Fair optimal tax with endogenous productivities*

Marc Fleurbaey† and Giacomo Valletta‡

April 22, 2014

Abstract

The literature on fair income taxation has so far adopted the sharply dual goal of eliminating inequalities due to unequal skills while accepting inequalities due to diverse consumption-leisure preferences. We introduce human capital choices and endogenous skills in order to be able to hold individuals partly liable for their skills and for their consumption-leisure preferences. With a maximin social objective embodying this approach, we analyze the evaluation of tax reforms and the properties of optimal linear and non-linear taxes. Social priority is granted to the situation of individuals with the most disadvantaged characteristics who work full time and spend a certain amount in human capital.

Keywords: tax reforms, optimal tax, fairness, endogenous skills, human capital.

JEL Classification: D63, D71.

1 Introduction

The literature on optimal taxation has recently started to study non-utilitarian social objectives, on the grounds that utilitarian social welfare functions fail to incorporate important value judgments that are part of the current public debates. Boadway [3], Piketty and Saez [27] , Weinzierl [34],

*We thank François Maniquet, Paolo Piacquadio and Emmanuel Saez for comments and suggestions, as well seminar and conference participants at CORE, Institute for Global Studies, Luiss, the Social Choice and Welfare Meeting 2012 (New Delhi) and the ASSA meeting 2013 (San Diego). A first draft of this paper was prepared while G. Valletta was visiting the Institute for Global Studies (Paris), to which he is grateful for hospitality and financial support.

†Princeton University, USA. Email: mfleurba@princeton.edu.

‡Maastricht University, Netherlands - Email: g.valletta@maastrichtuniversity.nl.

and, most comprehensively, Saez and Stantcheva [32] provide a list of relevant considerations of equity that are ignored by the classical approach. This list includes the popular idea that, unlike disadvantageous circumstances, personal preferences and effort should not be a target of redistribution. This idea has been applied to the problem of optimal taxation by Fleurbaey and Maniquet ([12], [13], [14]), with results suggesting a zero or even negative marginal tax rate for low incomes, reflecting the focus on the hardworking poor induced by the dual compensation-responsibility objective.[1] Jacquet and Van De Gaer [20] have also studied the compensation-responsibility approach for the extensive margin. A controversial aspect of these works, however, is that they assume that the workers' earnings abilities are totally a matter of circumstances, while their preferences over consumption and labor are totally a matter of responsibility. This may be viewed as a convenient starting point, but in this paper we propose to examine how the analysis is modified when the individuals can be held partly responsible for their earning abilities (due to their training or health choices) and partly non-responsible for their preferences.

We do this in the following way. We introduce a third dimension alongside consumption and leisure, namely, human capital, which is meant to encapsulate both skills and health conditions and has a positive impact on earning ability.[2] The individuals will choose their human capital in addition to their leisure time and consumption, under a budget constraint. In this way, a partial responsibility for earning ability is introduced. It is partial only because individuals may differ in the cost of human capital (due to genetic and epigenetic characteristics,[3] social background, health contingencies) and in their earning ability for a given level of human capital (due to social networks, location, wage differences between industries). On the other hand, the presence of human capital as an argument of individual utility may explain certain differences in preferences. Apparent aversion to work may be due to bad health, or to low skills giving access to low-quality jobs. Insofar as a low human capital is partly due to circumstances, the individuals' consumption-leisure preferences can also be taken to be only partly a matter of responsibility.[4]

[1] Boadway et al. [4] had already obtained the possibility of such results with weighted utilitarian criteria in which the weights could vary with the degree of "desert" of agents with different preferences, but they had not proposed a methodology for determining the weights.

[2] This is true not just for education but also for health care as shown, among others, by Mushkin [26], Grossman and Benham [15], Luft [22].

[3] Christensen et al. [8] show that approximately a quarter of the variation in self-reported health and in the number of hospitalizations could be attributed to genetic factors. Currie [9] studies the epigenetic determinants of human capital.

[4] In the literature (e.g., Roemer [31]) it is sometimes argued that individuals should not be held responsible for the part of their preferences that is influenced by their circumstances (e.g., their upbringing). This sort of reasoning makes sense when the goal is to give people equal access to a material outcome such as income, because an upbringing that makes one less hardworking is indeed an obstacle on the road to income. But in our paper, welfare is at stake, so we stick to the idea that individual preferences are equally respectable no matter how they were formed, and leave the study of paternalism for another occasion.

Our aim is to study how the results obtained by the "fair taxation" approach are modified by this important refinement of the framework. We adopt a social welfare ordering that incorporates the compensation and responsibility fairness principles and examine how the evaluation of income taxes and the design of optimal taxes is affected by the introduction of the new dimensions of heterogeneity and the new features of partial responsibility. We examine income taxation but also introduce the possibility to subsidize (or tax) human capital expenditure. We supplement the analysis of the optimal non-linear tax with a study of linear taxes.

The main results are the following. On the technical side, the assumptions that were made in the simpler framework by Fleurbaey and Maniquet ([12], [14]) are less realistic in our setting and, without such assumptions, we develop new techniques in order to obtain results. These techniques can be useful in other models.

On the economic side, we look first at the problem of assessing tax reforms. The focus on the hardworking poor is now diffused to a specific range of earnings that takes account of a certain level of human capital that the most disadvantaged types of individuals would achieve under certain conditions. More precisely, in order to evaluate a certain tax scheme the policy maker should look primarily at the part of the budget set that is attainable by an agent with the worst earning ability and the worst human capital disposition[5]. We show that within this region the focus should always be on the earnings of agents working full time, with a productivity corresponding to low or high expenses in human capital depending on the shape of the tax function. When subsidies on human capital are absent (or very low) the focus is more likely to go on individuals spending the maximum affordable amount on human capital (and consuming very little). However, if human capital is already to some extent subsidized then the focus shifts toward lower levels of human capital expenditures (and hence lower earnings). Subsidies on human capital therefore affect the level of earnings that gets most priority, making it possible to better target the population that should be the focus of the reform.

We also study the shape of the optimal income schedule. We consider both linear and non-linear taxes. In the former case, it turns out that in certain instances taxing human capital expenditure might be optimal. Interestingly the occurrence of a negative subsidy rate on human capital expenditure is closely related to the distribution of preferences across the population. For example a high sensitivity of human capital expenditures to subsidies together with a low sensitivity of earnings to tax will bring high tax rates and low subsidy rates. The same happens if the worst-off agents spend considerably less than average in human capital. On the other hand, when we turn our attention to the nonlinear case we find that the agents who are at the focus of social preferences are typically subsidized, although their marginal subsidy may be null. The agent who receives the

[5] However the worst-off agent at the allocation generated by a certain tax policy does not actually need to be an agent whose circumstances are the worst in society.

highest subsidy is typically some disadvantaged agent who works full time and who, among all the agents with the poorest human capital disposition, has a high (sometimes the highest) human capital expenditure.

The paper is organized as follows. Section 2 briefly reviews the related literature. Section 3 introduces the model and the social preferences used in the paper. Section 4 proposes a way to compare different arbitrary tax policies, with and without subsidies on human capital expenditures. Section 5 examines the optimal non-linear tax schedule. Section 6 briefly examines two other issues: the case of observable human capital, and the case of linear taxes. Section 7 concludes.

2 Related literature

The generality of our model connects it with different strands of the literature on optimal income taxation depending on how one interprets human capital.

One can think of human capital as the level of education of agents. A first idea found in the literature on education and taxation is that appropriate education subsidies may offset the disincentives to work and acquire education induced by a progressive income tax (Krueger and Ludwig [21]). However this efficiency effect may be partly counterbalanced by a redistribution effect due to the fact that education may be associated with greater ability, as analyzed by Maldonado [23] and Bovenberg and Jacobs [6]. A striking result in these papers is that the optimal policy will simply deduct education expenditures from taxable income under specific conditions about the earnings function. These conditions are studied in detail by Jacobs and Bovenberg [19] and involve comparing the degree of complementarity between education and labor on one hand and between education and ability on the other hand. When earnings are equal to the product of labor and a wage rate that depends on education expenditures and individual ability, as in Maldonado [23] and in our model, the condition for full tax deduction is that the elasticity of the wage rate with respect to education expenditures be the same for all individuals (i.e., be independent of ability) —if it increases with ability, then education is not fully deductible.

This result is obtained in models in which individual preferences are indifferent to the education level. In such models efficiency dictates treating education as an intermediate good used as a mean to maximize consumption. In our paper we allow individuals to care about their level of human capital. Our results on the optimal rate of subsidy (or tax) on human capital expenditures therefore depend not only on properties of the earnings function, but also on the distribution of preferences in the population.

One can also interpret human capital as the level of health of the individuals. Most of the literature focusing on the taxation/subsidization of health care does not allow for endogenous earning abilities. Sickness is simply considered as a loss in the resources available to an individual (see, among others, Blomqvist and Horn [2], Cremer and Pestieau [7], Henriet and Rochet [17] and Rochet [30]). The main objective of these papers is to understand whether covering people against such a risk, by means of a public health insurance, is welfare improving or not from an ex ante-perspective. In this respect, health insurance can indeed be used as a redistribution device provided that the probability of being ill is comparatively larger for agents at the bottom of the income distribution. Our setting is different by making productivity and preferences being influenced by health, and also we look at the ex post distribution of health after individuals make choices about treatments.[6]

Our paper also relates to the literature on commodity taxation. Atkinson and Stiglitz [1]'s seminal result on useless commodity taxation (see Boadway and Pestieau [5] for an overview) does not hold in our framework if applied to human capital expenditures, because human capital here affects the agents' productivity, and agents have to face unequal costs in order to acquire it.

All the papers quoted so far have focused on social objectives defined in terms of utilitarian-type social welfare functions. Such social welfare functions are typically not precisely specified, and they generally assume that all individuals have the same utility function. As explained in the introduction, we follow the fair tax approach which relies on specific representations of preferences embodying fairness principles, and which gives an absolute priority to the worst-off. We examine how to extend it in order to introduce partial individual responsibility for skills and for consumption-labor preferences. A previous exploration of this extension has been made by Valletta [33] with a much simpler version of our model in which health influences productivity, the choice of health status is dichotomous, and there are only two types of earning abilities and health dispositions in the population. He provides an axiomatic characterization of a social ordering function that can be easily extended to our model. This is the social ordering that we retain here, without repeating the axiomatic analysis. But his analysis of taxation cannot be extended to our model in which we have an arbitrarily large number of types of individuals and a continuum of human capital values. Moreover, in addition to the simpler framework, he makes specific assumptions about the profile of characteristics that enable him to follow the techniques used in Fleurbaey and Maniquet [12]. We avoid these assumptions and develop new techniques.

3 Model and social preferences

Consider an economy with a finite set of agents $N \subset \mathbb{N}$. There are three goods: consumption, labor and human capital. A bundle for agent $i \in N$ is a triple $z_i = (c_i, l_i, h_i)$, where c_i is consumption, l_i is labor, and h_i is human capital. In particular, $c_i \in \mathbb{R}_+$ will be interpreted here as the expenditure on ordinary consumption goods, excluding human capital expenditure. As usual for this kind of

[6] For a justification of this focus on the ex post allocation, see Fleurbaey (2010).

analysis, $l_i \in [0,1]$. Human capital is also a continuous variable, and for simplicity we assume $h_i \in [0,1]$. To sum up, for each $i \in N$, $z_i \in Z := \mathbb{R}_+ \times [0,1] \times [0,1]$. An allocation describes each agent's bundle, and will be denoted by $z = (z_i)_{i \in N}$.

Agents have three characteristics: their personal preferences, their earning ability and their human capital disposition.

For each agent $i \in N$, preferences are denoted R_i and $z_i' R_i z_i$ (resp $z_i' P_i z_i$, $z_i' I_i z_i$) means that bundle z_i' is weakly preferred (resp. strictly preferred, indifferent) to bundle z_i. We restrict our attention to preferences which are continuous, strictly monotonic (increasing in c_i and h_i,[7] decreasing in l_i) and convex. Let $R = (R_i)_{i \in N}$ denote the profile of preferences of the whole population.

The wage rate is assumed to be an increasing function of human capital, $w_i(h_i)$ with $w(0) \geq 0$. It is measured in consumption units per full time labor, so that for any l_i, $w_i(h_i)l_i$ is the agent's pre-tax income (earnings). Agents are endowed with different such functions. For $i, j \in N$ we say that agent i is more productive than agent j if i's productivity function dominates j's, that is, if $w_i(h) \geq w_j(h)$ for all h. Let $w(.) = (w_i(.))_{i \in N}$ denote the profile of individual productivity functions for the whole population.

Finally, every individual i has a mapping $m_i(h_i)$ describing the amount of money she has to spend in order to attain the human capital h_i. We assume that this function satisfies $m_i(h) = 0$ for $h \leq \underline{h}_i$, increases over $\left[\underline{h}_i, \overline{h}_i\right)$, and is equal to $+\infty$ for $h \geq \overline{h}_i$. One can interpret $\underline{h}_i$ as the level of h in absence of expenses, and $\overline{h}_i$ as the maximum attainable level for i. We define the inverse function $m_i^{-1} : \mathbb{R}_+ \to \left[\underline{h}_i, \overline{h}_i\right)$ by $m_i^{-1}(0) = \underline{h}_i$ for $m = 0$ and $m_i^{-1}(m) = h$ such that $m_i(h) = m$ for $0 < m$. For $i, j \in N$ we say that agent i has a (weakly) worse human capital disposition than agent j if $m_i(h) \geq m_j(h)$ for all h. Let $m(.) = (m_i(.))_{i \in N}$ denote the profile of human capital dispositions for the whole population.[8]

An economy is denoted $e = (R, w(.), m(.))$. Let $\mathcal{D}$ denote the set of economies complying with our assumptions.

An allocation is feasible if

$$\sum_{i=1}^{n} c_i + \sum_{i=1}^{n} m_i(h_i) \leq \sum_{i=1}^{n} w_i(h_i) l_i.$$

[7] Individuals have preferences for human capital so that they may choose a certain level of human capital just because they care about it and not only because this choice is instrumental to the attainment of a higher level of consumption (via a greater productivity).

[8] When the acquisition of human capital costs not just money but also labor time, as in the case of higher education, this time cost can be taken into account in this model via the preferences over (c, l, h). For instance, if an individual cares only about h because of this time cost, his preferences on (c, l, h) actually bear on the pair $(c, l + t(h))$, where $t(h)$ is the time needed to acquire l. But we prefer to interpret the model as bearing on a working period that comes after the choice of h has been made, and the time budget for paid work is now the same for all.

In absence of redistribution, the budget set of each agent $i \in N$ is equal to the possible combinations of consumption, labor and human capital that are attainable for her, given her earning ability and her human capital disposition. In the first best context, one can use lump-sum transfers in order to redistribute income across agents. Then, agent i's first-best budget set is, letting t_i denote the transfer:

$$B(t_i, w_i(.), m_i(.)) = \{(c_i, l_i, h_i) \in \mathbb{R}_+ \times [0,1] \times [0,1] \mid c_i + m_i(h_i) \leq t_i + w_i(h_i)l_i\}.$$

This notion of budget set is important in the criterion for social welfare that is used in this paper and that we now introduce. This criterion applies the leximin ordering to the vector of individual indices, and these individual indices are equal to the lump-sum transfers which, combined with average $w\,(.)$ and $m\,(.)$ dispositions, would give individuals the same satisfaction as in the allocation to be evaluated.

Formally, let $\overline{w}(.) = \frac{1}{|N|}\sum_{j \in N} w_j(.)$ and $\overline{m}(.) = \frac{1}{|N|}\sum_{j \in N} m_j(.)$ denote respectively the average earning ability function and the average human capital disposition function. Then, the implicit transfer associated with an agent's bundle z_i is defined by[9]

$$IT_i(z_i) = t \Leftrightarrow z_i \ I_i \ \max|_{R_i} B(t, \overline{w}(.), \overline{m}(.)).$$

This expression, as a function of z_i, corresponds to a particular money-metric utility function. This measure of individual well-being does not require any information about individuals' subjective utility, it relies only on information about ordinal non-comparable preferences.

For any given allocation one can compute the vector of implicit transfers associated with the bundles received by each agent. Two different allocations are then ranked by the leximin criterion to the vector of the corresponding implicit transfers. **Social welfare criterion:** For all $e \in \mathcal{D}$, $z, z' \in Z$,

$$z'\overline{R}(e)z \iff (IT_i(z'_i))_{i \in N} \geq_{lex} (IT_i(z_i))_{i \in N},$$

where $\overline{R}(e)$ denotes a social ordering function, namely, a mapping from the set of economies to the set of complete orderings over allocations, and $\geq_{lex}$ denotes the leximin ordering of real vectors.[10]

The key fairness properties that are satisfied by this particular criterion and underlie it are the following. First, one would like to compensate agents for differences in their circumstances that are beyond their responsibility. In our framework this amounts to saying that inequalities deriving

[9] The expression $\max|_{R_i} B(t, \overline{w}(.), \overline{m}(.))$ denotes the subset of $B(t, \overline{w}(.), \overline{m}(.))$ that contains the best allocations for R_i. Under our assumptions, this subset is always non-empty.

[10] The leximin lexicographically evaluates a real vector by looking at the lower components before the greater components.

solely from someone's human capital disposition or someone's earning ability are not acceptable. In other words:

It is a strict social improvement to make a Pigou-Dalton transfer on the consumptions of two agents i and j who have identical preferences $R_i = R_j$, the same amount of labor time and the same level of human capital, i.e. to change their consumption levels from c_i,c_j to c_i',c_j' such that

$$c_i - \Delta = c_i' > c_j' = c_j + \Delta,$$

for Δ a strictly positive real number.

Redistribution should however have a limit: inequalities solely due to different choices are acceptable if individuals are held responsible for their goals. If all agents had the same circumstances, i.e., the same human capital disposition mapping and the same earning ability mapping, then they should be free to choose a different amount of labor, a different level of human capital and hence, indirectly, a different productivity, and the optimal policy is the laissez-faire. By extension, in a first-best context with lump-sum transfers, reducing inequality in the lump-sum transfers should be seen as improving the social situation, because different preferences do not justify redistribution.

It is a strict social improvement, in an economy in which all agents have the same circumstances, to change an allocation obtained via lump-sum transfers $(t_i)_{i\in N}$ by making Pigou-Dalton transfers on the t_i's, i.e., for any two agents i and j, by changing t_i,t_j into t_i',t_j' such that

$$t_i - \Delta = t_i' > t_j' = t_j + \Delta,$$

for Δ a strictly positive real number.

The social ordering function introduced above can be axiomatically characterized by these two fairness requirements, together with efficiency, informational and robustness requirements.[11]

Before examining tax implications, let us briefly examine an interesting line of objection to the framework proposed here. Our approach makes a sharp distinction between individual preferences, which are left to the agents' responsibility, and individual earning abilities and human capital dispositions, for which the objective is full compensation. In practice, it may seem hard to separate abilities and dispositions from preferences. However, insofar as the variables c (net income), l (labor hours), h (education and/or health status), m (education and/or medical expenditures), w (hourly wage rate) have an empirical meaning, the approach is applicable. Note that the m function does not make a difference between individuals who need a lot of investment in order to obtain a given level of h because of a genuinely unfavorable disposition or because of ill-will (e.g., laziness in their studies, negligence in following medical guidance). This is similar to the Mirrlees [25] model

[11] Informational and robustness requirements make connections between the social orderings applied in economies with different population profiles. See Valletta [33] for the complete list of axioms and a formal characterization.

which does not distinguish between those who have a low wage rate for various reasons. As it is not obvious that individuals should be held liable for apparent "laziness" (which may come from hidden costs), it is probably sensible to err in this charitable direction.

Before examining tax implications, it may also be helpful for the reader have a quick glance at the laissez-faire allocation and the first-best allocation. The laissez-faire gives every i the budget $c_i + m_i(h_i) \leq w_i(h_i)l_i$, implying that for given preferences R_i, it is unambiguously bad to have a high $m_i(.)$ function and a low $w_i(.)$ function. We will later focus on economies in which there is a subpopulation whose $m_i(.)$ and $w_i(.)$ functions are worse than in the rest of the population. This subpopulation, in the laissez-faire, has a smaller budget set, in the sense of inclusion, than anyone else in the population.

The first-best allocation will seek to equalize $IT_i(z_i)$ across all i. The first-best allocation can be obtained by making lump-sum transfers from those with favorable $m_i(.)$ and $w_i(.)$ functions to those with disadvantageous functions. These are the main transfers. Additionally, people with identical $m_i(.)$ and $w_i(.)$ functions but different preferences will also sometimes receive different transfers (unless all $m_i(.)$ and $w_i(.)$ functions are identical in the population), with the individuals with greater taste for h and greater aversion to l receiving a lower lump-sum transfer if their $m_i(.)$ and $w_i(.)$ functions are flatter than $\overline{w}(.), \overline{m}(.)$, and a greater lump-sum transfer if their $m_i(.)$,$w_i(.)$ functions are steeper than $\overline{w}(.), \overline{m}(.)$.[12]

Note that equalizing $IT_i(z_i)$ across all individuals is not always possible if inequalities in $m_i(.)$,$w_i(.)$ functions are severe (and preferences have certain properties). For instance, if some individuals are in good health without spending anything, while others have a low $\overline{h}_i$, it may even happen that $IT_i(z_i)$ is infinite for the former agents at all feasible allocations if they feel always better off than choosing from $B(t, \overline{w}(.), \overline{m}(.))$ for any t, because the $\overline{m}(.)$ function necessarily goes to $+\infty$ at $\min_{i \in N} \overline{h}_i$, so that from $B(t, \overline{w}(.), \overline{m}(.))$ it is impossible to choose a (c, l, h) bundle with $h \geq \min_{i \in N} \overline{h}_i$.

4 Tax evaluation

The notion of social welfare just described can serve for the evaluation of arbitrary tax policies. As it is well known in the taxation literature since Feldstein [10], the reform problem is often more relevant to policy makers than knowing the features of the optimal tax policy. In such a case the policy maker is primarily interested in determining which part of the tax policy should be changed first in order to obtain a social improvement.

[12] Intuitively, they will have a greater h and lower l in the budget set $B(t, \overline{w}(.), \overline{m}(.))$ than people with lower taste for h and aversion to l, which, in the former case (flatter), corresponds to starting with a lower consumption at $h = \underline{h}_i$ and $l = 0$ with their flatter $m_i(.)$ and $w_i(.)$ functions which make h less costly and l less profitable.

As in Mirrlees [25], we assume that the policy maker knows the distribution of characteristics of the population. In such a case, he is actually able to forecast the statistical distribution of $IT_i(z_i)$ over the population for any tax schedule that is enforced. But it is important to seek if the evaluation of tax reforms can be done with less information and in a more convenient way than computing how each type of individual will be affected by the tax. The fact that the social criterion focuses on the worst-off has a great simplification power in this respect.

4.1 Incentive-compatible allocations

Consider a given economy $e = (R, w(.), m(.))$. The policy maker is assumed to know the distribution of types in the population but does not observe the characteristics of any particular agent. We assume that, in this second best context, only earned income, $y_i = w_i(h_i)l_i$, and human capital expenditure, $m_i = m_i(h_i)$, are observable.

A tax policy is a function $T(y, m)$ defining a transfer of income depending on the level of earnings and on the human capital expenditure. Later we will also examine the special case in which this is only an income tax $T(y)$. The tax turns into a subsidy when $T(y, m) < 0$. Individuals are free to choose their labor time and their human capital in the budget set modified by the tax function, namely, the set of bundles $(c, l, h) \in \mathbb{R}_+ \times [0, 1] \times [0, 1]$ such that

$$c_i \leq w_i(h_i)l_i - m_i(h_i) - T(w_i(h_i)l_i, m_i(h_i)).$$

Let $B_i(T)$ denote this set. In what follows we will focus on the space of consumption, earnings, human capital expenditure where agent's i budget set becomes

$$c_i \leq y_i - m_i - T(y_i, m_i).$$

In addition to the budget constraint, every agent is submitted to the constraints $c, m \geq 0$ and $y \leq w_i^*(m)$, where the function $w_i^*(m) = w_i \circ m_i^{-1}(m)$ determines the earning ability that i obtains with any amount of human capital expenditure m.

Let R_i^* define agent i's preferences over consumption, earnings and human capital expenditure. These are derived from the ordinary preferences R_i defined in the (c, l, h) space as follows:

$$(c, y, m) \, R_i^* \, (c', y', m') \Leftrightarrow \left(c, \frac{y}{w_i^*(m)}, m_i^{-1}(m)\right) R_i \left(c', \frac{y'}{w_i^*(m')}, m_i^{-1}(m')\right).$$

These preferences are continuous, convex, increasing in c, non-decreasing in m, and decreasing in y. In addition, they satisfy the following restriction:

$$(c, y, m) \, R_i^* \, (c, y', m') \text{ if } \frac{y}{w_i^*(m)} \leq \frac{y'}{w_i^*(m')} \text{ and } m \geq m'. \tag{1}$$

This restriction comes from the fact that in the (c, l, h) space, (1) amounts to

$$(c, l, h) \, R_i \, (c, l', h') \text{ if } l \leq l' \text{ and } h \geq h',$$

which is a direct consequence of monotonicity of preferences in l and h.[13]

The restriction described by (1) has important consequences on the agents' behavior which will significantly affect our analysis. Agents will never choose a bundle (c, y, m) if they are given the possibility to choose another bundle which entails the same labor supply $y/w^*(m)$, a greater m (therefore greater h), and no lower c. It is important to stress that an agent might be confronted with this kind of choice in many ordinary situations. For instance, in the laissez-faire budget set where $T(y, m) \equiv 0$, the bundle $(w^*(m) - m, w^*(m), m)$, corresponding to working full time and spending m in human capital, is dominated by another bundle $(w^*(m') - m', w^*(m'), m')$ if $m' > m$ and $w^*(m') - m' \geq w^*(m) - m$. In such a situation the extra human capital expenditure is more than repaid by the extra earnings it makes possible: $w^*(m') - w^*(m) \geq m' - m$. This restriction is a clear consequence of the fact that we are assuming endogenous productivity and imposes quite important changes in the analysis compared to the simpler model in which productivity is exogenous.

An allocation $z \in Z$ is incentive compatible if and only if no agent envies the bundle of any other agent provided that such a bundle is feasible for her: for all $i, j \in N$,

$$(c_i, y_i, m_i)\, R_i^* \,(c_j, y_j, m_j) \text{ or } y_j > w_i^* (m_i)\,.$$

In other words agent i has to receive an allocation that she prefers to the allocation received by agent j unless it is not possible for her to mimic agent j because j earns more than she can. This implies that any incentive-compatible allocation can be obtained by letting every agent $i \in N$ choose her best bundle, under the constraint $y \leq w_i^* (m)$, in a budget set modified by a well chosen tax function $T(y, m)$ such that the locus of points

$$S(T) = \left\{(c, y, m) \in \mathbb{R}_+^3 \mid c \leq y - m - T(y, m)\right\},$$

lies nowhere above the envelope curve of the indifference curves in the (c, y, m) space, and intersects this envelope curve at all points (c_i, y_i, m_i) for each $i \in N$. Conversely, any allocation obtained by letting all agents choose from a budget set $S(T)$, under the constraint, $y \leq w_i^* (m)$, is incentive compatible. In other words, the taxation principle (Guesnerie [16], Rochet [29]) holds in this model. An incentive compatible allocation so obtained is feasible if and only if $\sum_{i=1}^N T(y_i, m_i) \geq 0$.

For every incentive-compatible allocation, there is a *minimal tax* that implements it, namely, the tax T such that $y - m - T(y, m)$ follows the lower envelope of agents' upper contour sets in the (c, y, m) space at the allocation. For such a tax, $S(T)$ coincides with the intersection of the closed lower contour sets of the agents.

[13] An additional restriction is that

$$(c, y, m)\, I_i^* \,(c, y, m') \text{ if } m, m' \geq m_i\left(\overline{h}_i\right),$$

because this corresponds to a situation in which the corresponding (c, l, h) bundles are the same.

However, in this model, because of (1), minimal taxes form a relatively narrow class of tax functions. In the classical Mirrlees model (with exogenous productivities), any non-decreasing function $y - T(y)$ can be arbitrarily close to the lower envelope of agents' upper contour sets at the allocation generated by the tax function for a sufficiently large population with sufficiently diverse preferences. In contrast, here, a function $y - m - T(y, m)$ that is non-decreasing in y and non-increasing in m may never be close to a minimal tax configuration, for the following reason. It holds true, by monotonicity of preferences R_i^*, that we can restrict attention to tax functions T such that $y - m - T(y, m)$ is non-decreasing in y and non-increasing in m. But the restriction (1) implies that some parts of such a budget set may never be chosen by any agent, because for a fixed amount of labor, consumption is increasing with human capital expenditure.

4.2 Estimating social welfare: a lower bound

The evaluation of a policy hinges on its social consequences. It turns out that evaluating the consequences of a certain policy is made easier by the fact that we are using a social ordering function of the leximin type. Indeed, given the allocation generated by a given tax policy we first need to spot the worst-off agents at such an allocation, and this is often sufficient to make the evaluation. Once we have this piece of information we know which parts of the budget set modified by the tax function have to be changed (and how) in order to obtain a social improvement.[14]

Let T be an arbitrary tax function such that $y - m - T(y, m)$ is non-decreasing in y and non-increasing in m. The main purpose of this section is to provide a measure of $\min_i IT_i(z_i)$ at the allocation z generated by T. We directly consider the case of a function $T(y, m)$ that may include subsidies on m rather than an income tax $T(y)$, because the latter is a special case.

The methodology developed, for Mirrlees' model, by Fleurbaey and Maniquet ([12],[13]) to measure social welfare under an arbitrary tax code, is not appealing here. It consists of (i) assuming that the least-skilled individuals have a sufficient diversity of preferences to cover the whole range of low incomes that more talented (but work-averse) individuals might choose, and (ii) focusing on minimal taxes, which under the preference diversity assumption espouse the envelope of the unskilled agents' upper contour sets. One can then use the budget curve (on the low-income bracket) as an indifference curve and find out the lowest implicit budget for the whole population. With this approach, knowing only the tax code and the level of the lowest skill is sufficient to compute social welfare.

Here the two key ingredients of this strategy fail, both because of (1). First, minimal taxes are

[14] As recalled in Saez and Stantcheva [32], the evaluation of small reforms only requires knowing the social marginal utility of money of subgroups of individuals being in the same observable situation (y, m). With a maximin criterion, one might hope that only one such individual situation will have full priority, but this cannot be guaranteed. The methodology we follow enables us to avoid this difficulty and to deal with large reforms.

not common in our model, as explained in the previous subsection. Second, diversity of preferences in the disadvantaged population may not enable them to cover the whole range of situations that could be chosen by less disadvantaged individuals (and are attainable to the disadvantaged). Indeed, the dominated parts of the disadvantaged individuals' budget may be undominated for more talented individuals (as will be explained in more detail below). It is then possible for more talented individuals to be amongst the worst-off, which is very hard to ascertain without using detailed information about the distribution of types.

We therefore explore how to bracket (instead of exactly measuring) the level of $\min_i IT_i(z_i)$ with minimal information about the population profile. We first focus on finding a lower bound.

Let us consider agent i's budget set, modified by the tax function T, in the (c,l,h) space. As explained above, the upper frontier of $B_i(T)$ may contain dominated parts, where increasing h entails an increase in productivity that pays more than it costs. It is therefore better to focus on the undominated parts of the budget set since this gives a more accurate picture of well-being opportunities. More precisely, let us define a new budget set which flattens the dominated parts of $B_i(T)$. For an arbitrary function $f(h)$, let $f^+(h)$ be the lowest non-increasing cover of f, i.e., the lowest function that is non-increasing and never below f. For a given $l \in [0,1]$ and T, let

$$b_{ilT}(h) = w_i(h)\,l - m_i(h) - T\left(w_i(h)\,l, m_i(h)\right).$$

The new (flattened) budget set, denoted $B_i^+(T)$, is defined as the set of (c,l,h) bundles such that $c \leq b_{ilT}^+(h)$. This construction is illustrated in figure 1, where the thin line depicts, for a given l, the increasing part of $b_{ilT}(h)$ and the thick line depicts $b_{ilT}^+(h)$.

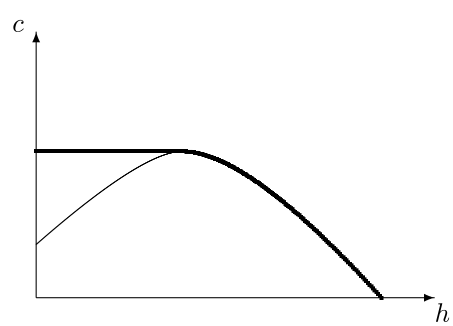

Figure 1: Budget set for given l

Let $B_\cap^+(T) = \bigcap_{i\in N} B_i^+(T)$. This is the intersection of all individuals' flattened budget set. The key property of this set is that every individual $i \in N$ is at least as well off as choosing from $B_\cap^+(T)$. This is because $B_\cap^+(T) \subset B_i^+(T)$, and even though $B_i^+(T)$ is larger than i's actual budget set $B_i(T)$, the individual cannot be better off with $B_i^+(T)$ since the flattened parts that make it larger than $B_i(T)$ cannot bring greater satisfaction for monotonic preferences.

Finally, consider the budget of the average type that would be obtained under a lump-sum transfer $t_0 \in \mathbb{R}$:

$$B(t_0, \overline{w}(.), \overline{m}(.)) = \{(c, l, h) \in \mathbb{R}_+ \times [0,1] \times [0,1] \mid c + \overline{m}(h_i) \leq t_0 + \overline{w}(h)l\}.$$

This is the hypothetical budget set used by our social welfare function to measure individual well being. For t_0 small enough (possibly negative), this budget $B(t_0, \overline{w}(.), \overline{m}(.))$ is contained in $B_\cap^+(T)$. Let t_0^* be the maximum level at which this property is satisfied (as illustrated in figure 2 for some $l \in [0,1]$).[15]

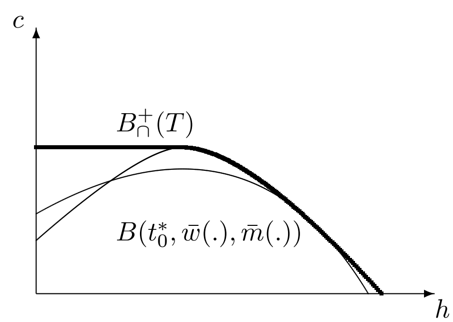

Figure 2: Budget tangency for some l

Individuals, given their preferences, choose their bundle on the budget set $B_i(T)$ modified by the tax function (see the indifference curves depicted in figure 3). Clearly the indifference surface passing through each bundle lies nowhere below $B_i^+(T)$, therefore nowhere below $B_\cap^+(T)$. Therefore, by construction, the allocation generated by the tax function T grants every agent an $IT_i(z_i)$ level no lower than t_0^*. This observation gives us a lower bound for $\min_i IT_i(z_i)$.[16]

Proposition 1 *Let z be an incentive-compatible allocation generated by the tax function T. Then*

$$t_0^* \leq \min_i IT_i(z_i),$$

where $t_0^* = \sup\left\{t_0 \mid B(t_0, \overline{w}(.), \overline{m}(.)) \subset B_\cap^+(T)\right\}$.

This result is not very satisfactory, because it requires substantial information about the profile, in order to estimate $B_\cap^+(T)$.[17] However, a much more limited amount of information is needed if one assumes that there is an unambiguously most disadvantaged type in the profile, just

[15] This maximum level is well defined because both $B_\cap^+(T)$ and $B(t_0^*, \overline{w}(.), \overline{m}(.))$ are compact, and the latter varies continuously with t_0.

[16] Observe on the figures that the boundary of $B(t_0^*, \overline{w}(.), \overline{m}(.))$ crosses the non-flattened part of $B_\cap^+(T)$ (i.e., $B_i(T)$). This explains why it is important to work with $B_i^+(T)$ rather than $B_i(T)$ in order to obtain a tighter lower bound.

[17] It also requires knowing the average functions $\overline{w}(.), \overline{m}(.)$, but this seems unavoidable given the social criterion.

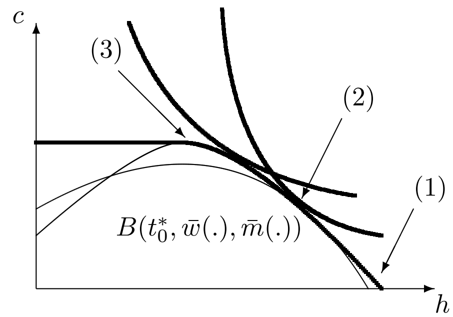

Figure 3: Individuals choosing a particular l

as in Mirrlees' model there is a lowest skill level. Here the greatest disadvantage is to have a lower wage rate and greater human capital expenditures than anyone else, at all levels of h.

Assumption 1 (Worst Type): There is a nonempty subset $P \subset N$ and two functions $\underline{w}(.), \underline{m}(.)$, such that for all $i \in P$, all $j \in N$, $w_i(h) = \underline{w}(h) \leq w_j(h)$ and $m_i(h) = \underline{m}(h) \geq m_j(h)$ for all h.[18]

This assumption makes it much easier to compute t_0^*, for the following reason. For every $i \in N$,

$$w_i(h)\, l - m_i(h) - T(w_i(h)\, l, m_i(h)) \geq \underline{w}(h)\, l - \underline{m}(h) - T(\underline{w}(h)\, l, \underline{m}(h)),$$

because the expression $wl - m - T(wl, m)$ is non-decreasing in w and non-increasing in m. This means that the worst type has a budget set that is always included in the budget sets of all other types of agents. For $i \in P$, the budget $B_i^+(T)$ will be denoted $\underline{B}^+(T)$. For every $i \in N$, $B_i^+(T)$ includes $\underline{B}^+(T)$ because whenever for two arbitrary functions f and g one has $f \geq g$, then necessarily $f^+ \geq g^+$. Therefore $\underline{B}^+(T) = B_\cap^+(T)$.

One therefore obtains the following simpler way of computing a lower bound, as the needed information about the profile is then reduced to the worst and average functions $w(.)$ and $m(.)$.

Corollary 2 *Let z be an incentive-compatible allocation generated by the tax function T. Under the Worst Type assumption,*

$$t_0^* \leq \min_i IT_i(z_i),$$

where $t_0^* = \sup\left\{t_0 \mid B(t_0, \overline{w}(.), \overline{m}(.)) \subset \underline{B}^+(T)\right\}$.

We will retain the Worst Type assumption in the rest of this paper.

More can be said about the lower bound by computing the characteristics of the intersection point between the two budget sets $B(t_0, \overline{w}(.), \overline{m}(.))$ and $\underline{B}^+(T)$. Locating this point is important

[18] The notation $\underline{m}(h)$ may look strange for the greatest function in the profile, but the lower bar is the symbol for the worst-off individuals.

because a reform can improve the budget set by raising this portion of the budget surface.[19]

Let us first examine the lower bound that is obtained when the non-flattened part of $\underline{B}^{+}(T)$ is touched by $B(t_0^*, \overline{w}(.), \overline{m}(.))$. At a point jointly belonging to the upper boundaries of $B(t_0, \overline{w}(.), \overline{m}(.))$ and $\underline{B}(T)$, one has

$$\overline{w}(h)\,l - \overline{m}(h) + t_0 = \underline{w}(h)\,l - \underline{m}(h) - T(\underline{w}(h)\,l, \underline{m}(h)),$$

therefore

$$t_0 = -\left(\overline{w}(h) - \underline{w}(h)\right)l - \left(\underline{m}(h) - \overline{m}(h)\right) - T(\underline{w}(h)\,l, \underline{m}(h)).$$

The value of t_0^* corresponds to the minimum of this expression, i.e., the lowest value of t_0 such that the upper boundaries of the two budget sets have a non-empty intersection.

We focus on tax functions such that the net wage rate of the worst type is not greater than the average gross wage rate.

Assumption 2 (No Net Wage Reranking): $\underline{w}(h)\,l - T(\underline{w}(h)\,l, \underline{m}(h)) - \overline{w}(h)\,l$ is decreasing in l.

Under this assumption, the minimum of t_0 is attained for $l = 1$. Therefore, as in previous results with fair income tax, the hardworking individuals of the worst type are singled-out. But their wage rate depends on h, and it remains to determine what level of h will be targeted.

There are two possibilities when $B(t_0, \overline{w}(.), \overline{m}(.))$ touches $\underline{B}(T)$ (i.e., the non-flattened part of $\underline{B}^{+}(T)$). In the first, the lowest t_0 is obtained at a point where consumption is null, so that

$$\overline{w}(h) - \overline{m}(h) + t_0 = \underline{w}(h) - \underline{m}(h) - T(\underline{w}(h), \underline{m}(h)) = 0,$$

which corresponds to a situation of a worst-type individual working full time and spending the greatest affordable amount on human capital.

In the second possibility, the two budgets are tangent (in the h dimension), so that

$$\overline{w}'(h) - \overline{m}'(h) = (1 - T_y)\,\underline{w}'(h) - (1 + T_m)\,\underline{m}'(h),$$

where T_y and T_m denote partial derivatives. This corresponds to a situation in which the consumption impact of an additional unit of h is the same for an individual with average characteristics and no tax distortion as for an individual from P and subject to tax T (both individuals working full time).[20]

[19] Admittedly, we only have a lower bound for $\min_i IT_i(z_i)$. But a good reform will be detected as one that puts the lower bound of the post-reform allocation above the upper bound of the pre-reform allocation. Hence the need to raise the lower bound.

[20] One should also have a second-order condition

$$\begin{aligned} \overline{w}''(h) - \overline{m}''(h) \quad &\leq \quad (1 - T_y)\,\underline{w}''(h) - (1 + T_m)\,\underline{m}''(h) \\ &\quad -T_{yy}\underline{w}'(h) - T_{mm}\underline{m}'(h) - T_{ym}\left(\underline{w}'(h) + \underline{m}'(h)\right). \end{aligned}$$

Let us now examine the possibility that the tangency occurs on a flattened part of $\underline{B}^{+}(T)$. Under Assumption 2, it is still correct to focus on the case $l = 1$, because the flattening occurs in the direction of h, not l. Then tangency with $\underline{B}^{+}(T)$ occurs at a point such that

$$\overline{w}'(h) - \overline{m}'(h) = 0 \leq (1 - T_y)\, \underline{w}'(h) - (1 + T_m)\, \underline{m}'(h).$$

But this point is not particularly interesting, what is more interesting is that improving $\underline{B}^{+}(T)$ so as to raise t_0^* would require raising the point of the budget where $(1 - T_y)\, \underline{w}'(h) - (1 + T_m)\, \underline{m}'(h) = 0$, which is on the edge of the flattened part of the budget.

In conclusion, we obtain three possible cases, implying that social evaluation may focus on different parts of the budget of P individuals. All of these cases are about P individuals working full time, under Assumption 2. The first case is when the focus is on the P individuals spending the maximum affordable amount on h (and consuming nothing). The second case puts the focus on P individuals whose returns to investing in h are negative[21] and equal to $\overline{w}'(h) - \overline{m}'(h)$. The third case puts the focus on P individuals whose returns on h are null (they maximize their consumption). These three cases are marked (1),(2),(3) on figure 3 (the position of $B(t_0, \overline{w}(.), \overline{m}(.))$ on that figure illustrates case 2).

It remains to examine which of the three cases is more likely to occur, depending on T. Consider first the context in which $\overline{w}'(h) > (1 - T_y)\, \underline{w}'(h)$ for all h and in which marginal subsidies on m are low so that $\overline{m}'(h) < (1 + T_m)\, \underline{m}'(h)$ for all h. In such a situation, for all h,

$$\overline{w}'(h) - \overline{m}'(h) > (1 - T_y)\, \underline{w}'(h) - (1 + T_m)\, \underline{m}'(h),$$

which implies that case 1 is prevailing.

On the other hand, if T_m is close to -1 (full marginal reimbursement of m), then case 3 may occur, because $(1 - T_y)\, \underline{w}'(h) - (1 + T_m)\, \underline{m}'(h)$ may remain positive at the value of h such that $\overline{w}'(h) - \overline{m}'(h) = 0$.

Case 1 will also occur if T_m is close to (but bounded away from) -1 but $\underline{m}'(h)$ goes to infinity at a low level of h (because the worst type may have great difficulty in raising h), so that consumption goes to zero quickly after some level of h for which $\overline{w}'(h) - \overline{m}'(h)$ is still not very low (even though $\overline{m}'(h)$ also has to go to infinity at the same level of h).

It is unsurprisingly more difficult to identify assumptions under which the intermediate case 2 will occur.

Let us take stock. When the marginal return to h is always lower for a P individual under T than $\overline{w}'(h) - \overline{m}'(h)$ (average characteristics, no marginal tax), then the target area of the budget set, for reform, is the situation of hardworking P individuals spending the maximum possible on h.

[21] When returns on h are positive, this part of the budget is flattened in $\underline{B}^{+}(T)$.

This situation is particularly likely to prevail under an income tax $T(y)$ that ignores m, because the inequality

$$\overline{w}'(h) - \overline{m}'(h) > (1 - T_y)\, \underline{w}'(h) - \underline{m}'(h)$$

is very plausible. Then, the level of earnings that gets the greatest social priority is the greatest full-time wage accessible to P individuals. Note that this level is endogenous and varies with the tax function.

If, on the other hand, the subsidy rate for m is high, the social priority may all be on the situation of P individuals who do not maximize their h, but do less, including, in the extreme, those who do the least that is rational, i.e., maximize their consumption. Note that, again, this is an endogenous point on the budget of P individuals, since it depends on the tax function.

These results appear quite sensible. Full-time low earnings will always be the locus of social priority, but corresponding to low or high expenses in human capital depending on whether such expenses are strongly subsidized or not. Subsidies on human capital therefore reduce the level of earnings that get most priority.

4.3 An upper bound

Let us now seek an upper bound for $\min_i IT_i(z_i)$. From the definition of $IT_i(z_i)$, for each $i \in N$, at the allocation z,

$$IT_i(z_i) \leq c_i - \overline{w}(h_i)\, l_i + \overline{m}(h_i).$$

Therefore

$$\min_i IT_i(z_i) \leq \min_i \left[c_i - \overline{w}(h_i)\, l_i + \overline{m}(h_i)\right].$$

This is not an interesting upper bound because it requires the computation of the (c, l, h) bundle of all types of agents induced by T. This is almost the same information as is needed to compute the actual distribution of $IT_i(z_i)$ —the only information that is not needed is the indifference sets at each bundle.

It is possible to obtain an upper bound that only requires knowledge of the observable (c, y, m) bundles as well as the worst and average functions w and m. In order to do so, one can focus on the undominated parts of the P agents' budget set and assume that, at least locally, their preferences are sufficiently diverse so that all bundles (c_i, y_i, m_i) in this area could be chosen by some agents from P (they can also be chosen by some other agent). This is formulated in the following assumption.[22]

[22] This assumption is much weaker than similar assumptions made in Fleurbaey ([12],[13]). It only applies to the particular budget generated by the tax function under consideration. The reason we introduce this weaker assumption is that the stronger assumption is very unlikely to be satisfied due to (1), as explained in the beginning of the previous subsection.

Assumption 3 (Preference Diversity): For all $i \in N$, if (i) $y_i \leq \underline{w}^*(m_i)$ and (ii) $(1 - T_y(y_i, m_i))\,\underline{w}^{*\prime}(m_i)\,\frac{y_i}{\underline{w}^*(m_i)} \leq 1 + T_m(y_i, m_i)$, then there exists $j \in P$ such that $(c_i, y_i, m_i)\,R_j^*\,(c_j, y_j, m_j)$.

Condition (ii) means that (y_i, m_i) does not lie in a dominated part of a P agent's budget (at the given labor level $y_i/\underline{w}^*(m_i)$, consumption is decreasing in m). Note that by incentive compatibility, necessarily $(c_j, y_j, m_j)\,R_j^*\,(c_i, y_i, m_i)$, so that we could as well write $(c_i, y_i, m_i)\,I_j^*\,(c_j, y_j, m_j)$ in the assumption.

With this assumption, one can obtain an informationally parsimonious upper bound. Intuitively, the subset of bundles lying in the subset defined in Preference Diversity gives an upper bound (possibly greater than if all bundles were taken into account). The problem is that some such bundles may belong to agents who are not in P and whose characteristics could be figured out only by using knowledge of the full statistical distribution of the population characteristics. But under Preference Diversity, we know that these bundles could as well be chosen by some P agents, therefore one still gets an upper bound by proceeding as if all these bundles belonged to P agents.

For notational convenience, let Z_T^* denote the set of (c, y, m) individual bundles observed under T and satisfying conditions (i) and (ii) of Preference Diversity.

Proposition 3 *Let z be an incentive-compatible allocation generated by the tax function T. Under Worst Type and Preference Diversity,*

$$\min_i IT_i(z_i) \leq \min_{(c,y,m)\in Z_T^*} \left[c - \frac{\overline{w} \circ \underline{m}^{-1}(m)}{\underline{w}^*(m)} y + \overline{m} \circ \underline{m}^{-1}(m) \right].$$

Proof. One has

$$c_i - \overline{w}(h_i)\,l_i + \overline{m}(h_i) = c_i - \frac{\overline{w} \circ m_i^{-1}(m_i)}{w_i^*(m_i)} y_i + \overline{m} \circ m_i^{-1}(m_i).$$

Under Worst Type and Preference Diversity, for all observed bundles $(c_i, y_i, m_i) \in Z_T^*$, there is $j \in P$ such that $(c_i, y_i, m_i)\,I_j^*\,(c_j, y_j, m_j)$. For such j, one has

$$IT_j(z_j) \leq c_i - \frac{\overline{w} \circ \underline{m}^{-1}(m_i)}{\underline{w}^*(m_i)} y_i + \overline{m} \circ \underline{m}^{-1}(m_i).$$

Therefore

$$\min_{j\in P} IT_j(z_j) \leq \min_{(c_i,y_i,m_i)\in Z_T^*} c_i - \frac{\overline{w} \circ \underline{m}^{-1}(m_i)}{\underline{w}^*(m_i)} y_i + \overline{m} \circ \underline{m}^{-1}(m_i).$$

The conclusion follows from $\min_{i\in N} IT_j(z_j) \leq \min_{j\in P} IT_j(z_j)$. ■

This result reduces the amount of information needed for policy evaluation to data that are easily available to the policy-maker. Knowing the worst and average functions $w(.)$ and $m(.)$, it suffices to look at bundles in a well-defined area and it is not difficult to locate and exclude the dominated part of the worst-type budget.

In our results there is no guarantee that the worst-off individuals actually belong to the worst type. But this is plausible and it would not be difficult to make assumptions to this effect (e.g., assuming that all types of preference orderings R_i found in the population are represented in P). However, this would not provide any different bounds than those obtained here, it would only probably make them closer to $\min_i IT_i(z_i)$.

5 Optimal tax: the non-linear case

Describing the optimal non-linear tax policy is extremely hard when the individuals differ in many dimensions and their behavior unfolds in a three-dimensional space. We will focus on a specific aspect of the optimal tax which is nevertheless quite central in understanding the shape of the optimal policy. Our goal is to determine what sort of agent (type, behavior) will receive the greatest subsidy at the second-best optimum and what that implies for tax rates around this agent.

5.1 Earnings tax

Let us first briefly describe what can be derived from the previous section about the optimal tax $T(y)$ when m is not taxed or subsidized. Assuming that No Net Wage Reranking holds at the optimal tax (which is very plausible if the difference between $\overline{w}(h)$ and $\underline{w}(h)$ is large), and that t_0^* is a good approximation of social welfare, the priority will be put on the level of earnings equal to $\underline{w}(h^*)$ for h^* such that $\underline{w}(h^*) - T(\underline{w}(h^*)) = \underline{m}(h^*)$, i.e., such that consumption is null for a worst-type individual at this level of earnings. This level is endogenous, but once it is determined, the optimal tax is similar to the tax that maximizes the net income of the working poor as described in Fleurbaey and Maniquet [13] and Saez and Stantcheva [32].[23] The only difference is, here, the endogeneity of the earnings level that receives the greatest subsidy, and we know that it is the level of full-time earnings of a worst-type person with the greatest human capital expenses.

Therefore the introduction of partial responsibility for skills and for consumption-leisure preferences implies *increasing* the level of earnings that receives the greatest support, from the lowest full-time earnings in the population to the greatest full-time earnings of the worst-type subpopulation. One can also interpret this result as meaning that when human capital expenses are not subsidized, income is used as a proxy to subsidize them via a subsidy at the level of earnings of the worst-type individuals who have the greatest level of expenses.

[23] In the Mirrlees model (with exogenous human capital), the optimal allocation for a similar social ordering (egalitarian-equivalent with reference wage equal to the average) features a marginal rate that is non-positive on average over income below the lowest wage, with a greatest subsidy granted to the least skilled individuals working full time.

One might have thought that reducing individual responsibility for consumption-leisure preferences (by the introduction of h) might have possibly *reduced* the level of earnings that benefits from the greatest support, at least in some configurations. This, indeed, could occur here in the following fashion. Suppose that in the original model with fixed skills, the differences in preferences are partly due to hidden differences in human capital that do not affect skills but affect the amount of labor one is able to perform. By making these differences explicit in the current model, social priority is shifted toward the individuals with the greatest difficulties to acquire human capital, and therefore to individuals who work less than full time if a low human capital is really a barrier to working full time. Our assumption, in the previous paragraph, that t_0^* is a good approximation of social welfare is then no longer valid, because the budget surface (which goes up to full-time work) diverges from the indifference surface of the worst-off.

5.2 Combined tax-subsidy on earnings and expenses

We now examine optimal policy with a general tax function $T(y, m)$, which is a more difficult topic. The idea that the locus of priority identified in the previous section will receive the greatest support remains valid, but this locus can be more precisely specified for the optimal tax.

Let z^* be an optimal incentive-compatible allocation implemented by the optimal tax function $T^*(y, m)$.[24] The following proposition identifies a way to cut subsidies above a certain level without reducing the value of t_0^* (as defined in section 4.2) as a lower bound for well-being as measured by IT_i.

Proposition 4 *Assume Worst Type holds. Let $T^*(y, m)$ implement z^*, let t_0^* be the greatest value of t_0 such that $B(t_0, \overline{w}(.), \overline{m}(.)) \subset B_\cap^+(T^*)$, and let r^* be equal to the maximum of $\overline{w}(h) - \underline{w}(h) - \overline{m}(h) + \underline{m}(h)$ for $h \in [0, 1]$ such that $t_0^* + \overline{w}(h) - \overline{m}(h) \geq 0$. Then the tax function*

$$T^{**}(y, m) = \max\{T^*(y, m), -t_0^* - r^*\}$$

is feasible and satisfies

$$t_0^* \leq \min_i IT_i(z_i^{**}),$$

for any allocation z^{**} that T^{**} induces.

Proof. The new tax, $T^{**}(y, m)$, reduces the budget set by cutting all subsidies above $t_0^* + r^*$. Consider any $i \in N$ and the couple (y_i^*, m_i^*) chosen by agent i under T^*. If $T^*(y_i^*, m_i^*) \geq -(t_0^* + r^*)$, then z_i^* is still an attainable option under T^{**}, therefore it is still the best choice for i. If $T^*(y_i^*, m_i^*) < -(t_0^* + r^*)$, then z_i^* is no longer accessible to i, and at the new best bundle z_i^{**} chosen under T^{**}, $T^{**}(y_i^{**}, m_i^{**}) \geq -(t_0^* + r^*)$, therefore $T^{**}(y_i^{**}, m_i^{**}) > T^*(y_i^*, m_i^*)$. Hence, if

[24] We ignore the case in which m is not used for taxation, as this is clearly suboptimal when it is observed.

there are agents $i \in N$ for whom z_i^* is no longer accessible under T^{**}, they will chose a new bundle such that the new allocation z^{**} generates a surplus. In any case, z^{**} is by construction feasible.

Suppose that $B(t_0^*, \overline{w}(.), \overline{m}(.)) \subset B_\cap^+(T^{**})$. Then, by Prop. 1, $t_0^* \leq \min_i IT_i(z_i^{**})$. It is therefore sufficient to prove that $B(t_0^*, \overline{w}(.), \overline{m}(.)) \subset B_\cap^+(T^{**})$.

By Worst Type, $B_\cap^+(T^{**}) = \underline{B}^+(T^{**})$. By construction,

$$\underline{B}^+(T^{**}) = \underline{B}^+(T^*) \cap \{(c,l,h) \in Z \mid c \leq \underline{w}(h)\, l - \underline{m}(h) + t_0^* + r^*\}.$$

As $B(t_0^*, \overline{w}(.), \overline{m}(.)) \subset \underline{B}^+(T^*)$, it is sufficient to prove that

$$B(t_0^*, \overline{w}(.), \overline{m}(.)) \subset \{(c,l,h) \in Z \mid c \leq \underline{w}(h)\, l - \underline{m}(h) + t_0^* + r^*\}.$$

Suppose this does not hold. Then there is $(c,l,h) \in B(t_0^*, \overline{w}(.), \overline{m}(.))$ such that $c > \underline{w}(h)\, l - \underline{m}(h) + t_0^* + r^*$, implying

$$t_0^* + \overline{w}(h)l - \overline{m}(h) \geq c > \underline{w}(h)\, l - \underline{m}(h) + t_0^* + r^*,$$

therefore

$$r^* < \overline{w}(h)l - \underline{w}(h)\, l - \overline{m}(h) + \underline{m}(h).$$

The right-hand side is increasing in l, so one must have

$$r^* < \overline{w}(h) - \underline{w}(h) - \overline{m}(h) + \underline{m}(h).$$

In addition, as $(c,l,h) \in B(t_0^*, \overline{w}(.), \overline{m}(.))$,

$$0 \leq t_0^* + \overline{w}(h)l - \overline{m}(h) \leq t_0^* + \overline{w}(h) - \overline{m}(h).$$

One therefore obtains a contradiction with the definition of r^*. ■

The previous proposition proves that constructing T^{**} from T^* does not necessarily entail a large welfare loss in the sense that t_0^* remains a lower bound for the worst off at the allocation generated by both tax functions. The value of r^* is picked so as to maximize the cut while preserving t_0^*, i.e., while keeping the boundary of $\underline{B}^+(T^{**})$ above $B(t_0^*, \overline{w}(.), \overline{m}(.))$.

The following corollary identifies conditions under which the two tax functions are welfare equivalent, namely, the conditions under which T^{**} is optimal as well.

Corollary 5 *Under the conditions of Proposition 4, if* $t_0^* = \min_i IT_i(z_i^*)$*, then* $t_0^* = \min_i IT_i(z_i^{**})$ *and* z^* *is implemented by* T^{**}*.*

Proof. This derives from the fact that by construction, $IT_i(z_i^{**}) \leq IT_i(z_i^*)$ for all i, and by Proposition 4, $t_0^* \leq \min_i IT_i(z_i^{**})$. In the proof of Proposition 4 it was shown that if T^{**} cannot implement z^* (because z_i^* is no longer affordable for some i), then z^{**} generates a surplus. But if this is the case, it is possible to distribute the surplus so as to raise IT_i for every i.[25] This would contradict the fact that z^* is optimal and therefore maximizes $\min_i IT_i(z_i^*)$. ■

These results suggest that it is interesting to study T^{**}. Note that even if $t_0^* < \min_i IT_i(z_i^*)$, one has

$$t_0^* \leq \min_i IT_i(z_i^{**}) \leq \min_i IT_i(z_i^*),$$

so that if t_0^* is close to $\min_i IT_i(z_i^*)$, the allocation z^{**} is close to being optimal. Therefore, when looking at the optimal tax scheme, there is no loss, or a limited loss, of social welfare if one restricts one's attention to taxes that share the salient features of T^{**}. In what follows we describe some of these features.

What is interesting about T^{**} is that it generates a budget frontier $c = y - m - T^{**}(y,m)$ which lies between the hyperplane $c = y - m + t_0^* + r^*$ and the manifold defined by

$$y \leq \underline{w}^*(m) \quad \text{and } c = y\frac{\overline{w} \circ \underline{m}^{-1}(m)}{\underline{w}^*(m)} - \overline{m} \circ \underline{m}^{-1}(m) + t_0^*. \tag{2}$$

More formally, one has

$$y - m - T^{**}(y,m) \leq y - m + t_0^* + r^* \quad \text{for all } y, m$$

and

$$y - m - T^{**}(y,m) \geq y\frac{\overline{w} \circ \underline{m}^{-1}(m)}{\underline{w}^*(m)} - \overline{m} \circ \underline{m}^{-1}(m) + t_0^* \qquad \text{for all } y, m \text{ s.t. } y \leq \underline{w}^*(m)$$

The former inequality is a direct consequence of $T^{**}(y,m) \geq -(t_0^* + r^*)$; the latter is nothing but the translation, in the (c,y,m) space, of the fact that $B(t_0^*, \overline{w}(.), \overline{m}(.)) \subset \underline{B}^+(T^{**})$. Indeed, in the (c,l,h) space the equation defining the upper boundary of $B(t_0^*, \overline{w}(.), \overline{m}(.))$ is $c = \overline{w}(h)\, l - \overline{m}(l) + t_0^*$. Substituting $l = y/\underline{w}(h) \leq 1$ and $h = \underline{m}^{-1}(m)$ yields the manifold described by (2).

The intersection between the hyperplane and the manifold determines the sort of individual situation that receives the greatest subsidy (i.e., $t_0^* + r^*$). The intersection is determined by the equation

$$y - m + t_0^* + r^* = y\frac{\overline{w} \circ \underline{m}^{-1}(m)}{\underline{w}^*(m)} - \overline{m} \circ \underline{m}^{-1}(m) + t_0^*,$$

which can also be written as

$$r^* = \frac{y}{\underline{w}^*(m)}\left(\overline{w} \circ \underline{m}^{-1}(m) - \underline{w}^*(m)\right) + \underline{m} \circ \underline{m}^{-1}(m) - \overline{m} \circ \underline{m}^{-1}(m).$$

[25] Doing such a distribution while preserving incentive compatibility is not trivial. See [12] for a rigorous proof in the Mirrlees model. The argument can be extended to the present model, as the dimension of the other goods than c does not matter.

Now, by definition of r^*, we know that it is the greatest value of the right-hand side among pairs (y, m) such that $t_0^* + \overline{w} \circ \underline{m}^{-1}(m) - \overline{m} \circ \underline{m}^{-1}(m) \geq 0$.

This implies that $y = \underline{w}^*(m)$ (full-time work for a P individual spending m) and, as far as m is concerned, there are two possibilities. The first possibility is when the maximum is obtained at a point m^* satisfying

$$\overline{w}' \circ \underline{m}^{-1}(m^*) - \underline{w}' \circ \underline{m}^{-1}(m^*) - \overline{m}' \circ \underline{m}^{-1}(m^*) + \underline{m}' \circ \underline{m}^{-1}(m^*) = 0,$$

which is a situation in which the gross productivity of h is the same for worst and average characteristics,

$$\overline{w}'(h^*) - \overline{m}'(h^*) = \underline{w}'(h^*) - \underline{m}'(h^*),$$

for $h^* = \underline{m}^{-1}(m^*)$.

The second possibility is when the constraint bites and the intersection occurs at a point m^{**} such that $t_0^* + \overline{w} \circ \underline{m}^{-1}(m^{**}) - \overline{m} \circ \underline{m}^{-1}(m^{**}) = 0$. The latter case will be obtained in particular if $\overline{w}'(h) > \underline{w}'(h)$ and $\underline{m}'(h) \geq \overline{m}'(h)$ for all h.

In conclusion, the greatest subsidy is obtained by some agent who belongs to P and works full time and either spends m^{**} as defined above or has a null consumption because of great human capital expenditures (the latter case is similar to the result of the previous subsection). The other agents who belong to P and also work full time, but have lower human capital expenditures, face a non-negative marginal rate of subsidy for human capital expenditures (on average over this part of their budget), whereas those who have greater expenditures face a non-positive rate of subsidy on average. This is due to the fact that their budget set under T^{**} has to lie below the hyperplane at which the rate of subsidy is null. Note that when $t_0^* + \overline{w}(h^*) - \overline{m}(h^*) = 0$ there are no such agents because consumption is below zero beyond m^*.

Similarly, the agents (from P or not from P) who spend m^* and earn less than $\underline{w}(h^*)$ face on average over this range of earnings a non-positive marginal tax rate. Again, note that when $t_0^* + \overline{w}(h^*) - \overline{m}(h^*) = 0$ there are no such agents because consumption is below zero in this area. This shows that Fleurbaey and Maniquet's [13] result obtained for the Mirrlees model (with exogenous human capital), according to which the marginal rate is non-positive on average over income below the lowest wage, becomes elusive in our model, with a general tax function. It is confirmed only for agents with a certain m^*, and may vanish when these agents have a very low consumption.

6 Additional results

In this section, we briefly examine optimal linear taxes and the case in which not only m but also h is observable.

6.1 Optimal tax: the linear case

Linear taxes are interesting in view of the results by Maldonado [23] and Bovenberg and Jacobs [18], [6]. Jacobs and Bovenberg [18] show that when individuals do not care about h_i directly, and when $w_i^* = w_i \circ m_i^{-1}$ has the same constant elasticity for all i, then m is simply tax deductible (it is subsidized at the same rate as earnings are taxed), independently of the degree of inequality aversion in the social objective. We checked that their result remains true here.[26] In this section, we briefly show that with a more general framework but our specific social objective, one may obtain very different results, including the possibility that m should be taxed.

Assume that $T(y,m) = \tau y - \rho m - \theta$,where $\theta \in \mathbb{R}$ is a universal lump-sum grant while $\tau \in \mathbb{R}$ and $\rho \in \mathbb{R}$ are the parameters for marginal income tax rate and human capital subsidy rate. This implies that, for each agent $i \in N$, the budget set modified by the tax function is $c_i \leq \theta + (1-\tau)\, y_i - (1-\rho)\, m_i$.

The general budget constraint requires

$$n\theta + \rho \sum_i m_i(h_i) \leq \tau \sum_i w_i(h_i)\, l_i. \tag{3}$$

Let $\theta(\tau,\rho)$ denote the maximum θ compatible with the budget constraint for a given pair (τ,ρ). Plugging $\theta(\tau,\rho)$ into equation (3) one obtains a budget identity that is function of two parameters only (τ,ρ). That is,

$$n\theta(\tau,\rho) + \rho M(\theta(\tau,\rho),\tau,\rho) = \tau Y(\theta(\tau,\rho),\tau,\rho), \tag{4}$$

where

$$\begin{aligned} M(\theta,\tau,\rho) &= \sum_i m_i\left(h_i\left(\theta,\tau,\rho,w_i(.),m_i(.)\right)\right), \\ Y(\theta,\tau,\rho) &= \sum_i w_i\left(h_i\left(\theta,\tau,\rho,w_i(.),m_i(.)\right)\right) l_i\left(\theta,\tau,\rho,w_i(.),m_i(.)\right). \end{aligned}$$

The budget set of any agent $i \in N$ can then simply be rewritten as $c_i = (1-\tau)w_i^*(m_i)l_i - (1-\rho)m_i + \theta(\tau,\rho)$.

Given our maximin social objective,[27] it is a convenient simplification to assume that there is a sufficient diversity among the worst-off agents, so that the optimal tax policy will actually be the optimal tax for one of them, say, i_0. Let individual indirect utility be defined as:

$$v_i(\theta,\tau,\rho) = \max\left\{IT\left((c_i,l_i,h_i),R_i\right) \mid c_i \leq \theta + (1-\tau)\, w_i(h_i)\, l_i - (1-\rho)\, m_i(h_i)\right\},$$

and the behavior functions be denoted $(c_i(\theta,\tau,\rho), l_i(\theta,\tau,\rho), h_i(\theta,\tau,\rho))$. By the envelope theorem, at the bundle chosen by the agent one has $\frac{\partial v_i}{\partial \theta} = \frac{\partial}{\partial c_i} IT\left((c_i,l_i,h_i),R_i\right)$, $\frac{\partial v_i}{\partial \tau} = -y_i \frac{\partial v_i}{\partial \theta}$, $\frac{\partial v_i}{\partial \rho} = m_i \frac{\partial v_i}{\partial \theta}$.

[26] Their proof is valid for agents with heterogeneous preferences, even if their model assumes identical preferences.

[27] Truly enough, the objective is a leximin, but the optimal tax for the leximin must also be optimal for the maximin.

The optimal tax for $i_0 \in P$ satisfies

$$\frac{\partial v_{i_0}}{\partial \tau} + \frac{\partial v_{i_0}}{\partial \theta}\theta_\tau = (\theta_\tau - y_{i_0})\,\frac{\partial v_{i_0}}{\partial \theta} = 0$$

and

$$\frac{\partial v_{i_0}}{\partial \rho} + \frac{\partial v_{i_0}}{\partial \theta}\theta_\rho = (\theta_\rho + m_{i_0})\,\frac{\partial v_{i_0}}{\partial \theta} = 0,$$

implying $\theta_\tau = y_{i_0}$ and $\theta_\rho = -m_{i_0}$.

Following the methodology of Piketty and Saez [27] applied to the simplified case in which only one individual situation gets positive weight in the social objective, one can obtain formulas for τ and ρ. Consider the program

$$\max_{\tau,\rho} u_{i_0}\left((1-\tau)w^*_{i_0}(m_{i_0}(\tau,\rho))l_{i_0}(\tau,\rho) - (1-\rho)m_{i_0}(\tau,\rho) + \theta\,(\tau,\rho)\,, l_{i_0}(\tau,\rho), m^{-1}_{i_0}\,(m_{i_0}(\tau,\rho))\right)$$

where u_{i_0} is the direct utility function of agent $i \in N$. From the FOC of this program, using the envelope theorem, one derives the following formulas (notation: $Y_\theta = \partial Y/\partial\theta$, and so on):[28]

$$\tau = \frac{(Y - ny_{i_0})(M_\theta\theta_\rho + M_\rho) + (M - nm_{i_0})(M_\theta\theta_\tau + M_\tau)}{(M_\theta\theta_\tau + M_\tau)\,(Y_\theta\theta_\rho + Y_\rho) - (M_\theta\theta_\rho + M_\rho)\,(Y_\theta\theta_\tau + Y_\tau)}, \tag{5}$$

$$\rho = \frac{(Y - ny_{i_0})(Y_\theta\theta_\rho + Y_\rho) + (M - nm_{i_0})\,(Y_\theta\theta_\tau + Y_\tau)}{(M_\theta\theta_\tau + M_\tau)\,(Y_\theta\theta_\rho + Y_\rho) - (M_\theta\theta_\rho + M_\rho)\,(Y_\theta\theta_\tau + Y_\tau)}. \tag{6}$$

Substituting $\theta_\tau = y_{i_0}$ and $\theta_\rho = -m_{i_0}$, one obtains

$$\frac{\tau}{\rho} = \frac{(M_\theta y_{i_0} + M_\tau)\,(M/n - m_{i_0}) + (-M_\theta m_{i_0} + M_\rho)\,(Y/n - y_{i_0})}{(Y_\theta y_{i_0} + Y_\tau)\,(M/n - m_{i_0}) + (-Y_\theta m_{i_0} + Y_\rho)\,(Y/n - y_{i_0})}, \tag{7}$$

and, in the case of quasi-linear preferences (for which $Y_\theta = M_\theta = 0$):

$$\frac{\tau}{\rho} = \frac{M_\tau\,(M/n - m_{i_0}) + M_\rho\,(Y/n - y_{i_0})}{Y_\tau\,(M/n - m_{i_0}) + Y_\rho\,(Y/n - y_{i_0})}. \tag{8}$$

We will focus on the standard situation in which

$$Y_\rho \geq 0;\ \ M_\rho \geq 0$$

and, moreover,

$$Y_\tau \leq 0;\ \ M_\tau \leq 0,$$

The first two inequalities can be justified by the assumption that if ρ increases, the total human capital expenditure increases too. This makes agents more productive so that they (eventually) work more and earn more. The latter inequalities can be justified by the assumption that if τ increases, agents work less and earn less, this also reduces the payoff of human capital and

[28] Still following Piketty and Saez [27] one can also express the previous formulas in terms of elasticities. We omit this for brevity.

hence the human capital expenditure. Moreover, if income effects are not too strong, the previous inequalities also yield

$$\begin{aligned} M_\theta y_{i_0} + M_\tau &\leq 0 \leq -M_\theta m_{i_0} + M_\rho, \\ Y_\theta y_{i_0} + Y_\tau &\leq 0 \leq -Y_\theta m_{i_0} + Y_\rho. \end{aligned}$$

Finally, if $m_{i_0} > M/n$ and $y_{i_0} < Y/n$, then τ and ρ have the same sign, which will be positive if[29]

$$\left(M_\theta y_{i_0} + M_\tau\right)\left(-Y_\theta m_{i_0} + Y_\rho\right) > \left(-M_\theta m_{i_0} + M_\rho\right)\left(Y_\theta y_{i_0} + Y_\tau\right),$$

a plausible condition. In particular, in the quasi-linear case, this condition boils down to $M_\tau Y_\rho > M_\rho Y_\tau$, which is very likely to occur because Y should be more sensitive to τ than to ρ, whereas the opposite holds for M.

So, under the assumptions we have listed so far, what does ultimately determine the mix of income redistribution and human capital subsidies? If M_ρ is much greater than $|M_\tau|$ and $|Y_\tau|$ is much greater than Y_ρ, and if M_θ and $|Y_\theta|$ are sufficiently small, the prominent terms in (7) and (8) form the ratio

$$\frac{M_\rho \left(Y/n - y_{i_0}\right)}{Y_\tau \left(M/n - m_{i_0}\right)}, \tag{9}$$

and provide a simple message. The optimal ratio between τ and ρ is bigger if human capital expenditures react strongly to subsidies and (or) total earnings react mildly to tax (this reflects the incentive concern). Moreover this ratio also increases with the gap between the average earnings and the earnings of the worst-off agent while it decreases with the gap between average human capital expenditure and human capital expenditure of the worst-off (this reflects the inequality concern).

This simple message is refined by adding the other components of the ratios in (7) and (8). In particular, more income redistribution in the mix will be pushed by a greater sensitivity of M to income tax and a lower sensitivity of Y to human capital subsidies. These results also suggest that it may be optimal to tax human capital expenditures if the worst-off agents spend less than average in human capital or, alternatively, earn more than average. The former case does not appear unrealistic in the context of education.

As this analysis relies on endogenous variables, one should check that the configurations discussed above can actually occur. The following simulations illustrate it. We consider a simple economy consisting of eight equally sized subgroups, varying in three dimensions: preferences, earning ability (function), human capital disposition. Preferences are expressed by the following utility functions: either $u^f(c,l,h) = c + \sqrt{(1-l)\,h}$ or $u^s(c,l,h) = c + 1.5\sqrt{(1-l)\,h}$; earning abilities are either $w^f = \sqrt{h}$ or $w^s = 2\sqrt{h}$; health dispositions is either $m^s(h) = h^2$ or $m^f(h) = 2h^2$.

[29] This determines the sign of the denominator in (5) and (6).

The average earning ability is therefore $\overline{w}(h) = 1.5\sqrt{h}$ whilst the average human capital disposition is $\overline{m}(h) = 1.5h^2$. In this economy the optimal policy is approximately $\tau = .35$ and $\rho = -.13$, with $\theta = .21$. The worst-off type's (first type of preferences, lowest earning ability, worst human capital disposition) human capital expenditure is .09 units below the average. The particular feature of this example is that the worst-off agents have the type of preferences with lowest concern for leisure and for human capital.

It must be emphasized, however, that even in this kind of economy, lower-than-average human capital expenditures on behalf of the worst-off is not sufficient to induce an optimal tax (i.e., a negative subsidy) on such expenditures, because the other terms in (8) can dominate. If, compared to the previous example we consider preferences represented by the following utility functions $u^f(c,l,h) = c + \sqrt{(1-l)\,h}$ and $u^s(c,l,h) = c + 2\sqrt{(1-l)\,h}$ and the earning abilities $w^f(h) = 2\sqrt{h}$ and $w^s(h) = 4\sqrt{h}$ (while keeping the same human capital dispositions), then the optimal policy is approximately $\tau = .39$ and $\rho = .03$, with $\theta = .69$. The worst-off type's human capital expenditure is .22 units below the average but ρ is positive even if strikingly low. This is because the term $Y_\tau\,(M/n - m_i)$ (in (8)) is substantially negative, even though it ends up being counterbalanced by the positive term $Y_\rho\,(Y/n - y_i)$ because of the great gap $Y/n - y_i$.

6.2 Observable human capital

Let us now assume that h is observed, together with c, y and m. This amounts to saying that, for instance, when it comes to education, the policy maker can observe the diplomas an agent has. Alternatively one could think of health. In this case our assumption implies that the social planner can rely on the physicians' evaluation in order to assess agents' health status. In such an informational framework the incentive-compatibility constraint becomes: for all i, j,

$$(c_i, y_i, m_i)\, R_i^*\, (c_j, y_j, m_j) \text{ or } y_j > w_i^*\,(m_i) \text{ or } m_i\,(h_j) \neq m_j\,(h_j)\,.$$

As in the previous setting, agent i still has to receive an allocation that she prefers to the allocation received by agent j unless it is not possible for her to mimic agent j. This occurs either if $y_j > w_i^*\,(m_i)$ (exactly as in the previous framework) or if $m_i\,(h_j) \neq m_j\,(h_j)$. That is, agent i can pretend to have agent j's human capital disposition only if her human capital disposition function crosses j's function at $h = h_j$.[30] To simplify the analysis and better analyze the consequences of

[30] An alternative specification would allow agents to "inflate" their expenditures and pretend they have a worse m function than they really have. In this case the incentive-compatibility constraint would become:for all i, j,

$$(c_i, y_i, m_i)\, R_i^*\, (c_j, y_j, m_j) \text{ or } y_j > w_i^*\,(m_i) \text{ or } m_i\,(h_j) > m_j\,(h_j)\,.$$

This alternative setting would give some protection to agents with a better disposition. However the practical implications would not be very different since we rely on an egalitarian social welfare function anyway. Hence we stick to the setting presented in the main text which is simpler.

using an egalitarian social objective we introduce the following assumption:

Assumption 4 (Nested Types) For all $i, j \in N$, either $m_i(.) = m_j(.)$ or for all h, $m_i(h) \neq m_j(h)$ (except when $m_i(h) = m_j(h) = 0$).

This assumption allows us to partition the population into different subgroups of agents having the same human capital disposition. Let K denote the set of subgroups resulting from such a partition. The fact that the human capital level is observable entails that one can conceive a different tax policy $T_k(y, m)$ for each $k \in K$.

We also introduce a further assumption which is meant to rule out a strict relation between having a good earning ability and a good human capital disposition. Whatever the human capital disposition, there is always some agent with the worst earning ability belonging to such group. Correlation is however permitted.

Assumption 5 (Uniformity): For every $k = 1, ..., K$, there is i in subgroup k such that $w_i(h) = \underline{w}(h)$.

This assumption just rules out the possibility for the policy maker to conceive a tax scheme that is particularly harsh to some specific subgroup k just because she happens to know that no unskilled agents belong to that subgroup. Let P_k denote the subset of i from subgroup k such that $w_i = \underline{w}$. Let also $\underline{B}_k^+(T)$ denote the budget set of some agent belonging to P_k, for $k \in K$.

Consider the budget $B(t_k, \overline{w}(.), \overline{m}(.))$ of a hypothetical agent with average circumstances, under laissez-faire except for a lump-sum transfer $t_k \in \mathbb{R}$:

$$c \leq \overline{w}(h)\, l - \overline{m}(h) + t_k.$$

For any $k \in K$ and for t_k small enough (possibly negative), this budget $B(t_k, \overline{w}(.), \overline{m}(.))$ is contained in $\underline{B}_k^+(T)$. Let t_k^* be the maximum level at which this property is satisfied.

We are now able to bracket the value of $\min_i IT(z_i, R_i)$, as stated below.

Proposition 6 *Let z be an incentive-compatible allocation generated by the tax function T. Then*

$$\min_k t_k^* \leq \min_i IT_i(z_i) \leq \min_k \min_{i \in P_k} \left[c_i - \frac{\overline{w} \circ m_i^{-1}(m_i)}{\underline{w} \circ m_i^{-1}(m_i)} y_i + \overline{m} \circ m_i^{-1}(m_i) \right].$$

Proof. Proposition 1 implies that , for every k,

$$t_k^* \leq \min_{i \in k} IT_i(z_i) \leq \min_{i \in P_k} \left[c_i - \frac{\overline{w}_N \circ m_i^{-1}(m_i)}{\underline{w} \circ m_i^{-1}(m_i)} y_i + \overline{m}_N \circ m_i^{-1}(m_i) \right].$$

The conclusion then follows from the fact that

$$\min_{i \in N} IT_i(z_i) = \min_k \min_{i \in k} IT_i(z_i),$$

and from the fact that when for all $k \in K$, $a_k \leq x_k \leq b_k$, then

$$\min_k a_k \leq \min_k x_k \leq \min_k b_k.$$

■

Note that in every subgroup k, the m_i function is known and identical across agents belonging to the subgroup. With some preference diversity assumption one can easily simplify the upper bound so as to take all bundles such that $y_i \leq \underline{w} \circ m_i^{-1}(m_i)$, in a similar fashion as done earlier.

As far as optimal tax is concerned, the result of the previous section applies to every subgroup k separately. What is new is that an optimal tax will equalize $\min_{i \in k} IT_i(z_i)$ across k. This is not the same as equalizing t_k^* across k, because in absence of preference diversity, one may have $t_k^* < \min_{i \in k} IT_i(z_i)$ for some k.

7 Conclusion

This paper proposes to extend the fair tax approach by letting individuals make choices that affect their productivity and by letting their preferences over consumption and labor be influenced by their human capital. The goal for redistribution is then to eliminate inequalities due to inter-individual differences in the intrinsic cost to acquire human capital and in earning ability conditional on human capital, while respecting individual choices on labor and human capital.

As far as tax reform is concerned, we found that the policy maker should primarily be interested in the part of the budget set that is attainable by agents endowed with the worst personal circumstances. However, the worst-off agent, at any arbitrary incentive compatible allocation, need not be one of them. In typical circumstances, the part of the budget set that should be the focus of attention corresponds to the full time earnings of an agent from the worst type, at a level of human capital expenditures that can vary depending on the rate of subsidies on human capital expenditures —in absence of subsidies, it will be the greatest affordable amount.

As far as optimal tax is concerned, we looked both at linear and non linear tax schemes. The main difference between the two cases is that in the former human capital expenditure might be taxed while in the latter case human capital expenditures are subsidized on the margin, up to a level of expenditures defined in reference to the agents who receive the greatest absolute amount of subsidy. Earnings are subsidized at low levels, up to the full-time wage of worst-type individuals who spend a certain amount on human capital —in absence of human capital subsidies, this is the greatest affordable amount.

Several extensions of this analysis can be considered. First, our analysis has ignored risk in the production of human capital and in the returns to human capital on the labor market. However, we believe that our analysis covers the most relevant case of pure idiosyncratic risk, i.e., when the policy-maker is able to predict the distribution of individual situations. It is then more respectful of the individuals' preferences to take account of this distribution rather than just the individual ex ante prospects, because what the individuals care about is their final situation (Fleurbaey [11]).

Another extension would consider more than one dimension of human capital. While our model can be applied to education or health, it cannot be applied to both dimensions simultaneously, unless they are lumped together into a single human capital variable. The extension of the social ordering function to dimensional human capital is straightforward, but the application to tax evaluation is less obvious because two kinds of expenditures can then be distinguished by the tax function.

A key feature of our approach, which helps a lot in obtaining results in such a general model, is the absolute priority granted to the worst-off. Some readers may find that indexing well-being by money-metric utilities $IT_i(z_i^*)$ is sensible but resist the absolute priority. It would be interesting to see what happens to the results when a strong but finite degree of priority replaces the maximin criterion in the evaluation of taxes. This would imply paying attention to levels of income above the levels accessible to the worst type.

Finally, actual policies are segmented and specific tax-subsidy functions operate separately on income and human capital expenditures. Our analysis of reform evaluation, fortunately, carries over to this case which is a subclass of the arbitrary tax functions studied here. The analysis of optimal linear tax, by construction, happens to satisfy this separation property. But such is not the case for optimal non-linear taxation. The methodology of Proposition 4 cannot be applied to separate taxes on earnings and human capital expenditures, because for an optimal tax function T^* that is additively separable in earnings and human capital expenditures, the new tax function T^{**} that cuts all subsidies above a fixed level loses this property.

References

[1] A. B. Atkinson, J.E. Stiglitz, The design of tax structure: direct versus indirect taxation, Journal of Public Economics, 6(1976), 55-75.

[2] A. Blomqvist, H. Horn, Public Health Insurance and Optimal Income Taxation, Journal of Public Economics, 24(1984), 352-371.

[3] R. Boadway, Recent Advances in Optimal Income Taxation, Hacienda Publica Espanola, 200(2012), 15-39.

[4] R. Boadway, M. Marchand, P. Pestieau, M. Racionero, Optimal Redistribution with Heterogeneous Preferences for Leisure, Journal of Public Economic Theory, 4(2002), 475-498.

[5] R. Boadway, P. Pestieau, Indirect Taxation and Redistribution: The Scope of the Atkinson-Stiglitz Theorem, in Richard Arnott, Bruce Greenwald, Ravi Kanbur and Barry Nalebuff (eds.), Economics for an Imperfect World: Essays in Honor of Joseph E. Stiglitz, MIT Press (2003).

[6] A. L. Bovenberg, B. Jacobs, Redistribution and Education Subsidies are Siamese twins, Journal of Public Economics, 89(2005) 2005-2035

[7] H. Cremer, P. Pestieau, Redistributive Taxation and Social Insurance, International Tax and Public Finance, 3(1996), 281-295.

[8] K. Christensen, N. V. Holm, M. Mcgue, L. Corder, J. W. Vaupel, A Danish Population-Based Twin Study on General Health in the Elderly, Journal of Aging and Health, 11(1999), 49-64.

[9] J. Currie, Inequality at Birth: Some Causes and Consequences, American Economic Review, 101(2011), 1-22.

[10] M.S. Feldstein, On the theory of tax reform, Journal of Public Economics 6(1976), 77-104.

[11] M. Fleurbaey, Assessing risky social situations, Journal of Political Economy, 118(2010), 649-680.

[12] M. Fleurbaey, F. Maniquet, Fair Income Tax, Review of Economic Studies, 73(2006), 55-83.

[13] M. Fleurbaey, F. Maniquet, Help the low-skilled or let the hardworking thrive? A study of fairness in optimal income taxation, Journal of Public Economic Theory 9(2007) 467-500.

[14] M. Fleurbaey, F. Maniquet, A Theory of Fairness and Social Welfare, Cambridge University Press (2011).

[15] M. Grossman, L. Benham, Health, Hours and Wages, M. Perlman (ed.), The Economics of Health and Medical Care. Macmillan and Co., London (1974).

[16] R. Guesnerie, A contribution to the pure theory of taxation, Cambridge University Press (1995).

[17] D. Henriet, J.C. Rochet, Is Public Health Insurance an Appropriate Instrument for Redistribution?, Annales d'economie et de statistique, 83-84(2006), 61-88.

[18] B. Jacobs, A. L. Bovenberg, Optimal Taxation of Human Capital and the Earnings Function, CESifo Working Paper No. 2250.

[19] B. Jacobs, A. L. Bovenberg, Optimal Taxation of Human Capital and the Earnings Function, Journal of Public Economic Theory 13(2011), 957-971.

[20] L. Jacquet, D. Van de gaer, A comparison of optimal tax policies when compensation and responsibility matter, Journal of Public Economics 95 (2011), 1248-1262.

[21] D. Krueger, A. Ludwig, Optimal Progressive Taxation and Education Subsidies in a Model of Endogenous Human Capital Formation, *mimeo*.

[22] H. Luft, The Impact of Poor Health on Earnings. Review of Economics and Statistics, 57 (1975), 43-57.

[23] D. Maldonado, Education Policies and Optimal Taxation, International Tax and Public Finance, 15(2008), 131-143.

[24] M.G. Marmot, G. Davey Smith, S. Stansfeld, C. Patel, F. North, J. Head, I. White, E. Brunner, A. Feeney, Health inequalities among British civil servants: the Whitehall II study, Lancet 337(1991), 1387-1393.

[25] J. Mirrlees, An exploration in the theory of optimum income taxation, Review of Economic Studies 38(1971), 175-208.

[26] S. Mushkin, Health as an Investment, Journal of Political Economy 70(1962), 129-157.

[27] T. Piketty, E. Saez, Optimal Labor Income Taxation, Handbook of Public Economics V (2013), 391-474 .

[28] D.D. Reid, G.Z. Brett, P.J.S. Hamilton, R.J. Jarrett, H. Keen, G. Rose, Cardio-respiratory disease and diabetes among middle-aged male civil servants: a study of screening and intervention. Lancet 303(1974), 469-473.

[29] J.C. Rochet, The taxation principle and multi-time Hamilton-Jacobi equations, Journal of Mathematical Economics, 14(1985), 113-128.

[30] J.C. Rochet, Incentives, Redistribution and Social Insurance, Geneva Papers on Risk and Insurance Theory, 16(1991), 143-165.

[31] J.E. Roemer, On several approaches to equality of opportunity, Economics and Philosophy 28(2012), 165-200.

[32] E. Saez, S. Stantcheva, Generalized social marginal welfare weights for optimal tax theory, mimeo (2013).

[33] G. Valletta, Health, fairness and taxation, Social Choice and Welfare, forthcoming

[34] M. Weinzierl, The Promise of Positive Optimal Taxation, Harvard Business School Working Paper, No. 14-013 (2013).