



Working Papers

www.cesifo.org/wp

Education: Optimal Choice and Efficient Policy

Wolfram F. Richter
Kerstin Schneider

CESIFO WORKING PAPER NO. 5352
CATEGORY 5: ECONOMICS OF EDUCATION
MAY 2015

An electronic version of the paper may be downloaded

- *from the SSRN website:* www.SSRN.com
- *from the RePEc website:* www.RePEc.org
- *from the CESifo website:* www.CESifo-group.org/wp

ISSN 2364-1428

Education: Optimal Choice and Efficient Policy

Abstract

The research on earnings determination is based on the Mincer-Becker assumption that individuals decide on schooling by maximizing income. This paper offers an alternative and less restrictive approach based on utility maximization. Using this approach, we analyze the efficiency of education policy in Ramsey's tradition. Distortive wage taxation is shown to provide an efficiency reason for subsidizing education in effective terms. Second-best policy is confronted with empirical evidence for OECD countries.

JEL-Code: J240, H210, I280.

Keywords: schooling choice and earnings functions, utility vs. earnings maximization, power law of learning, second-best taxation in Ramsey's tradition, empirical evidence.

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

Kerstin Schneider
University of Wuppertal
Wuppertal Research Institute for the
Economics of Education
Germany – 42119 Wuppertal
Schneider@wiwi.uni-wuppertal.de

May 2015

This is a fully revised and extended version of a paper written by the first author and previously circulating under the title “Mincer equation, power law of learning, and efficient education policy”.

1. Introduction

The traditional approach to modeling schooling choice relies on the assumption that individuals maximize the present value of lifetime earnings. Although appealing at first sight, the idea that the choice of schooling results from strict income maximizing behavior is challenged by the persuasive evidence of significant nonpecuniary returns and costs of education. Summarizing the literature on the nonpecuniary returns, Oreopoulos and Salvanes (2011, p. 180) conclude that the returns “are both real and important”. As to the costs of education, Heckman et al. (2006, p. 436) suggest that psychic costs “play a very important role” and describe the evidence against strict income maximization as “overwhelming”.

The obvious problem raised is the black-box character of nonpecuniary returns and costs. Rather than explaining schooling choice, the reference to psychic costs, not further defined, concedes the limits in understanding schooling choice. As Heckman et al. (2006, p. 436) put it: “Explanations based on psychic costs are intrinsically unsatisfactory”.

The present paper contributes to the debate by offering an alternative model of schooling choice which is shown to be empirically promising and theoretically more convincing than the income maximizing framework. Our approach is based on: (i) utility maximization rather than income maximization, (ii) the recourse to learning theory, and (iii) the shift in focus away from optimal choice of schooling towards the analysis of efficient education policy. Each of these components is well established in the literature; progress comes from combining the three, as we hope to convince the reader.

The obvious appeal of utility maximization is that it is the standard assumption in the neoclassical paradigm of individual behavior. Relying on this assumption serves as a basis for efficiency analysis, which is a major objective of the present paper. At first sight, utility and income maximization are concepts with equivalent behavioral implications; hence replacing one concept by the other is not expected to have major effects. However, utility and income maximization lead to different conclusions if the earnings function fails to be concave. With a strictly convex earnings function, as is suggested by empirical evidence, utility maximization implies that the cost of foregone leisure exceeds the cost of foregone income. Hence, the marginal internal rate of return to

schooling is systematically overestimated when using the observable cost of foregone earnings as a proxy for the unobservable cost of foregone leisure.

The derivation of specific *a-priori* properties of earnings functions is one major contribution of the paper. By recourse to learning theory, we justify our key assumption that earnings functions feature increasing elasticity in the amount of schooling. An increasing elasticity is not a restrictive assumption and in fact nests the Mincerian earnings function as the special case with a proportionally increasing elasticity. While functional flexibility is one advantage of the present approach, another appealing property is that the estimated (Mincer) coefficient of years of schooling in a regression of log earnings does not have to be interpreted as a rate of discount. Hence, if marginal internal rates of return to schooling are regularly estimated to exceed the costs of funds, this is no evidence against the model developed in the present paper. Moreover, assuming an increasing elasticity is shown to be pivotal when characterizing efficient education policy in Ramsey's tradition: Distortive wage taxation requires subsidizing education effectively, if the earnings function displays increasing elasticity in the amount of schooling.

The recommendation of effectively subsidizing education is finally confronted with empirical evidence for OECD countries. It is shown that education policies in OECD countries tend towards effective subsidization of education, as optimal Ramsey policy suggests. Furthermore, there is evidence that the extent of subsidization increases with the public share of the benefits of education.

The paper is organized as follows. Section 2 briefly reviews the related literature. Section 3 applies learning theory to justify our key assumption of an increasing elasticity of earnings functions. Section 4 sets up a standard model of a representative individual who invests in education by maximizing lifetime utility. Section 5 characterizes efficient education policy in Ramsey's tradition. Section 6 confronts second-best policy with empirical evidence from a sample of OECD countries. Section 7 concludes.

2. Related literature

This paper unifies two strands of the literature. The older strand has emerged from labor and education economics. It has been initiated by Mincer (1958) and Becker (1964) and is positive theoretic in spirit. The focus is on schooling choice and earnings determination. The other strand has grown out of the public economics literature. It is normative theoretic and it is the starting point for the analysis of the optimal taxation of education. Examples are Bovenberg and Jacobs (2005), Anderberg (2009), and Richter (2009).

The older literature, following Becker and Mincer, estimates earnings functions with schooling being modeled as continuous time spent in education. The bottom line of this strand of literature, well surveyed by Card (1999), is that the growth rate of earnings as a function of schooling is higher than a typically assumed real rate of discount. This raises the puzzling question of why individuals do not continue schooling despite the high returns.

More recent contributions follow Roy (1951) and Willis and Rosen (1979) in modeling schooling choice as a problem of self-selection. In line with the theory of comparative advantage, the individual is assumed to make a discrete choice between continuing or not continuing schooling. However, the estimated marginal internal rates of return to schooling still substantially exceed the level of real interest rates (Heckman et al., 2006; Heckman et al., 2008). One possible, and often suggested, explanation refers to liquidity constraints. However, even though public concerns about credit constraints are strong, the impact of the latter on tertiary education is estimated to be relatively weak (Carneiro and Heckman, 2002). All this has led Heckman et al. (2008) to challenge the assumption that individuals simply maximize income when making schooling decisions. They suggest accounting for heterogeneity and including psychic costs in the analysis. As compared to low ability individuals, more able individuals have lower psychic costs of attending college.

A seminal paper by Carneiro, Heckman and Vytlačil (2011) presents returns to education, explicitly accounting for individual observed and unobserved heterogeneity as well as sorting issues. The average treatment effect is lower than the treatment effect on the treated but substantially higher than the treatment effect on the untreated. And interestingly, the effect on the untreated is below a typically assumed discount rate. Carneiro et al. (2011) estimate the distribution of the marginal treatment effects and an MP RTE (marginal policy relevant treatment effect) resulting from a small

change in education policy. The magnitude of the MP RTE varies, depending on the policy intervention, between 0.087 for a policy that changes the probability to attend college by a small proportion and 0.015, for a policy that expands each individual's probability of attending college by the same proportion. However, the focus of Carneiro et al. is not on efficient education policy, but on the estimation of returns to tertiary education. From the perspective of the present paper, the relevance of unobservables on the decision to attend college is a key feature. The "unobserved component of the desire to go to college" (Carneiro et al., 2011 p. 2758) refers to the importance of utility rather than income maximization.

The public economics literature concerned with the choice of schooling and education policy has developed fairly independently from the labor economics literature. In fact, there is hardly any cross acknowledgment between the two literatures. A notable exception is a recent paper by Findeisen and Sachs (2014). The authors calibrate a model combining optimal nonlinear income taxation in the tradition of Mirrlees (1971) with discrete schooling choice in the tradition of Roy (1951), Willis and Rosen (1979), Heckman et al. (2006), and others.

Both the paper of Findeisen et al. and the present paper assess the efficiency of education policy but choose different modeling strategies. While Findeisen et al. follow Mirrlees (1971) and Bovenberg et al. (2005) in allowing for individual heterogeneity, the present paper is grounded in Ramsey's tradition and studies the efficiency of educational incentives in the framework of a representative taxpayer. Both modeling strategies have advantages and disadvantages.

The model of Findeisen et al. is rich enough to incorporate multidimensional heterogeneity, idiosyncratic risk, and borrowing constraints. The downside of this complexity is simplicity in modeling details. For example, individual preferences are assumed to be quasi linear. Furthermore, psychic costs, which are not well understood, are pivotal for explaining schooling choice. By contrast, the present paper does not explicitly rely on psychic costs and builds on arbitrary utility functions. This level of generality comes at the cost of neglecting heterogeneity. However, we argue that disregarding individual heterogeneity is rather appropriate when analyzing policy issues. After all, tax and education policy is not designed for individuals or small groups characterized by distinct social criteria. Tax and education policy must set efficient incentives for individuals at large. Although we cannot determine the efficiency frontier with our data, we can provide valuable

inferences for policy makers by exploiting differences in tax and education policy between countries. For instance, we can check whether countries effectively subsidize education, which is the efficient policy. While most countries in our sample effectively subsidize education, countries such as Ireland or Australia tax education effectively. Other countries like Belgium pursue an education policy which strongly deviates from the majority of OECD countries and raises policy questions as well.

3. The power law of learning and earnings curves

Most tasks get faster with routine. This observation is not surprising. However, and this is in fact surprising, the rate of improvement appears to follow a 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). In neuroscience this is known as the *power law of learning* (Newell and Rosenbloom, 1981; Anderson, 2005). One of the early studies reporting detailed data is Blackburn (1936). The study reports the productivity of seven individuals accomplishing five specific tasks repeatedly. The individuals were asked to sort packs of 42 cards, to cross out all occurrences of the letter *e* in a nonsense text, to transform short texts by some rather complicated code substitution, to add digits, and to learn a stylus maze. Crossman (1959) confirms the power law of learning for the first four experiments, while raising doubts about the applicability for maze learning. Figure 1 displays the learning curves of three individuals when crossing out e’s, doing code substitution, and adding digits, respectively.

The empirical evidence on learning curves suggests to define individual productivity by $H = H(D, E)$, with log productivity being a linear function of log experience

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

The variable E measures experience, while D denotes 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 .

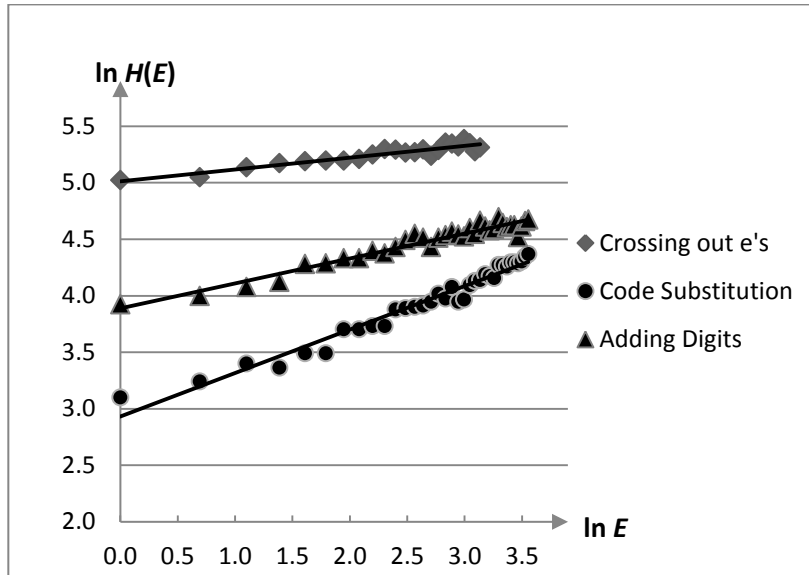


Figure 1: Learning curves when crossing out e's, doing code substitution, and adding digits. Data taken from Blackburn (1936)². Logarithmic scaling. E is the number of trials and $H(E)$ is the average number of achievements in 100 seconds.

In this paper, the power law of learning is assumed to extend on two counts. First, we assume that eq. (1) does not only apply to simple tasks but also to the acquisition of earnings related skills (“human capital”) at large. Consider for instance the study of economics. The assumption is that students become better at doing economics by solving economics problems again and again. Repetition enhances the productivity with an elasticity, not necessarily the same for all students, but constant for each individual student. Extending the assumption of a constant elasticity from simple tasks to the acquisition of earnings related skills is justified by the observation that the power law reflects a behavioral regularity which is “ubiquitous” (Newell and Rosenbloom, 1981; Ritter and Schooler, 2001). It seems to be tied to the neurological functioning of the human brain. The extension of the power law to the acquisition of earnings related skills suggests interpreting the variable D in eq. (1) as a *subject* or a *discipline* to be studied and interpreting the variable E as the time spent on *learning* or *education*.

The second extension is that eq. (1) is assumed to apply to the monetary compensation H of the physical productivity H^p . This is an extension as the power law of learning features a behavioral

² 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.

regularity in the acquisition of physical productivity. The extension is justified whenever the market valuation w defined as the ratio of compensation and physical productivity H/H^p equally satisfies eq. (1). This could be meant in a non-trivial sense with $w = w(D, E)$, i.e. the wage depending on both the subject and the experience, or in the trivial sense with $w = w(D)$, i.e., the wage only depending on the subject. For the sake of simplicity, we assume $w = w(D)$, so that $H(D, E)$ can be written as $w(D) \cdot H^p(D, E)$. This implies a price for disciplinary skills, such as a law degree or a degree in engineering; the time needed for acquiring some subject-related skill is not valued. Education, or time in education, E , is remunerated only via the enhancing effect on physical productivity, H^p . This model allows for differences in the market valuations of disciplinary degrees, while income differences between graduates of the same subject are explained by differences in real productivity. Clearly, this feature describes primarily tertiary education, where the choice of disciplines matters. In secondary education, the acquired skills are more homogenous. In fact, whenever central university entrance exams or high school exit exams are administered, it is implicitly assumed that secondary education produces homogenous education or human capital.

Given that $H(D, E)$ in (1) can be interpreted as monetary compensation, we follow Willis and Rosen (1979) in assuming that individuals maximize $H(D, E)$ in D for given E . Assume $D(E)$, the optimal discipline, to exist for all E . The resulting function $W(E) \equiv H(D(E), E)$ is called the *earnings function*. The following proposition characterizes the shape of earning functions when learning functions are isoelastic. The proposition is just as trivial as it is fundamental for the subsequent analysis.

Proposition 1: The power law of learning implies that the elasticity $\gamma(E) \equiv \eta(D(E))$ of the earnings function $W(E) \equiv H(D(E), E)$ is increasing in education.

The proof is straightforward. Assume $E_2 > E_1$ and D_i being the optimal choice at E_i ($i=1, 2$). Hence $W(E_2) = H(D_2, E_2) \geq H(D_1, E_2)$ and $W(E_1) = H(D_1, E_1) \geq H(D_2, E_1)$. Eq. (1) implies

$$h(D_2) + \eta(D_2) \ln E_2 \geq h(D_1) + \eta(D_1) \ln E_2$$

and

$$h(D_1) + \eta(D_1) \ln E_1 \geq h(D_2) + \eta(D_2) \ln E_1 .$$

Adding these two inequalities, yields

$$[\ln E_2 - \ln E_1] \eta(D_2) \geq [\ln E_2 - \ln E_1] \eta(D_1), \text{ i.e.}$$

$$\gamma(E_2) = \eta(D(E_2)) = \eta(D_2) \geq \eta(D_1) = \eta(D(E_1)) = \gamma(E_1). \quad \square$$

The proof is illustrated in Figure 2. The figure shows linear but possibly intersecting learning functions for subjects D_1 and D_2 . The slope of each individual learning function is constant by assumption. The slope of the upper envelope is then necessarily increasing, which is stated in Proposition 1.

The following analysis relies on assuming the elasticity of the earnings function W to be increasing in education, E , which will – in a later part of the analysis – turn out to be crucial for proving that it is efficient to subsidize education. An earnings function with a proportionally increasing elasticity, $\gamma(E) \equiv mE$, is the simplest case of an earnings function with an increasing elasticity. This is equivalent to assuming a constant growth rate and a log-linear earnings function, which is the basic Mincer equation.

$$\gamma(E) \equiv mE \Leftrightarrow \ln W = W_0 + mE \text{ with } m > 0. \quad (2)$$

A log-linear function is strictly convex in E . However, convexity is not implied by assuming an increasing elasticity. An earnings function with an increasing elasticity may well be strictly concave. An example is $W(E) = W_0 E^\gamma + W_1$ with $W_1 > 0$ and $\gamma < 1$.

The particular appeal of the recourse to learning theory is to rationalize log linearity of earnings without interpreting the Mincer coefficient m as a discount rate. In order to derive eq. (2) with some arbitrary value of m , one only has to assume (i) the power law of learning, (ii) individuals who maximize earnings over D for each given E , and finally, (iii) functional simplicity in the sense that the elasticity of the earnings function is not only increasing but proportionally increasing. Clearly,

functional simplicity is not easily justified by purely economic or neuro-scientific reasoning. It is therefore important to note that the following analysis does not rely on log linearity of earnings. All that is needed is the assumption that the elasticity of the earnings function is increasing. The log-linear case, as the most popular and empirically estimated form, is just an example illustrating the approach.

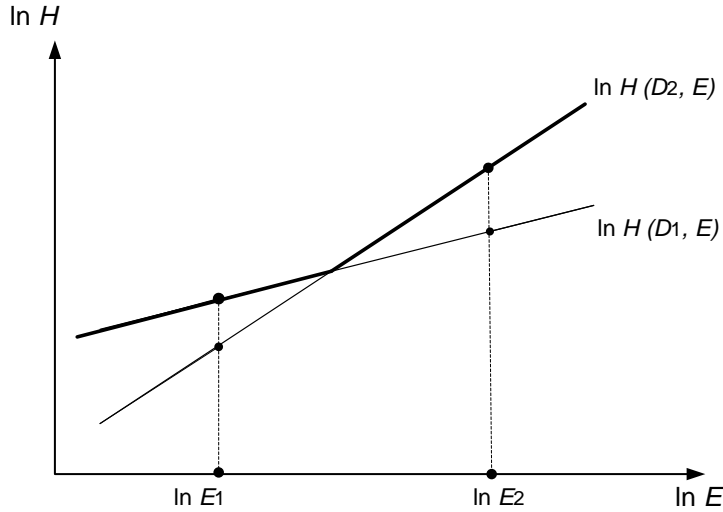


Figure 2: The upper envelope of linear functions is convex.

4. Household behavior

In the following, household behavior is modelled in a standard way. The focus is on a representative taxpayer living for two periods and deriving strictly increasing utility, U , from consumption, C_i , and strictly decreasing utility from non-leisure time, L_i , in periods $i=1,2$. The function $U = U(C_1, C_2, L_1, L_2)$ is strictly quasi-concave. Non-leisure in period 2, L_2 , is identical to the second-period labor supply. By contrast, in period 1 only $L_1 - E$ is the time spent working, while E is the time spent on education. First-period labor supply earns a constant wage rate, ω_1 ; the return to second-period labor, however, depends on the amount of education. Workers get paid $W(E) \equiv \omega_2 G(E)$ per unit of time, where ω_2 is constant and $G = G(E)$ is interpreted as an *earnings function* assumed to be twice differentiable and with an increasing elasticity $\gamma(E) \equiv EG'/G$. Hence, as noted

above, G may be either convex or concave. The wage rate $W(E)$ is written as the product of ω_2 and $G(E)$ to account for the taxation of labor income. In the absence of taxation, ω_2 equals one and $W(E)$ equals $G(E)$. Introducing labor taxes with $\omega_2 \leq 1$, renders $W(E) \leq G(E)$. Given a positive choice of education, $E > 0$, second-period labor is interpreted as qualified labor. Likewise, the quantities $L_1 - E$ and L_1 are interpreted as nonqualified labor and nonqualified non-leisure, respectively. Education may cause an opportunity cost in the form of foregone earnings and the direct cost of education, like tuition. Both costs are assumed to be linear in time spent on education. The cost of foregone earnings is modelled by $\omega_1 E$ and the cost of tuition is denoted by φE . The share of first-period income that is spent neither on education nor on consumption is first-period savings:

$$S = \omega_1(L_1 - E) - \varphi E - C_1 = \omega_1 L_1 - (\varphi + \omega_1)E - C_1 \quad (3)$$

By way of normalization, the price of consumption is set equal to one. The gross rate of return to saving is denoted by ρ and we assume perfect capital markets. In particular, there are no credit constraints, hence negative savings are no problem. The only inefficiency modelled in this analysis comes from taxation.

All second-period income is spent on consumption:

$$C_2 = \rho S + \omega_2 G(E)L_2 \quad (4)$$

Substituting for S in (3) and (4) yields the lifetime budget constraint:

$$C_1 + C_2/\rho = \omega_1 L_1 + \omega_2 G(E)L_2/\rho - (\varphi + \omega_1)E \quad (5)$$

Maximizing utility $U(C_1, C_2, L_1, L_2)$ in $C_1, C_2, L_1, L_2, E \geq 0$ subject to (5) and $L_1 \geq E$ requires maximizing the surplus income generated by education,

$$\begin{aligned} Y(L_1, L_2) &= Y(L_1, L_2; \omega_2/\rho(\varphi + \omega_1)) \\ &\equiv \max_{0 \leq E \leq L_1} [\omega_2 G(E)L_2/\rho - (\varphi + \omega_1)E] . \end{aligned} \quad (6)$$

Eq. (6) looks like a discrete version of income maximization *à la* Mincer and Becker. Note, however, that eq. (6) assumes linear costs of education, while the standard Mincer schooling model

implicitly assumes increasing costs. This has implications for the characterization of optimal behavior and needs some careful analysis.

When maximizing (6), three scenarios are of interest. In the first one, it is optimal for the taxpayer to remain unqualified, i.e. $E = 0$. This is the case when the incentive to invest in education is too weak. This could be the case because, for instance, the wage premium is low or the tax on qualified labor is high. Although this is empirically relevant, in the following the sole focus is on $E > 0$. In the second scenario, maximizing the net income of education has an interior solution with $E \in (0, L_1)$. Obviously, this scenario requires the earnings function to be concave. Concavity, however, is not ensured by assuming an earnings function with an increasing elasticity; hence the earnings function might well be convex. The following analysis therefore differentiates between the two following scenarios. The *interior solution* assumes $0 < E < L_1$ and a concave earnings function with $G''(E) < 0$, while the *upper corner solution* assumes $E = L_1$ and a convex earnings function $G''(L_1) \geq 0$. Note that the taxpayer's demands and supplies only depend on ω_1 if utility maximization has an interior solution. In this case, the costs of foregone earnings and foregone leisure are identical, $\omega_1 = MRS^1 \equiv -U_{L_1}/U_{C_1}$. By contrast $\omega_1 \leq MRS^1$, i.e., the cost of foregone leisure exceeds the cost of foregone earnings, if utility is maximized at an upper corner solution. (Subindices of functions indicate partial derivatives.)

Note, that in both scenarios, the optimal choice of education can be characterized by the equality of the *private marginal internal rate of return to education* and the private rate of discount,

$$IRR_{priv} \equiv \frac{\omega_2 G'(E) L_2}{\varphi + MRS^1} = \rho . \quad (7)$$

In eq. (7) the private marginal internal rate of return to education is equal to the ratio of the return to education in the second period, $\omega_2 G'(E) L_2$, and the opportunity cost of education in period 1, $\varphi + MRS^1$. Eq. (7) shows the pivotal difference to the standard Mincer schooling model. In the standard Mincer model, the cost of education simply equals the cost of foregone earnings. Other monetary costs like the cost of tuition have been included in extended versions. However, the point to be stressed is that the maximization of income requires all costs to be reflected in market prices. The same holds for (7), if the cost of foregone earnings equals the cost of foregone leisure, $\omega_1 = MRS^1$. This equality, however, holds only if maximizing utility yields an interior solution for the

optimal E . This in turn implies that the earnings function is concave. By contrast, if G is convex, as is strongly supported by empirical evidence, maximizing utility generates a corner solution with the cost of foregone leisure exceeding the cost of foregone earnings. Thus, IRR_{priv} is systematically overestimated when the cost of education is estimated by the cost of foregone earnings rather than by the cost of foregone leisure.

Maximizing the surplus income from education, Y , generates increasing returns. This can hardly surprise if the earnings function is convex. However, increasing returns also result in the concave case. More precisely, interior solutions generate increasing returns to qualified labor,

$$\frac{d^2Y}{dL_2^2} = \frac{\omega_2}{\rho} G' \frac{dE}{dL_2} = -\frac{\omega_2}{\rho} \frac{G'^2}{G''L_2} > 0.$$

Y is convex in L_2 as optimal education increases monotonically in L_2 . By contrast, upper corner solutions generate increasing returns with respect to nonqualified non-leisure,

$$\frac{d^2Y}{dL_1^2} = \frac{\omega_2}{\rho} G''L_2 \geq 0.$$

Thus, the convexity of Y is implied by the convexity of G . The marginal net return to nonqualified non-leisure, $\omega_2 G' L_2 / \rho - (\varphi + \omega_1)$, increases in $L_1 = E$.

The convexity of surplus income 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 strictly positive solutions may fail to exist. Still, the following analysis only looks at first-order conditions. The implicit assumption is, firstly, that the taxpayer discards all solutions of the first-order conditions which fail to be globally optimal and, secondly, that a global optimum exists at positive values of the choice variables. The latter requires that the supply of non-leisure is sufficiently inelastic. More precisely, the convexity of Y must be dominated by the convexity of the cost of foregone leisure.

5. Second-best policy

We now turn to optimal policy design. The government needs to raise revenue. There are four possible linear tax instruments, each of which is distorting the individual's decision. The taxes can be levied on labor income in the first and the second period, on the cost of tuition, and on the returns 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 $\omega_1, \omega_2, \varphi, \rho$. The prices before taxes and subsidies are exogenous and denoted by w_1, w_2, f, r .³ The tax on labor income in period $i = 1, 2$ 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 be negative, i.e., a subsidy. Government's net revenue amounts to

$$T \equiv (w_1 - \omega_1)(L_1 - E) + (\varphi - f)E + [(w_2 - \omega_2)G(E)L_2 + (r - \rho)S]/r. \quad (8)$$

In order to characterize second-best tax policy it is convenient to work with the taxpayer's expenditure function, which is defined by

$$e(\omega_1, \omega_2, \varphi, \rho; u) \equiv \min[\rho C_1 + C_2 + \rho(\varphi + \omega_1)E - \rho\omega_1 L_1 - \omega_2 G(E)L_2] \quad (9)$$

in C_1, C_2, L_1, L_2, E subject to $U(C_1, C_2, L_1, L_2) \geq u$ and $L_1 \geq E$. Assume that the expenditure function is twice differentiable. Relying on Hotelling's lemma yields the identities $e_{\omega_1} = -\rho(L_1 - E)$, $e_{\omega_2} = -G(E)L_2$, $e_{\varphi} = \rho E$, and $e_{\rho} = C_1 + \varphi E = -S$, where the variables L_i, E, S , and C_i have to be interpreted as Hicksian supply and demand functions to be evaluated at $\omega_1, \omega_2, \varphi, \rho$, and u . Note that the expenditure function is independent of ω_1 , $e_{\omega_1} \equiv 0$, whenever $L_1 = E$, i.e., the individual spends all non-leisure time on education. Using these definitions, eq. (8) can be written as

$$T = \frac{1}{\rho}(\omega_1 - w_1) e_{\omega_1} + \frac{1}{\rho}(\varphi - f)e_{\varphi} + \frac{1}{r}[(\omega_2 - w_2)e_{\omega_2} + (\rho - r)e_{\rho}]. \quad (10)$$

³ It has been suggested above to interpret $G(E)$ as monetary productivity which then requires $w_2 = 1$. If one chose instead to interpret education as a labor augmenting activity and $G(E)L_2$ as *effective qualified labor*, w_2 would equal the latter's marginal productivity. It is a straightforward exercise to endogenize the prices before taxes and subsidies in this case. However, endogenization does not produce interesting new insights. Assuming no pure profit to accrue to the private sector so that the production efficiency theorem applies, endogenizing has no structural effect on efficient education policy.

The planner's objective is to maximize revenue T in $\omega_1, \omega_2, \varphi, \rho$ subject to the taxpayer's budget constraint, $e(\omega_1, \omega_2, \varphi, \rho; u) = 0$. In the Appendix 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 first-order conditions:

$$\widehat{E} = \widehat{L}_1 = \widehat{GL}_2 = \widehat{C}_1, \quad (11)$$

where the hat notation denotes relative changes, $\widehat{X} \equiv \Delta X/X$. The total differentiation operator Δ is defined on arbitrary functions $X = X(\omega_1, \omega_2, \varphi, \rho; u)$ by

$$\Delta X \equiv \frac{1}{\rho}(\omega_1 - w_1) X_{\omega_1} + \frac{1}{\rho}(\varphi - f) X_{\varphi} + \frac{1}{r}(\omega_2 - w_2) X_{\omega_2} + \frac{\rho-r}{r} X_{\rho} \quad (12)$$

According to (12), 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 the Appendix the equations in (11) are shown to imply

$$\widehat{C}_2 = \widehat{E}. \quad (13)$$

By applying hat calculus, one obtains $\widehat{GL}_2 = \widehat{L}_2 + \widehat{G} = \widehat{L}_2 + \gamma \widehat{E}$ where γ is the elasticity of the earnings function. Together with (11) this implies

$$\widehat{L}_2 = (1 - \gamma) \widehat{L}_1. \quad (14)$$

Summarizing (11), (13), and (14) yields:

Proposition 2: Second-best efficiency requires reducing education, consumption, nonqualified labor, and effective qualified labor equi-proportionately. Qualified labor, however, is reduced less than equi-proportionately.

According to the proposition it is second-best to reduce all quantities E, C_1, C_2, L_1 , and GL_2 entering the taxpayer's budget constraint by the same proportion, when all these demand and supply functions are interpreted in the Hicksian sense. The equi-proportionate reduction is clearly in line with Ramsey's (1927) characterization of efficient taxation. The less standard result concerns the

change in qualified labor supply, L_2 . Efficiency requires reducing qualified labor relatively less than non-qualified labor. The factor is $1 - \gamma$ and hence it is decreasing in γ , the elasticity of the earnings function. In other words, the more elastic the individual earnings function, the less should qualified labor be reduced relative to nonqualified labor. While this result is quite intuitive, it is clearly in contrast to Ramsey's Rule of reducing all household choices equi-proportionately. In the model with endogenous education, effective qualified labor GL_2 is reduced equi-proportionately. Qualified labor, however, should be reduced less than proportionately, as $G = G(E)$ reacts elastically. For earnings functions with elasticity γ greater than one, L_2 even increases (cf. eq. (14)).

The optimal choice of education is characterized in eq. (7). It states the equality of the private marginal internal rate of return to education and the private rate of discount. This condition is equivalent to the condition that the *marginal return to education* equals the (*effective*) *marginal cost of education*,

$$MR \equiv G'(E)L_2 = \rho(\varphi + MRS^1)/\omega_2 \equiv MC . \quad (7')$$

Applying hat calculus to the left-hand side of eq. (7') yields

$$\widehat{MR} = \frac{\Delta MR}{MR} = \frac{\Delta(G'L_2)}{G'L_2} = \frac{\Delta G'}{G} + \frac{\Delta L_2}{L_2} = \left[\frac{EG''}{G'} + 1 - \gamma \right] \widehat{E} = \gamma_\gamma \widehat{E} \quad (15)$$

where $\gamma_\gamma \equiv E\gamma'/\gamma$ denotes the second-order elasticity of the earnings function. This second-order elasticity is necessarily positive as the elasticity of the earnings function is assumed to be increasing in E (Proposition 1). γ_γ equals one if the earnings function is log-linear. As \widehat{E} is negative, given that taxation is to raise positive revenue, it follows from (15) that the efficient change in the marginal return to education, \widehat{MR} , is necessarily negative as well. Since MR equals MC , the efficient change in the marginal cost of education has to be negative as well. Applying hat calculus to the right-hand side of eq. (7') yields

$$\begin{aligned} 0 > \widehat{MC} &= \frac{\Delta MC}{MC} = \frac{\Delta(\rho\varphi) + \Delta(\rho MRS^1)}{\rho(\varphi + MRS^1)} - \frac{\Delta\omega_2}{\omega_2} \\ &= \frac{(\varphi - f) + (\rho - r)(\varphi + MRS^1)/r + \rho\Delta MRS^1}{\rho(\varphi + MRS^1)} - \frac{\omega_2 - w_2}{r\omega_2} \end{aligned}$$

$$= \frac{w_2/r}{\omega_2} - \frac{f + MRS^1 - \rho \Delta MRS^1}{\rho(\varphi + MRS^1)} \quad (16)$$

If $MRS^1 = \omega_1$ holds, this implies $\Delta MRS^1 = \Delta \omega_1 = (\omega_1 - w_1)/\rho$ and $MRS^1 - \rho \Delta MRS^1 = w_1$. However, as argued above, MRS^1 equals ω_1 only if the earnings function is concave. If the earnings function is convex, it is nevertheless suggestive to write $MRS^1 \equiv \omega_1^s$ and $MRS^1 - \rho \Delta MRS^1 \equiv w_1^s$ but to interpret ω_1^s and w_1^s as the *private* and *social shadow costs of foregone nonqualified leisure*. Hence, eq. (16) can be restated as

$$0 > \rho \widehat{MC} = \frac{w_2/r}{\omega_2/\rho} - \frac{f + w_1^s}{\varphi + \omega_1^s} = \frac{w_2 G' L_2 / r - (f + w_1^s)}{\varphi + \omega_1^s} \equiv \Delta_E \quad (17)$$

where Δ_E is interpreted as *effective wedge on education* and eq. (7) has been used. The inequality in (17) is equivalent to $w_2 G' L_2 / r < f + w_1^s$. Figure 3 visualizes such a scenario for concave and convex earnings functions.

The effective wedge on education is negative, which can be restated in various equivalent and more common ways. For instance, a negative effective wedge is equivalent to saying that the *social marginal internal rate of return to education*, $IRR_{soc} \equiv w_2 G' L_2 / (f + w_1^s)$, falls short of the rate of discount r . Alternatively, the (*effective*) *marginal social cost*, $r(f + w_1^s) / w_2$, can be required to exceed the *marginal social return*, $G' L_2$. If those conditions hold, *education* is said to be (*effectively*) *subsidized*. This definition differs from the conventional definition, according to which education is subsidized when the cost of tuition is subsidized, $\varphi < f$. Reducing the discussion on the cost of tuition is, however, too restrictive and partial. The entire tax and transfer system affects the incentives for education. Clearly, a negative value of Δ_E may result from subsidizing the cost of tuition. But there are other components in the tax and transfer system that result in negative values of Δ_E . For instance, reducing the tax on qualified labor increases the statutory return to education, w_2 . A tax on the return to saving, r , reduces the cost of education and works in the same direction. If the earnings function is concave and the taxpayer supplies nonqualified labor $L_1 - E > 0$ in the first period, another way of encouraging education is to tax first-period nonqualified wage income, which reduces the cost of foregone earnings. However, if the earnings function is convex so that the taxpayer's optimization implies $L_1 = E$, there are no foregone earnings but only non-taxable costs of foregone leisure.

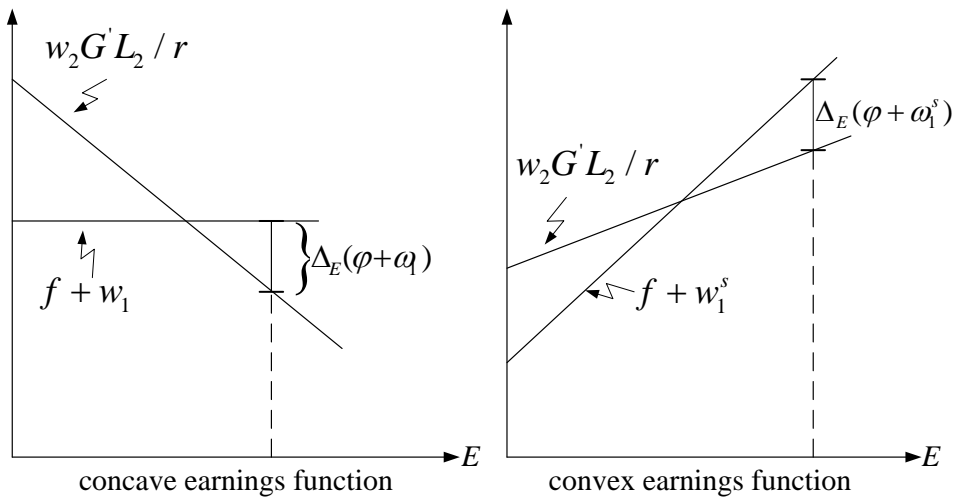


Figure 3: The effective wedge, Δ_E , when education is subsidized

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

Note that Proposition 3 holds for any particular utility function. The key assumption is the increasing elasticity of the earnings function. The utility function may be arbitrary except for the assumptions needed to guarantee that the planner's optimization is well behaved. This is noteworthy when comparing Proposition 3 with results characterizing the efficient taxation of savings. In the Ramsey model with finite periods, the question of whether it is efficient to tax savings or not critically depends on the choice of the utility function. This is a remarkable difference which can be explained as follows.

Savings result in wealth generating capital income without requiring extra effort. By contrast, education enhances productivity. This increase in productivity results only in higher income if combined with labor, which requires additional effort. Hence earning qualified labor income involves a double margin, educational choice and labor supply, while earning capital income does not. This difference explains and justifies differential taxation.

The theoretical analysis produces an optimal policy rule. It is inviting to look at the education policy of OECD countries and to check whether they pursue efficient education policies. In the following, we rewrite the efficiency condition derived in the analysis in a way that makes it suitable for empirical analysis.

Efficiency is characterized by the equality of \widehat{MR} and \widehat{MC} . Using (15), (17), (7), and $MRS^1 \equiv \omega_1^s$, this condition implies

$$\rho\gamma\widehat{E} = \rho\widehat{MR} = \rho\widehat{MC} = \frac{w_2G'L_2/r-(f+w_1^s)}{\varphi+\omega_1^s} = \frac{\left(\frac{w_2}{r}-\frac{\omega_2}{\rho}\right)G'L_2-(f-\varphi)-(w_1^s-\omega_1^s)}{\omega_2G'L_2/\rho}. \quad (18)$$

The translation of the efficiency condition in an empirically testable condition has to cope with the fact that not all variables in eq. (18) are observable. In particular, the efficient reduction in education, \widehat{E} , is not observable, neither is the difference between the social and the private costs of foregone leisure, $w_1^s - \omega_1^s$, whenever earnings functions are convex. The idea to separate observable from non-observable variables suggests the following notation. Let

$$PB \equiv \omega_2G'EL_2/\rho$$

be the *private benefit* of education. Subtracting the direct cost yields the *net private benefit*

$$NPB \equiv PB - \varphi E.$$

Similarly, we define the *net government benefit*

$$NGB \equiv \left(\frac{w_2}{r} - \frac{\omega_2}{\rho}\right)G'EL_2 - (f - \varphi)E.$$

And the ratio at which net returns to education are shared between the government and the individual is the *net benefit sharing ratio*

$$NBR \equiv \frac{NGB}{NPB}.$$

Optimal behavior requires NPB to be equal to the indirect cost of education which, in our simple model, is determined by the private cost of foregone nonqualified leisure, $\omega_1^s E$. However, translating the theory in an empirical framework requires a broader understanding of indirect costs of

education. Observed values of NPB are so high that it is implausible to assume that they only reflect cost of leisure. And indeed, Heckman et al. (2006) and others point to high risks of schooling decisions. There is the risk of failure, either because higher education cannot be successfully completed, does not result in employment, or individuals switch disciplines which also involves costs. It is useful to think of those costs of risk and re-optimization as being indirect and convex, just as the cost of leisure is indirect and convex. Hence we suggest interpreting the costs modelled in the theoretical section just as an example of any indirect convex costs, which determine the propensity to invest in education. Let

$$ICR \equiv \frac{w_1^s - \omega_1^s}{\omega_1^s}$$

be the *indirect cost ratio* which is the ratio at which indirect costs of education are shared between the government and the individual. Using all these definitions, eq. (18) can finally be rewritten as

$$NBR = ICR + \rho\gamma_\gamma \frac{PB}{NPB} \hat{E} . \quad (19)$$

Our objective is to check whether and to what degree OECD countries pursue efficient education policies. For this purpose three notions of efficiency have to be kept apart. Unconstrained efficiency requires $0 = \hat{E} = NBR = ICR$. In a world with taxation, unconstrained efficiency cannot hold and hence is entirely irrelevant for policy analysis. However, even in real world settings, education policy may well be efficient in the partial-analytical sense characterized by $NBR = ICR$. Finally, second-best policy requires the difference between the net benefit sharing ratio and the indirect cost ratio to be negative. This follows from eq. (19) and the analysis of the preceding sections. By Proposition 2, the relative change in education, \hat{E} , is negative in second-best. And by Proposition 1, we have reason to assume an increasing elasticity of the earnings function, $\gamma_\gamma > 0$.

6. Second-best tertiary education policy: An empirical application

The empirical research on earnings determination has a positive-theoretical focus. It aims to estimate the effect of a policy intervention on the marginal internal rate of return to schooling. For a recent discussion of the challenges in estimating the treatment effect see Carneiro et al. (2011). The

present paper follows a different research strategy. It is normative-theoretic in nature and tries to assess the relative efficiency of education policy. Such an undertaking is, no doubt, ambitious. Hence, the following analysis can only serve to open the discussion. We do this by studying the education/tax policies of OECD countries.

The data used is a panel of country level data. This has advantages, but also drawbacks. The advantage of country level data, as opposed to less aggregated data for individual countries, is more variation in the policy variables between countries as compared to variation within a country over time. The disadvantage of using aggregate data instead of individual level data is clearly the loss of variation in the national labor markets, as the heterogeneity of individuals with and without tertiary education is averaged out in aggregate data.

The data used is from various OECD publications and comprises six years between 2005 and 2010.⁴ Table A.1 in the Appendix describes the variables and the data sources, while Table A.2 summarizes the data. The focus is on *NBR*, the ratio at which net returns to education are shared between the government and the individual. The data for computing *NBR* are taken from the OECD data on net public and net private benefits from tertiary education for men.⁵ On average *NBR* is 0.55 but there is quite some variation, from a low of 0.08 up to 1.21. As a proxy for the indirect cost ratio *ICR* we take a marginal tax wedge associated with a suitably chosen marginal tax rate τ . This is suggested by the definition $ICR = (w_1^s - \omega_1^s)/\omega_1^s = \tau/(1 - \tau)$, with τ being the shadow tax rate on foregone leisure. Among the potential candidates we use the marginal tax wedge for an average income, single worker with no children denoted by MTW_{100} . The marginal tax wedge is computed from the net personal marginal tax rate reported by the OECD. Employer contributions to social security and net transfers are accounted for. Since our model focusses on individual education decisions, i.e., young individuals, we work with the marginal tax rates for single individuals with no children. Note that MTW_{100} varies between a low of 0.27 in Korea in 2010 and a high of 2.28 in 2005 in Hungary. The high value for Hungary results from a high marginal tax rate of 69.5 percent. MTW_{100} is a reasonable proxy for *ICR*, because workers who finished secondary schooling are clearly neither

⁴ The information needed to construct *NBR* is not available for earlier years.

⁵ The variables net public and net private benefits from tertiary education for men are not available for 2005 and 2006. They had to be computed from gross benefits and costs.

low nor high income workers. As an alternative measure we use $MTW_{67,100}$ (average of MTW_{67} and MTW_{100}).

Clearly, the marginal tax wedge is only an imperfect proxy of all indirect costs of tertiary education, like the risk of an investment in education. Hence, we include additional controls in the analysis. For instance, to account for the risk of unemployment and the supply of workers with completed tertiary education, we include the unemployment rate of individuals with tertiary education as well as the percentage of workers with tertiary education in the labor force. Again, there is substantial variation. The average unemployment rate for workers with tertiary education is 4 percent. However, it varies between a low of 1 and a maximum of 12 percent. We also find high variation regarding the percentage of individuals with tertiary education, which is between 12 and 51 percent (average is 30 percent). The ratio of private benefit and net private benefit is 1.05 and can be as high as 1.28. This points to differences regarding the private direct cost of getting tertiary education within the group of OECD countries. To control for the relative income position of the highly educated, the relative earnings of individuals with less than tertiary education is added. The average earnings premium for tertiary education is, at 53 percent, substantial and again, there are differences between the countries. To have a proxy of the (economic) ability of private households to invest in education, we also include the private savings rate.

Besides variables to assess indirect costs of tertiary education, we also account for the general inefficiency of public policy and the tax and transfer system. To control for political preferences for redistribution and taxation, which is typically associated with efficiency costs, we use the percentage of seats in parliament for leftist parties as well as the Gini-coefficient for disposable income. Moreover, the percentage of social expenditure from GDP describes how a country values and implements redistribution and is included as a proxy for the inefficiency of the tax system. GDP growth rates and year dummies serve as a general measure of economic development. And finally, we exclude outliers from the following analyses. The criterion used is Cook's D.⁶

⁶ When using MTW_{100} as our proxy for ICR , we drop BEL in 2006 and 2007, CZE in 2005 DNK in 2006, HUN in 2005 and 2006, and ITA in 2008. If $MTW_{67,100}$ is used, the excluded observations are CZE in 2005, DNK in 2006, HUN in 2005 and 2006, ITA in 2008, and SVN in 2009.

Second-best policy requires eq. (19) to hold where, clearly, all three terms appearing in eq. (19) are determined simultaneously. The left panel of Figure 4a shows the scatter plot of NBR and the proxy MTW_{100} for ICR , and the right panel uses $MTW_{67,100}$ instead. The first thing to note is the positive correlation between NBR and our alternative proxies MTW_{100} and $MTW_{67,100}$. Low values of NBR are found in Korea, whereas Belgium and Germany have high values indicating that the government strongly benefits from higher education. While it might be tempting to interpret this as evidence for demanding more public support for tertiary education, our model points to the relationship between NBR and ICR instead. Perhaps not surprisingly, the picture for NBR is mirrored in the graphs relating NBR to MTW_{100} and $MTW_{67,100}$, the marginal tax wedge of those without tertiary education. The tax wedge is low in Korea and high in Germany and Belgium. Thus, policy conclusions based on NBR only are in fact misleading, as they only partially account for variation in tax policy relevant for investment in higher education.

Recall that second-best policy requires countries to subsidize education. As argued in the preceding section, this means that the residual term in eq. (19), $\rho\gamma_V \frac{PB}{NPB} \hat{E}$, is negative. Thus we expect the observations to be below the 45° line. Observations above the line indicate inefficiency. According to Figure 4a, examples are Ireland and Australia, while the vast majority of the observation is below the 45° line. However, being below the 45° line is only a first indicator for efficiency. Figure 4a also shows the linear regression lines, with slopes being significantly less than one in both panels.

To learn more about the relative position of the countries over the entire sample period, we plot the country averages instead of observations for each year in Figure 4b.⁷ Germany, for instance, is on the regression curve, thus showing an average relationship between NBR and MTW . To get closer to the 45° line, Germany could either increase NBR and/or decrease MTW . Acknowledging the mobility of high skilled and the immobility of low skilled labor in an open economy, the policy advice would rather be to decrease MTW , that is, to lower the net marginal tax rate on the less educated. Note that this argument is not based on equity considerations, but results from an efficiency argument.

⁷ The regressions are based on all observations in the panel and not on the country averages. For ease of exposition we plot country averages only.

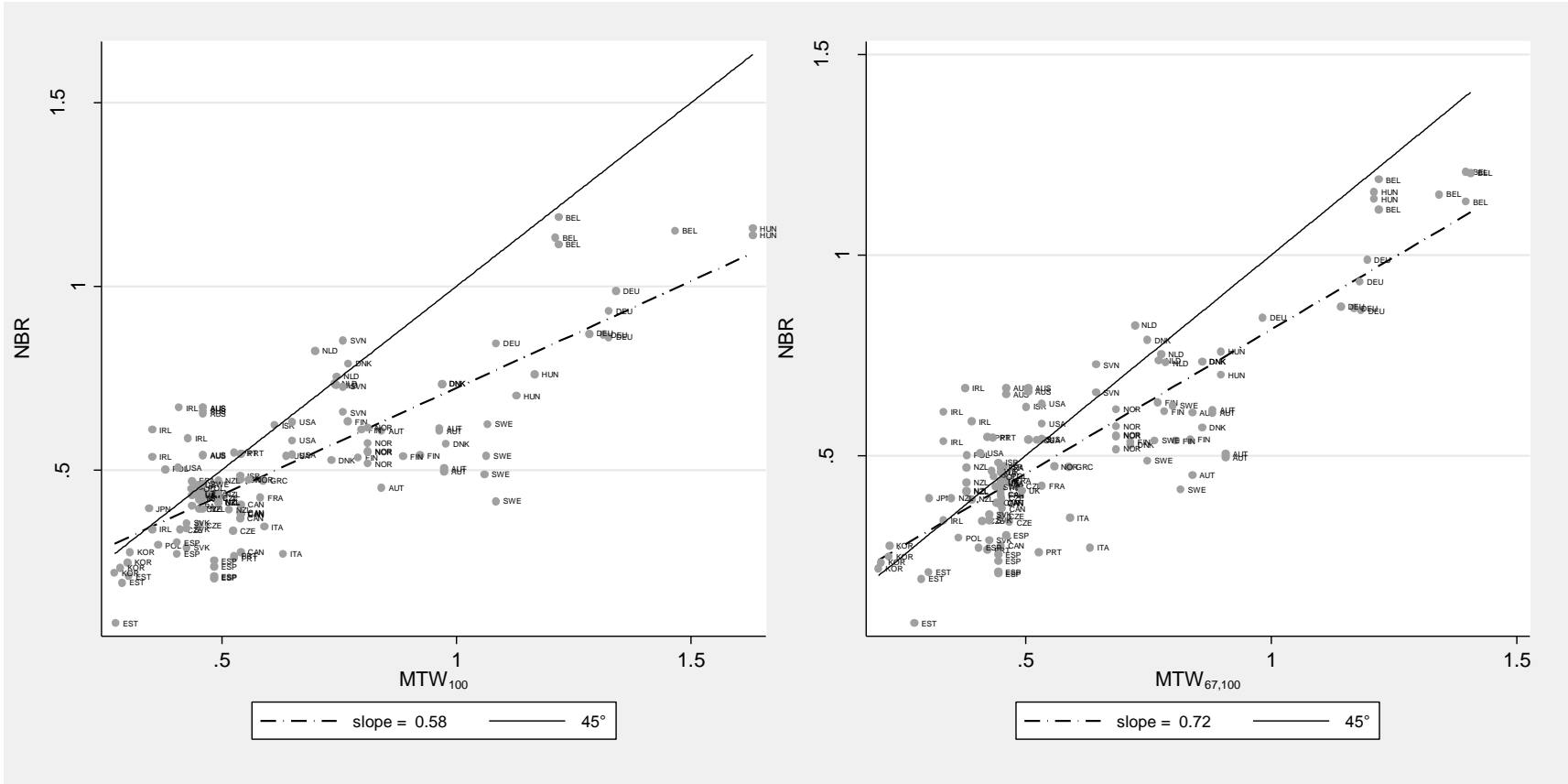


Figure 4a: Correlation between NBR and MTW

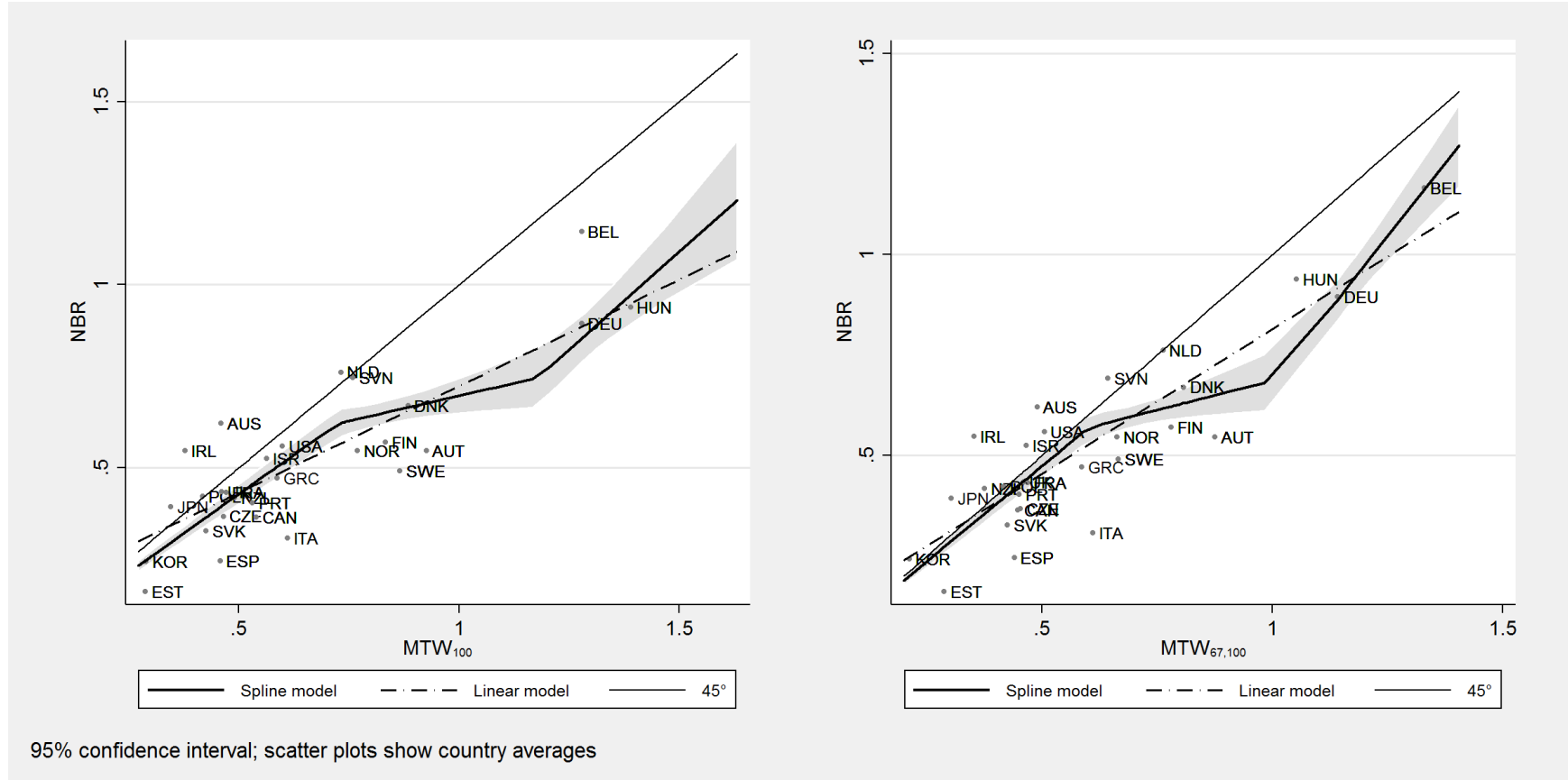


Figure 4b: Correlation between NBR and MTW, spline models

According to eq. (19), the strength of efficient subsidization of education increases in the efficient reduction of education, $-\hat{E}$. Hence it is plausible to assume that the locus of efficiency is a curve passing through the origin and bending away from the 45° line for large values of MTW . For small values of MTW one may even expect the deviations from the 45° line to be insignificant, as \hat{E} vanishes with lump-sum taxation. A tax system which is broadly based and which relies on a small tax rate τ is, however, close to a lump-sum tax. If, in addition, the tax base is consumption exempting the direct cost of education, $\varphi = (1 - \tau)f$, we obtain $NBR = \tau/(1 - \tau) = MTW$. Hence deviations from the 45° line should be statistically insignificant.

To check the empirical evidence on this, we add a spline model with 3 knots (cf. Figure 4b). It turns out, that for small values of MTW , the regression line is in fact very close to the 45° line in both panels of Figure 4b (the confidence interval of the first spline interval encloses the 45° line), and the slope of the first segment is statistically not different from one. This supports our theoretical reasoning.

Figures 4a and 4b rely on the assumption that the marginal tax wedge on average non-skilled labor can be used to proxy the indirect cost ratio. This choice of proxy is suggested by the theoretical model equating indirect costs of education with costs of foregone leisure. As noted, an empirical test of the theoretical analysis requires some broader interpretation of indirect costs. We mentioned the risk of failure to which schooling decisions are exposed. Hence in a next step we control for the additional indirect costs of tertiary education and the inefficiency of the tax and transfer system by adding the control variables described in Tables A.1 and A.2 and year dummies. Controlling for those variables we expect an added variable regression line closer to the 45° line compared to the earlier analysis. Moreover, the difference in the slope coefficient between countries with low values of MTW and higher values of MTW should become smaller. In Figure 4c we find in fact a stronger relationship between $MTW_{67,100}$ and NBR , but the slope of the regression line is still significantly less than one. Moreover, the spline model shows more similar slope coefficients than the spline model without controls. The linear regression coefficient varies between 0.56 and 0.79, depending on the chosen proxy for ICR and is significantly less than one. Note that in the plots without controls, very few countries are above the 45° line. More countries are in this group as we control for the general inefficiency of the tax and transfer system. Above the 45° line and outside of the

confidence interval of the spline model are for instance Australia, the Netherlands, Poland, Slovenia and Ireland. Hence those countries subsidize education less than the average of the OECD and, in fact, tax tertiary education.

One missing feature of the analysis is the lack of a benchmark. Since we use proxy variables for the theoretical variables of the model, we can compare the policy of the countries only with respect to the 45° line and with respect to the average of the OECD countries. Hence, we redo the analysis and differentiate between countries with a successful and less successful educational system. We expect successful countries to be closer to the 45° line compared to those with a less successful system. Clearly, it is not trivial to assess the quality of the educational system. However, it is common in the economics of education literature to use the results of the PISA study and other large scale assessments as a proxy for the quality of education. In fact, Hanushek and Wößmann (2010, 2015) have argued that the performance on large scale assessments is in fact a good predictor for economic growth. Since the optimality condition (19) describes welfare maximizing tax systems, we expect well performing countries to be on average closer to the 45° line than the low performing countries. Figure 5 shows the added variables plots for countries with above and below average performance on the PISA math test. It turns out that the high performing countries are much closer to the 45° line in the added-variable plots. Thus high performing countries do not only outperform the others with respect to academic achievement, they also implement a more efficient education tax system. In fact, in the right panel of Figure 5, the slope of the regression for countries with high PISA scores is statistically not different from one. Put differently, countries that are successful in achieving their educational goals are also closer to the optimal policy rule compared to the low performing countries.

Thus, and this is the conclusion from our empirical exercise, when evaluating education policy, direct and indirect costs and benefits of education as well as information about the tax and transfer system ought to be considered. However, so far, no testable condition for efficient policy has been derived. Based on a utility maximizing framework, we derive a theoretical optimality condition and suggest how to relate this condition to data. Clearly, this is only a first step, but we suggest a tentative analysis to describe the relative efficiency of OECD countries.

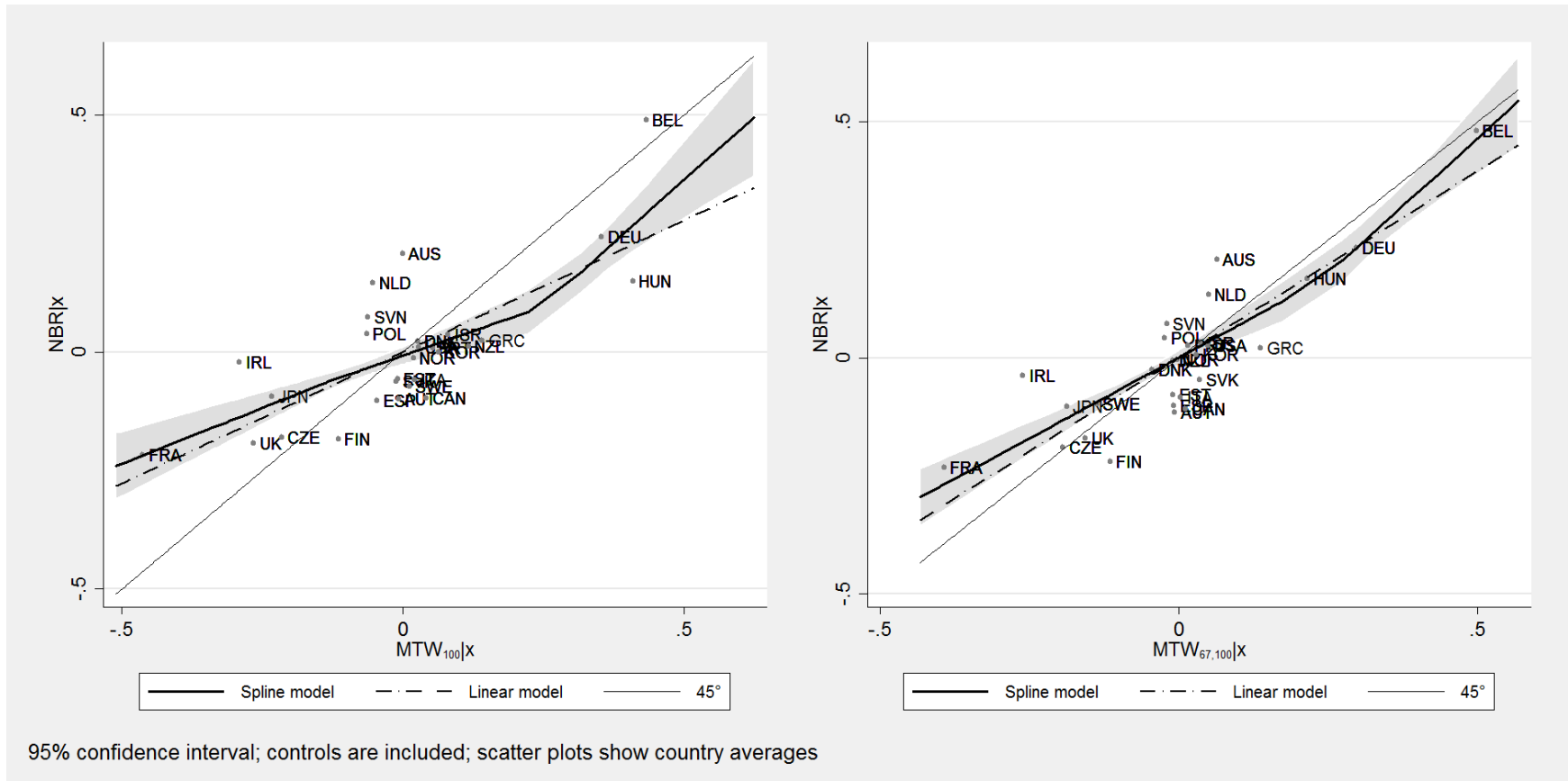


Figure 4c: Correlation between NBR and MTW, controls are included

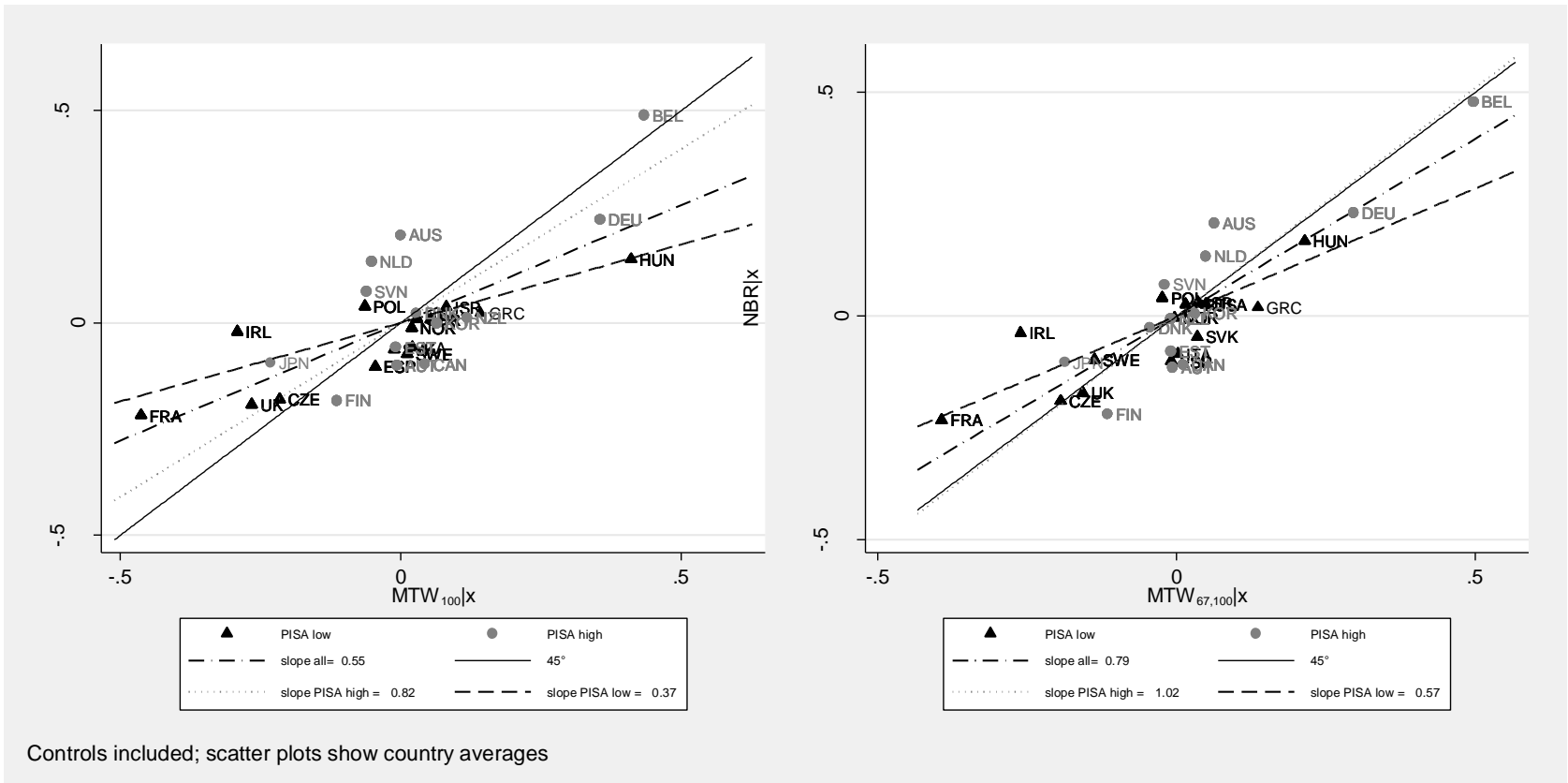


Figure 5: Correlation between NBR and MTW in countries with high and low PISA math scores

7. Conclusions

The presented analysis contributes to the literature methodologically, theoretically, and empirically. The methodological contribution refers to the modelling of schooling choice. The Mincer-Becker assumption of income maximization is replaced by utility maximization as is common in the optimal tax literature. The appeal of utility maximization is that it provides the basis for doing efficiency analysis. The two approaches, income maximization and utility maximization, are shown to have non-equivalent behavioral implications for non-concave earnings functions. When earnings are convex in education and individuals maximize utility, the cost of foregone leisure exceeds the cost of foregone earnings. Hence, the marginal internal rate of return is necessarily overestimated if the estimation uses data of foregone earnings. This is one possible reason, why the marginal internal rate of return to schooling is regularly estimated to exceed the cost of funds.

A critical implication of replacing income maximization by utility maximization is that it is no longer clear which properties of the earnings function should be assumed to hold on *a priori* grounds. However, by relying on learning theory this paper argues that there is good reason to assume earnings functions to display increasing elasticity in the amount of education, which is a rather weak assumption. Functions of increasing elasticity can well be concave or convex and the standard case of a Mincerian earnings function is the special case of an earnings function with a proportionally increasing elasticity. A disadvantage of increasing elasticity is that results from optimal tax theory do not automatically apply, since the optimal tax literature assumes concave earnings functions. In the present paper, however, we show that a result derived in the Ramsey literature for concave earnings functions *cum grano salis* extends to the case of convex earnings functions. The increasing elasticity of the earnings function turns out to be the pivotal assumption for proving that it is second-best to subsidize education effectively. Subsidizing education is optimal because it alleviates the social cost from taxing qualified labor. In other words, a double margin requires effective subsidization of education. Effective subsidization can be implemented in different ways, all of which are equivalent in the partial analytical sense. It might for instance result from subsidizing the cost of tuition. Alternatively, the tax on qualified labor income could be reduced relative to the tax on non-qualified labor.

In the last part of the paper, we confront theory with empirical evidence for OECD countries. This exercise suffers from the fact that key variables determining the choice of education are not directly observable. In particular, indirect costs of education are not observable. We solve this problem by using marginal tax wedges as proxies for *ICR*, the ratio at which indirect costs of education are shared between the government and the individual. Although the analysis is admittedly tentative, the results are promising. It is shown that the vast majority of OECD countries subsidize education as is suggested by the theoretical model. There is even some evidence that the strength of subsidization increases when the government reaps a larger share of the benefits of education.

Our analysis does not allow us to identify the efficiency frontier of education policy. Still, the analysis is informative. For instance, if a country can be shown to tax education effectively, this clearly raises policy questions. Examples are Ireland and Australia.

In this paper, the recommendation of effectively subsidizing education is derived from imperfections in taxation. This is in contrast to traditionally discussed justifications in the literature, which are based on arguments of market failure. The empirical evidence of externalities and liquidity constraints is, however, mixed (Heckman et al., 1998; Lange et al., 2006; Carneiro et al., 2002). Even if the evidence is considered to be supportive of some subsidization, it can at most rationalize subsidization to the extent that the marginal social costs and benefits of education are equated. The argument presented in the present paper, however, goes beyond arguments of possible market failure. We show that in a second-best world, the marginal social cost of education exceeds the marginal social benefit when labor is taxed and the elasticity of the earnings function is increasing.

Unlike the often heard arguments in the public debate for subsidizing tertiary education for equity reasons, this paper focuses purely on efficiency. Equity considerations can certainly justify subsidies to education. An important paper analyzing the close connection between equity and the subsidization of expenses for education in Mirrlees' tradition is Bovenberg and Jacobs (2005). The point made by the present paper goes beyond this and argues that labor taxation provides a strong efficiency reason for effectively subsidizing education in Ramsey's tradition.

8. Appendix

Proposition 2 has been proven by Richter (2009) for concave earnings functions. *A priori* it is not clear whether the proof extends to convex functions as convexity implies equality of L_1 and E and lacking disposability of ω_1 as a policy instrument. In what follows, it is, however, shown that the proof for concave earnings functions extends to the convex case. The proof relies on taking partial derivatives of the Lagrange function $T - \lambda e$ with respect to φ , ω_2 , and ρ :

$$\frac{\partial}{\partial \varphi} [T - \lambda e] = 0 \Leftrightarrow \left(\lambda - \frac{1}{\rho} \right) e_\varphi = \Delta e_\varphi. \quad (23)$$

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

$$e_\varphi = \rho E \quad \text{and} \quad \Delta e_\varphi = \Delta(\rho E) = \rho \Delta E + \frac{\rho^{-r}}{r} E. \quad (24)$$

Plugging (24) into (23) yields $\lambda - 1/r = \Delta E/E = \hat{E}$. Taking partial derivatives of the Lagrange function with respect to ω_2 , and ρ allows one to derive the equations

$$\lambda - \frac{1}{r} = \frac{\Delta(GL_2)}{GL_2} = \frac{\Delta C_1}{C_1}.$$

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

$$\rho C_{1x} + C_{2x} + \rho \varphi E_x = \omega_2 (GL_2)_x \quad \text{for } x = \varphi, \omega_2, \text{ and } \rho. \quad (25)$$

The relationship (25) extends to the Δ -notation:

$$\rho \Delta C_1 + \Delta C_2 + \rho \varphi \Delta E = \omega_2 \Delta(GL_2) \quad (26)$$

Eq. (14) is now easily proved by relying on (26), (12), and (6):

$$\begin{aligned} \Delta C_2 &= \omega_2 GL_2 \frac{\Delta(GL_2)}{GL_2} - \rho C_1 \frac{\Delta C_1}{C_1} - \rho \varphi E \frac{\Delta E}{E} \\ &= [\omega_2 GL_2 - \rho C_1 - \rho \varphi E] \frac{\Delta E}{E} = C_2 \frac{\Delta E}{E}. \quad \square \end{aligned}$$

Table A.1. Data description and sources

		Source:
Net benefit sharing ratio	$\frac{\text{Public benefits} - \text{public direct costs}}{\text{private benefits} - \text{private direct costs}}$ (men)	OECD: Education at a Glance 2009-2013
MTW_{100}	Marginal tax wedge for a single person at 100% of average earnings, no children	OECD iLibrary
$MTW_{67,100}$	Average marginal tax wedge for a single person at 67 and 100% of average earnings, no children	OECD iLibrary
Unemployed academics	Unemployment rates among men with tertiary education	OECD: Education at a Glance 2009-2013
Tertiary education	Share of the population with tertiary education, aged 25-64 years	OECD: Education at a Glance 2007-2011
Relative earnings (tertiary)	Relative earnings of 25-64 year-olds with income from employment (tertiary education)	OECD: Education at a Glance 2009-2013
Private savings rate	Net private saving rate in percent of GDP	National Accounts at a Glance 2014
Private benefit	Total private benefit	OECD: Education at a Glance 2009-2014
Net private benefit	Total private benefit - direct private cost	OECD: Education at a Glance 2009-2014
Social democrats (share of seats in parliament)	Share of seats for the party classified as a social party	The Comparative Political Data Set III 1990-2011 (University of Bern)
Social expenditure	Total public social expenditure/ GDP	OECD 2012 Social issues: Key tables from OECD
Gini (disposable income)	Gini-coefficient of disposable income	OECD iLibrary
Real GDP growth rate		OECD Economic Outlook

Table A.2 The data

	N	Mean	Standard dev.	Min	Max
<i>NBR</i>	130	0.55	0.25	0.08	1.21
MTW_{100}	130	0.70	0.36	0.27	2.28
$MTW_{67,100}$	130	0.63	0.29	0.20	1.43
Unemployed academics	130	0.04	0.02	0.01	0.12
Tertiary education	130	0.30	0.10	0.12	0.51
Relative earnings (tertiary)	130	1.53	0.23	1.18	2.13
Private savings rate	130	0.07	0.06	-0.10	0.28
Private benefit / net private benefit	130	1.05	0.05	1.01	1.28
Social democrats (share of seats in parliament)	130	0.30	0.16	0.00	0.63
Social expenditure	130	0.22	0.05	0.08	0.32
Gini (disposable income)	130	0.29	0.04	0.23	0.37
Real GDP growth rate	130	0.02	0.04	-0.14	0.10

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.
- Becker, G.S., 1964, *Human capital: A theoretical and empirical analysis, with special reference to education*, New York, National Bureau of Economic Research.
- Blackburn, J.M., 1936, *The acquisition of skill: An analysis of learning curves*, IHRB Rep. No. 73, London.
- Bovenberg, A.L., and B. Jacobs, 2005, Redistribution and learning subsidies are Siamese twins, *Journal of Public Economics* 89, 2005–2035.
- 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. and J.J. Heckman, 2002, The Evidence on credit constraints in post-secondary schooling, *Economic Journal*, 112, 705–734.
- Carneiro, P., J.J. Heckman, and E.J. Vytlačil, 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.
- Findeisen, S. and D. Sachs, 2014, *Designing efficient college and tax policies*, Universities of Mannheim and Cologne, mimeo.
- Hanushek, E.A. and L. Wößmann, 2010, Education and Economic Growth, P. Peterson, E. Baker, and B. McGaw, eds., *International Encyclopedia of Education*. Vol. 2, 245-252, Oxford: Elsevier.
- Hanushek, E.A. and L. Wößmann, 2015, *The knowledge capital of nations, Education and the economics of growth*, The MIT Press.
- Heckman, J.J. and P.J. Klenow, 1998, Human capital policy, in: M. Boskin, ed., *Policies to promote human capital formation*, Hoover Institution.
- Heckman, J.J., Lochner, L.J., and P.E. Todd, 2006, Earnings functions, rates of return and treatment effects: The Mincer equation and beyond, in: E.A. Hanushek and F. Welch, eds., *Handbook of the economics of education*, Elsevier, Vol. 1, Chap. 7, 307-458.
- Heckman, J.J., Lochner, L.J., and P.E. Todd, 2008, Earnings functions and rates of return, *Journal of Human Capital* 2, 1–31.

- Lange, F. and R. Topel, 2006, The social value of education and human capital, in: E.A. Hanushek and F. Welch, eds., *Handbook of the economics of education*, Elsevier, Vol. 1, Chap. 8, 459–509.
- Mincer, J., 1958, Investment in human capital and personal income distribution, *Journal of Political Economy* 66, 281–302.
- Mirrlees, J.A., 1971, An exploration in the theory of optimum income taxation, *Review of Economic Studies* 38, 175–208.
- 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.
- OECD, 2013, *Education at a Glance*.
- Oreopoulos, P. and K.G. Salvanes, 2011, Priceless: The nonpecuniary benefits of schooling, *Journal of Economic Perspectives*, Winter, 159-184.
- Ramsey, F.P., 1927, A contribution to the theory of taxation, *Economic Journal* 37, 47-61.
- Richter, W.F., 2009, Taxing education in Ramsey’s tradition, *Journal of Public Economics* 93, 1254–1260.
- 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.
- Roy, A., 1951, Some thoughts on the distribution of earnings, *Oxford Economic Papers* 3, 135–146.
- Willis, R.J. and S. Rosen, 1979, Education and Self-Selection, *Journal of Political Economy* 87, S7–S36.