, many Western countries face the consequences of a greying population. The results presented here can also be interpreted as a case for lowering marginal taxes on labor incomes in order to stimulate the accumulation of human capital and to reduce adverse incentives to retire early.

References

- Ashenfelter, Orley, and James J. Heckman (1973). "Estimating Labor-Supply Functions", in: Cain, G.G. and H.W. Watts (eds) (1973). *Income Maintenance and Labor Supply*, Chicago: Rand McNally.
- Ashenfelter, Orley, and Alan Krueger (1994). "Estimates of the Economic Return to Schooling from a New Sample of Twins", *American Economic Review*, 84, (5), 1157-1173.
- Atkinson, Anthony B. (1973). "How Progressive Should the Income Tax Be?", in Parkin, M., and A.R. Nobay (eds). *Essays in Modern Economics*, London: Longman.
- Atkinson, Anthony B. (1990). "Public Economics and the Economic Public", *European Economic Review*, 34, (2-3), 225-248.
- Atkinson, Anthony B. (1995). *Public Economics in Action. The Basic Income/Flat Tax Proposal*, Oxford: Oxford University Press.
- Atkinson, Anthony B., and Joseph E. Stiglitz (1976). "The Design of Tax Structure: Direct versus Indirect Taxation", *Journal of Public Economics*, 6, 55-75.
- Atkinson, Anthony B., and Joseph E. Stiglitz (1980). *Lectures on Public Economics*, New York: McGraw-Hill.
- Becker, Gary S. (1965). "A Theory of the Allocation of Time", *Economic Journal*, 75, 493-517.
- Becker, Gary S. (1964). *Human Capital: A Theoretical and Empirical Analysis with Special Reference to Education*, Third edition 1993, Chicago: Chicago University Press.
- Ben-Porath, Yoram (1967). "The production of Human Capital and the Life Cycle of Earnings", *Journal of Political Economy*, 75, (4), 352-365.
- Boadway, Robin, Nicolas Marceau, and Maurice Marchand (1996). "Investment in Education and the Time Inconsistency of Redistributive Tax Policy", *Economica*, 63, 171-189

- Boskin, Michael (1975). "Notes on the Tax Treatment of Human Capital", NBER Workingpaper 116.
- Bovenberg, A. Lans, and Casper van Ewijk (1997). "Progressive Taxes, Equity, and Human Capital Accumulation in an Endogenous Growth Model with Overlapping Generations", *Journal of Public Economics*, 64, 154-179.
- Bovenberg, A. Lans, and Bas Jacobs (2001). "Redistribution and Education Subsidies are Siamese Twins", CEPR Discussion Paper No. 3309.
- Davies, James, and John Whalley (1989). "Taxes and Human Capital Formation: How Important is Human Capital?", NBER Working Paper 2899.
- Diamond, Peter A. (1998). "Optimal Income Taxation: An Example with a U-Shaped Pattern of Optimal Marginal Tax Rates", *American Economic Review*, 88 (1), 83-95.
- Dixit, Avinash K., and Agnar Sandmo (1977). "Some Simplified Formulae for Optimal Income Taxation", *Scandinavian Journal of Economics*, 79, 417-423.
- Driffil, E. John, and Harvey S. Rosen (1983). "Taxation and Excess Burden: A Life Cycle Perspective", *International Economic Review*, 3, 671-683.
- Eaton, Jonathan and Harvey S. Rosen (1980). "Taxation, Human Capital and Uncertainty", *American Economic Review*, 70, (4), 705-715.
- Fischer, Stanley (1980). "Dynamic Inconsistency, Cooperation, and the Benevolent Dissembling Government", *Journal of Economic Dynamics and Control*, 2, 93-107.
- Haley, William J. (1976). "Estimation of the Earningsprofile from Optimal Human Capital Accumulation", *Econometrica*, 44, 1223-1238.
- Hansson, Ingemar, and Charles Stuart (1985). "Tax Revenue and the Marginal Costs of Public Funds in Sweden", *Journal of Public Economics*, 27, 331-353.
- Hare, Paul G., and David T. Ulph (1979). "On Education and Distribution", *Journal of Political Economy*, 87 (5), S193-S212.

- Harmon, Colm, and Ian Walker (1999). "The Marginal and Average Returns to Schooling", *European Economic Review*, 43, 879-887.
- Heckman, James J. (1976). "A Life-Cycle Model of Earnings, Learning and Consumption", *Journal of Political Economy*, 4, S11-S44.
- Heckman, James J., Lance Lochner, and Christopher 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.
- Hendricks, Lutz (1999). "Taxation and Long-Run Growth", *Journal of Monetary Economics*, 43, 411-434.
- Jacobs, Bas (2000). "A Note on Taxation and Human Capital Formation", mimeo: University of Amsterdam.
- Jacobs, Bas (2001). "Optimal Taxation of Human Capital and Credit Constraints", mimeo: University of Amsterdam.
- 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. (1985). "Redistributive Taxation in a Simple Perfect Foresight Model", *Journal of Public Economics*, 28, 59-83.
- Judd, Kenneth L. (1999). "Optimal Taxation and Spending in General Competitive Growth Models", *Journal of Public Economics*, 71, 1-26.
- Killingsworth, Mark R., and James J. Heckman (1986). "Female labor Supply: A Survey", in: Ashenfelter, Orley and Richard Layard (eds) (1986). *Handbook of Labor Economics - Vol. I*, Amsterdam: Elsevier Science
- King, Robert G., Charles I. Plosser, and Sergio Rebelo (1988), "Production, Growth and Business Cycles: I. The Basic Neoclassical Model", *Journal of Monetary Economics*, 21, 195-232.
- 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.

- Lazear, Edward P. (1977). "Education: Consumption or Production?", *Journal of Political Economy*, 85, 569-597.
- Lord, William (1989). "The Transition from Payroll to Consumption Receipts with Endogenous Human Capital", *Journal of Public Economics*, 38, 53-73.
- Lucas, Robert E. Jr. (1990). "Supply-Side Economics: An Analytical Review", *Oxford Economic Papers*, 42, 293-316.
- Milesi-Ferretti, Gian Maria, and Nouriel Roubini (1998). "On the Taxation of Human and Physical Capital in Models of Endogenous Growth", *Journal of Public Economics*, 70, 237-254.
- Mincer, Jacob (1962). "On the Job Training: Costs, Returns and Some Implications", *Journal of Political Economy*, 70, 50-79.
- Mirrlees, James A. (1971). "An Exploration in the Theory of Optimum Income Taxation", *Review of Economic Studies*, 38, 175-208.
- Nerlove, Marc, Assaf Razin, Efraim Sadka, Robert K. von Weizsäcker (1993). "Comprehensive Income Taxation, Investments in Human and Physical Capital, and Productivity", *Journal of Public Economics*, 50, 397-406.
- 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.
- Pecorino, Paul (1993). "Tax Structure and Growth in a Model with Human Capital", *Journal of Public Economics*, 52, 251-271.
- Pencavel, John (1986). "Labor Supply of Men: A Survey", in: Ashenfelter, Orley, and Richard Layard (eds) (1986). *Handbook of Labor Economics - Vol I*. Amsterdam: Elsevier Science Publishers BV.
- Rebelo, Sergio (1991). "Long-run Policy Analysis", *Journal of Political Economy*, 99, 500-521.
- Rosen, Harvey S. (1980). "What is Labor Supply and Do Taxes Affect It?", *American Economic Association - Papers and Proceedings*, 70 (2), 171-176.

- Saez, Emanuel (2001). “Using Elasticities to Derive Optimal Income Tax Rates”, *Review of Economic Studies*, 68, 205-229.
- Sheshinski, Eytan (1972). “The Optimal Linear Income Tax”, *Review of Economic Studies*, 39, 297-302.
- Stern, Nicholas H. (1976). “On the Specification of Models of Optimum Income Taxation”, *Journal of Public Economics*, 6, 123-162.
- Stokey, Nancy L., and Sergio T. Rebelo (1995). “Growth Effects of Flat-Rate Taxes”, *Journal of Political Economy*, 103, 519-550.
- Trostel, Philip A. (1993). “The Effect of Taxation on Human Capital”, *Journal of Political Economy*, 101, 327-350.
- Tuomala, Matti (1985). “Simplified Formulae for Optimal Linear Income Taxation”, *Scandinavian Journal of Economics*, 87 (4), 668-672.
- Tuomala, Matti (1986). “On the Optimal Income Taxation and Educational Expenditures”, *Journal of Public Economics*, 30, 183-198.
- Tuomala, Matti (1990). *Optimal Income Tax and Redistribution*, Oxford: Clarendon Press.
- Ulph, David T. (1977). “On the Optimal Distribution of Income and Educational Expenditure”, *Journal of Public Economics*, 8, 341-356.
- Weiss, Yoram. (1986). “The Theory of Life-Cycle Earnings”, in: Ashenfelter, Orley and Richard Layard (1986). *Handbook of Labor Economics*. Amsterdam: Elsevier Science Publishers BV.

Appendix

Household optimization

Consolidating the household budget constraint yields:

$$c = (1 - t)w(1 - x) - py + (1 - t)wl\phi(.) + 2g.$$

Substitution of the household budget constraint in the utility function yields and unconstrained maximization problem:

$$\max_{\{l,x,y\}} u = \ln \left((1-t)w(1-x) - py + (1-t)wl\phi(\cdot) + 2g - \frac{l^{1+1/\varepsilon}}{1+1/\varepsilon} \right).$$

This is equivalent to:

$$\max_{\{l,x,y\}} u^* = (1-t)w(1-x) - py + (1-t)wl\phi(\cdot) + 2g - \frac{l^{1+1/\varepsilon}}{1+1/\varepsilon}.$$

First-order conditions are:

$$\begin{aligned} \frac{\partial u^*}{\partial l} &= (1-t)w\phi(\cdot) - l^{1/\varepsilon} = 0, \\ \frac{\partial u^*}{\partial x} &= (1-t)wl\phi_x(\cdot) - (1-t)w = 0, \\ \frac{\partial u^*}{\partial y} &= (1-t)wl\phi_y(\cdot) - p = 0. \end{aligned}$$

Rewriting yields:

$$\begin{aligned} l &= [(1-t)w\phi(\cdot)]^\varepsilon, \\ \frac{\phi_x}{\phi_y} &= \frac{\gamma y}{vx} = \frac{(1-t)w}{p}, \\ l\phi_x &= 1. \end{aligned}$$

We can solve for the optimal values of l , x , and y . First, use the marginal rate of technical substitution for goods and time invested in education to get:

$$y = \frac{vw(1-t)}{\gamma p}x.$$

Substitute the last result in the equation for labor supply then we get l as a function of x only:

$$l = \left(\frac{v}{p\gamma} \right)^{\varepsilon v} h(\alpha)^\varepsilon (w(\alpha)(1-t))^{\varepsilon(1+v)} x^{\varepsilon(\gamma+v)}.$$

Second, we have:

$$\gamma l\phi(\cdot) = x,$$

which follows from the arbitrage condition. Use the expression for l to substitute l out and to arrive at the equation for x :

$$x^* = \gamma^{\frac{1}{\mu}} h(\alpha)^{\frac{1+\varepsilon}{\mu}} w(\alpha)^{\frac{\varepsilon+v(1+\varepsilon)}{\mu}} \left(\frac{v}{\gamma p} \right)^{\frac{v(1+\varepsilon)}{\mu}} (1-t)^{\frac{\varepsilon+v(1+\varepsilon)}{\mu}}.$$

y^* and l^* follow from plugging the value for x^* into the equations for y and l above.

To check the second-order conditions we first derive the utility function as a function of x only. Then, we evaluate the second derivative of the utility function at the optimum. If this second derivative is negative we know that utility reaches a maximum in (x, y, l) space and we do not encounter a saddle-point, since optimum values of y and l are positive transformations of x . Substitution of the optimal values of y and l yields utility as a function of x only:

$$v^* = -(1-t)w(1-x) - p\Delta x + \frac{1}{1+\varepsilon} [(1-t)w\Phi(x)]^{1+\varepsilon} + 2g.$$

Where we used $y = \frac{v(1-t)w}{\gamma p} x \equiv \Delta x$, and $\frac{1}{1+1/\varepsilon} l^{1+1/\varepsilon} = \frac{1}{1+1/\varepsilon} ((1-t)w\phi(\cdot))^{1+\varepsilon}$, $\Phi(x) \equiv \phi(x, \Delta x)$. The second derivative of the utility function at the optimum values for y and l is:

$$\frac{\partial^2 v^*}{\partial x^2} = ((1-t)w\Phi(x))^\varepsilon \frac{((1-t)w)^2}{1+\varepsilon} \left(\varepsilon \frac{\Phi_x^2}{\Phi} + \Phi_{xx} \right).$$

For utility to reach a maximum this must be negative, i.e.

$$\varepsilon \frac{\Phi_x^2}{\Phi} + \Phi_{xx} < 0.$$

since all other terms are positive. Next use the properties of Φ :

$$\begin{aligned} \Phi &= h(\alpha) \Delta^v x^{\gamma+v}, \\ \Phi_x &= h(\alpha) (\gamma+v) \Delta^v x^{\gamma+v-1}, \\ \Phi_{xx} &= h(\alpha) (\gamma+v)(\gamma+v-1) \Delta^v x^{\gamma+v-2}. \end{aligned}$$

Upon substitution in the last equality we derive:

$$(1+\varepsilon)(\gamma+v) < 1.$$

I.e. the elasticity of labor supply must not be too high and the private returns to inputs must not be too high.

Note further that the elasticities of x and y w.r.t. t are constant, $\varepsilon_{xt} \equiv -\frac{\partial x}{\partial t} \frac{(1-t)}{x} = \frac{\varepsilon + v(1+\varepsilon)}{1-(1+\varepsilon)(\gamma+v)} > 0$, $\varepsilon_{yt} \equiv -\frac{\partial y}{\partial t} \frac{(1-t)}{y} = \frac{(1+\varepsilon)(1-\gamma)}{1-(1+\varepsilon)(\gamma+v)} > 0$. The elasticity of labor supply is constant as well: $\varepsilon_{lt} \equiv -\frac{\partial l}{\partial t} \frac{(1-t)}{l} = \frac{\varepsilon(1-\gamma)}{1-(1+\varepsilon)(\gamma+v)} > 0$. (If goods are tax-deductible we would have the elasticities: $\varepsilon_{xt} = \varepsilon_{yt} = \frac{\varepsilon}{1-(1+\varepsilon)(\gamma+v)} > 0$ and $\varepsilon_{lt} \equiv \frac{\varepsilon(1-\gamma-v)}{1-(1+\varepsilon)(\gamma+v)} > 0$).

Derivation optimum tax rate

The Lagrangian for maximization of social welfare is given by:

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

where η is the Lagrange multiplier associated with the government budget constraint. First we have:

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

Note that there are no income effects on broad labor supply so that $\frac{\partial H}{\partial G} = 0$. This equation can be rewritten using the definition of b :

$$\int_{\underline{\alpha}}^{\infty} (b - 1) dF = 0,$$

so that:

$$\int_{\underline{\alpha}}^{\infty} b dF = \bar{b} = 1.$$

Second, we have:

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

(If goods invested in education are deductible we would have an additional $-py \frac{\partial y}{\partial t}$ in the last term in brackets). Rewriting yields:

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

This formula can be simplified in four steps. First, use the definition of ξ to rewrite the first term:

$$\int_{\underline{\alpha}}^{\infty} \left(-\frac{\Psi'}{\eta} + 1 \right) (wl\phi(\cdot) + w(1-x)) dF = \xi \int_{\underline{\alpha}}^{\infty} (wl\phi(\cdot) + w(1-x)) dF.$$

Second, rewrite the second term:

$$\int_{\underline{\alpha}}^{\infty} tw\phi(\cdot) \frac{\partial l}{\partial t} dF = \frac{t}{1-t} \int_{\underline{\alpha}}^{\infty} wl\phi(\cdot) \frac{1-t}{l} \frac{\partial l}{\partial t} dF = -\frac{t}{1-t} \varepsilon_{lt} \int_{\underline{\alpha}}^{\infty} wl\phi(\cdot) dF.$$

Third, note that

$$twl\phi_x \frac{\partial x}{\partial t} - tw \frac{\partial x}{\partial t} = 0.$$

as a consequence of the first order condition for x . And, fourth, rewrite the last term:

$$\begin{aligned}\int_{\underline{\alpha}}^{\infty} twl \left(\phi_x \frac{\partial x}{\partial t} + \phi_y \frac{\partial y}{\partial t} \right) - tw \frac{\partial x}{\partial t} dF &= \int_{\underline{\alpha}}^{\infty} twl\phi_y \frac{\partial y}{\partial t} dF \\ &= \frac{t}{1-t} \int_{\underline{\alpha}}^{\infty} wl\phi(\cdot) \frac{\phi_y y}{\phi(\cdot)} \frac{1-t}{y} \frac{\partial y}{\partial t} dF = -\frac{t}{1-t} v\varepsilon_{yt} \int_{\underline{\alpha}}^{\infty} wl\phi(\cdot) dF.\end{aligned}$$

(If goods invested in education are deductible this term is zero). Collecting all terms gives:

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

where $\omega \equiv \int_{\underline{\alpha}}^{\infty} wl\phi(\cdot) dF / \int_{\underline{\alpha}}^{\infty} wl\phi(\cdot) + w(1-x) dF$ is the average ratio of second period income in total income.

Sensitivity analysis

In Tables 2 and 3 we change production elasticities. Here, it must be noted that the range over which the parameters can be varied is limited. Too high

Table 2: Optimal tax rates (%) - Changing production parameters

Base: $\gamma = .3$	CES		CELS	
	End	Ex	End	Ex
$\gamma = .1$	16.8	27.3	18.6	25.3
$\gamma = .2$	16.9	28.6	17.6	26.8
$\gamma = .3$	17.4	29.8	17.4	29.8
$\gamma = .4$	18.3	32.3	14.2	29.5
$\gamma = .5$	20.0	33.5	12.1	30.9
$\gamma = .6$	23.2	34.6	10.9	32.0

Table 3: Optimal tax rates (%) - Changing production parameters

Base: $v = .1$	CES		CELS	
	End	Ex	End	Ex
$v = .05$	19.5	28.2	18.0	26.8
$v = .1$	17.4	29.8	16.1	28.2
$v = .15$	15.9	31.4	14.5	29.4
$v = .2$	-	-	13.0	30.5

values violate second-order conditions, so there are limits on the returns to private inputs so as to rule out perverse behavior. From Tables 2 and 3 we can see that changing the elasticities of production yields only small effects in the optimal tax rates. So the results are robust with respect to the technology parameters of the production function of human capital. We note that optimum taxes increase when the elasticity of time in production of human capital increases for the CES utility function. This is somewhat strange at first sight, but can be attributed to the backward bending labor supply curve. Taxation induces agents to work more and rates of return to investments in human capital rise. More time is spend learning, the tax base increases, and costs of redistribution falls accordingly. This effect is stronger if the elasticity is larger.

Finally, in table 4 we compute optimum income taxes in the cases where either the revenue requirement of the government is increased or the elasticity of inequality aversion. In both cases optimum tax rates increase as expected. The earlier conclusion that optimum tax rates are lower with endogenous

Table 4: Optimal tax rates (%) - Changing government parameters

Base: $\nu = 0$	CES		Base: $\Lambda = 0$	CES		CELS	
	End	Ex		End	Ex	End	Ex
$\nu = .99$	31.0	43.7	$\Lambda = .1$	18.1	31.7	17.9	30.3
$\nu = 2$	38.3	50.2	$\Lambda = .3$	19.8	36.1	23.3	35.9
$\nu = 3$	42.9	54.2	$\Lambda = .5$	21.6	41.5	37.5	46.6

learning decisions is also confirmed here.²³

²³The optimization routine was not able to compute optimum taxes for the CELS utility function with positive inequality aversion as utility levels were negative for some agents due to the scaling parameter θ . This gave the problem that 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.