ANNALS OF ECONOMICS AND FINANCE **6**, 209–228 (2005)

Optimal Tax and Education Policy When Agents Differ in Altruism and Productivity

Helmuth Cremer

University of Toulouse (IDEI and GREMAQ)

Pierre Pestieau

CREPP, University of Liège, CORE, and Delta

Emmanuel Thibault

University of Toulouse (GREMAQ) and University of Perpignan (GEREM)

and

Jean-Pierre Vidal

European Central Bank

This paper studies the design of education policies in a setting of overlapping generations with heterogeneous individuals. Individuals differ in productivity (high and low earning ability) and in altruism (altruists and non altruists). Only altruistic parents invest in education out of some joy of giving. Their investment determines the probability that a child has high ability. Education policies consist of a subsidy on private educational investments and of public education. We show that when an income tax is available, the subsidy on education should not depend on redistribution. Instead, it is determined by the following terms. First, a Pigouvian term which arises because under warm glow altruism parents' utility does not properly account for the impact of education on future generations. The second term captures a "merit good" effect, which arises when the warm glow term is not fully included in social welfare (possibility of laundering out). Third, depending on the information structure there may be a substitution term that arises because the demand for second period consumption and for education transfer are interdependent. The first two terms are of opposite sign and the optimal subsidy may be positive or negative. Finally, we derive conditions under which public education is desirable. Public education affects also the probability of being highly productive for the altruists and the non altruists. Its desirability will in part depend on its substitutability with private educational investment.

1529-7373/2005

Key Words: Tax; Education; Altruism.
JEL Classification Numbers: E6, E2, H2

1. INTRODUCTION

In almost all societies education is one of the most heated political issues. It is also one of the largest items of expenditure, both in the public sector (along with health care and social security) and in households' budgets. Underlying a number of debates there is the basic question of what should be the role of government and families in the production and financing of education. In this paper, we consider two broad questions. First, why not let families choose their optimal amount of investment in education? In other words, should the government subsidize the private provision of education and if so, how? We are naturally concerned by the traditional trade-off between equity and efficiency. The second question concerns the provision of public education.

The direct (public provision) and indirect (subsidies) intervention of the government in education is motivated by the simple fact that education is not a "pure private good". It is often claimed there are important externalities associated with having an educated society. Recently, the literature on endogenous growth has focused on the idea that education financed by altruistic parents or by the students themselves (through borrowing) has a social return superior to its private return. Consequently the level of education that individuals would privately choose to undertake were there no government subsidy would be insufficient.

It remains that this efficiency argument is much less agreed upon than the equity argument. The primary justification for public support of education indeed arises from distributional concerns particular at the elementary and secondary-school levels. The equity argument is particularly strengthened when taking into account the distortionary nature of traditional income tax policies. Even with optimal non linear income taxation we can show that some public provision is desirable without separability between leisure and other consumption goods. With separability, following Atkinson and Stiglitz (1976) propositions, public provision is not needed at least for redistributive purposes.

The complexity of models with education is well-known; we already mentioned the technological externality in endogenous growth models and the second-best argument for public provision.[1] We could also mention imperfect capital markets, peer groups effects, competition among schools,

[1] See Glomm and Ravikumar (1992) and De Fraja (2002).

screening education, school vouchers, etc.[2] In this paper we have chosen to focus on two aspects that we believe important. The first one is that even altruistic parents do not necessarily provide the right amount of education to their children because they are concerned by their mere joy of giving and not by the incidence of their choice on the society as a whole. The second one is that societies consist of individuals who differ not only in ability but also in altruism. Redistribution among workers with different ability has been studied by Mirrlees (1971). When workers further differ in altruism, redistribution policy is not as simple. First, there is the issue of how to deal with the altruistic component of the individual utility function when aggregating utilities. Second, when there is no altruism parents do not invest in education and thus educational subsidies are of no use. In that alternative, the case for public provision is overwhelming.

2. THE MODEL

We consider an overlapping generations model with individuals characterized by a level of productivity which can take only two values and a level of altruism which can also take only two values. Individuals live for two periods. They work only in the first but consume in both periods. In the second period they can also provide an investment in an education technology which affects their children's probability to be highly productive. We consider a small open economy for which both interest and wage rates are given; we further assume that the interest rate is consistent with the modified golden rule; see condition (7) below.[3]

2.1. The household's problem

All individuals have the following utility:

$$\begin{aligned} U_t^{ij} &= U\left(c_t^{ij}, d_{t+1}^{ij}, \ell_t^{ij}, x_{t+1}^{ij}\right), \\ &= u\left(c_t^{ij}\right) + \beta u\left(d_{t+1}^{ij}\right) - v\left(\ell_t^{ij}\right) + \varepsilon^j h\left(x_{t+1}^{ij}\right), \end{aligned} \tag{1}$$

where ij denotes an individual of productivity $i = 1, 2$ and altruism $j = a, n$; c_t^{ij} is first period consumption, d_{t+1}^{ij}, second period consumption, ℓ_t^{ij}, the labor supply and x_{t+1}^{ij}, the investment in education. Separability between labor supply and consumption is assumed to keep in line with Atkinson and Stiglitz' result. The functions u and h are increasing and

[2] See e.g., Loury (1981), Galor and Zeira (1993), Benabou (2002), Fernandez and Rogerson (1996).

[3] We have also developed a closed economy version (inspired by Cremer et al. (2003)) of this model with public debt. It is more complex than the open economy model and does not bring additional insights other than a justification for the modified golden rule.

concave while v is increasing and convex. The parameters β and ε^j represent the individual time preference factor and the degree of (joy of giving) altruism respectively. There are two levels of altruism: $\varepsilon^a = \varepsilon > 0$ for the altruists and $\varepsilon^n = 0$ for the non altruists. There are also two levels of productivity ψ^1 and ψ^2 with $\psi^2 > \psi^1$.

In a world without taxation, an individual of type ij working in period t maximizes (1) subject to the budget constraints:

$$\begin{aligned} w^i \ell_t^{ij} &= c_t^{ij} + s_t^{ij}, \\ (1+r)\, s_t^{ij} &= d_{t+1}^{ij} + (1+n) x_{t+1}^{ij}, \end{aligned}$$

where $w^i = w\psi^i$; w and r are the rates of wage and of interest, $1+n$ is the number of children (n is the rate of population growth) and s_t^{ij} represents savings. Under some regulatory conditions, we expect interior (and positive) solutions for c, ℓ, s, d as well as for x^{ia}. The non altruists on the other hand will obviously set $x^{in} = 0$.

We assume children have the same level of altruism as their parents (i.e., altruists have altruistic children and non altruists have non altruistic children).[4] Consequently, the distribution of altruists and non altruists is fixed with proportion λ^a and $\lambda^n = 1 - \lambda^a$. The relative number of high and low productivity individuals,

$$\pi^{1j} \text{ and } \pi^{2j} = 1 - \pi^{1j} \qquad (j = a, n)\,,$$

is common knowledge but endogenous.[5] At a given period of time t, the probability (proportion) π_t^{ij} results from investment in human capital.

For altruists, the investment may be private but it can also be public. For the non altruists, it is only public. The educational technology which specifies the probability that a child of individual ij is of high productivity can be written as:

$$H\left(x_t^{ij}, e_t\right),$$

with x being private investment and e public investment (per young individual). Given that the π's are probabilities we restrict $H(\cdot)$ to be also included in the interval $[0,1]$. The partial derivative of H are denoted $H_1 > 0$, $H_2 > 0$ and we assume that it is concave in x (i.e., $H_{11} < 0$). Further, we know that for non altruists, we have $H(0, e)$. Note that e is by

[4]Rather then assuming that parents and children have the same level of altruism we could have made the weaker assumption that level of altruism and productivity are independently distributed. This would complicated the writing of some expression but otherwise leave the analysis and results unchanged.

[5]Troughout the paper we assume that the number of individuals is sufficiently large so that actual proportions are equal to probabilities.

assumption identical for all children; in other words whether parents are highly productive or not, altruistic or not, does not make a difference. The government cannot discriminate families by type when providing public education. Finally we write:

$$\pi_t^{2j} = \pi_{t-1}^{1j} H\left(x_t^{1j}, e_t\right) + \pi_{t-1}^{2j} H\left(x_t^{2j}, e_t\right). \tag{2}$$

In words, the probability that a child of generation t is productive is a weighted sum of the productivity that parents of type 1 and of type 2 have a productive child; the weights being the probabilities (proportions) in generation $t-1$. This rule applies for the altruists $j = a$ and for the non altruist $j = n$.

2.2. First-best

As a reference we start by deriving the first-best optimality conditions. The social planner's objective is the discounted sum of utilities with a time preference factor $\gamma < 1$. To allow for alternative treatments of the altruistic utility term $\varepsilon h(x)$ we use a parameter ν with $0 \leqslant \nu \leqslant 1$. When $\nu = 0$, the social planner does not include the joy of giving in its welfare criterion. This is the position advanced by, e.g., Hammond (1987) who has advocated "excluding all external preferences, even benevolent ones for one social utility function". When $\nu = 1$ the social planner adopts a pure utilitarist position.

In the first-best we write the social planner's problem as the maximization of the following Lagrangian:

$$\begin{aligned}\mathcal{L} \;=\; \sum_t \gamma^t \Bigg\{ & \sum_{ij} \lambda^j \; \pi_t^{ij} \left[u\left(c_t^{ij}\right) + \beta \; u\left(d_{t+1}^{ij}\right) - v\left(\ell_t^{ij}\right) + \nu \; \varepsilon^j \; h\left(x_{t+1}^{ij}\right)\right] \\ & -\mu_t \sum_{ij} \lambda^j \left[\pi_t^{ij}\left(c_t^{ij} - w^i \ell_t^{ij}\right) + \pi_{t-1}^{ij}\left(\frac{d_t^{ij}}{1+n} + x_t^{ij}\right) + e_t\right] \\ & -\sum_{ij} \eta_t^j \left[\pi_{t+1}^{2j} - \pi_t^{ij} \; H\left(x_{t+1}^{ij}, e_{t+1}\right)\right] \Bigg\},\end{aligned}$$

where μ and η are the multipliers associated with the resource constraint

$$\sum_{ij} \lambda^j \pi_t^{ij} w^i \ell_t^{ij} = \sum_{ij} \lambda^j \left[\pi_t^{ij} c_t^{ij} + \pi_{t-1}^{ij}\left(\frac{d_t^{ij}}{1+n} + x_t^{ij}\right) + e_t\right], \tag{3}$$

and the human capital technology (2) respectively.

Differentiating $\mathcal{L}$ with respect to the first-best control variables c_t^{ij}, d_t^{ij}, ℓ_t^{ij}, x_t^{ia}, π_t^{2j} and $e_t \geqslant 0$ and evaluating in the steady-state yield the following optimality conditions:

$$
\begin{aligned}
c^{ij}: &\quad u'\left(c^{ij}\right) - \mu = 0, && (4a)\\
d^{ij}: &\quad \beta(1+n)u'\left(d^{ij}\right) - \mu\ \gamma = 0, && (4b)\\
\ell^{it}: &\quad -v'\left(\ell^{ij}\right) + \mu\ w^i = 0, && (4c)\\
x^{ia}: &\quad \varepsilon\nu h'\left(x^{ia}\right) - \mu\gamma + \frac{\eta^a}{\lambda^a} H_1\left(x^{ia}, e\right) = 0, && (4d)\\
\pi^{2j}: &\quad \lambda^j[\left(U^{2j} - U^{1j}\right) - \mu(c^{2j} - w^2\ \ell^{2j} + \frac{\gamma d^{2j}}{1+n} + \gamma x^{2j} - c^{1j} \\
&\quad + w^1\ \ell^{1j} - \frac{\gamma d^{1j}}{1+n} - \gamma x^{1j})] - \frac{\eta^j}{\gamma} + \eta^j[H(x^{2j}, e) - H(x^{1j}, e)] = 0, && (4e)\\
e: &\quad -\mu\gamma + \sum_{ij} \eta^j \pi^{ij}\ H_2\left(x^{ij}, e\right) \leqslant 0 \quad (= 0 \text{ if } e > 0). && (4e)
\end{aligned}
$$

Non altruists will never make any educational investment and thus $x^{in} = 0$. [6] For all other variables an interior solution is assumed. Rearranging one obtains:

$$
\begin{aligned}
c^{1a} &= c^{2a} = c^{1n} = c^{2n} = c^*, && (5)\\
d^{1a} &= d^{2a} = d^{1n} = d^{2n} = d^*. && (6)
\end{aligned}
$$

According to (4a) and (4b), $\gamma = \beta(1+n)$ is a necessary and sufficient condition to have $c^* = d^*$. Remark that, with the modified golden rule assumption

$$
1 + r = \frac{(1+n)}{\gamma}, \qquad (7)
$$

this is the case when the rate of time preference is equal to the rate of interest $(1/\beta = 1 + r)$; throughout the paper we shall assume for simplicity that this condition holds.

From (4a) and (4c):

$$
w^i\ u'\left(c^{ij}\right) = v'\left(\ell^{ij}\right),
$$

[6] Alternatively we could allow x^{in} to take positive values in a setting of first-best or of non linear taxation. Even though x^{in} does not bring any utility to parents of type in, it has a positive effect on human capital. By assuming $x^{in} = 0$, we consider that non altruistic parents are unable to accomplish such an action. This would have more sense with parental love than with financial transfers.

and thus:

$$\ell^{1a} = \ell^{1n} = \ell^{1*} < \ell^{2a} = \ell^{2n} = \ell^{2*}. \tag{8}$$

Condition (4d) implies that

$$x^{1a} = x^{2a} = x^{*}. \tag{9}$$

Given the above equalities (5)–(9), one can rewrite (4e) as:

$$\eta^{j}/\lambda^{j} = \gamma[v\left(\ell^{1*}\right) - v\left(\ell^{2*}\right) + v'(\ell^{2*})\ \ell^{2*} - v'(\ell^{1*})\ \ell^{1*}]. \tag{10}$$

Finally, combining (4d) and (4e), the condition for public education becomes:

$$-\varepsilon\nu\, h'\left(x^{*}\right) + \eta^{n}\, H_2\left(0, e\right) + \eta^{a}\, H_1\left(x^{*}, e\right)\left[\frac{H_2\left(x^{*}, e\right)}{H_1\left(x^{*}, e\right)} - \frac{1}{\lambda^{a}}\right] \leqslant 0\ (= 0 \text{ if } e > 0). \tag{11}$$

The intuition behind these conditions is easily understood. With a utilitarian objective and separable utility functions, consumptions are type-independent and the more able work more than the less able. Investment in education is independent of productivity. Condition (10) provides the expression for the multiplier associated with the probability of being productive. In the first-best we have

$$\eta^{a}/\lambda^{a} = \eta^{n}/\lambda^{n},$$

namely the multipliers divided by the relative size of the two types are equal.

Finally equation (11) characterizes the optimal level of public education. To be more precise it measures the variation in social welfare induced by an increase in e which is compensated for budget balance by a decrease in x^{*}. Observe that budget balancing requires $dx^{*} = -de/\lambda^{a}$ (because e is per person while x^{*} is per altruist). The first term is negative for $\nu > 0$. It reflects the fact that private investment in education besides its effect on the educational technology brings some joy of giving that is accounted for in social welfare. The second term measures the direct impact of the increase in public education on (the probability of having high productivity of) the children of the non altruist (π^{2n}). The third term measures the *net* effect of the considered variation via its impact on the children of the altruists. To see this observe that the term in brackets represents the difference between the "technical rate of substitution" (between private and public education) and the "marginal cost" of e (in terms of x). This term can be positive or negative. Consider the case when e and x are perfect substitutes so that $H\left(x, e\right) = G\left(x + e\right)$ and $H_1 = H_2$. Then, it is negative for $\lambda^{a} < 1$ while it vanishes for $\lambda^a = 1$. The term can be positive when the two education inputs are not perfect substitutes and when H_2/H_1 is sufficiently large.

Returning to the case of perfect substitutes, one can write (11) as:

$$-\varepsilon\nu\ h'(x^*) + \eta^n\ [G'(e) - G'(e + x^*)] \leqslant 0\ (= 0 \text{ if } e > 0). \tag{12}$$

This expression shows that when $\eta^n = 0$ (which is the case in particular when $\lambda^n = 0$) we always have $e = 0$. When all individuals are altruistic (as in Cremer and Pestieau (2005)) there is of course no role for public education in the case of perfect substitutes. On the other hand, when $\eta^n > 0$, e is surely positive for $\nu = 0$ or for $G'(0) = \infty$. This point illustrates the idea that public education is the only way to foster the probability to achieve a high productivity of the children of the non altruists who represent a fraction λ^n of society.

Equation (12) also points to the role played by the parameter ν that shows how the joy of giving is accounted for in social welfare. We know that when x and e are perfect substitutes, $\lambda^n > 0$ and $\nu = 0$, there is no room for x and e is most likely positive. This is in line with the intuition according to which one would expects e decreasing with ν. A closer examination of the comparative statics properties of e shows, however, that this conjecture has to be qualified. Even for the first-best, the relationship between ν and e appears to be complex as it depends on the third derivative of $G(e + x)$. Consequently there is no hope to find a monotonic relationship under simple and fairly general conditions.

2.3. Laissez-faire and decentralization

In a decentralized economy with an income tax function $T_{ij}(w^i\ell^{ij})$, a tax/subsidy τ_x^i on x and a level of public education e we can achieve the above first-best optimum. With these instruments, the budget constraints of an individual are:

$$\begin{aligned}
&w^i\ell_t^{ij} - T_{ij}(w^i\ell_t^{ij}) = c_t^{ij} + s_t^{ij},\\
&(1+r)s_t^{ij} = d_{t+1}^{ij} + (1+n)(1+\tau_x^i)x_{t+1}^{ij}.
\end{aligned}$$

The FOC are given by

$$-u'\left(c^{ij}\right) + \beta(1+r)u'\left(d^{ij}\right) = 0, \tag{13}$$

$$v'\left(\ell^{ij}\right) - w^i\left(1 - T'_{ij}\left(w^i\ell^{ij}\right)\right)u'\left(c^{ij}\right) = 0, \tag{14}$$

$$-\beta(1+n)(1+\tau_x^i)u'\left(d^{ia}\right) + \varepsilon h'\left(x^{ia}\right) = 0. \tag{15}$$

Equation (13) is the same as in the first-best. Equation (14) yields the first-best allocation if $T'_{ij}(\cdot) = 0$; in other words, we need lump-sum redistributive taxes across the four types so as to equate the consumption levels. Equation (15) along with (4d) yields:

$$1+\tau_x^i = \frac{\varepsilon h'\left(x^{ia}\right)}{\beta(1+n)u'\left(d^{ia}\right)} = \frac{\varepsilon h'\left(x^{ia}\right)}{\varepsilon\nu\ h'\left(x^{ia}\right)+\frac{\eta^a}{\lambda^a}H_1\left(x^{ia},e\right)},$$

or

$$\tau_x^* = \frac{\lambda^a\varepsilon\left(1-\nu\right)h'\left(x^*\right)-\eta^a H_1\left(x^*,e\right)}{\lambda^a\varepsilon\nu\ h'\left(x^*\right)+\eta^a H_1\left(x^*,e\right)} = \frac{\lambda^a\varepsilon\left(1-\nu\right)h'\left(x^*\right)-\eta^a H_1\left(x^*,e\right)}{\lambda^a\gamma\mu}. \tag{16}$$

Equation (16) shows that a tax or subsidy on x is required to decentralize the optimal allocation, even in a first-best setting. For $\nu=1$, τ_x^i is negative which means that a subsidy is needed to internalize the externality of private education on the probability to be productive. However, when $\nu<1$ and particularly when $\nu=0$, namely if the social planner does not include the joy of giving in its welfare measure, then a tax is not impossible, especially if the marginal social return of private education is low.

Finally, one needs the government to supply the amount of public education consistent with condition (11). This supply will depend on the educational technology $H\left(x,e\right)$ and on the relative number of altruists, λ^a.

3. SECOND-BEST OPTIMALITY

We now turn to the case where individual types are not publicly observable so that the lump-sum taxes required to decentralize the (utilitarian) first-best are no longer feasible. We distinguish two cases. In the first one, private education is observable at the individual level and can thus be controlled by the central planner through a non linear tax/subsidy function. In the second, private education is not observable at the individuals' level (only anonymous transactions are observable). Consequently, it can only be controlled through a linear tax or subsidy.

3.1. Private education is individually observable

In this problem the only variables not observed by the social planner are individuals' productivity and labor supply. Instead the social planner only observes their product: gross earning $I^{ij}=w^i\ \ell^{ij}$. The social planner problem consists in finding the values of c, d, x, I and e that maximize the following Lagrangian:

$$\begin{aligned}\mathcal{L} \;=\; & \sum_t \gamma^t \Bigg\{ \sum_{ij} \lambda^j \pi_t^{ij} \left[u\left(c_t^{ij}\right) + \beta u\left(d_{t+1}^{ij}\right) - v\left(\frac{I_t^{ij}}{w^i}\right) + \nu\varepsilon^j h\left(x_{t+1}^{ij}\right)\right] \\ & -\mu_t \sum_{ij} \lambda^j \left[\pi_t^{ij}\left(c_t^{ij} - I_t^{ij}\right) + \pi_{t-1}^{ij}\left(\frac{d_t^{ij}}{1+n} + x_t^{ij}\right) + e_t\right] \\ & -\sum_j \eta_t^j \left[\pi_{t+1}^{2j} - \sum_i \pi_t^{ij} H\left(x_{t+1}^{ij}, e_{t+1}\right)\right] \\ & +\sum_{ij} (-1)^i \varphi_t^j \left[u\left(c_t^{ij}\right) + \beta\, u\left(d_{t+1}^{ij}\right) - v\left(\frac{I_t^{ij}}{w^2}\right) + \varepsilon^j h\left(x_{t+1}^{ij}\right)\right]\Bigg\},\end{aligned}$$

where the φ^j's ($j = a, n$) are the multipliers associated with the self-selection constraints. For simplicity we consider only the self-selection constraints within a given category $j = a, n$. Put differently, we assume that altruists cannot mimic non-altruists an vice-versa.[7]

Deriving the FOC and evaluating them in the steady-state yields:

$$c^{ij}: \quad (\lambda^j \pi^{ij} + (-1)^i \varphi^j) u'\left(c^{ij}\right) - \mu\, \lambda^j \pi^{ij} = 0, \tag{17a}$$

$$d^{ij}: \quad (\lambda^j \pi^{ij} + (-1)^i \varphi^j)\, \beta u'\left(d^{ij}\right) - \frac{\gamma\mu\lambda^j\pi^{ij}}{1+n} = 0, \tag{17b}$$

$$I^{ij}: \quad \mu\lambda^j\pi^{ij} - \frac{\lambda^j\pi^{ij}}{w^i} v'\left(\frac{I^{ij}}{w^i}\right) - (-1)^i \frac{\varphi^j}{w^2} v'\left(\frac{I^{ij}}{w^2}\right) = 0, \tag{17c}$$

$$x^{ia}: \quad (\lambda^a \pi^{ia}\nu + (-1)^i \varphi^a)\varepsilon h'\left(x^{ia}\right) + \pi^{ia}(\eta^a H_1\left(x^{ia}, e\right) - \gamma\mu\lambda^a) = 0, \tag{17d}$$

$$e: \quad -\mu\gamma + \sum_{ij} \eta^j \pi^{ij}\; H_2\left(x^{ij}, e\right) \leqslant 0 \quad (= 0 \text{ if } e > 0). \tag{17e}$$

We do not present the FOC with respect to the π^{ij}'s as they are not needed. We assume that there exist steady-state values for the variables

[7] The fact that type n cannot mimic type a follows directly from the assumption that $x^{in} = 0$. Conversely, to ensure that altruists never want to mimic non altruists it is sufficient to assume that $h'(0)$ is sufficiently large.

c, d, I, x and e. Conditions (17a)–(17d) can be rearranged as follows:

$$v'\left(\frac{I^{2j}}{w^2}\right) = u'\left(c^{2j}\right) w^2,$$
$$\frac{v'\left(\frac{I^{1j}}{w^1}\right)}{u'\left(c^{1j}\right) w^1} - 1 = \left[\frac{v'\left(\frac{I^{1j}}{w^2}\right)}{u'\left(c^{1j}\right) w^2} - 1\right] \frac{\varphi^j}{\lambda^j \pi^{1j}},$$
$$u'\left(c^{ij}\right) = \frac{(1+n)\beta}{\gamma} u'\left(d^{ij}\right),$$
$$\frac{\varepsilon h'\left(x^{ia}\right)}{(1+n)\beta u'\left(d^{ia}\right)} = 1 - \frac{\eta^a H_1(x^{ia}, e)}{\lambda^a \gamma \mu} + \frac{(1-\nu)\,\varepsilon h'\left(x^{ia}\right)}{\gamma\mu}. \tag{18}$$

The first two expressions are standard conditions of optimal income taxation with two types: no distortions at the top for types $2j$; positive marginal tax for types $1j$. The third expression is consistent with Atkinson and Stiglitz's proposition: no taxation of second period consumption. Finally, the fourth expression gives a condition for taxing or subsidizing private education.

Using (15), one can rewrite (18) as follows:

$$\tau_x^i = \frac{(1-\nu)\,\varepsilon h'\left(x^{ia}\right)}{\gamma\mu} - \frac{\eta^a H_1\left(x^{ia}, e\right)}{\lambda^a \gamma \mu}. \tag{19}$$

As in the first-best, the tax/subsidy on education consists of two terms: the first term is positive (as long as $\nu < 1$) and the second is negative. The first term reflects the idea that if the government does not think that the joy of giving has a full social value then it finds desirable to tax private education. The second term is an externality or a Pigouvian term: private education has a positive effect on the probability of the next generation being highly productive and this effect is not internalized by the altruistic parents. Unlike in the first-best, we now have a subsidy which is type specific. This is because the second best levels of x differ between types (while they are equal at the first-best). One can assume reasonably that $x^{1a} < x^{2a}$ and thus $h'\left(x^{1a}\right) > h'\left(x^{2a}\right)$ and $H_1\left(x^{1a}, e\right) > H_1\left(x^{2a}, e\right)$; consequently for $\nu = 1$, $-\tau_x^1 > -\tau_x^2$. In words, when the joy of giving altruism is fully accounted for in welfare, the (marginal) subsidy is higher for lower skill individuals. Note that if we abstract for the type specificity, the second best subsidy rule (19) is the same as the first-best rule (16). This is in line with the Atkinson and Stiglitz property: because of the separability of preferences there is no incentive term in the second-best rule.

Turning to the level of public education, combining the FOC with respect to x and to e yields

$$
\begin{aligned}
& -\nu \sum_i \pi^{ia} \varepsilon h'\left(x^{ia}\right) + \eta^a \bar{H}_1^a \left[\frac{\overline{H}_2^a}{\overline{H}_1^a} - \frac{1}{\lambda^a}\right] \\
& + \eta^n \bar{H}_2^n - \frac{\varphi^a \varepsilon}{\lambda^a}\left[h'\left(x^{2a}\right) - h'\left(x^{1a}\right)\right] \leqslant 0 \; (= 0 \text{ if } e > 0).
\end{aligned} \tag{20}
$$

where

$$
\overline{H}_k^j = \sum_i \pi^{ij} H_k\left(x^{ij}, e\right). \tag{21}
$$

Like its first-best counterpart, (11), the second-best expression (20) measures the impact on social welfare due to an increase in e associated with a budget balancing decrease in x^{ia}'s. Its four terms can easily be interpreted and signed. The first three terms are the same as their counterparts in (11) except that levels of x now vary across individuals. Following the discussion in Subsection 2.2, it does appear that the first term measures the warm glow effect associated with *private* education spending; it is negative when $\nu > 0$. The second term can be positive as well as negative depending on the degree of substitutability between x and e. The third term is positive and measures the direct impact on the children of non altruists. The last term is specific to the second-best nature of the problem; it is positive as long as $x^{2a} > x^{1a}$ (which we assume). The presence of the Lagrange multiplier φ^a suggests that it is an “incentive term”. This may appear surprising for two reasons. Firstly because e does not appear in the incentive constraints and secondly because we are in a setting where the Atkinson-Stiglitz property holds. To understand this term one has to keep in mind that we are considering an increase compensated by a (uniform) decrease in x^{1a} and x^{2a}. The increase in e has not impact on the incentive constraint, but the decrease in x does have an effect. Specifically, when $x^{2a} > x^{1a}$ we have $h'\left(x^{2a}\right) - h'\left(x^{1a}\right) < 0$; consequently, a uniform increase in x^{ia} would violate the incentive constraint and conversely, a uniform decrease relaxes the incentive constraint.

With $\nu = 0$ and perfect substitutability, only the second term is negative and equal to

$$
\frac{\lambda^a - 1}{\lambda^a} \sum_i \pi^{ia} G'\left(x^{ia} + e\right).
$$

Like in the first-best solution, we then obtain that e is necessarily positive when $H'(0) = \infty$, as long as η^n. From that perspective the presence of non-altruist has the expected effect of fostering a positive level of public education. For the rest, the exact relationship between e and λ^n is quite complex. However, it is clear that one cannot expect a monotonic relationship. To see this we continue to assume perfect substitutes and $\nu = 0$. It then follows that, when $\lambda^a = 1$ the LHS of (20) is always positive. This means that as long all individuals have an interior solution for x, welfare is always increasing in e. Consequently the optimal solution requires total crowding our of private education expenditures for one of the types; specifically, one expects $x^{1a} = 0$. Interestingly, this is no longer true when there are non-altruists in the society.

3.2. Private education is not individually observable

We now turn to the case where individual private investment in education is not observable. Aggregate and anonymous transactions are however observable and can be subject to a linear tax (with the same uniform rate applying to all individuals). For consistency we make the same assumption for consumption (and specifically for second period consumption).[8] We have a tax on education at rate θ_t^x and a tax on second period consumption at rate θ_t^d.

Let $R_t^{ij} = I_t^{ij} - T(I_t^{ij})$ the disposable income of an individual ij. If we introduce an interest factor for education $p_t^x = (1+\theta_t^x)/(1+r)$ and another one for second period consumption $p_t^d = (1+\theta_t^d)/(1+r)$ we write[9]

$$c_t^{ij} = R_t^{ij} - p_{t+1}^d d_{t+1}^{ij} - (1+n)p_{t+1}^x x_{t+1}^{ij}.$$

Altruist and non-altruists determine their consumption pattern by solving

$$V_t^{ij} = \max_{x,d} u(R_t^{ij} - p_{t+1}^d\, d_{t+1}^{ij} - (1+n)p_{t+1}^x\, x_{t+1}^{ij}) + \beta u\left(d_{t+1}^{ij}\right) + \varepsilon^j h\left(x_{t+1}^{ij}\right).$$

This yields the demand functions

$$x_{t+1}^{ia} = x^{ia}\left(R_t^{ia}, p_{t+1}^d, p_{t+1}^x\right),\ d_{t+1}^{ia} = d^{ia}\left(R_t^{ia}, p_{t+1}^d, p_{t+1}^x\right),\ d_{t+1}^{in} = d^{in}\left(R_t^{in}, p_{t+1}^d\right),$$

and the indirect utility functions

$$V_t^{ia} = V^{ia}\left(R_t^{ia}, p_{t+1}^d, p_{t+1}^x\right) \quad \text{and} \quad V_t^{in} = V^{in}\left(R_t^{in}, p_{t+1}^d\right).$$

To state the government's problem, we have to specify the resource constraint. We show in the Appendix that in each period t the resource constraint coincides with the government budget constraint:

$$e_t = \sum_{ij} \lambda^j \left[\pi_t^{ij}(I_t^{ij} - R_t^{ij}) + \pi_{t-1}^{ij}\left[\theta_t^d \frac{d_t^{ij}}{1+n} + \theta_t^x x_t^{ij}\right]\right] \tag{22}$$

[8] If both c and d were observable, x would necessarily be observable also.

[9] We have set the tax on capital income to zero but this is just a matter of normalization. A uniform tax on x and d is effectively equivalent to a tax on interest income.

As it will appear below the self-selection constraints are not involved in the determination of either indirect tax which is the focus of this section. To simplify the presentation we thus assume the same constraints as in the case when education expenditures are observed. In other words, we exclude the possibility that non altruists could mimic altruists and vice versa.

With these assumptions, the Lagrangian for the government's problem is now given by:

$$\begin{aligned}\mathcal{L} \;=\; & \sum_t \gamma^t \Bigg\{ \sum_{ij} \lambda^j \pi_t^{ij} \left[V_t^{ij} - v(\frac{I_t^{ij}}{w^i}) + (\nu - 1)\varepsilon^j h(x_{t+1}^{ij}) \right] \\ & +\mu_t \sum_{ij} \lambda^j \left[\pi_t^{ij} (I_t^{ij} - R_t^{ij}) + \pi_{t-1}^{ij} \left(\theta_t^d \frac{d_t^{ij}}{1+n} + \theta_t^x x_t^{ij} \right) - e_t \right] \\ & - \sum_j \eta_t^j \left(\pi_{t+1}^{2j} \; - \sum_i \pi_t^{ij} \; H(x_{t+1}^{ij}, e_{t+1}) \right) \\ & + \sum_{ij} (-1)^i \varphi_t^j \left[V_t^{ij} - v(\frac{I_t^{ij}}{w^2}) \right] \Bigg\} \end{aligned} \tag{23}$$

We are interested by the values of θ^d and θ^x which can be obtained by differentiating $\mathcal{L}$ with respect to R_t^{ij}, θ_{t+1}^d, θ_{t+1}^x and e_{t+1}.[10] The FOC are stated in the Appendix. To obtain the formula for θ^x and θ^d we evaluate these FOC in the steady-state and combine them to obtain the compensated effects, denoted by a tilde and defined as

$$\begin{aligned} \frac{\partial \tilde{\mathcal{L}}}{\partial \theta^x} &= \frac{\partial \mathcal{L}}{\partial \theta^x} - \sum_i x^{ia} \frac{\partial \mathcal{L}}{\partial R^{ia}}, \\ \frac{\partial \tilde{\mathcal{L}}}{\partial \theta^d} &= \frac{\partial \mathcal{L}}{\partial \theta^d} - \sum_{ij} d^{ij} \frac{\partial \mathcal{L}}{\partial R^{ij}}. \end{aligned}$$

[10] As we are not interested by the income tax formula, we do need to differentiate the Lagrangean with respect to I_t^{ij}.

Using this way of rearranging the FOC, we have:

$$\begin{aligned}\frac{\partial\tilde{\mathcal{L}}}{\partial\theta^x} =&\lambda^a\sum_i \pi^{ia}\left[(\nu-1)\,\varepsilon h'\left(x^{ia}\right)\frac{\partial\tilde{x}^{ia}}{\partial\theta^x}+\ \gamma\mu\left(\frac{\theta^d}{1+n}\frac{\partial\tilde{d}^{ia}}{\partial\theta^x}+\theta^x\frac{\partial\tilde{x}^{ia}}{\partial\theta^x}\right)\right]\\ &+\eta^a\sum_i\pi^{ia}\frac{\partial\tilde{x}^{ia}}{\partial\theta^x}H_1\left(x^{ia},e\right)=0,\\ \frac{\partial\tilde{\mathcal{L}}}{\partial\theta^d} =&\sum_{ij}\lambda^j\pi^{ij}\left[(\nu-1)\,\varepsilon^j h'\left(x^{ij}\right)\frac{\partial\tilde{x}^{ij}}{\partial\theta^d}+\gamma\mu\left(\frac{\theta^d}{1+n}\frac{\partial\tilde{d}^{ij}}{\partial\theta^d}+\theta^x\frac{\partial\tilde{x}^{ij}}{\partial\theta^d}\right)\right]\\ &+\eta^a\sum_i\pi^{ia}\frac{\partial\tilde{x}^{ia}}{\partial\theta^d}H_1\left(x^{ia},e\right)=0.\end{aligned}$$

where the compensated variation of a variable $z=x,d$ with respect to θ^k, $k=x,d$ is defined by

$$\frac{\partial\tilde{z}^{ij}}{\partial\theta^k}=\frac{\partial z^{ij}}{\partial\theta^k}-\sum_{ij}k^{ij}\frac{\partial z^{ij}}{\partial R^{ij}}.$$

Finally, rewriting these tax formulas using the expectation operator E_j $(j=a,n)$ yields[11]

$$\begin{aligned}&-\lambda^a\left[\frac{\theta^d}{1+n}E_a\frac{\partial\tilde{d}}{\partial\theta^x}+\theta^x E_a\frac{\partial\tilde{x}}{\partial\theta^x}\right]\\ =&\frac{(\nu-1)\,\varepsilon\lambda^a}{\gamma\mu}E_a h'(x)\,\frac{\partial\tilde{x}}{\partial\theta^x}+\frac{\eta^a}{\gamma\mu}E_a\frac{\partial\tilde{x}}{\partial\theta^x}H_1(x,e)\,, \qquad (24)\\ &-\left[\frac{\theta^d}{1+n}\sum_j\lambda^j E_j\frac{\partial\tilde{d}}{\partial\theta^d}+\lambda^a\theta^x E_a\frac{\partial\tilde{x}}{\partial\theta^d}\right]\\ =&\frac{(\nu-1)\,\varepsilon\lambda^a}{\gamma\mu}E_a h'(x)\,\frac{\partial\tilde{x}}{\partial\theta^d}+\frac{\eta^a}{\gamma\mu}E_a\frac{\partial\tilde{x}}{\partial\theta^d}H_1(x,e)\,. \qquad (25)\end{aligned}$$

Two comments are in order on these two formulas. First if $\eta^a=0$ and $\nu=1$, namely if there is no Pigouvian externality nor laundering out, then $\theta^x=\theta^d=0$. We are then placed in the framework of Atkinson-Stiglitz proposition. Second, when this is not the case, i.e., when $\eta^a\neq 0$ or $\nu<1$, then one should have a nonzero tax or a subsidy *both on private education and on second period consumption.* This is a surprising result to

[11] For any variable z we define

$$E_j\frac{\partial\tilde{z}}{\partial\theta^k}=\sum_i\pi^{ij}\frac{\partial\tilde{z}^{ij}}{\partial\theta^k}.$$

emerge within an Atkinson-Stiglitz type setting. It means that taxing (or subsidizing) second period consumption becomes now desirable in spite of the separability of preferences and even though it was not for redistributive reasons but as a way of achieving the optimal level of private education. Take the case where private education should be subsidized ($\nu = 1$) and assume that θ^d has a positive effect on the compensated demand for private education. One can expect a tax on second period consumption. Note that one can easily interpret θ^x and θ^d as a tax on saving and a subsidy on the net-of-tax amount of resources saved for education.

The second point concerns the differences between the subsidy on education in the two informational alternatives. Here we have a single rate for both types of productivity. Assuming zero cross derivatives, (25) implies $\theta^d = 0$ while (24) yields

$$\theta^x = \frac{(1-\nu)\,\varepsilon\lambda^a E_a h'(x)\,\dfrac{\partial\tilde{x}}{\partial\theta^x} - \eta^a E_a \dfrac{\partial\tilde{x}}{\partial\theta^x} H_1(x,e)}{\gamma\mu\ \lambda^a E_a \dfrac{\partial\tilde{x}}{\partial\theta^x}}. \tag{26}$$

This formula can be compared with (19). It has the same components as (19) but averaged over the two types of productivity. In the numerator the externality term pushes for a subsidy and the possibility of laundering out ($\nu < 1$) pushes for a tax.

Turning to the level of public education, we combine the FOC with respect to e and (26) and use the notation defined by (21) to obtain:

$$\begin{aligned}\frac{\partial\mathcal{L}}{\partial e} = {}& (\nu-1)\,\varepsilon\frac{E_a h'(x)\,\dfrac{\partial\tilde{x}}{\partial\theta^x}}{\theta^x E_a \dfrac{\partial\tilde{x}}{\partial\theta^x}} + \eta^a \frac{E_a H_1(x,e)\,\dfrac{\partial\tilde{x}}{\partial\theta^x}}{\theta^x\lambda^a E_a \dfrac{\partial\tilde{x}}{\partial\theta^x}} \\ & + \eta^a \overline{H}_2^a(x,e) + \eta^n \overline{H}_2^n(0,e)\end{aligned} \tag{27}$$

To interpret equation (27), we have to keep in mind that the sign of θ_x is ambiguous. It is likely to be positive with laundering out ($\nu = 0$) but negative when the government gives a sufficiently high weight to the parent's joy of giving (ν close to 1). In other words, with low value of ν, the first term of the RHS of (27) is negative and the second term is positive. With high value of ν, it is the other way around. Naturally, the last two terms are positive.

As we have noted for the first-best solution, one cannot find a clear relation between public education and the laundering out parameter ν. When this latter parameter is zero or close to zero, one cannot exclude taxing x.

When compared to its counterparts in the first-best (11) and in the second-best with individually observable expenditures (20), expression (27) appears to have a rather different structure. However, this is mainly due to the way we have rearranged the FOCs. If we return to the original conditions (4e), (17e) and (A.1) it appears that they have all a similar structure. Assuming interior solutions, we have in either case:

$$\frac{\partial \mathcal{L}}{\partial e} = \sum_{i,j} \eta^i \pi^{ij} H_2\left(x^{ij}, e\right) - \mu\gamma = 0. \tag{28}$$

To go from (28) to (27) we have replaced the cost term $\mu\gamma$ from (26) and this yields the first two terms in (27). These terms thus represent the social benefits associated with the change in θ_x required to maintain budget balance. The effect of θ_x in turn is assessed through its impact on x.[12]

Benefits (or costs) of a variation in x of two types: joy of giving (first term) and human capital (second term). When there is no laundering out (standard utilitarian approach), the joy of giving does not influence the choice of e. With laundering out and taxation ($\theta^x > 0$) the joy of providing education has a negative effect on public education. The second term of the RHS of (27) gives the effect of taxing/subsidizing private education on human capital. With a tax (subsidy) this effect is positive (negative). It is thus clear that ν has an ambiguous effect on expression (27). More generally, it is rather difficult to assess under which conditions public education ought to be provided in this setting. When compensated derivatives are equal for both types of productivity and e and x are perfect substitutes, we have:

$$\begin{aligned}\frac{\partial \mathcal{L}}{\partial e} &= \frac{(\nu - 1)\,\varepsilon E_a h'(x)}{\theta^x} + \eta^a \frac{1 + \lambda^a \theta^x}{\lambda^a \theta^x} E_a G'(x + e) \\ &\quad + \eta^n E_n G'(e).\end{aligned}$$

The key factor for a positive public education are clearly: ν close to 1 (laundering out), low λ^a (minority of altruists) and $G'(0)$ very high.

4. CONCLUSION

In a society with just altruists, if public education is a perfect substitute of private education and the government adopts a utilitarian viewpoint, there is no ground for public education. Redistribution is achieved through a non linear tax on income. There is just one reason for public intervention

[12] This is where the linear case differs from the two other cases in which x is controlled directly.

through an educational subsidy: individuals don't internalize the effect of their educational choice on next generation's probability to be productive.

If public and private education are not substitutes, there can be a case for public investment. If the government does not value properly individuals' joy of giving, the subsidy may become a tax. More importantly, if a fraction of individuals are not altruistic, namely they do not get any utility out of their children's education, then public education is unavoidable. This is the main message of this paper.

APPENDIX A

A1 First-order conditions of problem (23)

Differentiating (23) yields the following expressions

$$\begin{aligned}
\frac{\partial \mathcal{L}}{\partial R_t^{ia}} =&\gamma^t \left\{ \lambda^a \pi_t^{ia} \left[\frac{\partial V_t^{ia}}{\partial R_t^{ia}} + (\nu - 1)\varepsilon \frac{\partial x_{t+1}^{ia}}{\partial R_t^{ia}} h'(x_{t+1}^{ia}) - \mu_t + \gamma \mu_{t+1} \left(\frac{\theta_{t+1}^d}{1+n} \frac{\partial d_{t+1}^{ia}}{\partial R_t^{ia}} \right.\right.\right. \\
&\left.\left.\left. + \theta_{t+1}^x \frac{\partial x_{t+1}^{ia}}{\partial R_t^{ia}} \right) \right] + \eta_t^a \pi_t^{ia} \frac{\partial x_{t+1}^{ia}}{\partial R_t^{ia}} H_1(x_{t+1}^{ia}, e_{t+1}) + (-1)^i \varphi_t^a \frac{\partial V_t^{ia}}{\partial R_t^{ia}} \right\} = 0, \\
\frac{\partial \mathcal{L}}{\partial R_t^{in}} =&\gamma^t \left\{ \lambda^n \pi_t^{in} \left[\frac{\partial V_t^{in}}{\partial R_t^{in}} - \mu_t + \gamma \mu_{t+1} \frac{\theta_{t+1}^d}{1+n} \frac{\partial d_{t+1}^{in}}{\partial R_t^{in}} \right] + (-1)^i \varphi_t^n \frac{\partial V_t^{in}}{\partial R_t^{in}} \right\} = 0, \\
\frac{\partial \mathcal{L}}{\partial \theta_{t+1}^d} =&\gamma^t \left\{ \sum_{ij} \lambda^j \pi_t^{ij} \left[\frac{\partial V_t^{ij}}{\partial \theta_{t+1}^d} + (\nu - 1)\varepsilon^j \frac{\partial x_{t+1}^{ij}}{\partial \theta_{t+1}^d} h'(x_{t+1}^{ij}) + \gamma \mu_{t+1} \left(\frac{d_{t+1}^{ij}}{1+n} \right.\right.\right. \\
&\left.\left. + \frac{\theta_{t+1}^d}{1+n} \frac{\partial d_{t+1}^{ij}}{\partial \theta_{t+1}^d} + \theta_{t+1}^x \frac{\partial x_{t+1}^{ij}}{\partial \theta_{t+1}^d} \right) \right] + \eta_t^a \sum_i \pi_t^{ia} \frac{\partial x_{t+1}^{ia}}{\partial \theta_{t+1}^d} H_1(x_{t+1}^{ia}, e_{t+1}) \\
&\left. + \sum_{ij} (-1)^i \varphi_t^j \frac{\partial V_t^{ij}}{\partial \theta_{t+1}^d} \right\} = 0, \\
\frac{\partial \mathcal{L}}{\partial \theta_{t+1}^x} =&\gamma^t \left\{ \sum_i \lambda^a \pi_t^{ia} \left[\frac{\partial V_t^{ia}}{\partial \theta_{t+1}^x} + (\nu - 1)\varepsilon \frac{\partial x_{t+1}^{ia}}{\partial \theta_{t+1}^x} h'(x_{t+1}^{ia}) + \gamma \mu_{t+1} \left(\frac{\theta_{t+1}^d}{1+n} \right.\right.\right. \\
&\left.\left. \frac{\partial d_{t+1}^{ia}}{\partial \theta_{t+1}^x} + x_{t+1}^{ia} + \theta_{t+1}^x \frac{\partial x_{t+1}^{ia}}{\partial \theta_{t+1}^x} \right) \right] + \eta_t^a \sum_i \pi_t^{ia} \frac{\partial x_{t+1}^{ia}}{\partial \theta_{t+1}^x} H_1(x_{t+1}^{ia}, e_{t+1}) \\
&\left. + \sum_i (-1)^i \varphi_t^a \frac{\partial V_t^{ia}}{\partial \theta_{t+1}^x} \right\} = 0, \\
\frac{\partial \mathcal{L}}{\partial e_{t+1}} =&\gamma^t \left\{ -\mu_{t+1}\gamma + \sum_{ij} \eta_t^j \pi_t^{ij} \; H_2\left(x_{t+1}^{ij}, e_{t+1} \right) \right\} \leq 0. \qquad \text{(A.1)}
\end{aligned}$$

A2 Specification of the resource constraint in expression (23)

With this notation, the resource constraint (3) can be reformulated as

$$\sum_{ij} \lambda^j \pi_t^{ij} I_t^{ij} = \sum_{ij} \lambda^j \left[\pi_t^{ij} \left(R_t^{ij} - p_{t+1}^d d_{t+1}^{ij} - (1+n) p_{t+1}^x x_{t+1}^{ij} \right) \right.$$
$$\left. + \pi_{t-1}^{ij} \left(\frac{d_t^{ij}}{1+n} + x_t^{ij} \right) + e_t \right].$$

Then, the aggregate resource constraint with discount rate γ becomes

$$\sum_t \gamma^t \left\{ \sum_{ij} \lambda^j \pi_t^{ij} I_t^{ij} - \sum_{ij} \lambda^j \left[\pi_t^{ij} \left(R_t^{ij} - p_{t+1}^d d_{t+1}^{ij} - (1+n) p_{t+1}^x x_{t+1}^{ij} \right) \right. \right.$$
$$\left. \left. + \pi_{t-1}^{ij} \left(\frac{d_t^{ij}}{1+n} + x_t^{ij} \right) + e_t \right] \right\} = 0,$$

or, using $1 + r = (1+n)/\gamma$,

$$\sum_t \gamma^t \left\{ \sum_{ij} \lambda^j \left[\pi_t^{ij} (I_t^{ij} - R_t^{ij}) + \pi_{t-1}^{ij} \left[\theta_t^d \frac{d_t^{ij}}{1+n} + \theta_t^x x_t^{ij} \right] - e_t \right] \right\} = 0$$

which in turn implies that (22) must be satisfied in each period t.

REFERENCES

Atkinson, A.B. and J. Stiglitz, 1976. The structure of indirect taxation and economic efficiency. *Journal of Public Economics* **1**, 97-119.

Benabou, R., 2002. Tax and education policy in a heterogenous-agent economy: What levels of redistribution maximize growth and efficiency? *Econometrica* **70**, 481-517.

Cremer, H., P. Pestieau, and J-Ch. Rochet, 2003. Capital income taxation when inherited wealth is not observable. *Journal of Public Economics* **87**, 2475-90.

Cremer, H. and P. Pestieau, 2005. Intergenerational transfer of human capital and optimal education policy. *Journal of Public Economic Theory*, forthcoming.

De Fraja, G., 2002. The design of optimal educational policies. *Review of Economic Studies* **69**, 437-66.

Fernandez, R. and R. Rogerson, 1996. Income distribution, communities and the quality of public education. *Quarterly Journal of Economics* **111**, 135-64.

Galor, O. and J. Zeira, 1993. Income distribution and macroeconomics. *Review of Economic Studies* **60**, 35-52.

Glomm, G. and B. Ravikumar, 1992. Public versus private investment in human capital: Endogenous growth and income inequality. *Journal of Political Economy* **100**, 818-34.

Hammond, P., 1987. Altruism, in The New Palgrave. *A Dictionary of Economics*, Macmillan, London.

Loury, G., 1981. Intergenerational Transfers and the Distribution of Earnings. *Econometrica* **49**, 843-67.

Mirrlees, J. A., 1971. An exploration in the theory of optimal taxation. *Review of Economic Studies* **38**, 175-208.