DISCUSSION PAPER SERIES

Mincer Equation, Power Law of Learning, and Efficient Education Policy

Wolfram F. Richter

March 2013

I Z A

Mincer Equation, Power Law of Learning, and Efficient Education Policy

Wolfram F. Richter
TU Dortmund University,
CESifo and IZA

ABSTRACT

Mincer Equation, Power Law of Learning, and Efficient Education Policy

The basis for the empirical research on earnings determination is the Mincer equation. Individuals are assumed to make schooling decisions by maximizing earnings. Leisure costs of schooling and labour supply are neglected which has some empirically implausible implications. This paper shows a way of deriving a Mincer-type earnings function from the more standard assumption of utility maximization. The implications are less questionable. The approach allows one to analyse the efficiency of education policy in Ramsey's tradition. Distortive wage taxation is shown to provide reason for subsidizing education in effective terms. Second-best policy is confronted with empirical evidence on OECD countries.

Corresponding author:

Wolfram F. Richter
Department of Economics
TU Dortmund University
44221 Dortmund
Germany
E-mail: Wolfram.Richter@tu-dortmund.de

1. Introduction

The objective of this paper is to unify two approaches that serve as models for the research of schooling choice and earnings determination.

The older approach has been initiated by Jacob Mincer (1958). It is positive theoretic in spirit and till this day the starting point for the empirical research on earnings determination. The approach relies on the assumption that individuals make schooling choices by maximizing the present value of their lifetime earnings. This has some strong testable implications when additionally assuming an infinite planning horizon and a constant rate of discount. Specifically, log earnings should be a linear function of schooling ("Mincer equation") and the marginal internal rate of return to schooling should equal the optimizing individual's discount rate. While the empirical evidence supporting linearity of mean log earnings is quite impressive, marginal internal rates of return to schooling are regularly estimated to be much higher than market rates of interest. Heckman, Lochner, and Todd (2008) interpret this as indicating that individuals do not simply maximize income when making schooling decisions.

The alternative approach to earnings determination cannot be associated with just one particular name. It is normative theoretic in spirit and the starting point for the analysis of the optimal taxation of education. The relevant literature is typically found in public finance journals. Some recent examples are Bovenberg and Jacobs (2005), Anderberg (2009), and Richter (2009). The approach relies on the assumption that individuals choose education by maximizing their lifetime utility. A major result states that education is neither taxed nor subsidized on a net basis in the policy optimum if only the earnings function displays specific features and if the government disposes of appropriate policy instruments.

Both approaches have their distinct merits and deficiencies. A direct comparison is provided in Section 6 below. Most disturbing is however to see that there is almost no cross acknowledgment in the literature. Each particular approach is obviously considered being irrelevant for the other. The reason cannot be the use of different terminologies, the Mincer approach talking about schooling where the public finance approach is talking about education. In fact, this paper adopts the practice which can be found in parts of the literature – e.g. in Cahuc and Zylberberg (2004) – and which uses schooling and education as interchangeable notions. The lack of cross referencing has rather to be explained by the fact that the two approaches rely on mutually exclusive functional specifications of the earnings function. The standard Mincerian earnings function is strictly convex thus displaying

increasing marginal returns to the absolute amount of schooling. By contrast, the standard public finance earnings function is strictly concave thus displaying decreasing marginal returns to education. *A priori* assumptions of the Mincer approach typically refer to the growth rate of earnings as a function of schooling while *a priori* assumptions of the public finance approach typically refer to the elasticity of earnings as a function of the amount of education.

Against this background the first objective of the present paper is to show a way of deriving Mincer-type earnings functions from an extended version of the neoclassical model of labour-leisure choice underlying the public finance literature. The derivation is achieved by replacing the Mincerian one-dimensional choice of schooling with a two-dimensional one. One dimension requires choosing the time of education at the cost of foregone leisure, and the other dimension requires choosing a particular subject ("discipline") out of a menu of competing ones. The returns to education from studying a particular discipline are modelled by an isoelastic concave learning function. By definition, the earnings function is the function of education which results when optimally adjusting the choice of discipline to the planned amount of education. More technically speaking, the earnings function is defined as the upper envelope of a set of discipline-specific learning functions.

This two-dimensional-choice approach has some critical advantages. First, it allows one to reconcile the notion of diminishing returns to learning with the notion of increasing returns to education. Second, even though the approach is compatible with a Mincerian earnings function it does not require equality of the marginal internal rate of return to schooling and the market rate of interest. Third, it is shown that the Mincerian earnings function is just a prominent special case within a set of functions compatible with the two-dimensional-choice approach. The specific feature characterizing this set of earnings functions is their increasing elasticity. The Mincerian earnings function is the special case where the elasticity of the earnings function is not only increasing but linear increasing. Such increasing elasticity of the earnings function is not simply assumed ad hoc. Instead, it is the logical implication of assuming the discipline-specific learning functions to be isoelastic. As argued below in Section 2, there is impressive empirical evidence that the productivity of learning displays such constant elasticity less than one. In neuroscience this is called the power law of learning (Newell and Rosenbloom, 1981; Anderson, 2005). See Section 2. The idea of deriving earnings functions from a two-dimensional choice is clearly not new. An early reference in

the Mincer literature is Willis and Rosen (1979). What is claimed to be innovative is the idea to derive convex earnings functions from concave but isoelastic learning functions.

The second objective of the present paper is to do normative analysis with convex earnings functions. In doing so, issues of equity are ignored. The focus will fully be on efficiency which is studied in Ramsey's tradition. The utilitarian foundation of schooling choice provided in Section 3 makes such an efficiency analysis meaningful. Earlier investigations by Richter (2009 and 2011) relied on strictly concave earnings functions. Hence they suffered from being not applicable to convex functions in general and Mincerian earnings functions in particular. They suffered additionally from generating inconclusive results in the key question of whether education should effectively be taxed or subsidized in second best. Against this background, the present paper makes twofold progress. First, it shows that the education elasticity rule derived by Richter (2011) *cum grano salis* extends to the empirically more appealing convex case (Proposition 4). The rule excels by particular simplicity in the case of a Mincerian earnings function. See Section 5. Secondly, the paper makes a clear case for effectively subsidizing education in second best (Proposition 3). Any effective taxation would conflict with the increasing elasticity of earnings functions which according to the reasoning of Section 2 can be considered being a robust empirical finding (Proposition 1). Even more, distortive wage taxation is shown to be the reason why education should be subsidized in effective terms relative to the first best. In Section 7 this clear-cut policy conclusion is confronted with the empirical evidence on tertiary education in OECD countries. It is shown that effective country policies more or less deviate from the optimal rule. Various countries, including the United States, effectively tax rather than subsidize tertiary education.

Section 8 summarizes while Section 6 highlights conceptual differences between the Mincer schooling model and the one developed in this paper. Two aspects deserve to be noticed more than others. While the Mincerian approach to earnings determination does not offer a truly convincing rationalization of why the mean of log earnings so nicely fits a linear function in schooling, the present approach allows one to conjecture that the growth rate of the geometric mean of individual earnings equals the ratio of two increasing functions. As a result, one should not be surprised finding empirical evidence for a ratio which is roughly constant in schooling. Another striking suggestion derived from this paper's model is that the empirical literature on earnings determination with its strong focus on the marginal internal rate of return to schooling is focussing on a variable the importance of which for education policy may be debated. Neither is the individual choice of education convincingly captured by the

simple equalization of the internal rate of return to schooling and the individual discount rate. Nor is the internal rate of return a key variable determining efficient education policy in the model to be presented. The true variables determining efficient policy turn out to be the second-order elasticity of the earnings function and the social cost of distortionary labour taxation.

2. The power law of learning and earnings curves

Most tasks get faster with routine. This observation is not surprising as such. What is surprising is that the rate at which people improve with practice appears to follow a similar pattern that is best fitted by a power function. "It has been seen in pressing buttons, reading inverted text, rolling cigars, generating geometry proofs, and manufacturing machine tools" (Ritter and Schooler, 2001). One of the early studies reporting detailed data is by Blackburn (1936). The study lists the productivities of seven subjects doing five specific tasks in repeated trials. The subjects were asked to sort packs of 42 cards, to cross out e's in nonsense French, to transform short texts by some rather complicated code substitution, to do addition exercises, and to learn a stylus maze. Crossman (1959) finds the first four experiments confirming the power law of learning, while the fit in maze learning is, according to him, more doubtful. Figure 1 displays the learning curves of three individuals when crossing out e's, doing code substitution, and adding digits, respectively.
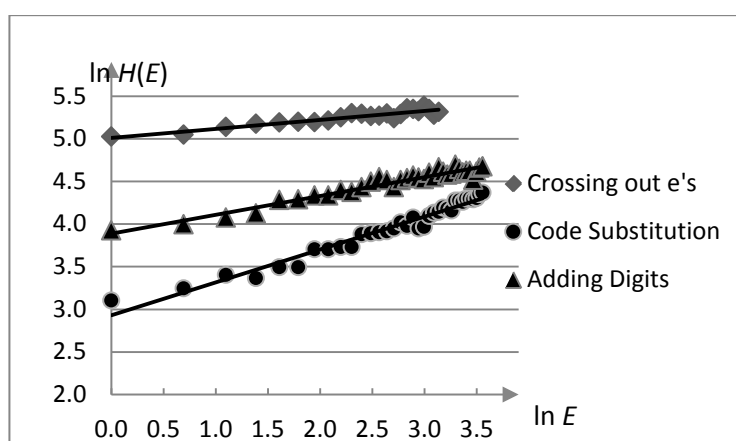
Figure 1: Learning curves when crossing out e's, doing code substitution, and adding digits. Data taken from Blackburn (1936)[1]. Logarithmic scaling. *E* measures trials and *H(E)* the average number of achievements in 100 seconds.

---

[1] The displayed learning curves are the one of subject 4 in the crossing-out-e's experiment, the one of subject 1 in the code-substitution experiment, and the one of subject 2 in the adding-digits experiment.

The empirical evidence on learning curves suggests defining individual productivity $H = H(D, E)$ by setting

$$\ln H(D, E) \ \equiv \ h(D) + \eta(D) \ln E \qquad \textit{(power law of learning)}. \qquad (1)$$

The variable E measures experience, while D stands for some particular task such as crossing out e's. The characteristic feature of the power law of learning is that the elasticity of productivity with respect to experience, $\eta(D)$, is constant in E. The following analysis relies on the assumption that the power law of learning does not only govern the doing of simple tasks but also the acquisition of more complex skills ("human capital"). To be more specific, consider the study of economics. The suggestion is that the command of economics is a skill acquired by studying economics as a subject which requires solving economics problems again and again. The repetition results in productivity enhancements the elasticity of which is not necessarily the same for different students but constant in each student's experience. Extending the assumption of constant elasticity to the acquisition of complex skills is justified by the observation that the power law reflects a behavioural regularity which is "ubiquitous" (Newell and Rosenbloom, 1981; Ritter and Schooler, 2001) and which can be considered to reflect the neurological functioning of human brain. In the generalized sense adopted in this paper, D stands for a *subject* or a *discipline* to be studied, while E measures *learning* or *education* in units of time.

The reference to the power law is empirically maintainable when productivity H^p is measured in physical terms. This paper however relies on the assumption that the power law is a valid description of the return to education even if productivity H is measured in monetary terms. Such an extension is logically defensible when the market valuation $w = H/H^p$ equally fits the power law, be it in the non-trivial sense with $w = w(D, E)$ or in the trivial sense with $w = w(D)$. The trivial case needs less justifying arguments. Hence assume that $H(D, E)$ can be written as $w(D) \cdot H^p(D, E)$. Thus excluding the case where the pricing of skills varies with E amounts to conceiving education as an activity which is remunerated only via the enhancing effect it has on physical productivity and not as a characteristic of own market value.

Given that $H(D, E)$ in (1) can be interpreted as monetary productivity, it makes sense to assume that individuals maximize $H(D, E)$ in D at each E they choose. (In the Mincer literature such a maximization is considered giving rise to a selection problem. See Willis and Rosen, 1979.) Let $D(E)$ denote the maximizing discipline assumed to exist for each E. The

function $G(E) \equiv H(D(E), E)$ is called the *earnings function*. The following proposition characterizing the shape of earning functions when learning functions are isoelastic is just as trivial as fundamental for the subsequent analysis.

Proposition 1: The power law of learning has the implication that the elasticity $\gamma(E) \equiv \eta(D(E))$ of the earnings function $G(E) \equiv H(D(E), E)$ is increasing in education. The property of increasing elasticity is inherited by the geometric mean of individual earnings functions.

The proof is straightforward. Consider an individual earnings function. Assume $E_1 > E_0$ and D_i to be the optimal choice at E_i (*i*=0,1). Hence $G(E_1) = H(D_1, E_1) \geq H(D_0, E_1)$ and $G(E_0) = H(D_0, E_0) \geq H(D_1, E_0)$. Eq. (1) implies

$$h(D_1) + \eta(D_1) \ln E_1 \geq h(D_0) + \eta(D_0) \ln E_1 \text{ and}$$

$$h(D_0) + \eta(D_0) \ln E_0 \geq h(D_1) + \eta(D_1) \ln E_0 \ .$$

Adding these two inequalities, yields

$$[\ln E_1 - \ln E_0] \ \eta(D_1) \geq [\ln E_1 - \ln E_0] \ \eta(D_0), \text{i.e.}$$

$$\gamma(E_1) = \eta(D(E_1)) = \eta(D_1) \geq \eta(D_0) = \eta(D(E_0)) = \gamma(E_0).$$

The geometric intuition is the following. Consider a menu of linear but possibly intersecting learning functions displayed in a diagram like Figure 1 with logarithmic coordinates. The slope of each individual learning function is constant by assumption. The slope of the upper envelope is then necessarily increasing.

Now assume that each individual *n*=1,2,…,*N* has her own learning function satisfying the power law, $\ln H^n(D, E) \equiv h_n(D) + \eta_n(D) \ln E$. Let $D^n(E)$ denote the productivity maximizing discipline chosen by *n* at *E* and let $\bar{G}(E)$ be the geometric mean of individual earnings,

$$\ln \bar{G}(E) \equiv \ln \sqrt[N]{\prod_n H^n(D^n(E))} = \frac{1}{N}\sum_n \ln H^n(D^n(E)).$$

As the slope of $\ln H^n (D^n(E))$ is increasing as a function of $\ln E$, the slope of $\ln \bar{G}(E)$ is equally increasing in $\ln E$. This proves that the elasticity of $\bar{G}(E)$ is increasing.□

The following analysis relies on assuming the elasticity of G to be increasing in E. This will later be the reason why it is efficient to subsidize education in second best. The simplest case of an earnings function with an increasing elasticity assumes linearity:

$$\gamma(D(E)) \equiv mE \Longleftrightarrow \ln G = g_0 + mE \text{ with } m>0. \tag{2}$$

An earnings function of type (2) is called *Mincerian*. In the present model, the sole appeal of the Mincerian specification arises from the fact that linearity is a first-order approximation to an increasing function. Note that while a Mincerian earnings function is strictly convex, $G'' = m^2 G > 0$, convexity is not implied by assuming increasing elasticity. A function displaying increasing elasticity may well be strictly concave. Richter (2009 and 2011) derives characterizations of second-best policy for strictly concave earnings functions when ignoring earnings risk. Sections 4 and 5 derive characterizations for convex earnings functions and earnings risk.

3. Household behaviour

Household behaviour is modelled by a representative taxpayer living for two periods and deriving strictly increasing utility U from consumption C_i and strictly decreasing disutility from non-leisure time L_i. The index i=1 refers to the first period while the indices $i = q, n$ refer to two possible states in the second period. The function $U = U(C_1, L_1, C_q, L_q, C_n, L_n)$ is strictly quasi-concave. L_i is identical with the labour supply in state $i = q, n$ of the second period. By contrast, L_1 equals the time E spent on education. Notice that with strictly convex earnings it would not be an earnings optimizing strategy to divert a positive share of L_1 on supplying labour at some constant wage rate. The return to second-period labour depends on the success and the amount of education. Education is a risky activity which only succeeds with probability π. In case of failure, labour is paid a constant wage rate ω_n. In case of success, labour is paid $\omega_q G(E)$, where ω_q is constant and where $G(E)$ is a twice differentiable convex function of increasing elasticity $\gamma(E)$. The sole reason for modelling the wage rate ω_q separately from G is taxation. If there were no taxation, one could set $\omega_q = 1$ without loss of generality. The quantity L_q is interpreted as qualified labour / non-leisure.

Likewise, L_n and L_1 are interpreted as nonqualified activities. Education causes a monetary cost of tuition assumed to be linear in time, φE. In the absence of any initial wealth, first-period expenditures imply negative savings:

$$S \equiv -\varphi E - C_1. \tag{3}$$

By way of normalization, the price of consumption is set equal to one. The gross rate of return to saving is denoted by ρ. Expected second-period consumption $\overline{C}_2$ is constrained by the income expected to be earned in the same period after servicing negative savings:

$$\overline{C}_2(C_q, C_n) \equiv \pi C_q + (1-\pi)C_n = \rho S + \pi \omega_q G(E) L_q + (1-\pi)\omega_n L_n. \tag{4}$$

Eq. (4) implicitly assumes that the risk of educational success is insurable. Informational imperfections are not modelled. Substituting for S in eq. (3) and eq. (4) yields the lifetime budget constraint:

$$C_1 + \varphi E = [\pi \omega_q G(E) L_q + (1-\pi)\omega_n L_n - \overline{C}_2] / \rho. \tag{5}$$

Maximizing utility in $C_1, E, C_q, L_q, C_n, L_n, \geq 0$ subject to (5) requires *inter alia* equating the *marginal return to education* with the *effective marginal cost of education*,

$$MR \equiv \pi G'(E) L_q = \rho(\varphi + MRS^1) / \omega_q \equiv MC \tag{6}$$

where $MRS^1 \equiv -U_{L_1} / U_{C_1}$ captures the marginal cost of foregone leisure. Throughout, subindices of functions indicate partial derivatives. Because of the assumed convexity of the earnings function, qualified wage income, $\pi \omega_q G(E) L_q / \rho$, is convex in E. Such convexity has implications for the taxpayer's optimization. Just assuming quasi-concavity of the utility function is clearly not sufficient to ensure that the taxpayer's optimization is well behaved. The second-order conditions are not necessarily satisfied and solutions may fail to be in the interior of the domain. Still, the following analysis only looks at interior solutions of the first-order conditions. The implicit assumption is that the taxpayer discards all solutions of the first-order conditions which fail to be global optimums and that a global optimum exists in the interior. The latter requires assuming that the supply of non-leisure is sufficiently inelastic. More precisely, the convexity of the earnings function must be dominated by the convexity of the disutility of non-leisure.

In its technical derivations the following analysis relies on the taxpayer's expenditure function. This function is defined as

$$e(\omega_n, \omega_q, \varphi, \rho; u) \equiv \min[\rho C_1 + \bar{C}_2(C_q, C_n) + \rho \varphi E - \pi \omega_q G(E)L_q - (1-\pi)\omega_n L_n] \quad (7)$$

in $C_1, E, C_q, L_q, C_n, L_n$ subject to $U(C_1, E, C_q, L_q, C_n, L_n) \geq u$. Assume that the expenditure function is twice differentiable. By relying on Hotelling's lemma one derives the identities $e_{\omega_n} = -(1-\pi)L_n$, $e_{\omega_q} = -\pi G(E)L_q$, $e_\varphi = \rho E$, and $e_\rho = C_1 + \varphi E = -S$. The capital letters L_i, E, S, and C_i have to be interpreted as Hicksian supply and demand functions. This means that they have to be evaluated at $\omega_n, \omega_q, \varphi, \rho$, and u.

4. Second-best policy

The government faces the need to raise revenue. Four linear tax instruments are available, each of which is distorting. The taxes are levied on labour income, on the cost of tuition, and on the return to saving. They are modelled implicitly as the difference between prices before and after taxes. The prices after taxes and subsidies are endogenous and denoted by $\omega_n, \omega_q, \varphi, \rho$. The prices before taxes and subsidies are exogenous and denoted by w_n, w_q, f, r.[2] The tax on labour income in state $i = q, n$ is modelled by $w_i - \omega_i$, the tax on capital income by $r - \rho$, and the tax on the cost of tuition by $\varphi - f$. It goes without saying that each tax can well take on a negative value so that it is effectively a subsidy. Government's net revenue amounts to

$$T \equiv (\varphi - f)E + [\pi(w_q - \omega_q)G(E)L_q + (1-\pi)(w_n - \omega_n)L_n + (r - \rho)S]/r. \quad (8)$$

By invoking Hotelling's lemma revenue can be written as

$$T = \frac{1}{\rho}(\varphi - f)e_\varphi + [(\omega_q - w_q)e_{\omega_q} + (\omega_n - w_n)e_{\omega_n} + (\rho - r)e_\rho]/r \quad (9)$$

---

[2] It has been suggested above to interpret $G(E)$ as monetary productivity which then requires $w_q = 1$. If one chose instead to interpret education as a labour augmenting activity and $G(E)L_q$ as effective qualified labour, then w_q would equal the latter's marginal productivity. It would be a straightforward exercise to endogenize the prices before taxes and subsidies in this case. However, such endogenization does not produce interesting new insights. Assuming that no pure profit accrues to the private sector so that the production efficiency theorem applies, endogenizing has no structural effect on efficient education policy.

The planner's dual objective is to maximize net revenue (9) in $\omega_n, \omega_q, \varphi, \rho$, subject to the individual budget constraint $e=0$. In Appendix A it is shown that taking partial derivatives with respect to $x = \varphi, \omega_n, \omega_q, \rho$, invoking Hotelling's lemma, and eliminating the Lagrange multiplier yields the following system of three first-order conditions:

$$\hat{E} \;=\; \widehat{L_n} \;=\; \widehat{GL_q} \;=\; \widehat{C_1}, \tag{10}$$

where the hat notation denotes relative changes, $\hat{X} \equiv \Delta X / X$, and where the total differentiation operator Δ is defined on arbitrary functions $X = X(\omega_n, \omega_q, \varphi, \rho; u)$ by

$$\Delta X \equiv \frac{1}{\rho}(\varphi - f)X_\varphi + \frac{1}{r}(\omega_q - w_q)X_{\omega_q} + \frac{1}{r}(\omega_n - w_n)X_{\omega_n} + \frac{\rho - r}{r}X_\rho. \tag{11}$$

According to (11), ΔX equals the weighted sum of the partial derivatives of X with the weights given by the tax wedges. It is an approximation of the total change in X when taxes are chosen efficiently. In Appendix A the equations in (10) are shown to imply

$$\widehat{\overline{C}_2} = \hat{E}. \tag{12}$$

By applying hat calculus one obtains

$$\widehat{GL_q} = \widehat{L_q} + \hat{G} = \widehat{L_q} + \gamma\hat{E} . \tag{13}$$

where γ is the elasticity of the earnings function. Summarizing (10), (12), and (13) yields:

Proposition 2: Second-best efficiency requires reducing

    (i) education, consumption, nonqualified labour, and effective qualified labour equi-proportionately while reducing

    (ii) qualified labour to a lesser degree in accordance with

$$\widehat{L_q} = (1 - \gamma)\widehat{L_n} .$$

According to statement (i) of the proposition it is second best to reduce all the quantities E, C_1, $\overline{C}_2$, L_n, and GL_q showing up in the taxpayer's budget constraint in the same proportion, when all these demand and supply functions are interpreted in the Hicksian sense. The equi-proportionate reduction is something one would clearly expect in view of Ramsey's (1927) characterization of efficient taxation. The nonstandard result concerns L_q. Obviously,

efficiency requires reducing qualified labour relatively less than non-qualified labour. The ratio equals $1 - \gamma$, and it decreases in γ. In other words, the more elastic the individual earnings function is, the less should qualified labour be reduced in relative terms. Although this makes good sense, it must be noted that it fails to agree with Ramsey's Rule of reducing *all* household choices equi-proportionately. Only the effective qualified labour GL_q is reduced equi-proportionately. As $G = G(E)$ reacts elastically, L_q is reduced to a lesser degree. In footnote 5 of section 7 it will be argued that it is empirically justified to assume $\gamma < 1$.

Proposition 2 makes no statement about the sign and the size of efficient tax rates. This may be considered being the unsurprising price one has to pay when not restricting the choice of the taxpayer's utility function. And yet, more can be said with regard to education policy. One only has to exploit the fact that the planner's optimization is constrained by the equality of the marginal return and the marginal cost of education. If this equality is exploited, effective subsidization of education turns out to be optimal irrespective of which particular utility function has been assumed. To see this, denote the second-order elasticity of the earnings function by $\gamma_\gamma \equiv E\gamma' / \gamma$ and apply hat calculus:

$$\gamma_\gamma \hat{E} = [\frac{EG''}{G'} + 1 - \gamma]\hat{E} = \frac{\Delta G'}{G'} + \frac{\Delta L_q}{L_q} = \frac{\Delta (G'L_q)}{G'L_q} = \frac{\Delta MR}{MR}$$

$$\underset{(7)}{=} \frac{\Delta MC}{MC} = \frac{\Delta(\rho\varphi) + \Delta(\rho MRS^1)}{\rho(\varphi + MRS^1)} - \frac{\Delta\omega_q}{\omega_q}$$

$$= \frac{(\varphi - f) + (\rho - r)(\varphi + MRS^1) / r + \rho\Delta MRS^1}{\rho(\varphi + MRS^1)} - \frac{\omega_q - w_q}{r\omega_q}$$

$$= \frac{w_q / r}{\omega_q} - \frac{f + MRS^1 - \rho\Delta MRS^1}{\rho(\varphi + MRS^1)}. \tag{14}$$

The left-hand side of eq. (14) – measuring the efficient relative change in the marginal return to education – is negative. More precisely, the factor γ_γ is positive as the elasticity of the earnings function is assumed to be increasing (Proposition 1). By way of contrast, the efficient relative reduction of education, $\hat{E}$, is necessarily negative given that taxation is to raise positive revenue. Hence, the left-hand side of (14) is negative.

Some more words are needed to understand that the right-hand side of (14) can be interpreted as an effective tax wedge on education. Note first that the term $MRS^1 - \rho \Delta MRS^1 \equiv w_1^s$ captures the social cost of foregone nonqualified leisure in the first period. This interpretation may not appear obvious and still, it follows from comparing the choice of education with the choice of second-period nonqualified labour. The taxpayer's optimal choice of L_n makes her equate the marginal rate of substitution $MRS^n \equiv -U_{L_n}/U_{C_1}$ with the after-tax wage rate, ω_n/ρ. When applying the Δ-operator to the equation $\omega_n = \rho MRS^n$ one obtains

$$\frac{\omega_n - w_n}{r} = \frac{\rho - r}{r} MRS^n + \rho\Delta(MRS^n) \iff \frac{w_n}{r} = \frac{\omega_n}{\rho} - \rho\Delta MRS^n .$$

The parallel structure suggests setting $\omega_1^s \equiv MRS^1$, $w_1^s \equiv \omega_1^s - \rho\Delta MRS^1$ and interpreting ω_1^s, w_1^s as shadow prices. With this notation, eq. (14) can be rewritten as

$$\Rightarrow \qquad \rho\gamma_\gamma\hat{E} = \rho\frac{\Delta MR}{MR} = \rho\frac{\Delta MC}{MC} = \frac{w_q/r}{\omega_q/\rho} - \frac{f + w_1^s}{\varphi + \omega_1^s} \equiv \Delta_E \tag{15}$$

where Δ_E is interpreted as *effective wedge on education*. Another way of displaying the wedge is

$$\Delta_E = \frac{\pi w_q G'L_q/r}{\pi\omega_q G'L_q/\rho} - \frac{f + w_1^s}{\varphi + \omega_1^s} = \frac{[\pi w_q G'L_q/r - (\varphi + \omega_1^s)] + [(\varphi + \omega_1^s) - (f + w_1^s)]}{\varphi + \omega_1^s} . \tag{16}$$

The right-hand side of eq. (16) reveals that Δ_E is the sum of two standard wedges. Non-standard is only the non-negative slope of the "demand curve" given that G has been assumed to be convex. Eq. (16) and eq. (6) imply that the effective wedge on education is positive if, and only if, the effective private marginal cost of education exceeds the effective social one:

$$\Delta_E \overset{>}{\underset{<}{=}} 0 \iff \rho(\varphi + \omega_1^s)/\omega_q \overset{>}{\underset{<}{=}} r(f + w_1^s)/w_q . \tag{17}$$

Efficient education policy requires a vanishing wedge, $\Delta_E = 0$. As the left-hand side of (15) is negative, it is however second best to *subsidize education effectively*, i.e. to set $\Delta_E < 0$.

Proposition 3: It is second best to subsidize education in effective terms.

A negative value of Δ_E may result from subsidizing the cost of tuition, f. This is however not the only way of causing Δ_E to be negative. Other means are (i) reducing the tax on qualified labour and thus increasing the statutory return to education, w_q, and also (iii) taxing the return to saving, r, and thus reducing the cost of education. If the earnings function were concave and if the taxpayer were to supply nonqualified labour $L_1 - E > 0$ in the first period, a forth way of encouraging education would be to tax first-period nonqualified wage income and thus to reduce the cost of foregone earnings. Since the present model assumes convex earnings so that the taxpayer's optimization implies $L_1 = E$, there are no foregone earnings but only non-taxable costs of forgone leisure.

Eq. (15) allows one to determine the efficient proportional change in the effective marginal cost of education that is needed to align the taxpayer's maximization of net qualified labour income with the Ramsey rule of equi-proportional reductions in demands and supplies. As the elasticity of the earnings function is increasing, efficiency requires the marginal return to education, MR, to decrease which in turn requires subsidizing the effective marginal cost, MC.

The generality of the formula (15) is striking. Applicability is however limited as long as one does not know which percentage reduction $\hat{E}$ is efficient and which values of $\gamma_\gamma, \omega_1^s$, and w_1^S are to be assumed. More can only be said when working with specific utility and earnings functions. This is why the focus is on particular functional specifications in what follows.

5. An elasticity rule

Consider the special case in which the taxpayer's utility function is quasi-linear in first-period consumption and additive in sub-utilities,

$$U(C_1, E, C_q, L_q, C_n, L_n) = C_1 - V(E) + \pi U^q(C_q, L_q) + (1 - \pi)[U_n(C_n) - V(L_n)] \ . \ (18)$$

Disutility of nonqualified non-leisure V is a strictly increasing and strictly convex function. Maximizing (18) subject to the taxpayer's budget constraint, (5), yields the following first-order conditions:

$$\pi \omega_q G'(E) L_q / \rho = \varphi + V'(E) \tag{19}$$

$$\pi\omega_q G(E)/\rho = -\pi U_L^q(C_q, L_q) \tag{20}$$

$$(1-\pi)\omega_n/\rho = (1-\pi)V'(L_n) \tag{21}$$

$$\rho = 1/U_C^q(C_q, L_q) = 1/U_n'(C_n) \tag{22}$$

When substituting for savings, S, one can restate the government budget constraint, (8):

$$T = (\varphi - f)E + \pi[\frac{w_q}{r} - \frac{\omega_q}{\rho}]G(E)L_q + (1-\pi)[\frac{w_n}{r} - \frac{\omega_n}{\rho}]L_n]$$

$$+ [\frac{1}{\rho} - \frac{1}{r}][\pi C_q + (1-\pi)C_n] \qquad (\alpha) \tag{23}$$

The bracketed variable α denotes a Lagrange multiplier associated with the tax planner's problem. The planner's primal objective is to maximize (18) subject to (5) and (19) – (23). Taking partial derivatives of the Lagrangean objective function with respect to ω_n, L_n and rearranging yields

$$(\alpha - 1)\nu = \alpha \Delta_n . \tag{24}$$

where $\nu \equiv L_n V''/V'$ is the elasticity of marginal disutility of nonqualified labour and where

$$\Delta_n \equiv \frac{w_n/r - \omega_n/\rho}{\omega_n/\rho}$$

is the tax wedge on second-period nonqualified labour. Δ_n/ν can be interpreted as the *social cost of taxation*. Taking partial derivatives with respect to φ, ω_q, E and rearranging yields

$$(\alpha - 1)\eta_\eta(\varphi + V'(E)) = \alpha[f - \pi w_q G'(E)L_q/r] + \alpha V'(E) + (\alpha - 1)EV''(E) . \tag{25}$$

Assume that the planner's optimization generates a constellation where the supplies of non-qualified leisure are equal in both periods, $E = L_n$. Then (25) can be rewritten as

$$(\alpha - 1)\eta_\eta(\varphi + V'(L_n)) = \alpha[f - \pi w_q G'(L_n)L_q/r] + \alpha w_n/r \tag{26}$$

Setting

$$\Delta_{E,n} \equiv \frac{w_q / r}{\omega_q / \rho} - \frac{f + w_n / r}{\varphi + \omega_n / \rho}$$

and combining eq. (24) and eq. (26) yields

Proposition 4:   If the utility function satisfies (18), if ω_q, ω_n, and φ are optimally chosen by the planner, and if nonqualified non-leisure choices are equal in both life periods, then second-best policy is characterized by the *education elasticity rule*,

$$\Delta_{E,n} = -\gamma_\gamma \Delta_n / v \tag{27}$$

Equation (27) suggests that the effective rate of subsidizing education should increase in the second-order elasticity of the earnings function and in the social cost of distortionary labour taxation. The latter in turn increases in the tax wedge on nonqualified labour and in the (compensated) wage elasticity of the nonqualified labour supply, $1/v$. It is worth noting that the rule holds even if the planner does not optimize with respect to ρ. Saving does not need to be taxed efficiently, and yet second-best education policy should respect eq. (27).

For a Mincerian earnings function eq. (27) takes on a particularly simple form. This is so because the elasticity of a Mincerian earnings function is linear in education by definition. Hence the second-order elasticity, γ_γ, equals one.

Corollary: If the earnings function is Mincerian, efficient education policy is characterized by

$$\Delta_{E,n} = -\Delta_n / v .\tag{28}$$

A final remark concerns logical consistency. The education elasticity rule implicitly assumes that the second-order condition of educational choice, $\pi\omega_q G''(E)L_q/\rho < V''(E)$, holds in the taxpayer's optimum. Assuming a Mincerian earnings function and $E = L_n$, this implies

$$mE = \frac{EG''}{G'} < \frac{EV''(E)}{\varphi + V'(E)} = \frac{L_n V''(L_n)}{\varphi + V'(L_n)} = \frac{\omega_n/\rho}{\varphi + \omega_n/\rho} v. \qquad (29)$$

The inequality (29) is a constraint which has to be checked for reasons of consistency when confronting eq. (28) with the empirical data on educational choice. Because $mE=\ln[G(E)/G(0)]$, the left-hand side of (29) can be interpreted as skill premium. The ratio on the right-hand side, $\omega_n/(\rho\varphi + \omega_n)$, measures the cost of foregone leisure spent on education as a share in the private statutory cost of education. This share has to exceed the product of the skill premium and the compensated elasticity of nonqualified labour if the second-order condition of optimal schooling choice is to hold.

6. The Mincer model in comparison

The present approach critically relies on the assumption that the elasticity of the earnings function $\gamma = EG'/G$ is increasing in E. By contrast, the Mincer schooling model relies on the assumption that the growth rate of earnings G'/G is weakly decreasing in E. Both assumptions are well grounded in empirical analysis. In this paper, the increasing elasticity of the earnings function has been justified by referring to the power law known to govern learning functions. By contrast, a weakly decreasing growth rate of earnings has been justified in the literature by comparing the internal rates of return to increasing years of schooling. Psacharopoulos (1985) a. o. provides evidence that the internal rate of return to finishing high school is at least as high as the internal rate of return to finishing college. See also the discussion in the influential text book of Borjas (2012). More contentious is the implication of the Mincer schooling model that estimated internal rates of return can be interpreted as the rates by which individuals discount future earnings. The empirical research regularly comes up with a significant mark-up of the marginal internal rate of return to schooling on the market rate of interest. Evidence is provided a. o. by Heckman, Lochner, and Todd (2008). Their "estimates of the return to high school and college completion for recent years are … substantially larger than real interest rates typically observed." According to Heckman et al.,

the evidence is suggestive for the interpretation that individuals do not simply maximize income when making schooling decisions.

The equalization of internal rates of return to schooling and discount rates has been rationalized by two competing arguments. One originates in Mincer (1958) and relies on arbitrage. The present value of earnings derived from the quantity of schooling E_2 must equal the present value of earnings derived from E_1 even if E_2 deviates from E_1. Foregone earnings are the only cost of schooling. Assuming an infinite working life, arbitrage requires the growth rate of earnings to equal the individual discount rate. Assuming that credit markets are perfect, the discount rate can be equated with the market rate of interest after tax, ρ. The equality of $G'/G = \rho$ for all E has the implication that the increase in earnings is totally unrelated to productivity differentials generated by education. Instead, the percentage increase in earnings reflects the cost of funds and the return to capital. As a result, there is no individual net benefit from accumulating more years at school. E_2 is just as gainful as E_1. Furthermore, the only cost of schooling is the cost of funds. This is highly implausible.

The competing rationalization of $G'/G = \rho$ is less demanding. It relies on interpreting the equality as one which only holds for a particular value of E. The equality for a particular E is derived as first-order condition when maximizing the present value of individual earnings $\int_E^\infty G(E)e^{-\rho t}dt$ in E. To guarantee a well behaved optimization, the growth rate G'/G has to be decreasing which is interpreted as reflecting diminishing returns. Reference to diminishing returns is however difficult to reconcile with the neoclassical framework. According to standard neoclassical reasoning the notion of diminishing returns should apply to the marginal return to earnings G' rather than to the growth rate, G'/G. Even more, the approach is unable to explain why the *mean* of log earnings, $\ln \bar{G}(E)$, so nicely fits a linear function in schooling. For an impressive visualization see Figure 2 of Card (1999). The present approach does better in this respect. According to Proposition 1 one should not be surprised to find empirical evidence for an elasticity of the geometric mean of earnings, $E\bar{G}'/\bar{G}$, which is increasing in schooling. As a result, the growth rate $\bar{G}'/\bar{G}$ is the ratio of two increasing functions and one should not be surprised to find empirical support for a ratio which is roughly constant in E.

In the present paper's model, a constant individual growth rate of earnings is the possible though not necessary result when individuals optimize on the choice of disciplines to be studied and when assuming that each discipline-specific learning curve is isoelastic as suggested by the power law of learning. In contrast to the Mincer model, the growth rate of

earnings is no variable governing the individual decision on the length of education. Instead, the length is determined by equating the marginal return to education, $\pi G' L_q$, to the effective marginal cost, $\rho(\varphi + MRS^1) / \omega_q$. When the earnings function G is convex, as suggested by the empirical evidence, then increasing marginal costs of foregone leisure stop the individual in the model from overly extending education.

Striking differences between the Mincer model and the model presented in this paper are additionally revealed when characterizing socially efficient education policy. In the Mincer model, the criterion of efficiency is $G'/G=r$. The only impediment to efficiency this criterion allows one to identify is a tax on saving. This contrasts with the present model. Here the criterion is the equality of the private and the social effective (marginal) costs of education as implied by eq. (17). A tax on saving, $\rho<r$, is no impediment to efficiency if it is compensated by other taxes and subsidies – a tax on qualified labour income in particular. Conceptual differences become even more manifest in the characterization of second-best policy. The Mincer model offers no clue to how to differentiate meaningfully between the first best and the second best. By contrast, the present model does. It suggests that it is second best to subsidize education effectively if wage income is taxed. The intuitive explanation is the following. Solving the planner's problem requires determining the efficient proportional change in the effective marginal cost of education such that the taxpayer's optimal choice of education is aligned with the Ramsey rule of equi-proportional reductions in demands and supplies. The alignment requires subsidizing the taxpayer's choice of education because there are good reasons to assume an increasing elasticity of the earnings function.

An immediate implication of the present model is that the strong focus of the empirical literature on earnings determination on the growth rate of earnings has to be questioned. Neither individual educational choice nor efficient educational policy is shown to be governed by this growth rate. The variables identified as characterizing efficient policy are the second-order elasticity of the earnings function and the social cost of distortionary labour taxation.

In later work, Mincer (1974) extended his model of 1958 to incorporate post-school work experience. The result was the so-called extended earnings function which is obtained when adding a quadratic experience term to the linear function of years of schooling in estimating log earnings. In follow-up estimations the quadratic experience term has occasionally been replaced with polynomials of higher degree in order to improve the fit with data. By contrast, the present model suggests adding a linear transformation of log experience to the linear

function of years of schooling. At least, this is what the power law of learning suggests when interpreting post-school work experience as the acquisition of experience in a given discipline.

7. Second-best tertiary education policy: Practice

It is inviting to confront the theoretical results derived in this paper with the empirical evidence. An immediate implication of Propositions 1 and 4 is that the tax wedges on nonqualified labour and education should be negatively correlated. This theoretical prediction is supported by some casual inspection of Figure 2. The figure uses OECD (2011) data to relate the tax wedge on nonqualified labour, Δ_n, with the wedge by which tertiary education is effectively subsidized, $-\Delta_{E,n}$.[3] The figure confirms the theoretical prediction of a positive relationship between Δ_n and $-\Delta_{E,n}$. The coefficient of correlation is 0.43. If Turkey and Denmark are treated as outliers and removed from the data set, the coefficient of correlation however drops to 0.21. Note that roughly one third of all countries are reported to tax tertiary education in effective terms. Effective taxation is clearly at variance with Proposition 3.
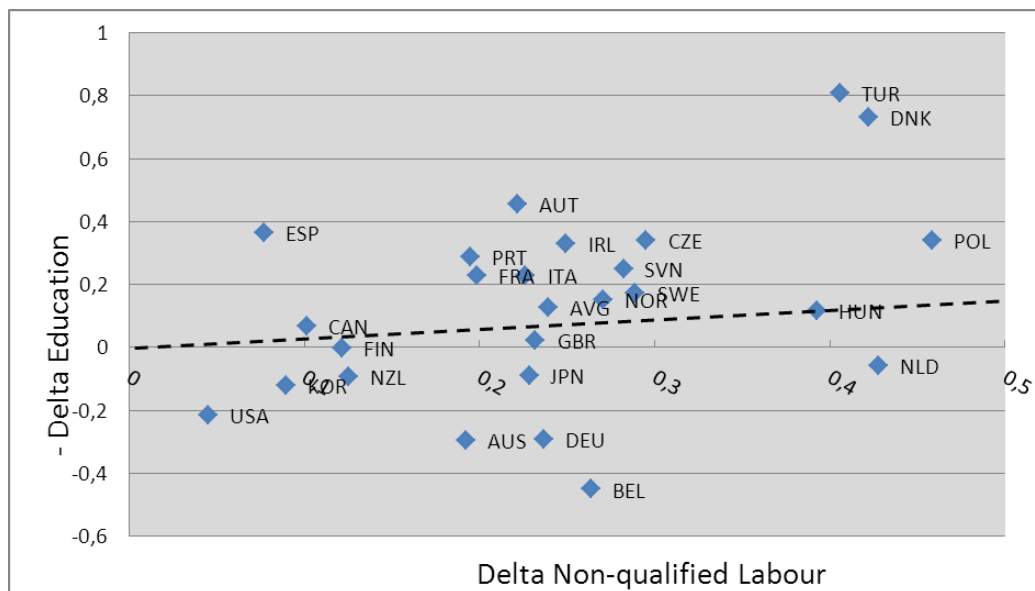

Figure 2: Measured and second-best tax and subsidy wedges Δ_n and $-\Delta_{E,n}$
in OECD countries. AVG marks the OECD average.

---

[3] The wedges are computed by relying on data on private and public direct costs of tertiary education, private and public foregone earnings, gross earnings benefits, income tax effects, social contribution effects, transfer effects, unemployment effects, and grant effects. Taxes on savings are ignored. Cash flows are discounted by a 3% interest rate. See Appendix *B* for details.

The dashed line in Figure 2 marks second-best policy. I.e. combinations of wedges on this line satisfy the simplified version (28) of the education elasticity rule. The underlying assumptions are (i) a Mincerian earnings function and (ii) a compensated elasticity of nonqualified labour supply with value of 0.3. Both assumptions are clearly debatable on empirical grounds. However, log linearity is an acceptable assumption when studying mean earnings (Carneiro, Heckman, and Vytlacil, 2011). And the value of 0.3 is justified with reference to Keane (2011, p. 1044).

If the dashed line is accepted as indicating efficient education policy, then Hungary comes closest to the optimum. The deviation from efficient policy is relatively small in Canada, Finland, Great Britain, Norway, and Sweden. By contrast, heavily overshooting incentives for tertiary education are displayed for Turkey and Denmark. The measured value of $-\Delta_{E,n}$ is at 81 per cent for Turkey while 12 per cent would be efficient. Some closer inspection of the data reveals that the large measured value for Turkey results because this is the country in the sample taxing qualified income most leniently. By contrast, the large measured value of 73 per cent results for Denmark because this is the country subsidizing the direct cost of tertiary education most strongly in absolute terms. An extreme case of insufficient incentives is Belgium for which the measured value of $-\Delta_{E,n}$ is at minus 45 per cent while plus 8 per cent would be efficient. According to the data, Belgium stands out for a policy characterized by a heavy tax on qualified income and by a public share in the cost of tertiary education which does not provide sufficient compensation.

Figure 2 lends itself to highlighting the rich variety of education policies pursued by OECD countries. In order to give some illustrating examples, the wedge on education is written as the difference of cost and benefit ratios,

$$-\Delta_{E,n} = \frac{f + w_n/r - (\varphi + \omega_n/\rho)}{\varphi + \omega_n/\rho} - \frac{w_q/r - \omega_q/\rho}{\omega_q/\rho}$$

$$\frac{\text{public total costs}}{\text{private total costs}} - \frac{\text{public total benefits}}{\text{private total net benefits}} \quad . \tag{30}$$

Countries setting similar incentives for tertiary education are GBR and FIN. The negative wedge on education equals 0.023 in the case of GBR and -0.003 in case of FIN. These almost vanishing wedges however result from rather different policies. FIN is a country in which the government assumes relatively high shares of costs and benefits while GBR is a country for which the shares are closer to the OECD average. More precisely, eq. (30) takes on the values $-0.003 = 0.74 - 0.74$ in the case of FIN and $0.023 = 0.49 - 0.47$ in the case of GBR. DEU and FRA are two countries with similar cost ratios of 0.73 and 0.75, respectively, but strongly deviating benefit ratios. More precisely, eq. (30) takes on the values $-0.29 = 0.73 - 1.03$ in the case of DEU and $0.23 = 0.75 - 0.52$ in the case of FRA. FRA and USA have similar benefit ratios of 0.52 and 0.53, respectively, but they strongly differ with respect to the cost ratios. Eq. (30) takes on the values $-0.21 = 0.31 - 0.53$ in the case of USA. Note that the reported cost and benefit ratios are on their part the result of particular policies. If one liked to better understand how countries end up with the incentives reported in Figure 2, one would have to study their policies in more detail. Lack of space forbids doing so. Each country seems to be a case of its own.

The numbers displayed by Figure 2 must clearly be taken with caution. This is not least because of the debatable data quality. An indication of the deficient quality is that for some countries the reported data strongly move over the years. An extreme example is Denmark for which the measurement of $-\varDelta_{E,n} = 0.73$ is based on OECD (2011). The same kind of calculations only yields -0.025 if based on OECD (2009). There are however no better data available and this may justify taking them as they are. [4]

The discussion of the empirical evidence is to be completed by a check of consistency. Assuming $v=0.3^{-1}$, the second-order condition (29) requires

$$mE < \tfrac{10}{3}\omega_n / (\rho\varphi + \omega_n) \ . \tag{31}$$

Recent estimates of wage premiums on tertiary education are provided by Strauss and de la Maisonneuve (2009). The estimates cover all countries entering Figure 2 except CZE, JPN, KOR, NZL, NOR, and TUR. The premiums estimated are gross, while the wage premiums guiding educational choice are net of taxes. As tax progression tends to compress wage premiums, one is however on the safe side when working with gross premiums instead of net premiums. Large gross premiums are reported for the United States (88.1) and Portugal

---

[4] This is why no figures are reported for females. Such figures are easily produced. However, the evaluation would require looking at country specific details. Scarcity of space does not allow doing so.

(81.9).[5] The given numbers refer to men, and they measure the average changes of upper-secondary degree holders in per cent. According to OECD data (2011, Table A9.3), $\omega_n / (\rho\varphi + \omega_n)$ equals 0.36 in the case of USA and 0.80 in case of PRT. As $\ln(1.881) = 0.63 < 1.2 = 3.6/3$, the United States pass the consistency check. The other countries, including Portugal, are less critical cases. Hence consistency seems to be no problem; the inequality (31) can safely be assumed to hold.

8. Summary

The basis for the empirical research on earnings determination is Mincer's equation. Individuals are assumed to make schooling decisions by maximizing earnings. Leisure costs of schooling and labour supply are neglected which fails to be convincing. Even more, the model produces empirically questionable implications. Specifically, the marginal internal rate of return to schooling would have to equal the optimizing individual's discount rate. The findings of empirical research are however different. Estimations of the return to high school and college completion regularly come up with a significant mark-up of the internal rate of return on the market rate of interest.

A first objective of the present study is to show a way of deriving a Mincer-type earnings function from the more standard assumption that individuals maximize lifetime utility. This objective is achieved by assuming that individuals do not only choose the amount of education but also a particular discipline out of a menu of competing disciplines where the elasticity of the return to studying any particular discipline is constant. Such constancy of elasticity is empirically well founded and is known in neuroscience as the power law of learning. An earnings function is derived on assuming that each level of education is supported by an individually optimal choice of discipline. The elasticity of the earnings function is shown to be increasing if the power law of learning is assumed to hold for the return to studying a particular discipline (Proposition 1). The Mincerian earnings function is obtained on assuming that the elasticity of the earnings function is not only increasing but linear increasing in the absolute amount of education.

---

[5] As *mE* equals the elasticity of a Mincerian earnings function, it is seen to be empirically justified to assume $\gamma < 1$.

The great advantage of the present model is that it allows one to analyse the efficiency of education policy in Ramsey's tradition. First steps towards this goal have been undertaken by Richter (2009; 2011). They suffered from explicitly assuming a strictly concave earnings function. The Mincerian earnings function is however convex in the absolute amount of education. Hence a second objective of the present paper is to show that the results of Richter (2009; 2011) largely extend to convex earnings functions. The greatest change comes from the modelling of the costs of education. When earnings are concave, it will be optimal for the taxpayer to divide first-period non-leisure between education and the supply of nonqualified labour. In such a situation foregone earnings add to the costs of education. As foregone earnings are reflected in market prices the tax planner can use them to optimize. This is not possible when earnings are convex. In the presented model convexity has the effect that it is not optimal for the taxpayer to divide non-leisure between education and labour supply. Hence there are no foregone earnings but only non-taxable costs of forgone leisure. As the latter are not reflected in market prices the planner is constrained in the choice of instruments. And yet, the essence of the results characterizing efficient policy carries over from the concave to the convex case. For the policy conclusions derived it is less relevant whether earnings functions are concave or convex. Pivotal is the question of whether the elasticity of the earnings function is increasing or decreasing. Increasing elasticity implies that education should be subsidized in effective terms because this helps to alleviate the distortions of labour taxation.

Rationalizing an effective subsidization of education by referring to the increasing elasticity of the earnings function contrasts with the reasons traditionally discussed in the literature. The traditional rationalization relies on assuming market failure. The empirical evidence of externalities and liquidity constraints is however mixed. Even if the evidence is considered to be supportive of some subsidization, it can at most rationalize subsidization to the extent that is needed to close the gap between the *laissez faire* and the first best. The rationalization discussed in the present paper, however, leads us beyond the first best. The marginal social cost of education should exceed the marginal social return from education when labour is taxed and when the elasticity of the earnings function is increasing. Such a rationalization of subsidization contrasts with the considerations suggested by the traditional derivation of the Mincer equation. The traditional derivation makes the individual's discount rate the focus of analysis. Subsidization is justified to the extent that the social discount rate exceeds the private one. Hence there is no direct connection to labour taxation.

If one is interested in a theory-based quantification of the efficient rate of subsidizing education, one has to assume particular functions of earnings and of utility. As the empirical evidence on average behaviour is quite supportive of a Mincerian earnings function, the real problem is the choice of the utility function. In section 5 the utility function is assumed to be quasi-linear. The sole justification for this choice comes from mathematical tractability. Assuming quasi-linearity and equality of nonqualified non-leisure choices across life periods, second-best policy is characterized by the education elasticity rule. The rule suggests that the effective rate of subsidizing education should increase in (i) the tax wedge on nonqualified labour and (ii) the elasticity of the nonqualified labour supply, in short: in the social cost of nonqualified-labour taxation. (If the earnings function were not assumed to be Mincerian, the efficient rate of effective subsidization would have to increase additionally in the second-order elasticity of the earnings function.) The derivation of the rule does not assume saving to be efficiently taxed. This is additionally indicative of the conceptual difference between the presented derivation of efficient education policy and the traditional Mincerian one with its focus on efficient discounting.

When confronting theory with effective policy, remarkable deviations become visible. OECD data suggest that various countries – including Belgium, Australia, Germany, and the United States – effectively tax tertiary education. There are other countries effectively subsidizing education, but they overshoot. Turkey, Denmark, and Austria are examples of overshooting. Efficient rates of effective subsidization of tertiary education range from one per cent (United States) to 14 per cent (Poland). By contrast, measured rates range from −45 per cent (Belgium) to 81 per cent (Turkey).

Such results may be questioned for various reasons. The debatable quality of the OECD data from which the numbers are computed may be one reason. However, the data are the best available and the sole role of the computations is to illustrate the theoretical derivations. One may equally question the partial-equilibrium flavour of the analysis and the underlying behavioural assumptions. As mentioned in footnote 2, much of the analysis may however be expected to extend beyond partial equilibrium. And as to the behavioural assumptions, those of a Mincerian earnings function and of a compensated elasticity of nonqualified labour supply of 0.3 have some backing in the literature. This is certainly not the case as far as quasi-linearity of the utility function is concerned. Working with other utility functions will, however, mean that one has to give up the hope of characterizing second-best policy by some

simple rule. One will have to resort to calibration techniques instead. Whether the quality of the results justifies the effort has to be seen.

A more critical shortcoming is the pure focus on efficiency. This contrasts with the weight that equity considerations receive in public discussions of education policy. On the other hand, one would not really be surprised to learn that equity gives reason to subsidize education effectively. For a paper analysing the close connection between equity and statutory subsidization of education in Mirrlees' tradition see Bovenberg and Jacobs (2005). The point made by the present paper is that labour taxation may provide strong reason for effectively subsidizing education if only certain assumptions on preferences and earnings are made which are not too far-fetched.

10. Appendices

Appendix A: The *proof of* (10) follows Richter (2009). It relies on taking partial derivatives of the Lagrange function $T - \lambda e$ with respect to $\varphi, \omega_n, \omega_q,$ and ρ:

$$\frac{\partial}{\partial \varphi}[T - \lambda e] = 0 \quad \Leftrightarrow \quad (\lambda - \frac{1}{\rho})e_\varphi = \Delta e_\varphi. \tag{32}$$

By Hotelling's lemma and by the definition of the Δ-operator, one obtains

$$e_\varphi = \rho E \quad \text{and} \quad \Delta e_\varphi = \Delta(\rho E) = \rho \Delta E + \frac{\rho - r}{r} E. \tag{33}$$

Plugging (33) into (32) yields $\lambda - 1/r = \Delta E / E$. Similarly one derives

$$\lambda - \frac{1}{r} = \frac{\Delta L_n}{L_n} = \frac{\Delta(GL_q)}{GL_q} = \frac{\Delta C_1}{C_1}. \square$$

By relying on the definition of the expenditure function and by invoking Hotelling's lemma one obtains

$$\rho C_{1x} + \overline{C}_{2x} + \rho \varphi E = \pi \omega_q (GL_q)_x + (1-\pi)\omega_n L_{nx} \quad \text{for } x = \varphi, \omega_n, \omega_q, \rho. \tag{34}$$

The relationship (34) extends to the Δ-notation:

$$\rho \Delta C_1 + \Delta \overline{C}_2 + \rho \varphi \Delta E = \pi \omega_q \Delta(GL_q) + (1-\pi)\omega_n \Delta L_n. \tag{35}$$

Eq. (12) is now easily proved by relying on (35) and (10):

$$\Delta \bar{C}_2 \underset{(35)}{=} \pi \omega_q GL_q \frac{\Delta(GL_q)}{GL_q} + (1-\pi)\omega_n L_n \frac{\Delta L_n}{L_n} - \rho C_1 \frac{\Delta C_1}{C_1} - \rho \varphi E \frac{\Delta E}{E}$$

$$\underset{(10)}{=} [\pi \omega_q GL_q + (1-\pi)\omega_n L_n - \rho C_1 - \rho \varphi E]\frac{\Delta E}{E} = \bar{C}_2 \frac{\Delta E}{E}. \ \square$$

Appendix B: Figure 2 is derived from OECD Indicators, Education at a Glance 2011, Tables A9.3 and A9.4. The numbers entering Figure 2 are computed as follows (the short form "i!J" refers to column J in Table A9.i):

$$\Delta_n = \frac{\text{taxes on nonqualified labour}}{\text{nonqualified labour income after tax}}$$

$$= \frac{\text{foregone taxes on earnings}}{\text{private foregone earnings}} = \frac{4!D}{2!F},$$

$$-\Delta_{E,n} \underset{(30)}{=} \frac{\text{public total costs}}{\text{private total costs}} - \frac{\text{public total benefits}}{\text{private total net benefits}}$$

$$= \frac{4!(C+D)}{3!(C+D)} - \frac{4!(E+F+G+H+I)}{3!(E+F+G+H+I+J)}.$$

References

Anderberg, D., 2009, Optimal policy and the risk-properties of human capital reconsidered, Journal of Public Economics 93, 1017–1026.

Anderson, J.R., 2005 (1980), Cognitive psychology and its implications, New York: Worth Publishers, 6th ed.

Blackburn, J.M., 1936, The acquisition of skill: An analysis of learning curves, IHRB Rep. No. 73, London.

Borjas, G.L., 2012 (1996), Labor Economics, New York: McGraw-Hill, 6th ed.

Bovenberg, A.L., and B. Jacobs, 2005, Redistribution and learning subsidies are Siamese twins, Journal of Public Economics 89, 2005–2035.

Cahuc, P., and A. Zylberberg, 2004, Labor Economics, Cambridge, Mass., MIT Press.

Card, D., 1999, The causal effect of education on earnings, in: Handbook of Labor Economics, O. Ashenfelter and D. Card., eds., Elsevier, Vol. 3, Chap. 30, 1801–1863.

Carneiro, P., J.J. Heckman, and E.J. Vytlacil, 2011, Estimating marginal returns to education, American Economic Review 101, 2754–2781.

Crossman, E.R.F.W., 1959, A theory of the acquisition of speed-skill, Ergonomics 2, 153–166.

Heckman, J.J., Lochner, L.J., and P.E. Todd, 2008, Earnings functions and rates of return, Journal of Human Capital 2, 1–31.

Keane, M.P., 2011, Labor supply and taxes: A survey, Journal of Economic Literature 49, 961-1075.

Mincer, J., 1958, Investment in human capital and personal income distribution, Journal of Political Economy 66, 281–302.

Mincer, J., 1974, Schooling, experience, and earnings, New York: Columbia University Press.

Newell, A. and P.S. Rosenbloom, 1981, Mechanisms of skill acquisition and the law of practice, in: Cognitive skills and their acquisition, J.R. Anderson, ed., Hillsdale, NJ: Erlbaum, 1–55.

Psacharopoulos, G., 1985, Returns to education, Journal of Human Resources 20, 583–604.

Richter, W.F., 2009, Taxing education in Ramsey's tradition, Journal of Public Economics 93, 1254–1260.

Richter, W.F., 2011, Efficient education policy – A second-order elasticity rule, FinanzArchiv / Public Finance Analysis 67, 1–7.

Ritter, F.E., and Schooler, L.J., 2001, The learning curve, in: International Encyclopedia of the Social and Behavioral Sciences, W. Kintch, N. Smelser, and P. Baltes, eds., Oxford, 8602−8605.

Strauss, H., and C. de la Maisonneuve, 2009, The wage premium on tertiary education: New estimates for 21 OECD countries, OECD Journal: Economic Studies.

Willis, R.J. and S. Rosen, 1979, Education and Self-Selection, Journal of Political Economy 87, S7–S36.