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

By

Wolfram F. Richter
TU Dortmund University, CESifo Munich, and IZA Bonn

July 2011
This version: December 2011

Abstract: The Mincer equation postulates log earnings to be a linear function of schooling. The linearity assumption is however increasingly challenged on empirical grounds. This paper offers an alternative to the Mincer framework. The approach allows one to reconcile increasing returns to education with the neoclassical notion of diminishing returns to learning. The approach also allows one to perform efficiency analysis in Ramsey’s tradition. Distortive wage taxation is shown to provide strong reason for subsidizing education in effective terms. Second-best policy is confronted with empirical evidence on OECD countries.

JEL Classification: (J24, H21, I28)

Keywords: endogenous choice of education; power law of learning; Mincer equation; earnings determination; second-best taxation in Ramsey’s tradition; education elasticity rule

Address:
Wolfram F. Richter, TU Dortmund University, Department of Economics, 44221 Dortmund, Germany. Wolfram.Richter@tu-dortmund.de, phone: +49 231 755 3146

+ Valuable comments and suggestions made by Kerstin Schneider and Holger Sieg are gratefully acknowledged.
1. Introduction

More than fifty years have passed since Jacob Mincer (1958) showed a way of modelling schooling choice. The approach is still the starting point for empirical research on earnings determination. A recent example is Carneiro, Heckman, and Vytlacil (2011). The Mincer equation postulates log earnings to be a linear function of schooling. For an impressive visualization of log linearity see, e.g., Figure 2 in Card (1999). The figure plots mean log hourly wages against mean years of completed education using US schooling data for a sample of men aged 40–55 in the 1994–1996 Current Population Survey. The explanatory power of the linearity assumption is, however, no longer undisputed. Heckman, Lochner, and Todd (2008) provide challenging evidence. The authors employ a general nonparametric approach to estimate marginal internal rates of return, taking into account tuition costs, income taxes, and nonlinearities in the earnings–schooling–experience relationship. They conclude that the marginal rates of return to schooling obtained by their approach are anything but constant across schooling levels and that the rates also differ substantially from those returns obtained by a standard Mincer estimation.

The Mincer model is not only challenged on empirical grounds. Even more critical is the lack of a truly convincing theoretical foundation. The Mincer equation is neither derived from an explicit model of learning nor well anchored in the neoclassical standard model of labour–leisure choice. The method by which Mincer (1958) rationalizes log linearity in schooling relies on applying techniques of capital budgeting to the costs and benefits resulting from schooling choice for earnings. If life never ended, the growth rate of earnings associated with an additional year of school would not only be constant in the amount of schooling but also equal to the discount rate. This has at least two implausible implications. The first is that the increase in earnings is unrelated to any productivity differential between skilled and unskilled labour. Instead, the growth rate of earnings is determined by the discount rate, which normally is assumed to reflect the cost of funds and the return to physical capital. The second implausible implication is that the finiteness of life is the only reason why an individual should stop attending school. Schooling is modelled as if it did not suffer from diminishing returns. In fact, the Mincerian earnings function is strictly convex in schooling.

Diminishing returns is a phenomenon which is strongly suggested by learning curves. When an individual is observed learning to execute some task by repeating it again and again, her productivity shows clear diminishing returns. There is even impressive evidence that the productivity of learning displays some constant elasticity less than one. In neuroscience this is
called the power law of learning (Newell and Rosenbloom, 1981; Anderson, 2005). Supportive data have been collected by Blackburn (1936) and others. See Section 2 below. One could expect that the question of how to connect the phenomenon of strictly concave learning curves with the phenomenon of a strictly convex earnings curve à la Mincer (1958) is a big topic in the literature. This is however not the case. There must be some implicit consensus that the two phenomena are unconnected.

A first objective of the present paper is to show a way of connecting the Mincer equation with the power law of learning. More precisely, it is shown how the Mincerian earnings function can be reconciled with the assumption of diminishing returns governing the neoclassical standard model of household behaviour. Reconciliation is enabled by replacing the choice of schooling with a two-dimensional one. One dimension is choosing the time of learning (“education”), and the other dimension is choosing a particular subject (“discipline”) out of a continuous menu of competing disciplines. The earnings function is the function of education which results when optimally adjusting the choice of discipline to the increasing amount of education. More technically speaking, the earnings function is the upper envelope of learning functions. If the power law of learning is assumed to hold for each learning function, the resulting earnings function displays increasing elasticity. The Mincerian earnings function is obtained as the special case when the elasticity of the earnings function is not only increasing but linear by assumption. See Section 3 below.

A second objective of the present paper is to characterize efficient education policy in Ramsey’s tradition and to confront theory with practice. The choice-theoretic foundation of the earnings function provided in Sections 3 and 4 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) extends to the empirically more appealing convex case. The rule excels by particular simplicity in the case of a Mincerian earnings function. See Section 7. Secondly, the paper makes a clear case for effectively subsidizing education in second best. Any effective taxation would conflict with the increasing elasticity of earnings functions which according to the reasoning of Section 3 can be considered being a robust empirical finding. Even more, distortive wage
taxation is shown to be the reason why education should be subsidized in effective terms relative to the first best. This clear-cut policy conclusion is confronted with the empirical evidence on OECD countries in Section 8. It is shown that most countries strongly deviate from the optimal rule. Many countries, including the United States, effectively tax rather than subsidize tertiary education. Before Section 10 summarizes, Section 9 highlights conceptual differences between the Mincer model and the one developed in this paper. The striking suggestion derived from this paper’s model is that the empirical literature on earnings determination with its strong focus on the growth rate of earnings is focussing on a variable the importance of which for education policy may be strongly debated. Neither is the individual choice of education convincingly captured by the simple comparison of the growth rate with the individual discount rate. Nor is the growth rate of earnings determining efficient education policy. 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

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

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

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 governs not only the learning of simple tasks but also the acquisition of more complex skills. The assumption 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 learned, while $E$ measures education in units of time.

---

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.
3. Household behaviour and the earnings function

The household model adapted from Richter (2009) assumes 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 \( L_i \) in periods \( i=1,2 \). The function \( U = U(C_1, C_2, L_1, L_2) \) is strictly quasi-concave. \( L_2 \) is identical with the second-period labour supply. By contrast, only \( L_1 - E \) is the time spent in the market, while \( E \) is the time spent on education. First-period labour supply earns a constant wage rate \( \omega_1 \); the return to second-period labour depends on the amount of education. It is paid \( \omega_2 H(D, E) \), where \( \omega_2 \) is constant. The wage rate \( \omega_2 \) is modelled separately from \( H(D, E) \), as the former but not the latter is varied by taxation. The quantity \( L_2 \) is interpreted as qualified labour. Likewise, the quantities \( L_1 - E \) and \( L_1 \) are interpreted as non qualified labour and non qualified non-leisure, respectively. Education causes an opportunity cost in foregone earnings and a monetary cost of tuition. Both costs are assumed to be linear in time and constant across different disciplines \( D \). The cost of foregone earnings is modelled by \( \omega_1 E \) and the cost of tuition is modelled by \( \varphi E \). The share of first-period income that is spent neither on education nor on consumption is saved:

\[
S = \omega_1 (L_1 - E) - \varphi E - C_1 = \omega_1 L_1 - (\varphi + \omega_1)E - C_1.
\]  

(2)

By way of normalization, the price of consumption is set equal to one. The gross rate of return to saving is denoted by \( \rho \). Second-period consumption is constrained by income earned:

\[
C_2 = \rho S + \omega_2 H(D, E)L_2.
\]  

(3)

Substituting for \( S \) in (1) and (2) yields the lifetime budget constraint:

\[
C_1 + C_2 / \rho = \omega_1 L_1 + \omega_2 H(D, E)L_2 / \rho - (\varphi + \omega_1)E.
\]  

(4)

Maximizing utility \( U(C_1, C_2, L_1, L_2) \) in \( C_1, C_2, L_1, L_2, D, E \geq 0 \) subject to (4) and \( L_1 \geq E \) requires maximizing

\[
\omega_2 H(D, E)L_2 / \rho - (\varphi + \omega_1)E
\]  

(5)
in $D \geq 0$ and $0 \leq E \leq L_1$.

Consider the choice of the discipline. Keeping the amount of education $E$ fixed and taking the derivative with respect to $D$ implies

$$\omega_2 H_D L_2 / \rho = 0 \iff H_D(D, E) = 0. \quad (6)$$

The condition of second order requires $H_{DD} < 0$. Eq. (6) is used to determine the choice of discipline, $D(E)$, as an implicit function of education, $E$.

The function $G(E) \equiv H(D(E), E)$ is called the earnings function. Because of the envelope theorem, the elasticity of $G$ equals the partial elasticity of $H$ with respect to $E$,

$$\frac{E}{G} \frac{dG}{dE} = E \frac{\partial H}{H \partial E} = \eta(D(E)). \quad (7)$$

The second equality in eq. (7) follows from the power law of learning (1). Note that constancy of the learning elasticity $\eta(D)$ in $E$ for each $D$ implies an increasing earnings elasticity $\eta(D(E))$ in $E$:

$$\frac{d}{dE} \eta(D(E)) = \eta \frac{dD}{dE} (6) = - \frac{d}{dD} \left( \frac{E}{H} \right) \frac{H_{DE}}{H_{DD}} = - \frac{EH_{ED}^2}{HH_{DD}} > 0. \quad (8)$$

The geometric intuition is the following. Consider a menu of linear and 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.

**Proposition 1**: The power law of learning implies that the elasticity of the earnings function $G(E) = H(D(E), E)$ is increasing in education.

In what follows, the elasticity of $G$ is assumed 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:
\[ \eta(D(E)) \equiv mE \Leftrightarrow \ln G = g_0 + mE. \] 

An earnings function of type (9) is called Mincerian. The 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 can well be strictly concave.

### 4. The reduced model of household optimization

In reduced form, the representative taxpayer maximizes

\[ U(C_1, C_2, L_1, L_2) \text{ in } C_1, C_2, L_1, L_2, E \geq 0 \]  
subject to \( L_1 \geq E \) and

\[ C_1 + C_2 / \rho = \omega_1 L_1 + \omega_2 G(E) L_2 / \rho - (\varphi + \omega_1)E. \]  

(4)

Three scenarios are of interest and need separate discussion. In the first one, the taxpayer finds it optimal to stay nonqualified, \( E=0 \). This happens when the incentive to engage in education is too weak. In the second scenario, maximizing the ability rent income,

\[ Y(L_1, L_2) = Y(L_1, L_2; \omega_2 / \rho (\varphi + \omega_1)) \equiv \max_{0 < E \leq L_1} \left[ \omega_2 G(E) L_2 / \rho - (\varphi + \omega_1)E \right], \]  

(11)

has an interior solution \( E \in (0, L_1) \). Clearly, such a scenario requires the earnings function to be strictly concave. Concavity shall not be assumed, however. The following analysis therefore differentiates between the two following scenarios. The interior solution assumes \( 0 < E < L_1 \) and \( G'(E) < 0 \), while the upper corner solution assumes \( E = L_1 \) and \( G'(E) \geq 0 \).

Maximizing the ability rent income generates increasing returns. The interior solution yields increasing returns to qualified labour,

\[ \frac{d^2 Y}{dL_2^2} = \frac{\omega_2}{\rho} G' \frac{dE}{dL_2} = -\frac{\omega_2}{\rho} G^2 L_2 > 0, \]  

(12)
and the upper corner solution yields increasing returns to nonqualified labour,
\[ \frac{d^2 Y}{dL_i^2} = \frac{\omega \rho}{\rho} G'' L_i \geq 0. \]

The convexity has implications for the taxpayer’s optimization. Just assuming quasi-concavity of the utility function \( U(C_1, C_2, L_1, L_2) \) is not sufficient to guarantee the second-order conditions of the optimization (10), (4′) to be satisfied. Instead, the disutility of non-leisure has to be sufficiently convex in order to obtain a well-behaved optimization. In other words, the supply of non-leisure has to be sufficiently inelastic by assumption. In what follows, this is assumed to hold.

5. The tax planner’s problem

The government faces the need to raise an exogenous amount of revenue \( T > 0 \). Four linear tax instruments are available, each of which is distorting. The taxes are levied on period \( i \)'s labour income, on the return to saving, and on the cost of tuition. They are modelled implicitly as the difference between prices before and after taxes. The prices after taxes and subsidies are endogenous and denoted by \( \omega_1, \omega_2, \rho, \phi \). The prices before taxes and subsidies are exogenous and denoted by \( w_i, w_2, r, f \).\(^2\) The tax on period \( i \)'s labour income is modelled by \( w_i - \omega_i \), the tax on capital income by \( r - \rho \), and the tax on the cost of tuition by \( \phi - f \). 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 has to balance the budget:
\[ (w_i - \omega_i)(L_i - E) + (\phi - f)E + [(w_2 - \omega_2)G(E)L_2 + (r - \rho)S]/r \]
\[ \equiv (w_i - \omega_i)(L_i - E) + (\phi - f)E + \left[ \frac{w_2}{r} - \frac{\omega_2}{\rho} \right] GL_2 + \left[ \frac{1}{\rho} - \frac{1}{r} \right] C_2 = T. \quad (\gamma) \quad (12) \]

The bracketed variable \( \gamma \) is a Lagrange multiplier associated with the tax planner’s problem to be set up. The planner’s objective is to maximize \( U(C_1, C_2, L_1, L_2) \) in \( C_1, C_2, L_1, L_2 \geq 0 \), and

\(^2\) It is a straightforward exercise to endogenize the prices before taxes and subsidies. However, endogenizing 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.
\( \omega_1, \omega_2, \rho, \varphi \) subject to \( L \geq E > 0 \), the government’s budget constraint (12), and the first-order conditions of the taxpayer’s maximization.

Assume that the planner’s maximization is well behaved, and assume also that nonqualified labour has to be taxed in the optimum. The sole objective of this paper is to characterize efficient policy for education in relation to the tax on nonqualified labour. Efficiency will be characterized in terms of wedges. Denote by

\[
\Delta_{e_i} = \frac{w_i - \omega_i}{\omega_i} > 0 \quad \text{the positive wedge on nonqualified labour} \quad (13)
\]

and by

\[
\Delta_{e} = \frac{w_2 / r}{\omega_2 / \rho} - \frac{f + w_1}{\varphi + \omega_1} \quad \text{the wedge on education.} \quad (14)
\]

While the definition in (13) is standard, the one in (14) is not. According to the definition, the wedge on education equals the difference between two ratios. The first ratio relates present returns before and after taxes and subsidies, and the second ratio relates costs before and after taxes and subsidies. The wedge is positive if, and only if, the effective private marginal cost of education exceeds the effective social one:

\[
\Delta_{e} > 0 \iff \frac{\rho(\varphi + \omega_1)}{\omega_2} \leq \frac{r(f + w_1)}{w_2}. \quad (15)
\]

Efficient education policy is characterized by a vanishing wedge, \( \Delta_{e} = 0 \). It is however the objective of this paper to argue that \( \Delta_{e} \) is negative in second best. This means that education should be effectively subsidized. A negative value of \( \Delta_{e} \) can result from the use of each of the four policy instruments. For instance, effective subsidization is clearly obtained by the statutory subsidization of the cost of tuition. This is however not the only way of causing \( \Delta_{e} \) to be negative. Other means are (i) increasing the tax on nonqualified labour and thus reducing the cost of foregone earnings, (ii) reducing the tax on qualified labour and thus increasing the return to education, and finally (iii) taxing the return to saving and thus increasing the return to education.
6. Second-best education policy for the interior solution

As has been shown by Richter (2009), it is second best to subsidize education effectively if the following two assumptions are made. The elasticity of the earnings function must be increasing and the choice of education must have an interior solution. The result requires no particular functional specification of utility. The utility function can be arbitrarily chosen except for conditions of regularity. In other words, one only has to assume that the taxpayer’s and the planner’s optimizations are mathematically well behaved.

As it turns out, the second-best rate of effective subsidization is proportional to the second-order elasticity of the earnings function:

\[ \eta_\eta \equiv \frac{E d\eta(D(E))}{\eta(D(E)) dE} \]  \hspace{1cm} (16)

More precisely, the following Ramsey-like formula is obtained in second best (Richter, 2009, eq. (11)):

\[ \Delta_e / \rho = \eta_\eta \cdot \hat{E} , \]  \hspace{1cm} (17)

where \( \hat{E} < 0 \) is the efficient percentage reduction in education. (Throughout, a hat indicates a proportional change.) Because of Proposition 1 it is empirically justified to assume a positive value for \( \eta_\eta \). Hence \( \Delta_e \) is negative and effective subsidization of education is efficient. Effective taxation is not. An intuitive and nontechnical explanation of (17) is the following. When solving the tax planner’s problem one has to determine a combination of policy instruments so that the induced proportional reductions in the quantities chosen by the taxpayer satisfy two sets of constraints. One set reflects the standard Ramsey problem of minimizing the social cost of distorted demands and supplies. Its solution requires equal proportional reductions in all those quantities entering the taxpayer’s lifetime budget constraint: \( \hat{E} = \hat{L}_1 = \hat{C}_1 = \hat{C}_2 = \hat{N} \) with \( N \equiv GL_2 \) denoting the effective second-period labour supply. The other set is just one constraint. This constraint is derived from respecting the first-order condition associated with the taxpayer’s optimal interior choice of education, \( \hat{M} = G' L_2 = \rho (\phi + \omega_1) / \omega_2 \equiv c \). More precisely, the planner’s solution must be such that the proportional reduction in the marginal return to education, \( \hat{M} \), equals the proportional reduction in the effective marginal cost, \( \hat{c} \). From \( \hat{N} = \hat{E} \) one obtains \( \hat{L}_2 = (1-\eta)\hat{E} \) which
then implies $\dot{M} = \eta \hat{E}$. Eq. (17) results from verifying that $\hat{c}$ equals $\Delta E/\rho$. In sum, eq. (17) determines the efficient proportional change in the effective marginal cost of education that is needed to align the taxpayer’s maximization of ability rent income with the Ramsey rule of equi-proportional reductions in demands and supplies. If the elasticity of the earnings function increases, efficiency requires the marginal return to education to decrease which then requires subsidizing the effective marginal cost.

The generality of the formula (17) is striking. However, applicability is limited as long as one does not know which percentage reduction $\hat{E}$ is efficient and which value of $\eta$ should be assumed. One can only say more when working with specific utility and earnings functions. This is why the focus is on particular functional specifications in what follows. The utility function assumed is quasi-linear in first-period consumption and additive in periodic sub-utilities:

$$\begin{align*}
U(C_1, C_2, L_1, L_2) &= C_1 - V_1(L_1) + U(C_2, L_2).
\end{align*}$$

(18)

The function $V_1$ is strictly increasing and strictly convex. Maximizing (18) subject to the taxpayer’s budget constraint (4′) and assuming an interior solution yields the following first-order conditions:

$$\begin{align*}
U_1(C_2, L_2) + \omega_2 G(E) / \rho &= 0 \quad (\lambda) \quad (19) \\
\rho &= 1/ U_1(C_2, L_2) \quad (20) \\
\omega_2 G'(E)L_2 / \rho &= \varphi + V_1'(L_1) \quad (\mu) \quad (21) \\
\omega_1 &= V_1'(L_1) \quad (22)
\end{align*}$$

Solving the tax planner’s problem for the utility specification (18) yields

$$\begin{align*}
\frac{\Delta E}{\Delta \varphi} &= -\frac{\eta}{v_1} \quad (\text{education elasticity rule}), \quad (23)
\end{align*}$$

where

$$v_1 \equiv L_1 V_1' / V_1 > 0 \quad (24)$$
is the elasticity of marginal disutility of nonqualified labour. Equation (23) suggests that the effective rate of subsidizing education should increase in (i) the second-order elasticity of the earnings function, (ii) the tax wedge on non-qualified labour, and (iii) the (compensated) wage elasticity of the non-qualified labour supply, $1/v_1$. It is worth noting that the rule holds even if the planner does not optimize with respect to $\rho$. Saving does not need to be taxed efficiently, and yet second-best education policy should respect eq. (23). The proof is straightforward (Richter, 2011). Just set up the Lagrange function resulting from maximizing (18) subject to the taxpayer’s budget constraint (4′), the government’s budget constraint (12), and the FOCs (19)–(22). Assuming an interior solution and taking partial derivatives with respect to $C_1, L_1, E, \omega_1, \omega_2$, and $\varphi$ allows one to derive eq. (23).

7. Second-best education policy for the upper corner solution

It is not obvious and has not been shown before that the elasticity rule for education equally holds for the upper corner solution. In order to prove this, consider the problem of maximizing the utility function (18) in $C_1, L_1 = E \geq 0$ and $\omega_2, \varphi$ subject to the government’s budget constraint (12), the taxpayer’s budget constraint (4′), and the first-order constraints (19)–(21). Note that this planner maximization does not require an optimal choice of $C_2, L_2, \omega_1, \rho$. Still, the same elasticity rule (23) can be derived. More precisely, maximizing the utility function

$$
\omega_2 G(L_1)L_2 / \rho - \varphi L_1 - C_2 / \rho - V_1(L_1) + U(C_2, L_2)
$$

in $L_1 \geq 0$ and $\omega_2, \varphi$, subject to the government’s budget constraint

$$
(\varphi - f)L_1 + \left[ \frac{w_2}{r} - \frac{\omega_2}{\rho} \right] G L_2 + \left[ \frac{1}{\rho} - \frac{1}{r} \right] C_2 = T
$$

and subject to the FOCs (19) and (21), after having substituted $L_1$ for $E$, yields the following first-order conditions:

$$
\frac{\partial}{\partial \varphi} : \mu = (\gamma - 1)L_1
$$
\[
\frac{\partial}{\partial \omega_2} : \lambda = (\gamma - 1)(1 - \eta)L_2
\]  
(26)

\[
\frac{\partial}{\partial L_1} : (\gamma - 1)[1 - \eta + \frac{L_1 G''(L_1)}{G'(L_1)}]\omega_2 G'(L_1) L_2 / \rho -(\gamma - 1)L_1 \gamma' \]
\[
\frac{\partial}{\partial L_1} = (25), (26)
\gamma[f - w_2 G'(L_1) L_2 / r] + \gamma[\omega_2 G'(L_1) L_2 / \rho - \varphi] \]
\[
(\gamma - 1)\eta_\varphi (\varphi + \gamma') = \gamma[f - w_2 G'(L_1) L_2 / r] + \gamma \gamma' + (\gamma - 1)L_1 \gamma' \]
(27)

It remains to argue that the elasticity rule (23) follows from eq. (27). This is not obvious and therefore needs some explanation.

It is a specific feature of the upper corner solution that the planner’s optimization does not explicitly depend on \( w_1 \). The productivity of nonqualified labour does not enter the planner’s optimization. It is therefore a feasible strategy to define \( w_1 \) by setting

\[
\gamma w_1 \equiv \gamma V'_1 + (\gamma - 1)L_1 \gamma'
\]  
(28)

and to interpret \( w_1 \) as the shadow productivity of nonqualified labour. On setting \( \omega_1 \equiv \gamma' \), the equation \( \Delta_1 = \frac{\gamma - 1}{\gamma} \)

\[
- \Delta_e = \frac{f + w_1}{\varphi + \omega_1} - \frac{w_2 / r}{\omega_2 / \rho} = \frac{f + w_1 - w_2 G'(L_1) L_2 / r}{\varphi + \gamma'} \]
(21)

(27)

\[
\gamma - 1 \eta_\varphi
\]

is obtained from eqs. (14), (21), and (27). On dividing \( \Delta_e \) through by \( \Delta_1 \), the elasticity rule (23) is obtained. Another and more elaborate proof of eq. (23) is relegated to Appendix A. Proposition 2 summarizes the theoretical discussion.

**Proposition 2**: If utility is assumed to be quasi-linear and additive and if \( C_1, L_1, E, \omega_1, \omega_2, \) and \( \varphi \) are optimally chosen in the planner’s problem, the education elasticity rule characterizes efficient policy irrespective of whether the taxpayer’s choice of education results in an interior or an upper corner solution.
For a Mincerian earnings function eq. (23) 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, $\eta_\eta$, equals one.

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

$$\Delta_e = -\frac{\Delta_{\nu}}{\nu_1}. \quad (29)$$

A final remark concerns logical consistency. The education elasticity rule implicitly assumes the second-order condition of optimal educational choice to hold. For the interior solution this follows by definition. For the upper corner solution it is not guaranteed, however. The second-order condition then requires $\omega_2 G''(L_1) L_2 / \rho < V_1'(L_1)$. Assuming a Mincerian earnings function, this implies

$$mE = mL_1 \frac{L_1 G''}{G'} < \frac{L_1 V_1'(L_1)}{\varphi + V_1'(L_1)} = \frac{\omega_1}{\varphi + \omega_1} \nu_1. \quad (30)$$

The inequality (30) is a constraint which has to be checked for reasons of consistency when confronting eq. (29) with the empirical data on educational choice. Because $\ln[G(E)/G(0)] = mE$, the left-hand side of (30) can be interpreted as a skill premium. The ratio on the right-hand side, $\omega_1 / (\varphi + \omega_1)$, is the share of foregone earnings in the private 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.

**8. Second-best tertiary education policy: Practice**

It is inviting to confront the education elasticity rule (23) with the empirical evidence. An immediate implication of the rule and of Proposition 1 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 data to relate the tax wedge on nonqualified labour, $\Delta L$, with the wedge by which tertiary education is effectively subsidized, $-\Delta E$. The figure suggests that the positive correlation of $\Delta L$ and $-\Delta E$ with a coefficient of 0.52 is an empirical fact that is however strongly driven by some specific countries. These countries are Turkey and Poland, which apparently subsidize tertiary education heavily in effective terms. If these two countries are treated as outliers and removed from the data set, the correlation becomes less than one percent. The obvious reason for the vanishing correlation is that there are roughly as many countries subsidizing tertiary education as there are countries effectively taxing tertiary education. Effective taxation is clearly at variance with the education elasticity rule.

Figure 2: Measured and second-best tax and subsidy wedges $\Delta L$, and $-\Delta E$ in OECD countries.

Source: Own computations on the basis of OECD data (2009, Tables A8.2 and A8.4) collected for tertiary education and males. See Appendix.

---

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, and unemployment effects. Taxes on savings are ignored. Cash flows are discounted by a 5% interest rate. See Appendix B for details.
The dashed line in Figure 2 indicates second-best policy. I.e. combinations of wedges on this line satisfy the simplified version (29) 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 one quarter. Both assumptions are clearly debatable on empirical grounds. However, log linearity of earnings is an acceptable assumption for describing mean behaviour.\(^4\) And the value of one quarter is justified with reference to Bargain, Orsini, and Peichl (2011). These authors survey the literature and provide own estimates for seventeen European countries and the US. The estimates clearly vary across countries, family status, and estimation method. However, the value of one quarter seems to be a good approximation of average behaviour.

If the dashed line is accepted as a characterization of efficient education policy, then Norway comes closest to the optimum. By contrast, heavily overshooting incentives for tertiary education are displayed for Turkey. The measured value of \(\Delta_e\) is at 90 percent while 13 percent would be efficient. An extreme case of insufficient incentives is Belgium for which the measured value of \(\Delta_e\) is at minus 42 percent while plus 6 percent would be efficient.

The present analysis implies that effective taxation is inefficient. This raises the question of how to interpret the evidence. Measured inefficiency may well be the result of some poor quality of data. As no better data are available, the OECD data are however accepted as a valid picture of country-specific policies. Even then is the interpretation of Figure 2 anything but straightforward. The reason is that the wedge on education is the net result of a whole bunch of policies. Thus it is not true that the countries relying more than others on private means in tertiary education are just the countries effectively taxing education. A clear counterexample is Korea. The country is known for relying predominantly on private sources of funds. The OECD (2009, Table B2.4) reports a private expenditure ratio of 76 percent. Still, the effective rate of taxing education comes close to zero. The US is another country known for its high university fees, and yet the rate of effective taxation there is less than one might have expected. By contrast, those countries known for relying predominantly on public funds, like Belgium, France, and Germany, are among those countries heavily taxing tertiary education in effective terms. One would have to go into country details in order to explain

\(^4\) The conditions which must be fulfilled so that the mean marginal return to schooling can be estimated by a Mincerian earnings function are stated in Carneiro et al. (2011, p. 2754).
which particular policy combination brings about the measured wedges displayed in Figure 2. Each country seems to be a case of its own.  

This conclusion is stressed when taking a closer look at the set of countries effectively taxing education. For this purpose $-\Delta_e$ is written as the difference between a cost ratio and a benefit ratio, where the former is defined as the ratio of public total costs to private total costs of education, and the latter is defined as the ratio of public total benefits to private total benefits of education. (See Appendix.) Obviously, negativity of $-\Delta_e$ can result from a small cost ratio and/or a large benefit ratio. Small values of the cost ratio – i.e., values less than 0.5 – are reported for AUS (0.41), FRA (0.45), IRL (0.42), KOR (0.19), NZL (0.42), and USA (0.37). Large values of the benefit ratio – i.e., values larger than 1.0 – are reported for BEL (1.32), CZR (1.07), DEN (1.82), and GER (1.12). The only two countries effectively taxing education but not captured by these \textit{ad hoc} groupings of countries with small cost ratios and/or large benefit ratios are FIN and SWE. In all the other cases, effective taxation results either from of a small cost ratio (AUS, FRA, IRL, KOR, NZL, USA) or a high benefit ratio (BEL, CZR, DEN, GER). An extreme case is Denmark. It is the country with the largest benefit ratio (1.82). As it is equally the country with the largest cost ratio (1.79), effective taxation of education is still rather moderate. Denmark stands for a policy where the government roughly takes two-thirds of both the return to and the cost of tertiary education. The opposite extreme case is Korea. This is the country where the smallest benefit ratio (0.23) is matched by the smallest cost ratio (0.19). Hence Korea stands for a policy where the government roughly takes one-fifth of the return to education and one-sixth of the cost.

The discussion of the empirical evidence is to be completed by a check of consistency. Assuming $\nu_1=4$, the second-order condition (30) requires

$$mE < 4 \frac{\omega_l}{\phi + \omega_i} .$$  (31)

Recent estimates of wage premia on tertiary education are provided by Strauss and de la Maisonneuve (2009). The estimates cover all countries appearing in Figure 2 except CZR, KOR, NOR, and TUR. The premia estimated are gross, while the wage premia guiding educational choice are net of taxes. As tax progression tends to compress wage premia, one is

---

5 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.
however on the safe side when working with gross premia instead of net premia. High gross premia are reported for the US (88.1) and Portugal (81.9). The given numbers refer to men, and they measure the average changes of upper-secondary degree holders in percent. According to OECD data (2009, Table A8.2), \( \omega_t/(\varphi+\omega_t) \) equals 0.38 in case of the US and 0.81 in case of Portugal. As \( \ln(1.881) = 0.63 < 1.52 \), the US passes the consistency check. The other countries, including Portugal, are less critical cases. Hence consistency seems to be no problem; the inequality (30) can safely be assumed to hold.

9. The Mincer model in comparison

The Mincer equation relies on the assumption that schooling choice can be modelled as a one-dimensional decision. The one-dimensionality is the reason why it is appropriate to speak of schooling rather than of education. Even though schools do offer choices of different subjects, the very notion of schooling conveys the idea that the primary decision to be taken concerns the number of years of study needed to earn specific credentials. If the present study replaces the key variable schooling ("S") with the variable education ("E"), this is meant not least to stress the rejection of one-dimensionality. Education is still treated as measurable in units of time, but differences in observed quantities can no longer be interpreted as differences in pure learning time. Instead, they also reflect differences in training disciplines. Stressing these conceptual differences might suggest employing different notation when comparing the present model with the one of Mincer. That would however unduly overburden the comparison. Hence schooling is denoted by \( E \) – instead of \( S \) – and earnings are denoted by \( G(E) \) even when referring to the Mincer model in what follows.

There are two standard ways of rationalizing log linearity of the earnings function in the Mincer model. One is dynamic and the other is static. The dynamic version 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 \( E_2 \) is greater than \( E_1 \). Foregone earnings are the only cost of schooling. Assuming an infinite working life, arbitrage requires the growth rate of earnings, \( G'/G \), 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, \( \rho \). The equating of \( G'/G \) with \( \rho \) has some highly implausible implications already mentioned in the introduction. By implication, 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 not only implausible, but also empirically refuted. The empirical research on earnings determination regularly comes up with a significant mark-up of the marginal internal rate of return to schooling on the market rate of interest. An example is 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 maximize income when making schooling decisions. The present analysis suggests a different conclusion. What is critical is not the income-maximizing assumption as such but the overly simplistic conception of educational choice reflected by the standard Mincer model.

The static version of rationalizing log linearity has been developed by Card (1995). It amounts to maximizing $U(E) \equiv \ln G(E) - \rho E$. This competing approach has the advantage of conceptualizing the schooling choice as the distinct result of maximization. The obvious disadvantage is that it is hard to justify the suggested utility function. The functional specification raises the question of why earnings are weighted by a logarithm while costs are not. The objective of maximization is perceptibly motivated by mathematical reasoning. The equality $G'/G = \rho$ is to result as a first-order condition.

In the model of the present paper, log linearity of the earnings function is the possible though not necessary result when individuals optimize on learning disciplines and when assuming that learning curves are isoelastic as suggested by the power law of learning. In contrast with the Mincer model, log linearity is not governing the decision on the length of education as such. Instead, the length is determined by equating the marginal return to education, $G'L_2$, to the effective marginal cost, $\rho(\varphi + V'_i) / \omega_2 = c$. If the earnings function, $G$, fails to be concave, as suggested by the empirical evidence, then increasing marginal disutility of non-leisure stops the individual 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 (marginal) rates of return to education as stated by eq. (15). 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, \( c \), such that the taxpayer’s maximization of ability rent income is aligned with the Ramsey rule of equi-proportional reductions in demands and supplies. The alignment requires subsidizing the taxpayer’s maximization of ability rent income if the elasticity of the earnings function is increasing as suggested by empirical evidence.

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 debated. 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 the logarithmic function 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 within a given discipline.
10. Summary

The dominating role of the Mincer equation in estimations of the return to schooling contrasts with its unimpressive theoretical foundation. The standard derivation relies on assuming the net present value of schooling to be constant in the amount of schooling. Such an assumption is not theoretically appealing, as it seems to negate the standard notion of decreasing returns. A first objective of the present study is to provide a model of education which is more consistent with the neoclassical standard model of household behaviour and which allows one to dissolve the apparent antagonism between the strict convexity of the Mincerian earnings function and the law of diminishing returns. It is shown that the Mincerian earnings function can be derived from a model in which the individual has to choose not only the amount of education but also a particular discipline out of a continuous menu of competing disciplines where the elasticity of learning the discipline is constant. Such constancy of the learning 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 each discipline (Proposition 1). The Mincerian earnings function is obtained on assuming that the elasticity of the earnings function is not only increasing but linear in education.

The great advantage of the present model is that it allows one to analyze 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. Hence a second objective of the present paper is to show that the results of Richter (2009; 2011) extend to earnings functions that fail to be concave. For the policy conclusions derived it is less relevant, however, whether the earnings function is concave or convex. Pivotal is the question of whether the elasticity of the earnings function is increasing or decreasing. As argued in the present paper it is empirically justified to assume an increasing elasticity. The normative conclusion then is that the objective to alleviate the distortion of labour taxation provides reason for subsidizing education in effective terms.

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 reference 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 this paper the utility function is assumed to be quasi-linear and additive. The sole justification for this choice comes from mathematical tractability. Assuming quasi-linearity and additivity, 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 many countries – including Belgium, Australia, France, Germany, and the US – effectively tax tertiary education. There are other countries effectively subsidizing education, but they overshoot. Turkey, Poland, and Spain are examples of overshooting. Efficient rates of effective subsidization of tertiary education range from zero percent (France) to 18 percent (Poland). By contrast, measured rates range from −42 percent (Belgium) to 90 percent (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. One may equally question the underlying behavioural assumptions. The assumptions of a Mincerian earnings function and of a compensated elasticity of nonqualified labour supply of one-half are, however, have some backing in the literature. This is certainly not the case as far as quasi-linearity and additivity of the utility function are 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 general-equilibrium models and 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 analyzing 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.

11. Appendices

*Appendix A*: Assume that there are two types of taxpayers. One type is called qualified, is indexed by \( q \), and finds it optimal to spend all non-leisure in period 1 on education: \( L_{1q} = E_q \) \( >0 \). The other type is called nonqualified, is indexed by \( n \), and finds it optimal to remain nonqualified: \( E_n = 0 \). The different behaviour may result from different earnings functions, \( G_q \neq G_n \). Consider the following assumptions:

(i) The second-period wage income of the nonqualified individual can be taxed independently of first-period income and of second-period qualified wage income. In other words, the government can set \( \omega_{2n} \) independently from \( \omega_1 \) and \( \omega_{2q} \).

(ii) In the policy optimum, both types choose to spend the same amount of first-period non-leisure so that \( L_{1n} = L_{1q} = E_q \) holds.

(iii) The two types of individuals do not differ in the function which specifies the disutility of non-leisure in first period, \( V_n = V_q = V \).
(iv) The planner maximizes the non-weighted sum of utilities. Hence the focus is on pure efficiency.

It is rather straightforward to derive the education elasticity rule (23) if the assumptions (i) – (iv) are made. Just set

\[ W^n \equiv W^n(C_{2n}, L_{1n}, L_{2n}) \equiv \omega_1 L_{1q} + \omega_2 G_{1n}(0) L_{2n} - C_{2n} / \rho - V(L_{1n}) + U^n(C_{2n}, L_{2n}), \]

\[ W^q \equiv W^q(C_{2q}, E_q, L_{2q}) \equiv \omega_2 G_q(E_q) L_{2q} - \varphi E_q - C_{2q} / \rho - V(E_q) + U^q(C_{2q}, L_{2q}), \]

and maximize \( W^n + W^q \) in \( L_{1n}, E_q \), and \( \omega_1, \omega_2, \varphi \) subject to a joint tax revenue constraint and the first-order conditions associated with the choices of \( L_{1n}, E_q, \) and \( L_{2n}, L_{2q} \). This can be interpreted as saying that the education efficiency rule (23) extends to the upper corner solution if (iv) the sole focus is on efficiency and not on equity, if (i) the government is sufficiently flexible in the choice of tax instruments, if (iii) qualified and nonqualified taxpayers do not differ by the valuation of non-leisure time in their first life period, and if (ii) qualified and nonqualified taxpayers end up spending the same amount of non-leisure time in their first life period on education and labour, respectively. Such assumptions are admittedly restrictive. It however speaks for them that they allow a reasonable interpretation.

Appendix B: Figures 2 and 3 are derived from OECD Indicators, Education at a Glance 2009, Tables A8.2 and A8.4. The numbers entering Figures 2 and 3 are computed as follows (the short form “i!J” refers to column J in Table A8.i):

\[ \Delta_{1i} = \frac{w_i - \omega_i}{\omega_i} = \text{foregone taxes on earnings / private foregone earnings} = \frac{4!F}{2!F}, \]

\[-\Delta_2 = \frac{f + w_i - (\varphi + \omega_i)}{\varphi + \omega_i} - \frac{w_z / r - \omega_z / \rho}{\omega_z / \rho} = \frac{\text{public total costs}}{\text{private total costs}} - \frac{\text{public total benefits}}{\text{private total net benefits}} = \frac{4!(D + F)}{2!(D + F)} - \frac{4!(H + J + L + N)}{2!(H + J + L + N + P)}.\]
References


