DISCUSSION PAPER SERIES

DP12276
(v. 3)

OPTIMAL INCOME TAXATION WITH COMPOSITION EFFECTS

Laurence M. Jacquet and Etienne Lehmann

PUBLIC ECONOMICS

ISSN 0265-8003

OPTIMAL INCOME TAXATION WITH COMPOSITION EFFECTS

Laurence M. Jacquet and Etienne Lehmann

Discussion Paper DP12276
First Published 06 September 2017
This Revision 13 December 2019

Centre for Economic Policy Research
33 Great Sutton Street, London EC1V 0DX, UK
Tel: +44 (0)20 7183 8801
www.cepr.org

OPTIMAL INCOME TAXATION WITH COMPOSITION EFFECTS

Abstract

Providing estimable sufficient statistics to give policy prescriptions has become a widespread approach over the recent years. A well-known limitation of this approach is the endogeneity of sufficient statistics to the policy. In this paper, using optimal tax policy as our field of application, we highlight a new source of endogeneity. It arises since, under multidimensional heterogeneity, optimal tax formulas are expressed as a function of weighted means of sufficient statistics computed at the individual level and the weights are endogenous to tax policy. We analytically show that ignoring these composition effects leads to underestimate the optimal linear tax and, under a restrictive set of assumptions, the optimal nonlinear tax as well. To relax these assumptions, we use an improved tax perturbation approach to study composition effects in the latter case. Our numerical simulations using U.S. data suggest the optimal tax rate may be underestimated by 6 percentage points for high incomes levels. As a secondary result, we show the equivalence between our improved tax perturbation method and the first order mechanism design method, two methods which have hitherto been used separately to derive optimal tax schedules.

JEL Classification: N/A

Keywords: optimal taxation, Composition Effects, sufficient statistics, multidimensional screening problems, tax perturbation

Laurence M. Jacquet - laurence.jacquet@u-cergy.fr
THEMA - University of Cergy-Pontoise and Oslo Fiscal Studies, Economics Department, University of Oslo

Etienne Lehmann - etienne.lehmann@gmail.com
CRED (TEPP) Université Panthéon-Assas Paris 2, IZA, CESifo and CEPR

Acknowledgements
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 the editor, Nicola Pavoni, three anonymous referees, Pierre Boyer, Craig Brett, Bas Jacobs, Guy Laroque, St´ephane Robin, Emmanuel Saez, Florian Scheuer, Kevin Spiritus, Stefanie Stantcheva, Alain Trannoy, Nicolas Werquin as well as participants at various seminars, conferences and workshops. All remaining errors are our own.

Optimal income taxation with composition effects *

Laurence JACQUET† Etienne LEHMANN‡

December 6, 2019

Abstract

Providing estimable sufficient statistics to give policy prescriptions has become a widespread approach over the recent years. A well-known limitation of this approach is the endogeneity of sufficient statistics to the policy. In this paper, using optimal tax policy as our field of application, we highlight a new source of endogeneity. It arises since, under multidimensional heterogeneity, optimal tax formulas are expressed as a function of weighted means of sufficient statistics computed at the individual level and the weights are endogenous to tax policy. We analytically show that ignoring these composition effects leads to underestimate the optimal linear tax and, under a restrictive set of assumptions, the optimal nonlinear tax as well. To relax these assumptions, we use an improved tax perturbation approach to study composition effects in the latter case. Our numerical simulations using U.S. data suggest the optimal tax rate may be underestimated by 6 percentage points for high incomes levels. As a secondary result, we show the equivalence between our improved tax perturbation method and the first order mechanism design method, two methods which have hitherto been used separately to derive optimal tax schedules.

Keywords: Optimal taxation, composition effects, sufficient statistics, multidimensional screening problems, tax perturbation.

*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 the editor, Nicola Pavoni, three anonymous referees, Pierre Boyer, Craig Brett, Bas Jacobs, Guy Laroque, Stéphane Robin, Emmanuel Saez, Florian Scheuer, Kevin Spiritus, Stefanie Stantcheva, Alain Trannoy, Nicolas Werquin as well as participants at various seminars, conferences and workshops. All remaining errors are our own.

†THEMA - University of Cergy-Pontoise. Address: THEMA, Université de Cergy-Pontoise, 33 boulevard du Port, 95011 Cergy-Pontoise Cedex, France. Email: laurence.jacquet@u-cergy.fr.

‡CRED (TEPP) University Panthéon Assas Paris II. Address: 12 place du Panthéon 75 231, Paris Cedex 05, France. Email: etienne.lehmann@gmail.com. Etienne Lehmann is also research fellow at IZA, CESifo and CEPR.

I Introduction

The recent years have seen an increase in empirical analyses that provide so-called "sufficient statistics"to give policy prescriptions that are easily implementable and relatively easy to explain to the general public. 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, political economy and in international trade (e.g., Chetty (2009), Hornstein et al. (2011), Arkolakis et al. (2012), Bierbrauer and Boyer (2018)). In particular, optimal tax researchers rely extensively on empirically meaningful sufficient statistics to express tax formulas (e.g., Saez (2001, 2002), Saez and Stantcheva (2018), Costinot and Werning (2018) and references in Chetty (2009) and in Kleven (2018)). For this reason, in the present paper, we select optimal tax policy as the field of choice to illustrate a more general point regarding the use of sufficient statistics[1]. The endogeneity of sufficient statistics to the policy is a well-known limitation: the values of sufficient statistics in the actual economy where they are estimated differ from their values in the optimal economy where they need to be computed to determine the optimal policy. We highlight a new source of endogeneity due to changes in the composition of the population at the different income levels when one shifts from the actual to the optimal policy.

We call *composition effects* this new source of endogeneity, the rationale of which is as follows. Optimal tax policy is expressed as a function of weighted means of sufficient statistics, the latter being computed at the individual level. Therefore, between the actual and the optimal economy, not only do the sufficient statistics of each agent vary, but so do the weights used to compute the optimal policy. For instance, if the weights of taxpayers with relatively high (low) values of a certain sufficient statistic decrease (increase) when one moves from the actual to the optimal economy, then the weighted means of sufficient statistics decrease, which impacts the optimal policy. We argue that ignoring these composition effects may lead to quantitatively important bias in the computation of optimal tax schedules. To understand how, we focus on composition effects in the elasticity of earnings with respect to the marginal net-of-tax rate and explain how they impact both linear and nonlinear tax schedules. We characterize in which direction composition effects bias the optimal linear and nonlinear tax schedules. Under maximin social preferences and quasilinear individual preferences, we analytically show that composition effects exacerbate the difference between the actual and the optimal linear tax rate. To sign the bias in the nonlinear case, we have to rely on a rather restrictive set of assumptions. We therefore provide several numerical examples and show, under what we think to be very plausible empirical assumptions, that the underestimation of optimal marginal tax rate at high income levels can easily reach six percentage points. These findings have a crucial implication for the empirical literature that provides sufficient statistics. One cannot simply rely on esti-

[1] In the optimal tax literature, the key sufficient statistics are (i) behavioral responses to tax reforms, (ii) the income distribution and (iii) the social welfare weights which summarize the social preferences for redistribution (see e.g., Diamond (1998) and Saez (2001)).

mates of the means of sufficient statistics estimated in the actual economy. One needs instead to estimate the joint distribution of sufficient statistics and income.

Composition effects play a role in a variety of applications of optimal taxation. For instance, it may affect the design of optimal joint income taxation of couples if the labor supply elasticity of wives is larger than that of husbands, which is empirically plausible (Bargain and Peichl, 2016). Among couples who earn the same total income, those where the wife earns relatively more and the husband earns relatively less are characterized by a larger elasticity of taxable income. When moving from the actual to the optimal economy, the share of such households at each income level changes, which modifies the weighted mean elasticity of taxable income. Neglecting composition effects then probably biases the optimal marginal tax rates of couples. Another situation where composition effects most probably play a role occurs when taxpayers earn both an income reported to the tax authority by a third party and self-reported income e.g., casual wages such as tips. Reasonably assuming that the earnings elasticity of self-reported incomes is larger than the one of third-party reported incomes (Kleven et al., 2011), the earning elasticity of each taxpayer increases with the share of self-reported income in her total income. Among taxpayers who earn the same total income, those whose self-reported income is larger are characterized by a larger elasticity of taxable income. When moving from the actual to the optimal economy, the share of such taxpayers at each income level may vary so that the weighted mean elasticity of taxable income is likely to be impacted by composition effects. Composition effects may also take place in countries like the U.S. with a comprehensive personal income tax, i.e. when a single tax schedule applies to the sum of different types of incomes earned by a household (salaries, financial income, rents, etc). One expects that subjacent elasticities are specific to each type of income. Therefore, as in the previous examples, the earning elasticity of the total income of each taxpayer, at a given level of income, depends on the share of her incomes with larger elasticities. When moving from the actual to the optimal economy, at each income level, the weighted mean elasticity of taxable income is then likely affected by composition effects.

We also contribute to the optimal tax literature by improving the tax perturbation approach initiated by Piketty (1997) and Saez (2001). This approach consists in computing all responses to small tax reforms, to sum them up and equate them to zero in order to obtain the optimal tax schedule. The initial approach of Piketty (1997) and Saez (2001) is very intuitive but is only heuristic since it relies on tax reforms that creates kinks for which effects are neglected. This is the reason why Saez (2001) has to check the consistency of his tax formula with the one obtained using the mechanism design approach of Mirrlees (1971). He, however, verifies this only in the case where unobserved heterogeneity is one-dimensional which excludes the empirically plausible case where taxpayers differ both in skills and behavioral responses. Other attempts at extending the method of Saez (2001) to a richer class of tax reforms without kinks have been made (e.g., Hendren (2019) and Sachs et al. (2019) for the most recent). However, they need to *assume* that tax revenues are differentiable functions of tax reforms. In contrast, we *show* that income decisions are a differentiable functions of tax reforms by applying the implicit function theorem. We then verify the conjecture of Saez (2001) that, under multidimensional heterogeneity, optimal marginal tax rates depend on the averages of sufficient statistics taken among taxpayers who earn the same income.[2] Last but not least, we show that our tax perturbation approach and the (first-order) mechanism design approach are the two faces of the same coin.[3] In the latter, one considers the effects of perturbations on allocations (within the class of incentive-compatible differentiable and increasing allocations), while the tax perturbation considers the effects of a tax reform that decentralizes these perturbations.

The paper is organized as follows. We introduce the framework in Section II. We begin our analysis in Section III with the simple linear tax model to explain what composition effects are and to illustrate the empirical bias they impose. In Section IV, we characterize the optimal nonlinear tax using the tax perturbation method and we shed the light on composition effects in that case. Section V numerically investigates the sensitivity of the optimal tax function to composition effects. Section VI shows the equivalence between the mechanism design and tax perturbation approaches. Section VII concludes.

II 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). The government levies a tax $T(.)$ which depends on income y only. Consumption c is related to income y through the tax function $T(y)$ according to $c = y - T(y)$. Individuals differ along their skill level $w \in \mathbb{R}_+^*$ and along some characteristics denoted $\theta \in \Theta$. We call a *group* a subset of individuals with the same θ.[4] We assume that the set of groups Θ is measurable with a cumulative distribution function (CDF) denoted $\mu(\cdot)$. The set Θ can be finite or infinite and may be of any dimension. The distribution $\mu(.)$ of the population across the different groups may be continuous, but it may also exhibit mass points. Among individuals of the same group θ, skills are distributed according to the conditional skill density $f(\cdot|\theta)$ which is positive and differentiable over the support $\mathbb{R}_+^*$. The conditional CDF is denoted $F(w|\theta) \stackrel{\text{def}}{\equiv} \int_0^w f(x|\theta)dx$. We do not make any restriction on the correlation between w or θ. We normalize to unity the total size of the population.

[2] Our method can easily be extended to include participation decisions (Saez, 2002, Kleven et al., 2009, Jacquet et al., 2013), migration decisions (Lehmann et al., 2014, Blumkin et al., 2015) or sectoral decisions (Rothschild and Scheuer, 2013, Scheuer, 2014, Gomes et al., 2018).

[3] The tax perturbation and mechanism design methods have been used separately to solve optimal income tax problems. While the latter method is widely used in various fields in economics, the former is more specific to the optimal taxation literature.

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

II.1 Individual choice

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

$$U(w, \theta) \stackrel{\text{def}}{\equiv} \max_y \quad \mathscr{U}(y - T(y), y; w, \theta). \tag{1}$$

We call $Y(w, \theta)$ the solution to program (1), $C(w, \theta) = Y(w, \theta) - T(Y(w, \theta))$ the consumption of a worker of type (w, θ) and $U(w, \theta)$ her utility.[5] When the tax function is differentiable, the first-order condition associated to (1) implies that:

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

where:

$$\mathscr{M}(c, y; w, \theta) \stackrel{\text{def}}{\equiv} -\frac{\mathscr{U}_y(c, y; w, \theta)}{\mathscr{U}_c(c, y; w, \theta)} \tag{3}$$

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

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

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

$$\mathscr{U}(c, y; w, \theta) = c - \frac{\theta}{1+\theta}\, y^{1+\frac{1}{\theta}}\, w^{-\frac{1}{\theta}} \qquad w \in \mathbb{R}_+^*, \theta \in \Theta. \tag{4}$$

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

II.2 Government

The government's budget constraint takes the form:

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

[5] To ease the notations, we do not make explicit the dependence of $Y(\cdot, \cdot)$, $C(\cdot, \cdot)$, $U(\cdot, \cdot)$ on $T(\cdot)$.

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 transformation $\Phi(U; w, \theta)$ of individuals' utility levels U:

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

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

The government's problem consists in finding the tax schedule $T(\cdot)$ that maximizes the social welfare objective (6) subject to the budget constraint (5). Let $\lambda > 0$ denote the shadow price of public funds. The Lagrangian (expressed in monetary terms) is:

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

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

$$g\left(w, \theta\right) \stackrel{\text{def}}{\equiv} \frac{\Phi_U\left(U\left(w, \theta\right); w, \theta\right) \mathscr{U}_c\left(C(w,\theta), Y(w,\theta); w, \theta\right)}{\lambda}. \tag{8}$$

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

III Optimal linear tax and composition effects

In this section, we illustrate how composition effects bias the empirical implementation of the optimal tax rate using the very simple case of linear taxation. The tax schedule is linear with a tax rate denoted τ and a demogrant D so: $T(y) = \tau \, y - D$. Let $y^M(w, \theta; \tau, D)$ denote the Marshalian solution to the taxpayer's program $\max_y \mathscr{U}((1-\tau)y + D, y; w, \theta)$. The budget constraint (5) can be rewritten as:

$$\tau \iint_{(w,\theta) \in \mathbb{R}_+^* \times \Theta} y^M(w, \theta; \tau, D) \, f(w|\theta) dw \, d\mu(\theta) - D = E. \tag{9}$$

Assuming leisure is a normal good, one has $y_D^M(w, \theta; \tau, D) \leq 0$, which ensures that the left hand-side of the previous equation is decreasing in D. Hence, for each tax rate τ, there exists a single demogrant denoted $\tilde{D}(\tau)$ that clears the budget constraint (9). We denote: $\tilde{y}(w, \theta; \tau) \stackrel{\text{def}}{\equiv} y^M\left(w, \theta; \tau, \tilde{D}(\tau)\right)$the pretax income of taxpayers of type (w, θ) when the tax rate is τ and the demogrant clears the budget constraint. The earnings elasticity of these taxpayers with respect to the net-of-tax rat $1-\tau$ is defined as:

$$\tilde{\varepsilon}(w,\theta;\tau) \stackrel{\text{def}}{\equiv} -\frac{1-\tau}{\tilde{y}(w,\theta;\tau)}\frac{\partial \tilde{y}(w,\theta;\tau)}{\partial \tau}.$$

Define aggregate earnings as the sum of all individual incomes:

$$\overline{Y}(\tau) \stackrel{\text{def}}{\equiv} \iint_{(w,\theta)\in\mathbb{R}_+^*\times\Theta} \tilde{y}(w,\theta;\tau)\, f(w|\theta)dw\, d\mu(\theta).$$

Following Piketty and Saez (2013), the optimal tax rate τ_L is such that (Appendix A):

$$\tau_L = \frac{1-\overline{g}(\tau_L)}{1-\overline{g}(\tau_L)+\overline{e}(\tau_L)}, \tag{10}$$

where the Lagrange multiplier verifies:[6]

$$\lambda = \iint_{(w,\theta)\in\mathbb{R}_+^*\times\Theta} \Phi_U \langle w,\theta\rangle\, \mathscr{U}_c \langle w,\theta\rangle\; f(w|\theta)dw\, d\mu(\theta), \tag{11}$$

where the mean social marginal welfare weight is given by:

$$\overline{g}(\tau) \stackrel{\text{def}}{\equiv} \iint_{(w,\theta)\in\mathbb{R}_+^*\times\Theta} g\,(w,\theta)\; \frac{\tilde{y}(w,\theta;\tau)}{\overline{Y}(\tau)}\; f(w|\theta)dw\, d\mu(\theta), \tag{12}$$

and the mean earnings elasticity $e(\tau)$ is defined as:

$$\overline{e}(\tau) \stackrel{\text{def}}{\equiv} -\frac{1-\tau}{\overline{Y}(\tau)}\frac{\partial \overline{Y}}{\partial \tau} = \iint_{(w,\theta)\in\mathbb{R}_+^*\times\Theta} \tilde{\varepsilon}(w,\theta;\tau)\, \frac{\tilde{y}(w,\theta,\tau)}{\overline{Y}(\tau)} f(w|\theta)dw\, d\mu(\theta). \tag{13}$$

The following proposition explains why composition effects affect the implementation of the optimal linear tax rate.

Proposition 1. *In the optimal linear tax formula given by* (10)*, the share of income from* (w,θ)*-taxpayers in the aggregate earnings,* $\tilde{y}(w,\theta;\tau)f(w|\theta)/\overline{Y}(\tau)$ *impacts the mean social marginal welfare weight* $\overline{g}(\tau)$ *in* (12) *and the mean earnings elasticity* $\overline{e}(\tau)$ *in* (13)*.*

With a linear tax schedule, the optimal tax depends on a weighted mean of social welfare weights and earnings elasticities, where the weights $\tilde{y}(w,\theta;\tau)f(w|\theta)/\overline{Y}(\tau)$ are equal to the shares of income from (w,θ)-taxpayers in aggregate earnings. We define composition effects as the variations of the weighted means of welfare weights and earnings elasticities due to the change in the weights $\tilde{y}(w,\theta;\tau)f(w|\theta)/\overline{Y}(\tau)$ when one shifts from the actual economy (where $\overline{g}(\tau)$ and $\overline{e}(\tau)$ are estimated) to the optimal economy. Hence, ignoring composition effects typically biases the implementation of the optimal tax formula.

Importantly, the welfare weight $g(w,\theta)$ and the earnings elasticity $\tilde{\varepsilon}(w,\theta,\tau)$ of each type of taxpayers are also endogenous to the tax policy. So, to better isolate the direction of the bias induced by composition effects, we now specify individual preferences and the social objective to make $g(w,\theta)$ and $\tilde{\varepsilon}(w,\theta,\tau)$ exogenous. For this purpose, we assume the government's

[6] The notation $\langle w,\theta\rangle$ is a shortcut to indicate that the arguments are evaluated at the bundle chosen by taxpayers of type (w,θ).

objective is maximin which is equivalent to maximizing tax revenue $\tau \cdot \overline{Y} - E$ (Boadway and Jacquet, 2008) with welfare weights equal to zero. Substituting $\overline{g}(\tau_L) = 0$ into (10), the tax rate that maximizes tax revenue (or the Laffer rate) is:

$$\tau_L = \frac{1}{1 + \bar{e}(\tau_L)}. \tag{14}$$

Moreover, we assume individual preferences are isoelastic as in (4). This implies that income are given by $\tilde{y}(w, \theta; \tau) = (1 - \tau)^{\theta} w$, so that $\tilde{\varepsilon}(w, \theta; \tau) = \theta$ which is tax policy-invariant and homogeneous within each group. The average earnings within group θ is then given by:

$$\overline{y}(\theta, \tau) \stackrel{\text{def}}{\equiv} (1 - \tau)^{\theta} \int_{w \in \mathbb{R}_+^*} w f\left(w \middle| \theta\right) dw.$$

and the aggregate earnings can be rewritten as:

$$\overline{Y}(\tau) \stackrel{\text{def}}{\equiv} \int_{\theta \in \Theta} \overline{y}(\theta, \tau)\, d\mu(\theta) = \int_{\theta \in \Theta} (1 - \tau)^{\theta} \int_{w \in \mathbb{R}_+} w f(w|\theta) dw\, d\mu(\theta).$$

Hence, the elasticity of aggregate earnings given by (13) thus takes the following simpler expression:

$$\bar{e}(\tau) = \int_{\theta \in \Theta} \theta \, \frac{\overline{y}(\theta, \tau)}{\overline{Y}(\tau)}\, d\mu(\theta). \tag{15}$$

Let τ_0 denote the tax rate in the actual economy. Intuitively, if one neglects composition effects, one would implement the optimal tax formula (14) using $\bar{e}(\tau_0)$ instead of $\bar{e}(\tau_L)$. If the optimal tax rate is larger than the actual one, the rise from the actual to the optimal tax rate decreases earnings in every group. This decrease is higher for groups with a larger θ since their behavioral responses are larger than the ones of groups with a lower θ. Consequently, the rise in the tax rate is going to decrease (increase) the weight of high (low) θ-groups in the computation of the elasticity of aggregate earnings in (15) and we eventually get that $\bar{e}(\tau_L) < \bar{e}(\tau_0)$. Therefore, neglecting composition effects leads to underestimate the optimal tax rate, as proved in Appendix B and stated in the following proposition.

Proposition 2. *If the optimal linear revenue maximizing tax rate is higher (lower) than the actual tax rate, neglecting composition effects leads to a downward (upward) bias in the computation of the optimal tax rate.*

As a back-of-the-envelope numerical illustration, consider the case where the economy is made of two groups, a high elasticity one with $\theta_H = 0.4$ and a low elasticity one with $\theta_L = 0.1$. Assume both groups are of equal size $\mu(\theta_L) = \mu(\theta_H) = 0.5$ and are characterized by the same average income $\overline{y}(\theta, \tau)$ in the actual economy where the tax rate is assumed to be $\tau_0 = 0.25$. Then, ignoring the heterogeneity in the elasticity of labor supply, one obtains a revenue maximizing linear tax rate equal to $1/(1 + 0.25) = 80.0\%$. By contrast, taking into account composition effects leads to a revenue maximizing linear tax rate which rises to 82.1% from (14) and (15). With $\theta_H = 0.6$, we obtain a larger discrepancy: the optimal linear tax rate without composition effects is equal to $1/(1 + 0.35) \simeq 74.1\%$, while it rockets to 78.5% with composition effects.

IV Optimal nonlinear tax and composition effects

In this section, we study composition effects when the tax schedule is nonlinear. For this purpose, we improve the tax perturbation method initiated by Piketty (1997) and Saez (2001). In Subsection IV.1, we propose a tax perturbation method when individual characteristics are multidimensional and state sufficient conditions for using it. We define empirically measurable sufficient statistics (Subsection IV.2) that we use for characterizing desirable tax reforms and for deriving the optimal tax formula (Subsection IV.3). We then tell the reader about composition effects (Subsection IV.4).

IV.1 Sufficient conditions for the tax perturbation method

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

$$U^R(m;w,\theta) \stackrel{\text{def}}{\equiv} \max_y \; \mathscr{U}\left(y - T(y) + m\, R(y), y; w, \theta\right). \tag{16}$$

We denote by $Y^R(m;w,\theta)$ the income of workers of types (w,θ) after the reform and her consumption becomes $C^R(m;w,\theta) = Y^R(m;w,\theta) - T(Y^R(m;w,\theta)) + m\, R(Y^R(m;w,\theta))$. When $m = 0$, we have $Y^R(0;w,\theta) = Y(w,\theta)$ and $C^R(0;w,\theta) = C(w,\theta)$. Applying the envelope theorem to (16), we get:

$$\frac{\partial U^R}{\partial m}(m;w,\theta) = \mathscr{U}_c\left(C^R(m;w,\theta), Y^R(m;w,\theta); w, \theta\right)\; R(y). \tag{17}$$

Using (3), the first-order condition associated to (16) equalizes to zero the following expression:

$$\mathscr{Y}^R(y,m;w,\theta) \stackrel{\text{def}}{\equiv} 1 - T'(y) + m\, R'(y) - \mathscr{M}\left(y - T(y) + m\, R(y), y; w, \theta\right). \tag{18}$$

For simplicity, we drop the superscript R and write $\mathscr{Y}_y(Y(w,\theta);w,\theta)$ for $\mathscr{Y}_y^R(Y(w,\theta),0;w,\theta)$.[7] We define behavioral responses to tax reforms of direction R by applying the implicit function theorem at $m = 0$. For this purpose, we need the following assumptions:

Assumption 2. *Sufficient conditions for the tax perturbation method.*

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: $\mathscr{Y}_y\left(Y(w,\theta);w,\theta\right) < 0$.

iii) For all $(w,\theta) \in \mathbb{R}_+^* \times \Theta$, the function $y \mapsto \mathscr{U}\left(y - T(y), y; w, \theta\right)$ admits a unique global maximum over $\mathbb{R}_+$.

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

Part *i)* of Assumption 2 ensures that first-order condition (18) is differentiable.[8] Part *ii)* guarantees it is invertible in income y. Under *i)* and *ii)*, one can apply the implicit function theorem to (18) to describe how a local maximum of the individual maximization program (16) 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 $y \mapsto \mathscr{U}(y - T(y) + mR(y), y; w, \theta)$ may in general admit several global maxima among which individuals of type (w, θ) 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 is automatically satisfied when the tax function $T(y)$ is restricted to be linear as the indifference curves associated to $\mathscr{U}(.,.; w, \theta)$ are assumed 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 when 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.[9]

Thanks to Assumption 2, we can apply the implicit function theorem to *prove* that income is differentiable with respect to m after a tax reform in the direction $R(\cdot)$ (see Equation (19) below). Conversely, Golosov et al. (2014) do *assume* that the income function is locally Lipschitz continuous in tax reforms, while Hendren (2019) do *assume* that aggregate tax revenue, $\iint_{\theta \in \Theta, w \in \mathbb{R}_+} T(Y(w, \theta)) \, f(w|\theta) dw \, d\mu(\theta)$ varies smoothly in response to changes in the tax schedule, which is rather ad-hoc since these responses are endogenous. The strength of our approach is therefore to give micro-foundations to the property of smooth responses to tax reforms. In contrast, Hendren (2019)'s assumption allows for discrete changes in individual behavior in response to small tax changes, which is more general than our property of differentiable income. We can also note that Assumption 2 bears on tax functions that are endogenous objects. Considering only tax functions that verify this assumption is a restriction similar to considering only smooth allocations with no bunching, as done in the first-order mechanism design approach introduced by Mirrlees (1971). We drop the "no jumping" restriction by assuming Part *iii*) of Assumption 2 and verify ex-post in the simulations that the obtained tax schedule does satisfy Assumption 2.

[8] In practice, most of real world tax schedules are piecewise linear. In theory, bunching should occur at convex kink points and gaps in the income distribution should occur at concave kink points. In practice, bunching is very rare (with the noticeable exception of Saez (2010)) and gaps as well. This discrepancy between theory and reality can be due to the fact that taxpayers do not optimize with respect to the exact tax schedule but with respect to some smooth approximation of it, which verifies *i*) of Assumption 2.

[9] As pointed out by a referee, if the compensated elasticity is not bounded from above, Assumption 2 becomes verified only if the tax schedule is weakly convex.

IV.2 Behavioral responses

We now define the behavioral responses to a tax reform. Applying the implicit function theorem to $\mathscr{Y}^R(y, m; w, \theta) = 0$ at $m = 0$ yields:

$$\frac{\partial Y^R}{\partial m}(0; w, \theta) = -\frac{\mathscr{Y}_m^R(Y(w,\theta), 0; w, \theta)}{\mathscr{Y}_y(Y(w,\theta), 0; w, \theta)} \tag{19}$$

where:

$$\begin{aligned}
\mathscr{Y}_y^R(y, m; w, \theta) &= -T''(y) - \mathscr{M}_y(y - T(y) + m\, R(y), y; w, \theta) && \text{(20a)}\\
&- \mathscr{M}(y - T(y) + m\, R(y), y; w, \theta)\ \mathscr{M}_c(y - T(y) + m\, R(y), y; w, \theta), \\
\mathscr{Y}_m^R(y, m; w, \theta) &= R'(y) - R(y)\ \mathscr{M}_c(y - T(y) + m\, R(y), y; w, \theta). && \text{(20b)}
\end{aligned}$$

Checking out (20b), a tax reform affects individuals' decisions either because of a change in marginal tax rate (the term $R'(y)$ in (20b)) or because of a change in tax liability (the term proportional to $R(y)$ in (20b)).

Along the nonlinear income tax schedule, we define the *compensated elasticity* of earnings with respect to the marginal retention rate $1 - T'(.)$ as the elasticity of earnings for individuals of type (w, θ) to a change in the marginal tax rate, while leaving unchanged the level of tax at $y = Y(w, \theta)$. This tax reform has direction $R(y) = y - Y(w, \theta)$ with $R(Y(w, \theta)) = 0$ since the tax level is not modified at $y = Y(w, \theta)$ and with $R'(Y(w, \theta)) = 1$ since the marginal tax rate is uniformly modified. Using (2) and (20b), the compensated elasticity of earnings is equal to:

$$\varepsilon(w, \theta) \stackrel{\text{def}}{\equiv} \frac{1 - T'(Y(w,\theta))}{Y(w,\theta)} \frac{\partial Y^c}{\partial m} = \frac{\mathscr{M}(C(w,\theta), Y(w,\theta); w, \theta)}{-Y(w,\theta)\ \mathscr{Y}_y(Y(w,\theta); w, \theta)} > 0 \tag{21a}$$

which is positive due to Assumption 2 and where the superscript "c" stands for "compensated".

Along the nonlinear income tax schedule, the *income effect* is defined as the behavioral response to a lump-sum change m in tax liability with direction $R(y) = 1$. Plugging $R(Y(w, \theta)) = 1$ and $R'(Y(w, \theta)) = 0$ into (3) and (20b), the income effect is equal to:

$$\eta(w, \theta) \stackrel{\text{def}}{\equiv} \frac{\partial Y^i}{\partial m} = \frac{\mathscr{M}_c(C(w,\theta), Y(w,\theta); w, \theta)}{\mathscr{Y}_y(Y(w,\theta); w, \theta)} \tag{21b}$$

where the superscript "i" stands for "income effect". We have $\eta(w, \theta) < 0$ if leisure is a normal good, since then $\mathscr{M}_c > 0$.

Combining (21a) and (21b) with (20b), the way income of individuals (w, θ) reacts to any tax reform $R(\cdot)$ is given by:

$$\left.\frac{\partial Y^R}{\partial m}(0; w, \theta)\right|_{m=0} = \varepsilon(w, \theta) \frac{Y(w,\theta)}{1 - T'(Y(w,\theta))} R'(Y(w,\theta)) + \eta(w,\theta)\ R(Y(w,\theta)). \tag{21c}$$

Under Assumption 2, one can compute the elasticity $\alpha(w; \theta)$ of earnings with respect to the skill level:[10]

$$\alpha(w, \theta) \stackrel{\text{def}}{\equiv} \frac{w}{Y(w,\theta)} \dot{Y}(w, \theta) = \frac{w\ \mathscr{M}_w(C(w,\theta), Y(w,\theta); w, \theta)}{Y(w,\theta)\ \mathscr{Y}_y(Y(w,\theta); w, \theta)} > 0. \tag{21d}$$

[10]The dot above a variable stands for the partial derivative of this variable with respect to skill w.

Assumption 1 ensures this elasticity is positive. Hence, bunching cannot occur under Assumptions 1 and 2. This might be surprising since Rochet and Choné (1998) shows that bunching is generic in multidimensional screening problems. However, the reason why bunching occurs in their multidimensional nonlinear pricing model is because of the interplay between participation and self-selection constraints. This argument does not apply in our optimal tax problem without participation constraints.

It is worth stressing that $\varepsilon(w,\theta)$, $\eta(w,\theta)$ and $\alpha(w,\theta)$ denote *total* responses of earnings since they take into account the nonlinearity of the tax schedule as in Jacquet et al. (2013), see also Scheuer and Werning (2017). In Appendix C, we make the link between total responses and *direct* responses, the latter assuming a linear tax function (e.g. Saez (2001)).

Let $h(y|\theta)$ denote the conditional income density within group θ at income y and $H(y|\theta) \stackrel{\text{def}}{\equiv} \int_0^y h(z|\theta)dz$ the corresponding conditional income CDF. According to (21d) and Assumption 1, income $Y(\cdot,\theta)$ is strictly increasing in skill within each group. We then have $H\left(Y(w,\theta)|\theta\right) \equiv F(w|\theta)$ for each skill level w. Differentiating both sides of this equality with respect to w and using (21d), the two densities are linked by:

$$h\left(Y(w,\theta)|\theta\right) = \frac{f(w|\theta)}{\dot{Y}(w,\theta)} \qquad \Leftrightarrow \qquad Y(w,\theta)\, h\left(Y(w,\theta)|\theta\right) = \frac{w\, f(w|\theta)}{\alpha(w,\theta)}. \tag{22}$$

Let $W(\cdot,\theta)$ denote the reciprocal of $Y(\cdot,\theta)$ so that, within each group θ, 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, for individuals in group θ, the first-order condition $1 - T'(y) = \mathscr{M}(y - T(y), y; w, \theta)$ is verified at income y. The *unconditional* income density is given by:

$$\hat{h}(y) \stackrel{\text{def}}{\equiv} \int_{\theta\in\Theta} h(y|\theta)\, d\mu(\theta). \tag{23a}$$

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

$$\hat{\varepsilon}(y) = \int_{\theta\in\Theta} \varepsilon\left(W(y,\theta),\theta\right) \frac{h(y|\theta)}{\hat{h}(y)} d\mu(\theta). \tag{23b}$$

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:

$$\hat{\eta}(y) = \int_{\theta\in\Theta} \eta\left(W(y,\theta),\theta\right) \frac{h(y|\theta)}{\hat{h}(y)} d\mu(\theta). \tag{23c}$$

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

$$\hat{g}(y) = \int_{\theta\in\Theta} g\left(W(y,\theta),\theta\right) \frac{h(y|\theta)}{\hat{h}(y)} d\mu(\theta). \tag{23d}$$

IV.3 Tax perturbation and optimal tax formula

We now study when a tax reform is desirable. A tax reform with direction $y \mapsto R(y)$ affects the tax liability of a (w,θ)-worker through mechanical and behavioral effects as follows:

$$\left.\frac{\partial T(Y^R(m;w,\theta)) - m\, R(Y^R(m;w,\theta))}{\partial m}\right|_{m=0} = -\underbrace{R(Y(w,\theta))}_{\text{Mechanical}} + \underbrace{T'(Y(w,\theta))\,\frac{\partial Y^R}{\partial m}(0;w,\theta)}_{\text{Behavioral}} =$$

$$\varepsilon(w,\theta)\frac{T'(Y(w,\theta))}{1-T'(Y(w,\theta))}Y(w,\theta)\;R'(Y(w,\theta)) - \left[1-\eta(w,\theta)\;T'(Y(w,\theta))\right]R(Y(w,\theta)). \tag{24}$$

where the second equality is obtained using (21c). Combining Equations (8), (17) and (22)-(24) and integrating by parts, we get (see Appendix D):

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:*

$$\begin{aligned}\left.\frac{\partial \mathscr{L}^R}{\partial m}\right|_{m=0} &= \int_{y=0}^{\infty}\left\{\left[\hat{g}(y) - 1 + T'(y)\,\hat{\eta}(y)\right]\hat{h}(y) - \frac{d}{dy}\left[\frac{T'(y)}{1-T'(y)}\hat{\varepsilon}(y)\;y\;\hat{h}(y)\right]\right\}R(y)dy \qquad (25)\\ &+ \lim_{y\mapsto\infty}\frac{T'(y)}{1-T'(y)}\,\hat{\varepsilon}(y)\;y\;\hat{h}(y)\;R(y) - \lim_{y\mapsto 0}\frac{T'(y)}{1-T'(y)}\,\hat{\varepsilon}(y)\;y\;\hat{h}(y)\;R(y).\end{aligned}$$

An important point to notice is that, in general, implementing a reform with direction $R(\cdot)$ 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 (25). One can then define a balanced-budget tax reform with magnitude m and direction $R(\cdot)$ by combining it with the lump-sum rebate required to bind the budget constraint. As stated in the next lemma, Expression (25) allows one to characterize desirable tax reforms when λ is determined to verify:

$$1 = \int_{y=0}^{\infty}\left[\hat{g}(y) + T'(y)\;\hat{\eta}(y)\right]\hat{h}(y)dy = 0. \tag{26}$$

Lemma 2. *Under (26), a tax reform with direction $R(\cdot)$, combined with a lump-sum transfer to keep it budget-balanced, is socially desirable if either $\left.\frac{\partial \mathscr{L}^R}{\partial m}\right|_{m=0} > 0$ and $m > 0$ or $\left.\frac{\partial \mathscr{L}^R}{\partial m}\right|_{m=0} < 0$ and $m < 0$.*

The proof is in Appendix D. To obtain the optimal tax formula, we note that if the tax schedule is optimal, any tax reform $R(.)$ should have no first-order effect on the Lagrangian (7), i.e. (25) should be nil for any direction $R(\cdot)$. This leads to the following proposition which is proved in Appendix.

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

$$\frac{T'(y)}{1-T'(y)} = \frac{1}{\hat{\varepsilon}(y)}\;\frac{1-\hat{H}(y)}{y\;\hat{h}(y)}\left(1 - \frac{\int_{y}^{\infty}\left[\hat{g}(z) + \hat{\eta}(z)\;T'(z)\right]\hat{h}(z)dz}{1-\hat{H}(y)}\right) \tag{27}$$

If income effects were assumed away, Equation (26) 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 (27) generalizes the ABC terms described in Diamond (1998) and Saez (2001): (a) the behavioral responses to taxes denoted by $1/\hat{\varepsilon}(y)$, which, in the vein of Ramsey (1927), is the inverse of the mean compensated elasticity; (b) the social preferences and income effects $1 - \left(\int_y^\infty \left[\hat{g}(z) + \hat{\eta}(z)\, T'(z)\right] \hat{h}(z) dz\right) / \left(1 - \hat{H}(y)\right)$, which indicates the distributional benefits of increasing the tax liability by one unit for all workers with incomes above y and (c) the shape of the income distribution measured by the inverse of the local Pareto parameter $(1 - \hat{H}(y))/(y\, \hat{h}(y))$ of the income distribution. 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.

IV.4 Composition effects

With a nonlinear tax schedule, the optimal tax depends at income y on weighted means of compensated responses, incomes responses and welfare weights. Plugging Equations (23c) and (23d) into (27) leads to:

$$\frac{T'(y)}{1 - T'(y)} = \frac{1}{\hat{\varepsilon}(y)} \frac{1 - \hat{H}(y)}{y\, \hat{h}(y)} \left(1 - \iint_{\theta \in \Theta, z \geq y} \left[g\left(W(z,\theta),\theta\right) + \eta\left(W(z,\theta),\theta\right) T'(z)\right] \frac{h\left(W(z,\theta)|\theta\right)}{1 - \hat{H}(y)} dz\, d\mu(\theta)\right)$$

According to Equation (23b), the relevant weighted mean of compensated response $\hat{\varepsilon}(y)$ is computed among taxpayers who earn y and the weights are equal to the relative proportion $h(y|\theta)/\hat{h}(y)$ of taxpayers of group θ among taxpayers who earn income y. We define composition effects on mean compensated elasticities as the variation of $\hat{\varepsilon}(y)$ due to the change in the weights $h(y|\theta)/\hat{h}(y)$, which gives the relative share of θ-taxpayers among those who earn y, when one shifts from the actual economy where $\hat{\varepsilon}(y)$ is estimated to the optimal economy. We define composition effects on the mean of income responses $\eta(w,\theta)\; T'(Y(w,\theta))$ and on the mean of welfare weights $g(w,\theta)$ as the variation due to the change in the weights $h\left(W(z,\theta)|\theta\right) / \left(1 - \hat{H}(y)\right)$ when one shifts from the actual economy where they are estimated to the optimal economy. The means are computed among taxpayers who earn more than y. The weights are equal to the relative proportion $h\left(W(z,\theta)|\theta\right) / \left(1 - \hat{H}(y)\right)$ of taxpayers in group θ who earn z among taxpayers who earn income higher than y. Hence, as with a linear tax, ignoring composition effects typically biases the implementation of the optimal tax formula.

To study how composition effects impact the optimal tax rates, we consider maximin (to shut down composition effects on welfare weights[11]) and quasilinear preferences (to shut down

[11] Cuff (2000), Boadway et al. (2002), Brett and Weymark (2003), Choné and Laroque (2010) and Lockwood and Weinzierl (2015) introduce in the Mirrlees (1971) model an additional source of heterogeneity, typically preferences for leisure, that matters only for the computation of social welfare weights. Introducing this additional source of heterogeneity into our model would generate composition effects on welfare weights.

composition effects on income responses). This allows one to rewrite (27) as:

$$\frac{T'(y)}{1-T'(y)} = \frac{1}{\hat{\varepsilon}(y)}\,\frac{1-\hat{H}(y)}{y\,\hat{h}(y)} = \frac{1}{\int\limits_{\theta\in\Theta} \varepsilon\left(W(y,\theta),\theta\right)\frac{h(y|\theta)}{\hat{h}(y)}d\mu(\theta)}\,\frac{1-\hat{H}(y)}{y\,\hat{h}(y)}. \tag{28}$$

where the second equality follows (23b). Consider that marginal tax rates are larger at the optimum than in the actual economy, which is very likely under maximin in the United States. In this case, taxpayers respond to the rise of marginal tax rates from their actual to their optimal levels by reducing their incomes. These responses are larger in groups where the compensated elasticity $\varepsilon(W(y,\theta),\theta)$ is larger. Consequently, the income densities $h(y|\theta)$ of groups with a high compensated elasticity are much more shifted to the left than the income densities of groups with a low compensated elasticity. If the relative proportion $h(y|\theta)/\hat{h}(y)$ of high elasticity groups among income y earners is increased (decreased), these composition effects increase (decrease) the mean compensated elasticity $\hat{\varepsilon}(y)$ at income y, which reduces (increases) optimal marginal tax rate in (28).

However, there are two additional sources of endogeneity in (28). First, for each taxpayer, the compensated elasticity $\varepsilon\left(W(y,\theta),\theta\right)$ is endogenous to the tax policy. Second, the local Pareto parameter of the income distribution $y\,\hat{h}(y)/(1-\hat{H}(y))$ is also sensitive to the tax policy. To better isolate the direction of composition effects, we now consider isoelastic preferences (4) and we assume that each group-specific skill density is Pareto, with the same Pareto parameter (i.e. the local Pareto parameter of each group-specific income density is constant across groups and incomes). The latter assumption is obviously much more realistic for the upper part of the income distribution. Under these specifications, we get:

Proposition 4. *Assume isoelastic preferences (Equation (4)) and Pareto densities with identical Pareto coefficient. If the nonlinear revenue maximizing marginal tax rate is higher (lower) than the actual marginal tax rate, neglecting composition effects leads to a downward (upward) bias in the computation of the optimal marginal tax rate.*

This is formally shown in Appendix F. We now rely on numerical simulations to study composition effects under more general specifications of group-specific skill densities.

V Numerical illustrations of composition effects

In this section, we illustrate, thanks to simulations, that even in the case with only a second dimension of heterogeneity –the labor supply elasticity– the derived optimal tax schedule is significantly distinct from the one obtained when ignoring these effects.

V.1 Calibration

We first describe how we calibrate the model, i.e. how we select social preferences, individual preferences and the distribution of types.

To focus on composition effects that take place through compensated elasticities, we consider a maximin social objective, so there is no heterogeneity in social welfare weights $g(w, \theta)$. We also assume away income effects by specifying individual preferences to be quasilinear (see Equation (4)).

The meta-analysis of Bargain and Peichl (2016) shows that the elasticity along the intensive margin is lower for men than for women. We thus consider two groups of taxpayers, women with a high elasticity denoted by θ_H and men with a low elasticity denoted by θ_L. We select the values of θ_L and θ_H with two objectives in mind. First, we want the mean elasticity computed over the whole population to take a plausible value given the literature that estimates elasticities of taxable income (Saez et al., 2012). Second, we want the ratio of male over female elasticities to be realistic. Based on German data, Hermle and Peichl (2018) show that income responses to taxes differ substantially by gender. Based on Swedish data, Blomquist and Selin (2010) find that the labor earnings elasticity of women is five times larger than the one of men. Bargain and Peichl (2016) finds that women elasticity is two to six times larger than the one of men. Given this, we take $\theta_L = 0.1$ for men and $\theta_H = 0.4$ for women. This leads to a mean elasticity computed over the whole population equal to 0.237, which lies in the range $[0.12, 0.40]$ that correspond to the best available estimates for the long-run elasticity according to the meta-analysis of Saez et al. (2012).

We use CPS 2016 to calibrate the distribution of types. To avoid labor supply interactions within couples and to consider taxpayers facing the same tax schedule, we only consider singles without dependents. The fraction of women is $\mu(\theta_L) = 0.459$. For each earning observation, we infer the corresponding skill level by inverting taxpayers' first-order condition (2), using the gender of the corresponding observation (thereby the corresponding θ) and an approximation of the US tax schedule (see Appendix G). We then estimate each gender-specific type density $f(\cdot|\theta)$ thanks to the Silverman kernel density estimator. Earnings in CPS are topcoded, so, the kernel estimation is not valid for the upper part of each gender-specific skill distribution. Diamond (1998) and Saez (2001) emphasize that the upper part of these densities are well approximated by a Pareto, and this leads to a positive asymptotic optimal tax rate, implying that the zero-top tax rate result of Sadka (1976) and Seade (1977) is a very local result. We therefore choose to extend the kernel density estimations by Pareto distributions. A Pareto density $k\ y^{-1-p}$ is characterized by a scale k parameter and a Pareto parameter p. Using an estimate for p, we determine k and the skill level where the extension takes place to ensure identical left- and right-derivatives at the extension as well as continuity. We finally normalize the obtained density to get a total mass of 1.

According to Piketty and Saez (2013), the Pareto coefficient for top US incomes is $p = 1.5$. However, Atkinson et al. (2018) find that the Pareto coefficient is lower for men than for women in a set of OECD countries that, unfortunately, does not include the US. We therefore consider two scenarios. In the first scenario, we take the same asymptotic Pareto parameter equal to 1.5 for men and women. In the second scenario, we take different Pareto parameters for men and women.

V.2 Scenario 1: Gender specific elasticities, same asymptotic Pareto parameter

The solid blue line in Figure 1 displays the optimal marginal tax rates with composition effects which have the usual U-shaped pattern (Diamond, 1998). The numerical algorithm is described in Appendix G. To quantify the magnitude of the composition effects, we compare these optimal marginal tax rates with the ones obtained without composition effects.

To do so, we propose two different ways to figure out what optimal marginal tax rates would have been without composition effects. A first benchmark without composition effects consists in studying the workers as a single group with an homogeneous θ. One may, however, object that assuming a fixed θ is not a fair way to ignore composition effects, because a sophisticated calibration should use the information about differing elasticities for male and female workers and about how the share of women varies with income. In this "more sophisticated" benchmark without composition effects, one calculates, under the actual tax schedule, the elasticity at each income level as a weighted average of male and female elasticities where weights are the densities of male and female workers.

No composition effects benchmark with a fixed mean elasticity

In the first benchmark without composition effects, optimal marginal tax rates are described by the dashed red curve in Figure 1. In this benchmark, all taxpayers are assumed to belong to the same group characterized by a fixed direct elasticity $\theta = \overline{\theta}$ defined as the mean direct elasticity over the whole population, i.e.:

$$\overline{\theta} = \mu_L \theta_L + \mu_H \theta_H \simeq 0.238 \tag{29}$$

As detailed in Appendix G, the skill density in this economy is calibrated in a way similar to the calibration of both gender-specific skill densities in the economy with composition effects.

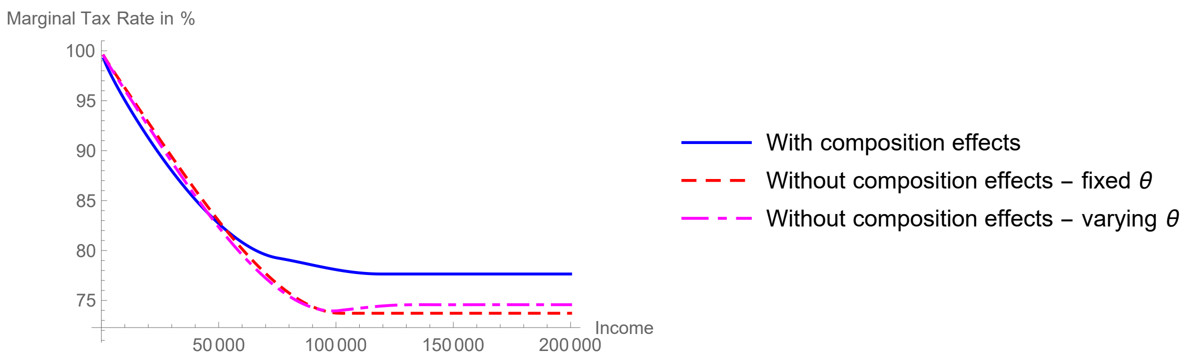

Figure 1: Optimal marginal tax rates with composition effects (solid blue line), without composition effects and fixed θ (dashed red lines), without composition effects and varying θ (dash-dotted pink lines)

We find that composition effects reduce marginal tax rates by as much as 1.5 percentage points below an income threshold around \$53,000 and increase them above this threshold, with a difference that rises to 4.4 percentage points at $y = \$97,000$. This can be seen on Figure 1 when comparing the solid blue curve (marginal tax rates with composition effects) with the dashed red curve (marginal tax rates without composition effects and fixed $\theta = \overline{\theta}$). Intuitively, marginal tax rates increase from the actual situation to the optimum. This induces taxpayers to reduce their labor supply and these behavioral responses are much larger for women than for men. Consequently, as described in Figure 2, both gender-specific income densities shift leftwards from the actual schedule (where dashed lines are used for the densities) to the optimal economy (where solid lines are used for the densities), but the shift is much larger for the income density of women (in blue) than for the one of men (in red). Figure 3 shows that the share of women at each income level is dramatically affected: the share of women rockets from 49% to 76% for the lowest income levels while it drops from 46% to 30% for the highest income levels. This pushes the mean compensated elasticities upwards (downwards) for low (high) income levels, thereby decreasing (increasing) optimal marginal tax rates through composition effects, as expected from Equation (28). This is illustrated in Figure 4 where the mean direct[12] elasticities with and without composition effects are displayed. The mean direct elasticity without composition effects is constant at $\overline{\theta} \simeq 0.238$ while the mean direct elasticity with composition effects markedly decreases with income.

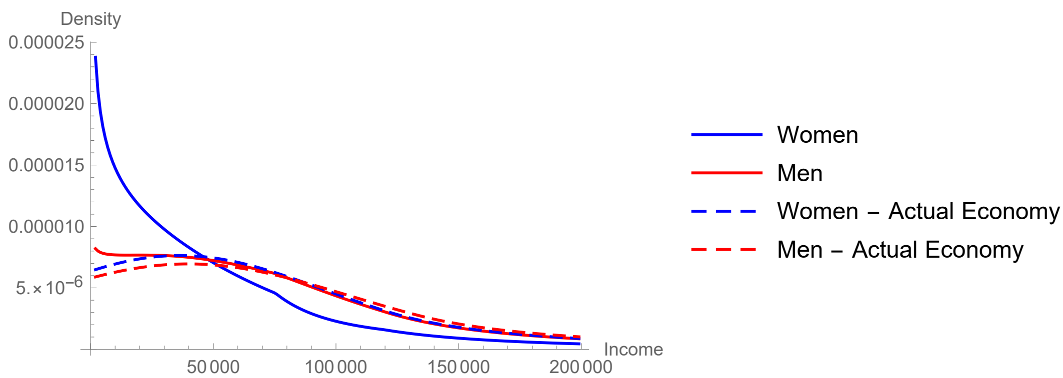

Figure 2: Gender-specific densities under the actual tax schedule (dashed red and blue lines) and under the optimal tax schedule (solid red and blue lines) in the economy with composition effects

We conduct sensitivity analysis. Differences between optimal marginal tax rates with and without composition effects can rapidly be magnified. For instance, when the male elasticity is $\theta_L = 0.1$ and female elasticity is $\theta_H = 0.6$ instead of 0.4, composition effects reduce marginal tax rates by as much as 3 percentage points around \$16,000 (instead of 1.5 percentage points), and increase them up to 8.1 percentage points at $y = \$97,000$ (instead of 4.4 percentage points).

[12] "Direct"means ignoring the impact of the curvature of tax function on the size of behavioral responses as detailed in Appendix C, i.e. substituting $T'' = 0$ in the definitions of the compensated and mean compensated elasticities.

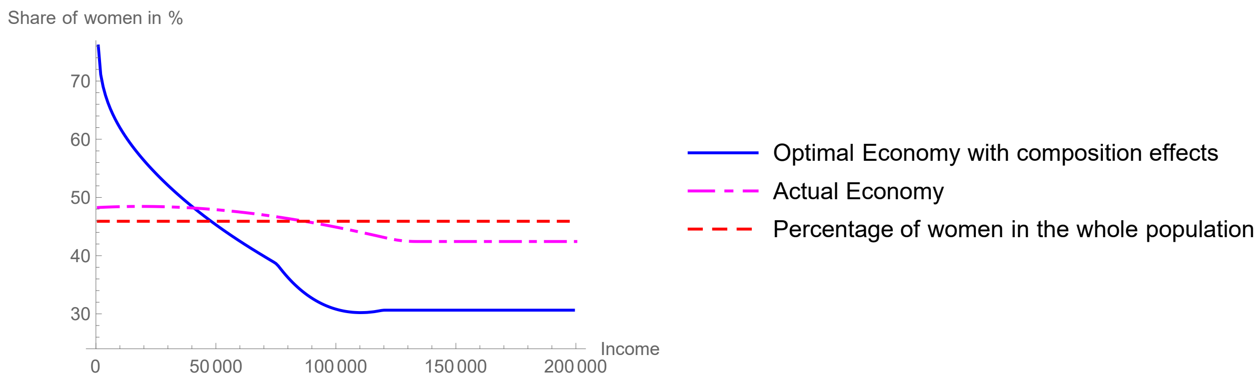

Figure 3: Share of women (in percentage) under the actual tax schedule (dash-dotted pink lines) and under the optimal tax schedule (in blue) in the economy with composition effects

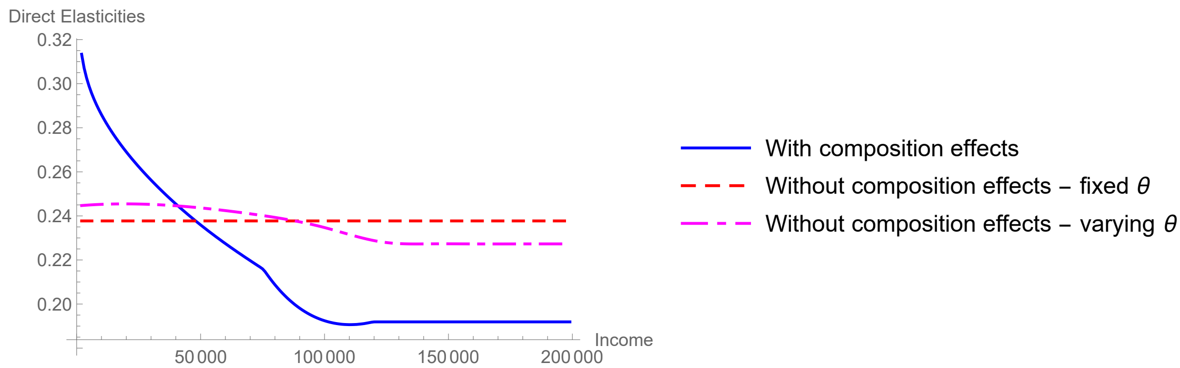

Figure 4: Mean direct elasticities with composition effects (solid blue line), without composition effects and fixed θ (dashed red lines), without composition effects and varying θ (dash-dotted pink lines)

No composition effects benchmark with an elasticity that varies with income

We now compare the optimal tax schedule with composition effects to the one obtained without composition effects but with θ varying along the income distribution. For this purpose, we compute optimal marginal tax rates by assuming, at each income level, a single value for the direct elasticity which depends on the share of women computed in our approximation of the actual economy with the actual tax schedule. At each income level, the direct elasticity $\widetilde{\theta}(y)$ is then calculated as:

$$\widetilde{\theta}(y) = \theta_L \frac{h_0(y|\theta_L)\ \mu(\theta_L)}{h_0(y|\theta_L)\ \mu(\theta_L) + h_0(y|\theta_H)\ \mu(\theta_H)} + \theta_H \frac{h_0(y|\theta_H)\ \mu(\theta_H)}{h_0(y|\theta_L)\ \mu(\theta_L) + h_0(y|\theta_H)\ \mu(\theta_H)} \tag{30}$$

where subscript 0 corresponds to our approximation of the income distribution in the actual economy. The share of women, $h_0(y|\theta_H)\ \mu(\theta_H) / (h_0(y|\theta_L)\ \mu(\theta_L) + h_0(y|\theta_H)\ \mu(\theta_H))$, decreases with income, as illustrated on Figure 3 (see the dash-dotted curve) so that the direct elasticity $\widetilde{\theta}(y)$ decreases with income. In Figure 4 where direct elasticities without composition effects are displayed, one sees that the mean direct elasticities are larger (lower) with varying θ than with fixed θ for low (high) incomes.

The optimal marginal tax rates for this second benchmark without composition effects are described by the dash-dotted pink lines in Figure 1 (see Appendix G for the detailed calculation). As expected from the way mean direct elasticities vary with income, optimal marginal tax rates are lower (larger) with varying θ than with fixed θ and for low (high) incomes. Hence, the differences between the optimal tax schedules with and without composition effects are slightly attenuated with the varying θ compared to the fixed θ but remain non negligible. The largest difference for low income is reduced to $1,2$ percentage point (instead of 1.5 percentage points) around $y = \$18,000$, while the largest difference for high incomes is 4.3 percentage points (instead of 4.4 percentage points) around income $y = \$93,000$.

V.3 Scenario 2: Gender specific elasticities and heterogeneous asymptotic Pareto parameters

Atkinson et al. (2018) find that the upper parts of men and women income distributions are characterized by different Pareto parameters in many OECD countries. For instance, they obtain $p_{\theta_L} \simeq 1.7$ for men and $p_{\theta_H} \simeq 2.1$ for women in the UK (see their Figure 3(B)). To investigate how our previous results are affected by the assumption of identical Pareto parameters for men and women, we choose $p_{\theta_H} = 1.8$ for women and $p_{\theta_L} = 1.4$ for men. With these values, Figure 5 indicates how the local Pareto parameter of the overall population varies with income. It confirms that our choice of $p_{\theta_H} = 1.8$ and $p_{\theta_L} = 1.4$ is consistent with the estimation of 1.5 in Piketty and Saez (2013).

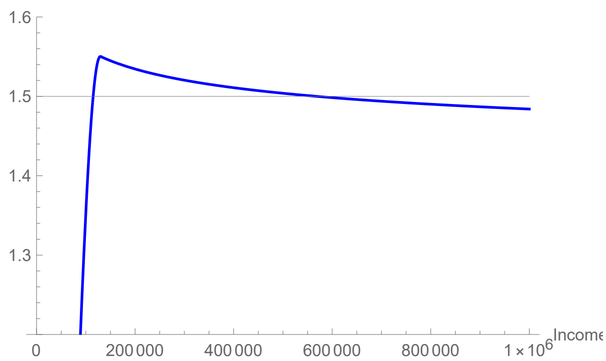

Figure 5: Local Pareto coefficient in our approximation of the actual economy with $p_{\theta_H} = 1.8$ and $p_{\theta_L} = 1.4$

Figure 6 displays the optimal marginal tax rates respectively with composition effects (solid blue line), without composition effects and with fixed θ (dashed red lines) and without composition effects and with varying θ (dash-dotted pink lines). Since the Pareto parameter of women is larger than the one of men, the share of women decreases with income and it tends to zero as income goes to infinity in our approximation of the actual economy. Consequently, the direct elasticity without composition effects but with varying θ decreases asymptotically to $\theta_L = 0.1$ (see Equation (30)). Figure 7, however, shows this convergence is actually very slow. This decrease in the direct elasticity induces that the optimal marginal tax rate without composition effects and with varying θ increases in income in the upper part of the distribution, while it remains constant without composition effects and with fixed θ.

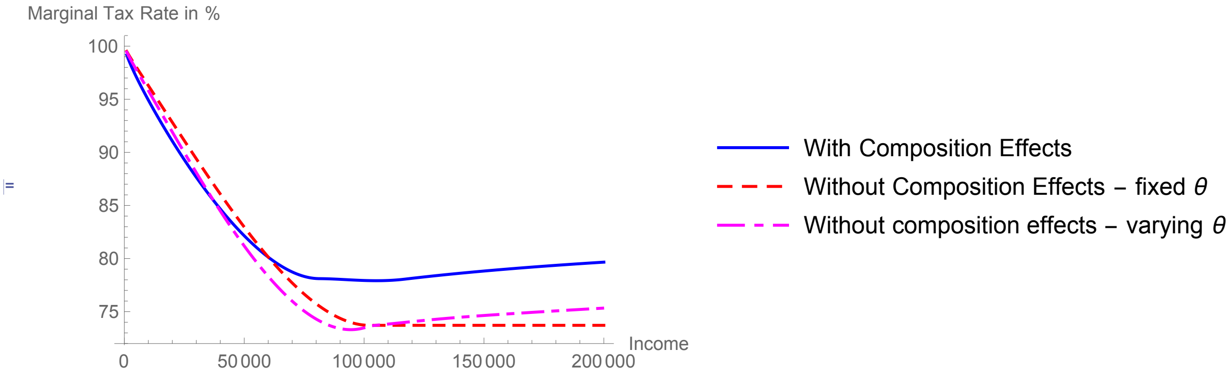

Figure 6: Optimal marginal tax rates with composition effects (solid blue line), without composition effects and fixed θ (dashed red lines), without composition effects and varying θ (dash-dotted pink lines)

The decrease in the share of women in the upper part of the distribution also affects the shape of optimal marginal tax rates with composition effects. Although the local Pareto parameter of the overall distribution remains close to 1.5 (see Figure 5), the gender-specificity in the Pareto parameters implies that the share of women is decreasing in income in the upper part of the distribution. The optimal marginal tax rate with composition effects is thereby increasing in income for high incomes. At income $\$200,000$, the optimal marginal tax rate with composition effects is 4.3 percentage points higher than optimal marginal tax rate without composition effects and varying θ and 5.9 percentage points higher than the optimal marginal tax rate without composition effects and fixed θ.

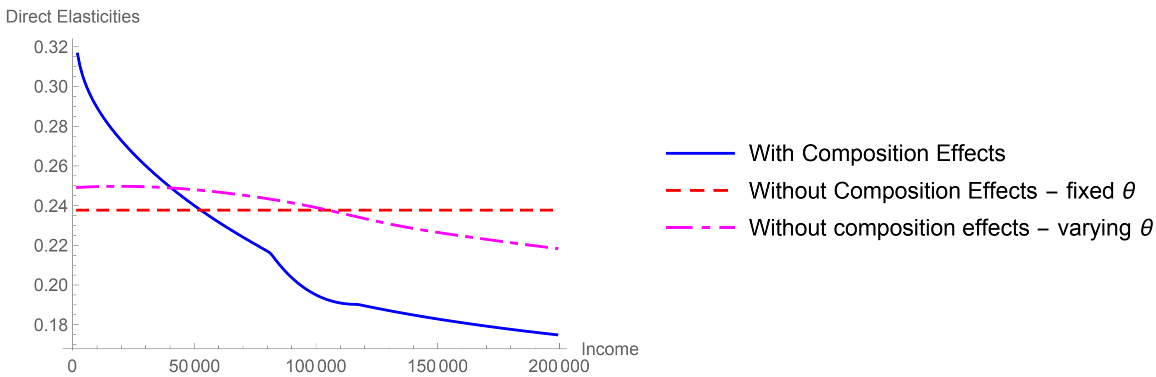

Figure 7: Assuming gender specific Pareto parameters, mean direct elasticities with composition effects (solid blue line), without composition effects and fixed θ (dashed red lines), without composition effects and varying θ (dash-dotted pink lines)

V.4 Optimal top marginal tax rates

Saez (2001) proposes two different approaches to compute optimal top marginal tax rates. One method consists in implementing the optimal nonlinear tax formula at a sufficiently high income level. We henceforth refer to the formula obtained in this way as the optimal *nonlinear* top tax rates formula and call the resulting tax rates, optimal nonlinear top tax rates. Another method consists in equating to zero the sum of mechanical and behavioral responses to a tax reform in which the marginal tax rate is increased by the same amount for all incomes above a given sufficiently high income threshold. As this approach is close to the one used to derive optimal linear tax rate, we say that it leads to the optimal *linear* top tax rates formula and optimal linear top tax rates.[13]

Both methods are generally believed to provide similar quantitative predictions. We now show that this equivalence does not necessarily hold true whenever unobserved heterogeneity is multidimensional. More specifically, under isoelastic individual preferences (4), maximin social preferences and group-specific skill densities that are Pareto with group-specific coefficients denoted p_θ, the optimal top tax rate increases with the weighted mean of the group-specific products θp_θ of the direct elasticity and Pareto coefficient, under both approaches. However, the weights take distinct values in each approach, as shown in Appendix H.

This can be understood intuitively. The optimal nonlinear top tax rate formula can be obtained by increasing marginal tax rate around a high income level y and increasing tax liability by a uniform amount for all incomes above y. Consequently, group-specific products θp_θ are weighted by the proportion of θ-*taxpayers* within taxpayers who earn more than y. In contrast, the optimal linear top tax rates formula consists in summing the responses when one increases the (marginal) tax rate for *all* taxpayers with incomes above an income threshold y. In this case, at income z larger than y, the tax liability varies in proportion to the difference $z - y$. Therefore, group-specific products θp_θ are weighted by the proportion of *income* above y earned by θ-taxpayers.

When the Pareto coefficients are identical across groups, the weights do not vary with income and are identical under both approaches. This is no longer true when Pareto coefficients vary across groups. To illustrate, take again $\theta_L < \theta_H$ and $p_{\theta_L} < p_{\theta_H}$. When the income level at which the optimal top tax formulas are evaluated tends to infinity, only the group θ_L with the lowest Pareto parameter prevails asymptotically since it is the group with the fatter income density. From (28), the optimal tax formula at the very top tends to $1/(1 + \theta_L\ p_{\theta_L})$. With the optimal linear top tax rates formula, the weight that multiplies $\theta_L p_{\theta_L}$ is, whatever the income level, always closer to one than with the optimal nonlinear top tax rates formula. This is due to the fact that the men skill distribution has a fatter tail. It implies that the proportion of incomes above y earned by men is larger than the proportion of men among taxpayers who earn more than y. Hence, at any income level, the optimal linear top tax rate is higher and closer to $1/(1 + \theta_L\ p_{\theta_L})$ than the optimal nonlinear top tax rate.

As an illustration, Figure 8 compares optimal asymptotic tax rates under both approaches using the calibration of the second scenario described in the preceding subsection, where in particular $\theta_L = 0.1$, $\theta_H = 0.4$, $p_{\theta_L} = 1.4$ and $p_{\theta_H} = 1.8$. As expected, the convergence of top tax rate towards $1/(1 + \theta_L\ p_{\theta_L}) \simeq 87.7\%$ is much faster when one uses the linear top tax rates formula. One can note that top tax rates converge towards $1/(1 + \theta_L\ p_{\theta_L})$ under both

[13] In Saez (2001), Section 3 is devoted to the linear top tax formula and Section 4 develops the nonlinear formula.

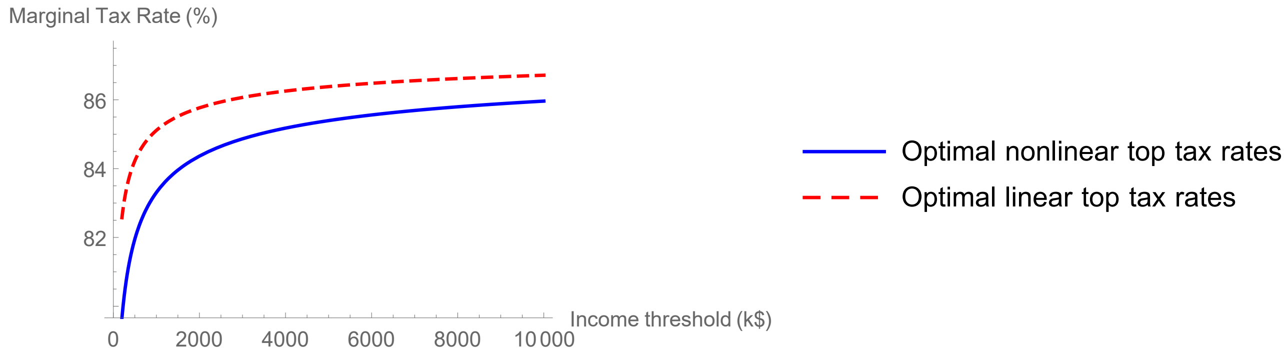

Figure 8: Top marginal tax rates calculated with the optimal nonlinear top tax rates formula vs with the optimal linear top tax rates formula

approaches only for extremely high income levels, so it is a very local result.

VI On the equivalence of tax perturbation and mechanism design methods

In this section, we show the equivalence between the tax perturbation approach (which relies on the sufficient conditions in 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 labor supply decisions such as those described in (1), or to directly select an allocation $(w,\theta) \mapsto (C(w,\theta), Y(w,\theta))$ that verifies the incentive constraints:

$$\forall w, \theta, w', \theta' \in (\mathbb{R}_+^* \times \Theta)^2 \qquad \mathscr{U}\left(C(w,\theta), Y(w,\theta); w, \theta\right) \geq \mathscr{U}\left(C(w',\theta'), Y(w',\theta'); w, \theta\right). \quad (31)$$

According to (31), individuals of type (w,θ) 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',θ').

In the mechanism design approach, it is usual to assume that the government selects among incentive-compatible allocations that are continuously differentiable (Salanié, 2005). Then, incentive constraints (31) imply the first-order incentive constraints, i.e.

$$\forall (w,\theta) \in \mathbb{R}_+^* \times \Theta \qquad \dot{U}(w,\theta) = \mathscr{U}_w\left(C(w,\theta), Y(w,\theta); w, \theta\right). \quad (32)$$

These first-order incentive constraints are necessary but not sufficient to verify the incentive constraints (31). The allocation also has to verify a monotonicity constraint according to which in each group, $Y(\cdot,\theta)$ is nondecreasing in skill. We define smooth allocations as follows:

Definition 1. *We say an allocation $(w,\theta) \mapsto (C(w,\theta), Y(w,\theta))$ is smooth if and only if it is continuously differentiable, it verifies (31) 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 the smooth allocation assumed in the first-order mechanism design approach. The proof is in Appendix I.

Proposition 5. *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.

ii) Any smooth allocation 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 Assumption 2 and assuming a smooth allocation are equivalent. The fact that, for each group θ, 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. In the mechanism design approach, the latter condition is related to the second order incentive constraints. 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.

Thanks to Proposition 5, first-order mechanism design and tax perturbation approaches are analogous.[14] The (first-order) mechanism design approach consists in choosing, among smooth allocations, 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 within the set of smooth and incentive compatible allocations. Since the allocation after perturbation has to be smooth, it is decentralized by a tax schedule that has to be smooth. Therefore, as stated in Proposition 5, the effects of a perturbation of the allocation that preserves its smoothness are equivalent to the responses of the allocation to a perturbation of the tax function that preserves smoothness. 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 unidimensional, the mechanism design approach can be developed without assuming smooth allocations. 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 the tax perturbation because of Assumption 2 but has been largely studied with the mechanism design approach. Note that the alternative "pathology" where individuals may be indifferent between two levels of income appears much

[14] In Jacquet and Lehmann (2016), optimal tax formula (28) is obtained thanks to a mechanism design method. To follow this method, one however needs to assume additive separable preferences.

more plausible under twice continuously differentiable tax schedules. 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).

With one dimension of heterogeneity, it is highly plausible that the optimal tax schedule does verify Assumption 2 or, equivalently, that the optimal allocation is smooth. With multidimensional heterogeneity, the plausibility of smooth optimal allocations is an open question. In our numerical calibrations, each group being characterized by a specific direct elasticity, optimal tax schedules are far from violating Assumption 2. Note that if the elasticity were continuously distributed and unbounded from above, Assumption 2 part *ii*) would imply that the marginal tax rate has to be nondecreasing, which would be an additional restriction on optimal tax schedules.

VII Concluding Comments

In this paper, using a new tax perturbation method, we provide formulas to calculate sufficient statistics in the presence of multidimensional individual heterogeneity. We also provide a set of sufficient conditions that guarantee the equivalence between the tax perturbation method and (first-order) mechanism design. Multidimensional heterogeneity generates a new channel through which sufficient statistics differ in the optimal and actual economies. We call this additional channel "composition effects ". These effects are due to the modification of the average behavioral response at each income level. We emphasize the key role they play in the calculation of sufficient statistics. We determine the sign of the bias that ignoring composition effects entails on the optimal linear and nonlinear tax schedules. We also run simulations to determine the direction and the size of the bias on the U.S. optimal tax schedule.

Our results call for more empirical studies on labor supply elasticities and distribution parameters for different demographic groups (e.g., according to age, ethnicity and gender), different types of workers (e.g., self-employed and salary workers) and sectors of activity.

Additionally, we expect that composition effects play a crucial role in many other applications beyond optimal income taxation. One such situation is that of social unemployment insurance (*à la* Baily (1978) and Chetty (2006)) when the job-search elasticities to unemployment benefits are heterogeneous. Another case is the optimal provision of public goods when the marginal rates of substitution between private and public goods are heterogeneous and endogenous. A final example is the regulation of a monopoly (*à la* Baron and Myerson (1982)) if uncertainty concerns not only the marginal cost but also the degree of convexity of the cost function, which is highly plausible empirically. Studying other applications where composition effects may play a role is part of our research agenda.

A Derivation of Equation (10)

Given (9), the government's program is:

$$\iint_{(w,\theta)\in\mathbb{R}_+^*\times\Theta} \Phi\left(\max_y \quad \mathscr{U}\left((1-\tau)y+\tau\,\overline{Y}(\tau)-E;w,\theta\right);w,\theta\right) f(w|\theta)dw\,d\mu(\theta).$$

The first-order condition is:

$$0=\iint_{(w,\theta)\in\mathbb{R}_+^*\times\Theta} \Phi_U\left\langle w,\theta\right\rangle\ \mathscr{U}_c\left\langle w,\theta\right\rangle\left[-\tilde{y}(w,\theta;\tau)+\overline{Y}(\tau)+\tau\,\frac{\partial\overline{Y}(\tau)}{\partial\tau}\right] f(w|\theta)dw\,d\mu(\theta).$$

Dividing this condition by $\lambda\,\overline{Y}(\tau)$ and using (8), (11) and (13) leads to:

$$\iint_{(w,\theta)\in\mathbb{R}_+^*\times\Theta} g(w,\theta)\,\frac{\tilde{y}(w,\theta;\tau)}{\overline{Y}(\tau)}\,f(w|\theta)dw\,d\mu(\theta)=1-\frac{\tau}{1-\tau}\overline{e}(\tau).$$

According to (12), the left-hand side is equal to $\overline{g}(\tau)$. Rearranging terms leads to (10).

B Proof of Proposition 2

Combining $\overline{y}(\theta,\tau)=(1-\tau)^\theta\int_{w\in\mathbb{R}_+^*} wf\left(w|\,\theta\right)dw$ and (15) the income-weighted average elasticity is given by:

$$e(\tau)=\frac{\displaystyle\int_{\theta\in\Theta}\theta\left(\frac{1-\tau}{1-\tau_0}\right)^\theta\overline{y}(\theta,\tau_0)\,d\mu(\theta)}{\displaystyle\int_{\theta\in\Theta}\left(\frac{1-\tau}{1-\tau_0}\right)^\theta\overline{y}(\theta,\tau_0)\,d\mu(\theta)}. \tag{33}$$

We now show that $e'(\tau)<0$.

$$\begin{aligned}
e'(\tau) &= -\frac{\displaystyle\int_{\theta\in\Theta}\theta^2\left(\frac{1-\tau}{1-\tau_0}\right)^\theta\overline{y}(\theta,\tau_0)\,d\mu(\theta)-e(\tau)\int_{\theta\in\Theta}\theta\left(\frac{1-\tau}{1-\tau_0}\right)^\theta\overline{y}(\theta,\tau_0)\,d\mu(\theta)}{\displaystyle(1-\tau)\int_{\theta\in\Theta}\left(\frac{1-\tau}{1-\tau_0}\right)^\theta\overline{y}(\theta,\tau_0)\,d\mu(\theta)}\\
&= -\frac{\displaystyle\int_{\theta\in\Theta}(\theta-e(\tau))\,\theta\left(\frac{1-\tau}{1-\tau_0}\right)^\theta\overline{y}(\theta,\tau_0)\,d\mu(\theta)}{\displaystyle(1-\tau)\int_{\theta\in\Theta}\left(\frac{1-\tau}{1-\tau_0}\right)^\theta\overline{y}(\theta,\tau_0)\,d\mu(\theta)}.
\end{aligned}$$

If $\theta<e(\tau)$, one has $(\theta-e(\tau))<0$, so that $(\theta-e(\tau))\theta>(\theta-e(\tau))e(\tau)$. The latter inequality also holds if $\theta>e(\tau)$. Hence, we get:

$$e'(\tau)\ <\ e(\tau)\frac{\displaystyle\int_{\theta\in\Theta}(\theta-e(\tau))\left(\frac{1-\tau}{1-\tau_0}\right)^\theta\overline{y}(\theta,\tau_0)\,d\mu(\theta)}{\displaystyle(1-\tau)\int_{\theta\in\Theta}\left(\frac{1-\tau}{1-\tau_0}\right)^\theta\overline{y}(\theta,\tau_0)\,d\mu(\theta)}=0.$$

In the absence of composition effects, there is a single group with an elasticity equal to $e(\tau_0)$. So, the optimal tax rate without composition effects is given by $1/(1+e(\tau_0))$ from (14). If the linear revenue maximizing tax rate τ_L is higher (lower) than the actual tax rate τ_0, one gets $e(\tau_0)>e(\tau_L)$ (resp. $e(\tau_0)>e(\tau_L)$), so that $1/(1+e(\tau_0))<1/(1+e(\tau_L))$ (resp. $1/(1+e(\tau_0))>1/(1+e(\tau_L))$). That is the optimal tax rate with compositions effects is larger (lower) than without composition effects.

C Total vs direct elasticities and income responses

The definitions of the behavioral responses given in (21a)-(21d) depend on the curvature of the tax function as emphasized by the term $T''(Y(w,\theta))$ in Equation (20a). Indeed, a circular process (Saez, 2001) is encapsulated into these definitions since any small tax reform or change in skill triggers a behavioral response that creates a change in marginal tax rate (whenever $T''(Y(w,\theta)) \neq 0$) that itself provokes a new behavioral response, etc. Behavioral responses (21a)-(21d) are therefore called *total*. In contrast, when the tax schedule is linear, one obtains the usual expressions for these responses that we call *direct* ones. Let $\varepsilon^\star(w,\theta)$, $\eta^\star(w,\theta)$ and $\alpha^\star(w,\theta)$ denote direct responses, i.e. (21a), (21b) and (21d) when $T'' = 0$ in (20a). From the implicit function theorem and Equation (19), for each type of behavioral response, the ratio of the total to the direct behavioral response is equal to the ratio of the value of $\mathscr{Y}_Y$ with $T''(Y(w,\theta))$ set to zero to the value $\mathscr{Y}_Y$, i.e.:

$$\frac{\mathscr{M}_y + \mathscr{M}\mathscr{M}_c}{T'' + \mathscr{M}_y + \mathscr{M}\mathscr{M}_c} = \frac{1 - T'(Y(w,\theta))}{1 - T'(Y(w,\theta)) + Y(w,\theta)\, T''(Y(w,\theta))\, \varepsilon^\star(w,\theta)}$$

where the second equality is obtained using (2), (3) and the definition of $\varepsilon^*(w,\theta)$ from (21a). We therefore obtain, as in Jacquet et al. (2013) that direct responses are timed by the above corrective term to obtain total responses as made explicit by the following equations:

$$\varepsilon(w,\theta) = \frac{1 - T'(Y(w,\theta))}{1 - T'(Y(w,\theta)) + Y(w,\theta)\, T''(Y(w,\theta))\, \varepsilon^\star(w,\theta)}\, \varepsilon^\star(w,\theta), \tag{34a}$$

$$\eta(w,\theta) = \frac{1 - T'(Y(w,\theta))}{1 - T'(Y(w,\theta)) + Y(w,\theta)\, T''(Y(w,\theta))\, \varepsilon^\star(w,\theta)}\, \eta^\star(w,\theta), \tag{34b}$$

$$\alpha(w,\theta) = \frac{1 - T'(Y(w,\theta))}{1 - T'(Y(w,\theta)) + Y(w,\theta)\, T''(Y(w,\theta))\, \varepsilon^\star(w,\theta)}\, \alpha^\star(w,\theta). \tag{34c}$$

D Proofs of Lemmas 1 and 2

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

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

Combining Equations (8), (17) and (24), the contribution of a (w,θ)-agent to the Lagrangian $\mathscr{L}^R$ varies with the magnitude of the tax reform by:

$$\left.\frac{\partial\left[T(Y^R(m;w,\theta)) - m\, R(Y^R(m;w,\theta)) + \frac{\Phi(U^R(m;w,\theta),w,\theta)}{\lambda}\right]}{\partial m}\right|_{m=0} = \tag{35}$$

$$\varepsilon(w,\theta)\frac{T'(Y(w,\theta))}{1 - T'(Y(w,\theta))} Y(w,\theta)\, R'(Y(w,\theta)) + \left[g(w,\theta) - 1 + \eta(w,\theta)\, T'(Y(w,\theta))\right] R(Y(w,\theta)).$$

Aggregating the latter expression over all types (w, θ), the partial (Gateaux) differential of the Lagrangian with respect to m, at $m = 0$, is equal to:

$$\begin{aligned}
\left.\frac{\partial \mathscr{L}^R}{\partial m}\right|_{m=0} &= \iint\limits_{\theta\in\Theta, w\in\mathbb{R}_+} \left\{ \frac{T'(Y(w,\theta))}{1-T'(Y(w,\theta))} Y(w,\theta)\, \varepsilon(w,\theta)\, R'(Y(w,\theta)) \right. \\
&+ \left. \left[T'(Y(w,\theta))\, \eta(w,\theta) - 1 + g(w,\theta)\right] R(Y(w,\theta)) \right\} f(w|\theta) dw\, d\mu(\theta) \\
&= \iint\limits_{\theta\in\Theta, y\in\mathbb{R}_+} \left\{ \frac{T'(y)}{1-T'(y)} y\, \varepsilon\left(W(y,\theta),\theta\right)\, R'(y) \right. \\
&+ \left. \left[T'(y)\, \eta\left(W(y,\theta),\theta\right) - 1 + g\left(W(y,\theta),\theta\right)\right] R(y) \right\} h(y|\theta) dy\, d\mu(\theta) \\
&= \int\limits_{y\in\mathbb{R}_+} \left\{ \frac{T'(y)}{1-T'(y)} y\, \hat{\varepsilon}(y)\, R'(y) + \left[T'(y)\, \hat{\eta}(y) - 1 + \hat{g}(y)\right] R(y) \right\} h(y) dy.
\end{aligned}$$

We use (35) to obtain the first equality. We use (22) for the change of variable from skill w to income y in the second equality. We use (23a)-(23d) 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 (25).

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 solve:

$$U^{\tilde{R}}(m; w, \theta) \stackrel{\text{def}}{\equiv} \max_y \qquad \mathscr{U}\left(y - T(y) + m\, R(y) + \ell(m), y; w, \theta\right). \tag{36}$$

Applying the envelope theorem to (36) at $m = 0$ yields:

$$\frac{\partial U^{\tilde{R}}}{\partial m}(0; w, \theta) = \left(R(y) + \ell'(0)\right)\; \mathscr{U}_c\left(C(w,\theta), Y(w,\theta); w, \theta\right). \tag{37}$$

Applying the implicit function theorem on the first-order condition

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

at $y = Y^{\tilde{R}}(m; w, \theta)$ and using (20b), (21b) and (21c) leads to:

$$\frac{\partial Y^{\tilde{R}}}{\partial m}(0; w, \theta) = \frac{\partial Y^{R}}{\partial m}(0; w, \theta) + \eta(w,\theta)\, \ell'(m). \tag{38}$$

We now denote respectively $\mathscr{B}^R(m)$, $\mathscr{S}^R(m)$ and $\mathscr{L}^R(m)$ 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 $\mathscr{L}^R(m) = \mathscr{B}^R(m) + (1/\lambda)\mathscr{S}^R(m)$. We symmetrically denote $\mathscr{B}^{\tilde{R}}(m)$, $SWF^{\tilde{R}}(m)$ and $\mathscr{L}^{\tilde{R}}(m)$ 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 for all m that $\mathscr{B}^{\tilde{R}}(m) = 0$ with:

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

We then obtain:

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

Using (38), we can then write:

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

so that:

$$\ell'(0) = \frac{1}{1 - \iint\limits_{(w,\theta)\in\mathbb{R}_+^*\times\Theta} T'(Y(w,\theta))\, \eta(w,\theta)\, f(w|\theta) dw\, d\mu(\theta)} \frac{\partial \mathscr{B}^R}{\partial m}(0). \tag{39}$$

Finally, using (37), we get:

$$\begin{aligned}
\frac{\partial \mathscr{S}^{\tilde{R}}}{\partial m}(0) &= \frac{\partial \mathscr{S}^R}{\partial m}(0) + \ell'(0) \iint\limits_{(w,\theta)\in\mathbb{R}_+^*\times\Theta} \Phi'_u(U(w,\theta);w,\theta)\, \mathscr{U}_c(C(w,\theta),Y(w,\theta);w,\theta)\, f(w|\theta) dw\, d\mu(\theta) \\
&= \frac{\partial \mathscr{S}^R}{\partial m}(0) + \frac{\iint\limits_{(w,\theta)\in\mathbb{R}_+^*\times\Theta} \Phi'_u(U(w,\theta);w,\theta)\, \mathscr{U}_c(C(w,\theta),Y(w,\theta);w,\theta)\, f(w|\theta) dw\, d\mu(\theta)}{1 - \iint\limits_{(w,\theta)\in\mathbb{R}_+^*\times\Theta} T'(Y(w,\theta))\, \eta(w,\theta)\, f(w|\theta) dw\, d\mu(\theta)} \frac{\partial \mathscr{B}^R}{\partial m}(0) \\
&= \lambda \frac{\partial \mathscr{L}^R}{\partial m}(0)
\end{aligned} \tag{40}$$

where the latter equality holds if and only if

$$\lambda = \frac{\int\limits_{(w,\theta)\in\mathbb{R}_+^*\times\Theta} \Phi'_u(U(w,\theta);w,\theta)\, \mathscr{U}_c(C(w,\theta),Y(w,\theta);w,\theta)\, f(w|\theta) dw\, d\mu(\theta)}{1 - \int\limits_{(w,\theta)\in\mathbb{R}_+^*\times\Theta} T'(Y(w,\theta))\, \eta(w,\theta)\, f(w|\theta) dw\, d\mu(\theta)} \tag{41}$$

which is equivalent to (26).

E Proof of Proposition 3

An optimal tax system implies that any tax reform $R(.)$ does not yield any first-order effect on the Lagrangian (7). That is (25) is nil at $m = 0$ for any direction $R(\cdot)$. This implies that $\lim\limits_{y\mapsto 0} \frac{T'(y)}{1-T'(y)}\, \hat{\varepsilon}(y)\, y\, \hat{h}(y) = \lim\limits_{y\mapsto\infty} \frac{T'(y)}{1-T'(y)}\, \hat{\varepsilon}(y)\, y\, \hat{h}(y) = 0$ and, for any income y, we have:

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

Integrating the latter equality for all income z above y and using $\lim\limits_{y\mapsto\infty} \frac{T'(y)}{1-T'(y)}\, \hat{\varepsilon}(y)\, y\, \hat{h}(y) = 0$ yields (27). Making y tends to 0 in (27) and using $\lim\limits_{y\mapsto 0} \frac{T'(y)}{1-T'(y)}\, \hat{\varepsilon}(y)\, y\, \hat{h}(y) = 0$ leads to (26).

F Proof of Proposition 4

Assume individual preferences (4). In the optimal economy, the skill level of taxpayers in group θ who earn income y is given by:

$$W(y,\theta) = (1 - T'(y))^{-\theta}\, y. \tag{42}$$

Under maximin ($\hat{g}(y) = g(w,\theta) = 0$) and using Equations (23a)-(23d) and (42), Equation (28) can be rewritten as:

$$\frac{T'(y)}{1-T'(y)} \int\limits_{\theta\in\Theta} \varepsilon(W(y,\theta),\theta)\frac{y\,h(y|\theta)}{1-H(y|\theta)} d\hat{\mu}(y,\theta) = 1 \tag{43}$$

where $\hat{\mu}$ is the CDF of θ among taxpayers earning an income larger than y, i.e.:

$$\hat{\mu}(y,\theta) \stackrel{\text{def}}{\equiv} \frac{\int\limits_{\theta'\in\Theta,\theta'\leq\theta} \left(1 - F\left((1-T'(y))^{-\theta'}\, y \middle|\, \theta'\right)\right)\, d\mu(\theta')}{\int\limits_{\theta'\in\Theta} \left(1 - F\left((1-T'(y))^{-\theta'}\, y |\, \theta'\right)\right)\, d\mu(\theta')} \tag{44}$$

so that $d\hat{\mu}(y,\theta) = \frac{1-H(y|\theta)}{1-\hat{H}(y)} d\mu(\theta)$. From (42), we have $H(y|\theta) \equiv F\left((1-T'(y))^{-\theta}\, y \middle|\, \theta\right)$. Differentiating both sides of this equality with respect to income y leads to:

$$\begin{aligned}
h(y|\theta) &= \left(1 + \frac{y\,T''(y)\,\theta}{1-T'(y)}\right)\, (1-T'(y))^{-\theta}\, f\left((1-T'(y))^{-\theta}\, y \middle|\, \theta\right) \\
y\,h(y|\theta) &= \left(\frac{1-T'(y) + y\,T''(y)\,\theta}{1-T'(y)}\right) W(y,\theta)\, f\left(W(y,\theta)|\,\theta\right) \\
\varepsilon(y,\theta)\, y\, h(y|\theta) &= \theta\, W(y,\theta)\, f\left(W(y,\theta)|\,\theta\right) \\
\varepsilon(y,\theta)\frac{y\,h(y|\theta)}{1-H(y|\theta)} &= \theta\,\frac{W(y,\theta)\, f\left(W(y,\theta)|\theta\right)}{1-F(W(y,\theta)|\theta)} = \theta\, p\left(W(y,\theta)|\theta\right)
\end{aligned} \tag{45}$$

where the third equality uses (34a) and $\varepsilon^{\star}(y,\theta) = \theta$ (with preferenecs preferences (4)) and the latest equality uses $H(y|\theta) = F(W(y,\theta)|\theta)$ and the following definition of the local Pareto parameter of the skill distribution:

$$p(w|\theta) = \frac{w\, f(w|\theta)}{1-F(w|\theta)}. \tag{46}$$

Plugging (45) into (43) leads to:

$$\frac{T'(y)}{1-T'(y)} = \frac{1}{\int\limits_{\theta\in\Theta} \theta\, p\left(W(y,\theta)|\theta\right)\, d\hat{\mu}(y,\theta)}. \tag{47}$$

Now assume that the conditional skill distribution in each group takes the form:

$$f(w|\theta) = k_\theta\, p\, w^{-(1+p)} \qquad \text{and} \qquad 1 - F(w|\theta) = k_\theta\, w^{-p} \qquad \text{if} \qquad w > \underline{w}_\theta. \tag{48}$$

Therefore the local Pareto parameter $p(w|\theta)$ (in (46)) is equal to p, provided that the income is high enough for $W(y,\theta)$ to remain above the positive lower bound of the skill distribution, which we henceforth assume. Substituting (42) into (47) yields:

$$\frac{T'(y)}{1-T'(y)} = \frac{\int_{\theta\in\Theta} k_\theta\, y^{-p}\, (1-T'(y))^{p\,\theta}\, d\mu(\theta)}{p\int_{\theta\in\Theta} \theta\, k_\theta\, y^{-p}\, (1-T'(y))^{p\,\theta}\, d\mu(\theta)}.$$

Define the weighted average direct elasticity as

$$\hat{\theta}(\zeta) \stackrel{\text{def}}{\equiv} \frac{\displaystyle\int_{\theta\in\Theta} \theta\, k_\theta\,(1-\zeta)^{p\,\theta}\, d\mu(\theta)}{\displaystyle\int_{\theta\in\Theta} k_\theta\,(1-\zeta)^{p\,\theta}\; d\mu(\theta)}$$

where ζ is the marginal tax rate. Substituting the latter equation in the previous one, we directly see that the optimal marginal tax rate $T'(y)$ is decreasing in the weighted-average direct elasticity $\hat{\theta}(T'(y))$. In the absence of composition effects, there is a single group with an elasticity equal to $\hat{\theta}(T_0'(y)) = \theta$. Consider, at a given level of income, that the optimal marginal tax rate $T'(y)$ is larger (lower) than the actual one $T_0'(y)$. In this case, $\hat{\theta}(T'(y))$ when composition effects prevail is lower (larger) than $\hat{\theta}(T_0'(y))$ obtained without composition effects since $\hat{\theta}'(\zeta) < 0$ as we now show.

Proof The previous equation can be rewritten as:

$$\begin{aligned}
\hat{\theta}'(\zeta) &= -\frac{\displaystyle\int_{\theta\in\Theta} p\,\theta^2\, k_\theta\,(1-\zeta)^{p\,\theta-1}\, d\mu(\theta) - \hat{\theta}(\zeta)\int_{\theta\in\Theta} p\,\theta\, k_\theta\,(1-\zeta)^{p\,\theta-1}\; d\mu(\theta)}{\displaystyle\int_{\theta\in\Theta} k_\theta\,(1-\zeta)^{p\,\theta}\; d\mu(\theta)} \\
&= -p\frac{\displaystyle\int_{\theta\in\Theta} \left(\theta-\hat{\theta}(\zeta)\right)\theta\, k_\theta\,(1-\zeta)^{p\,\theta-1}\, d\mu(\theta)}{\displaystyle\int_{\theta\in\Theta} k_\theta\,(1-\zeta)^{p\,\theta}\; d\mu(\theta)}.
\end{aligned}$$

If $\theta < \hat{\theta}(\zeta)$, then $\left(\theta-\hat{\theta}(\zeta)\right)\theta > \left(\theta-\hat{\theta}(\zeta)\right)\hat{\theta}(\zeta)$. The same inequality holds when $\theta > \hat{\theta}(\zeta)$. We can conclude that:

$$\hat{\theta}'(\zeta) < -p\,\hat{\theta}(\zeta)\frac{\displaystyle\int_{\theta\in\Theta} \left(\theta-\hat{\theta}(\zeta)\right)\; k_\theta\,(1-\zeta)^{p\,\theta-1}\, d\mu(\theta)}{\displaystyle\int_{\theta\in\Theta} k_\theta\,(1-\zeta)^{p\,\theta}\; d\mu(\theta)} = 0.$$

$\square$

Combining the latter result ($\hat{\theta}(T'(y))$ is lower (larger) than $\hat{\theta}(T_0'(y))$) with $T'(y)$ decreasing with $\hat{\theta}(T'(y))$, we can conclude that the optimal marginal tax rate with composition effect is higher (lower) than the optimal one obtained when one neglects composition effects.

G Numerical algorithm

Rewriting Equation (27) taking into account the specifications we use for our empirical exercise (maximin and isoelastic individual preferences) yields Equation (47) that we use for the simulations. This equation is very convenient because it implicitly defines the optimal marginal tax rate at income y independently of marginal tax rates at other incomes. It is a quasi-closed form. For each income y, the numerical algorithm starts from the actual marginal tax rate schedule and iterates the following steps until convergence:

1. Given $T'(y)$, the algorithm finds the skill level that corresponds to each observed income y, for each group θ using (42).
2. For each group θ, it finds the local Pareto parameter of the gender-specific skill distribution $p(w|\theta)$ using (46) and finds $d\hat{\mu}(y,\theta)$ using (44).
3. It updates the marginal tax rate using (47).

One needs to distinguish these three steps because the non-parametric calibration of the gender-specific skill densities prevents from numerically solving simultaneously (42), (44), (46) and (47). We finally check that the obtained tax schedule does verify Assumption 2.

For the economy without composition effects and fixed θ, we use the same algorithm, except that in Step 1, in (42), we use the mean direct elasticity $\overline{\theta} = \mu(\theta_L)\,\theta_L + \mu(\theta_H)\,\theta_H$ in the whole population instead of θ. As soon as each skill level is obtained from the income y found in the data, the algorithm approximates the density by a kernel density approximation. It then expands the latter density by a Pareto density with parameter 1.5 and rescales the obtained function to ensure the total mass is 1. Another difference for the economy without composition effects and fixed θ is that, in Step 3, we use:

$$\frac{T'(y)}{1-T'(y)} = \frac{1}{\overline{\theta}\; p\,(W(y))}$$

instead of (47).

For the economy without composition effects and varying θ, we first compute at each income level the weighted mean of direct elasticities in the actual economy as follows:

$$\widetilde{\theta}(y) = \frac{\theta_L\; h_0(y|\theta_L)\;\mu(\theta_L) + \theta_H\; h_0(y|\theta_H)\;\mu(\theta_H)}{h_0(y|\theta_L)\;\mu(\theta_L) + h_0(y|\theta_H)\;\mu(\theta_H)}.$$

We then follow the same steps as in the economy with composition effects and fixed θ but use $\widetilde{\theta}(y)$ in Step 1 and

$$\frac{T'(y)}{1-T'(y)} = \frac{1}{\widetilde{\theta}(y)\; p\,(W(y))}$$

in Step 3. In particular, we approximate the direct elasticity in the optimal economy at a given income level from the direct elasticity in the actual economy at the same income level. The skill distribution is calibrated by inferring from observed income levels the corresponding skill level from (42) using $\widetilde{\theta}(y)$. As in the other economy without composition effects, we approximate the density by a kernel density approximation, we expand the obtained density by a Pareto density with parameter 1.5 and rescale the obtained function to ensure the total mass is 1.

To calibrate the various skill densities, we approximate the tax schedule for singles without dependents by the US federal Income tax schedule, as follows:

Income tax bracket	Marginal tax rate
\$0 to \$9,225	10%
\$9,225 to \$37,450	15%
\$37,450 to \$90,750	25%
\$90,750 to \$189,300	28%
\$189,300 to \$411,500	33%
Above \$411,500	35%

Table 1: Income tax schedule used for calibration

H Top tax rate with a linear versus a nonlinear tax schedule

When the group-specific skill distribution is Pareto, the group-specific skill density takes the form:

$$f(w|\theta) = k_\theta\; p_\theta\; w^{-(1+p_\theta)} \qquad \text{and} \qquad 1 - F(w|\theta) = k_\theta\; w^{-p_\theta} \qquad \text{if} \qquad w \geq \underline{w}_\theta.$$

Hence, following (46), one has $p(w|\theta) = p_\theta$ whatever the income level. Assuming isoelastic individual preferences (4) and maximin social objective, the optimal nonlinear income tax schedule is given by (47), which given the assumption of Pareto group-specific skill distribution, simplifies to:

$$\frac{T'(y)}{1-T'(y)} = \frac{1}{\int\limits_{\theta\in\Theta} \theta \; p_\theta \; d\hat{\mu}(y,\theta)}.$$

The optimal nonlinear asymptotic marginal tax rate is therefore a weighted mean of $\theta \; p_\theta$ where, according to (44), the weights are given by the fraction of taxpayers in group θ within taxpayers who earn more than y.

Under the linear approach, one neglects the nonlinearity of the tax schedule for all incomes above y and one considers reforms of direction $z \mapsto (z-y)\mathbb{1}_{z\geq y}$. Then, applying (35) with the approximations $\varepsilon(w,\theta) = \theta$ and $T'(Y(w,\theta)) = \tau$ for all income above y, the optimal tax has to verify:

$$\frac{\tau}{1-\tau}\int_{\theta\in\Theta}\theta\left[\int_{w\geq W(y,\theta)} Y(w,\theta)f(w|\theta)dw\right]d\mu(\theta) = \int_{\theta\in\Theta}\left[\int_{w\geq W(y,\theta)}(Y(w,\theta)-y)\,f(w|\theta)dw\right]d\mu(\theta)$$

Let $y_m(y,\theta)$ denote the mean of income above y among taxpayers of group θ. One can rewrite the preceding condition as:

$$\frac{\tau}{1-\tau}\int_{\theta\in\Theta}\theta\; y_m(y,\theta)\;(1-F(W(y,\theta)|\theta))d\mu(\theta) = \int_{\theta\in\Theta}(y_m(y,\theta)-y)\;\;(1-F(W(y,\theta)|\theta))d\mu(\theta)$$

Under Pareto group-specific skill density, one has $y_m(y,\theta) = y\; p_\theta/\;(p_\theta - 1)$. Plugging this equality in the latter optimal tax condition leads to:

$$\frac{\tau}{1-\tau}\int_{\theta\in\Theta}\theta\; p_\theta\,\frac{1-F(W(y,\theta)|\theta)}{p_\theta-1}d\mu(\theta) = \int_{\theta\in\Theta}\frac{1-F(W(y,\theta)|\theta)}{p_\theta-1}d\mu(\theta)$$

Using $y_m(y,\theta) - y = y/\;(p_\theta - 1)$, the sum of income above y earned by taxpayers in group θ is equal to $y\;\;(1-F(W(y,\theta)|\theta))\;/\;(p_\theta-1)$. Hence the linear optimal asymptotic tax rate is a weighted mean of $\theta\; p_\theta$ where the weights are the shares of incomes above y earned by taxpayers who belong to group θ.

I Proof of Proposition 5

Part *i)* of Proposition 5.

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 θ. Moreover, $Y(\cdot,\theta)$ admits a positive derivative according to (21d). Finally, from (1) we get that:

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

Taking $y' = Y(w',\theta')$ leads to $C(w',\theta') = y' - T(y')$, so that the latter inequality leads to (31). Therefore the allocation $w \mapsto (C(\cdot,\theta),Y(\cdot,\theta))$ induced by $T(\cdot)$ verifies (31), thereby is smooth.

Part *ii)* of Proposition 5

Let $(w,\theta) \mapsto (C(w,\theta),Y(w,\theta))$ be a mapping defined over $\mathbb{R}_+^* \times \Theta$ which is smooth. 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,θ) and (w',θ') 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,θ) would be better off with the bundle $(C(w'),Y(w'))$ designed for individuals of type (w',θ'), which would be in contradiction with (31). A symmetric argument applies if $C(w,\theta) > C(w',\theta')$ by inverting the role of (w,θ) and of (w',θ'). We can then unambiguously define the tax schedule denoted $T(\cdot)$ that decentralizes this allocation by:

$$\forall y \in \mathbb{Y} \qquad T(y) \stackrel{\text{def}}{\equiv} Y(w,\theta) - C(w,\theta) \qquad \text{where } (w,\theta) \text{ are such that: } y = Y(w,\theta). \tag{49}$$

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

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

the solution of the latter is (w,θ) since $(w,\theta) \mapsto (C(w,\theta),Y(w,\theta))$ verifies the incentive constraints (31). Therefore, the tax schedule $T(\cdot)$ defined by (49) decentralizes the given allocation.[15]

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 (49) can be rewritten on $\mathbb{Y}(\theta)$ by:

$$T(y) = y - C\left(Y^{-1}(y,\theta),\theta\right). \tag{50}$$

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 θ such that $T(\cdot)$ verifies (50) in the neighborhood of y.

To show that $T(\cdot)$ verifies Part *i*) of Assumption 2, note that from (50), $T(\cdot)$ is continuously differentiable as $Y^{-1}(\cdot,\theta)$ and $C(\cdot,\theta)$ are continuously differentiable. Moreover, from (2), we have:

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

As $T(\cdot)$ and $Y^{-1}(\cdot,\theta)$ are continuously differentiable in y, and $\mathscr{M}(\cdot,\cdot;\cdot,\theta)$ is continuously differentiable in (c,y,w), $y \mapsto \mathscr{M}(y - T(y), y; Y^{-1}(w,\theta),\theta)$ 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, note that the first-order condition (18) can be rewritten as $\mathscr{Y}\left(Y(w,\theta);w,\theta\right) \equiv 0$ for all skill levels. Differentiating this equality with respect to skill leads to: $\mathscr{Y}_y\left(Y(w,\theta);w,\theta\right)\dot{Y}(w,\theta) + \mathscr{Y}_w\left(Y(w,\theta);w,\theta\right) = 0$. As $\mathscr{Y}_w\left(Y(w,\theta);w,\theta\right) = -\mathscr{M}_w\left(C(w,\theta),Y(w,\theta);w,\theta\right)$ which is positive from Assumption 1 and $\dot{Y}(w,\theta) > 0$ since allocations are assumed smooth, then one must have $\mathscr{Y}_y\left(Y(w,\theta);w,\theta\right) < 0$, 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^*,θ) 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 of 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 $\mathscr{C}(u,y;w,\theta)$ the consumption an individual of type (w,θ) should get to enjoy utility u while earning income y. Function $\mathscr{C}(\cdot,y;w,\theta)$ is the reciprocal of function $\mathscr{U}(\cdot,y;w,\theta)$. We get: $\mathscr{C}_u = 1/\mathscr{U}_c$, $\mathscr{C}_y = -\mathscr{U}_y/\mathscr{U}_c = \mathscr{M}$ and $\mathscr{C}_w = -\mathscr{U}_w/\mathscr{U}_c$. Let us denote:

$$\mathscr{Q}(w) \stackrel{\text{def}}{\equiv} \mathscr{C}\left(U(w,\theta),y';w,\theta\right) - y' + T(y').$$

[15] We have here followed Hammond (1979) very closely.

To be indifferent between earning income $Y(w,\theta)$ and income y', individuals of type (w,θ) have to receive after-tax income $\mathscr{C}\left(U(w,\theta),y';w,\theta\right)$ when they earn income y'. Therefore, $\mathscr{Q}(w)$ is a measure in monetary units of the difference in well-being for individuals of type (w,θ) 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 $\mathscr{Q}(w^*)=0$. We obtain:

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

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

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

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," *Journal of Economic Inequality*, 2018, *16(2)*, 225–256.

Baily, Martin Neil, "Some aspects of optimal unemployment insurance," *Journal of Public Economics*, 1978, *10* (3), 379 – 402.

Bargain, Olivier and Andreas Peichl, "Own-wage labor supply elasticities: variation across time and estimation methods," *IZA Journal of Labor Economics*, 2016, *5* (1), 1–31.

Baron, David P. and Roger B. Myerson, "Regulating a Monopolist with Unknown Costs," *Econometrica*, 1982, *50* (4), 911–930.

Bierbrauer, Felix J. and Pierre C. Boyer, "Politically feasible reforms of non-linear tax systems," CEPR Discussion Paper 13059 2018.

Blomquist, Sören and Håkan Selin, "Hourly wage rate and taxable labor income responsiveness to changes in marginal tax rates," *Journal of Public Economics*, 2010, *94* (11), 878 – 889.

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, R. and L. Jacquet, "Optimal Marginal and Average Income Taxation under Maximin," *Journal of Economic Theory*, 2008, *143*, 425–441.

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, "A general formula for the optimal level of social insurance," *Journal of Public Economics*, 2006, *90* (10), 1879 – 1901.

— , "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.

Costinot, Arnaud and Iván Werning, "Robots, Trade and Luddism: A Sufficient Statistic Approach to Optimal Technology Regulation," 2018.

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.

Ebert, Udo, "A Reexamination of the Optimal Nonlinear Income Tax," *Journal of Public Economics*, 1992, *49* (1), 47–73.

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*, 2018, *85(1)*, 511–557.

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, "Efficient Welfare Weights," NBER Working Papers issued in 2014, revised in August 2019 20351, 2019.

Hermle, Johannes and Andreas Peichl, "Jointly Optimal Taxes for Different Types of Income," CESifo Working Paper Series 7248, CESifo Group Munich 2018.

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 and Etienne Lehmann, "Optimal Taxation with Heterogeneous Skills and Elasticities: Structural and Sufficient Statistics Approaches," THEMA Working Papers 2016-04, THEMA, Université de Cergy-Pontoise 2016.

— **,** — **, 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, "Sufficient Statistics Revisited," mimeo 2018.

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

— **, Martin B. Knudsen, Claus Thustrup Kreiner, Søren Pedersen, and Emmanuel Saez**, "Unwilling or unable to cheat? Evidence from a tax audit experiment in Denmark," *Econometrica*, 2011, *79* (3), 651–692.

Lehmann, Etienne, Laurent Simula, and Alain Trannoy, "Tax Me If You Can! Optimal Nonlinear 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.

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.

Rochet, Jean-Charles 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.

Sachs, Dominik, Aleh Tsyvinski, and Nicolas Werquin, "Nonlinear Tax Incidence and Optimal Taxation in General Equilibrium," *Econometrica*, 2019, *Forthcoming.*

Sadka, Efraïm, "On Income Distribution, Incentive Effects and Optimal Income Taxation," *Review of Economic Studies*, 1976, *42*, 261–268.

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* (3), 180–212.

__ **and Stefanie Stantcheva**, "A Simpler Theory of Optimal Capital Taxation," *Journal of Public Economics*, 2018, *162*, 120–142.

Saez, Emmanuel Saez, 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 Contracts*, Cambridge, MA: MIT Press, 2005.

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

__ **and Iván Werning**, "The Taxation of Superstars," *Quarterly Journal of Economics*, 2017, *132*, 211–270.

Seade, J. K., "On the shape of optimal tax schedule," *Journal of Public Economics*, 1977, *7*, 203–235.

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