

**THE DESIGN OF OPTIMAL EDUCATION POLICIES  
WHEN INDIVIDUALS DIFFER IN INHERITED WEALTH  
AND ABILITY**

Darío Maldonado

**SERIE DOCUMENTOS DE TRABAJO**

No. 21

Julio 2007

# The design of optimal education policies when individuals differ in inherited wealth and ability

Darío Maldonado \*

Universidad del Rosario, Department of Economics and CeiBA-Complejidad

August 8, 2007

**Abstract:** In this paper I consider the role of optimal education policies in redistribution when individuals differ in two aspects: ability and inherited wealth. I discuss the extent to which the rules that emerge in unidimensional settings apply also in the bidimensional setting considered in this paper. The main conclusion is that, subject to some qualifications, the same type of rules that determine optimal education policies when only ability heterogeneity is considered apply to the case where both parameters of heterogeneity are considered. These rules imply a widening of the education gap between high- and low-ability individuals in second-best with respect to the first-best gap. The qualifications regard the implementation of the optimal allocation of resources to education and not on the way the optimal allocation in first- and in second-best differ.

**Keywords:** Optimal taxation, education, public provision, multidimensional screening

**JEL codes:** H21, H23, H52, I28, J31

## 1 Introduction

Education policy is usually justified in terms of heterogeneity in learning ability and in exogenous wealth. The first parameter affects the efficiency of education investments, the second determines individual resources available for consumption and investment in education. Heterogeneity in ability to learn motivates patterns of resources devoted to education in which the bulk of education resources are allocated to the individuals with the highest ability to learn in order to maximize the returns from these investments. Heterogeneity in exogenous wealth motivates patterns in which, on grounds of equality of opportunity,

---

\* Calle 14 # 4-69, Bogotá, Colombia, dario.maldonadoca@urosario.edu.co

public resources for education are allocated to all individuals regardless of their ability to pay.

In this paper I address the design of education policies when individuals differ in both of these parameters. A proper consideration of this question requires the consideration of other aspects of policy design. The first of them is the fact that education policy never appears as an isolated policy instrument. Almost all modern economies have governments that use other instruments to improve welfare. Probably the most important among these instruments is the tax code. The second element that must be considered is the non-linearity of welfare improving policies which is also a characteristic of most observed policies.

Consequently, in this paper I consider the optimal design of education policies when the government also sets optimally the income tax code and when taxes are non-linear. The main issue in this paper is how education policy depends on information available to a utilitarian government. The main conclusion is that the the second-best allocation of resources for education differs from the first-best one in a very similar way when inherited wealth heterogeneity is considered than when it is not considered. In this respect the conclusion is the same as in Maldonado (2007): the education gap between high- and low-ability individuals is wider in the first- than in the second-best if the education elasticity of wage function is decreasing in ability. This conclusion has the important implication that heterogeneity in exogenous wealth does not affect the way education policy differs in first- and in second-best. Heterogeneity in ability to pay, however, does have implications on the optimal education policies, it affects the way education policy is implemented.

Traditionally, economists recognize that income taxes have affect labor supply, however, it also affects education choices. Consequently one cannot look at education subsidies independently from income taxes.<sup>1</sup> Education taxes or subsidies can be used to reinforce or reduce the effect of income taxes on education. Thus, depending on whether one wants to reinforce or reduce this effect one can tax or subsidize education. It is in this respect that the implementation of the optimal distribution of education subsidies differ when the double heterogeneity is considered from the case when it is not considered. In a two-class economy where individuals differ only in ability, even if ability and education are complements, low-ability individuals may receive an education subsidy.<sup>2</sup> This will happen if the education elasticity of wage decreases with ability. This result holds as well when individuals differ in ability and in exogenous wealth as long as the marginal tax rate on labor income for low-ability individuals is positive. Additionally, in the multidimensional case, high-ability individuals may face a negative marginal income tax (a subsidy). Under

---

1. The effects of income taxes on education choice is quantified by Blöndal, Field, and Girouard (2002).

2. The traditional non-distortion at the top result holds in the two-class economy version of this model. Consequently marginal tax rates on labor income and on education for high-ability individuals are equal to zero

the stated condition on the education elasticity of wage, these individuals will also face a marginal tax on education. Other wise, if low ability individuals face a subsidy on labor market income the implementation of the optimal allocation of education resources implies a marginal education tax for low ability individuals and a marginal education subsidy for high-ability individuals.

Related problems have been analyzed by Arrow (1971), Ulph (1977), Hare and Ulph (1979), Toumala (1986), De Fraja (2002) and Bovenberg and Jacobs (2005). However, either they make important restrictions on functional forms (particularly on the wage function) and on the instruments used by the government (for example linear taxes) or they do not consider the double heterogeneity considered in this paper. This paper, like Maldonado (2007), departs from these assumptions. In terms of the conclusions the most important comparison with that of this paper is with Bovenberg and Jacobs (2005), they argue that the Atkinson-Stiglitz theorem holds for the education policy problem. In other words, education policy has the only role of restoring efficiency undoing the effect of income taxes on education choice. This papers argues that if one allows for a more general wage function<sup>3</sup> the role of education subsidies is to restore efficiency only when individuals only differ in exogenous wealth, otherwise, the Atkinson-Stiglitz efficiency theorem does not hold. In this paper I highlight a different role for education policies. By affecting relative wages education policy can relax the incentive constraints and make redistribution easier.

The rest of the paper is organized as follows. Section 2 sets up the model. Section 3 studies the optimal education policies. Section 4 shows numerical simulations of the model studied in Section 3. The last section concludes.

## 2 The model and the laissez-faire solution

The model I analyze is an extended version of the one used in Maldonado (2007) which is a modified version of the Stiglitz (1982) optimal taxation model. The two main modifications with respect to these two papers is that the model accounts for investments in education and for exogenous wealth heterogeneity. This second parameter of heterogeneity does not affect the benefits from investments in education but do affect the resources available for consumption and education.

Individuals derive utility directly from consumption and labor supply. Labor market income and inherited wealth are used to pay for consumption and for the investments in education. Individuals differ in their ability to learn and their inherited wealth. The first of these two parameters, together with labor supply and investment in education determine labor market income. In the absence of government intervention the decision problem of

---

3. See Maldonado (2007) for a discussion on the existing empirical support of more general wage functions.

an individual of type  $i$  can be expressed as follows:

$$\begin{aligned} \max_{c^i, l^i, q^i} & u(c^i) - v(l^i) \\ \text{s.t. } & q^i + c^i \leq \theta^i + \omega(\phi^i, q^i)l^i \end{aligned} \quad (1)$$

where  $u(c^i) - v(l^i)$  is the utility the individual gets from consumption  $c^i$  and labor supply  $l^i$ ; labor market income results from the combination of labor supply and investments on education  $q^i$ . Individuals differ in ability to learn  $\phi^i$  and ability to pay  $\theta^i$ . Utility is increasing in consumption and decreasing in labor supply and labor market productivity,  $\omega$ , is increasing in both arguments. Utility and labor market productivity are concave in choice variables.

Throughout this paper I will be interested in the analysis of the general tax function  $T(Y^i, q^i)$  when it is optimally set by a utilitarian government. Given the tremendous amount of degrees of freedom that this formula gives to the planner any policy that can be implemented by the government can be understood with its use.<sup>4</sup> When subject to this tax function, the problem of individuals becomes

$$\begin{aligned} \max_{c^i, Y^i, q^i} & u(c^i) - v\left(\frac{Y^i}{\omega(\phi^i, q^i)}\right) \\ \text{s.t. } & q^i + c^i \leq \theta^i + Y^i - T(Y^i, q^i) \end{aligned} \quad (2)$$

where the traditional change of variables in optimal tax literature has been made. The first order conditions of the individual problem yield the following arbitrage conditions:<sup>5</sup>

$$MRS_{cl}^i = \omega(\phi^i, q^i)[1 - T_Y(Y^i, q^i)], \quad (3)$$

$$MRS_{eq}^i = [1 + T_q(Y^i, q^i)] \quad (4)$$

and

$$MRT_{lq}^i = \frac{1}{\omega(\phi^i, q^i)} \frac{1 + T_q(Y^i, q^i)}{1 - T_Y(Y^i, q^i)}. \quad (5)$$

4. To see more clearly that this formula can subsume many different policies note that I am not making any assumptions about the slope of this formula. Consequently this formula includes the case of linear subsidies ( $T_q$  constant and negative) or the case where education policy takes the form of in kind transfers ( $T_q$  taking infinite values in all  $q$  except in the education level that the government wants to achieve); for more details on the relation of non-linear taxation and in-kind transfers see Cremer and Gahvari (1997) and for more details on the tax function see Maldonado (2007)

5. For simplicity, in these expressions, some abuse of notation is been done. It is well known that the optimal non-linear tax functions are not differentiable, however, these expressions allow to see the wedge between the MRS and the MRT introduced by taxes. Introducing the non differentiability at this stage will only subtract to simplicity without any gain.

In this expressions  $MRS$  and  $MRT$  stand for Marginal Rate of Substitution and Marginal Rate of Transformation, as usual. For the following analysis it is important to note from (3)-(5) that the MRT between labor supply and investment in education depends on both marginal tax rates,  $T_Y$  and  $T_q$ . These expressions will also be used to understand the implementation of optimal policies further in this paper.

### 3 The optimal education policy

Before addressing the optimal design of the tax function  $T(Y, q)$  in asymmetric information conditions, a few words about the optimal policies under first-best conditions are in order. The first-best problem of a utilitarian planner is

$$\begin{aligned} \max_{\{T^i, Y^i, q^i\}_i} \quad & \sum_i \pi^i \left[ u(Y^i - T^i - q^i) - v \left( \frac{Y^i}{\omega(\phi^i, q^i)} \right) \right] \\ \text{s.t.} \quad & \sum_i \pi^i T^i \geq 0. \end{aligned} \tag{6}$$

In the first-best the government directly observes all parameters and decision variables. Consequently it can set any tax-transfer scheme and faces no restrictions on the policies it uses to maximize welfare. A utilitarian government would set the tax function  $T(Y^i, q^i)$  so that all individuals receive the same consumption level regardless of their ability or their inherited wealth. However low ability individuals would work less and receive less education than high ability individuals. Moreover, as argued by the second theorem of welfare economics, the government does not need to introduce any distortions on individual choices. This means that marginal tax rates (the derivatives of  $T(Y^i, q^i)$ ) will be equal to zero. The balance between labor supply and education investment for each type of individuals will be set according to the same first order conditions as in the laissez-faire.

Under asymmetries of information things are different. As in Maldonado (2007) the main issue in this paper is the characterization of the function  $T(Y^i, q^i)$  when the government has limited information on individuals' characteristics and choice variables. In this paper I look for the qualifications that the double heterogeneity imposes on the conclusions obtained in my previous paper. As in most of the optimal non-linear tax literature this paper adopts the mechanism design approach to characterize  $T(Y^i, q^i)$ .

In line with the optimal taxation literature (Stiglitz (1987)) I will assume that the government observes labor market income  $Y^i = \omega(\phi^i, q^i) \times l^i$  but it does not observe ability, productivity, or labor supply. I will also assume that the government observes investments in education and consumption level. Inherited wealth is not observed either. The tax function  $T(Y^i, q^i)$ , or the public education system will be characterized for a four class economy where  $\phi^i$  and  $\theta^i$  take only two values. Specifically,  $\phi^i \in \{\phi^L, \phi^H\}$  with

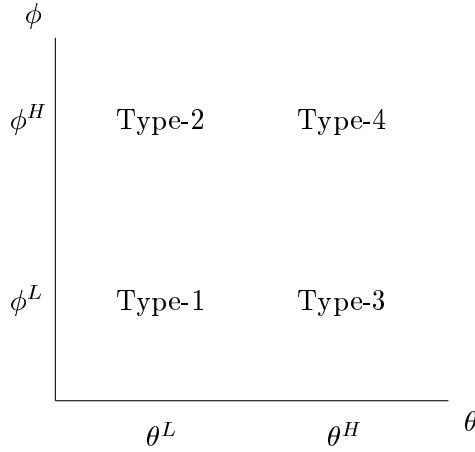


Figure 1: Types of individuals

$\phi^L < \phi^H$  and  $\theta^i \in \{\theta^L, \theta^H\}$  with  $\theta^L < \theta^H$ . Individuals will be labeled according to Figure 1.

As a consequence of asymmetric information the government must rely on direct revelation mechanisms; the Revelation Principle guarantees that there is no loss of generality associated to this choice and that any optimal indirect mechanism (such as a tax function) would be equivalent to the direct mechanism. The key element of this analysis is the incentive constraints that induces truthful revelation. Let  $(R^i, Y^i, q^i)$  be the second-best allocation implemented by the planner, where  $R^i$  is after-tax income income,  $Y^i$  is labor market income and  $q^i$  is the education investment of an individual of type  $i$ . Note that the allocation only includes variables observable to the government. Asymmetric information forces the government to design the mechanism  $(R^i, Y^i, q^i)$  in order to induce individuals to self-select themselves through the choice of the allocation designed for their type. This means the planner will face 12 incentive constraint of the form

$$u(\theta^i + R^i - q^i) - v\left(\frac{Y^i}{\omega(\phi^i, q^i)}\right) \geq u(\theta^k + R^k - q^k) - v\left(\frac{Y^k}{\omega(\phi^i, q^k)}\right). \quad (7)$$

The problem of the planner is problem (6) with the additional incentive constraints, (7). Note that there are twelve incentive constraints restricting the allocations that the government can choose.<sup>6</sup> Letting  $\lambda$  be the Lagrange multiplier of the resource constraint and  $\mu^{ik}$

---

6. The technical difficulties involved in solving multidimensional screening problems are well known. A general analysis of the mechanism design problem in a multidimensional setting has been done by Armstrong

the multiplier of the constraint for type- $i$  individuals not to mimic type- $k$  individuals the Lagrangian of this maximization problem is:

$$\begin{aligned}\Lambda = & \sum_{i=1}^4 \pi^i \left[ u(\theta^i + R^i - q^i) - v\left(\frac{Y^i}{\omega(\phi^i, q^i)}\right) \right] + \lambda \left\{ \sum_{i=1}^4 \pi^i [Y^i - R^i] \right\} \\ & + \sum_{i,k:i \neq k} \mu^{ik} \left[ u(\theta^i + R^i - q^i) - v\left(\frac{Y^i}{\omega(\phi^i, q^i)}\right) \right. \\ & \quad \left. - u(\theta^k + R^k - q^k) + v\left(\frac{Y^k}{\omega(\phi^k, q^k)}\right) \right].\end{aligned}\tag{8}$$

The first order conditions of this problem are:

$$\begin{aligned}[Y^i] : & \pi^i \left[ -v' \left( \frac{Y^i}{\omega(\phi^i, q^i)} \right) \frac{1}{\omega(\phi^i, q^i)} + \lambda \right] \\ & + \sum_{k:k \neq i} \left[ -\mu^{ik} v' \left( \frac{Y^i}{\omega(\phi^i, q^i)} \right) \frac{1}{\omega(\phi^i, q^i)} + \right. \\ & \quad \left. \mu^{ki} v' \left( \frac{Y^i}{\omega(\phi^k, q^i)} \right) \frac{1}{\omega(\phi^k, q^i)} \right] = 0,\end{aligned}\tag{9}$$

$$\begin{aligned}[R^i] : & \pi^i \left[ u'(\theta^i + R^i - q^i) - \lambda \right] \\ & + \sum_{k:k \neq i} \left[ \mu^{ik} u'(\theta^i + R^i - q^i) - \mu^{ki} u'(\theta^k + R^k - q^k) \right] = 0\end{aligned}\tag{10}$$

and Rochet (1999). The results they derive cannot be applied directly here since the optimal tax problem requires the introduction of a budget constraint and I am not assuming quasi-linearity of the utility function. The difficulty in analyzing the optimization problem in this paper is that if one wants to do the full Kuhn-Tucker analysis one would have to compare the optimal solutions of the 144 different optimization problems that emerge for all the possible combinations of binding incentive constraints. The treatment I give to this problem is similar to that in Cremer, Pestieau, and Rochet (2001).

and

$$\begin{aligned}
[q^i] : \quad & \pi^i \left[ -u'(\theta^i + R^i - q^i) + v' \left( \frac{Y^i}{\omega(\phi^i, q^i)} \right) \frac{Y^i \omega_q(\phi^i, q^i)}{(\omega(\phi^i, q^i))^2} \right] \\
& + \sum_{k:k \neq i} \mu^{ik} \left[ -u'(\theta^i + R^i - q^i) \right. \\
& \quad \left. + v' \left( \frac{Y^i}{\omega(\phi^i, q^i)} \right) \frac{Y^i \omega_q(\phi^i, q^i)}{(\omega(\phi^i, q^i))^2} \right] \\
& - \sum_{k:k \neq i} \mu^{ki} \left[ -u'(\theta^k + R^i - q^i) \right. \\
& \quad \left. + v' \left( \frac{Y^i}{\omega(\phi^k, q^i)} \right) \frac{Y^i \omega_q(\phi^k, q^i)}{(\omega(\phi^k, q^i))^2} \right] = 0. \tag{11}
\end{aligned}$$

To rearrange this expressions first define

$$\eta(\phi^i, q^i) \equiv \frac{q \times \omega_q(\phi^i, q^i)}{\omega(\phi^i, q^i)};$$

$\eta(\phi^i, q^i)$  is the education elasticity of the wage function.<sup>7</sup> The expressions for the marginal tax rates are obtained combining and rearranging (9), (10) and (11). Using the notation  $c^i = \theta^i + R^i - q^i$ ,  $c^{ki} = \theta^i + R^k - q^k$ ,  $l^i = Y^i / \omega(\phi^i, q^i)$ , and  $l^{ki} = Y^i / \omega(\phi^k, q^i)$ . These expressions are

$$-MRS_{cl}^i = \omega(\phi^i, q^i) \frac{1 + \sum_{k:k \neq i} \frac{\mu^{ik}}{\pi^i} - \sum_{k:k \neq i} \frac{\mu^{ki} u'(c^{ki})}{\pi^i u'(c^i)}}{1 + \sum_{k:k \neq i} \frac{\mu^{ik}}{\pi^i} - \sum_{k:k \neq i} \frac{\mu^{ki} \omega(\phi^i, q^i) v'(l^{ki})}{\pi^i \omega(\phi^k, q^i) v'(l^i)}}, \tag{12}$$

---

7. Note that one can see  $\omega(\phi^i, q^i)$  as a Mincer equation. Consequently,  $\eta(\phi^i, q^i)$  corresponds to the parameter that accompanies years of education in this type of equations. The education literature has called this parameter the return to education; even if with the specification adopted in this paper it would be correct to call  $\eta(\phi^i, q^i)$  this way, the fact that the term *return to education* has been widely misused I prefer to use for  $\eta(\phi^i, q^i)$  its more direct meaning. I could also call  $\eta(\phi^i, q^i)$  the growth rate of income with education. This point is elaborated by Heckman, Lochner, and Todd (2005).

and

$$MRT_{lq}^i = \frac{1}{\omega(\phi^i, q^i)} \frac{1 + \sum_{k:k \neq i} \frac{\mu^{ik}}{\pi^i} - \sum_{k:k \neq i} \frac{\mu^{ki} \omega(\phi^i, q^i) v'(l^{ki})}{\pi^i \omega(\phi^k, q^i) v'(l^i)}}{1 + \sum_{k:k \neq i} \frac{\mu^{ik}}{\pi^i} - \sum_{k:k \neq i} \frac{\mu^{ki} \eta(\phi^k, q^i) \omega(\phi^i, q^i) v'(l^{ki})}{\pi^i \eta(\phi^i, q^i) \omega(\phi^k, q^i) v'(l^i)}}. \quad (13)$$

Equations (12) and (13) define the marginal tax rates on  $Y^i$  and  $q^i$ . In this case, because of the impossibility to know which are the binding incentive constraints it is difficult to know a priori the signs of the marginal tax rates (some restrictions on marginal tax rates are provided in the appendix). However, it is possible to characterize the overall distortion on education with respect to labor supply.

From equation (13), it can be seen that the key to the analysis of the distortions on education is the form of  $\eta(\phi^i, q^i)$ . The direction of the overall distortion on education with respect to labor supply will be downwards, flat or upwards as  $MRT_{lq}^i \gtrless 1/\omega(\phi^i, q^i)$ . From (13) it is seen that this happens according to the sign of

$$\sum_{k:k \neq i} \frac{\mu^{ki} \omega(\phi^i, q^i) v'(l^{ki})}{\pi^i \omega(\phi^k, q^i) v'(l^i)} \left[ 1 - \frac{\eta(\phi^k, q^i)}{\eta(\phi^i, q^i)} \right]. \quad (14)$$

If this expression is negative the marginal rate of transformation between education and labor is greater than the inverse of the wage rate; if it is positive is smaller. Notice that for each individual  $i$  the ratio  $\eta(\phi^k, q^i)/\eta(\phi^i, q^i)$  can take only two values: if the individual is a low-ability one the ratio can be one or  $\eta(\phi^H, q^i)/\eta(\phi^L, q^i)$ ; if it is a high-ability individual the ratio is one or  $\eta(\phi^L, q^i)/\eta(\phi^H, q^i)$ . According to this, the distortion on education will be related to whether  $\eta(\phi^H, q^i) \gtrless \eta(\phi^L, q^i)$ . This means that, if distorted at all, individuals with different ability parameters will face opposite distortions on education.<sup>8</sup>

The effect of the assumptions about the form of  $\omega(\phi^i, q^i)$  on optimal education policy has already been introduced in Maldonado (2007). The three possible cases analyzed there are

- A1**  $\eta(\phi^i, q)$  increasing in  $\phi^i$  for all  $q$ ,
- A2**  $\eta(\phi^i, q)$  decreasing in  $\phi^i$  for all  $q$ , and
- A3**  $\eta(\phi^i, q)$  is independent of  $\phi^i$  for all  $q$ .

Consider the case in which A1 holds ( $\eta(\phi^i, q)$  is increasing in  $\phi^i$ ). Individuals with low-ability parameter will have  $MRT_{lq}^i \geq 1/\omega(\phi^L, q^i)$  and individuals with high-ability parameter will have  $MRT_{lq}^i \leq 1/\omega(\phi^H, q^i)$ .<sup>9</sup> This means that the education level of low-ability

8. By “distorted at all” I mean that for a type- $i$  individual at least one of the multipliers  $\mu^{ki}$  with  $k$  being such that  $\phi^k \neq \phi^i$  is different from zero.

9. I do not use strict inequality to consider also the case in which that type of individual is not distorted.

individuals will be distorted downwards (if distorted) and that of high-ability individuals upwards (if distorted). If it is A2 that holds the opposite pattern is found. Similarly, when  $\eta(\phi^L, q^i) = \eta(\phi^H, q^i)$  (14) will be equal to zero and there will be no distortion on the education level. The presence of a distortion requires that one of the incentive constraints that links individuals who differ in ability binds. If the only binding constraints are those linking individuals of equal abilities there would be no distortions on education. These results are summarized in Proposition 1 which restates Proposition 1 in Maldonado (2007) to take into account the double heterogeneity of individuals considered in this paper.

**Proposition 1** *Suppose A1 (A2) holds. Consider the low-ability individuals, i.e.  $i = 1, 3$ . If  $\mu^{2i} \neq 0$  or  $\mu^{4i} \neq 0$  then type- $i$  individuals will face a downward (upward) distortion on education with respect to labor supply. Consider the high-ability individuals, i.e.  $i = 2, 4$ . If  $\mu^{1i} \neq 0$  or  $\mu^{3i} \neq 0$ , then type- $i$  individuals will face an upward (downward) distortion on education with respect to labor supply. If A3 holds education will not be distorted with respect to labor supply.*

The results in Proposition 1 depend on at least one of the incentive constraints linking individuals differing in ability binds. If none of the incentive constraints is binding the solution would be identical to that in the first-best. If the only incentive constraints that bind are those linking individuals differing in exogenous wealth, there are no distortions on education with respect to labor supply. If at least one of the incentive constraints that link individuals of different ability binds and  $\frac{\partial}{\partial \phi^i} \eta(\phi^i, q^i) > 0$  (A1) the education gap between high- and low-ability individuals will be widened with respect to the first-best efficient gap. If  $\frac{\partial}{\partial \phi^i} \eta(\phi^i, q^i) < 0$  (A2) holds the same gap will be narrowed. In the case in which  $\frac{\partial}{\partial \phi^i} \eta(\phi^i, q^i) = 0$  (A3) the gap will be kept to its first-best level. Moreover, if A1 holds the optimal education policies are marginally input-regressive but if A2 holds they are marginally input-progressive.<sup>10</sup>

The reason that explains this result is that the government uses distortions in education levels to deter mimicking behavior. If the indifference curves of individuals of different abilities in the  $(l^i, q^i)$  plane are parallel distortions on this plane are useless to separate individuals. This happens if  $\eta(\phi^i, q^i)$  does not depend on  $\phi^i$ . Note that this is not changed by the multidimensional heterogeneity since these indifference curves do not depend on  $\theta^i$ . However, if  $\eta(\phi^i, q^i)$  depends on  $\phi^i$  distorting education choice becomes useful to separate individuals of different types.

Notice the effect of putting together the two parameters of heterogeneity ( $\phi^i$  and  $\theta^i$ ). The bi-dimensional heterogeneity breaks the traditional result of positive marginal tax

---

10. The results in Proposition 1 can not be extended easily to a more than two-types case. In such a case any individual different from the one with the highest or lowest abilities could be mimicked by individuals with higher and lower abilities making the term  $1 - \frac{\eta(\phi^k, q^i)}{\eta(\phi^i, q^i)}$  take positive and negative values for a given individual, and thus making (14) to have an ambiguous sign.

rates for low-ability individuals and zero marginal tax rates for high-ability ones. Now, the marginal income tax rate can have negative or positive signs and high-ability individuals may face non-zero marginal taxes. An analysis of the relations between the marginal tax rate on labor market income and on education is presented in the appendix. It is interesting to note here that positive and negative marginal tax rates on labor income can emerge in this model. Moreover, not always a marginal tax on labor income of a given sign is coupled with a marginal tax rate on education expenditure of the same sign.

A particularly interesting case emerges if A2 holds. In this case the optimal allocation is implemented with a positive tax on labor market income and a marginal subsidy on education for low-ability individuals together with a negative tax on labor income and a positive tax on education for high-ability individuals. This leads to the possibility of having something that can be called over-qualification (or under-activity): low-ability individuals are encouraged to acquire more education than it may be optimal to do in a first-best situation. Nevertheless, they may still be asked for lower labor supply than in a first-best situation.

As argued before, in Maldonado (2007), I show that if individuals differ only in learning ability, i.e. if  $\pi^3 = \pi^4 = 0$ ,  $\pi^1 > 0$  and  $\pi^2 > 0$ , the optimal tax function satisfies zero marginal tax rates on labor and education for high ability individuals, a positive marginal tax rate on labor supply for low ability individuals and that low ability individuals may receive a marginal tax or subsidy on education

A different polar case is that in which individuals differ only in inherited wealth, i.e.  $\pi^2 = \pi^4 = 0$  and  $\pi^1 + \pi^3 = 1$ . In the first-best solution of this problem the planner will set equal levels of labor supply and education and a lump-sum transfer from type-3 to type-1 individuals. Since the planner is pure utilitarian it will want to redistribute from type-3 to type-1 individuals. This means that there is only one binding constraint: the one preventing type-3 individuals to mimic type-1 individuals. Consequently among the Lagrange multipliers of the incentive constraints only  $\mu^{31}$  will differ from zero, and the first-order conditions (9) to (11) yield

$$\begin{aligned} -MRS_{cl}^1 &= \omega(\phi, q^1) \frac{\frac{\mu^{31} u'(c^{31})}{\pi^1} - \frac{u'(c^1)}{\pi^1}}{1 - \frac{\mu^{31}}{\pi^1}}, \\ MRT_{lq}^1 &= \frac{1}{\omega(\phi, q^1)}, \\ -MRS_{cl}^3 &= \omega(\phi, q^3) \end{aligned}$$

and

$$MRT_{lq}^3 = \frac{1}{\omega(\phi, q^3)}.$$

The above expressions imply undistorted choices for type-3 individuals and an upward distortion on labor supply for type-1 individuals. Consequently the marginal rate on labor income is zero for type-3 individuals and negative for type-1 individuals. Simultaneously, there will be a positive marginal tax rate on education equal to the negative of the marginal income tax so there will be no distortions on education. Thus, there is no effect of the structure of the returns to labor on optimal education policy. In this case we have that education should be taxed for high-ability individuals and that the only role for this tax is to restore efficiency in the relation between labor supply and education. The reason for this is that, since individuals differ only in exogenous wealth, indifference curves in the  $(l^i, q^i)$  plane are parallel, thus, there is no use for distortions on education with respect to labor supply.

Recapping, the main result in this section is that the education gap between high- and low-ability individuals will be widened or narrowed depending on which of the two possibilities helps deterring mimicking behavior. If the education elasticity of the wage function is increasing in ability, mimicking behavior will be deterred by a widening of the education gap. In the case the education elasticity of the wage function is decreasing in ability it will be a narrowing of the education gap that will deter mimicking behavior. The effect of this tax on individual education levels may possibly be reversed using the marginal tax on education. This is true if the education elasticity of wage decreases with ability. This rule implies that education of low-ability individuals must be subsidized in some cases, particularly if they face a positive marginal income tax.

## 4 Numerical simulations

In this section I show some numerical simulations for the problem stated in (8). The simulations have three objectives: they shed light to the role of the size-differences between the parameters, they show some specific cases where it can be seen which of the incentive constraints are binding, and they show the workings of the qualitative features of the model previously discussed.

All the simulations share some features, these are the proportion of each type of individuals and the utility function of consumers. I assume that there are equal numbers of each type of individual in the economy ( $\pi^i = 0.25$  for all  $i \in \{1, 2, 3, 4\}$ ) and that the utility function is

$$u(c) - v(l) = \sqrt{c} - l^2.$$

I show simulations for three different types of wage functions that were chosen according to the behavior of their education elasticity with respect to the ability parameter. According to the theoretical results in the previous sections, this amounts to whether the education elasticity of the wage function increases, decreases or does not change with the ability parameter.

Table 1 shows results for the wage function

$$\omega(\phi, q) = 10q^\phi$$

which satisfies A1, i.e.  $\eta(\phi, q)$  is increasing in  $\phi$ . Table 2 shows results for the wage function

$$\omega(\phi, q) = \phi + q - q^2$$

which satisfies A2, i.e  $\eta(\phi, q)$  is decreasing in  $\phi$ . Table 3 shows the results for the function

$$\omega(\phi, q) = 10 \log(q)$$

which satisfies A3. For each of the wage functions I show two different simulations, in the first the difference between the exogenous incomes are “small” and in the second one the difference is “big”. For each set of simulations sharing the same wage function the only thing that changes is the value of  $\theta^H$ . In each case I report consumption, labor supply, and education level for the first- and second-best, additionally I report the marginal tax rates, the compound effect of both marginal tax rates on education and the binding incentive constraints for the second-best.<sup>11</sup>

Some of the standard features of the optimal income tax problem can be seen in these tables. Most important, because of the separable utility function in the first-best consumption is equalized among all types of individuals. However, labor supply and education are higher for high-ability individuals than for low-ability ones. This causes first-best allocations not to be implementable since utility decreases with ability and is independent of exogenous wealth. Moreover, utility in the first-best is independent of exogenous income, which also contributes to the first-best not being implementable. As a consequence, in the second-best, consumption must depend on ability and exogenous wealth and distortions appear in the picture.

In the second-best marginal taxes stop being equal to zero. The tables report the binding incentive constraints that generate this fact. These patterns of binding incentive constraints show that the intuitive result of only “downward” incentive constraints and no bunching that holds in one dimension cases no longer holds when individuals differ in more than one parameter. In all the simulations there is at least one “upward” incentive constraint that binds.<sup>12</sup> Table 3 shows two cases where there is bunching, the upward and downward incentive constraints that link two given types of individuals bind simultaneously.

11. Quantitative comparisons of the results of sets of simulations with different wage functions should not be made here. This explains why the numerical values for the parameters do not coincide, there is no gain from choosing the same values and, due to computational constraints, there could be considerable costs in terms of time need to find interior solutions for each example.

12. Upward incentive constraints are those that avoid an individual with low ability or low exogenous wealth to mimic an individual with high ability or high exogenous wealth.

Table 1: Education elasticity increasing in  $\phi$ ,  $\omega(\phi, q) = 10q^\phi$

Parameters: $\theta^L = 5, \theta^H = 10, \phi^L = \frac{1}{4}, \phi^H = \frac{1}{3}$ (small difference in $\theta^i$ )								
Type	First-best			Second-best				
	$c$	$l$	$q$	$c$	$l$	$q$	$T_Y$	$T_q$
1	18.19	0.710	2.148	13.60	0.861	2.605	0.0004	0.049
2	18.19	1.146	7.462	18.79	1.109	7.107	0.0000	0.000
3	18.19	0.710	2.148	17.20	0.738	2.265	0.0010	-0.001
4	18.19	1.146	7.462	21.33	0.968	5.793	0.0050	-0.005
Binding incentive constraints: 1 → 3, 2 → 4 and 4 → 1								

Parameters: $\theta^L = 5, \theta^H = 30, \phi^L = \frac{1}{4}, \phi^H = \frac{1}{3}$ (big difference in $\theta^i$ )								
Type	First-best			Second-best				
	$c$	$l$	$q$	$c$	$l$	$q$	$T_Y$	$T_q$
1	24.62	0.566	1.587	14.56	0.924	2.510	-0.120	0.297
2	24.62	0.846	4.738	20.41	1.087	6.896	-0.032	0.032
3	24.62	0.566	1.587	32.20	0.462	1.213	0.000	0.000
4	24.62	0.846	4.738	35.79	0.652	3.426	-0.034	-0.011
Binding incentive constraints: 3 → 4, 4 → 1 and 4 → 2								

Table 2: Education elasticity decreasing in  $\phi$ ,  $\omega(\phi, q) = \phi + q - q^2$

Parameters: $\theta^L = 30, \theta^H = 60, \phi^L = 100, \phi^H = 200$ (small difference in $\theta^i$ )								
Type	First-best			Second-best				
	$c$	$l$	$q$	$c$	$l$	$q$	$T_Y$	$T_q$
1	241.6	1.611	0.190	204.2	1.708	0.211	0.026	-0.038
2	241.6	1.933	0.241	253.1	1.888	0.235	0.000	0.000
3	241.6	1.611	0.190	229.0	1.653	0.197	0.001	-0.001
4	241.6	1.933	0.241	274.2	1.812	0.224	0.002	-0.002
Binding incentive constraints: 1 → 3, 2 → 4 and 4 → 1								

Parameters: $\theta^L = 30, \theta^H = 300, \phi^L = 100, \phi^H = 200$ (Big difference in $\theta^i$ )								
Type	First-best			Second-best				
	$c$	$l$	$q$	$c$	$l$	$q$	$T_Y$	$T_q$
1	332.5	1.373	0.136	207.3	1.769	0.232	-0.017	-0.036
2	332.5	1.647	0.196	269.5	1.892	0.230	-0.034	0.057
3	332.5	1.373	0.136	410.7	1.235	0.095	0.000	0.000
4	332.5	1.647	0.196	474.8	1.455	0.143	-0.056	0.096
Binding incentive constraints: 1 → 2, 3 → 4 and 4 → 1								

Table 3: Education elasticity constant in  $\phi$ ,  $\omega(\phi, q) = \phi \log(q)$ 

Parameters: $\theta^L = 30, \theta^H = 60, \phi^L = 100, \phi^H = 120$ (small difference in $\theta^i$ )								
Type	First-best			Second-best				
	$c$	$l$	$q$	$c$	$l$	$q$	$T_Y$	$T_q$
1	2070	3.163	316.3	1733	3.407	340.7	0.0270	-0.0270
2	2070	4.084	490.1	2320	3.815	457.9	0.0000	0.0000
3	2070	3.163	316.3	1763	3.407	340.7	0.0190	-0.0190
4	2070	4.084	490.1	2337	3.797	455.7	0.0004	-0.0004
Binding incentive constraints: $1 \rightarrow 3, 3 \rightarrow 4, 3 \rightarrow 1, 4 \rightarrow 1$ and $4 \rightarrow 3$								

Parameters: $\theta^L = 30, \theta^H = 2100, \phi^L = 100, \phi^H = 120$ (small difference in $\theta^i$ )								
Type	First-best			Second-best				
	$c$	$l$	$q$	$c$	$l$	$q$	$T_Y$	$T_q$
1	2725	2.677	267.7	1752	3.527	352.7	-0.007	0.007
2	2725	3.465	415.9	2447	3.744	449.3	-0.011	0.011
3	2725	2.677	267.7	3085	2.482	248.2	0.000	0.000
4	2725	3.465	415.9	3822	2.939	352.7	-0.033	0.033
Binding incentive constraints: $1 \rightarrow 4, 3 \rightarrow 1, 3 \rightarrow 4, 4 \rightarrow 1$ and $4 \rightarrow 2$								

With respect to the marginal taxes the main regularity across the tables is that there is always a type of individuals who is not distorted. It turns out that this is either type-2 or type-3 individual. This was expected since these are the type of individuals with bigger and lower marginal rates of substitution. Moreover, when facing a non-zero tax, it is negative for type-2 individuals and positive for type-3 individuals. This is in line with the one dimensional cases. In the case where individuals differ only in ability, low-ability individuals face a positive marginal income tax and in the case where they differ only in exogenous wealth low-exogenous wealth individuals face negative marginal income taxes.

Education marginal taxes are in line with what has been argued all along the paper and particularly with the discussion in the appendix. Low-ability individuals may receive subsidies on their education expenditure. In all cases at least one low-ability individual (type-1 or type-3) has  $T_q < 0$ . If A2 holds (Table 2), whenever a low-ability individual faces a positive marginal income tax he also faces a marginal subsidy on education. If A1 holds, a positive marginal income tax can go together with a subsidy or a tax on education (respectively type-1 and type-3 in top panel of Table 1). When A3 holds (Table 3) marginal taxes on education are always equal to the marginal tax on labor income.

The education levels and the education gap between high- and low-ability individuals are also according to the results in the paper. The last column of Table 2 shows that,

under assumption A2, type-1 individuals face an upward distortion on education and in the bottom panel type-2 and type-4 individuals face downward distortions on education. In Table 3 the opposite is found. It is only in Table 3, where  $\eta(\phi, q)$  does not depend on  $\phi$ , that there are no distortions on education with respect to labor supply. Finally note that the simulation exercises constitute examples not only of policies that can be input-regressive or input-progressive in marginal terms but also in absolute terms (in the sense of Arrow, 1971). In Table 1 type-1 individuals have more education than type-3 and type-4 individuals. In Table 3 type-1 individuals have exactly the same level of education than type-4 individuals.

## 5 Concluding comments

In this paper I have discussed the design of optimal education policies when income taxation is also designed optimally and when individuals differ in learning ability and in inherited wealth. I have treated the problem carefully according to insights given by the traditional theories of welfare economics and of optimal taxation. This meant the introduction of cash transfers besides education policy and the need of asymmetric information. Also non-separability of education and labor supply is needed.

In a model where the government observes directly the individual education level, I have shown that the distortion on the education level may not have the same direction as the distortion on labor supply. Education subsidies may go together with income taxes. But this is just one of the possible cases that can emerge in this model. I have shown the condition under which it emerges, namely, if the education elasticity of the wage function is decreasing in ability, and it is optimal to set a positive marginal tax on low-ability individuals these individuals will also face a subsidy on education.

The main purpose of the paper was to contrast the optimal policies that emerge when the two dimensions of heterogeneity are considered with those that emerge when only one dimension is considered. The main conclusion is that the result of the difference of the education gap in first and second-best is very similar when only heterogeneity in ability and the double heterogeneity are considered. However, important differences between the way to implement this education gap emerge between the two models. With respect to the model where the only dimension of heterogeneity is inherited wealth, the model with the double heterogeneity shows important differences; in the first the education gap is kept to its first-best level (conditional on the levels of labor supply that do differ in both situations). The consideration of the model where only inherited wealth heterogeneity is introduced highlights the use of education policy in the models where heterogeneity in ability is introduced. In the first the only use for education policy is to restore the inefficiency brought by income taxes on the labor supply-education margin; in the second one education policy has a real role in redistribution.

## Appendices

### 1 The marginal tax rates

The specific signs of the marginal tax rates on labor income and on education expenses are ambiguous in this model because of the multidimensionality assumption that makes difficult to know which of the incentive constraints are binding. However, some restrictions on the possible marginal tax rates can be obtained.

From (12) and (13), the two conditions to understand these relations are:

$$\begin{aligned} \text{C1} \quad & \sum_{k:k \neq i} \frac{\mu^{ki}}{\pi^i} \frac{u'(c^{ki})}{u'(c^i)} - \sum_{k:k \neq i} \frac{\mu^{ki}}{\pi^i} \frac{\omega(\phi^i, q^i)}{\omega(\phi^k, q^i)} \frac{v'(l^{ki})}{v'(l^i)} \geq 0 \quad \text{and} \\ \text{C2} \quad & \sum_{k:k \neq i} \frac{\mu^{ki}}{\pi^i} \frac{u'(c^{ki})}{u'(c^i)} - \sum_{k:k \neq i} \frac{\mu^{ki}}{\pi^i} \frac{\eta(\phi^k, q^i)}{\eta(\phi^i, q^i)} \frac{\omega(\phi^i, q^i)}{\omega(\phi^k, q^i)} \frac{v'(l^{ki})}{v'(l^i)} \geq 0. \end{aligned}$$

Conditions C1 and C2 define four cases under which different configurations of marginal tax rates can appear. Notice that the two conditions differ in the presence of the relative education elasticity of the wage function between the mimicked individual  $i$  and the mimicker  $k$  in the second of them. This makes it possible for the left hand side of both conditions to have different signs.

From equation (3) it can be seen that  $T_Y(Y, q) \geq 0$  depending in whether  $-MRS_{cl}^i \leq \omega(\phi^i, q^i)$ . This means that the marginal tax rate on labor supply faced by individual  $i$  will be positive if the left hand side of C1 is strictly greater than zero, it will be negative if it is strictly less than zero. Similarly, (3) and (5) imply  $T_q(Y, q) \leq 0$  depending on whether  $-MRS_{cl}^i \times MRT_{lq}^i \leq 1$ . Which means that marginal tax rates on education will be negative if the left hand side of C2 is strictly greater than zero, it will be positive if it is strictly smaller than zero.

Generally the marginal tax rates faced by a type- $i$  individual will be zero when the corresponding multipliers  $\mu^{ki}$  (i.e., the multiplier of the constraint that prevents type- $k$  individual to mimic a type- $i$  individual) are all zero. Therefore if one of the tax rates is zero the other one will also be zero. Only in very special cases (depending on  $\eta(\phi, q)$ ) only one of the marginal tax rates will be different from zero.

Table 4 shows the possible pattern of marginal tax rates in each of these cases, it also labels the cases for further reference. In the table  $CI(>)$ ,  $CI(<)$ , refers to the left hand side of conditions C1 or C2 being strictly greater or strictly smaller than zero respectively.

From Proposition 1 and equation (5) it can be seen that if  $\frac{\partial}{\partial \phi} \eta(\phi, q) > 0$ ,  $T_q^i < -T_Y^i$  for low-ability individuals and  $T_q^i > -T_Y^i$  for high-ability individuals. The opposite pattern

	C1(>)	C1(<)
C2(>)	Case 1: $T_Y^i > 0, T_q^i < 0$	Case 2: $T_Y^i < 0, T_q^i < 0$
C2(<)	Case 3: $T_Y^i > 0, T_q^i > 0$	Case 4: $T_Y^i < 0, T_q^i > 0$

Table 4: Possible combinations of marginal tax rates

	$\frac{\eta(\phi^k, q^i)}{\eta(\phi^i, q^i)} < 1$	$\frac{\eta(\phi^k, q^i)}{\eta(\phi^i, q^i)} > 1$	$\frac{\eta(\phi^k, q^i)}{\eta(\phi^i, q^i)} = 1$
$T_Y^i < 0$	$T_q^i \leq 0$	$T_q^i > 0$	$T_q^i = -T_Y^i$
$T_Y^i > 0$	$T_q^i < 0$	$T_q^i \leq 0$	$T_q^i = -T_Y^i$
$T_Y^i = 0$	$T_q^i = 0$	$T_q^i = 0$	$T_q^i = 0$

Table 5: The relation between the marginal tax rates

will be found if  $\frac{\partial}{\partial \phi} \eta(\phi, q) < 0$ . Assuming  $\frac{\partial}{\partial \phi} \eta(\phi, q) = 0$  it will always be that  $T_q^i = -T_Y^i$ . Thus, depending on which of these assumptions hold the possible cases that will happen from Table 4 are restricted. Table 5 shows the resulting possible relations between marginal tax rates under the different assumptions on  $\eta(\phi, q)$ .

Which of these possibilities will be the case is difficult to know a priori. This depends on which of the multipliers of the incentive constraints are different from zero, on their size and on the size of the differences between the marginal rates of substitution between labor and consumption of mimickers and mimicked. The only thing that can be said is that, if different from zero, the marginal tax rate on labor income of type-2 individuals will be non-positive and for type-3 individuals will be non-negative since for the former the marginal rate of substitution will always be greater or equal to that of its mimickers whereas for the latter it will always be less or equal.

## References

- Armstrong, M. and J. C. Rochet (1999). Multi-dimensional screening: a user's guide. *European Economic Review* 43, 959–979.
- Arrow, K. J. (1971). A utilitarian approach to the concept of equality in public expenditures. *Quarterly Journal of Economics* 85, 409–415.
- Blöndal, S., S. Field, and N. Girouard (2002). Investment in human capital through post-compulsory education and training: selected efficiency and equity aspects. OECD, Economics Department Working Paper No. 333.
- Bovenberg, A. L. and B. Jacobs (2005). Redistribution and education subsidies. *Journal of Public Economics* 89, 2005–2035.
- Cremer, H. and F. Gahvari (1997). In-kind transfers, self-selection and optimal tax policy. *European Economic Review* 41, 97–114.
- Cremer, H., P. Pestieau, and J. C. Rochet (2001). Direct versus indirect taxation: the design of the tax structure revisited. *International Economic Review* 42, 781–799.
- De Fraja, G. (2002). The design of optimal education policies. *Review of Economic Studies* 69, 437–466.
- Hare, P. G. and D. T. Ulph (1979). On education and distribution. *Journal of Political Economy* 87, S193–S212.
- Heckman, J., L. Lochner, and P. Todd (2005). Earnings functions, rates of return and treatment effects: The mincer equation and beyond. Mimeo.
- Maldonado, D. (2007). Education policies and optimal taxation. *International Tax and Public Finance*, forthcoming.
- Stiglitz, J. E. (1982). Self-selection and Pareto efficient taxation. *Journal of Public Economics* 17, 213–240.
- Stiglitz, J. E. (1987). Pareto efficient and optimal taxation and the new new welfare economics. In A. J. Auerbach and M. Feldstein (Eds.), *Handbook of Public Economics*. Amsterdam: North-Holland.
- Toumala, M. (1986). Optimal income taxation and educational decisions. *Journal of Public Economics* 30, 183–198.
- Ulph, D. T. (1977). On the optimal distribution of income and educational expenditure. *Journal of Public Economics* 8, 341–356.