

Jacquet, Laurence; Lehmann, Etienne

**Working Paper**

## Optimal Income Taxation with Composition Effects

IZA Discussion Papers, No. 11019

**Provided in Cooperation with:**

IZA – Institute of Labor Economics

*Suggested Citation:* Jacquet, Laurence; Lehmann, Etienne (2017) : Optimal Income Taxation with Composition Effects, IZA Discussion Papers, No. 11019, Institute of Labor Economics (IZA), Bonn

This Version is available at:

<https://hdl.handle.net/10419/171003>

**Standard-Nutzungsbedingungen:**

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

**Terms of use:**

*Documents in EconStor may be saved and copied for your personal and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.*

DISCUSSION PAPER SERIES

IZA DP No. 11019

**Optimal Income Taxation with  
Composition Effects**

Laurence Jacquet  
Etienne Lehmann

SEPTEMBER 2017

## DISCUSSION PAPER SERIES

IZA DP No. 11019

# Optimal Income Taxation with Composition Effects

**Laurence Jacquet**

*THEMA, University of Cergy-Pontoise and University of Oslo*

**Etienne Lehmann**

*CRED (TEPP), University Panthéon-Assas Paris II, IZA, CESifo and CEPR*

SEPTEMBER 2017

Any opinions expressed in this paper are those of the author(s) and not those of IZA. Research published in this series may include views on policy, but IZA takes no institutional policy positions. The IZA research network is committed to the IZA Guiding Principles of Research Integrity.

The IZA Institute of Labor Economics is an independent economic research institute that conducts research in labor economics and offers evidence-based policy advice on labor market issues. Supported by the Deutsche Post Foundation, IZA runs the world's largest network of economists, whose research aims to provide answers to the global labor market challenges of our time. Our key objective is to build bridges between academic research, policymakers and society.

IZA Discussion Papers often represent preliminary work and are circulated to encourage discussion. Citation of such a paper should account for its provisional character. A revised version may be available directly from the author.

## ABSTRACT

---

# Optimal Income Taxation with Composition Effects\*

We study the optimal nonlinear income tax problem with multidimensional individual characteristics on which taxes cannot be conditioned. We obtain an optimal tax formula that generalizes the standard one by averaging, with specific weights, the sufficient statistics of individuals who earn the same income. Our first main contribution consists in showing that multidimensional heterogeneity brings a new source of endogeneity to the sufficient statistics that we call composition effects. We highlight that composition effects may substantially affect optimal marginal tax rates. Our results put the stress on the need for empirical studies on sufficient statistics for different demographic groups e.g., according to gender, age, ethnicity. As a second main contribution, we show the equivalence between the tax perturbation and mechanism design approaches which bridges the gap between both methods that have, so far, been used separately in the literature.

**JEL Classification:** H21

**Keywords:** optimal taxation, multidimensional screening problems, tax perturbation, allocation perturbation, sufficient statistics

**Corresponding author:**

Etienne Lehmann  
CRED (TEPP)  
University Panthéon-Assas Paris 2  
12 Place du Panthéon  
75 231, Paris Cedex 05  
France  
E-mail: [etienne.lehmann@gmail.com](mailto:etienne.lehmann@gmail.com)

---

\* Part of the material we propose in this paper has been presented in working papers under the following titles "Optimal Nonlinear Income Taxation with Multidimensional Types: The Case with Heterogeneous Behavioral Responses" (2014), "Optimal Income Taxation when Skills and Behavioral Elasticities are Heterogeneous" (2015) and "Optimal Taxation with Heterogeneous Skills and Elasticities: Structural and Sufficient Statistics Approaches" (2016). We thank Pierre Boyer, Craig Brett, Bas Jacobs, Guy Laroque, Stéphane Robin, Emmanuel Saez, Florian Scheuer, Stefanie Stantcheva, Alain Trannoy, Nicolas Werquin as well as participants at various seminars, conferences and workshops. All remaining errors are our own.

# I Introduction

The recent years have seen an increase in empirical analyses that provide so-called “sufficient statistics” (e.g., reduced-form elasticities) to give policy prescriptions. As a compromise between reduced-form and structural analyses, the approach based on sufficient statistics has applications in macroeconomics, labor economics, development economics, industrial organization and in international trade (e.g., [Chetty \(2009\)](#), [Hornstein et al. \(2011\)](#), [Arkolakis et al. \(2012\)](#)). It is however in public economics that the sufficient statistics approach has been the most widespread so far. The optimal tax literature in particular has seen to it that its tax policy prescriptions are easily implementable and relatively easy to explain to the general public. This has led researchers in this field to rely extensively on empirically meaningful sufficient statistics to express tax formulas (e.g., [Saez \(2001, 2002\)](#), [Kroft et al. \(2016\)](#) and references in [Chetty \(2009\)](#)). For this reason, in the present paper, optimal tax policy will be the field of choice to illustrate a more general point regarding the limitations of sufficient statistics in the presence of multidimensional heterogeneity.

In the optimal tax literature, the key sufficient statistics are behavioral responses to tax reforms, the income distribution and the social welfare weights which summarize the social preferences for redistribution (See e.g., [Diamond \(1998\)](#) and [Saez and Stantcheva \(2016\)](#)). One important and well-known limitation of sufficient statistics is their dependence to the tax schedule. Their values are distinct in the actual and optimal economies, which can bias the recommended tax schedule.

This problem is even more nagging if one wants to allow the unobserved individual heterogeneity to vary along multiple dimensions. Going beyond one dimension of unobserved heterogeneity is a requisite in order to obtain more empirically-meaningful tax formulas. In the case of the income tax, for instance, individuals who differ in many dimensions (such as health, gender, marital status and ethnicity), as is the case in real-world economies, are more likely to respond differently to any tax reform, even if they initially earn the same income.

The first contribution of this paper is to show that allowing for multidimensional heterogeneity in a model *à la* [Mirrlees \(1971\)](#) exacerbates the aforementioned bias in the tax schedule. More precisely, in the presence of multidimensional individual heterogeneity, we identify a new source of endogeneity in the sufficient statistics, that we call *composition effects*. Composition effects are changes in the composition of the population that may occur, at each income level, between the actual and optimal economies, when individuals at the same level have distinct multidimensional characteristics and therefore distinct behavioral responses.

As part of our first contribution, we highlight that the sufficient statistics which should be computed to calibrate optimal tax profiles are weighted averages of *total*<sup>1</sup> sufficient statistics in each group of workers who earn this income level, where a group of workers consists in

---

<sup>1</sup>Total sufficient statistics encapsulate the circularity process of optimal tax systems, i.e. they are evaluated along the nonlinear income tax schedule (as in [Jacquet et al. \(2013\)](#) and [Scheuer and Werning \(2016\)](#)) and not along a linearized one.

individuals who share all characteristics, except skill levels. What is crucial here is that the weights are given by group-specific income densities. In addition to this theoretical finding, we provide several examples of drastic changes that the US optimal marginal tax rates undergo when one neglects composition effects. In our numerical illustrations, top tax rates are modified up to 16.5 percentage points.

These findings have two implications for the empirical literature that provides sufficient statistics: First, they call for more empirical evidence regarding the distribution of income in various subgroups of the population. Second, they make clear that due to composition effects, what matters is to obtain the behavioral responses in subgroups of population and at distinct income levels to implement optimal tax schedules. Consequently, multidimensional heterogeneity and the implied composition effects appear as an important issue for policy design.

Our second contribution consists in bridging the gap between the mechanism design method, which is widely used in various fields in economics, and the tax perturbation method, which is more specific to the optimal taxation literature (although it may find applications in industrial organization as well, e.g. in nonlinear monopoly pricing problems). The optimal tax policy literature uses either method to determine optimal tax policies, but has so far treated the two methods as two separate ways to achieve the same objective. In the present paper, we show the equivalence between these methods, which had not been done until now.

To do so, we first provide the conditions under which the tax perturbation method of [Piketty \(1997\)](#) and [Saez \(2001\)](#) is valid with multidimensional heterogeneity. In a nutshell, these conditions imply that the marginal tax rate does not decrease too fast with income which ensures that individual behaviors respond smoothly to a tax reform. We then proceed to show (in Proposition 4) that, under multidimensional as well as under one-dimensional individual heterogeneity, the assumptions required by the tax perturbation and first-order mechanism design<sup>2</sup> approaches (Assumption 2 and Assumption 4, respectively) are equivalent. Therefore, both approaches are two faces of the same coin; they are equivalent. On the one hand, since [Mirrlees \(1971\)](#), the mechanism design method allows one to derive optimal tax profiles by finding the incentive-compatible allocation that maximizes the social objective subject to a resources constraint. The optimal allocation is obtained by verifying (usually using a Hamiltonian or a Lagrangian) that any incentive-compatible perturbation of the allocation does not lead to any first-order improvement. On the other hand, the tax perturbation approach seeks the tax reform that decentralizes such a perturbation.

As a by-product of the analysis, we derive the optimal tax schedule in terms of policy-invariant functions (i.e. skill density and the derivatives of the individual and social utility functions) instead of endogenous sufficient statistics. This formula is not closed-form because it depends on the allocation where these functions are evaluated, but it is much more convenient

---

<sup>2</sup>With multidimensional heterogeneity, we say that the mechanism design approach is first-order when income admits a strictly positive derivative with respect to skill, in each group of workers. This assumption implies (and plays a role similar to) the satisfaction of the second-order incentive compatibility condition when heterogeneity is one-dimensional.

to implement numerically. This rewriting of the tax formula also offers the possibility to sign marginal tax rates. To obtain this tax formula, beside the tax perturbation method, we propose a new mechanism design method, that we call an allocation perturbation, to solve a large class of tax problems with multidimensional heterogeneity.

The paper is organized as follows. Section II reviews the literature related to this paper. In Section III, we present the model. Section IV derives the optimal tax formula in terms of sufficient statistics using the tax perturbation method and sheds the light on composition effects. Section V numerically investigates the sensitivity of the optimal tax function to composition effects. Section VI provides the structural optimal tax formula and a sufficient condition for optimal marginal tax rates to be positive. Section VII shows the equivalence between the mechanism design and tax perturbation approaches. Section VIII concludes.

## II Related literature

The use of heuristic tax reform arguments in order to derive optimal nonlinear tax formulas goes back to Piketty (1997), Roberts (2000) and Saez (2001). Saez (2001) rigorously shows the consistency of these tax formulas with the ones obtained thanks to the first-order mechanism design approach of Mirrlees (1971), when heterogeneity is one-dimensional. Saez (2001) conjectures that his tax formula expressed in terms of sufficient statistics is also valid in the multidimensional environment. In this paper, we confirm that it is indeed the case by providing properly-constructed average measures of the usual sufficient statistics. We do so by following the general tax perturbation approach of Golosov et al. (2014), which is also used by Sachs et al. (2016), Gerritsen (2016) and Spiritus (2017). Scheuer and Werning (2016), who show that the optimal linear commodity tax model of Diamond and Mirrlees (1971) encapsulates the model of Mirrlees (1971), also point out that including multidimensional heterogeneity in the Mirrlees (1971) model leads to an optimal tax formula with simple averages of the sufficient statistics. Werning (2007) derives a condition for Pareto efficient tax schedules when heterogeneity is one-dimensional and also writes that his condition can simply be extended to multidimensional heterogeneity by averaging the sufficient statistics by group of individuals. Hendren (2014) develops inequality deflators which assess Pareto improvements when making second-best transfers through the income tax schedule. The method he uses to derive these inequality deflators is close to the tax perturbation approach and also allows for multidimensional heterogeneity. Beyond deriving the optimal tax formula under multidimensional heterogeneity, we contribute to this literature where multidimensionality prevails by studying the role of composition effects in the determination of optimal tax schedules and by connecting the first-order mechanism design and tax perturbation approaches.

Cuff (2000), Boadway et al. (2002), Brett and Weymark (2003), Choné and Laroque (2010), Lockwood and Weinzierl (2015) introduce in the Mirrlees (1971) model an additional source of heterogeneity, typically preferences for leisure or work opportunity cost, that matters only

for the computation of social welfare weights. The heterogeneity of preferences raises ethical questions which challenge the design of redistributive schemes. Individuals that have the same skill but distinct preferences for leisure, will earn, at the laissez-faire, different levels of income, with individuals having higher preferences for leisure earning less. The fraction of individuals with a high preference for leisure can be relatively higher among low income earners than among high income earners. The government may then be reluctant to redistribute towards the former, since some of them are also enjoying a higher quantity of leisure. In this context, the mean social welfare weight (the mean being taken across social welfare weights of individuals earning the same level of income) may, in contrast to the one-dimensional tax model, become less decreasing (Lockwood and Weinzierl, 2015) and even become increasing (e.g., Choné and Laroque (2010)) with income which opens the possibility for optimal marginal tax rates to be negative. In these papers, although agents differ in productivity and preference for leisure, their behavior is assumed to only depend on a unidimensional combination of the two underlying parameters. This aggregation implies that all individuals with a given income are constrained to respond identically to any tax reform despite the heterogeneity in social welfare weights. Therefore, to the contrary of our paper, composition effects on behavioral responses are assumed away.

Random participation models make up another strand of the literature where multidimensional heterogeneity is taken into account, although in a very specific way. In these models, individuals differ in skill and in a cost of participation (Rochet and Stole, 2002, Kleven et al., 2009, Jacquet et al., 2013) or of migration (Lehmann et al., 2014, Blumkin et al., 2015) and this latter dimension of heterogeneity matters only for the participation/migration margin. Scheuer (2013, 2014), Rothschild and Scheuer (2013), Gomes et al. (2017) consider optimal income tax models with different sectors<sup>3</sup> where agents can migrate from one sector to the other.<sup>4</sup> This is a form of random participation across sectors. Again, once individuals choose in which sectors to work (or which combination in Rothschild and Scheuer (2016)), income depends only on a single variable. While departing from this restriction, our model can readily be extended to include a participation margin simply following Jacquet et al. (2013). The tax formulas then simply incorporate new terms with the behavioral elasticities implied by the participation margin (see Jacquet et al. (2013)).

### III Model

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). Following Mirrlees (1971), the government

---

<sup>3</sup>Differing from our paper and from Scheuer (2013) and Rothschild and Scheuer (2013), nonlinear income taxation is sector-specific in Gomes et al. (2017) and in an extension of Scheuer (2014).

<sup>4</sup>In Gomes et al. (2017), types that derive the same utility across different sectors while supplying different labor are pooled together, whereas, in this paper, we will pool together individuals that generate the same income.



levies a non-linear tax  $T(\cdot)$  which depends on income  $y$  only. Consumption  $c$  is related to income  $y$  through the tax function  $T(y)$  according to  $c = y - T(y)$ . Individuals differ along their skill level  $w \in \mathbb{R}_+^*$  and along a vector of characteristics denoted  $\theta \in \Theta$ . We call a *group* a subset of individuals with the same  $\theta$ .<sup>5</sup> We assume that the set of groups  $\Theta$  is measurable with a cumulative distribution function (CDF) denoted  $\mu(\cdot)$ . The set  $\Theta$  can be finite or infinite and may be of any dimension. The distribution  $\mu(\cdot)$  of the population across the different groups may be continuous, but it may also exhibit mass points. Among individuals of the same group  $\theta$ , 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  $F(w|\theta) \stackrel{\text{def}}{=} \int_0^w f(x|\theta)dx$ . We do not make any restriction on the correlation between  $w$  or  $\theta$ . We normalize to unity the total size of the population.

### III.1 Individual choice

Individuals of type  $(w, \theta)$  have a twice continuously differentiable utility function with respect to  $c$  and  $y$  which is specified as  $\mathcal{U}(c, y; w, \theta)$  with  $\mathcal{U}_c > 0 > \mathcal{U}_y$ . We also assume that for each type  $(w, \theta)$ , indifference curves associated to  $\mathcal{U}(\cdot, \cdot; w, \theta)$  are strictly convex in the income-consumption space. Earning a given income requires less effort to a more productive worker, so  $\mathcal{U}_w > 0$ . A worker of type  $(w, \theta)$ , facing  $y \mapsto T(y)$ , solves:

$$U(w, \theta) \stackrel{\text{def}}{=} \max_y \mathcal{U}(y - T(y), y; w, \theta) \quad (1)$$

We call  $Y(w, \theta)$  the solution to program (1),<sup>6</sup>  $C(w, \theta) = Y(w, \theta) - T(Y(w, \theta))$  the consumption of a worker of type  $(w, \theta)$  and  $U(w, \theta)$  her utility. When the tax function is differentiable, the first-order condition associated to (1) implies that:

$$1 - T'(Y(w, \theta)) = \mathcal{M}(C(w, \theta), Y(w, \theta); w, \theta) \quad (2)$$

where:

$$\mathcal{M}(c, y; w, \theta) \stackrel{\text{def}}{=} -\frac{\mathcal{U}_y(c, y; w, \theta)}{\mathcal{U}_c(c, y; w, \theta)} \quad (3)$$

denotes the marginal rate of substitution between (pre-tax) income and consumption (after-tax income). 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 within each group of workers endowed with the same  $\theta$  that the marginal rate of substitution function is a decreasing function of the skill level, i.e. more skilled workers find it less costly to increase their income  $y$ :

**Assumption 1** (Within-group single-crossing condition). *For each  $\theta \in \Theta$ , and each  $(c, y) \in \mathbb{R}_+ \times \mathbb{R}_+$ , function  $w \mapsto \mathcal{M}(c, y; w, \theta)$  is differentiable with  $\forall w \in \mathbb{R}_+^*$ ,  $\mathcal{M}_w < 0$ .*

<sup>5</sup>Our definition of "group" is identical to the one in [Werning \(2007, p.15\)](#).

<sup>6</sup>If the maximization program (1) admits multiple solutions, we make the tie-breaking assumption that individuals choose among their best options the income level preferred by the government, i.e. the one with the largest tax liability.

Assumption 1 is for instance verified in the case where  $\mathcal{U}(c, y; w, \theta)$  is specified as:

$$\mathcal{U}(c, y; w, \theta) = u(c) - \frac{\theta}{1+\theta} \left(\frac{y}{w}\right)^{1+\frac{1}{\theta}} \quad \text{with} \quad \theta > 0 \quad \text{and} \quad u'(\cdot) > 0 \geq u''(\cdot). \quad (4)$$

We henceforth refer to this specification of preferences as the isoelastic ones. There  $\theta$  stands for the Frisch labor supply elasticity. The marginal rate of substitution equals  $\mathcal{M}(c, y; w, \theta) = y^{\frac{1}{\theta}} / [u'(c) w^{1+\frac{1}{\theta}}]$  and is decreasing in  $w$  from infinity to zero.

### III.2 Government

The government's budget constraint takes the form:

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

where  $E \geq 0$  is an exogenous amount of public expenditures. The objective of the planner is to maximize a general social welfare function that sums over all types of individuals an increasing and weakly concave transformation  $\Phi(U; w, \theta)$  of individuals' utility levels  $U$ :

$$\iint_{\theta \in \Theta, w \in \mathbb{R}_+^*} \Phi(U(w, \theta); w, \theta) f(w|\theta) dw d\mu(\theta) \quad (6)$$

This welfarist specification allows  $\Phi$  to vary with type  $(w, \theta)$  which makes it very general. *Weighted utilitarian* preferences 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 *utilitarian* if  $\varphi(w, \theta)$  is constant and  $\Phi(U; w, \theta) \equiv U$  and it turns out to be *maximin (or Rawlsian)* if  $\varphi(0, \theta) > 0$  while  $\varphi(w, \theta) = 0 \forall w > 0$ . When  $\Phi(U; w, \theta)$  does not vary with its two last arguments and  $\Phi_{UU} \leq 0$ , we obtain a Bergson-Samuelson criterion which is a concave transformation of utility and does not depend on individual characteristics.

The government's problem consists in finding the tax schedule  $T(\cdot)$  that maximizes social welfare function (6) subject to the budget constraint (5). Let  $\lambda > 0$  be the Lagrange multiplier associated with the budget constraint (5). The Lagrangian is:

$$\mathcal{L} \stackrel{\text{def}}{=} \iint_{\theta \in \Theta, w \in \mathbb{R}_+^*} \left[ T(Y(w, \theta)) + \frac{\Phi(U(w, \theta); w, \theta)}{\lambda} \right] f(w|\theta) dw d\mu(\theta) \quad (7)$$

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

$$g(w, \theta) \stackrel{\text{def}}{=} \frac{\Phi_U(U(w, \theta); w, \theta) \mathcal{U}'_c(C(w, \theta), Y(w, \theta); w, \theta)}{\lambda} \quad (8)$$

The government values giving one extra dollar to a worker  $(w, \theta)$  as a gain of  $g(w, \theta)$  dollar(s) of public funds.<sup>7</sup>

<sup>7</sup>We can easily extend our analysis to *non-welfarist* social criteria following the method of generalized marginal social welfare weights developed in [Saez and Stantcheva \(2016\)](#) to reflect non-welfarist views of justice which can be particularly relevant with heterogeneous preferences. Complementary to their approach, [Fleurbaey and Maniquet \(2011, 2017\)](#) connect the axioms of fair income tax theory and optimal income taxation and emphasize that it is not always straightforward to derive generalized marginal social welfare weights by income level.

## IV Optimal tax schedule using tax perturbation

Piketty (1997) and Saez (2001) were the first to derive the optimal income tax formula using a tax perturbation approach. However, the tax perturbation approach is valid only under some unclarified circumstances and this is the reason why Saez (2001) needed to show that his tax formula was consistent with the one of Mirrlees (1971). However, he did so only when individual heterogeneity is one-dimensional. In this section, we first state sufficient conditions for using a tax perturbation method and using the latter, we derive the optimal tax system when individual characteristics are multidimensional. We express the optimal tax formulas in terms of empirically measurable sufficient statistics. The mechanism design approach, which is the alternative method to obtain the optimal tax system, will be presented in Section VII. In the latter section, we will bridge the gap between both approaches.

### IV.1 Sufficient conditions for a tax perturbation

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

$$U^R(m; w, \theta) \stackrel{\text{def}}{=} \max_y \mathcal{U}(y - T(y) + m R(y), y; w, \theta) \quad (9)$$

We denote  $Y^R(m; w, \theta)$  the income of workers of types  $(w, \theta)$  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 (9), we get:

$$\frac{\partial U^R}{\partial m}(m; w, \theta) = \mathcal{U}_c \left( C^R(m; w, \theta), Y^R(m; w, \theta); w, \theta \right) R(y) \quad (10)$$

Using (3), the first-order condition associated to (9) equalizes to zero the following expression:

$$\mathcal{Y}^R(y, m; w, \theta) \stackrel{\text{def}}{=} 1 - T'(y) + m R'(y) - \mathcal{M}(y - T(y) + m R(y), y; w, \theta). \quad (11)$$

For simplicity, we drop the superscript  $R$  and write  $\mathcal{Y}_y(Y(w, \theta); w, \theta)$  for  $\mathcal{Y}_y^R(Y(w, \theta), 0; w, \theta)$  since at  $m = 0$ ,  $\mathcal{Y}_y^R$  does no longer depend on the direction  $R$  of the tax reform. To use the tax perturbation method, one needs the following assumptions:

**Assumption 2.** *Sufficient conditions for a tax perturbation.*

- i) *The tax function  $T(\cdot)$  is twice differentiable.*
- ii) *For all  $(w, \theta) \in \mathbb{R}_+^* \times \Theta$ , the second-order condition holds strictly:  $\mathcal{Y}_y(Y(w, \theta); w, \theta) < 0$ .*
- iii) *For all  $(w, \theta) \in \mathbb{R}_+^* \times \Theta$ , the function  $y \mapsto \mathcal{U}(y - T(y), y; w, \theta)$  admits a unique global maximum over  $\mathbb{R}_+$ .*

Part *i*) of Assumption 2 ensures that first-order condition (11) is differentiable. Part *ii*) guarantees it is invertible in income  $y$ . Under *i*) and *ii*), one can apply the implicit function theorem to (11) to describe how a local maximum of the individual maximization program (9) changes after a tax reform. Part *iii*) ensures that after an incremental tax reform or change in skill, the maximum remains global. Indeed since the tax function is nonlinear, the function  $y \mapsto \mathcal{U}(y - T(y) + mR(y), y; w, \theta)$  may in general admit several global maxima among which individuals of type  $(w, \theta)$  are indifferent. Any small tax reform may then lead to a jump in individual's choice from one maximum to another one (which is associated to a jump in the supply of effort). Part *iii*) prevents this situation and ensures the allocation changes in a differentiable way with the magnitude  $m$  of a tax reform. Assumption 2 plays the same role as e.g., Assumption 1 in Hendren (2014).

Let us emphasize circumstances under which the tax perturbation approach can be used because Assumption 2 is automatically satisfied. This is the case when the tax function  $T(y)$  is restricted to be linear as the indifference curves associated to  $\mathcal{U}(\cdot, \cdot; w, \theta)$  are strictly convex. Similarly, Assumption 2 is also satisfied when the tax function  $T(y)$  is convex ( $y \mapsto y - T(y)$  being concave, Parts *ii*) and *iii*) are then verified). By continuity, Assumption 2 is also verified when  $y \mapsto T(y)$  is “not too concave”, more precisely when  $y \mapsto y - T(y)$  is less convex than the indifference curve with which it has a tangency point in the  $(y, x)$ -plane (so that Part *ii*) of Assumption 2 is satisfied) and that this indifference curve is strictly above  $y \mapsto y - T(y)$  for all other  $y$  (so that Part *iii*) of Assumption 2 is satisfied). In a nutshell, Assumption 2 is satisfied whenever the marginal tax rate does not decrease too rapidly with income.

## IV.2 Perturbations and elasticities

In this subsection, we derive the behavioral responses to small perturbations of the tax function. We use them to formulate the way income reacts to any tax reform. The elasticities we obtain are also helpful to define the sufficient statistics relevant to the optimal tax formula with multidimensional heterogeneity (see Subsection IV.4 and onwards).

Apply the implicit function theorem to  $\mathcal{Y}^R(y, m; w, \theta) = 0$  at  $(y = Y^R(m; w, \theta), m = 0; w, \theta)$ . We obtain:

$$\frac{\partial Y^R}{\partial m} = -\frac{\mathcal{Y}_m^R}{\mathcal{Y}_y^R}$$

with:

$$\begin{aligned} \mathcal{Y}_y^R(y, m; w, \theta) &= -T''(y) - \mathcal{M}_y(y - T(y) + mR(y), y; w, \theta) \\ &\quad - \mathcal{M}(y - T(y) + mR(y), y; w, \theta) \mathcal{M}_c(y - T(y) + mR(y), y; w, \theta) \end{aligned} \quad (12a)$$

$$\mathcal{Y}_m^R(y, m; w, \theta) = R'(y) - R(y) \mathcal{M}_c(y - T(y) + mR(y), y; w, \theta). \quad (12b)$$

We consider two types of tax perturbations to derive the behavioral elasticities. First, we analyze the effects of small changes in the marginal tax rates to capture substitution effects. Second, the income effects are isolated thanks to a uniform transfer to all workers.

Consider first a change in the marginal tax rate by a constant amount  $m$  around income  $Y(w, \theta)$  and leave unchanged the level of tax at this income level. This reform is therefore given the label *compensated*. Formally, the direction of this compensated reform is  $R(y) = y - Y(w, \theta)$ ; it does not modify the tax level in  $y = Y(w, \theta)$  (i.e.  $R(Y(w, \theta)) = 0$ ) and it uniformly modifies the marginal tax rate as can be seen from  $R'(Y(w, \theta)) = 1$ . Using (2) and (3), we define the *compensated elasticity* of earnings with respect to the marginal retention rate  $1 - T'(\cdot)$  as:

$$\varepsilon(w; \theta) \stackrel{\text{def}}{=} \frac{1 - T'(Y(w, \theta))}{Y(w, \theta)} \frac{\partial Y^c}{\partial m} = \frac{\mathcal{M}(C(w, \theta), Y(w, \theta); w, \theta)}{-Y(w, \theta) \mathcal{Y}_y(Y(w, \theta); w, \theta)} > 0 \quad (13a)$$

where the superscript "c" emphasizes that the change of  $Y(w, \theta)$  is due to the compensated tax reform. The compensated elasticity is positive from Assumption 2.

To capture income effects, we consider a uniform transfer of money to all workers and call this reform a *lump-sum* one. This reform is obtained thanks to  $R(y) \equiv 1$ . Define the *income effect* as:

$$\eta(w; \theta) \stackrel{\text{def}}{=} \frac{\partial Y^L}{\partial m} = \frac{\mathcal{M}_c(C(w, \theta), Y(w, \theta); w, \theta)}{\mathcal{Y}_y(Y(w, \theta); w, \theta)} \quad (13b)$$

where the superscript "L" stresses that the change of  $Y(w, \theta)$  is due to the lump-sum reform. If leisure is a normal good, one has  $\mathcal{M}_c > 0$ , in which case  $\eta(w, \theta) < 0$ .

Combining (13a) and (13b) with (12b), the way income of individuals  $(w, \theta)$  reacts to any tax reform  $R(\cdot)$  is given by:

$$\frac{\partial Y^R}{\partial m}(0; w, \theta) = \varepsilon(w, \theta) \frac{Y(w, \theta)}{1 - T'(Y(w, \theta))} R'(Y(w, \theta)) + \eta(w, \theta) R(Y(w, \theta)) \quad (13c)$$

where the compensated elasticity and income effect show up.

Another relevant elasticity, the *elasticity of earnings with respect to skill  $w$* , can be built up under Assumption 2. Apply the implicit function theorem to (11) with respect to skill  $w$ . Note that this ensures that income  $Y(\cdot, \theta)$  is a continuously differentiable function in skill. Using  $\mathcal{Y}_w^R = -\mathcal{M}_w$ , the elasticity of earnings with respect to skill  $w$  is defined as:

$$\alpha(w; \theta) \stackrel{\text{def}}{=} \frac{w}{Y(w, \theta)} \dot{Y}(w, \theta) = \frac{w \mathcal{M}_w(C(w, \theta), Y(w, \theta); w, \theta)}{Y(w, \theta) \mathcal{Y}_y(Y(w, \theta); w, \theta)} > 0 \quad (13d)$$

which is positive from Assumption 1. Note that Assumptions 1 and 2 rule out bunching.<sup>8</sup> This is because, under Assumption 1, bunching can only be decentralized by a kink in the tax function with increasing marginal tax rates (Saez, 2010). However, a tax function with such a kink violates Part i) of Assumption 2.

Denote  $h(y|\theta)$  the conditional income density within group  $\theta$  at income  $y$  and  $H(y|\theta) \stackrel{\text{def}}{=} \int_{z=0}^y h(z|\theta) dz$  the corresponding conditional income CDF. According to (13d) and Assumption 1, income  $Y(\cdot, \theta)$  is strictly increasing in skill within each group. We then have  $H(Y(w, \theta)|\theta) \equiv$

<sup>8</sup>In our context of multidimensional characteristics, *bunching* refers to the specific situation where individuals who earn the same income belong to the same group  $\theta$  but have distinct skills. In contrast, *pooling* refers to a situation where individuals who earn the same income belong to distinct groups. Since we address multidimensional problems, we can study pooling and neglect bunching without any loss in generality.

$F(w|\theta)$  for each skill level  $w$ . Differentiating both sides of this equality with respect to  $w$  and using (13d) leads to:

$$h(Y(w, \theta)|\theta) = \frac{f(w|\theta)}{\dot{Y}(w, \theta)} \quad \Leftrightarrow \quad Y(w, \theta) h(Y(w, \theta)|\theta) = \frac{w f(w|\theta)}{\alpha(w, \theta)} \quad (14)$$

### IV.3 Total vs direct elasticities and income responses

Our definitions of elasticities and income response (13a)-(13d) account for the nonlinearity of the income tax schedule. In the denominators of these definitions, the term  $T''(Y(w, \theta))$  (which is incorporated in  $\mathcal{Y}_Y$  (see Equation (12a))) emphasizes the role played by the local curvature of the tax schedule. We refer to these elasticities and income response as representing *total* responses i.e., including the circular process induced by the endogeneity of marginal tax rates. Recent papers in the literature e.g., [Jacquet et al. \(2013\)](#) and [Scheuer and Werning \(2017\)](#) use total elasticities and income responses which help streamline tax formulas. In contrast, the empirical literature estimates responses that do not take into account the local curvature of the tax function, which we refer to as *direct* responses.<sup>9</sup> Let  $\varepsilon^*(w; \theta)$ ,  $\eta^*(w; \theta)$  and  $\alpha^*(w; \theta)$  denote these direct responses, i.e. the compensated elasticity of earnings with respect to the marginal retention rate, the income effect and the elasticity of earnings with respect to skill, when  $T'' = 0$  in (12a) and thereby in (13a)-(13d). These would be the relevant concepts if the tax function were linear.

To better grasp the distinction between direct and total responses, consider an exogenous change in  $w$ , or a tax reform inducing a *direct* change in earnings  $\Delta_1 y$  proportional to the direct response  $\varepsilon^*(w; \theta)$ ,  $\eta^*(w; \theta)$  and  $\alpha^*(w; \theta)$ . However, when the tax schedule is nonlinear, this direct response in earnings  $Y$  modifies the marginal tax rate by  $\Delta_1 T' = T''(Y(w, \theta)) \times \Delta_1 y$ , thereby inducing a further change in earnings  $\Delta_2 y = -Y(w, \theta) \frac{T''(Y(w, \theta))}{1 - T'(Y(w, \theta))} \varepsilon^*(w, \theta) \Delta_1 y$ . This second change in earnings, in turn, induces a further modification in the marginal tax rate  $T''(Y(w, \theta)) \times \Delta_2 y$  which induces an additional change in earnings. Therefore, a *circular process* takes place. The income level determines the marginal tax rate through the tax function, and the marginal tax rate affects the income level through the substitution effects. Using the identity

<sup>9</sup>To estimate the behavioral responses to tax reforms, there are two main methodologies. The first one uses actual tax reforms as a quasi-experimental design. What allows empirical researchers to identify the causal effect of tax on individual behavior is the exogenous shock induced by a tax reform. This shock makes some taxpayers face a change in their marginal tax rate while others do not ([Feldstein, 1995](#), [Auten and Carroll, 1999](#), [Gruber and Saez, 2002](#)), see [Saez et al. \(2012\)](#). By using two-stages least squares estimators, this approach implicitly assumes that marginal tax rates do locally not depend on taxable income. It thus identifies *direct* behavioral responses. The second method identifies behavioral responses from discontinuities in the distribution of taxable income around kinks ([Saez, 2010](#)) or notches ([Kleven and Waseem, 2013](#)) observed in tax schedules. In particular, around a convex kink where  $T''(\cdot) = +\infty$ , the *total* skill elasticity  $\alpha(w, \theta)$  is nil from (15c) which triggers bunching around the kink. One then uses the relation between the magnitude of this bunching and the *direct* compensated elasticity to identify the latter.

$1 + x + x^2 + x^3 + \dots = \frac{1}{1-x}$ , the total effect is given by:

$$\begin{aligned}\Delta y &= \sum_{i=1}^{\infty} \Delta_i y = \Delta_1 y \sum_{i=1}^{\infty} \left( -Y(w, \theta) \frac{T''(Y(w, \theta))}{1 - T'(Y(w, \theta))} \varepsilon^*(w, \theta) \right)^{i-1} \\ &= \Delta_1 y \frac{1 - T'(Y(w, \theta))}{1 - T'(Y(w, \theta)) + Y(w, \theta) T''(Y(w, \theta)) \varepsilon^*(w, \theta)}\end{aligned}$$

where the ratio in the latter equality is positive whenever  $-Y(w, \theta) \frac{T''(Y(w, \theta))}{1 - T'(Y(w, \theta))} \varepsilon^*(w, \theta)$  is lower than 1, i.e. whenever the second-order condition holds strictly. This ratio is the *corrective term* by which direct responses must be timed to obtain total responses as made explicit by the following equations:

$$\varepsilon(w, \theta) = \frac{1 - T'(Y(w, \theta))}{1 - T'(Y(w, \theta)) + Y(w, \theta) T''(Y(w, \theta)) \varepsilon^*(w, \theta)} \varepsilon^*(w, \theta) \quad (15a)$$

$$\eta(w, \theta) = \frac{1 - T'(Y(w, \theta))}{1 - T'(Y(w, \theta)) + Y(w, \theta) T''(Y(w, \theta)) \varepsilon^*(w, \theta)} \eta^*(w, \theta) \quad (15b)$$

$$\alpha(w, \theta) = \frac{1 - T'(Y(w, \theta))}{1 - T'(Y(w, \theta)) + Y(w, \theta) T''(Y(w, \theta)) \varepsilon^*(w, \theta)} \alpha^*(w, \theta) \quad (15c)$$

Equalities (15a)-(15c) are obtained from the definitions of elasticities, income responses and from (2).<sup>10</sup> In the real world, most of income tax schedules are piecewise linear. In this case, direct and total responses differ at the kinks of tax schedules which makes the distinction between both particularly relevant.<sup>11</sup>

#### IV.4 Sufficient statistics

We can now define the sufficient statistics that will play a role in the tax formula. Let  $W(\cdot, \theta)$  denote the reciprocal of  $Y(\cdot, \theta)$  so that, within each group  $\theta$ , individuals of type  $(w = W(y, \theta), \theta)$  earn income  $y$ . According to Assumption 1,  $W(y, \theta)$  is the unique skill level  $w$  such that the individual first-order condition  $1 - T'(y) = \mathcal{M}(y - T(y), y; w, \theta)$  is verified at income  $y$ . The *unconditional* income density is given by:

$$\hat{h}(y) \stackrel{\text{def}}{=} \int_{\theta \in \Theta} h(y|\theta) d\mu(\theta) \quad (16a)$$

The mean *total* compensated elasticity at income level  $y$  is:

$$\hat{\varepsilon}(y) = \int_{\theta \in \Theta} \varepsilon(W(y, \theta), \theta) \frac{h(y|\theta)}{\hat{h}(y)} d\mu(\theta) \quad (16b)$$

<sup>10</sup>From (12a) and (13a) we can write:

$$\frac{\varepsilon(y, \theta)}{\varepsilon^*(y, \theta)} = \frac{\mathcal{M}_y + \mathcal{M} \mathcal{M}_c}{T'' + \mathcal{M}_y + \mathcal{M} \mathcal{M}_c}$$

Substituting (3) into (2) and using the definition of  $\varepsilon^*(y, \theta)$  yields (15a). The same goes for Equations (15b) and (15c).

<sup>11</sup>Within tax brackets, the linearity of the tax implies  $T''(y) = 0$  so that total and direct responses are identical. At kinks, two possibilities can occur. First, the marginal tax rate can increase, i.e.  $T''(y) = +\infty$ , which implies that total responses are nil and bunching prevails (since  $\alpha(w, \theta) = 0$  from (15c)). Second, the marginal tax rate can decrease at kinks, i.e.  $T''(y) = -\infty$ , so that the second-order condition which is necessary for using tax perturbations (see Assumption 2) is no longer satisfied. Intuitively, one has  $\alpha(w, \theta) = \infty$  and income jumps.

where each within-group total elasticity 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  $y$  is:

$$\hat{\eta}(y) = \int_{\theta \in \Theta} \eta(W(y, \theta), \theta) \frac{h(y|\theta)}{\hat{h}(y)} d\mu(\theta) \quad (16c)$$

Finally, the mean marginal social welfare weight at income level  $y$  is:

$$\hat{g}(y) = \int_{\theta \in \Theta} g(W(y, \theta), \theta) \frac{h(y|\theta)}{\hat{h}(y)} d\mu(\theta) \quad (16d)$$

## IV.5 Desirable tax reforms

Having defined the general tax reforms and described their impact on individual income, we now study when a given tax reform is desirable. To do so, we locally perturb the tax system in a direction  $R(y)$  with magnitude  $m$ . The initial tax system can be optimal or suboptimal. If the initial tax schedule  $T(\cdot)$  is optimal, such a perturbation should not yield any first-order effect on the Lagrangian (7).

**Lemma 1.** *Under Assumptions 1 and 2, reforming the tax schedule in the direction  $R(\cdot)$  triggers first-order effects on the Lagrangian (7) equal to:*

$$\begin{aligned} \frac{\partial \mathcal{L}^R}{\partial m} &= \int_{y=0}^{\infty} \left\{ [\hat{g}(y) - 1 + T'(y) \hat{\eta}(y)] \hat{h}(y) - \frac{d}{dy} \left[ \frac{T'(y)}{1 - T'(y)} \hat{\varepsilon}(y) y \hat{h}(y) \right] \right\} R(y) dy \quad (17) \\ &+ \lim_{y \rightarrow \infty} \frac{T'(y)}{1 - T'(y)} \hat{\varepsilon}(y) y \hat{h}(y) R(y) - \lim_{y \rightarrow 0} \frac{T'(y)}{1 - T'(y)} \hat{\varepsilon}(y) y \hat{h}(y) R(y) \end{aligned}$$

The proof which is relegated to Appendix A.1 is in the vein of Golosov et al. (2014). It is based on studying perturbations of a given non-linear tax system taking into account the partial (Gateaux) differential of government tax revenue and social welfare with respect to tax reforms in the direction  $R(\cdot)$ .

An important point to notice is that, in general, implementing a reform with direction  $R$  implies a budget surplus or deficit. A first-order approximation of this budget surplus (or deficit) can be computed by putting social welfare weights  $\hat{g}(\cdot)$  equal to zero in (17). One can then define a balanced-budget tax reform with magnitude  $m$  and direction  $R$  by combining it with the lump-sum rebate required to bind the budget constraint. Appendix A.1 shows that the first-order effect of this balanced-budget tax reform on the social objective is positively proportional to the first-order effect of the tax reform with magnitude  $m$  and direction  $R$  on the Lagrangian. Expression (17) is therefore useful to determine which tax reforms are desirable. If (17) is positive, it is socially desirable to implement a tax reform with direction  $R(\cdot)$  and a positive magnitude  $m$  and to combine this reform with a lump-sum rebate to keep the government's budget balanced. Symmetrically, if Expression (17) is negative, it is socially desirable to implement a tax reform with direction  $-R$  and positive magnitude  $m$  combined with a lump-sum transfer.



## IV.6 Optimal tax schedule and composition effects

We now characterize the optimal tax schedule in the model with multidimensional types. The proof is in Appendix A.2.

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

$$\frac{T'(y)}{1 - T'(y)} = \frac{1}{\hat{\varepsilon}(y)} \frac{1 - \hat{H}(y)}{y \hat{h}(y)} \left( 1 - \frac{\int_y^\infty [\hat{g}(z) + \hat{\eta}(z) T'(z)] \hat{h}(z) dz}{1 - \hat{H}(y)} \right) \quad (18a)$$

$$1 = \int_0^\infty [\hat{g}(z) + \hat{\eta}(z) T'(z)] \hat{h}(z) dz. \quad (18b)$$

If income effects were assumed away, Equation (18b) would imply that the weighted sum of social welfare weights is equal to 1. In the presence of income effects, a uniform increase in tax liability induces a change in tax revenue proportional to the marginal tax rate which explains the presence of  $\hat{\eta}(z) \cdot T'(z)$ .

The optimal tax rate given in Equation (18a) consists in three terms: *i*) the behavioral responses to taxes  $\frac{1}{\hat{\varepsilon}(y)}$ , which, in the vein of Ramsey (1927), is the inverse of the mean compensated elasticity; *ii*) the shape of the income distribution measured by the relative hazard rate or local Pareto parameter  $\frac{1 - \hat{H}(y)}{y \hat{h}(y)}$  of the income distribution and *iii*) the social preferences and income effects  $1 - \frac{\int_y^\infty [\hat{g}(z) + \hat{\eta}(z) \cdot T'(z)] \hat{h}(z) dz}{1 - \hat{H}(y)}$ . This term indicates the distributional benefits of increasing the tax liability by one unit for all workers with incomes above  $y$ . Saez (2001) discusses how the optimal tax rate is affected by each of these three terms in the one-dimensional case. Shifting from the model with one dimension of heterogeneity to the model with multiple dimensions leads to replacing the marginal social welfare weight, the compensated elasticity and the income effect by their means calculated at a given income level. It is the mean of the *total* (rather than *direct*) compensated elasticity and income effect that must be computed.<sup>12</sup> Note that the averaging procedure is a far cry from the simple extension of the unidimensional case that would consist in computing the simple average of every estimated sufficient statistic and then multiplying each average by the same corrective term. Instead, every optimal sufficient statistic at any income level is a weighted average that requires as many corrective terms as there are groups in which individuals earn this income level and group-specific densities as weights (as described in Equations (15a)-(15b) and (16b)-(16d)).

The optimal tax formulas in Equations (18a) and (18b) are functions of sufficient statistics  $\hat{h}(y)$ ,  $\hat{\varepsilon}(y)$ ,  $\hat{\eta}(y)$  and  $\hat{g}(y)$  which are endogenous to the tax schedule. Therefore, plugging these sufficient statistics estimated in the actual economy into the optimal tax formulas (18a) and (18b) may lead to biased results. This is because these sufficient statistics typically take different values in the actual economy and in the optimal one. This well-known limit of sufficient statistics formulas already prevails when unobserved heterogeneity is one-dimensional.

<sup>12</sup>This is more intuitive than using the direct elasticity and income effect, which implies to encapsulate the circularity (described by (15a)-(15c)) in a so-called “virtual density” as in Saez (2001), Equation (13).

However, when heterogeneity is multidimensional, additional mechanisms appear which exacerbate the discrepancies between sufficient statistics in both economies. We now describe these additional mechanisms that we call *composition effects*. In a nutshell, composition effects stem, at every income level, from the prevalence of a distinct composition of population in the actual and optimal economies.

According to Equation (16a)-(16d), sufficient statistics  $\hat{\varepsilon}(y)$ ,  $\hat{\eta}(y)$  and  $\hat{g}(y)$  at a given income level  $y$  are weighted averages of group-specific sufficient statistics  $\varepsilon(y|\theta)$ ,  $\eta(y|\theta)$  and  $g(y|\theta)$ , where the weights are given by the relative proportion  $h(y|\theta)/\hat{h}(y)$  of individuals of group  $\theta$  among individuals who earn income  $y$ . Composition effects take place since these relative proportions are also endogenous to the tax schedule. Therefore, among individuals earning a given income  $y$ , there may be relatively fewer (more) individuals with lower compensated elasticity  $\varepsilon(y|\theta)$ , income response  $\eta(y|\theta)$  or welfare weight  $g(y|\theta)$  in the optimal economy than in the actual one. Therefore,  $\hat{\varepsilon}(y)$ ,  $\hat{\eta}(y)$  or  $\hat{g}(y)$  rise (shrink) when moving from the actual to the optimal economy. In the next section, we numerically illustrate that the optimal marginal tax rates may be highly sensitive to composition effects. It also allows us to carefully explain how composition effects play a role at all income levels in tax formulas.

## V A quantitative investigation of the composition effects

In this section, we numerically compare optimal tax rates when taking or not composition effects into account. As usual in the optimal tax literature, we assume income effects away with quasilinear preferences of the form:

$$\mathcal{U}(c, y; w, \theta) = c - \frac{\theta}{1+\theta} y^{1+\frac{1}{\theta}} w^{-\frac{1}{\theta}} \quad w \in \mathbb{R}_+^*, \theta \in \Theta$$

This specification implies a constant direct compensated elasticity of earnings with respect to the marginal retention rate,  $\varepsilon^*(w, \theta) = \theta$  and a direct elasticity of earnings with respect to skill normalized to 1,  $\alpha^*(w, \theta) = 1$ . The first-order condition (2) simplifies to:

$$1 - T'(y) = \left(\frac{y}{w}\right)^{\frac{1}{\theta}} \quad \Leftrightarrow \quad Y(w, \theta) = (1 - T'(y))^{\theta} w$$

At the optimum, a taxpayer of group  $\theta$  earning income  $y$  is endowed with skill:

$$W(y, \theta) = (1 - T'(y))^{-\theta} y \tag{19}$$

The social objective is weighted utilitarian. As detailed earlier, it encapsulates two interesting subcases: maximin and utilitarianism. Under these specifications, Appendix A.3 shows that the optimal tax formula (18a) simplifies to:

$$\frac{T'(y)}{1 - T'(y)} = \frac{1 - \int_{\theta \in \Theta} g^+(W(y, \theta)|\theta) d\hat{\mu}(y, \theta)}{\int_{\theta \in \Theta} \theta p(W(y, \theta)|\theta) d\hat{\mu}(y, \theta)} \tag{20}$$

where  $g^+(w|\theta)$  denotes the average of welfare weights for all skill levels above  $w$  in group  $\theta$ :

$$g^+(w|\theta) \stackrel{\text{def}}{=} \frac{\int_{x=w}^{\infty} g(x, \theta) f(x|\theta) dx}{1 - F(w|\theta)}, \quad (21)$$

where  $p(w|\theta)$  is the local Pareto parameter of the conditional skill distribution within group  $\theta$ :

$$p(w|\theta) = \frac{w f(w|\theta)}{1 - F(w|\theta)} \quad (22)$$

and where  $\hat{\mu}(y, \theta)$  is the distribution of groups  $\theta$  among individuals earning an income larger than  $y$ . According to (19) the mass of individuals who earn an income larger than  $y$  in group  $\theta$  is  $1 - H(y|\theta) = 1 - F(W(y, \theta)|\theta)$ . Using (19), the distribution of groups  $\theta$  among individuals who earn an income larger than  $y$  at the optimum is thus described by the cumulative distribution function:

$$\hat{\mu}(y, \theta) \stackrel{\text{def}}{=} \frac{\int_{\theta' \in \Theta, \theta' \leq \theta} \left(1 - F\left((1 - T'(y))^{-\theta'} y \mid \theta'\right)\right) d\mu(\theta')}{\int_{\theta' \in \Theta} \left(1 - F\left((1 - T'(y))^{-\theta'} y \mid \theta'\right)\right) d\mu(\theta')} \quad (23)$$

In Equation (20), the optimal marginal tax rate at any income level depends negatively on a weighted average, across groups, of social welfare weights (evaluated above the endogenous skill level  $W(y, \theta)$ ). It also depends negatively on a weighted average, across groups, of the products of the direct compensated elasticity and the local Pareto parameter within each group. It is clear that the marginal tax rate  $T'(y)$  affects these sufficient statistics through the skill level  $W(y, \theta)$  at which they need to be evaluated. This makes sufficient statistics distinct in the actual and optimal situations. This first mechanism of endogeneity is already well established in the standard model where individuals differ along a single dimension (their skills), i.e. where there is a single  $\theta$ -group. In this model, the optimal marginal tax rate is given by:

$$\frac{T'(y)}{1 - T'(y)} = \frac{1 - g^+(W(y))}{\theta p(W(y))} \quad (24)$$

where the sufficient statistics  $g^+(W(y))$  and  $p(W(y))$  are endogenous to  $T'(y)$ . Indeed behind the same income level in the actual and optimal economies, we have distinct skill levels because  $T'(y)$  is distinct in both economies (see e.g. [Chetty \(2009\)](#)).<sup>13</sup> In tax formula (20), a second mechanism of endogeneity appears, because the distribution of groups among individuals who earn an income larger than  $y$ ,  $\hat{\mu}(y, \theta)$ , is endogenous to  $T'(y)$  (see Equation (23)). Behind a given income level in the multidimensional model, we have individuals with distinct skills  $w$  and distinct other characteristics summarized by  $\theta$ . As we move from the actual to the optimal economy, the composition of population behind the same income level is different because  $T'(y)$  is distinct in both economies. Note that as  $T''$  does not show up in the right-hand sides of Equations (20)-(23), composition effects also prevail in the case of linear taxation.

<sup>13</sup>In the model with one dimension of heterogeneity, a taxpayer who earns income  $y$  is endowed with skill  $W(y) = (1 - T'(y))^{-1}y$  which is distinct, for a given income  $y$ , if marginal tax rates are different in the actual and optimal economies.

For our numerical examples, we will consider a simple setting consisting of one economy with only low- and high-elasticity individuals  $0 < \theta_L < \theta_H$ , those two groups having equal size,  $\mu(\theta_L) = \mu(\theta_H) = 1/2$ , and of another economy with a single group (hence without composition effects). To understand how composition effects work in this simple setting, consider the situation where at a given income level, the optimal marginal tax rate is higher (lower) than in the actual economy. According to (19), the  $\theta_L$ -workers behind income  $y$  are not the same in the optimal and actual economies. Their skill level  $W(y, \theta)$  is larger (lower) in the optimal economy. This also applies to  $\theta_H$ -workers. However, the skill level of  $\theta_H$ -workers needs to increase (decrease) more than the one of  $\theta_L$ -workers because the labor supply elasticity  $\theta_H$  is larger than  $\theta_L$ . Consequently, among workers who earn more than  $y$ , the share of low-elasticity workers  $\hat{\mu}(y, \theta_L)$  tends to increase (decrease) while the share of high-elasticity workers  $\hat{\mu}(y, \theta_H)$  tends to decrease (increase). From optimal tax formula (20), the sufficient statistics evaluated among low-elasticity workers, i.e.  $\theta_L$ ,  $g^+(W(y, \theta)|\theta_L)$  and  $p(W(y, \theta)|\theta_L)$ , have therefore a higher (lower) impact in the determination of marginal tax rates than the corresponding sufficient statistics evaluated among high-elasticity workers.

We assume maximin social preferences to neglect the impact the composition of population may have on social welfare weights since this has already been discussed in Cuff (2000), Boadway et al. (2002), Brett and Weymark (2003), Choné and Laroque (2010) and Lockwood and Weinzierl (2015). Calibrating labor supply elasticities  $\theta_L$  and  $\theta_H$ , is a challenge, because it is very difficult to identify heterogeneity in labor supply elasticity among individuals earning similar incomes. However, it is well established that male and female workers respond differently to changes in tax rate (see e.g. Bargain and Peichl (2017)). We take  $\theta_L = 0.1$  and  $\theta_H = 0.8$  based on estimates of male and female labor supply elasticities in Blau and Kahn (2007). However, we consider that these distinct levels of elasticity can be explained by other sources of heterogeneity, e.g. health, marital status, ethnicity and we simply distinguish between a high- and a low-elasticity group.<sup>14</sup> For the sake of comparison, we take  $\theta = 0.45 = (0.1 + 0.8)/2$  for the economy with a single group. The distribution of the skill levels is based on the CPS data (2007) and on an approximation of the actual tax schedule by a linear one with a constant marginal tax rate of 40% as in Saez (2001, 2002). The excellent Pareto fit of the top tail of the distribution has been well known for over a century since the pioneering work of Pareto (1896) and verified in many countries and many periods, as summarized in Atkinson et al. (2011). From estimates in Diamond and Saez (2011, p.170), and Piketty and Saez (2013, p. 424), we know that the top of the income distribution in the US is extremely well approximated by a Pareto distribution and that the implicit Pareto parameter is 1.5. Even though the value of the Pareto parameter for the top of the income distribution is well known for the entire population, there is no empirical evidence on its values in subgroups of population. Therefore, we believe the best we can do is to consider several scenarios regarding the Pareto parameters within

---

<sup>14</sup>In reality, we most probably have many groups with distinct elasticities and the latter do also differ with income as well.

groups always making sure that the Pareto parameter for top incomes, all groups together, is 1.5.

We now investigate two scenarios. In the first one, the group-specific income distributions in the actual economy are identical. We approximate each of them by a log-normal distribution, extended by a Pareto distribution with parameter 1.5. In the second scenario, the group-specific income distributions in the actual economy are Pareto with distinct coefficients,  $p_L = 1.3$  for the low-elasticity group and  $p_H = 2$  for the high-elasticity group.<sup>15</sup>

The optimal marginal tax rates obtained in the first scenario are shown on Figure 1. Marginal tax rates (in solid, blue line) are below the ones obtained without composition effects (in red dotted lines) for incomes below \$15,662. Conversely, marginal tax rates with composition effects are above the ones without composition effects, for incomes above \$15,662. Top marginal tax rate increases from  $1/(1 + 1.5 \times 0.45) = 59.7\%$  (from (24)) in the economy without composition effects to 65.5% from (20) in the economy with composition effects.

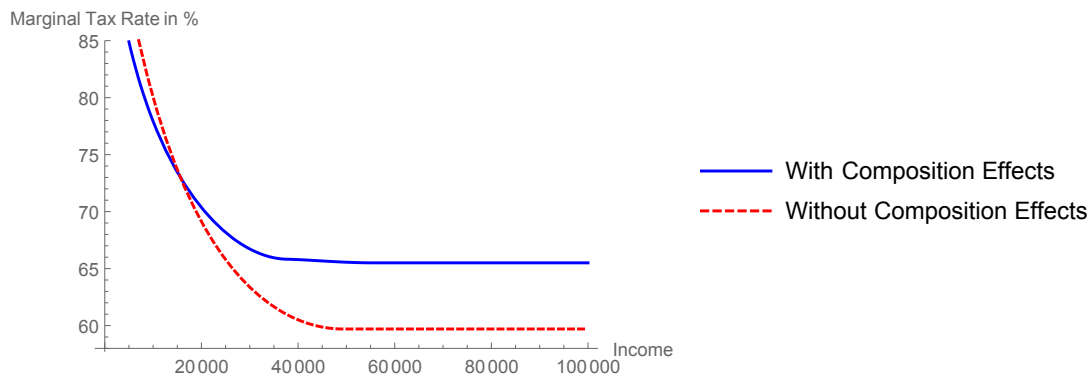


Figure 1: Optimal marginal tax rates with (solid blue line) and without composition effects (dashed red lines) with Log-normal income distribution extended by a Pareto distribution.

Figure 2 provides the actual income density which is identical for low and high-elasticity workers (dashed-dotted pink curve) as well as the density of low-elasticity workers (solid blue curve) and the one of high-elasticity workers (dotted red curve) at the optimum. The density of low-elasticity workers is almost not modified which is not surprising since their elasticity  $\theta_L$  is quite low. In contrast, we see on Figure 2 that more high-elasticity workers end up at the bottom of the income distribution in the optimal economy. This drastic distortion of the density of high-elasticity workers goes hand in hand with the changes of marginal tax rates on Figure 1. At any income level where the marginal tax rate increases (decreases), the relatively large behavioral responses, due to their large elasticity  $\theta_H$ , of this group of workers reduce (increase) their density. To say it another way, larger (lower) marginal tax rates are optimal at income levels where the density of high-elasticity workers is relatively low.

Regarding the second scenario, the left and right panels of Figure 3 confirm that the chosen Pareto parameters of 1.3 and 2 within groups make the Pareto parameter, across all groups,

<sup>15</sup>The group of workers with a low elasticity (high elasticity) has then a fatter (thinner) upper tail. This is for instance consistent with the estimates obtained by [Atkinson et al. \(2016\)](#), for men and women in the UK.

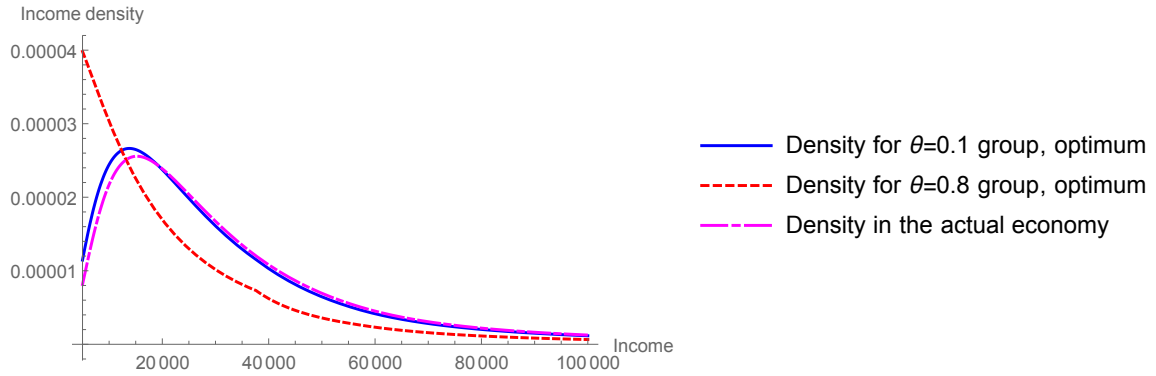


Figure 2: Densities in the actual economy (log-normal extended by a Pareto, identical for each group) and optimal income densities

at each income level above \$350,500 (the threshold for the top 1% income earners), consistent with the empirical estimates of [Piketty and Saez \(2013, p. 424\)](#). Interestingly, while the terms  $\theta p(\cdot|\theta)$  are independent of income, the weights  $\hat{\mu}(\cdot|\theta)$  now depend on income. As can be seen from Equation (20), composition effects are therefore the only mechanisms that make sufficient statistics endogenous. When income goes to infinity, the group of low-elasticity workers whose density has the lowest Pareto coefficient (i.e. whose density has the fattest upper tail) is the only one left. Therefore, only the elasticity and Pareto coefficient of these workers matter to compute the asymptotic optimal marginal tax rate.

Interestingly, the impact of composition effects extends beyond the top income. For large income levels, the share of workers with a low elasticity and a low Pareto parameter increases with income. More precisely, the share of these workers is 0.5 among individuals who earn an income higher than the threshold of \$350,500 and it increases to 1 when the threshold goes to infinity. Since the share of low-elasticity people increases with income, we expect optimal marginal tax rates to increase with income. Top optimal marginal tax rates are displayed, as a function of income levels, on the left panel of [Figure 4](#) and, in terms of percentiles, on the right panel of the same figure. Note that the top 1, 0.5 and 0.1 percentiles correspond to substantially distinct income levels: \$350,050, \$537,100 and \$1,528,500, respectively. The solid (blue) curves highlight that indeed top optimal marginal tax rate increases quite substantially with income. It rises from 61.0% for the top 1% of the population to 68.7% for the top 0.5% and it reaches 82.3% for the top 0.1%. At the very top it rockets to  $1/(1 + 1.5 \times 0.1) \simeq 86.9\%$ . These results again emphasize the crucial role played by composition effects. Optimal marginal tax rates (solid blue curve in [Figure 4](#)) undergo a drastic change, up to 20.3 percentage points, at the top of the income distribution compared to the top tax rate when heterogeneity is one-dimensional (dashed red curve in [Figure 4](#)). Along the latter curve, the marginal tax rate is constant at 59.7% (since  $\theta = 0.45$  and the value of the Pareto parameter is 1.5). To give an even better grasp of the exact role played by composition effects in this scenario, we can also consider two distinct group-specific Pareto income distributions and make them exogenous, identical

to what they are in the actual economy.<sup>16</sup> In this case, two groups of workers with distinct elasticities co-exist but no composition effect modifies their distribution functions. Since we allow for distinct Pareto income distributions, marginal tax rates are however not constant with income as highlighted by the dashed-dotted pink curves in Figure 4. Compared to this situation, optimal marginal tax rates undergo a large increase, up to 16.5 percentage points, at the top of the income distribution when including composition effects (solid blue curve in Figure 4).

Our results also highlight that the standard analysis of top tax rates can lead to misleading policy prescriptions. Indeed, the optimal top tax rate is usually derived for all incomes above a threshold as a function of the mean taxable income elasticity above this threshold and of the Pareto coefficient, see e.g. Saez et al. (2012) and Piketty and Saez (2013). The implicit assumption is that the elasticity of taxable income and the local Pareto coefficient are roughly constant beyond a threshold so that the formula for the top tax rate (which corresponds to Equation (24)) is robust to change in the chosen threshold, as can be seen from the dashed red curves in Figure 4. However, our results point out that the choice of threshold is far from innocuous. If the Pareto coefficients are different between high- and low-elasticity groups, slightly modifying the chosen threshold can substantially modify the optimal top tax rate. This can be seen in Figure 4, when comparing optimal top tax rates that include composition effects (solid blue curve) either with the constant top tax rate obtained under one-dimensional heterogeneity (dashed red curve) or with the top tax rates under the assumption of exogenous income distributions that exclude composition effects (dashed-dotted pink curve).

To put it in a nutshell, in the light of our numerical investigations, we call for more empirical evidence regarding the distribution of income in various subgroups of the population and the distribution of the relevant sufficient statistics. A first step towards the latter would be to recover estimates of behavioral responses in subgroups of population.

## VI Optimal structural tax formula

In this section, we first derive a so-called optimal *structural* tax formula, expressed in terms of the skill density and of direct (rather than total) behavioral elasticities. We highlight that this tax formula is more convenient when working on real data than the sufficient statistics tax formula of (18a). Beyond this first advantage, we show that this formulation is necessary to sign optimal marginal tax rates.

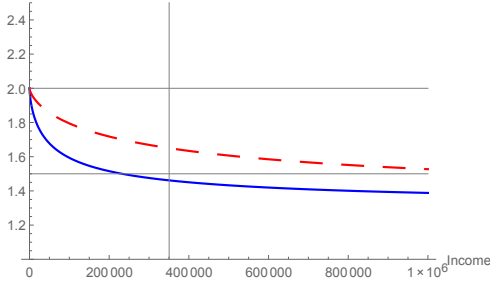
Following Mirrlees (1971), we assume individual preferences are additively separable. Like in Mirrlees (1971), this assumption allows one to write the optimal structural tax formula in

---

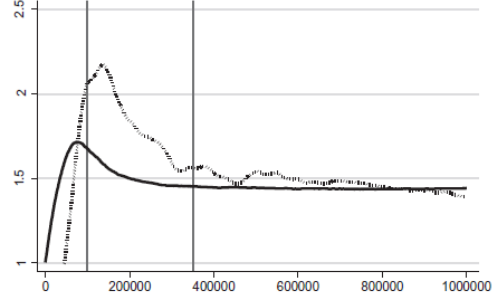
<sup>16</sup>The endogeneity of  $\hat{\mu}(y, \theta_L)$  and of  $\hat{\mu}(y, \theta_H)$  is then neglected. To do so, we fix  $\hat{\mu}(y, \theta_L)$  and  $\hat{\mu}(y, \theta_H)$  at their values in the actual economy. The optimal tax formula (20) becomes:

$$\frac{T'(y)}{1 - T'(y)} = \frac{1}{1 + \theta_L p_L \hat{\mu}_0(y, \theta_L) + \theta_H p_H \hat{\mu}_0(y, \theta_H)}$$

where the subscripts zero denote values taken in the actual economy.



(a) Our calibrated economy



(b) From Piketty and Saez (2013), p. 424

Figure 3:  $y \mapsto \frac{y_m(y)}{y_m(y)-y}$  (solid lines) where  $y_m(y)$  is the average income above  $y$  and  $y \mapsto \frac{y \hat{h}_0(y)}{1-\hat{H}_0(y)}$  (dotted lines) where the subscript zero stands for actual economy. The gray vertical lines depict the 99<sup>th</sup> percentile threshold.

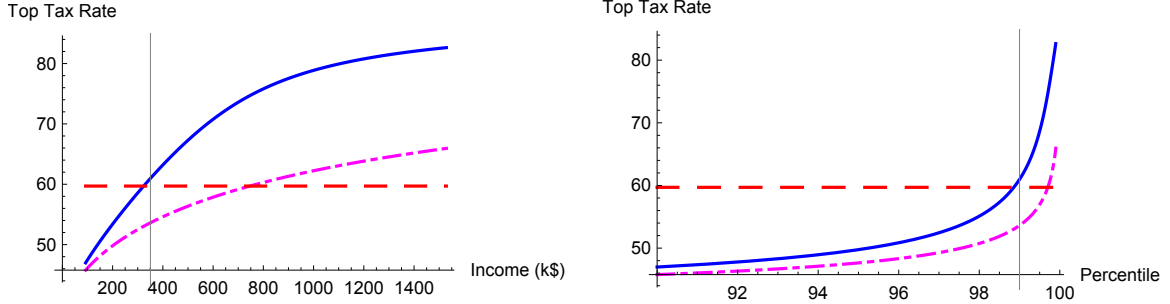


Figure 4: Optimal marginal tax rates, with Pareto income distributions when  $p_L = 1.3$  and  $p_H = 2$  and  $\theta_L = 0.1$  and  $\theta_H = 0.8$ , with (solid blue curve) and without (dashed-dotted pink curve) composition effects. The dashed-dotted pink curve corresponds to the case with exogenous income distributions that exclude composition effects. The dashed red line is the optimal tax rate with a single group, i.e. without composition effect ( $\theta = 0.45$  and  $p = 1.5$ ). The gray vertical line correspond to the 99<sup>th</sup> percentile.

a simple way (Proposition 2). Moreover, as in the model with one-dimensional heterogeneity, this assumption is necessary to sign optimal marginal tax rates (Proposition 3).

**Assumption 3.** *The utility function is additively separable and takes the form:*

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

Assuming  $u', v_y > 0 > v_w$  is necessary to retrieve  $\mathcal{U}_c, \mathcal{U}_w > 0 > \mathcal{U}_y$  (our initial assumption). The convexity of the indifference curve associated to  $\mathcal{U}$  under additively separable preferences is ensured by assuming that  $v_{y,y} > 0 \geq u''$ . Under Assumption 3, the marginal rate of substitution between (pre-tax) income and consumption is given by  $\mathcal{M}(c, y; w, \theta) = v_y(y; w) / u'(c)$ . Assumption 1 (the within-group single-crossing condition) requires  $v_{y,w} < 0$  with additively separable preferences. Therefore, Assumption 3 is more restrictive than Assumption 1 since the latter does not require additive separability.



## VI.1 Optimal tax schedule

In Appendix A.4, we simply rearrange the optimal tax formula of Proposition 1 (which was derived using the tax perturbation approach) to obtain a characterization of the optimal marginal tax rates in terms of the direct elasticities and of the structural parameters of the model. By definition, the latter (conditional skill densities, utility function and social welfare function) are policy invariant. In Appendix A.7, we alternatively propose a mechanism design method to obtain this structural tax schedule.<sup>17</sup> We call this method an *allocation perturbation approach* since it is based on a perturbation of the allocation. More precisely, 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 the single-crossing condition within each group (Assumption 1), we can fully characterize an incentive-compatible allocation. Applying variational calculus to a small perturbation of this allocation, we derive the necessary conditions of the optimal allocation. Rearranging terms gives the optimal structural tax formula.

**Proposition 2.** *Under Assumptions 2 and 3, the optimal structural tax formula verifies:*

$$\begin{aligned} & \frac{T'(y)}{1 - T'(y)} \int_{\theta \in \Theta} \frac{\varepsilon^*(W(y, \theta), \theta)}{\alpha^*(W(y, \theta), \theta)} W(y, \theta) f(W(y, \theta) | \theta) d\mu(\theta) \\ &= u'(y - T(y)) \iint_{\theta \in \Theta, w \geq W(y, \theta)} \left( \frac{1}{u'(C(w, \theta))} - \frac{\Phi_U(U(w, \theta); w, \theta)}{\lambda} \right) f(w | \theta) dw d\mu(\theta) \end{aligned} \quad (25a)$$

for all income  $y$  and:

$$\iint_{\theta \in \Theta, w \in \mathbb{R}_+} \left( \frac{\Phi_U(U(w, \theta); w, \theta)}{\lambda} - \frac{1}{u'(C(w, \theta))} \right) f(w | \theta) dw d\mu(\theta) = 0. \quad (25b)$$

This optimal tax formula is much more convenient to implement numerically than the tax formula in terms of sufficient statistics provided by Proposition 1 (in Equation (18a)) and hence circumvents a significant limit of the latter. First, Equation (25a) is expressed in terms of the conditional skill density  $f(\cdot | \theta)$  which is policy-invariant, while Equation (18a) is expressed in terms of the conditional income density  $h(\cdot | \theta)$  which depends on the tax schedule. Second, Equation (25a) makes use of direct elasticities, which are the ones obtained in the empirical literature (see Footnote 9). Equation (25a) is a second-order differential equation that depends simply on the tax function and on its first-order derivative (i.e. the marginal tax rate). In contrast, the left-hand side of (18a) displays total elasticities (hence the second-order derivative of

<sup>17</sup>The allocation perturbation method requires that preferences verify single-crossing within group and additive separability, which are imposed by Assumption 3, but also that  $\lim_{w \rightarrow 0} v_y(y; w) = +\infty$  and  $\lim_{w \rightarrow +\infty} v_y(y; w) = 0$ . These Inada conditions are not needed when one uses the tax perturbation approach.

the tax function).<sup>18</sup> Therefore, to implement this equation numerically, one has to deal not only with the first order derivative of tax function (i.e. with the marginal tax rate) but also with the second-order derivative  $T''$  that appears in Equations (16b) and (16c). Equation (18a) is thus a third-order differential equation. As such, it is much less convenient to implement numerically than the second-order differential equation (25a). Third, additive separability (Assumption 3) allows one to rewrite the income effects  $\hat{\eta}(z) T'(z)$  in Equation (18a) in terms of the derivatives of the individual utility of consumption,  $u'(\cdot)$ , and of the derivative of welfare with respect to utility,  $\Phi_u$ . Again, these two functions are policy-invariant which is extremely convenient for the numerical implementation.<sup>19</sup>

## VI.2 Signing optimal marginal tax rates

Thanks to Proposition 2, one can now provide a sufficient condition for positive marginal tax rates.

**Proposition 3.** *Under Assumptions 2 and 3, and 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). The proof is in Appendix A.5.

With multidimensional heterogeneity, the literature has highlighted that negative marginal tax rates can be optimal. In Cuff (2000), Boadway et al. (2002), 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. Proposition 3 shows that optimal marginal tax rates are positive as soon as all individuals who earn the same income  $y$  are characterized by the same marginal utility of consumption  $\mathcal{U}_c$ , which is ensured by the additive separability assumption, and by the same marginal social welfare  $\Phi_u$ , 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 emphasized in Cuff (2000), Boadway et al. (2002), Choné and Laroque (2010) and Lockwood and Weinzierl (2015) does not apply. In other words,

<sup>18</sup>More precisely, what appears in Equation (25a) is the ratio of the direct compensated elasticity  $\varepsilon$  to the direct skill density  $\alpha$ . According to (14), the skill elasticity  $\alpha$  shows up because the left-hand side is expressed in terms of the (policy-invariant) conditional skill density  $f(\cdot|\theta)$  instead of the endogenous conditional income density  $h(\cdot|\theta)$  found in Equation (18a). Note that, according to Equations (15a) and (15c), the total compensated  $\varepsilon$  and skill  $\alpha$  elasticities differ from their direct counterparts  $\varepsilon^*$  and  $\alpha^*$  by the same corrective term. Hence, the ratio of  $\varepsilon/\alpha$  is equal to the ratio of the direct compensated elasticity  $\varepsilon^*$  to the direct skill elasticity  $\alpha^*$ .

<sup>19</sup>Of course, (25a) and (25b) are not a closed-form solution. It depends on the arguments of functions  $\varepsilon^*/\alpha^*$ ,  $u'(\cdot)$  and  $\Phi_u$ , and are thus functions of the allocation  $(w, \theta) \mapsto (C(w, \theta), Y(w, \theta))$ . Therefore one needs to iterate when implementing tax formula (25a). For given values of  $\varepsilon^*/\alpha^*$ ,  $u'(\cdot)$  and  $\Phi_u$ , Equations (25a) and (25b) provide an approximation for the tax schedule. This, in turn, provides an approximation of the optimal allocation by solving maximization program (1). This allows one to update the evaluations of  $\varepsilon^*/\alpha^*$ ,  $u'(\cdot)$  and  $\Phi_u$ . However, because the second-order derivatives of the tax function do not show up in (25a) and (25b), this iteration process is much easier than the one that hinges on the sufficient statistics formula (Equations (18a) and (18b)).

introducing heterogeneous behavioral responses for individuals who earn the same level of income is not sufficient to obtain negative optimal marginal tax rates if the population earning a given income level remains homogeneous in terms of social welfare weights.

## VII Tax perturbation method versus mechanism design approach

In this section, we show the equivalence between the tax perturbation approach (which relies on Assumption 2) and the mechanism design approach, assuming individual characteristics are multidimensional. This equivalence is established under the within-group single-crossing condition (Assumption 1).

The mechanism design approach relies on the taxation principle (Hammond, 1979, Guesnerie, 1995) according to which it is equivalent for the government to select a nonlinear tax schedule taking into account the labor supply decisions as the ones described in (1), or to directly select an allocation  $(w, \theta) \mapsto (C(w, \theta), Y(w, \theta))$  that verifies the incentive constraints,

$$\forall w, \theta, w', \theta' \in (\mathbb{R}_+^* \times \Theta)^2 \quad \mathcal{U}(C(w, \theta), Y(w, \theta); w, \theta) \geq \mathcal{U}(C(w', \theta'), Y(w', \theta'); w, \theta). \quad (26)$$

According to (26), individuals of type  $(w, \theta)$  are better off with the bundle  $(C(w, \theta), Y(w, \theta))$  designed for them than with bundles  $(C(w', \theta'), Y(w', \theta'))$  designed for individuals of any other type  $(w', \theta')$ .

In the mechanism design approach, it is usual to assume that the government selects among incentive-compatible allocations that are continuously differentiable. Then, incentive constraints (26) imply the first-order incentive constraints, i.e.

$$\forall (w, \theta) \in \mathbb{R}_+^* \times \Theta \quad \dot{U}(w, \theta) = \mathcal{U}_w(C(w, \theta), Y(w, \theta); w, \theta) \quad (27)$$

These first-order incentive constraints are necessary but not sufficient to verify the incentive constraints (26). A sufficient condition is that the allocation also verifies a monotonicity constraint according to which in each group,  $Y(\cdot, \theta)$  is nondecreasing in skill. We adopt a slightly more restrictive assumption.

**Assumption 4.** *The allocation  $(w, \theta) \mapsto (C(w, \theta), Y(w, \theta))$  is smooth if and only if it is continuously differentiable, it verifies (26) and  $w \mapsto Y(w, \theta)$  admits a positive derivative for any group  $\theta \in \Theta$  and at any skill level  $w \in \mathbb{R}_+^*$ .*

We get the following connection between Assumption 2 required for the tax perturbation approach and Assumption 4 in the first-order mechanism design approach. The proof is in Appendix A.6.

**Proposition 4.** *Under Assumption 1,*

- i) *Any tax schedule  $y \mapsto T(\cdot)$  verifying Assumption 2 (i.e. the conditions for the tax perturbation) induces a smooth allocation that verifies Assumption 4.*

ii) Any smooth allocation verifying Assumption 4 can be decentralized by a tax schedule that verifies Assumption 2.

Intuitively, under Assumption 1 (which states the single-crossing condition within group), elements of Assumptions 2 and 4 are equivalent. The fact that, for each group  $\theta$ , the second-order condition of the individual program (1) holds strictly (Part ii of Assumption 2) is equivalent to  $Y(\cdot, \theta)$  admitting a strictly positive derivative in skill as required in Assumption 4. Moreover, the uniqueness of the global maximum from the individual maximization program (1) (Part iii of Assumption 2) is equivalent to  $Y(\cdot, \theta)$  being continuous in skill as stated in Assumption 4.

Thanks to Proposition 4, first-order mechanism design and tax perturbation approaches are analog. The (first-order) mechanism design approach consists in choosing, among the allocations that verify Assumption 4, the one that maximizes the social objective (6) subject to the budget constraint (5). It involves computing the first-order effect, on the Lagrangian (7), of a small perturbation of the optimal allocation. Since the allocation after perturbation has to verify Assumption 4, it is decentralized by a tax schedule that has to verify Assumption 2. Therefore, as stated in Proposition 4, the effects of a perturbation of the allocation that preserves Assumption 4 are equivalent to the responses of the allocation to a perturbation of the tax function that preserves Assumption 2. In other words, the mechanism design approach focuses on the effects of an allocation perturbation whereas the tax perturbation approach focuses on the effects of the tax reform that decentralizes this perturbation of the allocation. For this reason, the mechanism design approach and the tax perturbation approach are the two faces of the same coin.

In the literature where the unobserved heterogeneity is one dimensional, the mechanism design approach can be developed under less restrictive assumptions than Assumption 4. In particular, Lollivier and Rochet (1983), Guesnerie and Laffont (1984), Ebert (1992), Boadway et al. (2000) study the case where individuals endowed with different skill levels choose the same consumption-income bundle. To decentralize such an allocation where *bunching* occurs, one would need a kink in the tax function. This is excluded with a tax perturbation because of Assumption 2 but can be studied with the mechanism design approach. Note that the alternative “pathology” where individuals may be indifferent between two levels of income appears much more plausible under twice continuously differentiable tax schedule. Surprisingly, this problem has attracted much less attention than bunching in the literature based on the mechanism design approach, a noticeable exception being Hellwig (2010).

## VIII Concluding Comments

In this paper, we provide formulas to calculate sufficient statistics in the presence of multi-dimensional individual heterogeneity. Multidimensional heterogeneity generates a new channel through which sufficient statistics differ in the optimal and actual economies. We call this

additional channel “composition effects”. Using optimal tax policy as the field of choice to illustrate our point, we emphasize the key role of composition effects in the calculation of sufficient statistics. We show that neglecting composition effects entails a potentially large bias on optimal tax schedules. We quantify this bias through a few examples. Our results stress the need for empirical studies on labor supply elasticities and distribution parameters for different demographic groups e.g., according to age, ethnicity and gender. This paper is clearly a call for more empirical evidence within sub-groups of population at distinct levels of income to clarify the importance of composition effects in the design of optimal tax schedules. This is part of our research agenda.

As a second main contribution, we prove the equivalence, when unobserved heterogeneity is multidimensional, between the tax perturbation method and (first-order) mechanism design. Both methods have been used separately to solve optimal income tax problems. While the former is widely used in various fields in economics, the latter is more specific to the optimal taxation literature (although it may find applications in industrial organization as well). Having ascertained their equivalence (and the conditions of this equivalence) is therefore an important result.

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 what  $y$ ,  $w$ ,  $\theta$  represent so that the interpretation of the results is straightforward. In the vein of this paper, we choose optimal tax problems but our framework even extends beyond optimal taxation, e.g., to nonlinear pricing problems where consumers differ along several unobserved dimensions.

### Optimal joint taxation of labor and non-labor income

Consider individuals 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  $\theta$  as the ability to earn non-labor income  $z$  and  $w$  as the skill. Individuals of type  $(w, \theta)$  solve:

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

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  $\mathcal{U}(c, y; w, \theta) \stackrel{\text{def}}{=} \max_z \mathcal{U}(c, y - z, z; w, \theta)$ . 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  $\mathcal{U}(\cdot, \cdot; w, \theta)$  verifies Assumption 1.

### 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  $\theta$  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  $\theta$  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 1 holds.

## A Appendix

### A.1 Proof of Lemma 1

Let  $\mathcal{L}^R$  be the Lagrangian that results from applying a reform with a direction  $R$  and magnitude  $m$  on the Lagrangian (7):

$$\mathcal{L}^R(m) \stackrel{\text{def}}{=} \iint_{\theta \in \Theta, w \in \mathbb{R}_+} \left[ T(Y^R(m; w, \theta)) - m R(Y^R(m; w, \theta)) + \frac{\Phi(U^R(m; w, \theta); w, \theta)}{\lambda} \right] f(w|\theta) dw d\mu(\theta)$$

Computing the partial (Gateaux) differential of the Lagrangian with respect to  $m$  at  $m = 0$  yields:

$$\begin{aligned} \frac{\partial \mathcal{L}^R}{\partial m} &= \iint_{\theta \in \Theta, w \in \mathbb{R}_+} \left\{ \frac{T'(Y(w, \theta))}{1 - T'(Y(w, \theta))} Y(w, \theta) \varepsilon(w, \theta) R'(Y(w, \theta)) \right. \\ &\quad \left. + [T'(Y(w, \theta)) \eta(w, \theta) - 1 + g(w, \theta)] R(Y(w, \theta)) \right\} f(w|\theta) dw d\mu(\theta) \\ &= \iint_{\theta \in \Theta, y \in \mathbb{R}_+} \left\{ \frac{T'(y)}{1 - T'(y)} y \varepsilon(W(y, \theta), \theta) R'(y) \right. \\ &\quad \left. + [T'(y) \eta(W(y, \theta), \theta) - 1 + g(W(y, \theta), \theta)] R(y) \right\} h(y|\theta) dy d\mu(\theta) \\ &= \int_{y \in \mathbb{R}_+} \left\{ \frac{T'(y)}{1 - T'(y)} y \hat{\varepsilon}(y) R'(y) + [T'(y) \hat{\eta}(y) - 1 + \hat{g}(y)] R(y) \right\} h(y) dy \end{aligned}$$

We use (8), (10) and (13c) to obtain the first equality. We use (14) for the change of variable from skill  $w$  to income  $y$  in the second equality. Note the role of the within-group single-crossing condition (Assumption 1) behind this change of variable. It implies that in each group, income is an increasing function of skill with a strictly positive derivative. Therefore, in each group, the income density is continuous without any mass point nor hole. We use (16a)-(16d) for the third equality. Integrating by parts the integral of  $\frac{T'(y)}{1 - T'(y)} y \hat{\varepsilon}(y) \hat{h}(y) R'(y)$  leads to (17).

We now show that the first-order effect on the Lagrangian (7) of a reform with magnitude  $m$  and direction  $R(\cdot)$  is positively proportional to the first-order effect on the social objective (6) of the reform denoted  $\tilde{R}(m)$ . The latter is a tax reform in the direction  $R(\cdot)$  with magnitude  $m$  where the induced net budget surplus is rebated in a lump-sum way. Let  $\ell(m)$  denote this budget surplus. Under the balanced-budget tax reform  $\tilde{R}(m)$  individuals solves:

$$U^{\tilde{R}}(m; w, \theta) \stackrel{\text{def}}{=} \max_y \mathcal{U}(y - T(y) + m R(y) + \ell(m), y; w, \theta) \quad (28)$$

Applying the envelope theorem to (28) at  $m = 0$  yields:

$$\frac{\partial U^{\tilde{R}}}{\partial m}(0; w, \theta) = (R(y) + \ell'(0)) \mathcal{U}_c(C(w, \theta), Y(w, \theta); w, \theta) \quad (29)$$

Applying the implicit function theorem on the first-order condition

$$1 - T'(y) + m R'(y) = \mathcal{M}(y - T(y) + m R(y) + \ell(m), y; w, \theta)$$

at  $y = Y^{\tilde{R}}(m; w, \theta)$  and using (12b), (13b) and (13c) leads to:

$$\frac{\partial Y^{\tilde{R}}}{\partial m}(0; w, \theta) = \frac{\partial Y^R}{\partial m}(0; w, \theta) + \eta(w, \theta) \ell'(m) \quad (30)$$

We now denote respectively  $\mathcal{B}^R(m)$ ,  $\mathcal{S}^R(m)$  and  $\mathcal{L}^R(m)$  the budget surplus, the social objective and the Lagrangian when the tax function is perturbed in the direction  $R$  as a function of the magnitude  $m$  with  $\mathcal{L}^R(m) = \mathcal{B}^R(m) + (1/\lambda)\mathcal{S}^R(m)$ . We symmetrically denote  $\mathcal{B}^{\tilde{R}}(m)$ ,  $\mathcal{S}^{\tilde{R}}(m)$  and  $\mathcal{L}^{\tilde{R}}(m)$  the budget surplus, the social objective and the Lagrangian when the tax function is perturbed by the balanced-budget tax reform in the direction  $R$  with magnitude  $m$ . We get

$$0 = \mathcal{B}^{\tilde{R}}(m) = \iint_{(w, \theta) \in \mathbb{R}_+^* \times \Theta} \left\{ T(Y^{\tilde{R}}(m; w, \theta)) - m R(Y^{\tilde{R}}(m; w, \theta)) \right\} f(w|\theta) dw d\mu(\theta) - \ell(m).$$

We then obtain:

$$\ell'(0) = \iint_{(w, \theta) \in \mathbb{R}_+^* \times \Theta} \left\{ T'(Y(w, \theta)) \frac{\partial Y^{\tilde{R}}}{\partial m}(0; w, \theta) - R(Y(w, \theta)) \right\} f(w|\theta) dw d\mu(\theta)$$

Using (30), we can then write:

$$\ell'(0) = \frac{\partial \mathcal{B}^R}{\partial m}(0) + \ell'(0) \iint_{(w, \theta) \in \mathbb{R}_+^* \times \Theta} T'(Y(w, \theta)) \eta(w, \theta) f(w|\theta) dw d\mu(\theta)$$

so that:

$$\ell'(0) = \frac{1}{1 - \iint_{(w, \theta) \in \mathbb{R}_+^* \times \Theta} T'(Y(w, \theta)) \eta(w, \theta) f(w|\theta) dw d\mu(\theta)} \frac{\partial \mathcal{B}^R}{\partial m}(0) \quad (31)$$

Finally, using (29), we get:

$$\begin{aligned} \frac{\partial \mathcal{L}^{\tilde{R}}}{\partial m}(0) &= \frac{\partial \mathcal{S}^R}{\partial m}(0) + \ell'(0) \iint_{(w, \theta) \in \mathbb{R}_+^* \times \Theta} \Phi'_u(U(w, \theta); w, \theta) \mathcal{U}_c(C(w, \theta), Y(w, \theta); w, \theta) f(w|\theta) dw d\mu(\theta) \\ &= \frac{\partial \mathcal{S}^R}{\partial m}(0) + \frac{\iint_{(w, \theta) \in \mathbb{R}_+^* \times \Theta} \Phi'_u(U(w, \theta); w, \theta) \mathcal{U}_c(C(w, \theta), Y(w, \theta); w, \theta) f(w|\theta) dw d\mu(\theta)}{1 - \iint_{(w, \theta) \in \mathbb{R}_+^* \times \Theta} T'(Y(w, \theta)) \eta(w, \theta) f(w|\theta) dw d\mu(\theta)} \frac{\partial \mathcal{B}^R}{\partial m}(0) \\ &= \lambda \frac{\partial \mathcal{L}^R}{\partial m}(0) \end{aligned} \quad (32)$$

where the latter equality holds if and only if

$$\lambda = \frac{\int_{(w,\theta) \in \mathbb{R}_+^* \times \Theta} \Phi'_u(U(w,\theta); w, \theta) \mathcal{Z}_c(C(w,\theta), Y(w,\theta); w, \theta) f(w|\theta) dw d\mu(\theta)}{1 - \int_{(w,\theta) \in \mathbb{R}_+^* \times \Theta} T'(Y(w,\theta)) \eta(w,\theta) f(w|\theta) dw d\mu(\theta)} \quad (33)$$

## A.2 Proof of Proposition 1

An optimal tax system implies that any tax reform  $R(\cdot)$  does not yield any first-order effect on the Lagrangian (7). That is (17) is nil at  $m = 0$  for any direction  $R(\cdot)$ . This implies that  $\lim_{y \rightarrow 0} \frac{T'(y)}{1-T'(y)} \hat{\varepsilon}(y) y \hat{h}(y) = \lim_{y \rightarrow \infty} \frac{T'(y)}{1-T'(y)} \hat{\varepsilon}(y) y \hat{h}(y) = 0$  and, for any income  $y$ , we have:

$$\frac{d}{dy} \left[ \frac{T'(y)}{1-T'(y)} \hat{\varepsilon}(y) y \hat{h}(y) \right] = [\hat{g}(y) - 1 + T'(y) \hat{\eta}(y)] \hat{h}(y)$$

Integrating the latter equality for all income  $z$  above  $y$  and using  $\lim_{y \rightarrow \infty} \frac{T'(y)}{1-T'(y)} \hat{\varepsilon}(y) y \hat{h}(y) = 0$  yields (18a). Making  $y$  tends to 0 in (18a) and using  $\lim_{y \rightarrow 0} \frac{T'(y)}{1-T'(y)} \hat{\varepsilon}(y) y \hat{h}(y) = 0$  leads to (18b).

## A.3 Derivation of Equation (20)

Without income effects  $\hat{\eta}(y) = \eta(w, \theta) = 0$  and using Equations (16a)-(16d), Equation (18a) can be rewritten as:

$$\begin{aligned} \frac{T'(y)}{1-T'(y)} \int_{\theta \in \Theta} \varepsilon(W(y, \theta), \theta) y h(y|\theta) d\mu(\theta) &= \iint_{\theta \in \Theta, y \leq z} (1 - g(W(z, \theta), \theta)) h(z|\theta) dz d\mu(\theta) \\ \frac{T'(y)}{1-T'(y)} \int_{\theta \in \Theta} \varepsilon(W(y, \theta), \theta) \frac{y h(y|\theta)}{1-H(y|\theta)} d\hat{\mu}(y, \theta) &= \int_{\theta \in \Theta} (1 - g^+(W(y, \theta), \theta)) d\hat{\mu}(y|\theta) \end{aligned} \quad (34)$$

where the second equality comes from (21) and  $d\hat{\mu}(y, \theta) = \frac{1-H(y|\theta)}{1-H(y)} d\mu(\theta)$ , the latter being derived from  $\hat{\mu}(y, \theta)$ .

In the optimal economy, from (19), we have  $H(y|\theta) \equiv F\left((1-T'(y))^{-\theta} y \mid \theta\right)$ . Differentiating both sides of this equality with respect to income  $y$  leads to:

$$\begin{aligned} y h(y|\theta) &= \left(1 + \frac{y T''(y) \theta}{1-T'(y)}\right) (1-T'(y))^{-\theta} y f\left((1-T'(y))^{-\theta} y \mid \theta\right) \\ y h(y|\theta) &= \left(\frac{1-T'(y) + y T''(y) \theta}{1-T'(y)}\right) W(y, \theta) f(W(y, \theta) \mid \theta) \\ \varepsilon(y, \theta) y h(y|\theta) &= \theta W(y, \theta) f(W(y, \theta) \mid \theta) \\ \varepsilon(y, \theta) \frac{y h(y|\theta)}{1-H(y|\theta)} &= \theta \frac{W(y, \theta) f(W(y, \theta) \mid \theta)}{1-F(W(y, \theta) \mid \theta)} = \theta p(W(y, \theta) \mid \theta) \end{aligned} \quad (35)$$

where the third equality uses (15a) and  $\varepsilon^*(y, \theta) = \theta$  (under quasilinear and isoelastic individual preferences) and the latter equality uses  $H(y|\theta) = F(W(y, \theta) \mid \theta)$ . Plugging (35) into (34) and using the definition of the Pareto parameter  $p(w|\theta)$  (Equation (22)) leads to (20).



#### A.4 Proof of Proposition 2: A rewriting of Equation (18a)

Let

$$X(y) \stackrel{\text{def}}{=} \int_y^\infty \{1 - \hat{g}(z) - \hat{\eta}(z) T'(z)\} \hat{h}(z) dz \quad (36)$$

denote the right-hand side of (18a). Equation (18a) can be rewritten as:

$$\frac{T'(y)}{1 - T'(y)} \hat{\varepsilon}(y) y \hat{h}(y) = X(y) \quad (37)$$

Combine (16a) and (16b) to write:

$$\hat{\varepsilon}(y) y \hat{h}(y) = \int_{\theta \in \Theta} \varepsilon(W(y, \theta), \theta) y h(y|\theta) d\mu(\theta) = \int_{\theta \in \Theta} \frac{\varepsilon(W(y, \theta), \theta)}{\alpha(W(y, \theta), \theta)} W(y, \theta) f(W(y, \theta)|\theta) d\mu(\theta)$$

where (14) is used to obtain the second equality. Plugging this equality in the left-hand side of (37) and using (15a)-(15c) leads to:

$$\frac{T'(y)}{1 - T'(y)} \int_{\theta \in \Theta} \frac{\varepsilon^*(W(y, \theta), \theta)}{\alpha^*(W(y, \theta), \theta)} W(y, \theta) f(W(y, \theta)|\theta) d\mu(\theta) = X(y) \quad (38)$$

From Assumption 3, one has  $\mathcal{M}(c, y; w, \theta) = v_y(y; w, \theta) / u'(c)$ , so we get that  $\mathcal{M}_c(c, y; w, \theta) / \mathcal{M}(c, y; w, \theta)$  simplifies to  $-u''(c) / u'(c)$ . Therefore, according to Equations (13a) and (13b), we get:

$$\eta(w, \theta) = \varepsilon(w, \theta) Y(w, \theta) \frac{u''(C(w, \theta))}{u'(C(w, \theta))}$$

Using (16b) and (16c) and the fact that individuals of different  $(w, \theta)$ -types who earn the same income  $y$  have to consume the same amount  $c$ , we get:

$$\hat{\eta}(y) = \hat{\varepsilon}(y) y \frac{u''(y - T(y))}{u'(y - T(y))} \quad (39)$$

where one uses additive separability. We now define  $J(y)$  by the equality  $X(y) \stackrel{\text{def}}{=} J(y) u'(y - T(y))$ . We get:

$$\begin{aligned} X'(y) &= J'(y) u'(y - T(y)) + X(y) \frac{u''(y - T(y))}{u'(y - T(y))} (1 - T'(y)) \\ &= J'(y) u'(y - T(y)) + T'(y) \hat{\varepsilon}(y) y \hat{h}(y) \frac{u''(y - T(y))}{u'(y - T(y))} \\ &= J'(y) u'(y - T(y)) + T'(y) \hat{\eta}(y) \hat{h}(y) \end{aligned}$$

where the second equality uses (37) (and also relies on additive separability) and the third equality uses (39). Differentiating in income both sides of (36) leads to:

$$\begin{aligned} J'(y) u'(y - T(y)) + T'(y) \hat{\eta}(y) \hat{h}(y) &= \{-1 + \hat{g}(y) + T'(y) \hat{\eta}(y)\} \hat{h}(y) \\ J'(y) u'(y - T(y)) &= \{-1 + \hat{g}(y)\} \hat{h}(y) \\ J'(y) &= \int_{\theta \in \Theta} \left\{ -\frac{1}{u'(y - T(y))} + \frac{\Phi_u \langle W(y, \theta), \theta \rangle}{\lambda} \right\} h(y|\theta) d\mu(\theta) \end{aligned}$$

where the latter equality uses (16a) and (16d). Integrating this equality for all incomes  $z$  above  $y$  gives:

$$\begin{aligned} J(y) &= \iint_{z \geq y, \theta \in \Theta} \left\{ \frac{1}{u'(z - T(z))} - \frac{\Phi_u \langle W(z, \theta), \theta \rangle}{\lambda} \right\} h(z|\theta) dz d\mu(\theta) \\ &= \iint_{w \geq W(y, \theta), \theta \in \Theta} \left\{ \frac{1}{u'(C(w, \theta))} - \frac{\Phi_u \langle w, \theta \rangle}{\lambda} \right\} f(w|\theta) dw d\mu(\theta) \end{aligned} \quad (40)$$

where the second equality uses (14) to change variables from income to skill. Plugging (40) in (38) leads to (25a). Using (36), Equation (18b) can be rewritten as  $X(0) = 0$ , thereby  $J(0) = 0$ . Using (40) leads to (25b).

### A.5 Proof of Proposition 3

Under utilitarian preferences,  $\Phi_u = 1$  and we get:

$$J(y) \stackrel{\text{def}}{=} \int_{z \geq y} \left( \frac{1}{u'(z - T(z))} - \frac{1}{\lambda} \right) \left( \int_{\theta} f(W(z, \theta) | \theta) d\mu(\theta) \right) dz$$

The derivative of  $J(z)$  has the sign of  $1/\lambda - 1/u'(z - T(z))$ , which is decreasing in  $w$  because of the concavity of  $u(\cdot)$ . Moreover,  $\lim_{y \rightarrow \infty} J(y) = 0$  and Equation (25b) imply that  $J(0) = 0$ . Therefore,  $J(\cdot)$  first increases and then decreases. It is thus positive for all (interior) skill levels. Since  $v_{yw} < 0$  from (1), optimal marginal tax rates are positive.

Under Maximin, one has  $U(x, \theta) > U(0, \theta)$  for all  $x > 0$  from (45a). Therefore, within each group, the most deserving individuals are those whose skill  $w = 0$ . The Maximin objective implies  $\Phi_U \langle x, \theta \rangle = 0$  for all  $x > 0$ . Thereby,

$$J(y) \stackrel{\text{def}}{=} \int_{z \geq y} \frac{1}{u'(z - T(z))} \left( \int_{\theta} f(W(z, \theta) | \theta) d\mu(\theta) \right) dz$$

for all  $x > 0$ , which leads to positive marginal tax rates.

### A.6 Proof of Proposition 4

#### Part i) of Proposition 4.

Let  $T(\cdot)$  be an income tax schedule satisfying Assumption 2. We already know that under Assumptions 1 and 2, one can apply the implicit function theorem to the first-order condition associated to (1). This implies that  $Y(\cdot, \theta)$ , thereby  $C(\cdot, \theta)$  is continuously differentiable in  $w$  within each group  $\theta$ . Moreover,  $Y(\cdot, \theta)$  admits a positive derivative according to (13d). Finally, from (1) we get that:

$$\forall w, \theta, y' \in \mathbb{R}_+^* \times \Theta \times \mathbb{R}_+ \quad \mathcal{U}(C(w, \theta), Y(w, \theta); w, \theta) \geq \mathcal{U}(y' - T(y'), y'; w, \theta)$$

Taking  $y' = Y(w', \theta')$  leads to  $C(w', \theta') = y' - T(y')$ , so that the latter inequality leads to (26). Therefore the allocation  $w \mapsto (C(\cdot, \theta), Y(\cdot, \theta))$  induced by  $T(\cdot)$  verifies (26), thereby Assumption 4.

#### Part ii) of Proposition 4

Let  $(w, \theta) \mapsto (C(w, \theta), Y(w, \theta))$  be a mapping defined over  $\mathbb{R}_+^* \times \Theta$  which verifies Assumption 4. Let  $\mathbb{Y}$  denote the set of incomes that are assigned to some individuals along this allocation. To define the tax schedule that decentralizes this allocation, we first show that if two types  $(w, \theta)$  and  $(w', \theta')$  of individuals earn the same income  $y = Y(w, \theta) = Y(w', \theta')$ , then they have to be assigned the same consumption  $C(w, \theta) = C(w', \theta')$ . Otherwise, if by contradiction one has:  $C(w, \theta) < C(w', \theta')$ , then one would get that individuals of type  $(w, \theta)$  would be better off with the bundle  $(C(w'), Y(w'))$  designed for individuals of type  $(w', \theta')$ , which would be in contradiction with (26). A symmetric argument applies if  $C(w, \theta) > C(w', \theta')$  by inverting the role of  $(w, \theta)$  and of  $(w', \theta')$ . We can then unambiguously define the tax schedule denoted  $T(\cdot)$  that decentralizes this allocation by:

$$\forall y \in \mathbb{Y} \quad T(y) \stackrel{\text{def}}{=} Y(w, \theta) - C(w, \theta) \quad \text{where } (w, \theta) \text{ are such that: } y = Y(w, \theta) \quad (41)$$

Given this tax schedule, Program (1) of individuals of type  $(w, \theta)$  is equivalent to:

$$\max_{(w', \theta') \in \mathbb{R}_+^* \times \Theta} \mathcal{U}(C(w', \theta'), Y(w', \theta'); w, \theta),$$

the solution of the latter is  $(w, \theta)$  since  $(w, \theta) \mapsto (C(w, \theta), Y(w, \theta))$  verifies the incentive constraints (26). Therefore, the tax schedule  $T(\cdot)$  defined by (41) decentralizes the given allocation.<sup>20</sup>

We now need to show a mathematical result. For each group  $\theta \in \Theta$ , as  $Y(\cdot, \theta)$  is continuously differentiable, it admits a reciprocal denoted  $Y^{-1}(\cdot, \theta)$  which is also continuously differentiable with a strictly positive derivative. Therefore the image of the (open) skill set  $\mathbb{R}_+^*$  by  $Y(\cdot, \theta)$  is an open set denoted  $\mathbb{Y}(\theta) \subset \mathbb{R}_+$ . Equation (41) can be rewritten on  $\mathbb{Y}(\theta)$  by:

$$T(y) = y - C\left(Y^{-1}(y, \theta), \theta\right) \quad (42)$$

Moreover, we get that  $\mathbb{Y} = \cup_{\theta \in \Theta} \mathbb{Y}(\theta)$  and is therefore an open set. Hence, for each income  $y \in \mathbb{Y}$ , there exists a group  $\theta$  such that  $T(\cdot)$  verifies (42) in the neighborhood of  $y$ .

To show that  $T(\cdot)$  verifies Part *i*) of Assumption 2, note that from (42),  $T(\cdot)$  is continuously differentiable as  $Y^{-1}(\cdot, \theta)$  and  $C(\cdot, \theta)$  are continuously differentiable. Moreover, from (2), we have:

$$T'(y) = 1 - \mathcal{M}(y - T(y), y; Y^{-1}(w, \theta), \theta)$$

As  $T(\cdot)$  and  $Y^{-1}(\cdot, \theta)$  are continuously differentiable in  $y$ , and  $\mathcal{M}(\cdot, \cdot; \cdot, \theta)$  is continuously differentiable in  $(c, y, w)$ ,  $y \mapsto \mathcal{M}(y - T(y), y; Y^{-1}(w, \theta), \theta)$  is continuously differentiable. Therefore,  $T'(\cdot)$  is continuously differentiable and  $T(\cdot)$  verifies Part *i*) of Assumption 2.

To show that  $T(\cdot)$  verifies Part *ii*) of Assumption 2,<sup>21</sup> note that the first-order condition (11) can be rewritten as  $\mathcal{Y}(Y(w, \theta); w, \theta) \equiv 0$  for all skill levels. Differentiating this equality with respect to skill leads to:  $\mathcal{Y}_y(Y(w, \theta); w, \theta) \dot{Y}(w, \theta) + \mathcal{Y}_w(Y(w, \theta); w, \theta) = 0$ . As  $\mathcal{Y}_w(Y(w, \theta); w, \theta) = -\mathcal{M}_w(C(w, \theta), Y(w, \theta); w, \theta)$  which is positive from Assumption 1 and  $\dot{Y}(w, \theta) > 0$  from Assumption 4, then one must have  $\mathcal{Y}_y(Y(w, \theta); w, \theta) < 0$ , which is Part *ii*) of Assumption 2.

To show that  $T(\cdot)$  verifies Part *iii*) of Assumption 2, we assume by contradiction that individuals of type  $(w^*, \theta)$  are indifferent between earning income  $Y(w^*, \theta)$  and earning an income level denoted  $y' \in \mathbb{Y}$ . We show that in such a case, some individuals with skill  $w$  close to  $w^*$  are better off with the bundle  $(y' - T(y'), y')$  than with the bundle  $(C(w, \theta), Y(w, \theta))$  designed for them, a contradiction. For this purpose, we denote  $\mathcal{C}(u, y; w, \theta)$  the consumption an individual of type  $(w, \theta)$  should get to enjoy utility  $u$  while earning income  $y$ . Function  $\mathcal{C}(\cdot, y; w, \theta)$  is the reciprocal of function  $\mathcal{U}(\cdot, y; w, \theta)$ . We get:  $\mathcal{C}_u = 1/\mathcal{U}_c$ ,  $\mathcal{C}_y = -\mathcal{U}_y/\mathcal{U}_c = \mathcal{M}$  and  $\mathcal{C}_w = -\mathcal{U}_w/\mathcal{U}_c$ . Let us denote:

$$\mathcal{Q}(w) \stackrel{\text{def}}{=} \mathcal{C}(U(w, \theta), y'; w, \theta) - y' + T(y')$$

To be indifferent between earning income  $Y(w, \theta)$  and income  $y'$ , individuals of type  $(w, \theta)$  have to receive after-tax income  $\mathcal{C}(U(w, \theta), y'; w, \theta)$  when they earn income  $y'$ . Therefore,  $\mathcal{Q}(w)$  is a measure in monetary units of the difference in well-being for individuals of type  $(w, \theta)$  between the bundle  $(C(w, \theta), Y(w, \theta))$  designed for them (from which they obtain utility  $U(w, \theta)$ ) and the utility they would get by earning income  $y'$  and consuming  $y' - T(y')$ . We have by assumption  $\mathcal{Q}(w^*) = 0$ . We obtain:

$$\mathcal{Q}'(w) = \frac{\mathcal{V}(U(w, \theta), Y(w, \theta), w, \theta) - \mathcal{V}(U(w, \theta), y', w, \theta)}{\mathcal{U}_c(\mathcal{C}(U(w, \theta), Y(w, \theta); w, \theta), Y(w, \theta); w, \theta)}$$

<sup>20</sup>We have here followed Hammond (1979) very closely.

<sup>21</sup>We are grateful to Kevin Spiritus for encouraging us to emphasize this result.

where  $\mathcal{V}(u, y; w, \theta) \stackrel{\text{def}}{=} \mathcal{U}_w(\mathcal{C}(u, y; w, \theta), y; w, \theta)$  describes how  $\mathcal{U}_w$  varies with income  $y$  along the indifference curve of individuals of type  $(w, \theta)$  with utility  $u$ . We get that  $\mathcal{V}_y = -\mathcal{U}_c \mathcal{M}_w$  which is strictly positive from Assumption 1. Therefore:

- If  $y' > Y(w^*, \theta)$ , then  $\mathcal{Q}'(w^*) < 0$ , which implies that for some skills  $w > w^*$  above  $w^*$  and sufficiently close to  $w^*$ ,  $\mathcal{Q}(w) < 0$ , i.e.  $U(w, \theta) < \mathcal{U}(y' - T(y'), y'; w, \theta)$ . Therefore, individuals of type  $(w, \theta)$  strictly prefers the bundle  $(y' - T(y'), y')$  rather than the bundle  $(C(w, \theta), Y(w, \theta))$  designed for them, a contradiction.
- If  $y' < Y(w^*, \theta)$ , then  $\mathcal{Q}'(w^*) > 0$ , which implies that for some skills  $w < w^*$  below  $w^*$  and sufficiently close to  $w^*$ ,  $\mathcal{Q}(w) < 0$ , i.e.  $U(w, \theta) < \mathcal{U}(y' - T(y'), y'; w, \theta)$ . Therefore, individuals of type  $(w, \theta)$  strictly prefers the bundle  $(y' - T(y'), y')$  rather than the bundle  $(C(w, \theta), Y(w, \theta))$  designed for them, a contradiction.

## A.7 Proof of Proposition 2: An allocation perturbation method

In Section VII, we show the equivalence between the tax perturbation method and the first-order approach in mechanism design. Let us remind that this approach is labeled “first-order”, as usually done in one-dimensional tax models (see e.g. Salanié (2011)), since it takes into account the first-order incentive-compatibility constraints (45a) and assumes the second-order incentive-compatibility constraints are satisfied. In this appendix, we propose a first-order mechanism design approach to derive the optimal tax formulas (of Proposition 2).<sup>22</sup> Our contribution is technical. Our new method, that we call “allocation perturbation”, allows one to solve the class of multidimensional screening problems at play in this paper, using mechanism design.

Utility is additively separable as in (3). The role of additive separability (standardly used in the adverse selection literature with multidimensional heterogeneity, e.g., Rochet (1985), Wilson (1993), Rochet and Choné (1998), Rochet and Stole (2002)) will become clear as we proceed. With additively separable utility, we can write:

$$\mathcal{E}_u(\hat{u}, y; w, \theta) = \frac{1}{u'(c)} \quad \text{and} \quad \mathcal{E}_y(\hat{u}, y; w, \theta) = \frac{v_y(y; w, \theta)}{u'(c)} \quad (43)$$

where the various derivatives are evaluated at  $c = \mathcal{C}(\hat{u}, y; w, \theta)$ .

The proof is structured as follows. In a first step, we define a function of pooling which allows us to characterize the set of incentive compatible allocations. In a second step, we retrieve the optimal tax formula thanks to our allocation perturbation method which builds up on step 1.

### Step 1. Pooling function and incentive-compatible allocations

In Section VII, Equation (26) gives the set of incentive constraints. An incentive-compatible allocation has to satisfy (26). It thus has to verify for each group  $\theta$  the following set of “within-group incentive constraints”.

$$\forall (w, \tilde{w}, \theta) \in \mathbb{R}_+^2 \times \Theta \quad \mathcal{U}(C(w, \theta), Y(w, \theta); w, \theta) \geq \mathcal{U}(C(\tilde{w}, \theta), Y(\tilde{w}, \theta); w, \theta). \quad (44)$$

Under the within-group single-crossing Assumption 1, the set of within-group incentive constraints can be transformed into a monotonicity constraint and a differential equation that we retrieve in Lemmas 2 and 3 below.

<sup>22</sup> To solve the problem of a firm that hires managers whose productivity is private information, Garrett and Pavan (2015) develop another variational approach, under a set of conditions on the primitives of their model (e.g. quadratic disutility of effort), as an alternative to the usual first-order approach.

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

The within-group single-crossing assumption (1) implies that, in the same group  $\theta$ , 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 2 which is a simple reformulation of the usual proof in the one dimensional framework.  $Y(\cdot; \theta)$  being nondecreasing, it may exhibit discontinuities over a countable set and it may also exhibit bunching where individuals in the same group but endowed with different skill levels earn the same income. Since we follow a first-order approach, we 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<sup>23</sup> 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  $\lim_{x \rightarrow \infty} a'(x) = +\infty$ .

**Assumption 5** (Smooth allocations). In each group  $\theta$ ,  $w \mapsto Y(w, \theta)$  is smoothly increasing.

According to Assumption 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  $\theta$  earn income  $y = Y(w, \theta)$ . The following lemma provides the first-order incentive constraints within group  $\theta$  reformulated as a differential equation.

**Lemma 3.** Under Assumptions 3 and 5, for each  $\theta$ , the mapping  $w \mapsto U(w, \theta)$  is differentiable with:

$$\dot{U}(w, \theta) = \mathcal{U}_w(C(w, \theta), Y(w, \theta); w, \theta) = -v_w(Y(w, \theta); w, \theta). \quad (45a)$$

Moreover, Equation (45a) is equivalent to:

$$\frac{\dot{C}(w, \theta)}{Y(w, \theta)} = \mathcal{M}(C(w, \theta), Y(w, \theta); w, \theta). \quad (45b)$$

We use a dot to denote the derivative with respect to  $w$  for a fixed  $\theta$ . We skip the proof since it simply consists in adapting the one of the one dimensional tax models, see e.g., Salanié (2011). Integrating (45a) leads to:

$$U(w, \theta) = U(0, \theta) - \int_0^w v_w(Y(x, \theta); x, \theta) dx. \quad (45c)$$

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 (26). This is the *pooling issue* that we now address.

### Pooling Types across $\theta$ -Groups at each Income Level

Choose a reference group  $\theta_0 \in \Theta$ , a skill level  $w$  and another group  $\theta$ . Individuals of type  $(w, \theta_0)$  earn income  $Y(w, \theta_0)$ . According to the smoothness Assumption 5, 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  $\theta$  endowed with that skill level  $W(w, \theta)$  must get the same income level  $Y(w, \theta_0)$  as individuals of type  $(w, \theta_0)$ , i.e.  $Y(W(w, \theta), \theta) = Y(w, \theta_0)$ . We call  $W(\cdot, \cdot)$  the *pooling function*. For each  $\theta \in \Theta$ , the pooling function combines two smoothly increasing functions, namely  $\omega \xrightarrow{Y(\cdot, \theta_0)} Y(\omega, \theta_0) \xrightarrow{Y^{-1}(\cdot, \theta)} W(\omega, \theta)$ . The pooling function is therefore

<sup>23</sup>A smoothly increasing (decreasing) function is also called an increasing (decreasing) diffeomorphism for which the derivative maps the positive real line onto itself.

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 (26) that individuals of type  $(W(w, \theta), \theta)$  and individuals of type  $(w, \theta_0)$  obtain the same income  $Y(w, \theta_0)$  but distinct consumption levels. Therefore, for each  $(w, \theta)$ , we must simultaneously have:

$$Y(W(w, \theta), \theta) \equiv Y(w, \theta_0) \quad \text{and} \quad C(W(w, \theta), \theta) \equiv C(w, \theta_0). \quad (46)$$

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. Moreover, to determine the pooling function, one only needs the allocation designed for the reference group. Intuitively, individuals who earn the same income level, whatever the group they belong to, face the same marginal tax rate. Hence, from the individual first-order condition (2), these workers must have the same marginal rate of substitution. Therefore, the skill levels of individuals who earn a given income level is implicitly determined by the equality of their marginal rates of substitution which is called the pooling condition.

**Lemma 4.** *Under Assumptions 1 and 5, 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  $(w, \theta_0)$ , where  $W(w, \theta)$  verifies the following **pooling condition**:*

$$\mathcal{M}(C(w, \theta_0), Y(w, \theta_0); w, \theta_0) = \mathcal{M}(C(w, \theta_0), Y(w, \theta_0); W(w, \theta), \theta). \quad (47)$$

**Proof** According to Assumption 1,  $\mathcal{M}(C(w, \theta_0), Y(w, \theta_0); w, \theta_0) = \mathcal{M}(C(w, \theta_0), Y(w, \theta_0); \omega, \theta)$  admits exactly one solution in  $\omega$ . Differentiating in  $w$  both sides of equalities in (46) and rearranging 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 3, Equation (45b) holds, which implies (47).  $\square$

### Incentive-compatible allocations

Thanks to the pooling condition, the following lemma, provides a sufficient condition for the allocation to be incentive-compatible.

**Lemma 5.** *Let  $w \mapsto (C(w, \theta_0), Y(w, \theta_0))$  be a within-group allocation that verifies Assumption 5 and the within-group incentive-compatible Equation (45b). For each  $w \in \mathbb{R}_+$  and each group  $\theta \in \Theta$ , let  $\underline{W}(w, \theta)$  be the unique skill level  $\omega$  that solves the pooling condition  $\mathcal{M}(C(w, \theta_0), Y(w, \theta_0); w, \theta_0) = \mathcal{M}(C(w, \theta_0), Y(w, \theta_0); \omega, \theta)$ . There exists a unique incentive-compatible allocation  $(w, \theta) \mapsto (\underline{C}(w, \theta), \underline{Y}(w, \theta))$  the restriction of which to group  $\theta_0$  is  $w \mapsto (C(w, \theta_0), Y(w, \theta_0))$  and it verifies Assumption 5 if and only if, for each  $\theta$ ,  $w \mapsto \underline{W}(w, \theta)$  is smoothly increasing.*

**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 5 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 (26).

**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  $\theta$ ,  $\underline{Y}(\cdot, \theta)$  verifies Assumption 5 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 5, there exists a unique skill level  $w$  such that  $y = Y(w, \theta_0)$ . Then according to Lemma 4, among individuals of group  $\theta$ , only those of skill  $\underline{W}(w, \theta)$  must be assigned to the income level  $y = Y(w, \theta_0)$  to verify incentive-compatibility.<sup>24</sup> Therefore,  $\underline{Y}^{-1}(\cdot, \theta)$  must be defined by:

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

<sup>24</sup>Hence function  $\underline{W}(\cdot, \theta)$  coincides with the pooling function  $W(\cdot, \theta)$ .

$\underline{Y}^{-1}(\cdot, \theta)$  is then smoothly increasing as a combination of two smoothly increasing functions. Moreover, since for each type  $(\omega, \theta)$ , there exists a single skill level  $w$  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 (47) holds. Differentiating in  $w$  both sides of these two equations and rearranging terms, we obtain

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

As  $w \mapsto (C(w, \theta_0), Y(w, \theta_0))$  is assumed to verify the within-group incentive constraints in Equation (45b), we know that the left-hand side of the above equation is equal to  $\mathcal{M}(C(w, \theta_0), Y(w, \theta_0); w, \theta_0)$ . Using the definition of  $\underline{W}(\cdot, \theta)$ , we have that  $w \mapsto (\underline{C}(w, \theta), \underline{Y}(w, \theta))$  also verifies Equation (45b). From Lemma 3, it thus verifies the within-group incentive constraints (44). We now check whether the inequality (26) 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 (26) are therefore equivalent to:

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

The latter inequality is verified as  $w \mapsto (\underline{C}(w, \theta), \underline{Y}(w, \theta))$  satisfies Equation (45b).  $\square$

Lemma 5 guarantees that if  $Y(w, \theta_0)$  is smoothly increasing in  $w$  and if, for each  $\theta$ , the pooling function denoted  $\underline{W}(w, \theta)$  is also smoothly increasing in  $w$ , then the allocation is incentive-compatible. Assumption 5 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  $\theta_0$  and assume that the triggered allocations for the other groups verify Assumption 5. Using Equation (3), the pooling condition (47) can be rewritten as:

$$\frac{v_y(Y(w, \theta_0); w, \theta_0)}{u'(C(w, \theta_0))} = \frac{v_y(Y(w, \theta_0); \underline{W}(w, \theta), \theta)}{u'(C(w, \theta_0))}$$

which can be simplified as:

$$v_y(Y(w, \theta_0); w, \theta_0) = v_y(Y(w, \theta_0); \underline{W}(w, \theta), \theta). \quad (48)$$

Therefore, the pooling function  $\underline{W}(\cdot, \theta)$  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 new mechanism design method (see Step 2). 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.

## Step 2. Allocation perturbation

Consider Assumption 5 is verified. Thanks to Lemma 5, the government then simply chooses, among the set of smooth allocations, the best one that verifies the first-order incentive constraint (45a) for each group, and the pooling condition (47). When the solution to this problem verifies Assumption 5 and when the implied pooling function is, for each group  $\theta$ , 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 multi-dimensional environment, the pooling condition (47) 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 (47) that we use to deduce how the allocation in the other groups are perturbed. Thanks to this condition, we can thus compute the partial (Gâteaux) derivatives of the Lagrangian (7) in the direction of these perturbations. Equating these Gâteaux derivatives to zero leads to an optimal structural tax formula which gives the optimal marginal tax rates as a function of the primitives of the model. To save on notations, we from now on use the more compact notation  $\langle w, \theta \rangle$  when the various functions are evaluated for types  $(w, \theta)$  at income  $Y(w, \theta)$ , utility  $U(w, \theta)$  and consumption  $C(w, \theta)$ .

We now derive (25b). Consider a set of allocation perturbations indexed by  $\Delta \in \mathbb{R}$  and denoted  $(\tilde{C}(w, \theta; \Delta), \tilde{Y}(w, \theta; \Delta), \tilde{U}(w, \theta; \Delta)) \stackrel{\text{def}}{=} \mathcal{U}(\tilde{C}(w, \theta; \Delta), \tilde{Y}(w, \theta; \Delta); w, \theta)$ , 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  $\Delta$ . Hence, we get for each  $\Delta$  that  $\tilde{U}(w, \theta; \Delta) \stackrel{\text{def}}{=} U(w, \theta) + \Delta$ ,  $\tilde{Y}(w, \theta; \Delta) \stackrel{\text{def}}{=} Y(w, \theta)$  and  $\tilde{C}(w, \theta; \Delta) \stackrel{\text{def}}{=} \mathcal{C}(\tilde{U}(w, \theta; \Delta), \tilde{Y}(w, \theta; \Delta); w, \theta)$ . These perturbations preserve incentive-compatibility (26). According to (7), the perturbed Lagrangian can be written:

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

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  $\mathcal{L}(\cdot)$  with respect to  $\Delta$ , at  $\Delta = 0$ , to zero, we obtain an equation that characterizes the optimal tax system. Using the first equality in (43), this Gâteaux derivative of the Lagrangian is:

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

Equating this derivative to zero leads to (25b).

To derive (25a) 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  $(\hat{C}(w, \theta; t, \delta), \hat{Y}(w, \theta; t, \delta), \hat{U}(w, \theta; t, \delta)) \stackrel{\text{def}}{=} \mathcal{U}(\hat{C}(w, \theta; t, \delta), \hat{Y}(w, \theta; t, \delta); w, \theta)$  where  $t$  stands for the size of the perturbation, and  $\delta$  is the length of the skill interval where, in the reference group, the perturbation of incomes takes place. Following Lemma 5, we define the allocation perturbations from their restriction to the reference group  $\theta_0$  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:

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

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 is nil 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 5. 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 5,  $Y(\cdot, \theta_0)$  admits a positive derivative everywhere, so  $\hat{Y}(\cdot, \theta_0)$  is bounded away from 0 for all  $x \in [w - \delta, w]$ . Therefore, provided that  $t$  is small enough, which we assume in



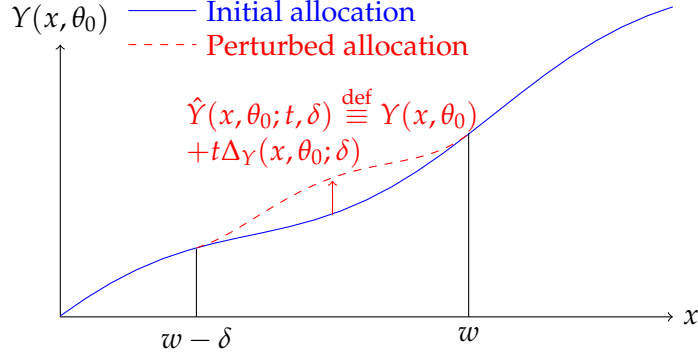


Figure 5: The perturbation of incomes in the reference group  $\theta_0$

the rest of the proof,  $\hat{Y}(\cdot, \theta_0; t, \delta)$  has also a positive derivative everywhere and therefore verifies Assumption 5.

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 (45c), the perturbed utility function in the reference group is:

$$\hat{U}(x, \theta_0; t, \delta) \stackrel{\text{def}}{=} U(0, \theta_0) - \int_0^x v_w(\hat{Y}(\omega, \theta_0; t, \delta); \omega, \theta_0) d\omega. \quad (49a)$$

From the pooling condition (47), 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  $\theta$  are not modified outside the skill interval  $(W(w - \delta, \theta), W(w, \theta))$ , and we must have (See Figure 6):

$$\hat{Y}(x, \theta; t, \delta) = Y(x, \theta) \quad \text{if} \quad x \in [0, W(w - \delta, \theta)] \cup [W(w, \theta), +\infty). \quad (49b)$$

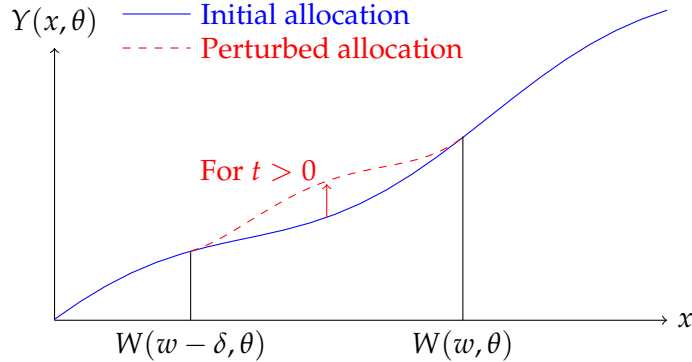


Figure 6: The perturbation of incomes in the other groups

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  $\theta$  that for all  $x \leq W(w - \delta, \theta)$ :

$$\hat{C}(x, \theta; t, \delta) = C(x, \theta) \quad \text{and} \quad \hat{U}(x, \theta; t, \delta) = U(x, \theta). \quad (49c)$$

For all skills  $x > W(w - \delta, \theta)$ , the change in utility obtained using the first-order incentive constraint (45c) is:

$$\hat{U}(x, \theta; t, \delta) - U(x, \theta) = - \int_0^x [v_w(\hat{Y}(\omega, \theta; t, \delta); \omega, \theta) - v_w(Y(\omega, \theta); \omega, \theta)] d\omega. \quad (49d)$$

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 (49b), we get:

$$\hat{U}(x, \theta; t, \delta) - U(x, \theta) = \int_{W(w-\delta, \theta)}^x [v_w(Y(\omega, \theta); \omega, \theta) - v_w(\hat{Y}(\omega, \theta; t, \delta); \omega, \theta)] d\omega. \quad (49e)$$

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

$$\hat{U}(x, \theta; t, \delta) - U(x, \theta) = \int_{W(w-\delta, \theta)}^{W(w, \theta)} [v_w(Y(\omega, \theta); \omega, \theta) - v_w(\hat{Y}(\omega, \theta; t, \delta); \omega, \theta)] d\omega. \quad (49f)$$

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

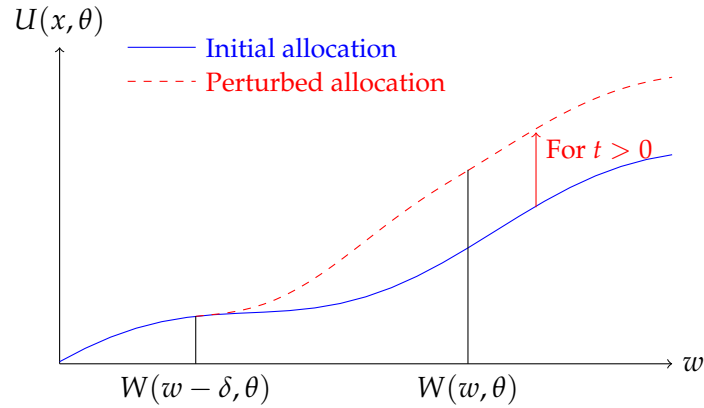


Figure 7: The perturbation of utilities

form amount above  $W(w, \theta)$ , as illustrated in Figure 7. 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 (49f), this utility therefore changes uniformly by:

$$u(\hat{C}(x, \theta; t, \delta)) - u(C(x, \theta)) = \int_{W(w-\delta, \theta)}^{W(w, \theta)} [v_w(Y(\omega, \theta); \omega, \theta) - v_w(\hat{Y}(\omega, \theta; t, \delta); \omega, \theta)] d\omega \quad (49g)$$

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 (46):

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

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:  $u(\hat{C}(x, \theta_0; t, \delta)) - u(C(x, \theta_0)) = u(\hat{C}(W(x, \theta), \theta; t, \delta)) - u(C(W(x, \theta), \theta))$ , and so, using (45c) and (49g), we obtain:

$$\begin{aligned} & \int_{w-\delta}^w [v_w(\hat{Y}(\omega, \theta_0; t, \delta); \omega, \theta) - v_w(Y(\omega, \theta_0); \omega, \theta)] d\omega \quad (49h) \\ &= \int_{W(w-\delta, \theta)}^{W(w, \theta)} [v_w(\hat{Y}(\omega, \theta; t, \delta); \omega, \theta) - v_w(Y(\omega, \theta); \omega, \theta)] d\omega. \end{aligned}$$

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:

$$\begin{aligned} \mathcal{L}(t, \delta) \stackrel{\text{def}}{=} & \iint_{\theta \in \Theta, w \in \mathbb{R}_+} \left[ \hat{Y}(w, \theta; t, \delta) - \mathcal{C}(\hat{U}(w, \theta; t, \delta), \hat{Y}(w, \theta; t, \delta); w, \theta) \right. \\ & \left. + \frac{\Phi(\hat{U}(w, \theta; t, \delta); w, \theta)}{\lambda} \right] f(w|\theta) dw d\mu(\theta). \end{aligned} \quad (50)$$

If the allocation is optimal, the derivative of this Lagrangian with respect to  $t$  must be nil at  $t = 0$ . We below show that the limit of  $\partial \mathcal{L} / \partial t$  when  $\delta$  goes to zero leads to:

$$\begin{aligned} & \int_{\theta \in \Theta} \frac{1 - \frac{v_y(Y(W(w, \theta), \theta); W(w, \theta), \theta))}{u'(C(W(w, \theta), \theta))}}{v_{yw}(Y(W(w, \theta), \theta); W(w, \theta), \theta))} f(W(w, \theta)|\theta) d\mu(\theta) \\ = & \iint_{\theta \in \Theta, x \geq W(w, \theta)} \left( \frac{\Phi_U < x, \theta >}{\lambda} - \frac{1}{u' < x, \theta >}} \right) f(x|\theta) dx d\mu(\theta). \end{aligned} \quad (51)$$

Using (3), (2),  $Y(W(w, \theta), \theta) \equiv Y(w, \theta_0)$  and  $C(W(w, \theta), \theta) \equiv C(w, \theta_0)$ , we can rewrite (51) as:

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

Using again (3), (2) and (46) leads to (25a).

### Proof : Derivation of Equation (51)

To derive (51), we must compute the various Gâteaux derivatives at  $t = 0$ . For  $A = C, Y, U$  and a given  $\delta$ , the Gâteaux derivative of  $A$  in the direction  $\Delta_Y(\cdot, \cdot; \delta)$  at  $t = 0$  is denoted  $\hat{A}(x, \theta; \delta)$ .<sup>25</sup> We derive  $\hat{Y}(x, \theta_0; \delta) = \Delta_Y(x; \delta)$ , and from (49b) we obtain:

$$\hat{Y}(x, \theta; \delta) = 0 \quad \text{if} \quad x \in [0, W(w - \delta, \theta)] \cup [W(w, \theta), +\infty). \quad (52a)$$

Equations (49c) 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 (49e) implies:

$$\hat{U}(x, \theta; \delta) = - \int_{W(w - \delta, \theta)}^x v_{yw}(Y(\omega, \theta_0); \omega, \theta_0) \hat{Y}(\omega, \theta_0; \delta) d\omega. \quad (52b)$$

For skill above  $W(w, \theta)$ , according to (49f), we have:

$$\hat{U}(x, \theta; \delta) = - \int_{W(w - \delta, \theta)}^{W(w, \theta)} v_{yw}(Y(\omega, \theta_0); \omega, \theta_0) \hat{Y}(\omega, \theta_0; \delta) d\omega. \quad (52c)$$

Moreover, Equation (49h) implies that the Gâteaux derivatives of income must verify:

$$\int_{w - \delta}^w v_{yw}(Y(\omega, \theta_0); \omega, \theta) \hat{Y}(\omega, \theta_0; \delta) d\omega = \int_{W(w - \delta, \theta)}^{W(w, \theta)} v_{yw}(Y(\omega, \theta); \omega, \theta) \hat{Y}(\omega, \theta; \delta) d\omega. \quad (52d)$$

<sup>25</sup>Let us remind its definition:

$$\hat{A}(x, \theta; \delta) \stackrel{\text{def}}{=} \lim_{t \rightarrow 0} \frac{\hat{A}(x, \theta; t, \delta) - A(w, \theta)}{t}$$

Using Equations (43), (52a) and (52c), the Gâteaux derivative of the Lagrangian (50) is:

$$\begin{aligned} \frac{\partial \mathcal{L}}{\partial t}(0; \delta) &= \int_{\theta \in \Theta} \left\{ \int_{W(w-\delta, \theta)}^{W(w, \theta)} \left( 1 - \frac{v_y(Y(x, \theta); x, \theta)}{u'(C(x, \theta))} \right) \hat{Y}(x, \theta; \delta) f(x|\theta) dx \right. \\ &+ \int_{W(w-\delta, \theta)}^{W(w, \theta)} \left( \frac{\Phi_U < w, \theta >}{\lambda} - \frac{1}{u' < x, \theta >} \right) \hat{U}(x, \theta; \delta) f(x|\theta) dx \\ &- \left( \int_{W(w-\delta, \theta)}^{W(w, \theta)} v_{yw}(Y(x, \theta); x, \theta) \hat{Y}(x, \theta; \delta) dx \right) \\ &\times \left. \left( \int_{W(w, \theta)}^{\infty} \left( \frac{\Phi_U < w, \theta >}{\lambda} - \frac{1}{u' < x, \theta >} \right) f(x|\theta) dx \right) \right\} d\mu(\theta). \end{aligned} \quad (53)$$

Dividing the first-order condition  $\frac{\partial \mathcal{L}}{\partial t}(0; \delta) = 0$  by  $\int_{w-\delta}^w v_{yw}(Y(x, \theta_0); x, \theta_0) \hat{Y}(x, \theta_0; \delta) dx$  implies, using (52b) and (52d), that

$$\begin{aligned} \int_{\theta \in \Theta} \frac{\int_{W(w-\delta, \theta)}^{W(w, \theta)} \left( 1 - \frac{v_y(Y(x, \theta); x, \theta)}{u'(C(x, \theta))} \right) \hat{Y}(x, \theta; \delta) f(x|\theta) dx}{\int_{W(w-\delta, \theta)}^{W(w, \theta)} v_{yw}(Y(x, \theta); x, \theta) \hat{Y}(x, \theta; \delta) dx} d\mu(\theta) &= \\ \int_{\theta \in \Theta} \left\{ \int_{W(w-\delta, \theta)}^{W(w, \theta)} \left( \frac{\Phi_U < w, \theta >}{\lambda} - \frac{1}{u' < x, \theta >} \right) \frac{\int_{W(w-\delta, \theta)}^x v_{yw}(Y(x, \theta); x, \theta) \hat{Y}(x, \theta; \delta) dx}{\int_{W(w-\delta, \theta)}^{W(w, \theta)} v_{yw}(Y(x, \theta); x, \theta) \hat{Y}(x, \theta; \delta) dx} f(x|\theta) dx + \right. \\ \left. \int_{W(w, \theta)}^{\infty} \left( \frac{\Phi_U < w, \theta >}{\lambda} - \frac{1}{u' < x, \theta >} \right) f(x|\theta) dx \right\} d\mu(\theta). \end{aligned} \quad (54)$$

We finally take the limit of the latter equality when  $\delta$  tends to 0. Let us consider the first term in the right-hand side of (54). Since

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

we get that:

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

As the right hand-side of the latter inequality tends to 0 when  $\delta$  tends to 0, the limit of (54) when  $t$  tends to zero leads to:

$$\begin{aligned} \lim_{\delta \rightarrow 0} \int_{\theta \in \Theta} \frac{\int_{W(w-\delta, \theta)}^{W(w, \theta)} \left( 1 - \frac{v_y(Y(x, \theta); x, \theta)}{u'(C(x, \theta))} \right) \hat{Y}(x, \theta; \delta) f(x|\theta) dx}{\int_{W(w-\delta, \theta)}^{W(w, \theta)} v_{yw}(Y(x, \theta); x, \theta) \hat{Y}(x, \theta; \delta) dx} d\mu(\theta) \\ = \iint_{\theta \in \Theta, x \geq W(w, \theta)} \left( \frac{\Phi_U < w, \theta >}{\lambda} - \frac{1}{u' < x, \theta >} \right) f(x|\theta) dx d\mu(\theta). \end{aligned} \quad (55)$$

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  $\delta$  tends to 0. As  $\Theta$  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:

$$\begin{aligned} \left( \frac{1 - v_y \langle W(w, \theta), \theta \rangle}{u'(C(W(w, \theta), \theta))} f(W(w, \theta) | \theta) - e \right) \hat{Y}(x, \theta; \delta) &\leq \left( \frac{1 - v_y \langle W(x, \theta), \theta \rangle}{u'(C(W(x, \theta), \theta))} f(x | \theta) \right) \hat{Y}(x, \theta; \delta) \\ &\leq \left( \frac{1 - v_y \langle W(w, \theta), \theta \rangle}{u'(C(W(w, \theta), \theta))} f(W(w, \theta) | \theta) + e \right) \hat{Y}(x, \theta; \delta) \end{aligned}$$

and

$$(v_{yw} \langle W(w, \theta), \theta \rangle - e) \hat{Y}(x, \theta; \delta) \leq v_{yw} \langle W(x, \theta), \theta \rangle \hat{Y}(x, \theta; \delta) \leq (v_{yw} \langle W(w, \theta), \theta \rangle + e) \hat{Y}(x, \theta; \delta) < 0$$

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

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

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

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

Hence, left-hand side of (55) is equal to the left-hand side of (51).  $\square$

## References

- Arkolakis, Arnaud Costinot, and Andres Rodriguez-Clare**, "New Trade Models, Same Old Gains?," *American Economic Review*, 2012, 102 (1), 94–130.
- Atkinson, A. B., A. Casarico, and S. Voitchovsky**, "Top Incomes and the Gender Divide," Melbourne Institute Working Paper Series Working Paper 27/16 2016.
- Atkinson, Anthony, Thomas Piketty, and Emmanuel Saez**, "Top Incomes in the Long Run of History," *Journal of Economic Literature*, 2011, 49 (1), 3–71.
- Auten, Gerald and Robert Carroll**, "The Effect of Income Taxes On Household Income," *The Review of Economics and Statistics*, 1999, 81 (4), 681–693.

- Bargain, Olivier and Andreas Peichl**, “Steady-State Labor Supply Elasticities: An International Comparison,” *IZA Journal of Labor Economics*, 2017, *Forthcoming*.
- Blau, F. and L. Kahn**, “Changes in the labor supply behavior of married women: 1980-2000,” *IZA Journal of Labor Economics*, 2007, 25, 393–438.
- 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, Katherine Cuff, and Maurice Marchand**, “Optimal Income Taxation with Quasi-linear Preferences Revisited,” *Journal of Public Economic Theory*, 2000, 2 (4), 435–60.
- , **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.
- Brett, Craig and John A. Weymark**, “Financing education using optimal redistributive taxation,” *Journal of Public Economics*, 2003, 87 (11), 2549–2569.
- Chetty, Raj**, “Sufficient Statistics for Welfare Analysis: A Bridge Between Structural and Reduced-Form Methods,” *Annual Review of Economics*, 2009, 1 (1), 451–488.
- 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,” *The Canadian Journal of Economics*, 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.
- **and Emmanuel Saez**, “The Case for a Progressive Tax: From Basic Research to Policy Recommendations,” *Journal of Economic Perspectives*, 2011, 25 (4), 165–90.
- **and James Mirrlees**, “Optimal Taxation and Public Production II: Tax Rules,” *American Economic Review*, 1971, 61 (3), 261–278.
- Ebert, Udo**, “A Reexamination of the Optimal Nonlinear Income Tax,” *Journal of Public Economics*, 1992, 49 (1), 47–73.
- Feldstein, Martin**, “Behavioral Responses to Tax Rates: Evidence from TRA86,” *American Economic Review*, 1995, 85, 170–174.
- Fleurbaey, Marc and François Maniquet**, *Theory of Fairness and Social Welfare*, Cambridge: Cambridge University Press, 2011.
- **and —**, “Optimal Taxation Theory and Principles of Fairness,” *Journal of Economic Literature*, 2017, *Forthcoming*.
- Garrett, Daniel and Alessandro Pavan**, “Dynamic Managerial Compensation: A Variational Approach,” *Journal of Economic Theory*, 2015, 159(B), 775–818.
- Gerritsen, Aart**, “Optimal nonlinear taxation: the dual approach,” *mimeo*, 2016.
- Golosov, Mikhail, Tsyvinski, and Nicolas Werquin**, “A Variational Approach to the Analysis of Tax Systems,” *Mimeo* (older version available as NBER working paper 20780) 2014.
- Gomes, Renato, Jean-Marie Lozachmeur, and Alessandro Pavan**, “Differential Taxation and Occupational Choice,” *The Review of Economic Studies*, 2017, *Forthcoming*.
- Gruber, Jon and Emmanuel Saez**, “The Elasticity of Taxable Income: Evidence and Implications,” *Journal of Public Economics*, 2002, 84, 1–32.
- Guesnerie, R.**, *A Contribution to the Pure Theory of Taxation*, Cambridge: Cambridge University Press, 1995.
- **and Jean-Jacques Laffont**, “A complete solution to a class of principal-agent problems with an application to the control of a self-managed firm,” *Journal of Public Economics*, 1984, 25,

329–369.

- Hammond, Peter J.**, “Straightforward Individual Incentive Compatibility in Large Economies,” *The Review of Economic Studies*, 1979, 46 (2), 263–282.
- Hellwig, Martin**, “Incentive Problems With Unidimensional Hidden Characteristics: A Unified Approach,” *Econometrica*, 2010, 78 (4), 1201–1237.
- Hendren, Nathaniel**, “The Inequality Deflator: Interpersonal Comparisons without a Social Welfare Function,” NBER Working Papers 20351, 2014.
- Hornstein, Andreas, Per Krusell, and Giovanni L. Violante**, “Frictional Wage Dispersion in Search Models: A Quantitative Assessment,” *American Economic Review*, 2011, 101 (7), 2873–2898.
- Jacquet, Laurence, Etienne Lehmann, 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 J. and Mazhar Waseem**, “Using Notches to Uncover Optimization Frictions and Structural Elasticities: Theory and Evidence from Pakistan,” *The Quarterly Journal of Economics*, 2013, 128 (2), 669–723.
- Kleven, Henrik Jacobsen, Claus Thustrup Kreiner, and Emmanuel Saez**, “The Optimal Income Taxation of Couples,” *Econometrica*, 2009, 77 (2), 537–560.
- Kroft, Kory, Kavan Kucko, Etienne Lehmann, and Johannes Schmieder**, “Optimal Income Taxation with Unemployment and Wage Responses: A Sufficient Statistics Approach,” IZA Discussion Papers 9719, 2016.
- Lehmann, Etienne, Laurent Simula, and Alain Trannoy**, “Tax Me If You Can! Optimal Non-linear Income Tax between Competing Governments,” *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.
- Lollivier, S. and J.-C. Rochet**, “Bunching and Second-Order Conditions: A Note on Optimal Tax Theory,” *Journal of Economic Theory*, 1983, 31, 392–400.
- Mirrlees, J. A.**, “An Exploration in the Theory of Optimum Income Taxation,” *The Review of Economic Studies*, 1971, 38, 175–208.
- Pareto, Vilfredo**, “La courbe de la répartition de la richesse,” *Ecrits sur la courbe de la répartition de la richesse*, 1896, pp. 1–15.
- Piketty, Thomas**, “La redistribution fiscale face au chômage,” *Revue Française d’Economie*, 1997, 12, 157–201.
- **and Emmanuel Saez**, “Optimal Labor Income Taxation,” in Martin Feldstein Alan J. Auerbach, Raj Chetty and Emmanuel Saez, eds., *Handbook of public economics*, Vol. 5, Elsevier, 2013, chapter 7, pp. 391 – 474.
- Ramsey, Frank**, “A Contribution to the Theory of Taxation,” *Economic Journal*, 1927, 37, 47–61.
- Robert, B Wilson**, *Nonlinear pricing*, Oxford: Oxford University Press, 1993.
- Roberts, Kevin**, “A Reconsideration of the Optimal Income Tax,” in P. Hammond and G. Myles, eds., *Incentives and Organization: Essays in Honour of Sir James Mirrlees*, 2000.
- Rochet, Jean-Charles**, “The taxation principle and multi-time Hamilton-Jacobi equations,” *Journal of Mathematical Economics*, 1985, 14 (2), 113 – 128.
- **and A. Stole Lars**, “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,” *The 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,” NBER Working Papers 22646, 2016.
- Saez, Emmanuel**, “Using Elasticities to Derive Optimal Income Tax Rates,” *The Review of Economic Studies*, 2001, 68, 205–229.
- , “Optimal Income Transfer Programs: Intensive versus Extensive Labor Supply Responses,” *Quarterly Journal of Economics*, 2002, 117, 1039–1073.
- , “Do Taxpayers Bunch at Kink Points?,” *American Economic Journal: Economic Policy*, 2010, 2 (1), 180–212.
- **and Stefanie Stantcheva**, “Generalized Social Marginal Welfare Weights for Optimal Tax Theory,” *American Economic Review*, 2016, 106 (1), 24–45.
- , **Joel Slemrod, and Seth H. Giertz**, “The Elasticity of Taxable Income with Respect to Marginal Tax Rates: A Critical Review,” *Journal of Economic Literature*, 2012, 50 (1), 3–50.
- Salanié, Bernard**, *The Economics of Taxation*, Cambridge, MA: MIT Press, 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*, 2014, 6 (2), 126–63.
- **and Iván Werning**, “Mirrlees meets Diamond-Mirrlees: Simplifying Nonlinear Income Taxation,” Mimeo (older version available as NBER working paper 22076) 2016.
- **and Ivan Werning**, “The Taxation of Superstars,” *Quarterly Journal of Economics*, 2017, 132, 211–270.
- Spiritus, Kevin**, “Characterizations for Optimal Multidimensional Taxation with Multidimensional Heterogeneity of Agents,” mimeo 2017.
- Werning, Iván**, “Pareto Efficient Income Taxation,” MIT Working Paper 2007.