Unequal wages for equal utilities[1]

Helmuth Cremer[2], Pierre Pestieau[3] and Maria Racionero [4]

December 2007, revised June 2009

[1] This paper has been presented at the TSE Public Economics Workshop, the IZA Workshop "The Economics of Labor Income Taxation" and in a number of seminars. We thank all the participants for their comments. We are also grateful to Robin Boadway, Georges Casamatta, Firouz Gavahri, Catarina Goulao, Martin Hellwig, Guy Laroque, Dario Maldonado, Bernard Salanié and Worlfram Richter for helpful remarks and suggestions.

[2] GREMAQ and IDEI, University of Toulouse, 21 Allée de Brienne, 31000 Toulouse, France; e-mail: helmut@cict.fr.

[3] CREPP, University of Liège, CORE, PSE and CEPR.

[4] Australian National University, Canberra.

Abstract

When educational policy is supplemented by a redistributive income tax, and when individuals differ in their ability to benefit from education, the optimal policy is typically rather regressive. Resources are concentrated on the most able individuals in order to get a "cake" as big as possible to share among individuals through income taxation. In this paper we put forward another reason to push for regressive education. It is not linked to heterogeneity in innate ability to benefit from education but to pervasive non-convexities that arise in the optimal income tax problem when individual productivities are endogenous. For simplicity we assume a linear education technology and a given total education budget. To give the equal wage outcome the best chance to emerge, we also assume that individuals have identical learning abilities. Nevertheless, it turns out that in the first-best wage inequality is always preferable to wage equality. Even more surprisingly, this conclusion remains valid in the second-best when the feasible degree of wage differentiation is sufficiently large. This is in spite of the fact that wage equalization would eliminate any need for distortionary income taxation.

JEL:H20, I20, D63

Keywords: education policy, optimal income taxation, equal opportunity.

1 Introduction

One of the main rationales for public involvement in education is redistribution.[1] In the literature, the design of an optimal educational policy is usually combined with that of a redistributive income tax. In that framework education is not the only channel of redistribution; there is also in a second stage an income tax that can alter social welfare. The government's problem is first to determine the amount and distribution of public education and then to design the structure of the income tax. The standard approach is to assume a population of individuals who differ in their ability to benefit from education. This heterogeneity typically implies a rather regressive distribution of public education: resources are concentrated on the most able individuals in order to get a "cake" as big as possible to share among individuals through income taxation; see e.g., Bruno (1976), Hare and Ulph (1979), Ulph (1977) and more recently, De Fraja (2002), Bovenberg and Jacobs (2005) and Maldonado (2008). Note that this "perverse" distribution effect just mentioned is mitigated when we introduce decreasing returns of educational spending.[2] Naturally, if we restrict the exercise to the first stage, the solution is also different and tends to be less regressive. This is shown for instance by Arrow (1971), who studies the optimal distribution of a given amount of public expenditure among individuals differing in their learning ability without accounting for the possibility of subsequent income redistribution.

In this paper we put forward another reason that may push for regressive education. It is not linked to heterogeneity in innate ability to benefit from education but to "non-convexities" that may arise in the optimal income tax problem when individual productivities are endogenous (depend on education). To make it as clear as possible, we assume away differences in learning abilities, decreasing returns of education and we suppose that the total amount of public expenditure on education is given. We shall discuss the implications of those three assumptions in the concluding section.

This point that unequal wages are to be preferred over equal wages has been ne-

[1] Poterba (1995). Other reasons include externalities and growth; see for instance Glomm and Ravikumar (1992) and Bénabou (2002).

[2] Bovenberg and Jacobs (2005) and Maldonado (2008).

glected so far, at least in the public finance literature.[3] It is however related to "specialization" results that have been derived in family economics. For instance, Becker (1985) shows that educational investment produces increasing returns so that there is a strong incentive for a division of labor even among basically identical persons.[4] Becker uses this argument to explain the traditionally observed division of labor in households and families (with one spouse specializing in market work while the other one specializes in domestic activities). Our analysis differs from his in two important respects. First our scope is not the cooperative family but the whole society. Second, we are concerned not only with first-best redistribution that is implicit in Becker's work, but also by second-best optimal solutions.

We show that when the general redistributive framework includes an education policy that determines the distribution of wage and an optimal income tax *à la Mirrlees*, then the most unequal distribution of wages happens to be desirable from the standpoint of social welfare maximization. In other words, in our setting, it is "more efficient" to redistribute income rather than to equalize wages.

This rather surprising and to some extent counter-intuitive result has to be interpreted with great care. It does have practical relevance because the effects that are at work in our simple setting (and the push for a corner solution) are also going to be relevant in much more complex (and realistic) frameworks. Consequently, the advocates of a progressive education policy have to show that there are benefits to wage equalization that outweigh the non-convexities we are putting forward. We shall revisit this issue in the concluding section. In the meantime let us just point out that our result is clearly at odds with the literature on equal opportunity, that would recommend wage equality.[5] In fact, we here have a trade-off between equal opportunity and utilitarian welfare maximization. It is also at odds with the observation that in many countries public spending in education, but also in health is more redistributive than income taxation. If such an

[3] Note however that it is implicitly acknowledged by Bovenberg and Jacobs (2005) who avoid it by making suitable assumptions on the education technology. There is also a recent paper by Brett and Weymark (2008) who study the impact of changing skill levels on everyone's optimal consumptions and optimal marginal tax rates.

[4] See also Becker (1991) chapter 2.

[5] Romer (1998), Athanasiou (2007).

observation is correct then our results suggest that current income taxation is far from being optimal.

To keep our arguments as simple as possible, we use the two-ability version of Mirrlees (1972)'s model by Stiglitz (1982). Stiglitz studies the optimal income tax schedule where there is a fraction n_i of workers with productivity w_i $(i = 1, 2)$. Traditionally, both (n_1, n_2) and (w_1, w_2) are given. The only restriction is that $w_2 > w_1$. Recently there have been some attempts to study the incidence of a change in n_1/n_2 over the optimal tax schedule and the extent of redistribution.[6] Yet there exist no study on the incidence of a change in w_1 and w_2.

Suppose now that through its educational policy the government can choose the vector (w_1, w_2) subject to some constraint $n_1w_1 + n_2w_2 = E$ (a constant to be specified below). This amounts to assuming that wages are determined by education expenditures through a linear technology and that the total education budget is given. The natural question to ask is what is the most desirable distribution of (w_1, w_2) given that whatever the vector (w_1, w_2) chosen there will be some redistributive taxation policy. At first glance, one is tempted to lean towards equal wages, $w_1 = w_2$. This appears to be the most "natural" solution at least with a concave social welfare function (when society cares for equality). Furthermore with equal wages, there is no need to resort to distortionary taxation to redistribute income. Consequently, a first-best policy is attainable.

We show in this paper that this intuitive argument is wrong and that the optimal distribution of abilities is the most unequal one. In the first-best, one easily shows that with quasi-linear preferences individual utility is a *convex* function of w so that the equal wages solution is a local social welfare minimum. By endowing one type of individual with maximum ability we get more aggregate utility than by endowing both types with average ability. Second-order conditions are more complicated with general preferences but we provide a simple argument to show that wage differentiation always dominates wages equalization.

All these results are obtained in a first-best context, when there is no asymmetric

[6] Boadway and Pestieau (2007).

information that hinders income distribution. However, this is not the most relevant setting to deal with the design of tax policy. The central element in the theory of optimal taxation is information. Lack of public information on personal characteristics prevents the government from levying optimum lump sum taxes and forces it to impose taxes on income. Now, when wages are not observable in the second stage one would expect the case for equal wages to be considerably strengthened. This is because when wages are equalized, distortionary redistribution of incomes is not needed. Conversely, under maximal wage differentiation efficiency losses of income redistribution are incurred. Consequently, when wage differentiation yields the bigger "cake", it is not clear if it remains bigger once the incomes have been redistributed. Quite surprisingly, though, it turns out that it does remain bigger! Specifically, we show that in a second-best setting, the most unequal distribution of productivity remains desirable if it implies zero productivity for all but one type. It is only when there is a positive and sufficiently large "productivity floor" (implying an exogenous limit to wage differentiation), that an equal distribution of productivity *may* become desirable in a second-best setting.

The rest of this paper is organized as follows. After introducing the basic model, we first deal with the first-best setting (Section 3) and then with the second-best one (Section 4). In Section 5, we present a numerical example.

2 The economy

In the line of Stiglitz (1982) we consider two types of individuals $i = 1, 2$ with relative size $n_i > 0$ (so that $n_1 + n_2 = 1$) and identical utilities

$$u(x_i, \ell_i), \tag{1}$$

where x_i is consumption of a numeraire good and ℓ_i labor supply.[7] Utility is increasing in x, decreasing in ℓ and strictly quasi-concave. Productivities (or wages) w_i are *endogenous* and satisfy what we call the educational technology. We characterize this

[7] It will become clear below that the two types are purely fictitious. We talk about transfers of earning abilities (wages) between individuals to make the link with the Stiglitz model. This is effectively equivalent to a setting where individuals are identical *ex ante*.

technology through Γ, the set of feasible wages that is defined by

$$\Gamma = \{(w_1, w_2) | \, n_1 w_1 + n_2 w_2 = 1 \text{ and } w_i \geq \underline{w}\} \text{ with } 1 \geq \underline{w} \geq 0. \tag{2}$$

In words we have a linear technology according to which earnings ability can be "transferred" between types on a one by one basis. To understand this formulation assume first that $\underline{w} = 0$. In that case the education technology has two extreme cases. One is the complete equalization of wages $w_1 = w_2 = 1$ and the other is the maximum wage inequality with say $w_1 = 0$ and $w_2 = 2/n_2$, where all educational resources are devoted to type 2 individuals. The role of $\underline{w}$ is to limit the scope of wages inequalities that is feasible.

We specify the education technology in terms of wages to make the link to the optimal taxation literature. The formalization we use may appear somewhat unusual, but can be derived from a more conventional specification of the education technologies as it is often found in the literature. To see this, we can specify individual wages following Ulph (1977) as $w_i = \phi_i(\theta_i, e_i)$, where θ_i is (exogenous) ability to benefit from educations and e_i is education expenditure. This is exactly equivalent to the technology we use if *(i)* individuals are identical ex ante (θ_i is the same for all), *(ii)* ϕ is linear in e_i (for instance $w_i = Ae_i$ where A is some constant) and *(iii)* the total budget for education expenditures is fixed at some level E (so that $n_1 e_1 + n_2 e_2 = E$).[8]

To evaluate welfare, we adopt a symmetric social criterion:

$$W = \sum n_i \psi(u_i),$$

where $\psi(\cdot)$ is concave. An extreme form of $\psi(\cdot)$ yields the Rawlsian criterion:

$$W = \min[u_1, u_2].$$

For the remainder of the paper we assume for notational convenience and without loss of generality that $n_1 = n_2 = 1/2$.

[8]For instance, we would have $w_1 = \theta_1 e_1^\alpha$ and $w_2 = \theta_2 e_2^\alpha$ where $\theta_1 \neq \theta_2$ reflects different learning ability and $\alpha \leq 1$ implies decreasing returns. In our paper $e_1 + e_2 = 2$, $\alpha = 1$ and $\theta_1 = \theta_2 = 1$. To be more precise, this specification yields (2) with $\underline{w} = 0$. To generate a positive level of $\underline{w}$ we would have to add a constraint specifying a minimum level of e_i.

3 First-best

3.1 General problem: statement

Assume full information and free use of lump-sum transfers. The optimal allocation is then determined by solving:

$$\max_{\ell_i, w_i, T_i} \quad W = \Psi[u(w_1\ell_1 - T_1, \ell_1)] + \Psi[u(w_2\ell_2 - T_2, \ell_2)], \tag{3}$$

$$\text{s.t.} \quad w_1 + w_2 = 2, \tag{4}$$

$$T_1 + T_2 = 0 \tag{5}$$

$$w_i \geq \underline{w}. \tag{6}$$

This is similar to the standard first-best benchmark considered by Stiglitz (1982), except that we have endogenous wages. We concentrate on the first-order condition with respect to w_i. Substituting for $w_2 = 2 - w_1$ and differentiating yields

$$\frac{\partial W}{\partial w_1} = (\ell_1 - \ell_2)\lambda, \tag{7}$$

where $\lambda > 0$ is the Lagrange multiplier of the resource constraint (5). This FOC is always satisfied for $w_1 = w_2 = 1$ (along with $\ell_1 = \ell_2$) so that one might be tempted to think that the solution implies equalization of wages. However, we cannot simply assume the SOC to be satisfied here. The SOCs condition for the general problem (with 6 decision variables) is complicated but the potential for nonconvexities can be illustrated by considering a special case.

3.2 Quasi-linear illustration

Assume quasi-linear preferences specified by $u_i(x_i, \ell_i) = x_i - v(\ell_i)$. With such preferences it is plain that Pareto efficiency requires the maximization of "surplus" given by

$$S = w_1\ell_1 + w_2\ell_2 - v(\ell_1) - v(\ell_2).$$

Furthermore individual labor supplies ℓ_i can be expressed solely as a function of w_i (there is no income effect). Substituting $w_2 = 2 - w_1$ we then reduce the problem to a single dimension and write $\widetilde{S}(w_1)$. Differentiating yields

$$\frac{d\widetilde{S}}{dw_1} = \ell_1(w_1) - \ell_2(2 - w_1), \tag{8}$$

$$\frac{d^2\widetilde{S}}{dw_1^2} = \frac{d\ell_1}{dw_1} + \frac{d\ell_2}{dw_2} > 0, \tag{9}$$

which shows that $\widetilde{S}(w_1)$ is a convex function. Consequently, with quasi-linear preferences equalization of wages is never optimal. Instead, we have a corner solution which involves "maximum differentiation". Specifically, when it is feasible (i.e, when $\underline{w} = 0$) one individual will have a zero wage (and not work at all) while the other one will have a positive wage (namely $w_i = 2$) and positive labor supply. When $\underline{w} > 0$, this extreme solution is not feasible. However, it continues to be true that wages ought to be differentiated as much as possible.

To sum up, the maximal wage differentiation maximizes the size of total surplus and thus applies at all Pareto efficient allocations. When Ψ is strictly concave, social welfare is maximized when utilities are equalized. This is achieved through lump-sump transfers. Recall that we are in a first-best world for the time being so that redistribution (of incomes) does not involve any efficiency loss.

We now return to the case with general preferences and show that maximum differentiation continues to be optimal. However, the proof is more complicated than in the quasi-linear case.

3.3 General problem: solution

With general preferences we can no longer reduce the problem to a single dimension and the SOCs are too complicated to be conclusive. However, we can use a different approach and directly compare the relevant levels of welfare, namely W^I (achieved when $w_1 = w_2 = 1$) and W^C (achieved when, say, $w_1 = (2 - \underline{w})$ and $w_2 = \underline{w}$), and show that $W^C \geq W^I$ so that maximum differentiation continues to dominate even with general preferences.

To see this, define the level of labor supply under wage (and consumption) equalization

$$\ell^I = \arg\max_{\ell} \ \Psi\left[u\left(\ell, \ell\right)\right] + \Psi\left[u\left(\ell, \ell\right)\right]. \tag{10}$$

Maximum welfare is then given by

$$W^I = \Psi\left[u\left(\ell^I, \ell^I\right)\right] + \Psi\left[u\left(\ell^I, \ell^I\right)\right]. \tag{11}$$

At a corner solution with $w_1 = (2 - \underline{w})$ and $w_2 = \underline{w}$, maximum welfare is given by

$$W^C = \max_{\ell_1, \ell_2, T} \Psi\left[u\left(w_1\ell_1 - T, \ell_1\right)\right] + \Psi\left[u\left(w_2\ell_2 + T, \ell_2\right)\right], \tag{12}$$

where T is a lump-sum transfer. By setting T to equalize consumption levels (which is generally *not* the optimal level)[9] we obtain

$$W^C \geq \max_{\ell_1, \ell_2} \Psi\left[u\left(\frac{w_1\ell_1 + w_2\ell_2}{2}, \ell_1\right)\right] + \Psi\left[u\left(\frac{w_1\ell_1 + w_2\ell_2}{2}, \ell_2\right)\right].$$

By setting $\ell_1 = \ell_2 = \ell^I$ (which is again generally not optimal) and using $w_1 = (2 - \underline{w})$ and $w_2 = \underline{w}$ we then obtain

$$W^C \geq \Psi\left[u\left(\ell^I, \ell^I\right)\right] + \Psi\left[u\left(\ell^I, \ell^I\right)\right] = W^I. \tag{13}$$

Consequently, the corner solution dominates wages equalization. This because for $w_1 = (2 - \underline{w})$ and $w_2 = \underline{w}$ there are feasible choices of labor supplies and transfers which yield (at least) the same level of welfare as under wage equalization.

Our argument effectively implies that *any* unequal wages allocation (on the wage frontier) dominates equal wages. To complete the proof that there is a corner solution, we have to show that there is no interior solution with $w_1 \neq w_2 < \underline{w}$. It follows from (7) that an interior solution must be such that $\ell_1 = \ell_2$. But by using an argument similar to the one used for $w_1 = w_2 = 1$ one can easily show that the corner solution $w_1 = (2 - \underline{w})$ and $w_2 = \underline{w}$ dominates the (candidate) interior solution.[10]

To sum up, we have the following proposition

[9] Unless for instance Ψ is linear while u is strongly separable.

[10] Using a tilde to refer to the variables at this interior solution we have

$$\widetilde{W} = \Psi\left[u\left(\widetilde{c}_1, \widetilde{\ell}\right)\right] + \Psi\left[u\left(\widetilde{c}_2, \widetilde{\ell}\right)\right],$$

with $\widetilde{c}_1 + \widetilde{c}_2 = \widetilde{w}_1\widetilde{\ell} + \widetilde{w}_2\widetilde{\ell} = 2\widetilde{\ell}$. This allocation remains feasible for $w_1 = (2 - \underline{w})$ and $w_2 = \underline{w}$; specifically, by setting $\ell_1 = \ell_2 = \widetilde{\ell}$, total aoutput remains at $2\widetilde{\ell}$. This established weak dominance. Strict dominance obtains whenver the initial allocation is not optimal under maximal wage differentiation (which is true except in very special cases).

Proposition 1 *Assume that preferences are represented by the general utility function $u_i(x_i, \ell_i)$, that wages can be chosen according to $w_1 + w_2 = 2$, with $w_i \geq \underline{w}$ and that the social welfare function is symmetric and concave. When individual types are observable (so that personalized lump-sum transfers are available) the highest level of welfare is achieved maximum wage differentiation (with $w_i = \underline{w}$ and $w_j = 2 - \underline{w}$).*

In words, in a first-best setting, equalization of wages (according to the linear technology considered here) is at best useless and generally yields a reduction in welfare. Roughly speaking, the best strategy is thus to redistribute incomes rather than abilities. Recall that with the linear education technology considered, redistribution of abilities does not involve any efficiency loss per se (the total cake does not become smaller). Yet, it proves not to be desirable.

Even though this result may at first seem surprising, the intuition becomes obvious under closer scrutiny. Consider the simplest case with $\underline{w} = 0$. Then, if we go from equal wages to say $w_2 = 2$, individual 2's wage is effectively doubled. This means that we can produce the same output as under equal wages by having only one individual work. The proof of Proposition 1 is constructed around this argument which becomes slightly more complicated when $\underline{w} > 0$. The main point is that if we differentiate wages from $w_1 = w_2 = 1$ to $w_1 = (2 - \underline{w})$ and $w_2 = \underline{w}$ (for whatever level of $\underline{w} \geq 0$) while keeping individual labor supplies unchanged (at ℓ^I) output does not change so that the level of welfare achieved under equalization remains feasible under differentiation.[11] In other words, differentiation is *at least as good* as equalization. But it is also plain that keeping labor supplies equal (and at their equal wage levels) is generally not the best that can be done under wage differentiation. Consequently the differentiated solution will be strictly better than the equal wage outcome (or any other interior allocation).

[11] We have

$$\begin{aligned} w_1\ell_1 + w_2\ell_2 &= (2 - \underline{w})\ell_1 + \underline{w}\ell_2 \\ &= (2 - \underline{w})\ell^I + \underline{w}\ell^I \\ &= 2\ell^I. \end{aligned}$$

4 The second-best

So far we have assumed full information under which redistribution of incomes through lump-sum taxes and transfers is possible and does not involve any efficiency loss (we move along the Pareto frontier). This first-best result is interesting, but it is only of limited relevance when it comes to the design of tax policies. We now turn to a setting of asymmetric information and assume that only income $y_i = w_i \ell_i$ is publicly observable, while wages and labor supplies are private information. This is the information structure on which most of the optimal tax literature is based. In this setting the case for equalization of wages appears to be considerably strengthened. Specifically, when wages are equalized, distortionary redistribution of incomes is not needed. On the other hand, when wages are different, redistribution along the first-best frontier may no longer be available.[12] Maximal wage differentiation may then, as shown in the previous section, yield a bigger "cake", but it is not clear if it remains bigger once the efficiency losses of income redistribution are incurred.

Quite surprisingly, when $\underline{w} = 0$, maximal wage differentiation continues to dominate, even in a second-best setting. We show this through two propositions. The first one is of limited scope and applies only to some classes of utility functions. However, it provides some interesting intuition. The second one applies to all utility functions but is established in a more indirect way.

To specify the second-best problem, we can start from the first-best problem, (3)–(6) to which we have to add the following two incentive constraints

$$u(x_i, \ell_i) \geq u(x_j, \widetilde{\ell}_j) \qquad i, j = 1, 2; i \neq j,$$

where $\widetilde{\ell}_j = w_j \ell_j / w_i$ is the labor supply of individual i mimicking individual j. Observe that when $w_j \ell_j > 0$, $\widetilde{\ell}_j$ is only defined if $w_i > 0$. An individual with zero wage can never mimick an individual with positive wage (and labor supply).

Now, assume $\underline{w} = 0$ and consider an allocation with maximum wage differentiation and say $w_1 = 0$ and $w_2 = 2$. Assume for the time being that the first-best allocation

[12] Formally, the policy space (in the second stage) is restricted to incentive compatible mechanisms and the tax functions that implement them.

(solution to (3)–(6)) implies

$$u(x_2, \ell_2) \geq u(x_1, 0), \tag{14}$$

so that the productive individual (who works) has a utility level that is at least as high as that of the non productive individual (who does not work). In that case it is plain that the first-best solution is also the second-best solution. Put differently, the full information optimum remains implementable under asymmetric information. To see this note that (14) implies that the incentive constraint of type 2 individuals is satisfied because with $w_1 = 0$ we have $\widetilde{\ell}_2 = \ell_1 = 0$.[13] Furthermore with $w_1 = 0$ type 1 individuals cannot mimick type 2 individuals. When inequality (14) is strict, it follows by continuity that the first-best solution continues to be implementable when $\underline{w} > 0$ and sufficiently small. We thus have the following proposition.

Proposition 2 *Assume $\underline{w} = 0$. When the first-best solution with $w_i = 0$ and $w_j = 2$ $(i, j = 1, 2; i \neq j)$ implies $u(x_j, \ell_j) \geq u(x_i, 0)$, then it can also be implemented if individual incomes are observable while wages and labor supplies are not. Furthermore, if $u(x_j, \ell_j) > u(x_i, 0)$ for $w_i = \underline{w}$ and $w_j = 2 - \underline{w}$ $(i, j = 1, 2; i \neq j)$, the first-best continues to be implementable when $\underline{w} > 0$ but sufficiently small.*

This proposition applies to a number of interesting settings. In particular, it holds for the Rawlsian welfare function or with quasi-linear preferences when Ψ is linear (simple utilitarian criterion). However, there are also many instances where it does *not* apply. A prominent and well known example is the case of separable preferences $u(x, \ell) = f(x) - g(\ell)$, where the first-best implies $x_1 = x_2$ so that (14) is violated under maximum wage differentiation (the productive individual is worse off than the unproductive individual).

When Proposition 2 applies, the first-best solution can be implemented in the second-best. Consequently first and second best solutions coincide (and both imply maximum wage differentiation). When (14) is violated, the first-best solution can no longer be implemented in the second best. However, as long as $\underline{w} = 0$ we can show that maximum wage differentiation continues to dominate the solution with equal wages.

[13] This continues to be true for $\underline{w} > 0$ when inequality (14) is strict.

Recall that with equal wages the first-best solution is achieved with $x_i = \ell_i = \ell^I$, where ℓ^I is defined by (10). Since individuals are treated identically this is both a first- and a second best solution. Now consider again an allocation with maximum wage differentiation and say $w_1 = 0$ and $w_2 = 2$. Set $\ell_2 = \ell^I$ and define $\widehat{T}$ such that

$$u(\widehat{T}, 0) = u(\ell^I, \ell^I). \tag{15}$$

Observe that $T < \ell^I$ (utility being decreasing in ℓ). It is then easy to verify that the allocation defined by $x_1 = \widehat{T}$, $\ell_1 = 0$, $x_2 = 2\ell^I - \widehat{T}$ and $\ell_2 = \ell^I$, is feasible, Pareto dominates the equal wage solution $x_i = \ell_i = \ell^I$ and is incentive compatible. Pareto dominance follows from (15) for individual 1. For individual 2, $\widehat{T} < \ell^I$ implies

$$u(x_2, \ell_2) = u(2\ell^I - \widehat{T}, \ell^I) > u(\ell^I, \ell^I). \tag{16}$$

Finally, combining (15) and (16) yields

$$u(x_2, \ell_2) > u(\widehat{T}, 0), \tag{17}$$

which implies incentive compatibility.[14] Observe that because of inequalities (16) and (17) the argument remains valid when $\underline{w} > 0$ and sufficiently small. To sum up, we have established the following proposition

Proposition 3 *Assume $\underline{w} = 0$. Under maximum wage differentiation with $w_i = 0$ and $w_j = 2$ $(i, j = 1, 2; i \neq j)$ there exists a feasible allocation defined by $x_i = \widehat{T}$, $\ell_i = 0$, $x_j = 2\ell^I - \widehat{T}$ and $\ell_j = \ell^I$ (where $\widehat{T}$ satisfies $u(\widehat{T}, 0) = u(\ell^I, \ell^I)$) that Pareto dominates the equal wage solution ($x_i = \ell_i = \ell^I$), and is incentive compatible. This property continues to apply with $\underline{w} > 0$ and sufficiently small.*

The allocation we have constructed to establish this proposition is of course in general not the optimal one but this can only strengthen the case against wage equalization.

To sum up, as long as $\underline{w} = 0$, wage equalization is never optimal, even in the second-best (i.e., even when it makes potentially costly ex post redistribution unnecessary). The case where $\underline{w} = 0$ is admittedly extreme. We have stated Propositions 2 and 3

[14] To establish (16) we use $w_2 = 2$ and thus $w_1 = 0$.

for this extreme case to obtain clean results. However, it is clear that as long as we get strict inequalities both results also hold (by continuity) for $\underline{w} > 0$ but sufficiently small. However, when $\underline{w} > 0$ and large enough the (second- best) dominance of maximum wage differentiation can then no longer be established at the same level of generality; we then can no longer rule out that wage equalization *may* be optimal in some situations (in particular when $\underline{w}$ is close to 1 so that only "local" differentiation is feasible). To illustrate this point we now turn to a numerical example.

5 Numerical example

To illustrate the argument at hand we use a quadratic disutility for labor so that the *laissez-faire* implies

$$u_i = w_i \ell_i - \ell_i^2/2 = w_i^2/2.$$

With a utilitarian social welfare function, social indifference curves in the plane (w_1, w_2) are then defined by

$$S = \frac{w_1^2}{2} + \frac{w_2^2}{2},$$

which corresponds to quarter circles as represented in Figure 1. When $w_1 + w_2 = 2$, we reach the same level $S = 2$ with $(w_1, w_2) = (2, 0)$ or $(0, 2)$, while wage equalization $w_1 = w_2 = 1$ yields $S = 1$.(a local minimum). To achieve $S = 2$ under wage equalization one would need $(w_1, w_2) = \left(\sqrt{2}, \sqrt{2}\right)$, which implies $w_1 + w_2 = 2\sqrt{2} > 2$. Consequently, to reach the same welfare with equal wages as with unequal wage distribution and lump-sum taxes, we need a larger total level of human capital.[15]

The level of (per capita) welfare as a function of w_2, with $w_1 = 2 - w_2$ over the range $[1, 2]$ is represented on Figure 2. Not surprisingly this curve is monotonically increasing. Thus, whatever the level of $\underline{w}$ the first-best solution always implies maximum wage differentiation. But of course this only confirms our theoretical results from Section 3.

We now turn to the second-best optimum. To tackle this problem we maximize (Rawlsian) welfare for a given level of w_2, with $w_1 = 2 - w_2$, subject to resource and

[15] A larger level of "total" human capital with equal wage than under unequal wages would be possible with decreasing returns to education. See the concluding section.

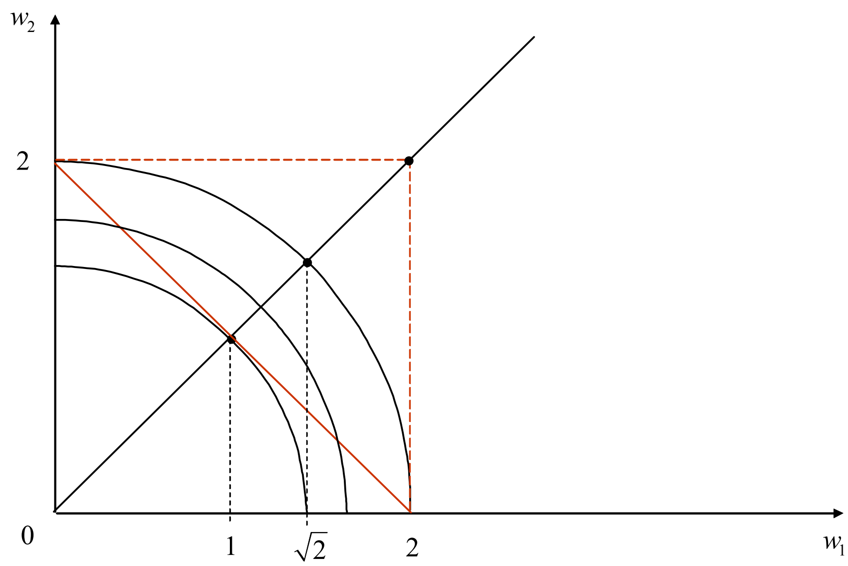

Figure 1: Illustration with quasi linear preferences and quadratic disutility of labor.

self selection constraints. The resulting level of welfare is represented on Figure 2. Not surprisingly first- and second-best coincide for $w_1 = w_2 = 1$. From Proposition 2 we also expect them to coincide in the neighborhood of $w_2 = 2$ (i.e., $w_1 = 0$). The really interesting feature about this welfare profile is that the curve is not monotonic. This means that while a large wage differential (when allowed by the level of $\underline{w}$) is always better than wage equalization, small wage differentials may not be second-best optimal. In other words, when $\underline{w}$ is large enough (more than roughly 0.8 in the example) wage equalization is effectively the second-best solution (and dominates $w_1 = \underline{w}$ and $w_2 = 2 - \underline{w}$). Still, the example also illustrates our result that maximum differentiation remains optimal when $\underline{w} > 0$ and not "too large". Our theoretical results do of course not tell us how large $\underline{w}$ must be in order to reverse the results. Our simulation provides an example of a case where $\underline{w}$ must be "quite large" for this to occur. While the orders of magnitude are of course of no practical relevance here the example nevertheless makes it clear that at least in some case the property does not just occur in the neighborhood of $\underline{w} = 0$.

Let us finally turn to the intuition behind the result that a small wage differentiation may not be desirable. In the second-best, differentiating wages has two (contradictory) effects. First it is efficiency enhancing. This is shown by the fact the first-best level of welfare is increasing in w_2. However, it also brings about an informational problem (unobservable types and the need for distortionary taxation). Interestingly the second-best curve is also convex which mean that there are "increasing returns" to wage differentiation. Even though the shape of the curve is example specific, it is not surprising; recall the intuition behind Proposition 1 presented in Section 3. Because of these increasing returns the efficiency gain associated with a small wage differentiation is not sufficient to compensate for the adverse informational impact. However, a large degree of differentiation yields efficiency gains that outweigh the negative effects of information asymmetries.

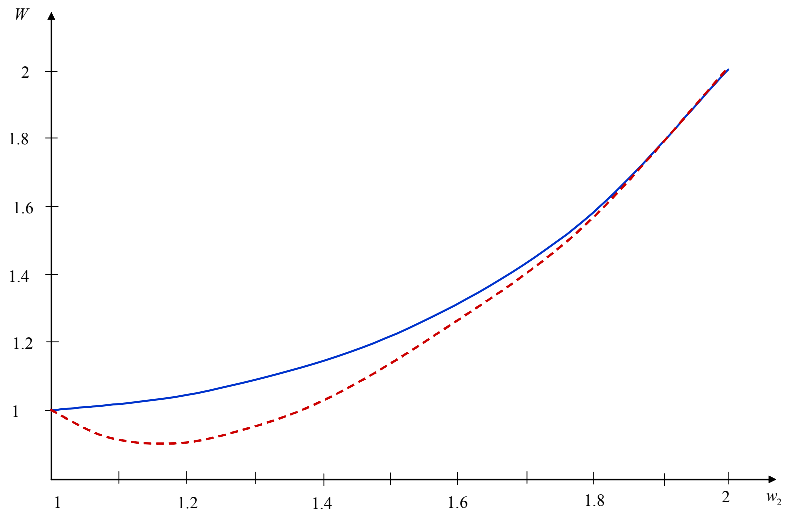

Figure 2: Welfare as a function of w_2, with $w_1 = 2 - w_2$ at the first best (solid curve) and the second-best (dotted curve) solutions.

6 Conclusion

When the productivities of otherwise identical individuals can be controlled through education expenses and when no subsequent redistribution via income taxation is available, it is clear the equal wage solution is very compelling. The question raised in this paper is different. We have studied the optimal distribution of productivity under the assumption that redistribution through an optimal income tax is also available. To give the equal wage outcome the best chance to emerge, we have assumed that individuals have identical learning abilities. Nevertheless, it turns out that in the first best inequality is *always* preferable to equality. This result does not appear to be known in the public finance literature even though it is reminiscent of "specialization" results obtained in family economics. Even more striking and unknown is the fact that this property remains valid in the second best for a large class of settings. Specifically, it always holds if the lower limit on productivity is not too high. On the other hand, the result *may* be reversed when wage inequalities are sufficiently restricted. Either way, the underlying problem is inherently non convex and a simple inspection of first-order conditions can be very misleading.

Admittedly, our model is at odds with most models of education. We assume that all individuals are ex ante identical; we also assume away the idea that educational spending has decreasing returns and the fact that the total amount of expenditure is endogenous. Let us see what would be the implications of relaxing these three assumptions. Making educational spending endogenous (and not equal to 2) would not change the main result. Decreasing returns could make a difference. Assume for the sake of illustration that instead of $w_1 + w_2 = 2$ we have $\max(w_1, w_2) = 2$, then it is clear that social optimality would imply $w_1 = w_2$; see Figure 1. This illustration also makes it clear that the degree of decreasing returns to scale must be sufficiently strong to reverse our result. When the set of feasible wages is convex but with a frontier sufficiently close to the line representing $w_1+w_2 = 2$ wage differentiation continues to be optimal. Roughly speaking the result is only reversed if the degree of concavity of the wage frontier is so large that its intersection with the 45 degree line is above the point $(\sqrt{2}, \sqrt{2})$. Finally, there is the assumption of uniform learning ability. If we were to assume different learning ability, then our results could only be strengthened. The main difference is that the individual with the higher wage level would be the more able.

The information structure in our model also differs from that considered by other authors (such as Bovenberg and Jacobs (2005) and Maldonado (2008)). Specifically, in our second-best setting we assume that wages are not observable, which may appear to be odd as they result from public education.[16] Implicitly we assume that for informational and administrative reasons the connection between the education department and the tax authorities is not perfect. Alternatively, our assumption can be justified by assuming that income taxes can only be based on contemporary variables (and not on past education records); this assumption is neither unrealistic, nor uncommon in the

[16] Bovenberg and Jacobs (2005) and Maldonado (2008) assume that individuals are heterogenous *ex ante* at the education stage. Their mechanisms (implemented by a tax function) screens for this unobservable type and transfers can be conditioned on education choices and income. Recall that we have no asymmetric information in the first (education) stage because individuals are identical *ex ante*.

literature. In any case, if tax authorities could use educational records, we would fall back on the first-best setting for which our results are even stronger.

To sum up, the somehow surprising result obtained in this paper is not just a technical curiosity, but would continue to hold in a more general setting. This calls for a number of remarks. First, it means that for people who are only concerned by the maximization of utilitarian social welfare the concept of "equality of opportunity", defined as equality of earning capacity, might be undesirable.

Second, one may find unsatisfactory a society leaves a good number of individuals without education, even when utility levels are equalized. If we think that education brings utility by itself and thus should be explicitly introduced in the utility function, our conclusion would be different. Finally, we have to keep in mind that for reasons pertaining to political economy redistribution through education may be easier than tax redistribution.

References

[1] Arrow, K. (1971), A utilitarian approach to the concept of equality in public expenditure, *Quarterly Journal of Economics*, 85, 409–415.

[2] Athanasiou, E. (2007), Educational outcomes and equality of opportunity, unpublished.

[3] Becker, G. (1985) Human capital, effort and the sexual division of labor, *Journal of Labor Economics*, 3, S33–S58.

[4] Becker, G. (1991) A *Treatise on the Family* (enlarged edition) Harvard University Press, Cambridge, Mass.

[5] Bénabou, R. (2002), Tax and education policy in a heterogenous-agent economy: What levels of redistribution maximize growth and efficiency?, *Econometrica*, 70, 481–517.

[6] Boadway, R. and P. Pestieau (2007), Tagging and Redistributive taxation, *Annales d'Econométrie*, 83–84, 123–147.

[7] Bovenberg, L. and B. Jacobs (2005), Redistribution and education subsidies are Siamese twins, *Journal of Public Economics*, 89, 2005–2035.

[8] Brett, C. and J. Weymark, (2008), The impact of changing skill levels on optimum nonlinear income taxes, *Journal of Public Economics*, forthcoming.

[9] Bruno, M. (1976), Equality, complementarity and the incidence of public expenditure, *Journal of Public Economics*, 5, 395–407.

[10] De Fraja, G. (2002), The design of optimal educational policies, *Review of Economic Studies,* 69, 437–466.

[11] Glomm, G. and B. Ravikumar (1992), Public versus private investment in human capital: endogenous growth and income inequality, *Journal of Political Economy*, 100, 818–834.

[12] Hare, P.G. and D. Ulph (1979), Education and income distribution, *The Journal of Political Economy*, 87, 5193-5212.

[13] Maldonado, M. (2008), Education policies and optimal taxation, *International Tax and Public Finance*, 15, 131–144.

[14] Mirrlees, J.A. (1971), An exploration in the theory of optimum income taxation, *Review of Economics Studies*, 38, 175–208.

[15] Poterba, J. (1995), Government Intervention in the Markets for Education and Health Care: How and Why?, NBER Working Paper 4916.

[16] Roemer, J. (1998), Equality of Opportunity, Cambridge, M.A., Harvard University Press.

[17] Stiglitz, J.E. (1982), Self-selection and Pareto efficient taxation, *Journal of Public Economics* 17, 213–240.

[18] Ulph, D. (1977), On the optimal distribution of income and educational expenditure, *Journal of Public Economics*, 8, 341–356.