CESifo WORKING PAPERS

8871
2021

February 2021

Optimal Tax Problems with Multidimensional Heterogeneity: A Mechanism Design Approach

Laurence Jacquet, Etienne Lehmann

CESifo

Impressum:

CESifo Working Papers
ISSN 2364-1428 (electronic version)
Publisher and distributor: Munich Society for the Promotion of Economic Research - CESifo GmbH
The international platform of Ludwigs-Maximilians University's Center for Economic Studies and the ifo Institute
Poschingerstr. 5, 81679 Munich, Germany
Telephone +49 (0)89 2180-2740, Telefax +49 (0)89 2180-17845, email office@cesifo.de
Editor: Clemens Fuest
https://www.cesifo.org/en/wp
An electronic version of the paper may be downloaded
· from the SSRN website: www.SSRN.com
· from the RePEc website: www.RePEc.org
· from the CESifo website: https://www.cesifo.org/en/wp

CESifo *Working Paper No. 8871*

Optimal Tax Problems with Multidimensional Heterogeneity: A Mechanism Design Approach

Abstract

We propose a new method, that we call an allocation perturbation, to derive the optimal nonlinear income tax schedules with multidimensional individual characteristics on which taxes cannot be conditioned. It is well established that, when individuals differ in terms of preferences on top of their skills, optimal marginal tax rates can be negative. In contrast, we show that with heterogeneous behavioral responses and skills, one has optimal positive marginal tax rates, under utilitarian preferences and maximin.

Keywords: optimal taxation, mechanism design, multidimensional screening problems, allocation perturbation.

Laurence Jacquet
THEMA - CY Cergy Paris Universite
Cergy-Pontoise Cedex / France
laurence.jacquet@u-cergy.fr

Etienne Lehmann
CRED (TEPP) University Pantheon
Assas Paris II / France
etienne.lehmann@gmail.com

We thank Craig Brett and two anonymous refererees, Pierre Boyer, Vidar Christiansen, Guy Laroque, Emmanuel Saez, Laurent Simula, Stefanie Stantcheva, Kevin Spiritus, Alain Trannoy and Nicolas Werquin. This research was partly realized while Laurence Jacquet was research associate at Oslo Fiscal Studies. She gratefully acknowledges support by Labex MME-DII. The usual disclaimer applies.

I Introduction

Since Mirrlees (1971), the mechanism design method has been largely used to obtain optimal tax formulas when individuals differ along a single dimension of heterogeneity, typically skills, and when income taxation cannot be conditioned on this dimension of heterogeneity.[1] In this paper, we extend Mirrlees (1971)'s mechanism design method to a situation where taxpayers differ along many characteristics that are unobserved or on which taxation cannot be conditioned.[2] For instance, taxpayers with the same income level can differ in terms of unobserved preferences for work, gender, number and age of the children, cultural background, geographical area where one lives, ethnicity or age. Taxpayers can then exhibit heterogeneous labor supply elasticities in addition to distinct skill levels.

In this paper, we develop a new method, that we call "allocation perturbation", to solve multidimensional screening problems. We show that individuals of different groups pooled at the same income level are characterized by the same marginal rate of substitution between pre-tax and after-tax income. Intuitively, individuals of distinct groups who earn the same income level face the same marginal tax rate. From the individual maximization program, we know that identical marginal tax rates imply identical marginal rates of substitution. Using this equality in marginal rates of substitution together with a single-crossing condition within each group, we can fully characterize an incentive-compatible allocation. We then apply variational calculus to small perturbations of this allocation and derive the necessary conditions of the optimal allocation. To do so, we take the Gâteaux derivatives of the associated Lagrangian and equate them to zero. Rearranging terms of these conditions gives the optimal tax formula.

In several recent papers, tax perturbation methods are proposed to solve multidimensional screening problems (e.g., Hendren (2020) and Sachs et al. (2020) and Jacquet and Lehmann (2020) for the most recent). In contrast to our approach, all these papers are built upon the tax perturbation approach initiated by Piketty (1997) and Saez (2001) which consists in computing all responses to small tax reforms, to sum them up and equate them to zero in order to obtain the optimal tax schedule.

Tax perturbation approaches easily allow one to obtain optimal tax formulas in terms of meaningful sufficient statistics. In contrast, we derive the optimal income tax schedule in terms of policy invariant functions (i.e. skill density and the derivatives of the individual and social utility functions) and rewrite it in terms of sufficient statistics. Just like the tax formula in terms of sufficient statistics, our formula in terms of policy invariant functions is not closed-form because it depends on the allocation where these functions are evaluated. We use it to sign the optimal marginal tax rates. With multidimensional heterogeneity, the literature has highlighted that negative marginal tax rates can be optimal. In Cuff (2000), Boadway et al.

[1]John Weymark has largely contributed to the literature on optimal income taxation, see e.g. Weymark (1986) and Weymark (1987) and Brett and Weymark (2011).

[2]Our paper studies the optimal tax system when individual characteristics, despite being observable by the tax authority, cannot be used as tags (Akerlof, 1978), due to legal and/or horizontal equity reasons.

(2002b), Brett and Weymark (2003), Choné and Laroque (2010) and Lockwood and Weinzierl (2015), individuals differ along their skills and preferences for leisure, and the social planner has weighted utilitarian preferences. In this context, individuals who pool at the same income level are characterized by different marginal social welfare weights. Therefore, the mean social welfare weight may not be decreasing with income as it is the case with one dimension of heterogeneity and e.g. utilitarian preferences. This opens the way to negative marginal tax rates. One may wonder whether heterogeneous labor supply elasticities for individuals who earn the same level of income could also open the way to negative marginal tax rates. We show that, under utilitarian preferences, marginal tax rates are positive. This result also prevails under maximin.

Our method can easily be extended to include participation decisions (Saez, 2002, Kleven et al., 2009, Jacquet et al., 2013), migration decisions (Lehmann et al., 2014, Blumkin et al., 2015) or sectoral decisions (Rothschild and Scheuer, 2013, Scheuer, 2014, Gomes et al., 2018), simply following Jacquet et al. (2013). The tax formulas then simply incorporate new terms implied by the participation margin.

The paper is organized as follows. In Section II, we present the model. In Section III, we characterize any incentive compatible allocation thanks to a monotonicity constraint, a differential equation and a pooling function. In Section IV, we detail the allocation perturbation method and derive the optimal tax schedule. In Section V, we give a sufficient condition for optimal marginal tax rates to be positive. Section VI concludes.

II Model

II.1 Individuals

Taxpayers differ along their skill level $w \in \mathbb{R}_+^*$ and along some characteristics denoted $\theta \in \Theta$. We call a *group* a subset of individuals with the same θ.[3] We assume that the set of groups Θ is measurable with a cumulative distribution function (CDF) denoted $\mu(\cdot)$. The set Θ can be finite or infinite and may be of any dimension and is compact. The distribution $\mu(.)$ of the population across the different groups may be continuous, but it may also exhibit mass points. Among individuals of the same group θ, skills are distributed according to the conditional skill density $f(\cdot|\theta)$ which is positive and differentiable over the support $\mathbb{R}_+^*$. The conditional CDF is denoted $F(w|\theta) \stackrel{\text{def}}{\equiv} \int_0^w f(x|\theta)dx$. We do not make any restriction on the correlation between w or θ. We normalize to unity the total size of the population.

Every worker derives utility from consumption $c \in \mathbb{R}_+$ and disutility from effort. Effort captures the quantity as well as the intensity of labor supply. More effort implies higher pre-tax income $y \in \mathbb{R}_+$ (for short, income hereafter). The government levies a tax $T(.)$ which depends on income y only. Consumption c is related to income y through the tax function $T(y)$ according to $c = y - T(y)$.

[3] Our definition of "group" is identical to the one in Werning (2007, p.13).

Assumption 1. *The utility function is additively separable and takes the form:*

$$\mathscr{U}(c, y; w, \theta) = u(c) - v(y; w, \theta) \qquad \textit{with}: \quad u', v_y, v_{y,y} > 0 > v_w \quad , \quad u'' \leq 0$$

The convexity of the indifference curve is ensured by assuming that $v_{y,y} > 0 \geq u''$. Additive separability is standardly used in the adverse selection literature with multidimensional heterogeneity, e.g., Rochet (1985), Wilson (1993), Rochet and Choné (1998), Rochet and Stole (2002). As in the model with one-dimensional heterogeneity (Mirrlees, 1971), this assumption is necessary to sign optimal marginal tax rates (Proposition 3). Note that additive separability can be seen as a rather restrictive assumption, in a multidimensional context, since the group θ matters only for the marginal disutility of income but not for the marginal utility of consumption.

A worker of type (w, θ), facing $y \mapsto T(y)$, solves:

$$U(w, \theta) \stackrel{\text{def}}{\equiv} \max_{y} \quad u(y - T(y)) - v(y; w, \theta) \tag{1}$$

We call $Y(w, \theta)$ the solution to program (1), $C(w, \theta) = Y(w, \theta) - T(Y(w, \theta))$ the consumption of a worker of type (w, θ) and $U(w, \theta)$ her gross utility (or maximized level of utility). When the tax function is differentiable, the first-order condition associated to (1) implies that:

$$1 - T'\left(Y\left(w, \theta\right)\right) = \mathscr{M}\left(C\left(w, \theta\right), Y\left(w, \theta\right); w, \theta\right) \tag{2}$$

where:

$$\mathscr{M}(c, y; w, \theta) \stackrel{\text{def}}{\equiv} \frac{v_y(y; w, \theta)}{u'(c)} \tag{3}$$

denotes the marginal rate of substitution between income and consumption. For a worker of a given type, the left-hand side of Equation (2) corresponds to the marginal gain of income after taxation while the right-hand side corresponds to the marginal cost of income in monetary terms.

We impose the single-crossing (Spence-Mirrlees) condition that, within each group of workers endowed with the same θ, the marginal rate of substitution is a decreasing function of the skill level, i.e. that higher-skilled workers find it less costly to increase their income y:

Assumption 2 (Within-group single-crossing condition)**.** *For each $\theta \in \Theta$, and each $(c, y) \in \mathbb{R}_+ \times \mathbb{R}_+$, function $w \mapsto \mathscr{M}(c, y; w, \theta)$ is differentiable with $\forall w \in \mathbb{R}_+^*$, $\mathscr{M}_w < 0$ i.e. $v_{y,w} < 0$, $\lim_{w \mapsto 0} \mathscr{M}(c, y; w, \theta) = +\infty$ and $\lim_{w \mapsto \infty} \mathscr{M}(c, y; w, \theta) = 0$, i.e. $\lim_{w \mapsto 0} v_{y,w}(y; w, \theta) = +\infty$ and $\lim_{w \mapsto \infty} v_{y,w}(y; w, \theta) = 0$.*

II.2 Government

The government's problem consists in finding the tax schedule $T(\cdot)$ that maximizes a social welfare function,

$$\iint_{\theta \in \Theta, w \in \mathbb{R}_+^*} \Phi\left(U(w, \theta); w, \theta\right) \; f(w|\theta) dw \; d\mu(\theta), \tag{4}$$

where $\Phi(U; w, \theta)$ is an increasing transformation of individuals' utility levels U. This welfarist specification allows Φ to vary with type (w, θ) which makes it very general. *Weighted utilitarian* preferences (as in Weymark (1987)) are obtained with $\Phi(U; w, \theta) \equiv \varphi(w, \theta) \cdot U$ with weights $\varphi(w, \theta)$ depending on individual characteristics. The objective is *utilitarist* if $\varphi(w, \theta)$ is constant and $\Phi(U; w, \theta) \equiv U$ and it turns out to be *maximin (or Rawlsian)* if $\varphi(w, \theta) = 0 \ \forall w > 0$. When $\Phi(U; w, \theta)$ does not vary with its two last arguments and is concave in individual utility ($\Phi_{UU} \leq 0$), we obtain a Bergson-Samuelson criterion which is a concave transformation of utility.

When maximizing social welfare, the government takes into account the individual labor supply decisions (1) and the budget constraint,

$$\iint_{\theta \in \Theta, w \in \mathbb{R}_+^*} T\left(Y(w, \theta)\right) \, f(w|\theta) dw \, d\mu(\theta) = E, \tag{5}$$

where $E \geq 0$ is an exogenous amount of public expenditures. Let $\lambda > 0$ denote the shadow price of public funds.

III Incentive constraints and pooling function

According to the taxation principle (Hammond, 1979, Guesnerie, 1995), it is equivalent for the government to select a nonlinear tax schedule taking into account labor supply decisions described in (1), or to directly select an allocation $(w, \theta) \mapsto (C(w, \theta), Y(w, \theta))$ that verifies the incentive constraints:

$$\forall w, \theta, w', \theta' \in (\mathbb{R}_+^* \times \Theta)^2 \qquad \mathscr{U}\left(C(w, \theta), Y(w, \theta); w, \theta\right) \geq \mathscr{U}\left(C(w', \theta'), Y(w', \theta'); w, \theta\right). \tag{6}$$

The incentive constraints (6) impose that workers of type (w, θ) prefer the bundle designed for them rather than the bundles $(C(w', \theta'), Y(w', \theta'))$ designed for individuals of any other type (w', θ').

With multidimensional heterogeneity, incentive constraints (6) can be rewritten as incentive constraints within each group θ and incentive constraints across distinct groups. In a first subsection, we show that the within-group incentive constraints can be reformulated as a monotonicity constraint and a differential equation, as in the model with one dimension of heterogeneity. In a second subsection, we show that, across groups, to guarantee incentive compatibility, taxpayers from distinct θ-groups who earn the same level of income need to face the same marginal rate of substitution between consumption and income. To satisfy this equalization of marginal rates of substitution across groups, we define a pooling function that describes the level of skill required in each θ group to obtain a given level of income. We show that we have an incentive-compatible allocation under the additional condition that, for each θ, the pooling function be smoothly increasing in skill.

III.1 Incentive constraints within groups

An incentive-compatible allocation has to satisfy (6). It thus has to verify for each group θ the following set of "within-group incentive constraints".

$$\forall (w, \widetilde{w}, \theta) \in \mathbb{R}_+^2 \times \Theta \quad \mathscr{U}\left(C(w,\theta), Y(w,\theta); w, \theta\right) \geq \mathscr{U}\left(C(\widetilde{w},\theta), Y(\widetilde{w},\theta); w, \theta\right). \tag{7}$$

Under the within-group single-crossing assumption 2, the set of within-group incentive constraints can be transformed into a monotonicity constraint and a differential equation that we give in Lemmas 1 and 2 below.

Lemma 1. *Under Assumption 2, the function $w \mapsto Y(w,\theta)$ is nondecreasing for each $\theta \in \Theta$.*

Assumption 2 (i.e. the within-group single-crossing assumption) implies that, in the same group θ, the indifference curves of workers with a lower skill are steeper than the ones of workers with a higher level of skill, as in one dimensional tax models. We therefore skip the proof of Lemma 1 which is a simple reformulation of the usual proof in the one dimensional framework. Note that $Y(\cdot;\theta)$ being nondecreasing, it may exhibit discontinuities over a countable set and it may also exhibit bunching if individuals in the same group but endowed with different skill levels earn the same income. We follow a first-order approach and consider only smooth allocations where these two pathologies do not arise.To do so, we make a smoothness assumption. As a preamble, we define smoothly increasing functions[4] and then give the smoothness assumption.

Definition 1. *A function $a : \mathbb{R}_+ \mapsto \mathbb{R}$ is "smoothly increasing"if it is differentiable with $\forall x \in \mathbb{R}_+$, $a'(x) > 0$, $a'(0) = 0$ and $\lim_{x \mapsto \infty} a'(x) = +\infty$.*

Assumption 3 (Smooth allocations)**.** *In each group θ, $w \mapsto Y(w,\theta)$ is smoothly increasing.*

According to Assumption 3,[5] for each income level $y \in \mathbb{R}^+$ and for each group $\theta \in \Theta$, there exists a single skill level w such that only individuals of that skill level within group θ earn income $y = Y(w,\theta)$. The following lemma provides the first-order incentive constraints within group θ reformulated as a differential equation, where the dot above a variable stands for the partial derivative of this variable with respect to skill w.

Lemma 2. *Under Assumptions 1 and 3, for each θ, the mapping $w \mapsto U(w,\theta)$ is differentiable with:*

$$\dot{U}(w,\theta) = -v_w\left(Y(w,\theta); w, \theta\right). \tag{8a}$$

Moreover, Equation (8a) is equivalent to:

$$\frac{\dot{C}(w,\theta)}{\dot{Y}(w,\theta)} = \mathscr{M}\left(C(w,\theta), Y(w,\theta); w, \theta\right). \tag{8b}$$

[4] A smoothly increasing (decreasing) function is also called an increasing (decreasing) diffeomorphism for which the derivative maps the positive real line onto itself.

[5] In Jacquet and Lehmann (2020, Proposition 5), we show that the assumption of a smoothly-increasing-in-types allocation amounts to assuming: (i) twice differentiability of the tax function $T(\cdot)$, that (ii) for all $(w,\theta) \in \mathbb{R}_+^* \times \Theta$, the second-order condition associated to the individual maximization program holds strictly and that (iii) for all $(w,\theta) \in \mathbb{R}_+^* \times \Theta$, the function $y \mapsto \mathscr{U}(y - T(y), y; w, \theta)$ admits a unique global maximum over $\mathbb{R}_+$.

We skip the proof since it simply consists in adapting that of the one-dimensional tax models, which consists in applying the envelope theorem to Equation (1), see e.g., Salanié (2011). Integrating (8a) leads to:

$$U(w,\theta) = U(0,\theta) - \int_0^w v_w \left(Y(x,\theta); x, \theta\right) \, dx. \tag{8c}$$

The "first-order" approach we follow is usual in one-dimensional tax models (see e.g. Salanié (2011)), since it takes into account the first-order incentive-compatibility constraints (8a) and assumes the second-order incentive-compatibility constraints are satisfied (Assumption 3) as standardly observed on data.[6].

III.2 Incentive constraints across groups and pooling function

We now describe how the various within-group allocations $\omega \mapsto (Y(\omega,\theta), C(\omega,\theta))$ need to be set to be mutually incentive-compatible and to verify the full set of incentive constraints (6). This is the *pooling issue* that we now address.

Choose a reference group $\theta_0 \in \Theta$, a skill level w and another group θ. Individuals of type (w,θ_0) earn income $Y(w,\theta_0)$. According to the smoothness assumption 3, each group-specific allocation $Y(\cdot,\theta) : w \mapsto Y(w,\theta)$ is an increasing one-to-one function that maps the positive real line onto itself. Therefore, there must exist a single skill level, hereafter denoted $W(w,\theta)$, so that individuals of the other group θ endowed with that skill level $W(w,\theta)$ must get the same income level $Y(w,\theta_0)$ as individuals of type (w,θ_0), i.e. $Y(W(w,\theta),\theta) = Y(w,\theta_0)$. We call $W(.,.)$ the *pooling* function. For each $\theta \in \Theta$, the pooling function combines two smoothly increasing functions, namely $\omega \overset{Y(\cdot,\theta_0)}{\longmapsto} Y(\omega,\theta_0) \overset{Y^{-1}(\cdot,\theta)}{\longmapsto} W(\omega,\theta)$. The pooling function is therefore also a smoothly increasing function in skill w. It obviously verifies $W(w,\theta_0) \equiv w$. Provided that the allocation is incentive-compatible, it is not possible from (6) that individuals of type $(W(w,\theta),\theta)$ and individuals of type (w,θ_0) obtain the same income $Y(w,\theta_0)$ but distinct consumption levels. Therefore, for each (w,θ), we must simultaneously have

$$Y(W(w,\theta),\theta) \equiv Y(w,\theta_0) \qquad \text{and} \qquad C(W(w,\theta),\theta) \equiv C(w,\theta_0). \tag{9}$$

Individuals who earn the same income level face the same marginal tax rate, according to Equation (2), hence have the same consumption level. The skill levels of individuals who earn a given income level are implicitly determined by the equality of their marginal rates of substitution which is our pooling condition.

Lemma 3. *Under Assumptions 2 and 3, along an incentive-compatible allocation, the bundle designed for individuals of type $(W(w,\theta),\theta)$ coincides with the bundle $(C(w,\theta_0), Y(w,\theta_0))$ designed for individuals of type (w,θ_0), where $W(w,\theta)$ verifies the following* ***pooling condition****:*

$$\mathscr{M}\left(C(w,\theta_0), Y(w,\theta_0); w, \theta_0\right) = \mathscr{M}\left(C(w,\theta_0), Y(w,\theta_0); W(w,\theta), \theta\right). \tag{10}$$

[6] For instance, we never found cases where the second-order incentive-compatibility constraints were violated in the large set of simulations we run on US data with taxpayers differing in terms of gender and labor supply elasticities, see Jacquet and Lehmann (2020).

Proof According to Assumption 2, $\mathscr{M}\left(C(w,\theta_0), Y(w,\theta_0); w,\theta_0\right) = \mathscr{M}\left(C(w,\theta_0), Y(w,\theta_0); \omega,\theta\right)$ admits exactly one solution in ω. Differentiating in w both sides of equalities in (9) and rearranging terms leads to:

$$\frac{\dot{C}(W(w,\theta),\theta)}{\dot{Y}(W(w,\theta),\theta)} = \frac{\dot{C}(w,\theta_0)}{\dot{Y}(w,\theta_0)}.$$

According to Lemma 2, Equation (8b) holds, which implies (10). □

One can retrieve the entire incentive-compatible allocation for all groups if one knows the pooling function $W(\cdot,\cdot)$ and the allocation $\omega \mapsto (Y(\omega,\theta_0), C(\omega,\theta_0))$ designed for the reference group.

Thanks to the pooling condition, in the following lemma, we provide a sufficient condition for the allocation to be incentive-compatible. The proof is in Appendix A.1.

Lemma 4. *Under Assumption (2), let $w \mapsto (C(w,\theta_0), Y(w,\theta_0))$ be a within-group allocation that verifies Assumption 3 and the within-group incentive-compatible Equation (8b). For each $w \in \mathbb{R}_+$ and each group $\theta \in \Theta$, let $\underline{W}(w,\theta)$ be the unique skill level ω that solves the pooling condition $\mathscr{M}\left(C(w,\theta_0), Y(w,\theta_0); w,\theta_0\right) = \mathscr{M}\left(C(w,\theta_0), Y(w,\theta_0); \omega,\theta\right)$. There exists a unique incentive-compatible allocation $(w,\theta) \mapsto (\underline{C}(w,\theta), \underline{Y}(w,\theta))$ the restriction of which to group θ_0 is $w \mapsto (\underline{C}(w,\theta_0), \underline{Y}(w,\theta_0))$ and it verifies Assumption 3 if and only if, for each θ, $w \mapsto \underline{W}(w,\theta)$ is smoothly increasing.*

Lemma 4 guarantees that if $w \mapsto Y(w,\theta)$ is smoothly increasing in w and if, for each θ, the pooling function denoted $\underline{W}(w,\theta)$ is also smoothly increasing in w, then the allocation is incentive-compatible. Assumption 3 together with the assumption that $\underline{W}(\cdot,\theta)$ is smoothly increasing plays here a role similar to the assumption that the second-order incentive compatibility condition is satisfied with one dimension of heterogeneity. In what follows, we therefore select the allocation only for the reference group θ_0 and assume that the triggered allocations for the other groups verify Assumption 3. From Assumption 1 and Equation (3), the pooling condition (10) can be rewritten as:

$$v_y\left(Y(w,\theta_0); w,\theta_0\right) = v_y\left(Y(w,\theta_0); W(w,\theta),\theta\right). \tag{11}$$

The pooling function $W(\cdot,\theta)$ that enables to retrieve $(C(\cdot,\theta), Y(\cdot,\theta))$ from the allocation of the reference group $(C(\cdot,\theta_0), Y(\cdot,\theta_0))$ depends on $Y(\cdot,\theta)$ and not on $C(\cdot,\theta)$ thanks to the additive separability of the utility function. This is necessary to apply our allocation perturbation method. Interestingly, the pooling function is endogenous so that individuals with the same income level can have distinct behavioral elasticities. This is a major difference with the previous literature where characteristics are aggregated along a single dimension (which can encapsulate general equilibrium effects on the wage distribution for instance), e.g. Boadway et al. (2002a), Brett and Weymark (2003), Rothschild and Scheuer (2013, 2016), Scheuer (2013) and Gomes et al. (2018). Such an aggregation indeed prevents individuals who earn the same income to differ in terms of behavioral elasticities.

IV Allocation perturbation and optimal tax schedule

In this section, in order to obtain the necessary conditions for the optimal tax schedule, we propose a new method that we call an "allocation perturbation". We first reformulate the government problem taking into account the pooling function we introduced in the previous section. We second motivate and summarize our methodology. We then present the optimal tax schedule and the technical details behind its derivation. Eventually, we rewrite our optimal tax formula in terms of sufficient statistics.

Let $\mathscr{C}(\hat{u}, y; w, \theta)$ denote the consumption level the government needs to provide to a worker of type (w, θ) who earns y to ensure she enjoys a given utility level $\hat{u}$. The function $\mathscr{C}(\cdot, y; w, \theta)$ is the reciprocal of $\mathscr{U}(\cdot, y; w, \theta)$ and:

$$\mathscr{C}_u(\hat{u}, y; w, \theta) = \frac{1}{u'(c)} \qquad \text{and} \qquad \mathscr{C}_y(\hat{u}, y; w, \theta) = \frac{v_y(y; w, \theta)}{u'(c)} \tag{12}$$

where the various derivatives are evaluated at $c = \mathscr{C}(\hat{u}, y; w, \theta)$. Define the Lagrangian associated to the planner's problem as

$$\mathscr{L} \stackrel{\text{def}}{\equiv} \iint \left[Y(w, \theta) - \mathscr{C}\left(U(w, \theta), Y(w, \theta); w, \theta\right) + \frac{\Phi\left(U(w, \theta); w, \theta\right)}{\lambda} \right] f(w|\theta) dw \, d\mu(\theta). \tag{13}$$

where all terms are expressed in monetary units.

Under Assumption 3, thanks to Lemma 4, the government then simply chooses, among the set of smooth allocations, the best one that verifies the first-order incentive constraint (8a) for each group, and the pooling condition (10). When the pooling function associated with the solution is, for each group θ, smoothly increasing in skill w, the found solution also solves the problem with all incentive constraints.

To solve this type of problem with one-dimensional unobserved heterogeneity, one typically constructs a Hamiltonian and one applies the Pontryagin principle. In our multidimensional environment, the pooling condition (10) induces constraints on state and control variables which hold at endogenous skill levels. In this context, we rather propose using the calculus of variation and consider a set of perturbations of the allocation in the reference group. The cornerstone of our method is the pooling condition (10) that we use to deduce how the allocations in the other groups are perturbed.

IV.1 Allocation perturbation: Methodology

Consider a perturbation of the allocation for a given "reference" group that preserves incentive compatibility. In this group, change all incomes that correspond to a small interval of skills $(w - \delta, w)$ and do not modify incomes outside this interval. More precisely, perturb incomes in the reference group $Y(x, \theta_0)$ to obtain $Y(x, \theta_0) + t\, \Delta_Y(x, \theta_0; \delta)$ where $\Delta_Y(\cdot, \theta_0; \delta)$ is a continuously differentiable function defined on $\mathbb{R}^+$ such that $\Delta_Y(\cdot, \theta_0; \delta)$ is positive for $x \in (w - \delta, w)$ and is nil otherwise, and where $t \in \mathbb{R}$ is an algebraic magnitude. Thanks to the pooling function and Lemma 4, one can deduce the impact of this perturbation in all groups (within and outside the skills interval where the perturbation takes place). Take the Gâteaux derivative of the perturbed Lagrangian associated to the government's optimization problem with respect to t, at $t = 0$, and equate it to zero. One obtains an equation that characterizes the optimal marginal tax rates. The other necessary condition of the optimal tax schedule is obtained by perturbating the government's Lagrangian as follows. Uniformly modify utility (without modifying incomes) of all taxpayers by an amount Δ. Then take the Gâteaux derivative of the perturbed Lagrangian with respect to Δ, at $\Delta = 0$, and equate it to zero.

IV.2 Derivation of the optimal tax schedule

To save on notations, we from now on use the more compact notation $[x, \theta]$ when the various functions are evaluated for types (x, θ) at income $Y(x, \theta)$, utility $U(x, \theta)$ and consumption $C(x, \theta)$. The optimality condition with multidimensional heterogeneity is presented in the following proposition.

Proposition 1. *Under Assumptions 1, 2 and 3, the optimal tax schedule verifies:*

$$\frac{T'\left(Y(w,\theta_0)\right)}{1-T'\left(Y(w,\theta_0)\right)} \int\limits_{\theta\in\Theta} \frac{v_y\left[W(w,\theta),\theta\right]}{-W(w,\theta)\; v_{yw}\left[W(w,\theta),\theta\right]}\, W(w,\theta)\; f(W(y,\theta)|\theta)\; d\mu(\theta) \tag{14a}$$
$$= \; u'\left(C(w,\theta_0)\right) \iint\limits_{\theta\in\Theta, x\geq W(w,\theta)} \left(\frac{1}{u'(C(x,\theta))} - \frac{\Phi_U(U(x,\theta);x,\theta)}{\lambda}\right) f(x|\theta)dx\; d\mu(\theta)$$

for all skill w and

$$\iint\limits_{\theta\in\Theta, w\in\mathbb{R}_+} \left(\frac{\Phi_U(U(w,\theta);w,\theta)}{\lambda} - \frac{1}{u'(C(w,\theta))}\right) f(w|\theta)dw\; d\mu(\theta) = 0. \tag{14b}$$

Proof (Allocation perturbation)

To derive (14a) at a given skill level w, we consider a set of allocation perturbations, indexed by $t \in \mathbb{R}$ and $\delta \in \mathbb{R}_+$, that we denote $\hat{C}(w,\theta;t,\delta)$,$\hat{Y}(w,\theta;t,\delta)$ and $\hat{U}(w,\theta;t,\delta) \stackrel{\text{def}}{\equiv} \mathscr{U}\left(\hat{C}(w,\theta;t,\delta), \hat{Y}(w,\theta;t,\delta); w,\theta\right)$ where t stands for the size of the perturbation, and δ is the length of the skill interval where, in the reference group, the perturbation of incomes takes place. Following Lemma 4, we define the allocation perturbations from their restriction to the reference group θ_0 and then study the impact of these perturbations on the allocation in every other group. The perturbations of incomes in the reference group are defined by:

$$\hat{Y}(x,\theta_0;t,\delta) \stackrel{\text{def}}{\equiv} Y(x,\theta_0) + t\,\Delta_Y(x,\theta_0;\delta)$$

where $\Delta_Y(\cdot,\theta_0;\delta)$ is a continuously differentiable function defined on $\mathbb{R}^+$ such that $\Delta_Y(\cdot,\theta_0;\delta) > 0$ for $x \in (w-\delta, w)$ and $\Delta_Y(\cdot,\theta_0;\delta) = 0$ otherwise. Incomes in the reference group remain unchanged outside the skill interval $(w-\delta, w)$ and are increased (decreased) inside the skill interval $(w-\delta, w)$ if $t > 0$ (if $t < 0$), as illustrated in Figure 1. It is worth noting that the perturbed income function remains differentiable with respect to skill w since $\Delta_Y(\cdot,\cdot,\delta)$ is differentiable. Moreover, from Assumption 3, $Y(\cdot,\theta_0)$ admits a positive derivative everywhere, so $\dot{Y}(\cdot,\theta_0)$ is bounded away from 0 for all $x \in [w-\delta, w]$. Therefore, provided that t is small enough, which

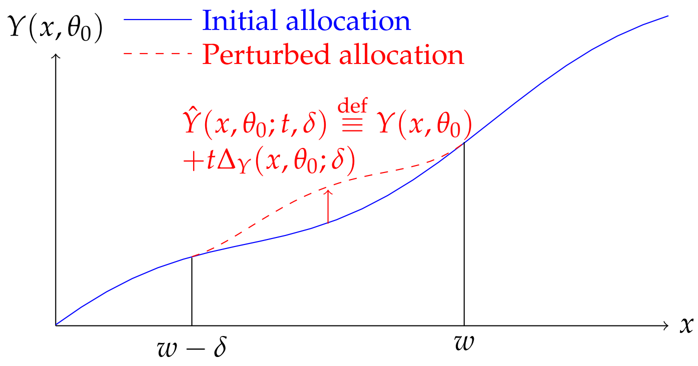

Figure 1: The perturbation of incomes in the reference group θ_0

we assume in the rest of the proof, $\hat{Y}(\cdot,\theta_0;t,\delta)$ has also a positive derivative everywhere and therefore verifies Assumption 3.

Let us in addition assume that the utility of the lowest skilled individuals in the reference group $U(0,\theta_0;t,\delta)$ is not perturbed and write it as $U(0,\theta_0)$. Therefore, according to (8c), the perturbed utility function in the reference group is

$$\hat{U}(x,\theta_0;t,\delta) \overset{\text{def}}{\equiv} U(0,\theta_0) - \int_0^x v_w\left(\hat{Y}(\omega,\theta_0;t,\delta);\omega,\theta_0\right) d\omega. \tag{15a}$$

From the pooling condition (10), as incomes $Y(\cdot,\theta_0;t,\delta)$ in the reference group remain unchanged outside the skill interval $(w-\delta,w)$, the pooling function $W(\cdot,\theta_0;t,\delta)$ is not perturbed outside the skill interval $(w-\delta,w)$. Therefore, incomes $Y(\cdot,\theta;t,\delta)$ in any group θ are not modified outside the skill interval $(W(w-\delta,\theta),W(w,\theta))$, and we must have (See Figure 2)

$$\hat{Y}(x,\theta;t,\delta) = Y(x,\theta) \qquad \text{if} \qquad x \in [0,W(w-\delta,\theta)] \cup [W(w,\theta),+\infty). \tag{15b}$$

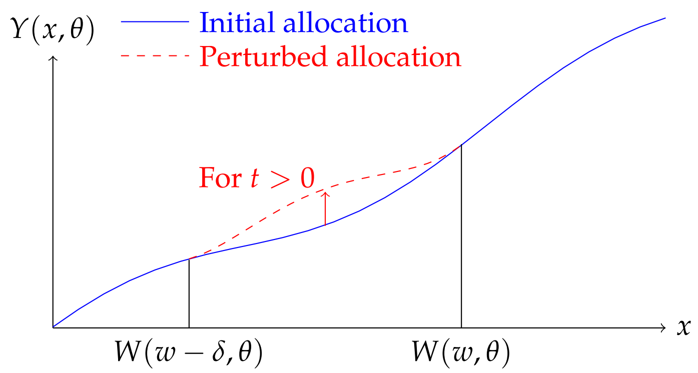

Figure 2: The perturbation of incomes in the other groups

Since incomes in the reference group are not perturbed for all skill x below $w-\delta$, the pooling function is also unchanged below $w-\delta$, so that the same types remain pooled together. Hence we get in group θ that for all $x \le W(w-\delta,\theta)$:

$$\hat{C}(x,\theta;t,\delta) = C(x,\theta) \qquad \text{and} \qquad \hat{U}(x,\theta;t,\delta) = U(x,\theta). \tag{15c}$$

For all skills $x > W(w-\delta,\theta)$, the change in utility obtained using the first-order incentive constraint (8c) is:

$$\hat{U}(x,\theta;t,\delta) - U(x,\theta) \quad = \quad -\int_0^x \left[v_w\left(\hat{Y}(\omega,\theta;t,\delta);\omega,\theta\right) - v_w\left(Y(\omega,\theta);\omega,\theta\right)\right] d\omega. \tag{15d}$$

Since incomes $\hat{Y}(\cdot,\theta;t,\delta)$ are only perturbed inside $(W(w-\delta,\theta)), W(w,\theta))$, for all skills x that belong to this interval, using (15b), we get:

$$\hat{U}(x,\theta;t,\delta) - U(x,\theta) \quad = \quad \int_{W(w-\delta,\theta)}^{x} \left[v_w\left(Y(\omega,\theta);\omega,\theta\right) - v_w\left(\hat{Y}(\omega,\theta;t,\delta);\omega,\theta\right)\right] d\omega. \tag{15e}$$

Moreover, for all skills x above $W(w,\theta)$, we have:

$$\hat{U}(x,\theta;t,\delta) - U(x,\theta) \quad = \quad \int_{W(w-\delta,\theta)}^{W(w,\theta)} \left[v_w\left(Y(\omega,\theta);\omega,\theta\right) - v_w\left(\hat{Y}(\omega,\theta;t,\delta);\omega,\theta\right)\right] d\omega. \tag{15f}$$

Hence utility in the other group does not change below $W(w-\delta,\theta)$ and changes by a uni-

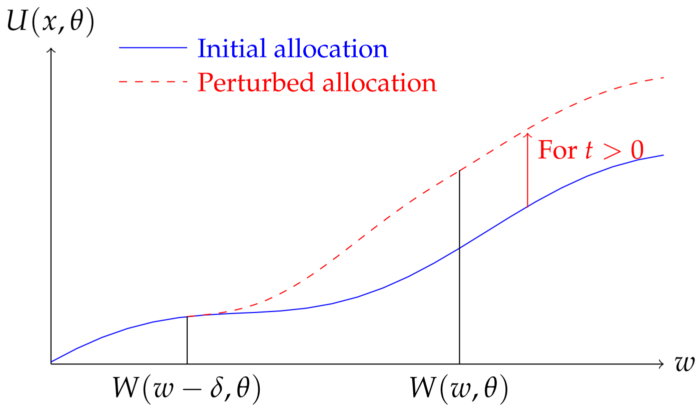

Figure 3: The perturbation of utilities

form amount above $W(w,\theta)$, as illustrated in Figure 3. As incomes above skill $W(w,\theta)$ are unchanged, this implies that, for all skill x above $W(w,\theta)$, the modifications in utility $U(x,\theta)$ occur only through changes of the utility $u(C(x,\theta))$ derived from consumption. Using (15f), this utility therefore changes uniformly by:

$$u\left(\hat{C}(x,\theta;t,\delta)\right) - u\left(C(x,\theta)\right) = \int_{W(w-\delta,\theta)}^{W(w,\theta)} \left[v_w\left(Y(\omega,\theta);\omega,\theta\right) - v_w\left(\hat{Y}(\omega,\theta;t,\delta);\omega,\theta\right)\right] d\omega \tag{15g}$$

which determines the perturbation of consumption for skill levels above $W(w,\theta)$. We now determine how the perturbations of incomes $Y(\cdot,\theta)$ in each group within the skill interval $(W(w-\delta,\theta), W(w,\theta))$ need to be set to ensure that the perturbed allocations remain incentive-compatible. For that purpose, we note that for all skill levels x above w, as incomes in the reference group are not perturbed, the pooling function is also unchanged, so that the same types remain pooled together. Hence, according to (9):

$$\forall t, \forall x \geq w \qquad \hat{Y}(W(x,\theta),\theta;t,\delta) = \hat{Y}(x,\theta_0;t,\delta) \qquad \text{and} \qquad \hat{C}(W(x,\theta),\theta;t,\delta) = \hat{C}(x,\theta_0;t,\delta).$$

This implies that, in all groups, the uniform change in utility that occurs for all skill levels above $W(w,\theta)$ must be identical across groups, so that: $u\left(\hat{C}(x,\theta_0;t,\delta)\right) - u\left(C(x,\theta_0)\right) = u\left(\hat{C}(W(x,\theta),\theta;t,\delta)\right) - u\left(C(W(x,\theta),\theta)\right)$, and so, using (8c) and (15g), we obtain:

$$\begin{aligned} &\int_{w-\delta}^{w} \left[v_w\left(\hat{Y}(\omega,\theta_0;t,\delta);\omega,\theta\right) - v_w\left(Y(\omega,\theta_0);\omega,\theta\right)\right] d\omega \\ = &\int_{W(w-\delta,\theta)}^{W(w,\theta)} \left[v_w\left(\hat{Y}(\omega,\theta;t,\delta);\omega,\theta\right) - v_w\left(Y(\omega,\theta);\omega,\theta\right)\right] d\omega. \end{aligned} \tag{15h}$$

The latter equation links the perturbed incomes $\hat{Y}(\cdot,\theta;t,\delta)$ in all groups within the interval of skills $(W(w-\delta,\theta), W(w,\theta))$ and the perturbed incomes $\hat{Y}(\cdot,\theta_0;t,\delta)$ in the reference group.

The perturbed Lagrangian is:

$$\hat{\mathscr{L}}(t,\delta) \stackrel{\text{def}}{\equiv} \iint_{\theta\in\Theta, w\in\mathbb{R}_+} \Big[\hat{Y}(w,\theta;t,\delta) - \mathscr{C}\left(\hat{U}(w,\theta;t,\delta),\hat{Y}(w,\theta;t,\delta);w,\theta\right) + \frac{\Phi\left(\hat{U}(w,\theta;t,\delta);w,\theta\right)}{\lambda}\Big] f(w|\theta)dw\, d\mu(\theta). \tag{16}$$

If the allocation is optimal, the derivative of this Lagrangian with respect to t must be nil at $t = 0$. In Appendix A.2, we show that the limit of $\partial\hat{\mathscr{L}}/\partial t$ when δ goes to zero leads to:

$$\int_{\theta\in\Theta} \frac{1 - \dfrac{v_y(Y(W(w,\theta),\theta);W(w,\theta),\theta)}{u'(C(W(w,\theta),\theta)}}{v_{yw}(Y(W(w,\theta),\theta);W(w,\theta),\theta)} f(W(w,\theta)|\theta)\, d\mu(\theta) \tag{17}$$
$$= \iint_{\theta\in\Theta, x\geq W(w,\theta)} \left(\frac{\Phi_U[x,\theta]}{\lambda} - \frac{1}{u'[x,\theta]}\right) f(x|\theta)dx\, d\mu(\theta).$$

Using (2), (3), $Y(W(w,\theta),\theta) \equiv Y(w,\theta_0)$ and $C(W(w,\theta),\theta) \equiv C(w,\theta_0)$, we can rewrite (17) as:

$$T'(Y(w,\theta_0)) \int_{\theta\in\Theta} \frac{1}{v_{yw}(Y(W(w,\theta),\theta);W(w,\theta),\theta)} f(W(w,\theta)|\theta)\, d\mu(\theta)$$
$$= \iint_{\theta\in\Theta, x\geq W(w,\theta)} \left(\frac{\Phi_U[x,\theta]}{\lambda} - \frac{1}{u'[x,\theta]}\right) f(x|\theta)dx\, d\mu(\theta).$$

Using again (2), (3) and (9) leads to (14a).

We now derive (14b). Consider a set of allocation perturbations indexed by $\Delta \in \mathbb{R}$ and denoted $(\tilde{C}(w,\theta;\Delta),\tilde{Y}(w,\theta;\Delta),\tilde{U}(w,\theta;\Delta) \stackrel{\text{def}}{\equiv} \mathscr{U}\left(\tilde{C}(w,\theta;\Delta),\tilde{Y}(w,\theta;\Delta);w,\theta\right))$, which consist, for each type $(x,\theta) \in \mathbb{R}_+ \times \Theta$, in no change in $Y(x,\theta)$ and in a uniform change in $U(x,\theta)$, therefore in $u(C(x,\theta))$ by an amount Δ. Hence, we get for each Δ that $\tilde{U}(w,\theta;\Delta) \stackrel{\text{def}}{\equiv} U(w,\theta) + \Delta$, $\tilde{Y}(w,\theta;\Delta) \stackrel{\text{def}}{\equiv} Y(w,\theta)$ and $\tilde{C}(w,\theta;\Delta) \stackrel{\text{def}}{\equiv} \mathscr{C}\left(\tilde{U}(w,\theta;\Delta),\tilde{Y}(w,\theta;\Delta);w,\theta\right)$. These perturbations preserve incentive-compatibility (6). According to (13), the perturbed Lagrangian can be written as

$$\tilde{\mathscr{L}}(\Delta) \stackrel{\text{def}}{\equiv} \iint \left[\tilde{Y}(w,\theta;\Delta) - \mathscr{C}\left(\tilde{U}(w,\theta;\Delta),\tilde{Y}(w,\theta;\Delta);w,\theta\right) + \frac{\Phi\left(\tilde{U}(w,\theta;\Delta);w,\theta\right)}{\lambda}\right] f(w|\theta)dw\, d\mu(\theta).$$

If the allocation is optimal, the above perturbations do not affect the Lagrangian. Thus, by equating the Gâteaux derivative of the Lagrangian in the direction described by the above perturbations (i.e. the derivative of the perturbed Lagrangian $\tilde{\mathscr{L}}(\cdot)$ with respect to Δ, at $\Delta = 0$) to zero, we obtain an equation that characterizes the optimal tax system. Using the first equality in (12), this Gâteaux derivative of the Lagrangian is

$$\tilde{\mathscr{L}}'(0) = \iint_{\theta\in\Theta, x\in\mathbb{R}_+} \left(\frac{\Phi_U\left(U(x,\theta);x,\theta\right)}{\lambda} - \frac{1}{u'\left(C(x,\theta)\right)}\right) f(x|\theta)dx\, d\mu(\theta).$$

Equating this derivative to zero leads to (14b).

□

The tax formula (14a) of Proposition 1 generalizes the Mirrlees (1971) formula and ABC formula described in Diamond (1998) to multidimensional individual characteristics with an efficiency, an equity and a distribution term. Using the Hamiltonian method, we show in Appendix A.3 that, under a single dimension of heterogeneity w, Equations (14a) and (14b) simplify to (18a)-(18b).

Lemma 5. *When the unobserved heterogeneity has only one dimension, the optimal tax schedule satisfies:*

$$\frac{T'(Y(w))}{1-T'(Y(w))} \cdot \frac{v_y\,[w]}{-w\; v_{yw}\,[w]}\; w\, f(w) \;=\; u'\,[w] \int_w^{\infty} \left(\frac{1}{u'\,[x]} - \frac{\Phi_U\,[x]}{\lambda} \right) f(x)\, dx \tag{18a}$$

$$0 \;=\; \int_0^{\infty} \left(\frac{1}{u'\,[x]} - \frac{\Phi_U\,[x]}{\lambda} \right) f(x)\, dx. \tag{18b}$$

Comparing these equations with Equations (14a) and (14b) makes clear that reducing the tax problem to one dimension of heterogeneity implies that the integrals over θ-groups disappear. With multidimensional heterogeneity, one needs to aggregate the terms of the formula for individuals of the different groups who pool at the same level of income. This is made possible thanks to our characterization of the pooling function in Lemmas 3 and 4.

Comparing Proposition 1 with Lemma 5 makes it clear that, with multidimensional heterogeneity, the composition of the population at each income level plays a role through the weight $W(y,\theta) f(W(y,\theta)|\theta)\; d\mu(\theta)$. This weight is not simply the density of people in each group θ who earn the relevant level of income but the product of the skill levels $W(y,\theta)$ times the corresponding densities $f(W(y,\theta)|\theta)\; d\mu(\theta)$. We know that taxpayers who earn the same income come from distinct groups θ have distinct skill levels $W(y,\theta)$. Moreover, at each income level, the composition of the population (e.g. the proportion of men versus women) usually varies between the actual economy where it is estimated and the optimal economy. Therefore, the weights $W(y,\theta) f(W(y,\theta)|\theta)\; d\mu(\theta)$ take distinct values in the actual and optimal economies. In Jacquet and Lehmann (2020), we show that ignoring these so-called composition effects significantly biases the optimal income tax schedule.

IV.3 Rewriting the optimal tax schedule in terms of sufficient statistics

In this section, we rewrite the optimal tax formula of Proposition 1 in terms of meaningful sufficient statistics. To obtain the latter, we first define a set of individual elasticities. For this purpose, we assume that the tax function $T(\cdot)$ is twice differentiable which ensures that the first-order condition associated to the individual maximization program (1) is differentiable. We also assume that, for all $(w,\theta) \in \mathbb{R}_+^* \times \Theta$, the second-order condition associated to the individual maximization program holds strictly which guarantees it is invertible in y. This allows one to apply the implicit function theorem to the first-order condition associated to the individual maximization program. We also assume that, for all $(w,\theta) \in \mathbb{R}_+^* \times \Theta$, the function $y \mapsto \mathscr{U}\,(y - T(y), y; w, \theta)$ admits a unique global maximum over $\mathbb{R}_+$. This ensures that after a change in the marginal tax rate or a change in skill, the maximum remains global.

Along the nonlinear income tax schedule, we define the *compensated elasticity* of earnings with respect to the marginal retention rate $1 - T'(.)$ as the elasticity of earnings for individuals of type (w, θ) to a change in the marginal retention rate by a constant amount τ, while leaving unchanged the level of tax at $y = Y(w, \theta)$, i.e.:

$$\varepsilon(w, \theta) \stackrel{\text{def}}{\equiv} \frac{1 - T'(Y(w, \theta))}{Y(w, \theta)} \frac{\partial Y^c}{\partial \tau} \tag{19a}$$

which is positive as shown in Appendix A.4 and where the superscript "c" stands for "compensated".

Along the nonlinear income tax schedule, the *income effect* is defined as the behavioral response to a lump-sum change m in tax liability:

$$\eta(w, \theta) \stackrel{\text{def}}{\equiv} \frac{\partial Y^i}{\partial m} \tag{19b}$$

where the superscript "i" stands for "income effect". We have $\eta(w, \theta) < 0$ if leisure is a normal good (see appendix).

Let us use the dot above a variable for the partial derivative of this variable with respect to skill w. One can define the elasticity $\alpha(w; \theta)$ of earnings with respect to the skill level:

$$\alpha(w, \theta) \stackrel{\text{def}}{\equiv} \frac{w}{Y(w, \theta)} \dot{Y}(w, \theta). \tag{19c}$$

which is positive thanks to Assumption 2. We can note that $\varepsilon(w, \theta)$, $\eta(w, \theta)$ and $\alpha(w, \theta)$ denote *total* responses of earnings since they take into account the nonlinearity of the tax schedule as in Jacquet et al. (2013), see also Scheuer and Werning (2016).

We define the social marginal welfare weights associated with workers of type (w, θ) expressed in terms of public funds by:

$$g(w, \theta) \quad \stackrel{\text{def}}{\equiv} \quad \frac{\Phi_U(U(w, \theta); w, \theta)\, \mathscr{U}_c(C(w, \theta), Y(w, \theta); w, \theta)}{\lambda}. \tag{20}$$

The government values giving one extra dollar to a worker (w, θ) as a gain of $g(w, \theta)$ dollar(s) of public funds.

Let $h(y|\theta)$ denote the conditional income density within group θ at income y and $H(y|\theta) \stackrel{\text{def}}{\equiv} \int_0^y h(z|\theta) dz$ the corresponding conditional income CDF. According to (19c) and Assumption 2, income $Y(\cdot, \theta)$ is strictly increasing in skill within each group. We then have $H(Y(w, \theta)|\theta) \equiv F(w|\theta)$ for each skill level w. Differentiating both sides of this equality with respect to w and using (19c), the two densities are linked by:

$$h(Y(w, \theta)|\theta) = \frac{f(w|\theta)}{\dot{Y}(w, \theta)} \qquad \Leftrightarrow \qquad Y(w, \theta)\, h(Y(w, \theta)|\theta) = \frac{w\, f(w|\theta)}{\alpha(w, \theta)}. \tag{21}$$

The *unconditional* income density at income $Y(w, \theta_0)$ is given by:

$$\hat{h}(Y(w, \theta_0)) \stackrel{\text{def}}{\equiv} \int_{\theta \in \Theta} h(Y(W(w, \theta), \theta)|\theta)\, d\mu(\theta). \tag{22a}$$

The mean (total) compensated elasticity at income level $y = Y(w, \theta_0)$ is:

$$\hat{\varepsilon}(Y(w, \theta_0)) = \int_{\theta \in \Theta} \varepsilon(W(w, \theta), \theta) \frac{h(Y(W(w, \theta), \theta)|\theta)}{\hat{h}(Y(w, \theta_0))} d\mu(\theta). \tag{22b}$$

where each within-group compensated elasticity of earnings is timed by the relative proportion $h(y|\theta)/\hat{h}(y)$ of individuals in the corresponding group among individuals who earn y. The mean (total) income effect at income level $Y(w,\theta_0)$ is:

$$\hat{\eta}(Y(w,\theta_0)) = \int_{\theta\in\Theta} \eta\left(W(w,\theta),\theta\right) \frac{h(Y(W(w,\theta),\theta)|\theta)}{\hat{h}(Y(w,\theta_0))} d\mu(\theta). \tag{22c}$$

Finally, the mean marginal social welfare weight at income level $y = Y(w,\theta_0)$ is:

$$\hat{g}(Y(w,\theta_0)) = \int_{\theta\in\Theta} g\left(W(w,\theta),\theta\right) \frac{h(Y(W(w,\theta),\theta)|\theta)}{\hat{h}(Y(w,\theta_0))} d\mu(\theta). \tag{22d}$$

As shown in Appendix A.4, we can now rearrange the first-order conditions (14a) and (14b) displayed in Proposition 1 to obtain the optimal marginal tax rate in terms of sufficient statistics, more precisely in terms of the mean compensated elasticity, mean income effect, mean marginal social weights and the unconditional income density.

Proposition 2. *Under assumption 2, with a twice differentiable $T(.)$, with the second order condition of the individual maximization program that holds strictly and the function $y \mapsto \mathscr{U}\left(y - T(y), y; w, \theta\right)$ that admits a unique global maximum, the optimal tax schedule satisfies:*

$$\frac{T'(y)}{1-T'(y)} = \frac{1}{\hat{\varepsilon}(y)} \cdot \frac{1-\hat{H}(y)}{y\hat{h}(y)} \cdot \left(1 - \frac{\int_y^\infty \left[\hat{g}(z) + \hat{\eta}(z)\cdot T'(z)\right]\cdot \hat{h}(z)dz}{1-\hat{H}(y)}\right) \tag{23a}$$

$$1 = \int_0^\infty \left[\hat{g}(z) + \hat{\eta}(z)\cdot T'(z)\right]\cdot \hat{h}(z)dz. \tag{23b}$$

The optimal tax rate given in Equation (23a) depends now on three terms that encapsulate sufficient statistics only: (a) the behavioral responses to taxes denoted by $1/\hat{\varepsilon}(y)$, which, in the vein of Ramsey (1927), is the inverse of the mean compensated elasticity; (b) the social preferences and income effects $1 - \left(\int_y^\infty \left[\hat{g}(z) + \hat{\eta}(z)\, T'(z)\right] \hat{h}(z)dz\right) / \left(1-\hat{H}(y)\right)$, which indicates the distributional benefits of increasing the tax liability by one unit for all workers with incomes above y and (c) the shape of the income distribution measured by the inverse of the local Pareto parameter $(1-\hat{H}(y))/(y\,\hat{h}(y))$ of the income distribution. In Equation (23a), the sufficient statistics' optimal tax formula of Saez (2001) is generalized to the multidimensional context.

Equations (23a) and (23b) can also be obtained, in a straightfoward way, using the tax perturbation approach as in Jacquet and Lehmann (2020). Thanks to these equations, which characterize the optimal tax schedules in terms of incomes, we know the weight that multiplies each sufficient statistic. In (23a), the unconditional income density $\hat{h}(Y(w,\theta_0))$ is the weight that multiplies each mean marginal social welfare weight and each mean income effect. In contrast, the weights that are used in the optimal tax formulas in terms of skills (i.e. Equations (14a) and (14b)) are distinct. These weights are the conditional skill densities times the corresponding skill.[7]

[7]More precisely, in the left-hand side of Equation (14a), the term $-\frac{v_y[w,\theta]}{w\cdot v_{yw}[w,\theta]}$ which is equal to the ratio of $\varepsilon(w,\theta)$ and $\alpha(w,\theta)$ (see Equation (35) in the appendix), is weighted by the conditional density times the skill, $W(w,\theta)\, f(W(y,\theta)|\theta)$. And, in the right-hand side of (14a), which encapsulates the mechanical and income effects, the weights are the conditional skill densities.

V Signing optimal marginal tax rates

In the literature, it has been highlighted that, with multidimensional heterogeneity, negative marginal tax rates can be optimal. In Cuff (2000), Boadway et al. (2002a), Brett and Weymark (2003), Choné and Laroque (2010) and Lockwood and Weinzierl (2015), individuals differ along their skills and preferences for leisure, and the social planner has weighted utilitarian preferences. In this context, individuals who pool at the same income level are characterized by different marginal social welfare weights, $\Phi_U(U(w,\theta);w,\theta)u'(C(w,\theta))/\lambda$. Therefore, the mean marginal social welfare weight, $\iint_{\theta\in\Theta,w\in\mathbb{R}_+}\left(\frac{\Phi_U(U(w,\theta);w,\theta)u'(C(w,\theta))}{\lambda}\right)f(w|\theta)dw\ d\mu(\theta)$, may not be decreasing with income which opens the way to negative marginal tax rates. One may wonder whether heterogeneous behavioral responses for individuals who earn the same level of income could also open the way to negative marginal tax rates. We now show that this is not the case (if the population earning a given income level remains homogeneous in terms of social welfare weights).

Proposition 3. *Under utilitarian or maximin social preferences, optimal marginal tax rates are positive.*

Proposition 3 generalizes to the multidimensional case Mirrlees (1971)'s result of positive optimal tax rates (which was obtained under additively separable preferences).[8] The proof, which relies on Proposition 1, can be found in Appendix A.5. Proposition 3 emphasizes that optimal marginal tax rates are positive as soon as all individuals who earn the same income y are characterized (*i*) by the same marginal utility of consumption, which is ensured by the additive separability assumption and (*ii*) by the same marginal social welfare Φ_u, which is ensured by utilitarian or maximin social objective. In such a case, all individuals who earn the same income are characterized by the same welfare weights. Therefore, the cause of negative marginal tax rates due to decreasing mean marginal social welfare weights emphasized in Cuff (2000) and follow-up papers does not apply.[9]

VI Concluding Comments

In this paper, we have proposed an allocation perturbation method to derive optimal nonlinear income tax schedules when taxpayers differ along several characteristics and when taxation cannot be conditioned on them. In this context, we have shown that marginal tax rates are positive under utilitarian preferences and maximin.

To illustrate the generality of our results in this concluding section, we now provide alternative tax problems that one can easily solve in our framework. For each of them, we explain

[8]In Hellwig (2007), under a utilitarian criterion, positive optimal tax rates are obtained with more general preferences.

[9]If the utility function $u(\cdot)$ in (1) were parameterized by type w and θ while $v(\cdot)$ were simply parameterized by w, individuals who earn the same income would have distinct social marginal welfare weights. This could drive negative marginal tax rates. Similarly, if both $u(\cdot)$ and $v(\cdot)$ were parameterized by w and θ, one would also expect negative marginal tax rates. Let us stress that our method could not be used in this framework since the pooling function (10) cannot depend simultaneously on Y and C.

what y, w, θ represent so that the interpretation of the results is straightforward.

Optimal joint taxation of labor and non-labor income

Consider individuals that have two sources of taxable income: a non-labor income z and a labor income $y - z$. Those incomes are jointly taxed and the tax function does not distinguish between both incomes. This applies, for instance, in countries like France where income received from renting property and entrepreneurial income are jointly taxed with labor income. As explained in Scheuer (2014), a single nonlinear tax schedule is also the system that is in place for employed workers and self-employed small business owners in many countries, including the U.S.. In this case, y is the total taxable income and we interpret θ as the ability to earn non-labor income z and w as the skill. Individuals of type (w, θ) solve:

$$\max_{y,z} \quad \mathcal{U}(y - T(y), y - z, z; w, \theta)$$

where two decision variables appear instead of one variable in the core of our paper. This program can be solved sequentially, the first step being the choice of non-labor income z for a given taxable income y which leads to $\mathscr{U}(c, y; w, \theta) \stackrel{\text{def}}{\equiv} \max_{z} \quad \mathcal{U}(c, y - z, z; w, \theta)$. The second step is the choice of y as in Equation (1). In the process, one simply needs to ensure the semi-indirect utility function $\mathscr{U}(\cdot, \cdot; w, \theta)$ verifies Assumption 2.

Optimal joint income taxation of couples

The joint income taxation of couples is a variant of the previous application, in which $y - z$ is the labor income of one individual and z is the one of his/her partner. The tax does not distinguish between $y - z$ and z and only depends on the sum of both incomes, y (as in France, Germany and the US). We redefine w and θ as the respective skill level of each member of the couple. The optimal tax schedules derived in this paper are then interpreted as the optimal tax schedules when the couple is the tax unit and each partner decides along the intensive margin. So far, previous attempts in the literature (Kleven et al. (2009) and Cremer et al. (2012)) have stopped short of obtaining these nonlinear tax schedules.

Optimal income taxation with tax avoidance

In this application, w is the skill and θ is the ability to avoid taxation. We assume that tax enforcement (penalty, monitoring, etc.) is given. We denote z the sheltered labor income (i.e. income that is not taxed at all) and $y + z$ the (total) labor income. The tax only depends on the taxable income y. Consumption becomes $c + z$, with $c = y - T(y)$ being the after-tax income. All results obtained in this paper are valid in this context when one simply makes sure that Assumption 2 holds.

A Appendix

A.1 Proof of Lemma 4

Proof The proof consists of two steps. In step (i), we show that there exists at most one incentive-compatible allocation $(w, \theta) \mapsto (\underline{C}(w, \theta), \underline{Y}(w, \theta))$ that verifies Assumption 3 and such that $(\underline{C}(w,\theta_0),\underline{Y}(w,\theta_0)) = (C(w,\theta_0),Y(w,\theta_0))$. In step (ii), we show that this allocation verifies the whole set of incentive constraints (6).

Step (i). To build up the entire incentive-compatible allocation $(w,\theta) \mapsto (\underline{C}(w,\theta),\underline{Y}(w,\theta))$, we must choose $(\underline{C}(w,\theta_0),\underline{Y}(w,\theta_0)) = (C(w,\theta_0),Y(w,\theta_0))$ at any skill level. For each group θ, $\underline{Y}(\cdot,\theta)$ verifies Assumption 3 if and only if its reciprocal $\underline{Y}^{-1}(\cdot;\theta)$ is smoothly increasing. Let $y \in \mathbb{R}_+$ be an income level. As $Y(\cdot,\theta_0)$ is smoothly increasing from Assumption 3, there exists a unique skill level w such that $y = Y(w,\theta_0)$. Then according to Lemma 3, among individuals of group θ, only those of skill $\underline{W}(w,\theta)$ must be assigned to the income level $y = Y(w,\theta_0)$ to verify incentive-compatibility.[10] Therefore, $\underline{Y}^{-1}(\cdot,\theta)$ must be defined by:

$$\underline{Y}^{-1}(\cdot,\theta): \qquad y \overset{Y^{-1}(\cdot,\theta_0)}{\longmapsto} w = Y^{-1}(y,\theta_0) \overset{W(\cdot,\theta)}{\longmapsto} Y^{-1}(y,\theta).$$

$\underline{Y}^{-1}(\cdot,\theta)$ is then smoothly increasing as a combination of two smoothly increasing functions. Moreover, since for each type (ω,θ), there exists a single skill level ω such that $\underline{Y}(\omega,\theta) = Y(w,\theta_0)$, incentive compatibility requires that $\underline{C}(\omega,\theta)$ also needs to be equal to $\underline{C}(w,\theta_0)$. This ends the proof of step (i).

Step (ii). Note that the allocation $(w,\theta) \mapsto (\underline{Y}(w,\theta),\underline{C}(w,\theta))$ is built in such a way that one has $\underline{Y}(\omega,\theta) = Y(w,\theta_0)$ and $\underline{C}(\omega,\theta) = C(w,\theta_0)$ if and only if $\omega = \underline{W}(w,\theta)$ and (10) holds. Differentiating in w both sides of these two equations and rearranging terms, we obtain

$$\frac{\dot{C}(w,\theta_0)}{\dot{Y}(w,\theta_0)} = \frac{\dot{\underline{C}}(\underline{W}(w,\theta),\theta_0)}{\dot{\underline{Y}}(\underline{W}(w,\theta),\theta_0)}.$$

As $w \mapsto (C(w,\theta_0),Y(w,\theta_0))$ is assumed to verify the within-group incentive constraints in Equation (8b), we know that the left-hand side of the above equation is equal to

$$\mathscr{M}(C(w,\theta_0),Y(w,\theta_0);w,\theta_0).$$

Using the definition of $\underline{W}(\cdot,\theta)$, we have that $w \mapsto (\underline{C}(w,\theta),\underline{Y}(w,\theta))$ also verifies Equation (8b). From Lemma 2, it thus verifies the within-group incentive constraints (7). We now check whether the inequality (6) is verified for any $(w,w',\theta,\theta') \in \mathbb{R}_+^2 \times \Theta^2$. We know there exists $\omega \in \mathbb{R}_+$ such that $\underline{Y}(\omega,\theta) = \underline{Y}(w',\theta')$ and $\underline{C}(\omega,\theta) = \underline{C}(w',\theta')$. The incentive constraints in (6) are therefore equivalent to:

$$\mathscr{U}(C(w,\theta),Y(w,\theta);w,\theta) \geq \mathscr{U}(C(\omega,\theta),Y(\omega,\theta);w,\theta).$$

The latter inequality is verified as $w \mapsto (\underline{C}(w,\theta),\underline{Y}(w,\theta))$ satisfies Equation (8b). □

A.2 Derivation of Equation (17)

Proof

To derive (17), we must compute the various Gâteaux derivatives at $t = 0$. For $A = C, Y, U$ and a given δ, the Gâteaux derivative of A in the direction $\Delta_Y(\cdot,\cdot;\delta)$ at $t = 0$ is denoted $\hat{\hat{A}}(x,\theta;\delta)$. Let us remind its definition:

$$\hat{\hat{A}}(x,\theta;\delta) \overset{\text{def}}{\equiv} \lim_{t\mapsto 0} \frac{\hat{A}(x,\theta;t,\delta) - A(w,\theta)}{t}.$$

By definition we get: $\hat{\hat{Y}}(x,\theta_0;\delta) = \Delta_Y(x;\delta)$, and from (15b) we obtain:

$$\hat{\hat{Y}}(x,\theta;\delta) = 0 \qquad \text{if} \qquad x \in [0, W(w-\delta,\theta)] \cup [W(w,\theta),+\infty). \tag{24a}$$

[10] Hence function $\underline{W}(\cdot,\theta)$ coincides with the pooling function $W(\cdot,\theta)$.

Equations (15c) imply that the Gâteaux derivatives of utilities are nil for skill below $W(w-\delta,\theta)$. For skills x between $W(w-\delta,\theta)$ and $W(w,\theta)$, Equation (15e) implies:

$$\hat{\hat{U}}(x,\theta;\delta) = -\int_{W(w-\delta,\theta)}^{x} v_{yw}\left(Y(\omega,\theta_0);\omega,\theta_0\right)\ \hat{\hat{Y}}(\omega,\theta_0;\delta)\,d\omega. \tag{24b}$$

For skill x above $W(w,\theta)$, according to (15f), we have:

$$\hat{\hat{U}}(x,\theta;\delta) = -\int_{W(w-\delta,\theta)}^{W(w,\theta)} v_{yw}\left(Y(\omega,\theta_0);\omega,\theta_0\right)\ \hat{\hat{Y}}(\omega,\theta_0;\delta)\,d\omega. \tag{24c}$$

Moreover, Equation (15h) implies that the Gâteaux derivatives of income must verify:

$$\int_{w-\delta}^{w} v_{yw}\left(Y(\omega,\theta_0);\omega,\theta\right)\ \hat{\hat{Y}}(\omega,\theta_0;\delta)\,d\omega = \int_{W(w-\delta,\theta)}^{W(w,\theta)} v_{yw}\left(Y(\omega,\theta);\omega,\theta\right)\ \hat{\hat{Y}}(\omega,\theta;\delta)\,d\omega. \tag{24d}$$

Using Equations (12), (24a) and (24c), the Gâteaux derivative of the Lagrangian (16) is:

$$\begin{aligned}\frac{\partial \mathscr{L}}{\partial t}(0;\delta) = \int_{\theta\in\Theta}\Bigg\{&\int_{W(w-\delta,\theta)}^{W(w,\theta)}\left(1-\frac{v_y(Y(x,\theta);x,\theta)}{u'(C(x,\theta))}\right)\hat{\hat{Y}}(x,\theta;\delta)f(x|\theta)dx \\ +&\int_{W(w-\delta,\theta)}^{W(w,\theta)}\left(\frac{\Phi_U[x,\theta]}{\lambda}-\frac{1}{u'[x,\theta]}\right)\hat{\hat{U}}(x,\theta;\delta)f(x|\theta)dx \\ -&\left(\int_{W(w-\delta,\theta)}^{W(w,\theta)} v_{yw}\left(Y(x,\theta);x,\theta\right)\ \hat{\hat{Y}}(x,\theta;\delta)\,dx\right) \\ \times&\left(\int_{W(w,\theta)}^{\infty}\left(\frac{\Phi_U[x,\theta]}{\lambda}-\frac{1}{u'[x,\theta]}\right)f(x|\theta)dx\right)\Bigg\}\,d\mu(\theta).\end{aligned} \tag{25}$$

Dividing the first-order condition $\frac{\partial \mathscr{L}}{\partial t}(0;\delta) = 0$ by $\int_{w-\delta}^{w} v_{yw}\left(Y(x,\theta_0);x,\theta_0\right)\ \hat{\hat{Y}}(x,\theta_0;\delta)\,dx$ implies, using (24b) and (24d), that

$$\begin{aligned}&\int_{\theta\in\Theta}\frac{\int_{W(w-\delta,\theta)}^{W(w,\theta)}\left(1-\frac{v_y(Y(x,\theta);x,\theta)}{u'(C(x,\theta))}\right)\hat{\hat{Y}}(x,\theta;\delta)f(x|\theta)dx}{\int_{W(w-\delta,\theta)}^{W(w,\theta)} v_{yw}\left(Y(x,\theta);x,\theta\right)\ \hat{\hat{Y}}(x,\theta;\delta)\,dx}\,d\mu(\theta) = \\ &\int_{\theta\in\Theta}\Bigg\{\int_{W(w-\delta,\theta)}^{W(w,\theta)}\left(\frac{\Phi_U[x,\theta]}{\lambda}-\frac{1}{u'[x,\theta]}\right)\frac{\int_{W(w-\delta,\theta)}^{x} v_{yw}\left(Y(x,\theta);x,\theta\right)\ \hat{\hat{Y}}(x,\theta;\delta)\,dx}{\int_{W(w-\delta,\theta)}^{W(w,\theta)} v_{yw}\left(Y(x,\theta);x,\theta\right)\ \hat{\hat{Y}}(x,\theta;\delta)\,dx}f(x|\theta)dx\ + \\ &\int_{W(w,\theta)}^{\infty}\left(\frac{\Phi_U[x,\theta]}{\lambda}-\frac{1}{u'[x,\theta]}\right)f(x|\theta)dx\Bigg\}\,d\mu(\theta).\end{aligned} \tag{26}$$

We finally take the limit of the latter equality when δ tends to 0. Let us consider the first term in the right-hand side of (26). Since

$$\frac{\int_{W(w-\delta,\theta)}^{x} v_{yw}\left(Y(x,\theta);x,\theta\right)\ \hat{\hat{Y}}(x,\theta;\delta)\,dx}{\int_{W(w-\delta,\theta)}^{W(w,\theta)} v_{yw}\left(Y(x,\theta);x,\theta\right)\ \hat{\hat{Y}}(x,\theta;\delta)\,dx} \in [0,1]$$

we get that:

$$\begin{aligned}&\left|\int_{\theta\in\Theta}\int_{W(w-\delta,\theta)}^{W(w,\theta)}\left(\frac{\Phi_U[x,\theta]}{\lambda}-\frac{1}{u'[x,\theta]}\right)\frac{\int_{W(w-\delta,\theta)}^{x} v_{yw}\left(Y(x,\theta);x,\theta\right)\ \hat{\hat{Y}}(x,\theta;\delta)\,dx}{\int_{W(w-\delta,\theta)}^{W(w,\theta)} v_{yw}\left(Y(x,\theta);x,\theta\right)\ \hat{\hat{Y}}(x,\theta;\delta)\,dx}f(x|\theta)dxd\mu(\theta)\right| \\ &\qquad\leq\left|\int_{\theta\in\Theta}\int_{W(w-\delta,\theta)}^{W(w,\theta)}\left(\frac{\Phi_U[x,\theta]}{\lambda}-\frac{1}{u'[x,\theta]}\right)f(x|\theta)dxd\mu(\theta)\right|.\end{aligned}$$

As the right hand-side of the latter inequality tends to 0 when δ tends to 0, the limit of (26) when δ tends to zero leads to:

$$\lim_{\delta\mapsto 0} \int_{\theta\in\Theta} \frac{\int_{W(w-\delta,\theta)}^{W(w,\theta)} \left(1 - \frac{v_y(Y(x,\theta);x,\theta)}{u'(C(x,\theta))}\right) \hat{\hat{Y}}(x,\theta;\delta) f(x|\theta) dx}{\int_{W(w-\delta,\theta)}^{W(w,\theta)} v_{yw}\left(Y(x,\theta);x,\theta\right)\ \hat{\hat{Y}}(x,\theta;\delta)\ dx} d\mu(\theta) \tag{27}$$
$$= \iint_{\theta\in\Theta, x\geq W(w,\theta)} \left(\frac{\Phi_U[x,\theta]}{\lambda} - \frac{1}{u'[x,\theta]}\right) f(x|\theta) dx\, d\mu(\theta).$$

By continuity, the variations of $f(x|\theta)$, $v_y(Y(x,\theta);x,\theta)$, $v_{yw}(Y(x,\theta);x,\theta)$ and $u'(c(x,\theta))$ within the skill intervals $[W(w-\delta,\theta), W(w,\theta)]$ are of second-order when δ tends to 0. As Θ and intervals $[W(w-\delta,\theta), W(w,\theta)]$ are compact, for any small $e > 0$, there always exists $\tilde{\delta}(e)$ such that for all $(x,\theta) \in [W(w-\tilde{\delta}(e),\theta), W(w,\theta)] \times \Theta$, one has:

$$\left(\frac{1 - v_y[W(w,\theta),\theta]}{u'(C(W(w,\theta),\theta)} f(W(w,\theta)|\theta) - e\right) \hat{\hat{Y}}(x,\theta;\delta) \leq \left(\frac{1 - v_y[W(x,\theta),\theta]}{u'(C(W(x,\theta),\theta)} f(x|\theta)\right) \hat{\hat{Y}}(x,\theta;\delta)$$
$$\leq \left(\frac{1 - v_y[W(w,\theta),\theta]}{u'(C(W(w,\theta),\theta)} f(W(w,\theta)|\theta) + e\right) \hat{\hat{Y}}(x,\theta;\delta)$$

and

$$\left(v_{yw}[W(w,\theta),\theta] - e\right) \hat{\hat{Y}}(x,\theta;\delta) \leq v_{yw}[W(x,\theta),\theta]\, \hat{\hat{Y}}(x,\theta;\delta) \leq \left(v_{yw}[W(w,\theta),\theta] + e\right) \hat{\hat{Y}}(x,\theta;\delta) < 0$$

so that for all $\delta < \tilde{\delta}(e)$:

$$\int_{\theta\in\Theta} \frac{\left(1 - \frac{v_y(Y(W(w,\theta),\theta);W(w,\theta),\theta)}{u'(C(W(w,\theta),\theta)}\right) f(W(w,\theta)|\theta) + e}{v_{yw}(Y(W(w,\theta),\theta);W(w,\theta),\theta) - e} \frac{\int_{W(w-\delta,\theta)}^{W(w,\theta)} \hat{\hat{Y}}(x,\theta;\delta) dx}{\int_{W(w-\delta,\theta)}^{W(w,\theta)} \hat{\hat{Y}}(x,\theta;\delta)\ dx} d\mu(\theta)$$
$$\leq \int_{\theta\in\Theta} \frac{\int_{W(w-\delta,\theta)}^{W(w,\theta)} \left(1 - \frac{v_y(Y(x,\theta);x,\theta)}{u'(C(x,\theta))}\right) \hat{\hat{Y}}(x,\theta;\delta) f(x|\theta) dx}{\int_{W(w-\delta,\theta)}^{W(w,\theta)} v_{yw}\left(Y(x,\theta);x,\theta\right)\ \hat{\hat{Y}}(x,\theta;\delta)\ dx} d\mu(\theta)$$
$$\leq \int_{\theta\in\Theta} \frac{\left(1 - \frac{v_y(Y(W(w,\theta),\theta);W(w,\theta),\theta)}{u'(C(W(w,\theta),\theta)}\right) f(W(w,\theta)|\theta) - e}{v_{yw}(Y(W(w,\theta),\theta);W(w,\theta),\theta) + e} \frac{\int_{W(w-\delta,\theta)}^{W(w,\theta)} \hat{\hat{Y}}(x,\theta;\delta) dx}{\int_{W(w-\delta,\theta)}^{W(w,\theta)} \hat{\hat{Y}}(x,\theta;\delta)\ dx} d\mu(\theta)$$

and therefore, for all $\delta < \tilde{\delta}(e)$:

$$\int_{\theta\in\Theta} \frac{\left(1 - \frac{v_y(Y(W(w,\theta),\theta);W(w,\theta),\theta)}{u'(C(W(w,\theta),\theta)}\right) f(W(w,\theta)|\theta) + e}{v_{yw}(Y(W(w,\theta),\theta);W(w,\theta),\theta) - e} d\mu(\theta)$$
$$\leq \int_{\theta\in\Theta} \frac{\int_{W(w-\delta,\theta)}^{W(w,\theta)} \left(1 - \frac{v_y(Y(x,\theta);x,\theta)}{u'(C(x,\theta))}\right) \hat{\hat{Y}}(x,\theta;\delta) f(x|\theta) dx}{\int_{W(w-\delta,\theta)}^{W(w,\theta)} v_{yw}\left(Y(x,\theta);x,\theta\right)\ \hat{\hat{Y}}(x,\theta;\delta)\ dx} d\mu(\theta)$$
$$\leq \int_{\theta\in\Theta} \frac{\left(1 - \frac{v_y(Y(W(w,\theta),\theta);W(w,\theta),\theta)}{u'(C(W(w,\theta),\theta)}\right) f(W(w,\theta)|\theta) - e}{v_{yw}(Y(W(w,\theta),\theta);W(w,\theta),\theta) + e} d\mu(\theta)$$

Hence, left-hand side of (27) is equal to the left-hand side of (17). □

A.3 Proof of Lemma 5

With one-dimensional heterogeneity, we only consider within-group incentive constraints. Adopting a first-order approach, only (8a) is considered when building up the Hamiltonian:

$$\left(Y(w,\theta) - \mathscr{C}\left(Y(w,\theta), U(w,\theta); w,\theta\right) + \frac{\Phi\left(U(w,\theta); w,\theta\right)}{\lambda}\right) \cdot f(w|\theta) - q(w|\theta) \cdot v_w\left(Y(w,\theta); w,\theta\right).$$

where $Y(w,\theta)$ and $U(w,\theta)$ are the control and state variables respectively. Using (12), the necessary conditions are:

$$0 = \left(1 - \frac{v_y\left[w,\theta\right]}{u'\left[w,\theta\right]}\right) \cdot f(w|\theta) - q(w|\theta) \cdot v_{yw}\left[w,\theta\right] \quad (28a)$$

$$-\dot{q}\left(w|\theta\right) = \left(\frac{\Phi_U\left[w,\theta\right]}{\lambda} - \frac{1}{u'\left[w,\theta\right]}\right) \cdot f(w|\theta) \quad (28b)$$

$$0 = q(0|\theta) \quad (28c)$$

$$0 = \lim_{w \mapsto \infty} q(w|\theta). \quad (28d)$$

Combining (28b) with (28d) leads to

$$q(w|\theta) = \int_w^\infty \left(\frac{\Phi_U\left[w,\theta\right]}{\lambda} - \frac{1}{u'\left[w,\theta\right]}\right) \cdot f(\omega|\theta) d\omega. \quad (28e)$$

Combining (3), (2), (28a) and (28e) leads to (18a). Combining (28c) with (28e) leads to (18b).

A.4 Proof of Proposition 2

Define a reform of a tax schedule $y \mapsto T(y)$ with its direction, which is a differentiable function $y \mapsto R(y)$ defined on $\mathbb{R}_+$, and with its algebraic magnitude $m \in \mathbb{R}$. After a reform, the tax schedule becomes $y \mapsto T(y) - m\, R(y)$ and the utility of an individuals of type (w,θ) is:

$$U^R(m; w,\theta) \stackrel{\text{def}}{\equiv} \max_y \quad u(y - T(y) + m\, R(y)) - v(y; w,\theta). \quad (29)$$

We denote by $Y^R(m; w,\theta)$ the income of workers of types (w,θ) after the reform and her consumption becomes $C^R(m; w,\theta) = Y^R(m; w,\theta) - T(Y^R(m; w,\theta)) + m\ R(Y^R(m; w,\theta))$. When $m = 0$, we have $Y^R(0; w,\theta) = Y(w,\theta)$ and $C^R(0; w,\theta) = C(w,\theta)$. Applying the envelope theorem to (29), we get:

$$\frac{\partial U^R}{\partial m}(m; w,\theta) = u_c\left(C^R(m; w,\theta)\right)\ R(y). \quad (30)$$

Using (3), the first-order condition associated to (29) equalizes to zero the following expression:

$$\mathscr{Y}^R(y, m; w,\theta) \stackrel{\text{def}}{\equiv} 1 - T'(y) + m\, R'(y) - \mathscr{M}\left(y - T(y) + m\, R(y), y; w,\theta\right). \quad (31)$$

For simplicity, we drop the superscript R and write $\mathscr{Y}_y(Y(w,\theta); w,\theta)$ for $\mathscr{Y}_y^R(Y(w,\theta), 0; w,\theta)$.[11] We define behavioral responses to tax reforms of direction R by applying the implicit function theorem to $\mathscr{Y}^R(y, m; w,\theta) = 0$ at $m = 0$, which yields:

$$\frac{\partial Y^R}{\partial m}(0; w,\theta) = -\frac{\mathscr{Y}_m^R(Y(w,\theta), 0; w,\theta)}{\mathscr{Y}_y(Y(w,\theta), 0; w,\theta)} \quad (32)$$

[11] Indeed, at $m = 0$, $\mathscr{Y}_y^R$ does no longer depend on the direction R of the tax reform.

where:

$$\mathscr{Y}_y^R(y,m;w,\theta) = -T''(y) - \mathscr{M}_y(y - T(y) + m\, R(y), y; w, \theta) \tag{33a}$$
$$- \mathscr{M}(y - T(y) + m\, R(y), y; w, \theta)\ \mathscr{M}_c(y - T(y) + m\, R(y), y; w, \theta),$$
$$\mathscr{Y}_m^R(y,m;w,\theta) = R'(y) - R(y)\ \mathscr{M}_c(y - T(y) + m\, R(y), y; w, \theta). \tag{33b}$$

Using (2) and plugging $R(Y(w,\theta)) = 0$ and $R'(Y(w,\theta)) = 0$ into (33b), the compensated elasticity of earnings (19a) can be rewritten as:

$$\varepsilon(w,\theta) = \frac{\mathscr{M}(C(w,\theta), Y(w,\theta); w, \theta)}{-Y(w,\theta)\ \mathscr{Y}_y(Y(w,\theta); w, \theta)} > 0 \tag{34a}$$

which is positive since $\mathscr{Y}_y\,(Y(w,\theta);w,\theta) < 0$. Plugging $R(Y(w,\theta)) = 1$ and $R'(Y(w,\theta)) = 0$ into (33b), the income effect (19b) can be rewritten as:

$$\eta(w,\theta) = \frac{\mathscr{M}_c(C(w,\theta), Y(w,\theta); w, \theta)}{\mathscr{Y}_y(Y(w,\theta); w, \theta)} \tag{34b}$$

which is negative if leisure is a normal good, since then $\mathscr{M}_c > 0$. The elasticity $\alpha(w;\theta)$ of earnings with respect to the skill level can be expressed as:

$$\alpha(w,\theta) = \frac{w\ \mathscr{M}_w(C(w,\theta), Y(w,\theta); w, \theta)}{Y(w,\theta)\ \mathscr{Y}_y(Y(w,\theta); w, \theta)} > 0. \tag{34c}$$

Dividing (34a) by (34c) we get:

$$\frac{\varepsilon(w,\theta)}{\alpha(w,\theta)} = -\frac{v_y\,[w,\theta]}{w \cdot v_{yw}\,[w,\theta]}. \tag{35}$$

Plugging (34a) into (34b) leads to:

$$\eta(w,\theta) = Y(w,\theta) \cdot \frac{u''\,[w,\theta]}{u'\,[w,\theta]} \cdot \varepsilon(w,\theta).$$

It is then straightforward to obtain:

$$\hat{\eta}(Y(w,\theta_0)) = Y(w,\theta_0) \cdot \frac{u''\,[w,\theta_0]}{u'\,[w,\theta_0]} \cdot \hat{\varepsilon}(Y(w,\theta_0)). \tag{36}$$

Let $y \in \mathbb{R}_+$. Since $\mathscr{Y}_y\,(Y(w,\theta);w,\theta) < 0$, there exists a single skill level w such that $y = Y(w,\theta_0)$. From (2), we know that:

$$1 - T'\,[w,\theta] = \frac{v_y\,[w,\theta]}{u'\,[w,\theta]}. \tag{37}$$

The term in the left-hand side integral of (14a) can be rewritten as:

$$\frac{v_y\,[W(w,\theta),\theta]}{-W(w,\theta)\ v_{yw}\,[W(w,\theta),\theta]}\ W(w,\theta)\ f(W(w,\theta)|\theta) = \frac{\varepsilon\,(W(w,\theta),\theta)}{\alpha\,(W(w,\theta),\theta)} \cdot W(w,\theta)\ f(W(w,\theta)|\theta)$$
$$= \varepsilon\,(W(w,\theta),\theta)\ \ Y(w,\theta_0)\ h(Y(w,\theta_0)|\theta).$$

The first equality is obtained using Equation (35). The second equality uses (21). It implies with (22b) that Equation (14a) can be rewritten as:

$$\frac{T'\,[w,\theta_0]}{1 - T'\,[w,\theta_0]} \cdot \hat{\varepsilon}\,(Y(w,\theta_0)) \cdot Y(w,\theta_0) \cdot \hat{h}(Y(w,\theta_0)) = J(w) \tag{38}$$

where $J(w)$ is defined by the right-hand side of (14a). $J(\cdot)$ admits for derivative $\dot{J}(w)$ where:

$$\dot{J}(w) = \dot{C}(w,\theta_0)\frac{u''[w,\theta_0]}{u'[w,\theta_0]}J(w) + \int\limits_{\theta\in\Theta}\left\{\frac{\Phi_U[W(w,\theta),\theta]\ u'[W(w,\theta),\theta]}{\lambda}-1\right\}\dot{W}(w,\theta)\,f(W(w,\theta)|\theta)\,d\mu(\theta)$$
$$= \int_{\theta\in\Theta}\{g(W(w,\theta),\theta)-1\}\cdot\dot{W}(w,\theta)\cdot f(W(w,\theta;\theta_0)|\theta)\cdot d\mu(\theta)+\dot{C}(w,\theta_0)\cdot\frac{u''[w,\theta_0]}{u'[w,\theta_0]}\cdot J(w)$$

where (20) has been used. Deriving with respect to the skill w both sides of (9) and of $C(w,\theta_0) = Y(w,\theta_0) - T(Y(w,\theta_0))$, we obtain:

$$\dot{W}(w,\theta) = \frac{\dot{Y}(w,\theta_0)}{\dot{Y}(W(w,\theta),\theta)} \qquad \text{and} \qquad \dot{C}(w,\theta_0) = \left(1-T'(Y(w,\theta_0))\right)\ \dot{Y}(w,\theta_0).$$

We thus obtain:

$$\dot{J}(w) = \left(\int\limits_{\theta\in\Theta}\{g(W(w,\theta),\theta)-1\}\frac{f(W(w,\theta)|\theta)}{\dot{Y}(W(w,\theta),\theta)}d\mu(\theta)+\left(1-T'[w,\theta_0]\right)\frac{u''[w,\theta_0]}{u'[w,\theta_0]}J(w)\right)\dot{Y}(w,\theta_0).$$

Using (21) and (38), $\dot{J}(w)$ can be rewritten as:

$$\begin{aligned}\dot{J}(w) &= \Bigg(\int\limits_{\theta\in\Theta}\{g(W(w,\theta),\theta)-1\}\,h(Y(w,\theta_0)|\theta)\,d\mu(\theta)\\ &+ T'(Y(w,\theta_0))\,Y(w,\theta_0)\frac{u''(C(w,\theta_0))}{u'(C(w,\theta_0))}\hat{\varepsilon}(Y(w,\theta_0))\hat{h}(Y(w,\theta_0))\Bigg)\dot{Y}(w,\theta_0).\end{aligned}$$

Using (36) and (22d), we get:

$$-\dot{J}(w) = \{1-\hat{g}(Y(w,\theta_0))-\hat{\eta}(Y(w,\theta_0))\cdot T'(Y(w,\theta_0))\}\cdot\hat{h}(Y(w,\theta))\cdot\dot{Y}(w,\theta_0).$$

As $J(w) = \int_{x\geq w}(-\dot{J}(x))dx$, we get

$$J(w) = \int_{x\geq w}\{1-\hat{g}(Y(x,\theta_0))-\hat{\eta}(Y(x,\theta_0))\cdot T'(Y(x,\theta_0))\}\cdot\hat{h}(Y(x,\theta))\cdot\dot{Y}(x,\theta_0)\cdot dx.$$

Changing variables by posing $z = Y(x,\theta_0)$, we get

$$J(w) = \int_{z\geq Y(w,\theta_0)}\{1-\hat{g}(z)-\hat{\eta}(z)\cdot T'(Y(z))\}\cdot\hat{h}(Y(x,\theta))\cdot dz. \tag{39}$$

Plugging (39) into (38) gives (23a). Combining (14b) and (39) leads to (23b).

A.5 Proof of Proposition 3

Let us denote

$$K(w) \stackrel{\text{def}}{\equiv} \iint\limits_{\theta\in\Theta,x\geq W(w,\theta)}\left(\frac{1}{u'(C(x,\theta))}-\frac{\Phi_U(U(x,\theta);x,\theta)}{\lambda}\right)f(x|\theta)dx\,d\mu(\theta) \tag{40}$$

the ratio of the right-hand side of (14a) at the skill level w divided by $u'(Y(w,\theta_0)-T(Y(w,\theta_0)))$. According to Proposition 1, Equation (14a) and $v_y > 0 > v_{yw}$, the sign of $T'(Y(w,\theta_0))$ is the sign of $K(w)$.

Under utilitarian preferences, $\Phi_u = 1$. Changing variable in (40) from x to t such that $x = W(t,\theta)$, (i.e. $Y(x,\theta) \equiv Y(t,\theta_0)$ and $C(x,\theta) \equiv C(t,\theta_0)$ according to (9)), we get:

$$K(w) = \int_{t \geq w} \left(\frac{1}{u'(C(t,\theta_0))} - \frac{1}{\lambda} \right) \left(\int_{\theta \in \Theta} \dot{W}(t,\theta) \, f\left(W(t,\theta)|\theta\right) d\mu\left(\theta\right) \right) dt$$

The derivative of $K(w)$ has the sign of $1/\lambda - 1/u'(C(w,\theta_0))$, which is decreasing in w because of the concavity of $u(\cdot)$. Moreover, $\lim_{w \mapsto \infty} K(w) = 0$ and Equation (14b) implies that $K(0) = 0$. Therefore, $K(\cdot)$ first increases and then decreases. It is thus positive for all (interior) skill levels. So, optimal marginal tax rates are positive.

Under maximin, one has $U(x,\theta) > U(0,\theta)$ for all $x > 0$ from (8a). Therefore, within each group, the most deserving individuals are those whose skill $w = 0$. The maximin objective implies $\Phi_U [x,\theta] = 0$ for all $x > 0$. Hence, Equation (40) simplifies to:

$$K(w) = \iint_{\theta \in \Theta, x \geq W(w,\theta)} \frac{1}{u'(C(x,\theta))} f(x|\theta) dx \, d\mu(\theta)$$

for all $w > 0$, which is always positive, thereby leading to positive optimal marginal tax rates.

References

Akerlof, George A., "The Economics of "Tagging" as Applied to the Optimal Income Tax, Welfare Programs, and Manpower Planning," *The American Economic Review*, 1978, *68* (1), 8–19.

Blumkin, Tomer, Efraim Sadka, and Yotam Shem-Tov, "International tax competition: zero tax rate at the top re-established," *International Tax and Public Finance*, 2015, *22* (5), 760–776.

Boadway, Robin, Maurice Marchand, Pierre Pestieau, and María Del Mar Racionero, "Optimal Redistribution with Heterogeneous Preferences for Leisure," *Journal of Public Economic Theory*, 2002, *4* (4), 475–498.

_, _, _, and María del Mar Racionero, "Optimal Redistribution with Heterogeneous Preferences for Leisure," *Journal of Public Economic Theory*, 2002, *4* (4), 475–498.

Brett, Craig and John A. Weymark, "Financing education using optimal redistributive taxation," *Journal of Public Economics*, 2003, *87* (11), 2549 – 2569.

_ and _, "How optimal nonlinear income taxes change when the distribution of the population changes," *Journal of Public Economics*, 2011, *95* (11), 1239 – 1247.

Choné, Philippe and Guy Laroque, "Negative Marginal Tax Rates and Heterogeneity," *American Economic Review*, 2010, *100* (5), 2532–47.

Cremer, Helmuth, Jean-Marie Lozachmeur, and Pierre Pestieau, "Income taxation of couples and the tax unit choice," *Journal of Population Economics*, 2012, *25* (2), 763–778.

Cuff, Katherine, "Optimality of workfare with heterogeneous preferences," *Canadian Journal of Economics/Revue canadienne d'économique*, 2000, *33* (1), 149–174.

Diamond, Peter, "Optimal Income Taxation: An Example with U-Shaped Pattern of Optimal Marginal Tax Rates," *American Economic Review*, 1998, *88* (1), 83–95.

Gomes, Renato, Jean-Marie Lozachmeur, and Alessandro Pavan, "Differential Taxation and Occupational Choice," *Review of Economic Studies*, 2018, *85* (1), 511–557.

Guesnerie, Roger, *A Contribution to the Pure Theory of Taxation*, Cambridge: Cambridge University Press, 1995.

Hammond, Peter, "Straightforward Individual Incentive Compatibility in Large Economies," *The Review of Economic Studies*, 1979, *46* (2), 263–282.

Hellwig, Martin F., "A contribution to the theory of optimal utilitarian income taxation," *Journal of Public Economics*, 2007, *91* (7), 1449 – 1477.

Hendren, Nathaniel, "Measuring Economic Efficiency Using Inverse-Optimum Weights," *Journal of Public Economics*, 2020 2020, *187.*

Jacquet, Laurence and Etienne Lehmann, "Optimal Income Taxation with Composition Effects," *Journal of the European Economic Association*, 06 2020. jvaa022.

_ , _ , and Bruno Van der Linden, "Optimal redistributive taxation with both extensive and intensive responses," *Journal of Economic Theory*, 2013, *148* (5), 1770–1805.

Kleven, Henrik Jacobsen, Claus Thustrup Kreiner, and Emmanuel Saez, "The Optimal Income Taxation of Couples," *Econometrica*, 2009, *77* (2), 537–560.

Lehmann, Etienne, Laurent Simula, and Alain Trannoy, "Tax Me if You Can! Otimal Nonlinear Income Tax Between Competing Governments," *The Quarterly Journal of Economics*, 2014, *129* (4), 1995–2030.

Lockwood, Benjamin B. and Matthew Weinzierl, "De Gustibus non est Taxandum: Heterogeneity in Preferences and Optimal Redistribution," *Journal of Public Economics*, 2015, *124*, 74–80.

Mirrlees, James, "An Exploration in the Theory of Optimum Income Taxation," *Review of Economic Studies*, 1971, *38* (2), 175–208.

Piketty, Thomas, "La Redistribution fiscale contre le chômage," *Revue française d'économie*, 1997, *12* (1), 157–203.

Ramsey, Franck P., "A Contribution to the Theory of Taxation," *The Economic Journal*, 1927, *37* (145), 47–61.

Rochet, Jean, "The taxation principle and multi-time Hamilton-Jacobi equations," *Journal of Mathematical Economics*, 1985, *14* (2), 113–128.

Rochet, Jean-Charles and Lars A. Stole, "Nonlinear Pricing with Random Participation," *The Review of Economic Studies*, 2002, *69* (1), 277–311.

_ and Philippe Choné, "Ironing, Sweeping, and Multidimensional Screening," *Econometrica*, 1998, *66* (4), 783–826.

Rothschild, Casey and Florian Scheuer, "Redistributive Taxation in the Roy Model," *The Quarterly Journal of Economics*, 2013, *128* (2), 623–668.

_ and _ , "Optimal Taxation with Rent-Seeking," *Review of Economic Studies*, 2016, *83* (3), 1225–1262.

Sachs, Dominik, Aleh Tsyvinski, and Nicolas Werquin, "Nonlinear Tax Incidence and Optimal Taxation in General Equilibrium," *Econometrica*, 2020, *88* (2), 469–493.

Saez, Emmanuel, "Using Elasticities to Derive Optimal Income Tax Rates," *Review of Economic Studies*, 2001, *68* (1), 205–229.

_ , "Optimal Income Transfer Programs: Intensive versus Extensive Labor Supply Responses," *Quarterly Journal of Economics*, 2002, *117*, 1039–1073.

Salanié, Bernard, *The Economics of Taxation (second edition)* 2011.

Scheuer, Florian, "Adverse selection in credit markets and regressive profit taxation," *Journal of Economic Theory*, 2013, *148* (4), 1333 – 1360.

_ , "Entrepreneurial Taxation with Endogenous Entry," *American Economic Journal: Economic Policy*, May 2014, *6* (2), 126–63.

_ and Iván Werning, "Mirrlees meets Diamond-Mirrlees," Working Paper 22076, National Bureau of Economic Research March 2016.

Werning, Ivàn, "Pareto Efficient Income Taxation," MIT Working Paper 2007.

Weymark, John A., "A reduced-form optimal nonlinear income tax problem," *Journal of Public Economics*, 1986, *30* (2), 199 – 217.

_ , "Comparative Static Properties of Optimal Nonlinear Income Taxes," *Econometrica*, 1987, *55* (5), 1165–1185.

Wilson, Robert B., *Nonlinear Pricing*, Oxford University Press, 1993.