Should Educational Policies be Regressive?∗

Daniel Gottlieb,† Humberto Moreira‡

January 31, 2005

Abstract

De Fraja (2002) has shown that when the ability to benefit from education is unobservable, optimal policies are regressive. We argue that his results follow from an unusual specification of the government’s budget constraint. When this constraint is replaced by the usual one, regressive policies are no longer optimal. The optimal educational policy can be decentralized through Pigouvian taxes and credit provision and is not regressive. Furthermore, when the utility function is not quasi-linear, education may not be monotonic in ability and progressivities of education are locally welfare improving.

Keywords: Education, Redistribution, Optimal Taxation, Pigouvian taxes.
JEL Classification: I22, H21, H23, H52.

∗We wish to thank Luis Henrique Braido, Carlos E. da Costa, Daniel Ferreira, Andrew Horowitz, Rodrigo Soares, Thierry Verdier and specially Antoine Bommier, Pierre Dubois, and James Poterba for helpful comments and suggestions. We would also like to thank seminar participants at Getulio Vargas Foundation, UFRJ, Toulouse, the 2004 North American Summer Meeting of the Econometric Society, the 2004 Meeting of the European Economic Association, and the 2003 Meeting of the Brazilian Econometric Society. As usual, all remaining errors are ours.
†Department of Economics, MIT.
‡Graduate School of Economics, Getulio Vargas Foundation. Corresponding Author. Postal Address: Praia de Botafogo, 190, sala 1121, CEP: 22253-900, Rio de Janeiro, RJ, Brazil. E-mail: humberto@fgv.br.
1 Introduction

The role of educational policies in the equalization of opportunities is a widely accepted issue in political debates. However, a remarkable feature of most educational systems in the world is the huge regressivity of spending per students. Economists have usually argued that a reduction in such regressivity would result in significant welfare benefits. The trade-off between redistribution and efficiency plays a fundamental role in the analysis of optimal tax policies. In the specific case of education, this issue was originally addressed by Hare and Ulph (1979) who introduced schooling in Mirrlees’ (1971) original model and assumed that the ability to benefit from it was only partially observable by the government. Fleurbaey et al (2002) extended the Hare and Ulph model to the case where the ability to benefit from education is unobservable and the government has a Rawlsian welfare function.

In a recent contribution, De Fraja (2002), hereafter DF, studied the optimal educational provision in an overlapping-generations model in the presence of externalities and imperfect capital markets. His surprising results suggest that educational policies: (i) should be regressive (in the sense that households with brighter children and higher incomes contribute less in absolute terms than those with less bright children and lower incomes), (ii) should not provide equality of opportunities in education (in the sense of the irrelevance of the household’s income to the education received by a child), and (iii) should be input-regressive (meaning that education should be increasing in ability). Therefore, the regressivity of educational systems in most countries may actually reflect the optimal educational policies and the provision of equality of opportunities in education may imply a great efficiency loss. We argue that the results obtained by DF critically rely on two specific assumptions which are hard to justify and are not common in the taxation literature: a nonstandard government budget constraint and quasi-linear utility functions combined with a utilitarian welfare function.

The regressivity of the financing mechanism and the inequality of opportunities (results i and ii mentioned above) are caused by a nonstandard restriction on the government’s budget constraint: it is assumed that deferred payments made by the older generation cannot be used to finance the education of the younger generation. However, the cross-subsidization of education is one important feature of public educational systems. Moreover, this restriction on the budget constraint generates a welfare loss. Hence, there seems to be neither positive nor normative justifications for this assumption.

When deferred payments are allowed to be used in education expenditures in the DF model, the optimal educational policy takes a very different form: it achieves first-best welfare (the maximum amount of welfare that could be reached under perfect information) and provides equality of opportunities in education. Moreover, it can be implemented in a decentralized way through a competitive equilibrium with Pigouvian taxes and the provision of credit.

In the decentralized mechanism, first-best welfare is reached through a subsidy on education to correct the externalities, a lump-sum tax proportional to the average education and the provision of credit (at the market interest rate). Such a mechanism is not regressive (i.e., wealthier households do not contribute less than poorer households and households with brighter children contribute more than those with less bright children) and can also be implemented in an environment where household’s wealth is unobservable. Unlike argued by DF, the reason for the regressivity and the inequality of opportunities of the second-best policy is not the trade-off between efficiency and equity but one between efficiency and rent-extraction. Since the utility function is quasi-linear and the government has a utilitarian welfare function, there is no preference for redistribution and the government simply minimizes deadweight loss. The need to use education in order to extract rents arises simply due to the unusual restriction that prevents the use of revenue from the deferred payments to finance education. Then, since poorer individuals are those who

---

1 See, for example, Fernandez and Rogerson (1996), Kozol (1991) or Psacharopoulus (1986).
2 Fernandez and Rogerson (1998) estimate the steady-state welfare increase of an equalization of spending across districts in American states increases welfare in the magnitude of 3.2 percent of steady-state income.
3 More specifically, they assumed that although the government could not condition financial contributions on the ability to benefit from education, it could use this information for allocational purposes.
4 According to DF: “The regressivity of the optimal education system derived in the paper can be interpreted as implying that pursuing redistributive goals using education policies is bound to have a substantial cost in terms of the sub-optimality of the education policy implemented: the fundamental message of the paper is that there is a stark conflict between equity and efficiency in education.”
5 Provided that the aggregate income is such that it is socially optimal for someone to leave positive bequests.
benefit most from education, minimizing deadweight loss implies that the education of poorer individuals should be taxed more.\footnote{This result is similar to the standard inverse elasticity rule of Ramsey (1927). The assumption of positive education externalities implies that the uniform commodity taxation result does not apply and it is optimal to tax education and consumption at different rates even though the utility function is separable.}

Since the utility function is assumed to be linear in the wealth left to the next generation, the optimal policy implies in a large inequality of the future generation’s wealth and no inequality of consumption. However, it seems counter-intuitive that society cares about the redistribution in the current period but does not care at all about inequality of the wealth of future generations.

We show that when the utility functions are concave in the wealth of the future generation input-regressive policies may no longer be incentive-compatible. Moreover, the first-best welfare will not be achievable through Pigouvian taxes and credit provision in this case. This follows from the impossibility of the decentralized mechanism to solve for both inefficiency and redistribution at the same time.

The non-implementability of input-regressive policies result is fairly general as it does not depend on the specific welfare function considered. Hence, it applies not only to the DF model, but also to the case where the welfare function is Rawlsian as in Fleurbaey et al (2001). Therefore, our model suggests that, unlike argued by DF and Fleurbaey et al, it may be the case that education is never increasing in ability and the attempt to implement a monotonic educational policy may actually cause welfare losses.

The remainder of the paper is organized as follows. Since the basic structure is the same as DF, we omit its presentation as well as the laissez-faire equilibrium. Section 2 contrasts the government budget constraint specified by DF with the usual specification. Section 3 presents the government-intervention solution when the ability to benefit from education is observable (first-best). Section 4 characterizes the optimal allocations when the ability to benefit from education is not observable (second-best) and discusses its implementation. Then, Section 5 studies the cases where parents’ utility function is concave in the wealth left to their children. Section 6 summarizes the main results of the paper.

## 2 The Government’s Budget Constraint

For notational convenience we define the expectations operator \( \bar{E} [\cdot] \) as

\[
\bar{E} [x] \equiv \sum_{i=1}^{n} h_i (w (T)) \int_{\theta_0}^{\theta_1} x_i (\theta) \phi (\theta) d\theta,
\]

where \( x = \{ x_i (\theta) : \theta \in [\theta_0, \theta_1] \, , \, i \in \{1, 2, \ldots, n\} \} \).\footnote{For notational simplicity, we shall omit the dependence of \( h_i (w (T)) \) on \( w (T) \).}

Following DF, we assume that the government can offer a tax schedule and an education schedule. A tax schedule consists of an income tax \( \tau_i \). An education schedule consists of an offer of education \( c_i (\theta) \), an up-front fee \( f_i (\theta) \) and a deferred payment \( m_i (\theta) \). Since \( f_i (\theta) \) and \( m_i (\theta) \) may be positive or negative, the government is able to offer loans to students. The parent’s wealth \( Y_i \) is observable but the daughter’s ability to benefit from education \( \theta \) is private information.

With no loss of generality, each household’s bequests can be normalized to zero. In that case, all bequests are left through up-front fees and deferred payments. The household’s budget constraint is

\[
Y_i = c_i (\theta) + \tau_i + f_i (\theta) . \tag{1}
\]

In each period a mother with wealth \( Y_i \) and whose child has ability \( \theta \) pays \( \tau_i + f_i (\theta) \) as up-front taxes and \( m_i (\theta) \) as deferred payments (due to the education received in her childhood) and receives \( k e_i (\theta) \) as education subsidies. In DF, it is imposed that the government is unable to use revenues from deferred payments to be spent in education or income tax. Hence, the budget constraint is

\[
\bar{E} [\tau + f] \geq \bar{E} [ke] , \quad \bar{E} [m] \geq 0 ,
\]

where the first equation states that the total amount of income taxes and up-front fees are used to finance the educational expenses while the second equation states that the total amount of deferred payments is non-negative.
As deferred payments and education are the only channels of transferring wealth between generations (since bequests have been normalized to zero), imposing that the aggregate amount of deferred payments be non-negative is equivalent to restricting all transfers between generations to be done through the provision of education. However, contributions for public education account for a substantial amount of intergenerational transfers. Bommier et al. (2004) have shown that when expenditures in public education are included in generational accounts along with pensions and health care, generations usually thought to have benefitted from intergenerational transfers were actually net losers.

The only justification provided for such specification is given in footnote 11:

"Of course future taxes could also be used to repay the debt, but this would reduce the amount available for the education of future generations, and must therefore be subject to a limit, normalized here to 0; this value is consistent with a steady state analysis, where the taxes levied on the future generation must correspond to those levied on the current generation to pay for the previous generation’s education."

Nevertheless, this does not explain why the government would not be able to use revenues from deferred payments to finance education (or income tax) since in each period all taxes are paid by the same generation (namely, by the parents). Hence the assumption that the budget constraint is balanced with each generation in each period is also consistent with a steady state analysis. Moreover, since the latter is the usual assumption in overlapping-generation models, is weaker than DF’s specification, and is consistent with the evidence on the importance of intergenerational transfers for public education, it does not seem reasonable to further restrict the government’s problem.\(^8\)

We will allow the government to finance current educational expenses through deferred payments. Therefore, the government’s budget constraint is

\[
\bar{E} [\tau + f + m] \geq \bar{E} [ke]. \tag{2}
\]

Equation (2) states that the net tax revenue (L.H.S.) is enough to finance the educational expenses (R.H.S.).

### 3 First-Best Solution

As in most public-finance literature, we take a utilitarian government. Moreover, we follow DF in assuming that the utility of future generations enter the welfare function only through the weight attributed to it by the current generation.

As usual, we shall refer to the outcome chosen by the government if ability were observable as a first-best solution. Then, a first-best solution is a profile of education and bequests \( \{ e_i(\theta), t_i(\theta); \theta \in [\theta_0, \theta_1], i \in \{1,2,\ldots,n\} \} \) that maximizes the sum of each parent’s utilities subject to the constraint the total level of education is the sum of every individual’s education:

\[
\max_{\{ e_i(\theta), t_i(\theta), E \}} \bar{E} [u(Y - t - ke) + y(\theta, e; E) + t] \\
\text{s.t.} \quad E = \bar{E} [e]. \tag{3}
\]

Notice that the social marginal benefit of education consists of the private marginal return of education \( y_e \) and the social return of education \( \bar{E} [y_E] \). Hence, the first-best amount of education should be such that the marginal benefit of education equals its marginal cost \( k \). Therefore, the first-best profile of education is implicitly determined by the relation:\(^9\)

\[
k = y_e(\theta, e^*_e(\theta); \bar{E}[e^*]) + \bar{E} [y_E(\theta, e^*_e(\theta), \bar{E}[e^*])].
\]

Since efficiency requires that marginal productivity of education must be equalized for all individuals, the amount of education received by an individual should depend only on her ability. Therefore, the first-best

\(^{8}\)Furthermore, the choice of 0 for the limit of deferred payments allowed to be spent in education is not just a normalization. As will be clear in Section 4, the first-best is obtained even under the unusual budget constraint specified by DF if this limit is lower than the total cost of first-best expenditures on education minus the first-best amount of income taxes.

\(^{9}\)The proofs of existence and uniqueness of the solutions obtained in this paper can be found at Gottlieb and Moreira (2003).
amount of education is independent of the parent’s wealth \(Y_i\) (we shall denote \(e^*_i(\theta)\) as \(e^*(\theta)\) in order to emphasize that it does not depend on \(Y_i\)). Moreover, since \(e^*\) is independent of \(Y_i\), the optimal consumption level is also independent of wealth.

Furthermore, as long as there are enough resources so that \(e^*\) and \(c^*\) are feasible, any profile of bequests yields the same welfare (since the utility function is linear in bequests and the welfare function is utilitarian). Then, an optimal profile of bequests consists of the amount of resources left to the household after the expenditures in education and consumption:

\[
t^*_i(\theta) = Y_i - ke^*_i(\theta) - c^*.
\]

As can be seen from the equation above, the marginal propensity to bequeath under the first-best policy is equal to one.

We will assume that resources are sufficiently high so that \(e^*\) and \(c^*\) are feasible. If this assumption did not hold, then it would not be socially optimal to have any individual leaving positive bequests.

**Assumption 1** \(\bar{E}[t^*] \geq 0\).

**Proposition 1** The first-best allocations are

\[
\{e^*, e^*(\theta), t^*_i(\theta); \theta \in [\theta_0, \theta_1], i \in \{1, 2, \ldots, n\}\}.
\]

**Proof.** See Appendix. ■

Because marginal productivity of education is increasing in ability, it follows that education provided in the first-best solution is also increasing in ability (i.e., the first-best equilibrium is input-regressive).\(^{10}\)

**Remark 1** The presence of positive externalities implies an inefficiently low amount of education provided in the laissez-faire equilibrium even for unconstrained households (since \(y\) is strictly concave in \(e\)).

4 The Second-Best Solution

Substituting the household’s budget constraint (1) in the utility function, it can be written as

\[
U_i(\theta) = u(Y_i - \tau_i - f_i(\theta)) + y(\theta, e_i(\theta); E) - m_i(\theta).
\]

(4)

From the revelation principle, the search for an optimal educational policy can be restricted to the class of incentive-compatible mechanisms with no loss of generality. The following lemma, whose proof can be found at DF, allows us to substitute the incentive-compatibility constraint for the local first- and second-order conditions.

**Lemma 1** A \(C^2\) by parts policy \(\{\tau_i, f_i(\theta), m_i(\theta), e_i(\theta); \theta \in [\theta_0, \theta_1], i \in \{1, 2, \ldots, n\}\}\) is incentive-compatible if, and only if, it satisfies

\[
\dot{U}_i(\theta) = y_\theta(\theta, e_i(\theta); E),
\]

(5)

\[
\dot{e}_i(\theta) \geq 0,
\]

(6)

for all \(\theta \in [\theta_0, \theta_1], i \in \{1, 2, \ldots, n\}\).

It is also assumed that individuals are not forbidden to purchase education in the private sector. Hence, they will only join the educational program when their utility exceeds the utility obtained if they purchase education privately. Then, the household’s utility must satisfy

\[
U_i(\theta) \geq P(\theta, Y_i - \tau_i, E),
\]

(7)

for all \(\theta \in [\theta_0, \theta_1], i \in \{1, 2, \ldots, n\}\).

As usual, we shall refer to the optimal contract chosen by the government when ability is not observable as the second-best solution:

\(^{10}\)Applying the implicit function theorem, we get \(\frac{\partial e^*_i(\theta)}{\partial \theta} = -\frac{y_{ee}}{y_{e\theta}} > 0\).
Definition 1 A second-best equilibrium is a policy \( \{ \tau_i, f_i(\theta), m_i(\theta), e_i(\theta) : \theta \in [\theta_0, \theta_1], i \in \{1, 2, ..., n\} \} \) solving

\[
\max_{\{c, \tau, f, m, E\}} \mathbb{E}[u(Y - \tau - f) + y(\theta, e; E) - m] \\
\text{s.t. } (2)-(7).
\]

The following proposition ensures that the government is able to implement the efficient level of education and consumption in an economy with private information and where an individual may choose not to join the public education system. Moreover, since this implementation does not require any additional resources, it achieves first-best welfare.

Proposition 2 The optimal educational policy implements the first-best amount of education and consumption \( \{e^*(\theta), c^*; \theta \in [\theta_0, \theta_1], i \in \{1, 2, ..., n\} \} \) and achieves first-best welfare. Furthermore, the optimal profile of deferred payments is such that \( m^*_i(\theta) = -t^*_i(\theta) \).

Proof. See Appendix. ■

The basic intuition behind this result is that when the government raises the income tax uniformly and decreases the up-front fee in the same amount, the indirect utility of an individual participating in the proposed scheme remains constant, whereas the indirect utility of an individual who purchases education privately decreases. Hence, the participation constraint (7) can be implemented at no cost. Moreover, since the utility function is quasi-linear and the social welfare function is utilitarianist, any redistribution of wealth does not change welfare.\(^{11}\)

Proposition 2 implies that when we consider budget constraint (2), a strictly higher welfare is achieved (since first-best is implemented). Moreover, contrary to the results of DF, the optimal policy provides equality of opportunities in education (since \( e^*(\theta) \) does not depend on \( Y_i \)).\(^{12}\)

Although Proposition 2 states that the first-best can be achieved, it does not show how. In the next section, we address the implementation of the optimal policy. It is shown that the first-best can be obtained through Pigouvian taxes and public provision of credit. Moreover, it does not require that the government observes wealth so that our results also hold in an environment where \( Y_i \) is unobservable.

4.1 Implementation through Pigouvian taxes

In this subsection, we will restrict the space of contracts to those consisting of lump-sum taxes, a linear up-front fee, and a deferred payment. Formally, let \( \tau_i, f_i(\theta) \) and \( m_i(\theta) \) be the income tax, up-front fee and deferred payments as defined in the last section. Define \( t_i(\theta) \) and \( \hat{k}_i(\theta) \) as

\[
t_i(\theta) = -m_i(\theta),
\hat{k}_i(\theta) = \frac{f_i(\theta) + m_i(\theta)}{e_i(\theta)}.
\]

In general, \( \hat{k} \) and \( t \) could depend on \( \theta \) and \( i \). However, as we show below, they are both constant for all \( \theta \) and \( i \) under the optimal policy. Hence, a contract consists of a lump-sum tax \( \tau \), a linear up-front fee \( t_i(\theta) + \hat{k}e_i(\theta) \) and a deferred payment \( -t_i(\theta) \) (which is a subset of the class of contracts considered previously). This mechanism can be alternatively interpreted as a lump-sum tax \( \tau \), a loan \( -t_i(\theta) \) and an up-front fee \( \hat{k}e_i(\theta) \).

\(^{11}\)In that sense, the model is similar to the regulation model when the shadow-cost of public funds is zero. However, it should be stressed that the shadow-cost of public funds is endogenous unlike in the standard regulation models [e.g. Laффont and Tirole (1993)]. In DF, it is only zero in the extreme case where the government is able to circumvent the restriction that deferred payments cannot be used to finance education and income taxes (Proposition 6). When deferred payments can be used to finance education, the results hold even when the extreme assumptions of DF are not satisfied.

\(^{12}\)Notice also that the second-best education profile is strictly greater than the one obtained in the laissez-faire equilibrium (see Remark 1). This follows from the positive externality of education.

Proposition 2 can be generalized to the case where the returns to education are random so that default may emerge [see Gottlieb and Moreira (2003)]. The intuition for this result is that quasi-linearity implies that individuals are risk-neutral with respect to bequests. Then, they are indifferent between lotteries with high deferred payments only when outcomes are good and fixed payments with the same expected value. Thus, the optimal policy is determined ex-ante but not ex-post and the first-best welfare can be implemented through income contingent payments.
Substituting the definitions of \( \hat{k} \) and \( t \) in the government budget constraint (2), it can be written as

\[
\bar{E} \left[ (k - \hat{k}) e - \tau \right] \leq 0. \tag{8}
\]

In each period, the government pays \( (k - \hat{k}) \) as a subsidy on each unit of education and receives \( \tau \) as a lump-sum tax. The government also loans \( \bar{E} [-t] \) in the first period and receives it in the next period. Since the market interest rate is normalized to 1, \( \bar{E} [-t] \) may take any value because it is always repaid in the following period. In steady-state, repayments are equal to loans in each period. Thus, as usual, the budget constraint simply states that the total expenses should not exceed the total revenues of the government.

Substituting the definitions of \( \hat{k} \) and \( t \) in the parent’s budget constraint (1), it follows that the total amount of consumption, loans repaid, and taxes must be equal to the household’s wealth:

\[
Y_i = c_i + t_i (\theta) + \tau + \hat{k} e_i (\theta). \tag{9}
\]

Hence, we can write the parent’s problem as:

\[
\max_{c_i(\theta), t_i(\theta)} u \left( Y_i - t_i (\theta) - \hat{k} e_i (\theta) \right) + y (\theta, e_i (\theta); E) + t_i (\theta) \quad \text{s.t. } E = \bar{E} [e]
\]

As there are no restrictions on \( t \), the solution must be such that the marginal utility of consumption is equal to the marginal utility of wealth left to the daughter. Hence, each parent must be consuming \( c^* \). Moreover, the private marginal benefit of education \( y_e \) must be equal to its marginal cost \( k \). Therefore, we get the following lemma:

**Lemma 2** The solution to the parent’s problem is \( \{ c_i^P (\theta), e_i^P (\theta), t_i^P (\theta); \theta \in [\theta_0, \theta_1], \ i \in \{1, 2, ..., n\} \} \), such that:

\[
c_i^P (\theta) = c^*, \quad \hat{k} = y_e (\theta, e_i^P (\theta); E), \quad t_i^P (\theta) = Y_i - c^* - \tau_i - \hat{k} e_i^P (\theta). \tag{10}
\]

**Proof.** The result follows from the first-order conditions (which are necessary and sufficient). \( \blacksquare \)

Now, we are ready to show that the first-best solution can be reached through a suitable choice of \( \tau \) and \( \hat{k} \). The price of education \( \hat{k} \) is chosen in order to internalize for the educational externalities. Hence, it must be equal to the private cost of education \( k \) minus the educational externalities \( \bar{E} [ye] \). The lump-sum tax is set in order to cover the expenses from the subsidies. Therefore, it must be equal to the average subsidy \( \bar{E} [eP] \bar{E} [ye] \). Substituting into the parent’s problem, we get the following result:

**Proposition 3** There exists a second-best equilibrium where the price of education \( \hat{k} \) and the tax \( \tau \) are both constant in \( \theta \) and \( Y_i \). Moreover, this equilibrium implements first-best welfare.

**Proof.** See Appendix. \( \blacksquare \)

The fact that the optimal policy can be implemented with a lump-sum tax \( \tau \), a constant price of education \( \hat{k} \) and a constant interest rate implies that it can also be implemented when household’s wealth is unobservable.

Define the household’s financial contribution as

\[
z_i (\theta) \equiv \tau + \hat{k} e^* (\theta). \]

As education is independent of wealth, it is clear that an individual’s financial contribution is independent of her income. Moreover, \( z_i (\theta) \) is strictly increasing in ability since \( \hat{k} > 0 \) and \( e^* > 0 \). Therefore, households with brighter children contribute more than households with less bright children. These results differ from

Let $x_i^P$ denote the wealth left to the daughter:

$$x_i^P(\theta) \equiv y(\theta, e_i^P(\theta); E) + t_i^P(\theta).$$

Then, it follows that the marginal propensity to bequeath under the Pigouvian scheme is equal to one (i.e., $\frac{\partial x_i^P}{\partial Y} = 1$). In other words, every additional amount of wealth is left to the future generation. Furthermore, individuals with higher ability receive more wealth through education than ones with less ability (since $\dot{x}_i^P = y_\theta > 0$). Therefore, the optimal policy generates large inequalities of wealth left to the future generation. However, the quasi-linearity of the utility function implies that such outcome is optimal. In the next section, we study how the results change when the utility function is not quasi-linear so that individuals care not only about redistribution of present consumption but also about redistribution of wealth left for future generations.

5 Preference for Redistribution

Throughout the previous sections, it has been assumed that parents’ preferences can be represented by a utility function linear in the wealth left to their children. This assumption implies that a utilitarian government does not have preference for redistribution of bequests. Then, as was shown in Subsection 4.1, the optimal policy implements first-best welfare but generates large inequalities of wealth left to the future generation.

In this section, we assume that parents’ utility function is concave in their children’s wealth:

$$U = u(c) + v(x),$$

$v' > 0$, $v'' < 0$.

This assumption implies that parents are risk-averse in the wealth of their children and a utilitarian government has preference for redistribution.

Substituting the parent’s budget constraint in her utility function, we get

$$U_i(\theta) = u(Y_i - r(\theta) - f_i(\theta)) + v(y(\theta, e_i(\theta); E) - m_i(\theta)).$$

Then, it follows that

$$U_{\epsilon}\theta = v'(y - m)[y_{e\theta} - r_A(\epsilon, y - m)y_{e\theta}].$$

where $r_A(\epsilon, y) \equiv -\frac{\partial v(y)}{\partial \epsilon} y$ is the absolute coefficient of risk-aversion and measures the concavity of $v$.

Notice that the sign of $U_{\epsilon}\theta$ is ambiguous since an increasing profile of education has two opposite effects in the parent’s utility. The first effect ($y_{e\theta} > 0$) concerns efficiency: an increasing profile of education benefits more those with higher marginal productivity of education. The second effect ($-r_A y_{e\theta} < 0$) concerns equity: an increasing profile of education gives more wealth to those with lower marginal utility. Then, the sign of $U_{\epsilon}\theta$ will depend on the preference for redistribution relative to efficiency (captured by the risk-aversion coefficient $r_A$).

The following lemma presents the standard necessary condition for incentive-compatibility.

**Lemma 3** An incentive-compatible $C^2$ by parts policy $\{\tau_i, f_i(\theta), m_i(\theta), e_i(\theta); \theta \in [\theta_0, \theta_1], i \in \{1, 2, ..., n\}\}$ satisfies

$$\frac{\partial^2 U_i(\theta)}{\partial e \partial \theta} e_i(\theta) \geq 0,$$

for all $\theta \in [\theta_0, \theta_1], i \in \{1, 2, ..., n\}$.

**Proof.** Omitted.

Thus, incentive-compatible policies are generally not input-regressive. When preference for redistribution is sufficiently small, an incentive-compatible policy is input-regressive ($\dot{e}_i > 0$) as in the quasi-linear environment. If the concern for redistribution is sufficiently high, input-regressive policies are no longer input-regressive.

---

13The preference for redistribution is ignored by DF since $r_A$ is zero when the utility function is quasi-linear.
incentive-compatible. Indeed, incentive-compatible educational policies may also be non-monotonic. This would be the case when the sign of $U_{\theta}$ changes (i.e., the single-crossing property does not hold).\footnote{Araujo and Moreira (2001) present a method for solving screening models where the single-crossing property does not hold.}

In any of those cases, however, a decentralized mechanism consisting of Pigouvian taxes and provision of credit can no longer implement the first-best. The intuition is that a decentralized mechanism is unable to redistribute wealth appropriately. Then, the equilibrium is sub-optimal since it maximizes efficiency instead of taking into account the trade-off between efficiency and redistribution.

The trade-off between efficiency and equity implies that optimal policies would typically be distorted towards more progressive policies as in the standard taxation models [e.g. Mirrless (1971)]. Indeed, starting from an undistorted economy (i.e. an economy with Pigouvian taxes and credit provision), it can be shown that a local increase in the progressivity of education is always welfare increasing. This follows from the well-known fact that, in a previously undistorted economy, distortions cause second-order losses and first-order gains in welfare (when the government cares about redistribution).

**Proposition 4** When $v$ is strictly concave, the first-best amount of welfare cannot be reached through credit loans and Pigouvian taxes. Moreover, starting from an undistorted economy, a local increase in the progressivity of education is strictly welfare enhancing.

**Proof.** See Appendix. ■

### 6 Conclusion

In this paper, we have shown that the regressivity results obtained by DF are driven by the assumptions that the government is not allowed to use deferred payments in education and that utility functions are quasi-linear and the welfare function is utilitarian (which imply that it minimizes deadweight loss). Then, the government’s problem is to minimize the inefficiency caused by this nonstandard restriction on its budget constraint. Moreover, contrary to the argument at DF, the relevant trade-off is not one between efficiency and redistribution but one between efficiency and rent-extraction. Not surprisingly, the solution involves taxing more heavily poorer individuals since they are the ones who benefit most from education.\footnote{Since the government can observe wealth, the education purchased by households with different wealth can be seen as different goods. Then, as shown by Ramsey (1927), taxes must be decreasing in the elasticity of demand for each good.}

When we allow the government to use deferred payments in education, the optimal educational policy achieves the same amount of welfare that could be reached if ability were observable (first-best welfare).

By not internalizing the effects that education bears on the rest of the economy, the amount of education that each household provides in the laissez-faire equilibrium is inefficiently low. We show that the first-best solution can be implemented when neither ability nor wealth are observable through Pigouvian taxes and credit provision. In this context, the appropriate Pigouvian taxes are educational subsidies that induce households to internalize for the (positive) externalities caused by education.

As first-best efficiency requires that marginal productivity of education be equalized across individuals, it follows that the amount of education received by a child does not dependent on his/her parent’s wealth. Therefore, equality of opportunities in education is provided. Furthermore, since the optimal amount of financial contribution does not depend on parental income and is increasing in the ability of the child, the optimal educational policy is not regressive.

If we do not impose that utility functions are quasi-linear so that individuals care not only about redistribution of present consumption but also about redistribution of wealth left for future generations, input-regressive policies may not be incentive-compatible. Then, the implementation through Pigouvian taxes and credit provision does not achieve the first-best welfare. This follows from the fact that the decentralized mechanism maximizes efficiency without taking into account the concern for redistribution. Moreover, starting from an undistorted economy, increases in the progressivity of education are locally welfare increasing.
Appendix

Proof of Proposition 1: Introducing the auxiliary variable \( S(\theta) \), (3) can be rewritten as

\[
\dot{S}(\theta) = \sum_{i=1}^{n} h_i e_i(\theta) \phi(\theta), \quad S(\theta_0) = 0, \quad S(\theta_1) = E.
\]

The optimal policy offered to an individual with wealth \( Y \) must solve the following Hamiltonian:

\[
H = \sum_{i=1}^{n} h_i \left[ u(Y_i - t_i(\theta) - ke_i(\theta)) + y(\theta, e_i(\theta); E) + t_i(\theta) \right] \phi(\theta) + \mu(\theta) \sum_{i=1}^{n} h_i e_i(\theta) \phi(\theta),
\]

where \( t_i \) and \( e_i \), \( i = 1, ..., n \), are control variables and \( S \) is the state variable. The first-order conditions are

\[
-u'(Y_i - t_i(\theta) - ke_i(\theta)) + 1 = 0, \quad k u'(Y_i - t_i(\theta) - ke_i(\theta)) + y_e(\theta, e_i(\theta), E) + \mu(\theta) = 0, \quad \mu(\theta) = \mu \text{ constant}.
\]

Let \( W \) be the welfare function. Then, as \( \frac{\partial W}{\partial E}|_{e=e^*, E=E^*} = \mu(\theta_1) = \mu \), it follows that

\[
\bar{E}[y_E(\theta, e; E)] = \mu.
\]

Substituting in the first-order conditions, concludes to proof.

Proof of Proposition 2: From (2), we can substitute \( m_i(\theta) \) in the welfare function:

\[
W = \bar{E}[u(Y - \tau - f) + y(\theta, e; E) + \tau + f - ke] = \bar{E}[u(Y - \tau - f) + y(\theta, e; E) + \tau + f - ke]
\]

Substituting (4) in the above expression, it follows that

\[
W = \int_{\theta_0}^{\theta_1} \sum_{i=1}^{n} h_i \left[ U_i(\theta) + m_i(\theta) + \tau_i + f_i(\theta) - ke_i(\theta) \right] \phi(\theta) d\theta.
\]

Introducing the auxiliary variable \( S(\theta) \), (3) can be rewritten as

\[
\dot{S}(\theta) = \sum_{i=1}^{n} h_i e_i(\theta) \phi(\theta), \quad S(\theta_0) = 0, \quad S(\theta_1) = E. \tag{11}
\]

For the moment, we will ignore the monotonicity condition (6). Latter, we will verify that it will be satisfied in the solution of this relaxed problem. We will also ignore equation (2) and verify that it is satisfied in the solution of the relaxed problem.

Fix an arbitrary wealth level \( Y_i \in \{Y_1, Y_2, ..., Y_n\} \). The optimal policy offered to an individual with wealth \( Y_i \) must solve the following Hamiltonian:

\[
H = \sum_{i=1}^{n} h_i \left[ U_i(\theta) + m_i(\theta) + \tau_i + f_i(\theta) - ke_i(\theta) \right] \phi(\theta) + \rho(\theta) \sum_{i=1}^{n} h_i e_i(\theta) \phi(\theta) + \gamma_i(\theta) y_\theta(\theta, e_i(\theta); E) + \lambda_i(\theta) [u(Y_i - \tau_i - f_i(\theta)) + y(\theta, e_i(\theta); E) - m_i(\theta) - U_i(\theta)] + \mu_i(\theta) [U_i(\theta) - P(\theta, Y_i - \tau_i, E)].
\]

The control variables are \( m_i, f_i, \tau_i, \) and \( e_i \) and the state variables are \( U_i \) and \( S_i \). \( \rho(\theta), \gamma_i(\theta), \lambda_i(\theta), \) and \( \mu_i(\theta) \) are the multipliers associated with (11), (5), (4), and (7), respectively.
The first-order conditions are\(^{16}\)

\[
\frac{\partial H}{\partial m_i (\theta)} = 0 \implies h_i (\phi (\theta) = \lambda_i (\theta), \tag{12a}
\]

\[
\frac{\partial H}{\partial f_i (\theta)} = 0 \implies Y_i - \tau_i - f_i (\theta) = c^*, \tag{12b}
\]

\[
\frac{\partial H}{\partial \tau_i} = 0 \implies \mu_i (\theta) P_Y (\theta, Y_i - \tau_i, E) = 0 \implies \mu_i (\theta) = 0, \tag{12c}
\]

\[
\frac{\partial H}{\partial e_i (\theta)} = 0 \implies h_i (\phi (\theta) [\rho (\theta) - k] + \gamma_i (\theta) y_{ie} + \lambda_i (\theta) y_e = 0, \tag{12d}
\]

\[
\frac{\partial H}{\partial U_i (\theta)} = -\gamma_i (\theta) \colon \gamma_i (\theta) = \gamma_i \text{ constant in } \theta, \tag{12e}
\]

\[
\frac{\partial H}{\partial S (\theta)} = -\rho (\theta) \colon \rho (\theta) = \rho \text{ constant in } \theta, \tag{12f}
\]

\[
0 = \min \{ \mu_i (\theta) \colon U_i (\theta) - P (\theta, Y_i - \tau_i, E) \}. \tag{12g}
\]

From equation (12a), it follows that the first-best amount of consumption is the solution for \(c\) in the relaxed problem.

**Lemma 4** \(\gamma_i = 0\) for all \(i\).

**Proof.** As \(U_i (\theta_1)\) is free for all \(i\), the transversality condition is \(\gamma_i (\theta_1) = 0\). Hence (12e) implies that \(\gamma_i = 0\).

Substituting \(\gamma_i = 0\) in equation (12d) yields

\[
y_e (\theta, e_i (\theta); E) = k - \rho. \tag{13}
\]

**Lemma 5** The amount of education solving the relaxed problem above is the same as in the first-best solution. That is, \(e_i (\theta) = e_i^* (\theta)\), for almost all \((\theta, Y_i) \in [\theta_0, \theta_1] \times \{Y_1, ..., Y_n\}\).

**Proof.** Let \(e^*, E^*\) be the amounts of education and externalities that solve the relaxed problem. As \(\frac{\partial H}{\partial e_i (\theta)} = 0\), it follows that

\[
\int_{\theta_0}^{\theta_1} \sum_{i=1}^{n} h_i (\phi (\theta, e_i^* (\theta); E^*); E) \phi (\theta) d\theta = \rho.
\]

Substituting in (13),

\[
y_e (\theta, e_i^* (\theta); E) = k - E [y_E (\theta, e_i^*; E^* [e^*])],
\]

which is the equation that implicitly defines \(e^*\). \(\blacksquare\)

Notice that \(P (\theta, 0, E) = 0\) and \(P_Y (\theta, Y_i - \tau_i, E) \geq 1\). Moreover, a unitary increase in \(\tau_i\) and a unitary decrease in \(f_i (\theta)\) leaves \(U_i (\theta)\) unchanged. Hence, it is always possible to choose \(\tau_i\) and \(f_i (\theta)\) such that equation (12g) is satisfied.

We have shown that the profiles of education and consumption solving the relaxed problem are the same as the first-best solution. Since the utility functions are linear in deferred payments \(m_i (\theta)\), it follows that any profile of deferred payments such that the government’s budget constraint is satisfied as an equality achieves the same welfare \(W\). Fix \(\tau_i\) and \(f_i (\theta)\) so that (12g) is satisfied and let \(m_i^* (\theta)\) be given by

\[
m_i^* (\theta) = -t_i^* (\theta) = ke^* (\theta) + c^* - Y_i.
\]

Then, by (12b),

\[
m_i^* (\theta) = ke^* (\theta) - \tau_i - f_i (\theta).
\]

\(^{16}\)We omit the dependence of \(y\) on \(\theta, e_i (\theta),\) and \(E\) for notational simplicity.
Taking the expectation of the equation above, we get
\[ \bar{E}[m^*] = \bar{E}[ke^* - \tau - f], \]
which is the government’s budget constraint (equation 2). Therefore, the first-best welfare is reached in the relaxed problem.

It remains to be shown that the monotonicity condition (6) is satisfied in the solution of the relaxed problem. But, since \( e^* (\theta) \) is increasing in \( \theta \), it follows that this condition is satisfied and, therefore, the solution of the relaxed problem also solves the second-best problem.

**Proof of Proposition 3:** Set \( \hat{k} \) as
\[ \hat{k} = k - \bar{E}[y_E(\theta, e^*, \bar{E}[e^*])]. \tag{14} \]
Substituting in the first-order conditions of the household’s problem (10), we get \( e^*_t (\theta) = e^* (\theta) \).

Set \( \tau \) as
\[ \tau = \bar{E}[e^*] \bar{E}[y_E(\theta, e^*, \bar{E}[e^*])]. \tag{15} \]

Then, it follows that
\[ \bar{E}[t^*] = \bar{E}[Y - \hat{k}e^* - c^* - \tau] = \bar{E}[Y - ke^* - c^*] = \bar{E}[t^*]. \]

Hence, as \( e^*, e^* (\theta) \), and \( \bar{E}[t^*] \) are the same as in the first-best solution (and utility is linear in \( t \)), first-best welfare is achieved. From equation (14),
\[ (k - \hat{k}) e^* (\theta) = \bar{E}[y_E(\theta, e^*, \bar{E}[e^*])] \ e^* (\theta). \]

Applying \( \bar{E} \) to both sides of the above expression yields
\[ \bar{E} \left[ (k - \hat{k}) e^* \right] = \bar{E}[y_E(\theta, e^*, \bar{E}[e^*])] \bar{E}[e^*]. \tag{16} \]

Hence, equations (15) and (16) imply that \( \bar{E} \left[ (k - \hat{k}) e^* - \tau \right] = 0 \). It follows that the government’s budget constraint (8) is satisfied.

**Proof of Proposition 4:** The first-best problem is
\[
\max_{\{c_i(\theta), t_i(\theta)\}} \int_{\theta_0}^{\theta_1} \sum_{i=1}^{n} h_i \left[ u(Y_i - t_i(\theta) - ke_i(\theta)) + v(y(\theta, e_i(\theta); E) + t_i(\theta)) \right] \phi(\theta) \, d\theta
\]
\[
s.t. \quad E = \bar{E}[e].
\]

The first-best allocations are \( \{c_i^{1b}(\theta), e_i^{1b}(\theta), t_i^{1b}(\theta); \theta \in [\theta_0, \theta_1], i = 1, ..., n\} \) that solve the problem above. As in Proposition 2, this problem can be solved introducing an auxiliary variable \( S \) and setting the Hamiltonian:
\[ H = \sum_{i=1}^{n} h_i \left[ u(Y_i - t_i(\theta) - ke_i(\theta)) + v(y(\theta, e_i(\theta); E) + t_i(\theta)) \right] \phi(\theta) + \mu(\theta) \sum_{i=1}^{n} h_i e_i(\theta) \phi(\theta) \]

The first-order conditions are
\[
\mu(\theta) = \mu \text{ constant,} \tag{17a}
\]
\[
u'(Y_i - t_i(\theta) - ke_i(\theta)) = v'(y(\theta, e_i(\theta); E) + t_i(\theta)), \tag{17b}
\]
\[ k = y_e(\theta, e_i(\theta); E) + \frac{\mu}{w'(Y_i - t_i(\theta) - ke_i(\theta))}. \tag{17c} \]
Since \( \frac{\partial W}{\partial e} |_{e=e^{1b}, t=t^{1b}, E^{1b}} = \mu \), it follows that
\[
\int_{\theta_0}^{\theta_1} \sum_{i=1}^{n} h_i v' \left( y \left( \theta, e_i^{1b} \left( \theta \right); E^{1b} \right) + t_i^{1b} \left( \theta \right) \right) y_E \left( \theta, e_i^{1b} \left( \theta \right); E^{1b} \right) \phi \left( \theta \right) d\theta = \mu > 0.
\]

Hence, the first-best profile is characterized by equation (17b) which equates the marginal cost of bequest with its marginal cost for each individual and
\[
k = y_e \left( \theta, e_i \left( \theta \right); E \right) + \bar{E} \left[ u' \left( Y - t - ke \right) y_E \left( \theta, e; E \right) \right] \frac{u' \left( Y_i - t_i \left( \theta \right) - ke_i \left( \theta \right) \right)}{u' \left( Y_i - t_i \left( \theta \right) - ke_i \left( \theta \right) \right)},
\]
which states that the marginal cost of education must be equal to its private marginal benefit plus its social marginal benefit.

Denote by \( c_i \left( \theta \right) \) the consumption associated with \( t_i \left( \theta \right) \) and \( e_i \left( \theta \right) \):
\[
c_i \left( \theta \right) = Y_i - t_i \left( \theta \right) - ke_i \left( \theta \right).
\]

**Lemma 6** The first-best profile of consumption \( c_i^{1b} \left( \theta \right) \) is not constant in \( \theta \).

**Proof.** Suppose \( c_i^{1b} \left( \theta \right) \) is constant in \( \theta \). Then, from (19),
\[
\frac{\partial c_i^{1b} \left( \theta \right)}{\partial t_i^{1b} \left( \theta \right)} = -\frac{1}{k},
\]
and, from equation (17b),
\[
0 = v'' \left( y \left( \theta, e_i^{1b} \left( \theta \right); E^{1b} \right) + t_i^{1b} \left( \theta \right) \right) \times \left( y_e \left( \theta, e_i^{1b} \left( \theta \right); E^{1b} \right) \frac{\partial c_i^{1b} \left( \theta \right)}{\partial t_i^{1b} \left( \theta \right)} + 1 \right).
\]
Therefore, since \( v'' < 0 \), it follows that \( y_e = k \). But this contradicts (18).  

The parents’ problem is:
\[
\max_{\{e,t\}} u \left( Y - t - \tilde{ke} \right) + v \left( y \left( \theta, e; E \right) + t \right)
\]
\[
s.t. \quad E = E \left[ e \right]
\]

The solution to this problem will be such that the marginal cost of bequests is equal to its marginal benefits and the marginal cost of education is equal to its (private) marginal benefits (first-order conditions, which are necessary and sufficient):
\[
u' \left( Y - t - \tilde{ke} \right) = v' \left( y \left( \theta, e; E \right) + t \right),
\]
\[
u' \left( Y - t - \tilde{ke} \right) \hat{k} = v' \left( y \left( \theta, e; E \right) + t \right) y_e \left( \theta, e; E \right).
\]

Therefore, the equilibrium is characterized by the following equations
\[
y_e \left( \theta, e^P; E \right) = \hat{k},
\]
\[
u' \left( Y - t^P - \tilde{ke^P} \right) = v' \left( y \left( \theta, e^P; E \right) + t^P \right).
\]
and the overall level of education \( E = E \left[ e^P \right] \).

From equation (18), in order to implement the first-best amount of education, it is necessary to set \( \hat{k} \) (independent of \( \theta \)) as
\[
\hat{k} = k - \frac{\bar{E} \left[ u' \left( c_i^{1b} \right) y_E \right]}{u' \left( c_i^{1b} \left( \theta \right) \right)}.
\]
But this would only be possible if the first-best amount of consumption did not depend on \( \theta \). However, as shown in Lemma 6, this is not the case. Hence, it follows that when \( v \) is concave, the first-best amount
of welfare cannot be reached through credit loans and Pigouvian taxes. This concludes the first part of the proof.

In order to establish the second part of the proposition, notice that the derivative of the welfare function with respect to $e_i(\theta)$ evaluated at the undistorted education profile is

$$\frac{\partial W}{\partial e_i(\theta)}|_{e^*,t^*} = h_i(\theta) \left[-k + v'(x_i(\theta))y_i(\theta, e^*(\theta); E)\right],$$

where $x_i(\theta) = y(\theta, e^*(\theta); E) + Y_i - ke^*(\theta) - c^*(\theta)$. The sign of the derivative above is the same as the sign of $v'(x_i(\theta))y_i(\theta, e^*(\theta); E) - k$, which is decreasing in $Y_i$. Therefore, a local increase in progressivity increases welfare. Moreover, since wealth is observable, there is no incentive-compatibility constraint associated with $Y_i$. ■

References


