THEMA Working Paper n°2016-04
Université de Cergy-Pontoise, France

Optimal Taxation with Heterogeneous Skills and Elasticities: Structural and Sufficient Statistics Approaches

Laurence JACQUET, Etienne LEHMANN

THEMA Working Paper 2014-01, revised March 2016

Optimal Taxation with Heterogeneous Skills and Elasticities: Structural and Sufficient Statistics Approaches *

Laurence JACQUET† Etienne LEHMANN‡

March 2016

Abstract

The usual approach to calibrate optimal tax formulae consists in using the observed sufficient statistics, which is only correct as long as they correspond to the optimal sufficient statistics. In the very general case where agents are heterogeneous in many dimensions, we propose a new structural method (based on an allocation perturbation) from which we derive the optimal income tax formula and its optimal sufficient statistics computed from the observed ones. This allows us to quantify the error in the marginal tax rates entailed by using observed rather than optimal sufficient statistics. On US data, we show that this error can be considerable (up to 10 percentage points). We also call for a change of focus in the empirical analysis of top tax rates. Since individuals are heterogeneous along multiple dimensions, one needs to estimate the elasticity of those whose income density has the fatter tail.

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

*Previous versions of this paper has circulated under the titles "Optimal Nonlinear Income Taxation with Multidimensional Types: The Case with Heterogeneous Behavioral Responses" (2014) and "Optimal Income Taxation when Skills and Behavioral Elasticities are Heterogeneous" (2015). The authors acknowledge comments made by Pierre Boyer, Craig Brett, Pierre Cahuc, Philippe Choné, Helmuth Cremer, Vidar Christiansen, Eric Danan, André Decoster, Olivier Donni, Robert Gary-Bobo, Stéphane Gauthier, Ani Guerdjikova, Bas Jacobs, Eckhard Janeba, Guy Laroque, Laurent Linnemer, François Maniquet, David Martimort, Jean-Baptiste Michau, Matias Nunez, Nicola Pavoni, Julien Prat, Stéphane Robin, Guillaume Rocheteau, Xavier Ruiz del Portal, Emmanuel Saez, Florian Scheuer, Laurent Simula, Johannes Spinnewijn, Kevin Spiritus, Stefanie Stantcheva, Alain Trannoy, Mark J. Tremblay, Aleh Tsyvinsky, Nicolas Werquin, John Weymark and Danny Yagan. Any remaining errors are ours. Laurence Jacquet's research has been supported by a Marie Curie FP7 Integration Grant within the 7th European Union Framework Program and is supported by a Labex MMEDII (ANR11-LBX-0023-01). A part of this research was undertaken while Etienne Lehmann was visiting UC Berkeley with financial support from the Center for Equitable Growth at UC Berkeley, Institut Universitaire de France and from the Fulbright Program.

†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. Laurence Jacquet is also research fellow at Oslo Fiscal Studies (UiO), CESifo and IRES- Université Catholique de Louvain.

‡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 IRES-Université Catholique de Louvain, IDEP, IZA and CESifo and is junior member of IUF.

I Introduction

The optimal tax formulae of Mirrlees (1971) have become a cornerstone of the public finance literature. A feature that makes these formulae particularly attractive is that they can easily be empirically implemented through the so-called "sufficient statistics"or so most researchers would argue (e.g., Saez (2001, 2002), Chetty (2012)). In this paper, we explain why the implementation of the optimal tax schedule using these sufficient statistics is generally improper, and we point out to their correct implementation in a general optimal tax framework that involves multidimensional individual heterogeneity. To do so, we propose a new structural method that uses calculus of variation based on an "allocation perturbation". It relies on a pooling function that characterizes individuals with distinct characteristics (e.g., skills and taxable income elasticities) who earn the same income. We treat the population as composed of distinct groups, which are subsets of individuals with the same vector of characteristics except for skills. Our allocation perturbation method allows us to derive a structural optimal tax formula (i.e. a formula expressed in terms of the policy-invariant primitives of the model) that we can reformulate as an ABC tax formula à la Diamond (1998), in which each term results from a very specific averaging procedure.

From this structural tax formula, we are able to find the income tax formula based on optimal sufficient statistics (which differ from the observed statistics). We then show how to recover this formula using the widespread "tax perturbation"method and proceed to compare both methods. While more intuitive, the tax perturbation requires restrictions not only on the tax function that is perturbed, but also on the way the allocation is affected by the tax perturbation. As the allocation is endogenous, imposing restrictions on it is ad-hoc. This is the *internal inconsistency* of the tax perturbation approach. Conversely, our allocation perturbation method requires a mild assumption on preferences and some restrictions on the perturbed allocation. It therefore does not suffer from the internal inconsistency of the tax perturbation approach.

More importantly, we show that simple extensions of the usual "tax perturbation"approach to the multidimensional case lead to improper definitions of the sufficient statistics and that the structural tax formula is needed for proper calibration. First, a group-specific corrective term is required to calibrate the tax schedule with estimated sufficient statistics. This term encapsulates the circular process which is inherent to the nonlinear tax schedule, and nevertheless neglected in the applied literature. Second, one has to follow a specific averaging procedure, which is a far cry from the simple extension of the unidimensional case that would consist in computing the simple average of every estimated sufficient statistic and then multiplying each average by the same corrective term. Instead, every optimal sufficient statistic at any income level is a weighted average that requires as many corrective terms as there are groups in which individuals earn this income level and group-specific densities as weights. Remarkably, this procedure highlights the importance of *composition effects* which make every weighted average of corrected sufficient statistic distinct in the actual and optimal economies.

Empirically, we show, using US Current Population Survey data, that implementing the optimal tax formula with sufficient statistics computed in the actual economy rather than in the optimal one can have dramatic consequences. Our results highlight that this is already true in the unidimensional case, but that multidimensional heterogeneity makes things worse, as the bias has *(i)* a greater magnitude and *(ii)* plays in different directions at different points of the income distribution. We also show that neglecting multidimensional heterogeneity strongly biases the optimal tax schedule. Last but not least, we call for a change of focus in the empirical analysis of top tax rates. Since individuals are heterogeneous along multiple dimensions, one needs to estimate the elasticity of those whose income density has the fatter tail. To illustrate this point, we assume that each group has a distinct taxable income elasticity and an unbounded Pareto conditional skill distribution (as observed empirically). If the asymptotic Pareto coefficients differ across groups, only the income elasticity of the group with the fatter-tailed Pareto distribution matters for calculating the asymptotic tax rate. This elasticity can be drastically different from the average taxable income elasticity among, e.g., the top 1%. In the literature, asymptotic tax rates are typically calibrated using this average elasticity among high income earners (Saez et al., 2012, Piketty and Saez, 2013), which may lead to erroneous recommendations.

Last but not least, a strength of our paper is that our method is general enough to solve a large set of adverse selection problems for which it is crucial, but challenging, to include multidimensional heterogeneity. We show that our framework encompasses many policy-oriented applications. It can be interpreted to derive the nonlinear optimal income tax schedules,[1] e.g., when individuals earn labor income and non-labor income (capital income, income from renting out property, etc.), or when the income of households is jointly taxed.

Related literature

The first contribution of our paper is theoretical: Our paper is the first to allow for a very general form of pooling in screening models where individuals differ along many unobserved characteristics and perform a single action (here, intensive labor supply decision). The usual approach to these models relies on the assumption that the action only depends on a one-dimensional aggregation of the multidimensional unobserved heterogeneity. This is the case, for instance, in models of optimal income taxation such as those proposed in Brett and Weymark (2003), Boadway et al. (2002), Choné and Laroque (2010), Lockwood and Weinzierl (2015), Rothschild and Scheuer (2013, 2016, 2014), Scheuer (2013), Gomes et al. (2014) and Scheuer (2014). By contrast, we relax this assumption (in Subsection III.1, we address its implications in details and compare our approach to those adopted in the aforemetioned literature). Since our method does no rely on an aggregator, we are able to simultaneously consider, in an optimal income tax model, heterogeneity in income and heterogeneity in behavioral elasticities.

[1] Beyond optimal tax problems, our model can be applied to the nonlinear monopoly pricing problem in a more general framework than Laffont et al. (1987).

Random participation models make up another strand of the literature where multidimensional heterogeneity is taken into account, although in a very specific way. In these models, individuals differ in skill and in a cost of participation (Rochet and Stole, 2002, Kleven et al., 2009, Jacquet et al., 2013) or of migration (Lehmann et al., 2014, Blumkin et al., 2014) and this latter dimension of heterogeneity matters only for the participation/migration margin. In these papers, the one-dimensional aggregation implies that individuals who earn the same income are characterized by the same level of aggregated characteristics and are therefore constrained to react identically to any tax reform. While departing from this restriction, we show (in an appendix available upon request) that our model can readily be extended to include a random participation constraint.

In parallel, a growing literature (e.g. Golosov et al. (2014), Kleven et al. (2009), Renes and Zoutman (2015)) studies screening problems when individuals differ along a number of characteristics that is equal to (or lower than) the number of actions they perform (e.g., labor income and saving decision). In contrast to our paper, this literature neglects pooling.[2]

The second contribution of our paper is empirical, and relates to the "sufficient statistics" literature (e.g., Piketty (1997), Saez (2001), Chetty (2009), Diamond and Saez (2011), Piketty and Saez (2013), Hendren (2014), Scheuer and Werning (2016)). In a nutshell, this approach consists in focusing on empirical combinations of the primitives of the model, known as "sufficient statistics", that can be estimated using data, rather than considering the full economic structure (Chetty, 2009). While this may be enough to indicate the direction of desirable tax reforms (Golosov et al., 2014), we show that a structural tax formula is required to correctly implement the tax schedule.

The sufficient statistics approach derives the optimal tax schedule by using a tax perturbation, i.e. by considering the effects of an infinitesimal tax reform on the government's objective. The tax perturbation method is a way to provide a clear intuition for the economics behind the optimal tax formula.[3] However, this method is unclear about the treatments of pooling and of the circularity process and it faces the above-mentioned internal inconsistency.[4] By contrast, we rely on an *allocation perturbation* method that avoids these mathematical weaknesses by directly optimizing over smooth incentive-compatible allocations. Our method also encapsulates the circularity in a clear-cut way.[5]

The remainder of the paper is organized as follows. In Section II, we present the model

[2] In Rochet and Choné (1998), individuals endowed with distinct characteristics choose the same action due to a strong conflict between the incentive constraints and a participation constraint. The latter arises because they consider a nonlinear pricing model where consumers have the same outside option and are then "bunched "in this outside option. This is irrelevant in our framework.

[3] Saez (2001) does not advocate the tax perturbation as a formal proof but calls it a "heuristic proof" and rigorously shows that a rewriting of Mirrlees (1971)'s structural tax formula leads to a tax formula in terms of sufficient statistics. His proof is however only valid under a one-dimensional unobserved heterogeneity. Our paper provides a formal derivation of the sufficient statistics-based tax formula with multidimensional heterogeneity.

[4] This circularity process is neglected in Piketty (1997) and Diamond and Saez (2011) and considered in Saez (2001), Hendren (2014), Golosov et al. (2014) and in the appendix of Piketty and Saez (2013).

[5] This involves incorporating the circularity process in the elasticities (as in Jacquet et al. (2013)) rather than in the income density (as in Saez (2001)). Our way is becoming more widespread, see e.g. Scheuer and Werning (2016).

and emphasize its flexibility by showing how it can easily be adapted to study several policy applications. In Section III, we derive the structural tax formula using our allocation perturbation method and we show that optimal marginal tax rates are positive under utilitarian and maximin social preferences. In Section IV, we express the optimal tax formula in terms of sufficient statistics, specify the correct averaging procedure of the latter and discuss the respective virtues and limitations of the tax perturbation and allocation perturbation approaches. We also study the asymptotic tax rate. In Section V, we numerically quantify the crucial role played by multidimensional heterogeneity and highlight the biases due to the typical miscalculations of the sufficient statistics. We conclude in the last section.

II Model

Individuals differ along their skill level $w \in \mathbb{R}_+$ and along a vector of characteristics denoted $\theta \in \Theta$. Labor supply elasticity can be one of these individual characteristics. We call a *group* a subset of individuals with the same θ. We assume that the set of groups Θ is compact and 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 continuously distributed according to the conditional skill density $f(\cdot|\theta)$ which is positive over the support $\mathbb{R}_+$. The conditional CDF is denoted $F(w|\theta) \stackrel{\text{def}}{\equiv} \int_0^w f(x|\theta)dx$. The size of the total population is normalized to one, so that:

$$\int_{\theta\in\Theta} \left\{ \int_0^{+\infty} f(w|\theta)dw \right\} d\mu(\theta) = 1.$$

Following Mirrlees (1971), the government levies a tax $T(.)$ which is a non-linear function of pre-tax income y (for short, income hereafter) and does not depend on individual types (w, θ).

II.1 Individual choices

Every worker of type (w, θ) derives utility from consumption c and disutility from effort. Effort captures the quantity as well as the intensity of labor supply. Let $v(y; w, \theta)$ be the disutility of a worker of type (w, θ) to obtain income $y \geq 0$ with $v_y(\cdot), v_{yy}(\cdot) > 0 > v_w(\cdot)$.[6] Disutility is increasing and convex in income, and decreasing in skill w. This is because earning a given income requires less effort to a more productive agent.[7] Individual preferences are described by the twice differentiable utility function:

$$\mathscr{U}(c, y; w, \theta) = u(c) - v(y; w, \theta) \qquad \text{with} \qquad u'(\cdot) > 0 \geq u''(\cdot). \tag{1}$$

[6] For any function f of a single variable, we denote f' its first derivative and f'' its second derivative. For any function g of multiple variables $x, y, ...$, we denote g_x its first-order partial derivative with respect to x and g_{xy} its second-order partial derivative with respect to x and y, etc.

[7] The latter assumption is standard. For instance, when income is equal to the product of effort and skill, $y = w \times \ell$ and when preferences depend on effort ℓ, we get $v(y; w, \theta) \equiv \mathcal{V}\left(\frac{y}{w}; \theta\right)$ with $\mathcal{V}_\ell(\cdot) > 0, \mathcal{V}_{\ell\ell}(\cdot) > 0$. The assumption $\mathcal{V}_\ell > 0$ implies $v_y > 0 > v_w$. The assumption $\mathcal{V}_{\ell\ell} > 0$ implies $v_{yy} > 0 > v_{yw}$.

Additive separable utility as in (1) is commonly assumed in optimal taxation and in the adverse selection literature with multidimensional heterogeneity (e.g., Rochet (1985), Wilson (1993), Rochet and Choné (1998), Rochet and Stole (2002)). The marginal rate of substitution between income y and consumption c is:

$$\mathscr{M}(c,y;w,\theta) \overset{\text{def}}{\equiv} -\frac{\mathscr{U}_y(c,y;w,\theta)}{\mathscr{U}_c(c,y;w,\theta)} = \frac{v_y(y;w,\theta)}{u'(c)}. \tag{2}$$

We impose a single-crossing (Spence-Mirrlees) condition within each group of individuals endowed with the same θ. Starting from any positive level of consumption and pre-tax income, more skilled workers need to be compensated with a smaller increase in their consumption to accept a unit rise in income. We therefore assume that for each $\theta \in \Theta$ and for any bundle (c,y), the marginal rate of substitution $\mathscr{M}(c,y;w,\theta)$ is a decreasing function of the skill level:

$$\forall (c,y,\theta) \in \mathbb{R}_+ \times \mathbb{R}_+ \times \Theta : \qquad \mathscr{M}_w(c,y;w,\theta) < 0 \quad \Leftrightarrow \quad v_{yw}(y;w,\theta) < 0. \tag{3}$$

We also impose that the marginal rate of substitution decreases from plus infinity to zero. This is a kind of INADA condition that will appear technically practical. For the ease of the presentation, we give the following definitions:

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

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

A smoothly increasing (decreasing) function is not only strictly increasing (decreasing), it is also differentiable with a strictly positive (negative) derivative everywhere and its derivative maps the positive real line onto itself. It is straightforward to verify that the combination of two smoothly increasing (decreasing) functions is a smoothly increasing function and that the reciprocal of a smoothly increasing (decreasing) function is a smoothly increasing (decreasing) function.[8] We therefore assume that:

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 smoothly decreasing in skill.*

This assumption is automatically verified in the case where (1) can be rewritten as:

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

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

[8]What we call a smoothly increasing (decreasing) function is also called an increasing (decreasing) diffeomorphism for which the derivative maps the positive real line onto itself.

Under preferences (1), an individual of type (w, θ), facing the nonlinear income tax $y \mapsto T(y)$, solves:

$$\max_{y} \quad \mathscr{U}\left(y - T(y), y; w, \theta\right) \tag{5}$$

We call $Y(w, \theta)$ the solution to program (5),[9] $C(w, \theta) = Y(w, \theta) - T(Y(w, \theta))$ the consumption of an individual of type (w, θ) and $U(w, \theta) = u(C(w, \theta)) - v\left(Y(w, \theta); w, \theta\right)$ her utility. When the tax function is differentiable, the first-order condition associated to (5) implies with (2) that:

$$1 - T'\left(Y\left(w, \theta\right)\right) = \mathscr{M}\left(C\left(w, \theta\right), Y\left(w, \theta\right); w, \theta\right) \tag{6}$$

II.2 The government

The government's budget constraint takes the form:

$$\int_{\theta \in \Theta} \left\{ \int_{0}^{+\infty} \left[Y(w, \theta) - C(w, \theta)\right] f(w|\theta) dw \right\} d\mu(\theta) \geq E \tag{7}$$

where $E \geq 0$ is an exogenous amount of public expenditures. Turning now to the government's objective function, we adopt a general *welfarist* criterion that sums over all types of individuals an increasing and weakly concave transformation $\Phi(U; w, \theta)$ of individuals' utility level U. The government's objective is:

$$\int_{\theta \in \Theta} \left\{ \int_{0}^{+\infty} \Phi\left(U(w, \theta); w, \theta\right)\ f(w|\theta) dw \right\} d\mu(\theta). \tag{8}$$

This welfarist specification allows Φ to vary with individual types (w, θ) which makes it very general. It encompasses the case of *weighted utilitarian* preferences with type-specific weights denoted $\varphi(w, \theta)$, so that $\Phi(U; w, \theta)) \equiv \varphi(w, \theta) \cdot U$. The social objective is then:

$$\int_{\theta \in \Theta} \left\{ \int_{0}^{+\infty} \varphi\left(w, \theta\right)\ U(w, \theta)\ f(w|\theta) dw \right\} d\mu(\theta). \tag{9a}$$

As particular cases, the latter 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 $\varphi(0, \theta) > 0$ while $\varphi(w, \theta) = 0\ \forall w > 0$. Our welfarist criterion also encompasses the Bergson-Samuelson criterion which is a concave transformation of utility that does not depend on individuals' type (w, θ), i.e. $\Phi(U; w, \theta)$ does not vary with its two last arguments. The Bergson-Samuelson criterion takes the form:

$$\int_{\theta \in \Theta} \left\{ \int_{0}^{+\infty} \Phi\left(U(w, \theta)\right)\ f(w|\theta) dw \right\} d\mu(\theta). \tag{9b}$$

Let $\lambda > 0$ be the Lagrange multiplier associated with the budget constraint (7) which can be interpreted as the shadow price of government's funds. Following Saez (2001), we define the marginal social welfare weight associated with workers of type (w, θ) by:

$$g(w, \theta) = \frac{u'(C(w, \theta))\ \Phi_U\left(U(w, \theta); w, \theta\right)}{\lambda}. \tag{10}$$

[9] If the maximization program (5) admits multiple solutions, we make the tie-breaking assumption that individuals choose among their best options the income level preferred by the government, i.e. the one with the largest tax liability.

The government values giving one extra dollar to a worker (w, θ) as a gain of $g(w, \theta)$ in terms of public funds.[10]

II.3 Possible applications

The framework presented in the previous two subsections is general enough to encompass a large set of applications. We now present several of these applications. Although our approach extends beyond tax theory[11], all the applications in this subsection concern optimal income taxation problems. In each case, we explain what y, w, θ represent so that the interpretation of the results of our general framework is straightforward. All proofs and propositions in the following sections are valid in each case, they are simply to be interpreted in a different way. Importantly, we state that Assumption 1 holds in each application, at the cost of at most very mild assumptions. The reader interested in the core model but not in its various applications can skip this subsection.

Optimal income taxation with heterogeneous skills and labor supply elasticities

In this case, all the tax schedules that we obtain in the paper are to be interpreted with y as the (pre-tax) labor income and, using isoelastic individual's preferences (4), with the following two dimensions of heterogeneity: skill w and labor supply elasticity θ. As previously mentioned, Assumption 1 is then automatically satisfied.

Optimal joint taxation of labor and non-labor income

In this case, individuals have two sources of taxable income: a non-labor income z and a labor income $y - z$. Those incomes are jointly taxed and the tax function does not distinguish between both incomes.[12] In this case, y is the total taxable income and we interpret θ as the ability to earn non-labor income z and w as the skill.

For an individual of skill w who belongs to group θ, let $V(y - z, z; w, \theta)$ be the joint disutility of earning $y - z$ and z, with $V_{y-z}, V_z > 0$. Individuals of type (w, θ) then solve:

$$\max_{y,z} \quad u(y - T(y)) - V(y - z, z; w, \theta).$$

where two decision variables appear instead of one variable in program (5). This program can be solved sequentially, the first step being the choice of non-labor income z for a given taxable income y. The disutility function in (1) is then retrieved by defining:

$$v(y; w, \theta) \stackrel{\text{def}}{\equiv} \min_{z} \quad V(y - z, z; w, \theta). \tag{11}$$

[10] We can easily extend our analysis to *non-welfarist* social criteria following the method of generalized marginal social welfare weights developed in Saez and Stantcheva (2016) to reflect non-welfarist views of justice.

[11] To illustrate this, we apply our model, in an appendix available upon request, to a nonlinear pricing model where a monopolist (the principal) observes a one-dimensional action (how much consumers are demanding of the single commodity it sells), and where the unobserved characteristics of the consumers (the agents) are multidimensional.

[12] This applies, for instance, in countries like France where entrepreneurial income and income received from renting property are jointly taxed with labor income.

Here, Assumption 1 is satisfied whenever the second-order derivatives of $V(\cdot)$ are such that $v_{yw} < 0$.[13] This inequality holds when z is exogenous. For instance, $z = \theta$ when θ are rents perceived by landlords who have inherited the property they rent. When z is endogenous, this inequality holds when $V(y-z, z; w, \theta) = V^{\ell}(y-z; w, \theta) + V^{z}(z; \theta)$ with $V^{\ell}(\cdot; w, \theta)$ and $V^{z}(\cdot; \theta)$ increasing and convex, and $V^{\ell}_{yw} < 0$.[14]

Optimal joint income taxation of couples

The joint income taxation of couples is a variant of the previous application, in which $y - z$ is the labor income of one individual and z is the one of his/her partner. The tax does not distinguish between $y - z$ and z and only depends on the sum of both incomes, y (as in France, Germany and the US). We redefine w and θ as the respective skill level of each member of the couple. The optimal tax schedules derived in this paper are then interpreted as the optimal tax schedules when the couple is the tax unit and each partner decides along the intensive margin. So far, previous attempts in the literature (Kleven et al. (2007) and Cremer et al. (2012)) have stopped short of obtaining these nonlinear tax schedules.

Optimal income taxation with tax avoidance

In this application, w is the skill and θ is the ability to avoid taxation. We assume that tax enforcement (penalty, monitoring, etc.) is given. We denote z the sheltered labor income (i.e. income that is not taxed at all) and $y + z$ the (total) labor income. The tax only depends on the taxable income y. Consumption becomes $c + z$, with $c = y - T(y)$ being the after-tax income.

Preferences (1) now become quasi-linear in consumption, $c + z - V(y + z, z; w, \theta)$, where $V_{y+z}, V_{y+z\,y+z} > 0$ and $V_z, V_{zz} > 0$.[15] The disutility function in (1) is then retrieved by defining:

$$v(y; w, \theta) \overset{\text{def}}{\equiv} \min_{z} \quad V(y + z, z; w, \theta) - z$$

assuming that the second-order derivatives of $V(\cdot)$ are such that $v_{yw}(y; w, \theta) < 0$ to ensure that Assumption 1 holds. We will get back to this application when deriving the optimal tax profile on US data, in Section V. In that section, θ will denote the taxable income elasticity, which depends on the individual ability to avoid taxation.

III A structural approach to the tax formula

This section studies the design of the optimal second-best allocation, while the next one will focus on the relation between the optimal tax schedule and empirically meaningful statistics.

[13] The envelope theorem induces that $v_y = V_{y-z}$ and $v_w = V_w$. Hence, one obtains $v_y > 0 > v_w$, whenever $V_{y-z} > 0 > V_w$, which are natural assumptions.

[14] We note $z^*(y; w, \theta)$ the solution to (11). Differentiating the first-order condition $V^{\ell}_y = V^z_z$ leads to $\partial z^*/\partial w = V^{\ell}_{yw}/(V^{\ell}_{yy} + V^z_{zz})$, which is negative by the convexity of $V^{\ell}(\cdot; w, \theta)$ and of $V^z(\cdot; \theta)$ and by $V^{\ell}_{yw} < 0$. Therefore, as $v_y(y; w, \theta) = V^z_z(z^*(w, \theta), \theta)$ from the envelope theorem and first-order condition, the convexity of $V^z(\cdot; \theta)$ induces that $v_{yw} < 0$.

[15] For a given labor income, increasing the amount of sheltered income is costly (i.e., requires more effort). This is a standard assumption in papers that incorporate avoidance effects for optimal tax design, see Piketty and Saez (2013, Section 4.3.).

III.1 Incentive-Compatible Allocations

In this subsection, we develop a method to characterizes incentive-compatible allocations when unobserved characteristics (w, θ) are multidimensional. We start by stating the incentive constraints. Since the individual's objective (5) is maximized for $y = Y(w, \theta)$, we have:

$$\forall (w, \theta, \tilde{y}) \in \mathbb{R}_+ \times \Theta \times \mathbb{R}_+ \quad \mathscr{U}\left(C(w,\theta), Y(w,\theta); w, \theta\right) \geq \mathscr{U}\left(\tilde{y} - T(\tilde{y}), \tilde{y}; w, \theta\right).$$

Taking $\tilde{y} = Y(\tilde{w}, \tilde{\theta})$ leads to the following set of incentive constraints:

$$\forall (w, \tilde{w}, \theta, \tilde{\theta}) \in \mathbb{R}_+^2 \times \Theta^2 \quad \mathscr{U}\left(C(w,\theta), Y(w,\theta); w, \theta\right) \geq \mathscr{U}\left(C(\tilde{w},\tilde{\theta}), Y(\tilde{w},\tilde{\theta}); w, \theta\right). \tag{12}$$

Equation (12) states that individuals of type (w, θ) prefer the bundle $(C(w, \theta), Y(w, \theta))$ they have chosen to any other bundle $(C(\tilde{w}, \tilde{\theta}), Y(\tilde{w}, \tilde{\theta}))$ intended for any other type $(\tilde{w}, \tilde{\theta})$ of workers. The usual taxation principle (Hammond, 1979, Guesnerie, 1995) holds. For the government, it is equivalent to choose a non-linear income tax, taking individual choices (5) into account or to directly select an allocation satisfying the incentive-compatible constraints (12). We follow the second approach in this section and characterize first the set of incentive-compatible allocations.

Within-Group Incentive Constraints

An incentive-compatible allocation has to satisfy (12). It thus has to verify for each group θ the following set of "within-group incentive constraints":

$$\forall (w, \tilde{w}, \theta) \in \mathbb{R}_+^2 \times \Theta \quad \mathscr{U}\left(C(w,\theta), Y(w,\theta); w, \theta\right) \geq \mathscr{U}\left(C(\tilde{w},\theta), Y(\tilde{w},\theta); w, \theta\right). \tag{13}$$

For each θ, characterizing the within-group allocations $w \mapsto (C(w, \theta), Y(w, \theta))$ that verify the within-group incentive constraints (13) is the same problem as characterizing incentive-compatible allocations when unobserved heterogeneity is one-dimensional. This is due to within-group single-crossing Assumption 1. Under this assumption, the set of incentive constraints can be transformed into a monotonicity constraint and a differential equation that we retrieve in Lemmas 1 and 2 below.

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

Appendix A.1 provides the proof. Note that $Y(\cdot; \theta)$ being nondecreasing, it may exhibit discontinuities over a countable set and it may also exhibit bunching where individuals in the same group but endowed with different skill levels earn the same income. It is however standard, in one-dimensional models, to consider only smooth allocations where these two pathologies do not arise and to follow the so-called "first-order approach" , e.g. Salanié (2011). We thus make the following smoothness assumption:

Assumption 2 (Smooth allocations)**.** *In each group θ, $w \mapsto Y(w, \theta)$ is smoothly increasing.*

According to Assumption 2, for each income level $y \in \mathbb{R}^+$ and for each group $\theta \in \Theta$, there exists a single skill level w such that only individuals of that skill level within group θ earn income $y = Y(w, \theta)$. In the present paper, we define *pooling* as the situation where the same income level is earned by individuals in different groups. As soon as one studies multidimensional heterogeneity in models with a single observable action, pooling cannot be neglected. Note that situations where people who earn the same income are identical in all their dimensions of heterogeneity except their skills are undoubtedly very rare empirically. Assumption 2 rules out the latter type of situations[16], while it implies that pooling is unavoidable. Assumption 2 holds, for instance, under the isoelastic individual preferences (4) when the income tax function is linear. Consequently, by continuity, as soon as the marginal tax rates do not vary too much, Assumption 2 holds. The following lemma provides the differential equation or first-order incentive constraint.[17]

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

$$\dot{U}(w, \theta) = \mathscr{U}_w(C(w, \theta), Y(w, \theta); w, \theta) = -v_w\left(Y(w, \theta); w, \theta\right). \tag{14a}$$

Moreover, Equation (14a) is equivalent to:

$$\frac{\dot{C}(w, \theta)}{\dot{Y}(w, \theta)} = \mathscr{M}\left(C(w, \theta), Y(w, \theta); w, \theta\right). \tag{14b}$$

Equation (14a) is the first-order incentive-compatible equation within group θ. The usual proof in the one dimensional context is adapted in Appendix A.2. Integrating (14a) leads to:

$$U(w, \theta) = U(0, \theta) - \int_0^w v_w\left(Y(x, \theta); x, \theta\right)\, dx. \tag{14c}$$

If the government were able to observe the group θ to which each taxpayer belongs to, the government would propose group-specific income tax schedules $T(\cdot; \theta)$. We would then only need to take into account the within-group incentive constraints (13).[18] The observation of θ would improve the possibility for the government to redistribute income as highlighted in the so-called *tagging* literature (see e.g., Akerlof (1978), Boadway and Pestieau (2006), Cremer et al. (2010), Mankiw and Weinzierl (2010)). In contrast, our paper does not consider tagging so that the government does not condition taxes on the group index θ. We thus need to describe how the various within-group allocations $\omega \mapsto (Y(\omega, \theta), C(\omega, \theta))$ need to be set to be mutually incentive-compatible and to verify the full set of incentive constraints (12). This is the *pooling issue* that we now address.

[16]In the vein of the model with one dimension of heterogeneity, this specif situation where individuals in the same group but with different skill levels earn the same income can be called bunching. This type of situation never appear in all our simulations (see Section V).

[17]We use a dot to denote the derivative with respect to w for a fixed θ.

[18]To be more precise, this remark holds only if the government was furthermore *allowed* to condition taxation on θ. For instance, despite the fact that the government can *observe* whether a taxpayer is a woman or a man, gender-based taxation (Alesina et al., 2011) is in practice ruled out for horizontal equity reasons. A similar issue arises when conditioning income taxation on individuals' height (Mankiw and Weinzierl, 2010).

Pooling Types across θ-Groups at each Income Level

Chose a reference group $\theta_0 \in \Theta$, a skill level w and another group θ. Individuals of type (w, θ_0) earn income $Y(w, \theta_0)$. According to the smoothness Assumption 2, each group-specific allocation $Y(\cdot, \theta) : w \mapsto Y(w, \theta)$ is an increasing one-to-one function that maps the positive real line onto itself. Therefore, there must exist a single skill level, hereafter denoted $W(w, \theta)$, so that individuals of the other group θ endowed with that skill level $W(w, \theta)$ must get the same income level $Y(w, \theta_0)$ as individuals of type (w, θ_0), i.e. $Y(W(w, \theta), \theta) = Y(w, \theta_0)$. We call $W(.,.)$ the *pooling* function. For each $\theta \in \Theta$, the pooling function combines two smoothly increasing functions, namely $\omega \overset{Y(\cdot,\theta_0)}{\longmapsto} Y(\omega, \theta_0) \overset{Y^{-1}(\cdot,\theta)}{\longmapsto} W(\omega, \theta)$. The pooling function is therefore also a smoothly increasing function in skill w. It obviously verifies $W(w, \theta_0) \equiv w$. Provided that the allocation is incentive-compatible, it is not possible from (12) that individuals of type $(W(w, \theta), \theta)$ and individuals of type (w, θ_0) obtain the same income $Y(w, \theta_0)$ but distinct consumption levels. Therefore, for each (w, θ), we must simultaneously have:

$$Y(W(w,\theta),\theta) \equiv Y(w,\theta_0) \qquad \text{and} \qquad C(W(w,\theta),\theta) \equiv C(w,\theta_0). \tag{15}$$

One can retrieve the entire incentive-compatible allocation for all groups if one knows the pooling function $W(\cdot, \cdot)$ and the allocation $\omega \mapsto (Y(\omega, \theta_0), C(\omega, \theta_0))$ designed for the reference group. Furthermore, to determine the pooling function, one only needs the allocation designed for the reference group, as explained by the following lemma.

Lemma 3. *Under Assumptions 1 and 2, along an incentive-compatible allocation, the bundle designed for individuals of type $(W(w, \theta), \theta)$ coincides with the bundle $(C(w, \theta_0), Y(w, \theta_0))$ designed for individuals of type (w, θ_0), where $W(w, \theta)$ verifies the following **pooling condition**:*

$$\mathscr{M}\left(C(w,\theta_0), Y(w,\theta_0); w, \theta_0\right) = \mathscr{M}\left(C(w,\theta_0), Y(w,\theta_0); W(w,\theta), \theta\right). \tag{16}$$

Proof According to Assumption 1, $\mathscr{M}\left(C(w,\theta_0), Y(w,\theta_0); w, \theta_0\right) = \mathscr{M}\left(C(w,\theta_0), Y(w,\theta_0); \omega, \theta\right)$ admits exactly one solution in ω. Differentiating in w both sides of each two equalities in (15) leads to:

$$\dot{Y}(W(w,\theta),\theta)\, \dot{W}(w,\theta) = \dot{Y}(w,\theta_0) \qquad \text{and} \qquad \dot{C}(W(w,\theta),\theta)\, \dot{W}(w,\theta) = \dot{C}(w,\theta_0)$$

where $\dot{W}(w, \theta)$ denotes the partial derivative of W with respect to the skill level. Hence,

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

According to Lemma 2, Equation (14b) holds, which implies (16). □

Intuitively, if individuals of type (w, θ_0) and of type $(W(w, \theta), \theta)$ choose the same income $Y(w, \theta_0)$, they must face the same marginal tax rate $T'(Y(w, \theta_0))$. Hence, from the first-order condition (6), they must face the same marginal rate of substitution, as highlighted in (16). A key point here is that, because of the within-group single-crossing condition (Assumption 1), the equation $\mathscr{M}\left(C(w,\theta_0), Y(w,\theta_0); w, \theta_0\right) = \mathscr{M}\left(C(w,\theta_0), Y(w,\theta_0); \omega, \theta\right)$ admits exactly one solution in ω. Hence, the previous equation fully characterizes the pooling function $W(\cdot,\theta)$ from the allocation $\omega \mapsto (C(\omega,\theta), Y(\omega,\theta))$ specific to the reference group θ_0. The following lemma, which is proved in Appendix A.3, provides a sufficient condition for the allocation to be incentive-compatible.

Lemma 4. *Let $w \mapsto (C(w,\theta_0), Y(w,\theta_0))$ be a within-group allocation that verifies Assumption 2 and the within-group incentive-compatible Equation (14b). For each $w \in \mathbb{R}_+$ and each group $\theta \in \Theta$, let $\underline{W}(w,\theta)$ be the unique skill level ω that solves the pooling condition $\mathscr{M}\left(C(w,\theta_0), Y(w,\theta_0); w,\theta_0\right) = \mathscr{M}\left(C(w,\theta_0), Y(w,\theta_0); \omega,\theta\right)$. There exists a unique incentive-compatible allocation $(w,\theta) \mapsto (\underline{C}(w,\theta), \underline{Y}(w,\theta))$ the restriction of which to group θ_0 is $w \mapsto (\underline{C}(w,\theta_0), \underline{Y}(w,\theta_0))$ and it verifies Assumption 2 if and only if, for each θ, $w \mapsto \underline{W}(w,\theta)$ is smoothly increasing.*

Lemma 4 guarantees that if $Y(w,\theta_0)$ is smoothly increasing in w and if, for each θ, the pooling function denoted $\underline{W}(w,\theta)$ is also smoothly increasing in w, then the allocation is incentive-compatible. Assumption 2 together with the assumption that $\underline{W}(\cdot,\theta)$ is smoothly increasing plays, in our analysis, a role similar to the monotonicity or second-order incentive-compatibility condition of the Mirrleesian "first-order approach" with one dimension of unobserved heterogeneity.

In what follows, we therefore select the allocation only for the reference group θ_0 and assume that the triggered allocations for the other groups verify Assumption 2. Using Equation (2), the pooling condition (16) can be rewritten as:

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

which can be simplified as:

$$v_y\left(Y(w,\theta_0); w,\theta_0\right) = v_y\left(Y(w,\theta_0); W(w,\theta),\theta\right). \tag{17}$$

Therefore, the pooling function $W(\cdot,\theta)$ that enables to retrieve $(C(\cdot,\theta), Y(\cdot,\theta))$ from the allocation of the reference group $(C(\cdot,\theta_0), Y(\cdot,\theta_0))$ depends on $Y(\cdot,\theta)$. This endogeneity of the pooling function is a major difference with the previous literature as we will discuss in the next subsection.[19]

Consider, as an illustration, the case where individual preferences are isoelastic as in (4). The equality in Equation (17) implies that the pooling function is:[20]

$$W(w,\theta) = \left(w^{\frac{1+\theta_0}{\theta_0}} \cdot \left(Y(w,\theta_0)\right)^{\frac{1}{\theta}-\frac{1}{\theta_0}}\right)^{\frac{\theta}{1+\theta}}.$$

The pooling function thus depends on the choice of $Y(\cdot,\theta_0)$ whenever the different groups are endowed with distinct labor supply elasticities (i.e. when $\theta \neq \theta_0$). Note that if one takes for the reference group $\theta_0 = \max\{\theta \in \Theta\}$, it is then sufficient to impose that $w \mapsto Y(w,\theta_0)$ is smoothly

[19]Moreover, the pooling function does not depend on $C(\cdot,\theta)$, a simplification that relies on the assumption that the utility function (1) is additively separable."

[20]Substituting (4) in (17) yields $Y(w,\theta_0)^{1/\theta_0} w^{-(1+\theta_0)/\theta_0} = Y(w,\theta_0)^{1/\theta} W(w,\theta)^{-(1+\theta)/\theta}$.

increasing to ensure that the pooling function $w \mapsto W(., \theta)$ is smoothly increasing as well for each θ.[21] Lemma 4 is then valid and Assumption 2 is verified.

Related literature

In many previous tax models with multidimensional unobserved heterogeneity, the decisions along the intensive margin are assumed to depend only on a one-dimensional aggregation of characteristics. This implies the counter-factual prediction that all individuals earning the same income level exhibit identical behavioral elasticities. To clarify this point, let **t** denote the vector of unobserved characteristics and assume that intensive decisions depend only on a one-dimensional aggregator denoted $w = \Xi(\mathbf{t})$, so that individuals of type **t** have preferences $\mathscr{U}(c, y; \Xi(\mathbf{t}))$ over consumption and income and solve $\max_y \mathscr{U}(y - T(y), y; \Xi(\mathbf{t}))$. All individuals with the same $w = \Xi(\mathbf{t})$ are thus facing the same decision program. They are therefore making the same intensive decisions and are equally responsive to tax reforms. Moreover, the pooling function is simply obtained by inverting the aggregator $\Xi(\cdot)$. Therefore, it does not depend on the chosen variables $Y(.,.)$ and $C(.,.)$. Therefore, if one wants a model where individuals differ also in their behavioral responses, the pooling function must depend on the allocation.

Brett and Weymark (2003), Boadway et al. (2002), Choné and Laroque (2010), Lockwood and Weinzierl (2015) explicitly assume that labor supply decisions depend only on an exogenous unidimensional combination $w = \Xi(\mathbf{t})$ of two unobserved characteristics **t**. Therefore, two individuals who earn the same income cannot have distinct labor supply elasticities despite their distinct characteristics. The additional heterogeneity only matters for the computation of social marginal weights.

Rothschild and Scheuer (2013, 2016, 2014), Scheuer (2013, 2014) and Gomes et al. (2014) study optimal income taxes with several sectors. In their models, individuals need to choose how to split their labor effort between different sectors. The productivity of individuals in each sector composes the vector of unobserved characteristics **t**. The private and social returns of labor effort in each sector are functions of the aggregate amount of labor in each sector, thereby allowing for rich patterns of technological complementaries and externalities between these sectors. However, individuals' preferences are specified in such a way that once the individual allocation of effort across sectors is chosen, the total amount of effort of an individual of characteristics **t** depends only on a one-dimensional aggregation $\Xi(\mathbf{t}; \mathbf{p})$ of types **t** and of prices **p**, i.e. private returns of effort in each sector. Hence, individuals who earn the same income cannot have distinct skills, thereby distinct labor supply elasticities.

In random participation models with endogenous participation (Rochet and Stole, 2002, Kleven et al., 2009, Jacquet et al., 2013) or in optimal income tax models with migration (Blumkin et al., 2014, Lehmann et al., 2014), individuals differ in skills and in costs of participation (mi-

[21] As $\theta_0 \geq \theta$, one has that $\frac{1}{\theta} > \frac{1}{\theta_0}$ so $(Y(w, \theta_0))^{\frac{1}{\theta} > \frac{1}{\theta_0}}$ is smoothly increasing in skill w for each group θ if and only if $Y(\cdot, \theta_0)$ is smoothly increasing. As $\theta, \theta_0 > 0$, $W(w, \theta)$ is therefore also smoothly increasing in skill w for each group θ.

gration). The cost of participation (migration) drives the individual participation (migration) decision while the level of skill determines the intensive labor supply decision. Therefore, people with an identical skill level earn the same income, whatever their participation (or migration) costs. The aggregator is then reduced to $w = \Xi(w,\theta)$ and again, workers earning the same income are constrained to react identically to any tax reform.

III.2 Optimal structural tax formula

We now derive the optimal marginal tax rates as a function of the policy-invariant primitives or structural parameters of the model, which are the individual utility function $\mathscr{U}(\cdot,\cdot;w,\theta)$, the government's objective function $\Phi(\cdot;w,\theta)$ and the distributions of characteristics $f(\cdot|\theta)$ and $\mu(\cdot)$. Like in the model with one dimension of heterogeneity (see e.g., Saez (2001)), obtaining such a structural tax formula is crucial if one wants to implement the model with data.[22] The government's problem consists in finding the incentive-compatible allocation that maximizes the social objective (8) under the budget constraint (7).

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

$$\mathscr{C}_u(\hat{u},y;w,\theta) = \frac{1}{u'(c)} \qquad \text{and} \qquad \mathscr{C}_y(\hat{u},y;w,\theta) = \frac{v_y(y;w,\theta)}{u'(c)} \tag{18}$$

where the various derivatives are evaluated at $c = \mathscr{C}(\hat{u},y;w,\theta)$. Let λ denote the multiplier associated to the budget constraint (7), the Lagrangian $\mathscr{L}$ of the government's problem is:

$$\mathscr{L} \stackrel{\text{def}}{\equiv} \iint \left[Y(w,\theta) - \mathscr{C}\left(U(w,\theta), Y(w,\theta); w,\theta\right) + \frac{\Phi\left(U(w,\theta); w,\theta\right)}{\lambda} \right] f(w|\theta) dw \, d\mu(\theta). \tag{19}$$

The government's problem consists in finding the best allocation that verifies the incentive constraints (12). Following the usual first-order approach, we consider a "relaxed" problem where the government maximizes over the set of allocations that verify the first-order incentive constraint (14a) for each group and the pooling condition (17).[23] According to Lemma 4, whenever the solution to this relaxed problem verifies Assumption 2 and the implied pooling function is, for each group θ, smoothly increasing in skill w, it also solves the problem with all incentive-compatible constraints.

When the unobserved heterogeneity is one-dimensional, the usual method to derive the necessary conditions is to construct a Hamiltonian and to apply the Pontryagin principle. In our multidimensional environment, the pooling condition (17) induces constraints on state and control variables which hold at endogenous skill levels. This is the reason why we use the

[22] According to (Saez, 2001, p. 223), the optimal tax formula derived by Saez (2001) "cannot be directly applied using empirical income distribution because the income distribution is affected by taxation. Therefore, it is useful to come back to the Mirrlees formulation and use an exogenous skill distribution to perform numerical simulations."

[23] To solve one-dimensional tax models, one usually assume that income is increasing with skill. This is called the *first-order approach*. Therefore, solving the multidimensional tax model under Assumption 2, as done with our allocation perturbation method, is (somewhat) also a first-order approach.

calculus of variation and consider a set of perturbations of the allocation in the reference group. The cornerstone of our method is the pooling condition (17) that we use to deduce how the allocation in the other groups are perturbed. Thanks to this condition, we can thus compute the Gâteaux derivatives of the Lagrangian (19) in the direction of these perturbations. Equating these Gâteaux derivatives to zero lead to an optimal structural tax formula which gives the optimal marginal tax rates as a function of the primitives of the model. To save on notations, we from now on use the more compact notation $\langle w,\theta\rangle$ when the various functions are evaluated for types (w,θ) at income $Y(w,\theta)$, utility $U(w,\theta)$ and consumption $C(w,\theta)$. We then get:

Proposition 1. *Under Assumptions 1 and 2, the optimal structural tax formula verifies:*

$$\begin{aligned} & \frac{T'\left(Y(w,\theta_0)\right)}{1-T'\left(Y(w,\theta_0)\right)} \int\limits_{\theta\in\Theta} \frac{v_y\left\langle W(w,\theta),\theta\right\rangle}{-W(w,\theta)\, v_{yw}\left\langle W(w,\theta),\theta\right\rangle}\, W(w,\theta)\, f(W(w,\theta)|\theta)\, d\mu(\theta) \\ = \;& u'\left(C(w,\theta_0)\right) \iint\limits_{\theta\in\Theta, x\geq W(w,\theta)} \left(\frac{1}{u'(C(x,\theta))} - \frac{\Phi_U(U(x,\theta);x,\theta)}{\lambda}\right) f(x|\theta) dx\, d\mu(\theta) \qquad \text{(20a)} \end{aligned}$$

for all $w\in\mathbb{R}_+$ and:

$$\iint\limits_{\theta\in\Theta, x\in\mathbb{R}_+} \left(\frac{\Phi_U(U(x,\theta);x,\theta)}{\lambda} - \frac{1}{u'(C(x,\theta))}\right) f(x|\theta) dx\, d\mu(\theta) = 0. \qquad \text{(20b)}$$

Proof To derive (20b), we consider a set of allocation perturbations indexed by $\Delta\in\mathbb{R}$ and denoted $(\tilde{C}(w,\theta;\Delta),\tilde{Y}(w,\theta;\Delta),\tilde{U}(w,\theta;\Delta) \stackrel{\text{def}}{\equiv} \mathscr{U}\left(\tilde{C}(w,\theta;\Delta),\tilde{Y}(w,\theta;\Delta);w,\theta\right))$, which consist, for each type $(x,\theta)\in\mathbb{R}_+\times\Theta$, in no change in $Y(x,\theta)$ and in a uniform change in $U(x,\theta)$, therefore in $u(C(x,\theta))$ by an amount Δ. Hence, we get for each Δ that $\tilde{U}(w,\theta;\Delta) \stackrel{\text{def}}{\equiv} U(w,\theta)+\Delta$, $\tilde{Y}(w,\theta;\Delta) \stackrel{\text{def}}{\equiv} Y(w,\theta)$ and $\tilde{C}(w,\theta;\Delta) \stackrel{\text{def}}{\equiv} \mathscr{C}\left(\tilde{U}(w,\theta;\Delta),\tilde{Y}(w,\theta;\Delta);w,\theta\right)$. These perturbations preserve incentive-compatibility (12). According to (19), the perturbed Lagrangian is:

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

If the allocation is optimal, the above perturbations do not affect the Lagrangian. Thus, by equating the Gâteaux derivative of the Lagrangian in the direction described by the above perturbations, i.e. the derivative of the perturbed Lagrangian $\mathscr{L}(\cdot)$ with respect to Δ, at $\Delta=0$, to zero, we obtain an equation that characterizes the optimal tax system. Using the first equality in (18), this Gâteaux derivative of the Lagrangian is:

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

Equating this derivative to zero leads to (20b).

To derive (20a) at a given skill level w, we consider a set of allocation perturbations, indexed by $t\in\mathbb{R}$ and $\delta\in\mathbb{R}_+$, that we denote $\hat{C}(w,\theta;t,\delta)$,$\hat{Y}(w,\theta;t,\delta)$ and $\hat{U}(w,\theta;t,\delta) \stackrel{\text{def}}{\equiv} \mathscr{U}\left(\hat{C}(w,\theta;t,\delta),\hat{Y}(w,\theta;t,\delta);w,\theta\right))$ where t stands for the size of the perturbation, and δ is the length of the skill interval where, in the reference group, the perturbation of incomes takes place. Following Lemma 4, we define the allocation perturbations from their restriction to the reference group θ_0 and then study the impact of these perturbations on the allocation in every other group. The perturbations of incomes in the reference group are defined by:

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

where $\Delta_Y(\cdot,\theta_0;\delta)$ is a continuously differentiable function defined on $\mathbb{R}^+$ such that $\Delta_Y(\cdot,\theta_0;\delta) > 0$ for $x \in (w-\delta,w)$ and is nil otherwise. Incomes in the reference group remain unchanged outside the skill interval $(w-\delta,w)$ and are smoothly increased (decreased) inside the skill interval $(w-\delta,w)$ if $t>0$ (if $t<0$), as illustrated in Figure 1. It is worth noting that the perturbed income function remains differentiable with respect to skill w since $\Delta_Y(\cdot,\cdot,\delta)$ is differentiable. Moreover, from Assumption 2, $Y(\cdot,\theta_0)$ admits a positive derivative everywhere, so $\dot{Y}(\cdot,\theta_0)$ is bounded away from 0 for all $x \in [w-\delta,w]$. Therefore, provided that t is small enough, which we assume in the rest of the proof, $\hat{Y}(\cdot,\theta_0;t,\delta)$ has also a positive derivative everywhere and therefore verifies Assumption 2.

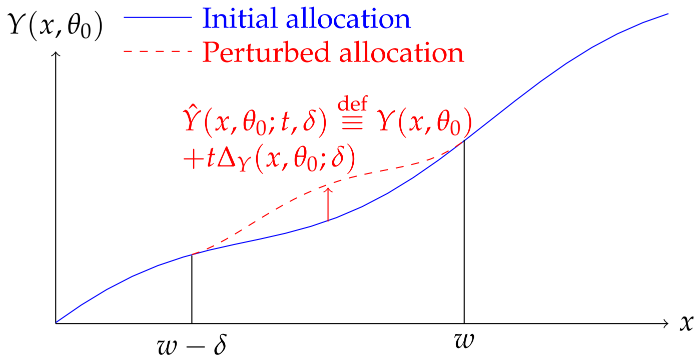

Figure 1: The perturbation of incomes in the reference group θ_0

Let us in addition assume that the utility of the lowest skilled individuals in the reference group $U(0,\theta_0;t,\delta)$ is not perturbed and write it as $U(0,\theta_0)$. Therefore, according to the first-order incentive constraint (14c), the perturbed utility function in the reference group is:

$$\hat{U}(x,\theta_0;t,\delta) \stackrel{\text{def}}{\equiv} U(0,\theta_0) - \int_0^x v_w\left(\hat{Y}(\omega,\theta_0;t,\delta);\omega,\theta_0\right)d\omega. \tag{21a}$$

From the pooling condition (17), as incomes $Y(\cdot,\theta_0;t,\delta)$ in the reference group remain unchanged outside the skill interval $(w-\delta,w)$, the pooling function $W(\cdot,\theta_0;t,\delta)$ is not perturbed outside the skill interval $(w-\delta,w)$. Therefore, incomes $Y(\cdot,\theta;t,\delta)$ in any group θ are not modified outside the skill interval $(W(w-\delta,\theta),W(w,\theta))$, and we must have (See Figure 2):

$$\hat{Y}(x,\theta;t,\delta) = Y(x,\theta) \qquad \text{if} \qquad x \in [0,W(w-\delta,\theta)] \cup [W(w,\theta),+\infty)\,. \tag{21b}$$

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

$$\hat{C}(x,\theta;t,\delta) = C(x,\theta) \qquad \text{and} \qquad \hat{U}(x,\theta;t,\delta) = U(x,\theta). \tag{21c}$$

For all skills $x > W(w-\delta,\theta)$, the change in utility obtained using the first-order incentive constraint (14c) is:

$$\hat{U}(x,\theta;t,\delta) - U(x,\theta) \quad = \quad -\int_0^x \left[v_w\left(\hat{Y}(\omega,\theta;t,\delta);\omega,\theta\right) - v_w\left(Y(\omega,\theta);\omega,\theta\right)\right]d\omega. \tag{21d}$$

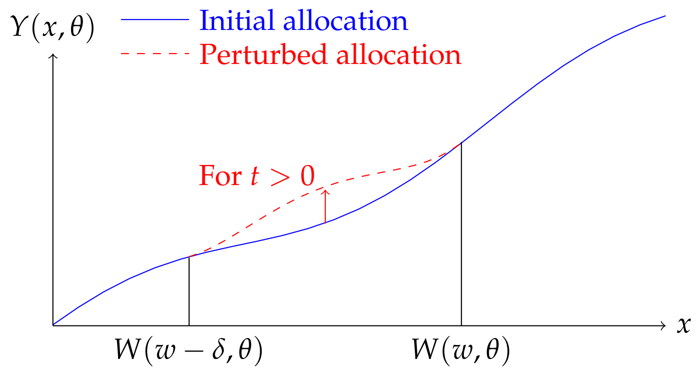

Figure 2: The perturbation of incomes in the other groups

Since incomes $\hat{Y}(\cdot,\theta;t,\delta)$ are only perturbed inside $(W(w-\delta,\theta)), W(w,\theta))$, for all skills x that belong to this interval, using (21b), we get:

$$\hat{U}(x,\theta;t,\delta) - U(x,\theta) \quad = \quad \int_{W(w-\delta,\theta)}^{x} \left[v_w\left(Y(\omega,\theta);\omega,\theta\right) - v_w\left(\hat{Y}(\omega,\theta;t,\delta);\omega,\theta\right)\right] d\omega. \tag{21e}$$

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

$$\hat{U}(x,\theta;t,\delta) - U(x,\theta) \quad = \quad \int_{W(w-\delta,\theta)}^{W(w,\theta)} \left[v_w\left(Y(\omega,\theta);\omega,\theta\right) - v_w\left(\hat{Y}(\omega,\theta;t,\delta);\omega,\theta\right)\right] d\omega. \tag{21f}$$

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

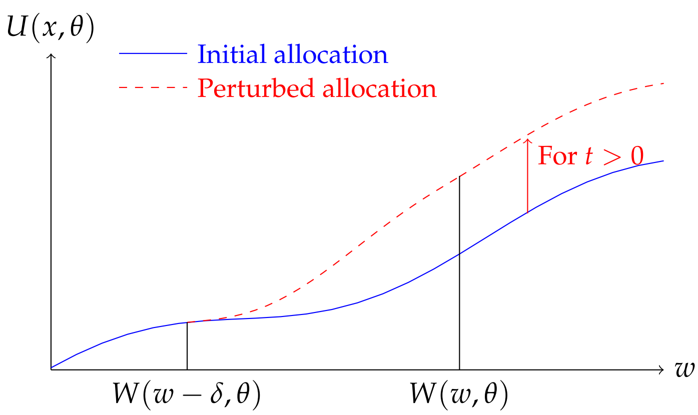

Figure 3: The perturbation of utilities

form amount above $W(w,\theta)$, as illustrated in Figure 3. As incomes above skill $W(w,\theta)$ are unchanged, this implies that, for all skill x above $W(w,\theta)$, the modifications in utility $U(x,\theta)$ occur only through changes of the utility $u(C(x,\theta))$ derived from consumption. Using (21f), this utility therefore changes uniformly by:

$$u\left(\hat{C}(x,\theta;t,\delta)\right) - u\left(C(x,\theta)\right) = \int_{W(w-\delta,\theta)}^{W(w,\theta)} \left[v_w\left(Y(\omega,\theta);\omega,\theta\right) - v_w\left(\hat{Y}(\omega,\theta;t,\delta);\omega,\theta\right)\right] d\omega \tag{21g}$$

which determines the perturbation of consumption for skill levels above $W(w,\theta)$. We now determine how the perturbations of incomes $Y(\cdot,\theta)$ in each group within the skill interval $(W(w-\delta,\theta), W(w,\theta))$ need to be set to ensure that the perturbed allocations remain incentive-compatible. For that purpose, we note that for all skill levels x above w, as incomes in the reference group are not perturbed, the pooling function is also unchanged, so that the same types remain pooled together. Hence, according to (15):

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

This implies that, in all groups, the uniform change in utility that occurs for all skill levels above $W(w,\theta)$ must be identical across groups, so that: $u\left(\hat{C}(x,\theta_0;t,\delta)\right) - u\left(C(x,\theta_0)\right) = u\left(\hat{C}(W(x,\theta),\theta;t,\delta)\right) - u\left(C(W(x,\theta),\theta)\right)$, and so, using (14c) and (21g), we obtain:

$$\begin{aligned}&\int_{w-\delta}^{w} \left[v_w\left(\hat{Y}(\omega,\theta_0;t,\delta);\omega,\theta\right) - v_w\left(Y(\omega,\theta_0);\omega,\theta\right)\right] d\omega \\ = &\int_{W(w-\delta,\theta)}^{W(w,\theta)} \left[v_w\left(\hat{Y}(\omega,\theta;t,\delta);\omega,\theta\right) - v_w\left(Y(\omega,\theta);\omega,\theta\right)\right] d\omega.\end{aligned} \tag{21h}$$

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

The perturbed Lagrangian is:

$$\begin{aligned}\hat{\mathscr{L}}(t,\delta) \stackrel{\text{def}}{\equiv} \iint_{\theta\in\Theta, w\in\mathbb{R}_+} &\left[\hat{Y}(w,\theta;t,\delta) - \mathscr{C}\left(\hat{U}(w,\theta;t,\delta), \hat{Y}(w,\theta;t,\delta);w,\theta\right)\right. \\ &\left. + \frac{\Phi\left(\hat{U}(w,\theta;t,\delta);w,\theta\right)}{\lambda}\right] f(w|\theta) dw\, d\mu(\theta).\end{aligned} \tag{22}$$

If the allocation is optimal, the derivative of this Lagrangian with respect to t must be nil at $t=0$. Appendix A.4 shows that the limit of Condition (22) when δ goes to zero leads to:

$$\begin{aligned}&\int_{\theta\in\Theta} \frac{1 - \dfrac{v_y(Y(W(w,\theta),\theta);W(w,\theta),\theta)}{u'(C(W(w,\theta),\theta)}}{v_{yw}(Y(W(w,\theta),\theta);W(w,\theta),\theta)} f(W(w,\theta)|\theta)\, d\mu(\theta) \\ = &\iint_{\theta\in\Theta, x\geq W(w,\theta)} \left(\frac{\Phi_U < x,\theta >}{\lambda} - \frac{1}{u'\langle x,\theta\rangle}\right) f(x|\theta) dx\, d\mu(\theta).\end{aligned} \tag{23}$$

An intuitive and short rephrasing of the proof follows. Consider the approximation where in each group θ, incomes within skill intervals $[W(w-\delta,\theta)(\theta), W(w,\theta)]$ are changed by a uniform amount denoted $t \times \Delta_Y(\theta)$ instead of the smooth perturbation $\hat{Y}(\cdot,\theta;t,\delta) - Y(\cdot,\theta)$. Appendix A.4 shows that this approximation has only second-order implications that can be neglected when t and δ tend to zero. Let $\delta_w(\theta) \stackrel{\text{def}}{\equiv} W(w,\theta) - W(w-\delta;\theta)$ denote the width of the skill interval where incomes are perturbed in group θ. Let $t \times \Delta U$ denote the uniform amount by which utility is changed for all types x above $W(w,\theta)$. Equation (21f) implies that the rate of change in incomes $\Delta_Y(\theta)$ and the size $\delta_w(\theta)$ of the skill interval over which incomes are perturbed are linked to the rate of change in utility ΔU through:

$$\forall \theta \in \Theta \qquad \Delta_U = -v_{yw}\langle W(w,\theta),\theta\rangle\, \Delta_Y(\theta)\, \delta_w(\theta). \tag{24}$$

Under the aforementioned approximation and using (18), the derivative of the Lagrangian (22) with respect to t can be approximated, when δ tends zero, by:

$$\begin{aligned}
\frac{\partial \mathscr{L}}{\partial t} &\simeq \int_{\theta\in\Theta} \left[1 - \frac{v_y\langle W(w,\theta),\theta\rangle}{u'(C(W(w,\theta);\theta))}\right] f(W(w,\theta)|\theta)d\mu(\theta)\ \Delta_Y(\theta)\ \delta_w(\theta) \\
&+ \Delta_U \iint_{\theta\in\Theta, x\geq W(w,\theta)} \left(\frac{\Phi_U\langle x,\theta\rangle}{\lambda} - \frac{1}{u'(C(x,\theta))}\right) f(x|\theta)dxd\mu(\theta) \\
&= \Delta_U \left\{ \int_{\theta\in\Theta} \frac{1 - \frac{v_y\langle W(w,\theta),\theta\rangle}{u'(C(W(w,\theta);\theta))}}{-v_{yw}\langle W(w,\theta),\theta\rangle} f(W(w,\theta)|\theta)\ d\mu(\theta) \right. \\
&+ \left. \iint_{\theta\in\Theta, x\geq W(w,\theta)} \left(\frac{\Phi_U\langle x,\theta\rangle}{\lambda} - \frac{1}{u'(C(x,\theta))}\right) f(x|\theta)dx\ d\mu(\theta) \right\}
\end{aligned}$$

where the second equality follows from (24). Dividing by Δ_U leads to (23). Using (2), (6) and (15), we can rewrite (23) as:

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

Using again (2), (6) and (15) leads to (20a). □

The tax formula of Proposition 1 generalizes the structural optimal income tax formula derived by Mirrlees (1971) to multidimensional individual characteristics. When the unobserved heterogeneity has only one dimension, we show in Appendix A.5, Equations (20a) and (20b) simplify to:

$$\frac{T'\langle w\rangle}{1-T'\langle w\rangle} \cdot \frac{v_y\langle w\rangle}{-w\ v_{yw}\langle w\rangle}\ w\ f(w) = u'\langle w\rangle \int_w^\infty \left(\frac{1}{u'\langle x\rangle} - \frac{\Phi_U\langle x\rangle}{\lambda}\right) f(x)\ dx \tag{25a}$$

$$0 = \int_0^\infty \left(\frac{1}{u'\langle x\rangle} - \frac{\Phi_U\langle x\rangle}{\lambda}\right) f(x)\ dx. \tag{25b}$$

Comparing these equations with Equations (20a) and (20b) makes clear that reducing the tax problem to one dimension of heterogeneity implies that the integrals over θ-groups disappear. With multidimensional heterogeneity, one needs to aggregate the terms of the formula for individuals of the different groups who pool at the same level of income. This is made possible thanks to our characterization of the pooling function in Lemmas 3 and 4.

The optimal tax formula of Proposition 1 depends only on structural parameters. To highlight its advantages, we now study this structural formula with isoelastic and quasilinear preferences. Isoelasticity implies that the term $v_y/(-w\ v_{yw})$ under the integral in the left-hand side of (20a) becomes $\theta/(1+\theta)$, which is policy-invariant. Under quasilinearity, $u(.)$ in Equation (1) is linear as in Diamond (1998), and we get:

Proposition 2. *If preferences are quasilinear and isoelastic, i.e., if* $\mathscr{U}(c,y;w,\theta) = c - \frac{\theta}{1+\theta}\left(\frac{y}{w}\right)^{1+\frac{1}{\theta}}$, *the optimal tax system is given by:*

$$\frac{T'(Y(w,\theta_0))}{1-T'(Y(w,\theta_0))} = \mathcal{A}(w)\times\mathcal{B}(w)\times\mathcal{C}(w) \qquad \textit{where:} \tag{26a}$$

$$\textit{Efficiency}: \quad \mathcal{A}(w) = \frac{\int_{\theta\in\Theta} W(w,\theta)f(W(w,\theta)|\theta)d\mu(\theta)}{\int_{\theta\in\Theta}\frac{\theta}{1+\theta}W(w,\theta)f(W(w,\theta)|\theta)d\mu(\theta)} \tag{26b}$$

$$\textit{Equity}: \quad \mathcal{B}(w) = \frac{\iint\limits_{\theta\in\Theta,x\geq W(w,\theta)}\left(1-\frac{\Phi_u(U(x,\theta);x,\theta)}{\lambda}\right)f(x|\theta)dx\,d\mu(\theta)}{\iint\limits_{\theta\in\Theta,x\geq W(w,\theta)} f(x|\theta)dx\,d\mu(\theta)} \tag{26c}$$

$$\textit{Distribution}: \quad \mathcal{C}(w) = \frac{\iint\limits_{\theta\in\Theta,x\geq W(w,\theta)} f(x|\theta)dx\,d\mu(\theta)}{\int_{\theta\in\Theta} W(w,\theta)f(W(w,\theta)|\theta)d\mu(\theta)}. \tag{26d}$$

In the absence of group heterogeneity, that is, if Θ is reduced to a singleton, then (26a)-(26d) correspond exactly to the ABC formula of Diamond (1998, Equation (10)) with an efficiency term, an equity term and a distribution one and the usual interpretation prevails. However, the generalization to our multidimensional case is not straightforward.

First, the *efficiency* term $\mathcal{A}(w)$ is not a mere rewriting of Diamond (1998)'s efficiency term, $(1+\overline{\theta})/\overline{\theta}$, where $\overline{\theta}$ would be the mean elasticity $\overline{\theta}$ across groups. Our tax formula reveals that $\mathcal{A}(w)$ is actually the *harmonic* mean,[24] in each θ group, of Diamond's efficiency term $(1+\theta)/\theta$. Interestingly, by the concavity of $x \mapsto x/(1+x)$ and the Jensen inequality, $\mathcal{A}(w)$ is lower than $(1+\overline{\theta})/\overline{\theta}$. Therefore, incorrectly taking heterogeneous elasticities θ into account leads, *ceteris paribus*, to an upward bias in the marginal tax rates when implementing the tax formula. Moreover, in the harmonic averaging procedure, the weights are not simply the densities of people who earn the relevant level of income in each group θ, but the products of a skill level and a density, $W(w,\theta)f(W(w,\theta)|\theta)$.

The *equity* term $\mathcal{B}(w)$ is the arithmetic mean of the equity terms $1-(\Phi_U(U(x,\theta);x,\theta)/\lambda)$ over all individuals whose skill x is above $W(w,\theta)$. The difference with Diamond (1998) is that the mean is computed across groups.

In our model with θ groups, the *distribution* term $\mathcal{C}(w)$ is the weighted mean of local Pareto parameters across groups at skill level $W(w,\theta)$.[25] Note that the weights are not simply the skill density at $W(w,\theta)$ but are again equal to the products of this skill level and the associated density, $W(w,\theta)f(W(w,\theta)|\theta)$.

[24] The harmonic mean is $\frac{\int p(x)dx}{\int x^{-1}\,p(x)dx}$ where x is a random variable and $p(x)$ are weights. The arithmetic mean is simply $\frac{\int x\,p(x)dx}{\int p(x)dx}$.

[25] $\mathscr{P}(\omega,\theta) \stackrel{\text{def}}{\equiv} \frac{\int_{x\geq\omega} f(x|\theta)dx}{\omega\, f(\omega|\theta)}$ denotes the local Pareto parameter of the skill distribution within group θ at skill ω. From the definition of local Pareto parameters, we have: $\int_{x\geq\omega} f(x|\theta)dx = \mathscr{P}(\omega)\;\omega\;f(\omega|\theta)$. Therefore, we get: $\mathcal{C}(w) = \int_{\theta\in\Theta}\mathscr{P}(W(w,\theta))\,\frac{W(w,\theta)\;f(W(w,\theta)|\theta)}{\int_{\theta\in\Theta}W(w,\theta)\;f(W(w,\theta)|\theta)\;d\mu(\theta)}d\mu(\theta)$. If heterogeneity were one-dimensional, the distribution term would be given by a single local Pareto parameter.

III.3 Signing optimal marginal tax rates

With multidimensional heterogeneity, the literature has highlighted that negative marginal tax rates can prevail. In Boadway et al. (2002), Choné and Laroque (2010) and Lockwood and Weinzierl (2015), individuals differ along their skills and preferences for effort, and the social planner has weighted utilitarian preferences (see (9a)). In this context, individuals who pool at the same income level $Y(w, \theta_0)$ are characterized by different social marginal welfare weights $g(W(w, \theta), \theta)$ (according to (10)). Within each group θ, the social marginal utility of consumption, $\Phi_U(U(w, \theta); w, \theta)$, is decreasing in skill due to the concavity of the general welfarist criterion $\Phi(U; w, \theta)$. However, the arithmetic mean of $\Phi_U(U; w, \theta)$, that appears in the right-hand side of (20a),[26] may not be decreasing in income because it aggregates $\Phi_U(U; w, \theta)$ across groups. This *composition effect* may reduce marginal tax rates (Lockwood and Weinzierl, 2015) and may even induce them to become negative (Boadway et al., 2002, Choné and Laroque, 2010). For instance, this happens when some groups undervalued in the social objective are overrepresented at low income levels. In this case, individuals at the bottom of the income distribution receive lower social welfare weights than individuals with larger income levels. This yields negative marginal tax rates at the bottom of the distribution.

Proposition 3 shows that the result of positive marginal tax rates found in Mirrlees (1971) with a single dimension of heterogeneity and utilitarian preferences still holds with an endogenous pooling function, provided that there is no composition effect. This is illustrated with utilitarian and maximin social preferences.

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

Proof. Let $u' \langle w, \theta_0 \rangle \, I(w)$ denote the right-hand side of (20a). Under utilitarian preferences, $\Phi_u = 1$ and we get: $I(w) \stackrel{\text{def}}{\equiv} \int_{x \geq W(w,\theta)} \left(\frac{1}{u' \langle x, \theta \rangle} - \frac{1}{\lambda} \right) \cdot \left(\int_\theta f(x|\, \theta) \, d\mu(\theta) \right) \, dx$. The derivative of $I(w)$ has the sign of $1/\lambda - 1/u' \langle x, \theta \rangle$, which is decreasing in w because of the concavity of $u(\cdot)$. Moreover, $\lim_{w \mapsto \infty} I(w) = 0$ and Equation (20b) imply that $I(0) = 0$. Therefore, $I(w)$ first increases and then decreases. It is thus positive for all (interior) skill levels. Since $v_{yw} < 0$ from (1), optimal marginal tax rates are positive.

Under maximin, one has $U(x, \theta) > U(0, \theta)$ for all $x > 0$ from (14a). Therefore, within each group, the most deserving individuals are those whose skill $w = 0$. The maximin objective implies $\Phi_U \langle x, \theta \rangle = 0$ for all $x > 0$. Thereby, $I(w) \stackrel{\text{def}}{\equiv} \int_{x \geq W(w,\theta)} \frac{1}{u' \langle x, \theta \rangle} \cdot \int_\theta f(x|\, \theta) \, d\mu(\theta) \, dx$ for all $x > 0$, which leads to positive marginal tax rates. □

IV A tax perturbation approach to the tax formula

In this section, we express the optimal tax formula in terms of *sufficient statistics*. In Appendix A.6, we derive this formula from the structural one (see Proposition 1) obtained with

[26] In the right-hand side of (20a), $1/u'(C(x, \theta)) - \Phi_U(U(x, \theta); x, \theta)/\lambda$ is equal to $(1 - g(x, \theta))/u'(C(x, \theta))$. Individuals who pool at the same income get the same consumption, hence the same marginal utility of consumption. As a result, their heterogeneity appears in the heterogeneous social welfare weights $g(x, \theta)$ of individuals who pool at the same income level.

our allocation perturbation method. In this section, we derive it following the tax perturbation approach of Saez (2001). The section is divided in three parts. We first define the required sufficient statistics and clarify the assumptions on which the tax perturbation approach relies. Second, using this approach, we derive the optimal tax formula in terms of these sufficient statistics, specify the correct averaging procedure of the latter (highlighting the importance of composition effects along the way) and discuss the respective virtues and limitations of the tax perturbation and allocation perturbation approaches. Third, we study the asymptotic tax rate.

IV.1 Sufficient statistics

The sufficient statitics we need are individual elasticities and income effects. We define the *compensated* elasticity using a *compensated* tax reform, which modifies the marginal tax rate by a constant amount τ around income $Y(w,\theta)$, while leaving unchanged the level of tax at this level of income. The income responses are defined as the responses to a small lump-sum change ρ in tax liability. The tax function then becomes $T(Y,\theta) - \tau(Y - Y(w,\theta)) - \rho$ and individuals of type (w,θ) solve the following program:

$$\max_{y} \quad u\left(y - T(y) + \tau(y - Y(w,\theta)) + \rho\right) - v(y;w,\theta). \tag{27}$$

The first-order condition can be written as $\mathscr{Y}(y,\tau,\rho;w,\theta) = 0$, where:

$$\mathscr{Y}(y,\tau,\rho;w,\theta) \stackrel{\text{def}}{\equiv} \left(1 - T'(y) + \tau\right) \cdot u'\left(y - T(y) + \tau(y - Y(w,\theta)) + \rho\right) - v'_y(y;w,\theta) \tag{28}$$

To apply the implicit function theorem to $\mathscr{Y}(Y(w,\theta),0,0;w,\theta) = 0$ in order to obtain the sufficient statistics, the tax perturbation approach requires the following assumptions:

Assumption 3. .

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

iii) For all $(w,\theta) \in \mathbb{R}_+ \times \Theta$, the function $y \mapsto u(y - T(y)) - v(y;w,\theta)$ admits a unique global maximum.

Part *i)* ensures that First-Order Condition (28) is differentiable. Part *ii)* guarantees it is invertible in income y. Under *i)* and *ii)*, one can apply the implicit function theorem to (28) and describe how a local maximum of Program (27) changes after a tax reform. In general, since the tax function is nonlinear, the function $y \mapsto u(y - T(y)) - v(y;w,\theta)$ may admit several global maxima among which individuals of type (w,θ) are indifferent. Each small tax reform may lead to a distinct unique global maximum. Moving from a global maximum to another in the wake of a tax reform is associated with a jump in the chosen labor supply and income. In this case, the optimal allocation does no longer smoothly respond to an infinitesimal tax perturbation. The tax perturbation approach relies on Part *iii)* of Assumption 3 to ensure that it is not the case and that the first-order condition corresponds to a unique global maximum. Applying the implicit function theorem to $\mathscr{Y}(Y(w,\theta),0,0;w,\theta) = 0$ yields $\partial Y/\partial x = -\mathscr{Y}_x/\mathscr{Y}_Y$ for $x = w, \tau, \rho$, where the various derivatives are evaluated at $(Y(w,\theta),0,0;w,\theta)$ with:[27]

$$\mathscr{Y}_y(Y,0,0;w,\theta) = -T''(Y)\cdot u'(Y-T(Y)) + (1-T')^2 \cdot u''(Y-T(Y)) - v_{yy}(Y;w,\theta). \tag{29}$$

From these expressions, we define sufficient statistics, which are elasticities and income responses for individuals in group θ whose skill w is such that they choose income $y = Y(w,\theta)$ in the absence of tax reform (i.e. $\tau = \rho = 0$). These elasticities and income responses encapsulate all responses that appear with a nonlinear income tax, so that we call them *total* elasticities or responses. The *total* compensated elasticity of earnings with respect to the marginal retention rate $1 - T'(.)$ is:

$$\varepsilon(y;\theta) \stackrel{\text{def}}{\equiv} \frac{1-T'(y)}{y}\frac{\partial Y}{\partial \tau} = -\frac{v_y}{y\cdot \mathscr{Y}_y} > 0 \tag{30a}$$

which is positive since $v_y > 0 > \mathscr{Y}_y$. The *total* elasticity of earnings with respect to skill w is:

$$\alpha(y;\theta) \stackrel{\text{def}}{\equiv} \frac{w}{y}\frac{\partial Y(w,\theta)}{\partial w} = \frac{w}{y}\dot{Y}(w,\theta) = \frac{w\, v_{yw}}{Y(w,\theta)\,\mathscr{Y}_y} > 0. \tag{30b}$$

This elasticity is positive from Assumption 1 and Equation (3). We define the *total* income response of earnings to a lump-sum change ρ in the level of income or, for short, the *total* income effect by:

$$\eta(y;\theta) \stackrel{\text{def}}{\equiv} \frac{\partial Y}{\partial \rho} = -\frac{u''\cdot v_y}{u'\cdot \mathscr{Y}_y} \leq 0 \tag{30c}$$

which is non-positive due to the additive separability of individual preferences (1) and $v_y > 0 > \mathscr{Y}_y$. Leisure is therefore a normal good.

Elasticities and income response (30a)-(30c) differ from those usually found in the optimal tax literature by the presence, in their denominators, of a term $T''(y)\cdot u'(Y-T(y))$ which is incorporated in $\mathscr{Y}_Y$ (see Equation (29)). This term accounts for the nonlinearity of the income tax schedule. To understand why, let $\varepsilon^\star(y;\theta)$, $\alpha^\star(y;\theta)$ and $\eta^\star(y;\theta)$ be the compensated elasticity of earnings with respect to the marginal retention rate, the elasticity of earnings with respect to skill and the income effect, when $T'' = 0$ in (30a)-(30c). These would be the relevant concepts if the tax function were linear. We call these terms the *direct* responses and denote them with stars for superscripts. An exogenous change in either w, τ or ρ induces a *direct* change in earnings $\Delta_1 y$ proportional to the direct response $\varepsilon^\star(y;\theta)$, $\alpha^\star(y;\theta)$ and $\eta^\star(y;\theta)$, respectively. However, when the tax schedule is nonlinear, the direct response in earnings Y modifies the marginal tax rate by $\Delta_1 T' = T''(y)\times\Delta_1 y$, thereby inducing a further change in earnings $\Delta_2 y = -y\frac{T''}{1-T'}\varepsilon^\star(y;\theta)\Delta_1 y$. This second change in earnings, in turn, induces a further modification in the marginal tax rate $T''(y)\times\Delta_2 y$ which induces an additional change in earnings. Therefore, a *circular process* takes place. The income level determines the marginal tax rate through the

[27] From (28), we have $\mathscr{Y}_w = v''_{y,w}(Y(w,\theta);w,\theta)$, $\mathscr{Y}_\tau = u'(C(w,\theta))$, $\mathscr{Y}_\rho = (1-T')u''(C(w,\theta))$.

tax function, and the marginal tax rate affects the income level through the substitution effects. Using the identity $1 + x + x^2 + x^3 + ... = \frac{1}{1-x}$, the total effect is given by:

$$\Delta y = \sum_{i=1}^{\infty} \Delta_i y = \Delta_1 y \sum_{i=1}^{\infty} \left(-y \frac{T''}{1-T'} \varepsilon^\star(y;\theta) \right)^{i-1} = \Delta_1 y \frac{1-T'(y)}{1-T'(y)+yT''(y)\varepsilon^\star(y;\theta)}.$$

Our definitions of elasticities and income responses capture the *total* effect, i.e., including the circular process, of slightly modifying either the marginal tax rate, the skill level or the income level. The term $T''(y) \cdot u'(Y - T(y))$ in $\mathscr{Y}_Y$ (see Equation (29)) testifies about this. The literature (e.g. Saez (2001), Golosov et al. (2014), Hendren (2014)) instead considers only the direct effects by assuming that marginal tax rates are exogenous in the computation of elasticities and income responses, thereby taking $T''(Y(w,\theta)) = 0$ in (30a)-(30c). In this case, the tax schedule is locally linear hence total and direct responses coincide. Equations (31a)-(31c) explicit the multiplicative term by which direct responses must be timed to obtain total responses.[28]

$$\varepsilon(y;\theta) = \frac{1-T'(y)}{1-T'(y)+y\,T''(y)\,\varepsilon^\star(y;\theta)}\,\varepsilon^\star(y;\theta) \tag{31a}$$

$$\alpha(y;\theta) = \frac{1-T'(y)}{1-T'(y)+y\,T''(y)\,\varepsilon^\star(y;\theta)}\,\alpha^\star(y;\theta) \tag{31b}$$

$$\eta(y;\theta) = \frac{1-T'(y)}{1-T'(y)+y\,T''(y)\,\varepsilon^\star(y;\theta)}\,\eta^\star(y;\theta) \tag{31c}$$

where the *corrective term* is

$$\frac{1-T'(y)}{1-T'(y)+y\,T''(y)\,\varepsilon^\star(y;\theta)}. \tag{31d}$$

Using various methods, it is the direct responses that the empirical literature estimates (e.g., Saez et al. (2012), Kleven and Waseem (2013)). Obtaining the proper total responses from these estimates is not straightforward because the corrective term in (31d) depends on *(i)* the curvature $T''(.)$ of the tax function which is different in the actual economy where the sufficient statistics are estimated and in the optimal economy and *(ii)* the direct compensated elasticity which is typically heterogeneous across individuals who pool at the same income level.

IV.2 Tax formula and discussion

Tax formula, averaging procedure and composition effects

We now derive the optimal tax formula using the tax perturbation approach. With $h(\cdot|\theta)$ denoting the conditional income density within group θ and $H(\cdot|\theta)$ the corresponding cumulative income distribution function, we obtain the equality $H(Y(w,\theta)|\theta) \equiv \int_{x=0}^{w} f(x|\theta)dx$, for all skills w and groups θ. Differentiating both sides of this equality with respect to w and using (30b), we note that the two densities are linked by:

$$h(Y(w,\theta)|\theta) = \frac{f(w|\theta)}{\dot{Y}(w,\theta)} \qquad \Leftrightarrow \qquad Y(w,\theta)\,h(Y(w,\theta)|\theta) = \frac{w\,f(w|\theta)}{\alpha(Y(w,\theta);\theta)}. \tag{32}$$

[28] These three equalities are obtained from the definitions of elasticities, income responses and from (6). From (28) and (30a) we can write:

$$\frac{\varepsilon(y,\theta)}{\varepsilon^\star(y,\theta)} = \frac{(1-T'(y))^2 u''(c) - v_{yy}(y;w,\theta)}{-T''(y)u'(c) + (1-T'(y))^2 u''(c) - v_{yy}(y;w,\theta)}$$

Substituting (2) into (6) and using the definition of $\varepsilon^\star(y,\theta)$ yields (31a). The same goes for Equations (31b) and (31c).

Therefore, $\frac{1-H(Y(w,\theta)|\theta)}{Y(w,\theta)\ h(Y(w,\theta)|\theta)} = \alpha(Y(w,\theta);\theta)\frac{\int_{x\geq w}^{\infty} f(x|\theta)dx}{w\ f(w|\theta)}$.

Proposition 4. *Under assumptions 1 and 3, the optimal tax schedule satisfies:*

$$\frac{T'(y)}{1-T'(y)} = \frac{1}{\hat{\varepsilon}(y)}\cdot\frac{1-\hat{H}(y)}{y\hat{h}(y)}\cdot\left(1-\frac{\int_y^{\infty}[\hat{g}(z)+\hat{\eta}(z)\cdot T'(z)]\cdot\hat{h}(z)dz}{1-\hat{H}(y)}\right) \tag{33a}$$

$$1 = \int_0^{\infty}[\hat{g}(z)+\hat{\eta}(z)\cdot T'(z)]\cdot\hat{h}(z)dz. \tag{33b}$$

where

$$\hat{\varepsilon}(y) \overset{def}{\equiv} \frac{\int_{\theta\in\Theta}\varepsilon(y;\theta)\ h(y|\theta)\ d\mu(\theta)}{\int_{\theta\in\Theta}h(y|\theta)\ d\mu(\theta)} \tag{34a}$$

$$\hat{h}(y) \overset{def}{\equiv} \int_{\theta\in\Theta}h(y|\theta)\ d\mu(\theta). \tag{34b}$$

$$\hat{g}(y) \overset{def}{\equiv} \frac{\int_{\theta\in\Theta}g\left(Y^{-1}(y,\theta),\theta\right)\ h(y|\theta)\ d\mu(\theta)}{\int_{\theta\in\Theta}h(y|\theta)\ d\mu(\theta)} \tag{34c}$$

$$\hat{\eta}(y) \overset{def}{\equiv} \frac{\int_{\theta\in\Theta}\eta(y;\theta)\ h(y|\theta)\ d\mu(\theta)}{\int_{\theta\in\Theta}h(y|\theta)\ d\mu(\theta)}. \tag{34d}$$

In Proposition 4, $\hat{\varepsilon}(y)$ is the mean *total* compensated elasticity at income level y, $\hat{h}(y)$ is the unconditional income density at income y, $\hat{g}(y)$ is the mean marginal social welfare weight at income y, and $\hat{\eta}(y)$ is the mean *total* income effect at income level y.[29]

Proof of Proposition 4. We consider an infinitesimal tax reform that consists in a uniform decrease $\Delta\tau$ of the marginal tax rates in a small interval $[y-\delta y, y]$ of the income distribution. It implies that the tax levels uniformly decrease by an amount $\Delta\rho = \Delta\tau\times\delta y$ for all income levels z above y.[30] Figure 4 depicts this tax reform. We now describe its effects in two steps.

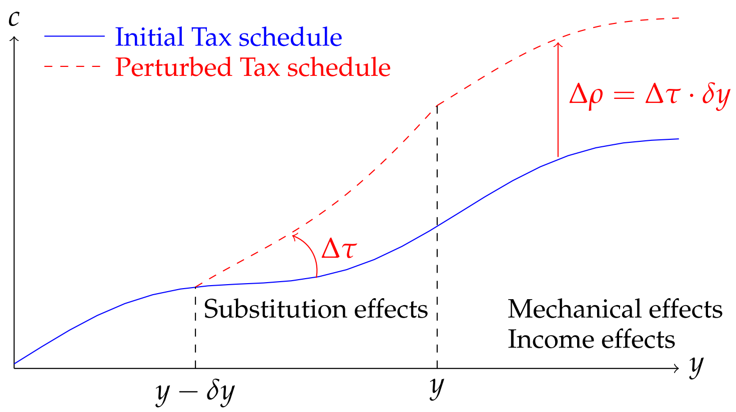

Figure 4: The tax reform

First, the lower marginal tax rate implies that individuals whose income before the tax reforms lies within $[y-\delta y, y]$ increase their income by $\Delta y(\theta)$ due to a *substitution effect*:

$$\Delta y(\theta) = \frac{y}{1-T'(y)}\cdot\varepsilon(y;\theta)\cdot\Delta\tau$$

[29]Note that $\hat{g}(y)$ is the mean of the marginal social welfare weights defined in Equation (10), for individuals in groups θ whose skills, denoted $w = Y^{-1}(y,\theta)$, are such that they all earn the same income $y = Y(w,\theta)$.

[30]We compute the first-order effects when $\Delta\tau, \Delta\rho > 0$. The case when $\Delta\tau, \Delta\rho < 0$ is symmetric. This proof neglects the bunching and gaps created by the kinks generated, at incomes $y-\delta$ and y, by the tax reform.

where (30a) has been used. Note that we here use the *total* compensated elasticity $\varepsilon(y;\theta)$ and not the *direct* one to take into account the circularity detailed in Section IV.1. This substitution effect has only a second-order effect on the utility of these individuals. It however increases their tax liability by:

$$T'(y)\cdot\Delta y = \frac{T'(y)}{1-T'(y)}\cdot\varepsilon(y;\theta)\cdot y\cdot\Delta\tau.$$

As there are $\int_{\theta\in\Theta} h(y|\theta)\cdot d\mu(\theta)\cdot\delta y$ affected taxpayers, these substitution effects lead to a a rise in tax revenue equal to:

$$\frac{T'(y)}{1-T'(y)}\int_{\theta\in\Theta}\varepsilon(y;\theta)\cdot y\cdot h(y|\theta)\cdot d\mu(\theta)\cdot\Delta\rho = \frac{T'(y)}{1-T'(y)}\cdot\hat{\varepsilon}(y)\cdot y\cdot\hat{h}(y)\cdot\Delta\rho \tag{35a}$$

where $\Delta\rho = \Delta\tau\cdot\delta y$, (34a) and (34b) have been used.

Second, individuals whose income levels were above y before the reform receive a transfer $\Delta\rho$ with no change in their marginal tax rate. This has two consequences. First, in the absence of any behavioral response from these workers, the government gets $\Delta\rho$ units of tax receipts less from each of them. Because of this tax reduction, the social objective rises by $g\left(Y^{-1}(z,\theta),\theta\right)$, from (10). The resulting *mechanical effect is equal to* $(-1+g\left(Y^{-1}(z,\theta),\theta\right))\Delta\rho$. Second, each taxpayer with income above y is induced to work less through *income effects*. The variation of the tax level $\Delta\rho$ triggers an income response of $\eta(y;\theta)\cdot\Delta\rho$ for each of them hence, a change $T'(y)\cdot\eta(y;\theta)\cdot\Delta\rho$ in their tax liability, where $\eta(y;\theta)$ is the total (rather than the direct) income effect, see Section IV.1. The sum of the mechanical and income effects for each of these individuals is:

$$\left[-1+g\left(Y^{-1}(z,\theta),\theta\right)+T'(y)\cdot\eta(y;\theta)\right]\cdot\Delta\rho.$$

Summing these effects for the mass $\hat{h}(z)$ of individuals who earn z, we get $[\hat{g}(z)+T'(z)\cdot\hat{\eta}(z)-1]\cdot\hat{h}(z)\cdot\Delta\rho$ where (34b)-(34d) have been used. Therefore, the sum of mechanical and income effects for all income levels above y gives:

$$\int_y^\infty\left[-1+\hat{g}(z)+T'(z))\cdot\hat{\eta}(z)\right]\cdot\hat{h}(z)dz\cdot\Delta\rho. \tag{35b}$$

The tax reform we consider should have no first-order effect at the optimum. This implies that the sum of (35a) and (35b) needs to be nil which leads to (33a).

To obtain (33b) using the tax perturbation approach, simply consider the effect of a uniform increase in tax liability across the entire income distribution. This reform triggers mechanical and income responses for all positive income as well as a loss of one in terms of tax revenue on the entire population. If the tax schedule is optimal, the sum of these mechanical, income responses and loss in tax revenue must be nil, i.e.

$$\int_0^\infty\left[\hat{g}(z)+\hat{\eta}(z)\cdot T'(z)\right]\cdot\hat{h}(z)dz-1=0$$

which gives (33b). □

Equations (33a)-(33b) generalize the optimal tax formula based on sufficient statistics to individuals with multidimensional characteristics. The optimal tax rate given in Equation (33a) consists in three terms: *i*) the behavioral responses to taxes $\frac{1}{\hat{\varepsilon}(y)}$, *ii*) the shape of the income distribution $\frac{1-\hat{H}(y)}{\hat{h}(y)}$ and *iii*) the social preferences and income effects $1 - \frac{\int_y^{\infty}[\hat{g}(z)+\hat{\eta}(z)\cdot T'(z)]\hat{h}(z)dz}{1-\hat{H}(y)}$. Saez (2001) discusses how the optimal tax rate is affected by each of these three terms in the one-dimensional case. Shifting from the model with one dimension of heterogeneity to the model with multiple dimensions leads to replacing the marginal social welfare weight, the compensated elasticity and the income effect by their means calculated at a given income level. Importantly, it is the mean of the *total* (rather than *direct*) compensated elasticity and income effect that must be computed. The weights then correspond to the income density in (34c)-(34d). This is more intuitive than using the direct elasticity and income effect, which implies to encapsulate the circularity (described by (31a)-(31c)) in a so-called "virtual density" as in Saez (2001), Equation (13).

Tax formula (33a) is not a closed-form expression because it depends on sufficient statistics that are generally endogenous and not policy-invariant. In particular, the curvature of the tax function $T''(\cdot)$ affects total compensated elasticities and income effects (see Equation (31)). In addition, multidimensional heterogeneity is a source of composition effects that add to the endogeneity of the tax rate. Composition effects arise as soon as individuals who earn a given income level are not the same in the actual and optimal economies. This implies that each sufficient statistic of Equation (33a) takes distinct values at the optimum and when one estimates it in the actual economy. Therefore, implementing Equation (33a) with real data is at the very least questionable. This issue will be thoroughly examined in Subsection IV.3 for top incomes and in Section V with real data for all income levels.

Equation (33b) is the sufficient statistics equivalent of the structural transversality condition (20b). If income effects were assumed away, this condition implies 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 the term $\hat{\eta}(z) \cdot T'(z)$.

Tax perturbation vs allocation perturbation

In Appendix A.6, we explain how the tax formula expressed in terms of sufficient statistics (Equations (33a)-(33b)) can be recovered from the structural tax formula (Equations (20a)-(20b)) derived using mechanism design. Our structural approach validates the sufficient statistics formula, which has never been done to this day in a multidimensional framework. More importantly, we identify the correct averaging procedure of the sufficient statistics ((34a)-(34d)) only hinted at in Saez (2001), p.220. This procedure is far from intuitive: First, every direct sufficient statistics, denoted with a $\star$ in Equations (31a)-(31c), has to be multiplied by the group-specific corrective term (31d) in order to obtain the *total* sufficient statistics. Then, one has to compute the weighted average of every total sufficient statistics across groups, the weights being the conditional income density for each group. This is a far cry from the simple extension of the unidimensional case that would consist in (*i*) computing the simple average of every direct sufficient statistic and (*ii*) multiplying each result by the corrective term.

An important consequence of the endogeneity of the sufficient statistics is that their values are bound to be different in the optimal economy and in the actual economy where they are estimated. Using a formula based on the policy-invariant primitives of the model (Proposition 1) allows one to get around this problem when computing the optimal tax formula. In addition, formulating the tax schedule in terms of structural parameters allows one to sign the optimal marginal tax rates (Proposition 3), which is far from obvious with the tax formula in terms of sufficient statistics.

We now discuss the respective virtues and limitations of the tax and allocation perturbation approaches and highlight how both approaches relate to each other. While more intuitive, the tax perturbation requires restrictions not only on the tax function that is perturbed (Part *i* of Assumption 3) but also on the way the allocation is affected by the tax perturbation (Parts *ii* and *iii* of Assumption 3). As the allocation is endogenous, imposing restrictions on it is ad-hoc. This is the internal inconsistency of the tax perturbation approach. Conversely, the allocation perturbation requires an assumption on preferences (Assumption 1) and restrictions on the perturbed allocation (Assumption 2). It therefore does not suffer the internal inconsistency of the tax perturbation approach. Indeed, the tax perturbation hinges on Assumption 3. As explained in Subsection IV.1, this assumption prevents jumps in the labor supply when a small tax reform occurs, and ensures that the individual first-order condition corresponds to the unique global maximum. Assumption 1 in Hendren (2014), local Lipschitz continuity of the income function in Golosov et al. (2014) and the assumption that incomes are differentiable with respect to tax reforms in Gerritsen (2016) play a similar role in their analysis. By contrast, our allocation perturbation approach relies on Assumption 2, which is less disputable because it restricts only the set of allocations to be perturbed. Note also that the latter assumption is standard when solving one-dimensional models (see Footnote 23). In addition, all our numerical simulations show that the no-bunching constraint imposed by Assumption 2 never binds. In other words, optimal income levels are always increasing in skills even when Assumption 2 is not imposed.

By adding Assumption 1 (within-group single-crossing), important connections between Assumptions 2 and 3 appear. On the one hand, it is possible to retrieve Assumption 2 from Assumptions 1 and 3, but this requires additional assumptions. Indeed, under Assumptions 1 and 3, Equation (30b) implies that in each group θ, $Y(\cdot,\theta)$ is differentiable with a positive derivative. From there, we can retrieve Assumption 2, if we, in addition, assume that, within each group, $Y(0,\theta) = 0$ and $\lim_{w\mapsto\infty} Y(w,\theta) = \infty$. On the other hand, Appendix A.7 shows the following lemma.

Lemma 5. *Assumptions 1 and 2 imply Assumption 3.*

Lemma 5 states that Assumptions 1 and 2 automatically guarantee Assumption 3. In Proposition 4, the tax schedule based on sufficient statistics has been derived under Assumption 1 and the fairly restrictive Assumption 3. Lemma 5 makes clear that Assumptions 1 and 2, used in our *allocation perturbation* method can also validate the *tax perturbation* approach. The allocation perturbation and the tax perturbation approaches clearly appear as the two faces of the same coin, each of them having advantages and drawbacks. In a nutshell, the tax perturbation method is more intuitive while our allocation perturbation approach is more rigorous and is internally consistent. It also clearly identifies the correct averaging procedures, the corrective term required to obtain the proper sufficient statistics and it guarantees the validity of the tax perturbation approach.

IV.3 Optimal tax rates on top incomes

To study the implications of multidimensional heterogeneity for the optimal asymptotic marginal tax rates, we follow the usual assumptions that lead to the asymptotic tax formula of Saez (2001) and Piketty and Saez (2013). We consider isoelastic individual preferences (see Equation (4))[31] and assume away income effects so that

$$\mathscr{U}(c, y; w, \theta) = c - \frac{\theta}{1+\theta}\left(\frac{y}{w}\right)^{\frac{1+\theta}{\theta}}. \tag{36}$$

We assume that the mean marginal social welfare weight is asymptotically nil (i.e. $\lim_{y \mapsto \infty} \hat{g}(y) = 0$). Taking optimal tax formula (33a) (which has been derived following both the allocation and the tax perturbation approaches) to its limit for high income levels, we obtain a tax formula equivalent to that of Piketty and Saez (2013):[32]

$$\tau_* = \frac{1}{1 + \hat{\varepsilon}_* \, p_*} \qquad \text{where} \qquad p_* = \lim_{y \mapsto \infty} \frac{1 - \hat{H}_*(y)}{y \cdot \hat{h}_*(y)} \tag{37}$$

where τ_* stands for the optimal asymptotic marginal tax rate and $\hat{\varepsilon}_*$ is the mean asymptotic compensated elasticity in the optimal economy. From now on, the variables at the optimum are marked with a subscript asterisk and we use the subscript zero to indicate that a variable is considered in the actual economy.

We now highlight the biases that occur in the asymptotic marginal tax rate when one does not take into account composition effects in the sufficient statistics $\hat{\varepsilon}_*$. In what follows, we adopt the usual assumption that in the actual economy, the upper part of the income density within group θ is described by a Pareto density of the form:

$$h_0(y|\theta) = k_\theta \cdot y^{-(1+p_\theta)} \tag{38}$$

where k_θ is the scale parameter and where the top income distribution term p_θ is the Pareto parameter, with $p_\theta > 1$. Both parameters can vary across groups. We first show that neglecting

[31] Note that the labor supply elasticity θ is a direct elasticity.

[32] A key difference is that Piketty and Saez (2013) aggregate all sufficient statistics across the top bracket earners assuming the top tax rate is constant for these individuals, whereas, in order to obtain the asymptotic tax rate, we average sufficient statistics at the highest income level following the correct procedure detailed in Proposition 4.

the heterogeneity of p_θ leads to major biases in the calculation of the asymptotic marginal tax rate, because such a neglect prevents one to correctly identify the composition of the population at the very top of the income distribution. We then show that even if p_θ does not vary across groups, following an incorrect averaging procedure (i.e. neglecting composition effects) to calculate the sufficient statistic $\hat{\varepsilon}_*$ also biases the asymptotic tax rate.

Identifying the composition of the population at the top of the income distribution

When the Pareto parameter p_θ varies across groups, very top income earners come only from a single θ group: the group whose Pareto distribution has the fatter tail, i.e. with the lowest p_θ. Therefore, from Equation (34a), the mean compensated elasticity $\hat{\varepsilon}_*$ is simply equal to the single θ. The latter can be dramatically different from the one estimated from the average response among, say, the top 1%, which is what Piketty and Saez (2013) suggest to do.

As an illustration, consider an economy composed of two groups of equal size in the top 1% of the population with $\theta_1 = 0.2$ and $\theta_2 = 0.8$. If the high-elasticity group has a Pareto parameter p_2 slightly above 1.5, while the low-elasticity group has a Pareto parameter p_1 slightly below 1.5, then, from (37), the optimal asymptotic marginal tax rate is $1/(1+1.5\times 0.8) = 45.5\%$. Conversely, if p_1 is slightly above 1.5 and p_2 slightly below, the optimal asymptotic marginal tax rate is $1/(1+1.5\times 0.2) = 76.9\%$. Now, if one fails to identify that there are two distinct groups, one will mistakenly consider a single p parameter and mistakenly estimate θ as the mean of θ_1 and θ_2 to implement the optimal asymptotic tax rate. In our example, this yields $\theta = (\theta_1+\theta_2)/2 = 0.4$, which leads to an optimal marginal tax rate of $1/(1+1.5\times 0.4) = 62.5\%$ with $p = 1.5$.

Given the lack of empirical evidence concerning difference in Pareto parameters across groups with different labor supply elasticities, one can be skeptical of asymptotic marginal tax rates calibrations based on the mean across the top percentile of the income distribution, see e.g. Saez et al. (2012) and Piketty and Saez (2013).[33] Our theoretical analysis thus calls for a change of focus in the empirical analysis: Since individuals are heterogeneous along multiple dimensions, one needs to estimate the elasticity of the group whose distribution has the fatter Pareto tail.

Sufficient statistics with and without composition effects

We now show that, even if the Pareto parameter p_θ did not vary across groups, it would be crucial to correctly compute the sufficient statistics $\hat{\varepsilon}_*$, taking composition effects into account. Assume $p_\theta = p$ for all groups. The mean compensated elasticity $\hat{\varepsilon}_*(y)$ is obtained, following Equation (34a), from the asymptotic Pareto income density in the optimal economy, which itself

[33] Saez Slemrod and Giertz (2012) and Piketty and Saez (2013) derive an optimal tax formula for all income above a threshold as a function of the mean taxable income elasticity above this threshold and of the Pareto coefficient. Their implicit assumption is that the elasticity of taxable income and the local Pareto coefficient are roughly constant, so their formula is robust to change in the threshold. Our argument is that such implicit assumptions can lead to misleading policy prescriptions, in particular if the Pareto coefficients are different between high-elasticity and low-elasticity groups.

derives from the following transformation of the density in the actual economy:

$$h_*(y|\theta) = k_\theta \cdot \left(\frac{1-\tau_*}{1-\tau_0}\right)^{\theta p} \cdot y^{-(1+p)} = \left(\frac{1-\tau_*}{1-\tau_0}\right)^{\theta p} \cdot h_0(y|\theta). \tag{39}$$

Proof First-order condition (6) and Equation (36) imply that individuals of type (w, θ) who face the asymptotic marginal tax rate τ earn income:

$$y(w, \theta) = (1-\tau)^\theta w^{1+\theta} \tag{40}$$

in the optimal economy. Inverting (40) , we can write the skill level of individuals belonging to group θ and earning income y in the optimal economy as:

$$w = y^{\frac{1}{1+\theta}} (1-\tau_*)^{-\frac{\theta}{1+\theta}}. \tag{41}$$

The latest two equations allow us to write the income earned in the actual economy by an individual who earns y in the optimal economy as:

$$\tilde{Y}_0(y, \theta) = \left(\frac{1-\tau_0}{1-\tau_*}\right)^\theta \cdot y \tag{42}$$

From the latter, we can write

$$H_*(y|\theta) = H_0(\tilde{Y}_0(y, \theta)|\theta) \tag{43}$$

Differentiating both sides of (43) in y and using (38) and (42), we obtain (39). □

Equation (39) highlights the difference in the conditional income densities driven by composition effects. This difference is larger when the labor supply elasticity θ is larger and results in distinct optimal and actual asymptotic mean compensated elasticities. According to (34a) and (39), the optimal asymptotic compensated elasticity is given by:

$$\hat{\varepsilon}(y) = \int_{\theta \in \Theta} \theta \cdot \frac{k_\theta \cdot \left(\frac{1-\tau^*}{1-\tau_0}\right)^{\theta p}}{\int_{\tilde{\theta} \in \Theta} k_{\tilde{\theta}} \cdot \left(\frac{1-\tau^*}{1-\tau_0}\right)^{\tilde{\theta} p} \cdot d\mu(\tilde{\theta})} \cdot d\mu(\theta). \tag{44}$$

To illustrate this second source of bias in the implementation of the asymptotic tax rate, consider again two groups of equal size, $k_{\theta_1}\mu(\theta_1) = k_{\theta_2}\mu(\theta_2)$. We set $\theta_1 = 0.2$ for the low-elasticity group, $\theta_2 = 0.8$ for the high-elasticity group and $p = 1.5$. If the actual asymptotic marginal tax rate is $\tau_0 = 35\%$, then numerically solving Equations (37) and (44) yields, in the optimal economy, an asymptotic compensated elasticity $\hat{\varepsilon}_* = 0.434$ and an optimal asymptotic marginal tax rate equal to $\tau_* = 60.6\%$.

Then, neglecting multidimensional heterogeneity leads to (44) being reduced to $\hat{\varepsilon}(y) = \theta$. When neglecting multidimensional heterogeneity, one does not estimate two distinct values $\theta_1 = 0.2$ and $\theta_2 = 0.8$ but simply an average value of $\theta = (\theta_1 + \theta_2)/2 = 0.5$ so that $\hat{\varepsilon}(y) = 0.5$. From (37), we obtain an optimal asymptotic marginal tax rate equal to $1/(1 + 1.5 \times 0.5) = 57.1\%$. These values are to be compared to the 0.434 and 60.6% found above, respectively. These differences appear because, in neglecting multidimensional heterogeneity, one relies on incorrect income densities ($h(y|\theta_1) = h(y|\theta_2)$ in the actual economy) rather than on the correct income densities derived in the optimal economy. Indeed, the latter densities incorporate the fact that the change in tax rates from the actual to the optimal economy modifies the income distribution and therefore the share of taxpayers endowed with a high elasticity.

V Numerical Illustration

In this section, we start by numerically implementing the optimal tax formula of Proposition 1, so as to quantify the crucial role played by multidimensional heterogeneity. In Subsection V.1, we document the possibly important quantitative impact on the optimal marginal tax rates of erroneously assuming identical behavioral elasticities across individuals who pool at the same income level. We then highlight the consequences for the tax formula of three usual miscalculations of the sufficient statistics. The first (Subsection V.2) consists in using the direct rather than the total responses when implementing the tax formula. The second and third (Subsection V.3) respectively consists in (i) directly plugging data from the actual economy in the density of income rather than determining its value in the optimal economy and (ii) neglecting the composition effects. While the first error has only a moderate impact on the optimal tax rates, the latter two may lead to serious biases.

V.1 One-dimensional versus multidimensional heterogeneity

We calibrate the model assuming an individual utility without income effects $u(c) = c$ (as in e.g., Atkinson (1990) and Diamond (1998)) and a disutility of income $v(y, w; \theta) = \frac{\theta}{1+\theta}(y/w)^{1+\frac{1}{\theta}}$ where θ is the *direct* (taxable income) elasticity. We assume Bergson-Samuelson social preferences with $\Phi(.; w, \theta) = \log(.)$. For our illustrative purposes, it is not necessary to go beyond two dimensions of individual heterogeneity. The two dimensions we choose are the direct elasticity θ and the earning ability (or skill) w.

We calibrate the model from a subsample of the March 2013 CPS that consists in single men or women without children. We consider two scenarii. In the first one, which we call the multidimensional scenario, there are two θ groups: wage earners on the one hand and self-employed on the other. Again, for illustrative purposes, it is not necessary to go beyond two θ groups. We assume these two groups have different taxable income elasticities because wage earners have much fewer possibilities to adjust their labor supply or to evade their income than the self-employed (see e.g., Sillamaa and Veall (2001), Saez (2010), Kleven et al. (2011)). We take realistic direct taxable income elasticities of $\theta = 0.8$ for the self-employed and $\theta = 0.2$ for salary workers. We then recover the skill distribution in each group from individuals' first-order condition (6) applied to the self-employed and salary workers' respective income data.[34]

[34] To approximate an unbounded skill distribution, we run simulations over the income range [\$0; \$1,000,000], exogenously adding a mass point at the highest income level to ensure that every conditional income density mimics a Pareto unbounded distribution for high income levels. Note that we show results only for income below \$250,000.

The second scenario, that we call the Mirrlees scenario, corresponds to the usual one-dimensional case where individuals differ only in skills. All individuals have the same direct elasticity θ which is computed as the sample mean, $\theta = 0.248$, of the direct elasticities used in the first scenario.

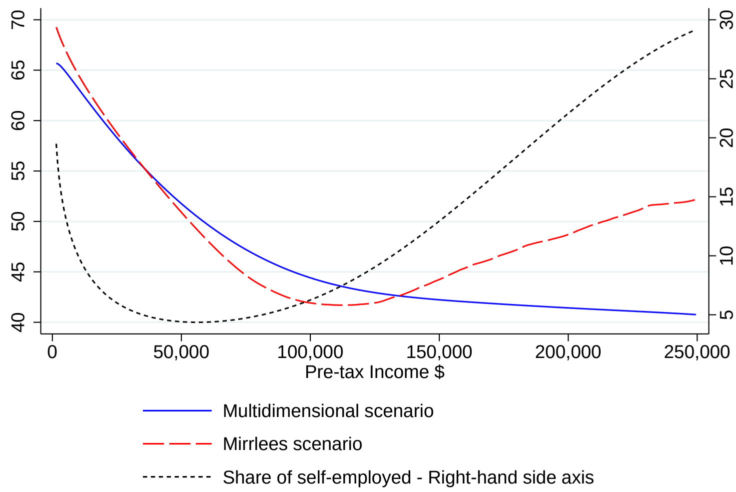

Figure 5: Optimal marginal tax rates

To obtain the optimal tax profiles, we implement our structural tax formula (Equation (20a)), using an algorithm which is detailed in Appendix B. The optimal marginal tax rates in the two scenarii are shown on Figure 5 with the percentage of the marginal tax rate on the left-hand side vertical axis. We observe significant differences between the shape of the tax profiles obtained in the Mirrlees scenario (dashed line) and the shape of those obtained in the multidimensional scenario (solid line). This is due to variations in the share of self-employed along the income distribution represented on the right-hand side vertical axis (dotted line). This share affects the mean compensated elasticity $\hat{\varepsilon}(y)$ in the scenario with heterogeneous elasticity. From the first term in the right-hand side of (33a), we know that a larger mean elasticity reduces the marginal tax rate, *ceteris paribus*. In the lower part of the income distribution, the share of self-employed is relatively large. This drives up the mean elasticities at these income levels hence, it slightly reduces the optimal marginal tax rates. Similarly, in the upper part of the income distribution, the share of self-employed is sharply increasing with income. Therefore, the marginal tax rates are drastically reduced. The reduction of marginal tax rate reaches up to 11 percentage points around $\$250,000$ when heterogeneous elasticities are taken into account.

V.2 Total versus direct sufficient statistics

Proposition 4 and Equations (31a)-(31c) makes it clear that the tax formula in terms of sufficient statistics has to be implemented with *total* and not *direct* responses. However, it is the direct elasticities that are estimated in practice, essentially because actual tax schedules are piecewise-linear. One may then be tempted[35] to implement the tax formula mistakenly using the direct instead of total elasticities. We now quantify the bias that such an approximation induces in the implementation of the optimal tax formula. Figure 6 contrasts, in the multidimensional scenario, the tax schedules obtained with the mean total compensated elasticity and the one obtained with the mean direct compensated elasticity (the differences in the elasticities themselves are presented in Figure 10 in Appendix B.1). Note that the mean total compensated elasticity is retrieved from Equation (34a) whereas the mean direct compensated elasticity is obtained by replacing $\varepsilon(y,\theta)$ with $\varepsilon^{\star}(y,\theta)$ in Equation (34a).[36] The differences in marginal tax rates induced by using *direct* rather than *total* compensated elasticities are relatively small. Mistakenly using the direct rather than the total elasticity leads to a difference of optimal marginal tax rates below 2 percentage points. The largest difference occurs around $50,000 where optimal marginal tax rates are decreasing at the highest pace. In Figure 6, we also add the optimal marginal tax rates obtained, in the Mirrlees scenario, with direct and total compensated elasticities. Again, the resulting differences in marginal tax rates are rather small. This would suggest that approximating total elasticities by direct ones might be an acceptable approximation for a numerical exercise which is reassuring for the literature at large. Finally, Figure 6 highlights that neglecting multidimensional heterogeneity and using direct compensated elasticity yields marginal tax rates that are very different from the ones generated in the multidimensional scenario. This is particularly striking above $120,000 where the share of self-employed becomes larger than the mean of self-employed across the entire population sample.

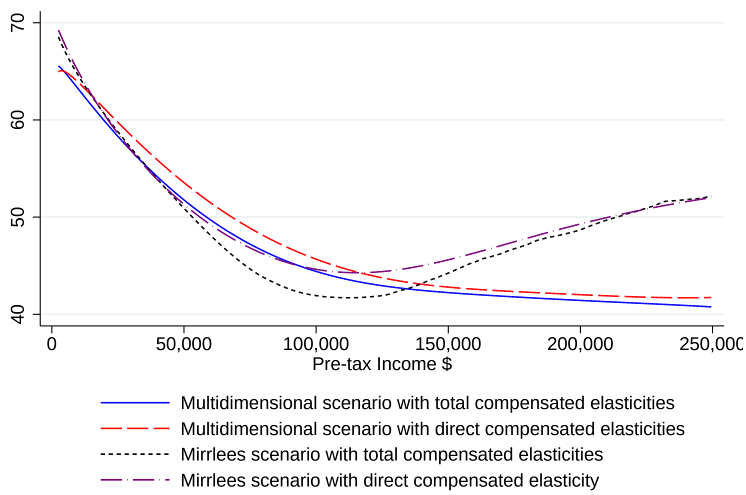

Figure 6: Optimal marginal tax rates with direct vs total elasticities

[35] Note that Saez (2001) uses, in order to implement the tax schedule, a structural tax formula and not a formula in terms of sufficient statistics. Piketty and Saez (2013, page 466) make a similar recommendation.

[36] To compute the mean total elasticity, we use the optimal marginal tax rate obtained from the implementation of the structural tax formula (Equation (20a)). The latter also allows us to calculate the optimal values of $\hat{h}(\cdot)$, $\hat{H}(\cdot)$, $\hat{g}(y)$. We then plug these optimal values into the sufficient statistics tax formula (Equation (33a)) to recover the mean total compensated elasticity $\hat{\varepsilon}(y)$.

V.3 Endogeneity bias and composition effects in sufficient statistics

Even though using direct rather than total elasticities only leads to small differences, one should use the structural tax formula and not the one based on sufficient statistics to compute optimal marginal tax rates. This is because all sufficient statistics should be computed in the optimal economy (which requires the structural tax formula) and not in the actual one. Figure 7 illustrates, in the Mirrlees and the multidimensional scenarii, the drastic differences in marginal tax schedules that result from evaluating the sufficient statistics at the optimum rather than in the actual economy. One of the sufficient statistics in Tax Formula (33a) is the mean marginal social welfare weight $\hat{g}(y)$ which is typically not evaluated in empirical applications. To avoid giving it an undeserved importance in the determination of optimal tax rates and to emphasize the importance of the other sufficient statistics, in what follows we systematically set $\hat{g}(y)$ at the optimal values found in the multidimensional scenario.[37] Therefore, the marginal tax rates vary only with the mean total compensated elasticity and with the income densities.

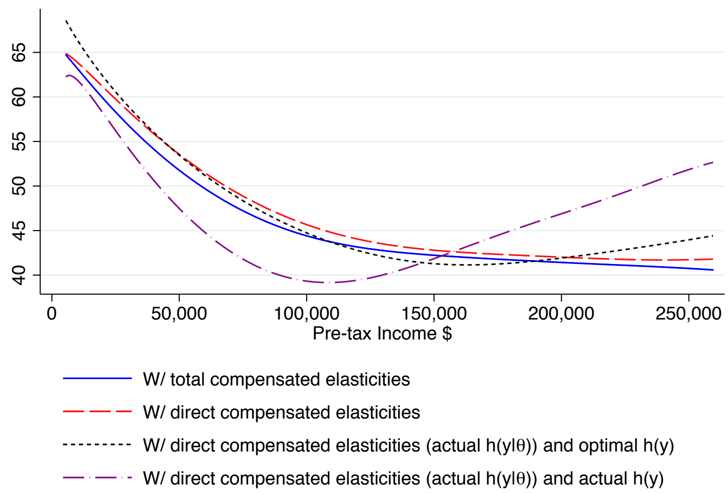

Figure 7: Marginal tax rates, in the multidimensional scenario, with sufficient statistics in the actual and optimal economies

In Figure 7, the blue and red curves are the same as in Figure 6. The black dotted curve depicts the marginal tax rates obtained when direct compensated elasticities are calibrated with $h(y|\theta)$ in the actual economy and when the density $\hat{h}(y)$ is evaluated at the optimum. Comparing this curve with the red one emphasizes the magnitude of the composition effects which imply distinct values for the mean compensated elasticity in the actual and optimal economies. The composition effect stems, at every income level, from the prevalence of the distinct shares of self-employed individuals in the actual and optimal economies. As can be seen, the composition effect implies differences in marginal tax rates up to 3 percentage points for lowest and highest incomes on Figure 7. This difference *a priori* seems to be of a small magnitude, but it

[37] For each curve of Figure 10, the welfare weights are normalized so that the sum of their product with the income density is equal to one.

actually takes other proportions when the marginal tax rates are obtained with both $h(y|\theta)$ and the density $\hat{h}(y)$ evaluated in the actual economy. Doing so leads to the purple curve in Figure 7. The differences between the purple and red curves in this figure highlights the bias induced by mistakenly neglecting the modification of the income distribution term $(1-\hat{H}(y))/(y\hat{h}(y))$ between the actual and optimal economies when calibrating the tax formula. As it can be seen, the error in terms of recommended marginal tax rates is then quite substantial, reaching up to 10 percentage points. When calibrating the tax formula with actual sufficient statistics rather than the optimal ones, marginal tax rates on incomes below (above) \$155,000 are drastically downward (upward) biased. Unfortunately, this improper implementation implies major mistakes in terms of tax policy recommendations.

In the standard Mirrlees' model, calibrating the model with actual direct compensated elasticities and income densities rather than the total compensated elasticities and optimal income densities already leads to a serious bias, as illustrated in Figure 8. In this figure, the difference between the blue and red curves shows the impact of neglecting the corrective term (31d). More remarkably, the difference between the red and black curves shows the dramatic downward bias in marginal tax rates when mistakenly using the initial income density rather than the optimal one. Differing from the multidimensional scenario, this mistake implies that marginal tax rates are downward biased at all income levels. In contrast, marginal tax rates are downward (upward) biased at low (high) income levels when one takes into account multidimensional heterogeneity (i.e., in our example, distinct elasticities for self-employed and wage earners). In other words, taking multidimensional heterogeneity into account aggravates the bias already observed in the unidimensional case, because the direction of the bias changes according to the income level. In addition, the magnitude of the bias is overall more important.

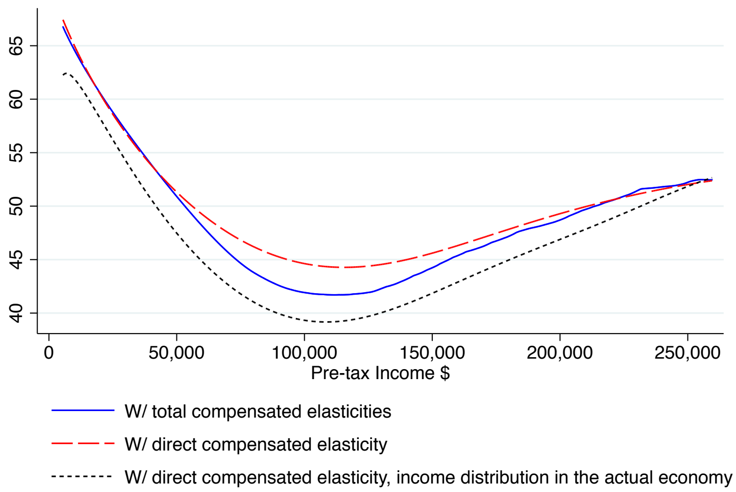

Figure 8: Marginal tax rates in the Mirrlees scenario with sufficient statistics in the actual and optimal economies

VI Concluding Comments

In this paper, we have proposed a new structural method, based on an allocation perturbation, to derive optimal tax schedules and their optimal sufficient statistics, in the very general case where agents are heterogeneous in many dimensions. After contrasting this method with the usual tax perturbation approach, we have quantified the bias in marginal tax rates entailed by using observed sufficient statistics rather than the optimal ones. Using US data, we have shown that, even in a simple illustration, this bias can reach up to 10 percentage points. Our structural tax formula allows us to avoid such a bias and is necessary to correct the observed sufficient statistics.

To illustrate the generality of our framework, we have provided four possible interpretations of our tax formulae: income taxation with heterogeneous skills and heterogeneous labor supply elasticities, joint taxation of labor and non-labor income, joint income taxation of couples and income taxation with tax avoidance. More generally, our approach applies to any taxation problem in which the tax function depends on as many different sources of income as one wishes. It even extends beyond optimal taxation, e.g., to nonlinear pricing problems where consumers differ along several unobserved dimensions. We intend to implement these applications on real data in our future research.

A Theoretical Proofs

A.1 Proof of Lemma 1

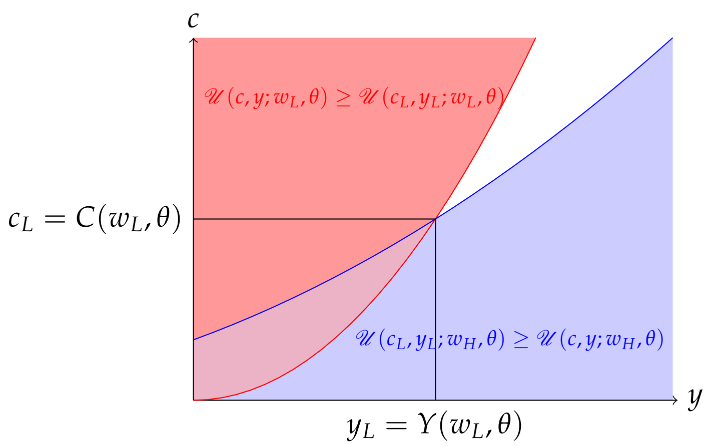

Figure 9: Proof of Lemma 1

Figure 9 displays the indifference curves of individuals belonging to the same group θ but endowed with two distinct skill levels $w_L < w_H$. These indifference curves intersect at the bundle $(C(w_L,\theta), Y(w_L,\theta))$ that the government designs for individuals of type (w_L,θ). The within-group single-crossing assumption implies that the indifference curve of the low-skilled workers is steeper than the one of the high-skilled worker. To respect the within group incentive constraints (13), the government needs to assign a bundle $(C(w_H,\theta), Y(w_H,\theta))$ to the high-skilled workers that is above the indifference curve of the high-skilled workers (otherwise, the individuals of type (w_H, θ) would prefer the bundle $(C(w_L, \theta), Y(w_L, \theta))$ to the bundle $(C(w_H, \theta), Y(w_H, \theta))$ designed for them) and below the indifference curve of the low-skilled workers (otherwise, individuals of type (w_L, θ) would prefer the bundle $(C(w_H, \theta), Y(w_L, \theta))$ to the bundle $(C(w_H, \theta), Y(w_H, \theta))$ designed for them). Consequently, the bundle $(C(w_H, \theta), Y(w_H, \theta))$ designed for the high-skilled workers should be located in the non-shaded area in Figure 9, which implies that $Y(w_L, \theta) \leq Y(w_H, \theta)$, $C(w_L, \theta) \leq C(w_H, \theta)$ and $Y(w_L, \theta) = Y(w_H, \theta)$ if and only if $C(w_L, \theta) = C(w_H, \theta)$.

A.2 Proof of Lemma 2

The steps we follow are standard, see e.g., Salanié (2005). From the taxation principle, individuals choose the type w', θ' that they want to mimic, i.e. they solve $\max_{w',\theta'} \mathscr{U}\left(C(w', \theta'), Y(w', \theta'); w, \theta\right)$. Function $(w', \theta') \mapsto \mathscr{U}\left(C(w', \theta'), Y(w', \theta'); w, \theta\right)$ admits a partial derivative with respect to w' that is equal to:

$$\dot{C}(w', \theta')\ \mathscr{U}_c\left(C(w', \theta'), Y(w', \theta'); w, \theta\right) + \dot{Y}(w', \theta')\ \mathscr{U}_y\left(C(w', \theta'), Y(w', \theta'); w, \theta\right).$$

The first-order condition implies that this expression must be nil at $(w', \theta') = (w, \theta)$. Using (2) leads to (14b). Differentiating in w both sides of $U(w, \theta) = \mathscr{U}\left(C(w, \theta), Y(w, \theta); w, \theta\right)$ leads to:

$$\begin{aligned}
\dot{U}(w, \theta) &= \dot{C}(w, \theta)\ \mathscr{U}_c\left(C(w, \theta), Y(w, \theta); w, \theta\right) + \dot{Y}(w, \theta)\ \mathscr{U}_y\left(C(w, \theta), Y(w, \theta); w, \theta\right) \\
&+ \mathscr{U}_w\left(C(w, \theta), Y(w, \theta); w, \theta\right) \\
&= \left(\frac{\dot{C}(w, \theta)}{\dot{Y}(w, \theta)} - \mathscr{M}\left(C(w, \theta), Y(w, \theta); w, \theta\right)\right) \mathscr{U}_c\left(C(w, \theta), Y(w, \theta); w, \theta\right)\ \dot{Y}(w, \theta) \\
&+ \mathscr{U}_w\left(C(w, \theta), Y(w, \theta); w, \theta\right)
\end{aligned}$$

where the second equality follows (2). Using $\mathscr{U}_w = -v_w$, (14a) holds if and only if (14b) holds.

A.3 Proof of Lemma 4

The proof consists of two steps, to show *(i)* that there exists at most one incentive-compatible allocation $(w, \theta) \mapsto (\underline{C}(w, \theta), \underline{Y}(w, \theta))$ that verifies Assumption 2 and such that $(\underline{C}(w, \theta_0), \underline{Y}(w, \theta_0)) = (C(w, \theta_0), Y(w, \theta_0))$, *(ii)* that this allocation verifies the whole set of incentive constraints (12).

Step (i). To build up the entire incentive-compatible allocation $(w, \theta) \mapsto (\underline{C}(w, \theta), \underline{Y}(w, \theta))$, we must choose $(\underline{C}(w, \theta_0), \underline{Y}(w, \theta_0)) = (C(w, \theta_0), Y(w, \theta_0))$ at any skill level. For each group θ, $\underline{Y}(\cdot, \theta)$ verifies Assumption 2 if and only if its reciprocal $\underline{Y}^{-1}(\cdot; \theta)$ is smoothly increasing. Let $y \in \mathbb{R}_+$ be an income level. As $Y(\cdot, \theta_0)$ is smoothly increasing from Assumption 2, there exists a unique skill level w such that $y = Y(w, \theta_0)$. Then according to Lemma 3, among individuals of group θ, only those of skill $\underline{W}(w, \theta)$ must be assigned to the income level $y = Y(w, \theta_0)$ to verify incentive-compatibility.[38] Therefore, $\underline{Y}^{-1}(\cdot, \theta)$ must be defined by:

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

$\underline{Y}^{-1}(\cdot, \theta)$ is then smoothly increasing as a combination of two smoothly increasing functions. We now show that $\underline{C}(w, \theta)$ is also uniquely determined for any skill level ω and group θ. This is because we know from above that for each type (ω, θ), there exists a single skill level ω such that $\underline{Y}(\omega, \theta) = Y(w, \theta_0)$. Incentive compatibility then requires that $\underline{C}(\omega, \theta)$ also needs to be equal to $\underline{C}(w, \theta_0)$. This ends the proof of step (i).

Step (ii). We now show that the aforementioned incentive-compatible allocation satisfies the within-group incentive constraints. Note that the allocation $(w, \theta) \mapsto (\underline{Y}(w, \theta), \underline{C}(w, \theta))$

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

is built in such a way that one has $\underline{Y}(\omega,\theta) = Y(w,\theta_0)$ and $\underline{C}(\omega,\theta) = C(w,\theta_0)$ if and only if $\omega = \underline{W}(w,\theta)$ and (16) holds. Differentiating in w both sides of these two equations, we obtain $\dot{\underline{Y}}(\underline{W}(w,\theta),\theta)\ \dot{\underline{W}}(w,\theta) = \dot{Y}(w,\theta_0)$ and $\dot{\underline{C}}(\underline{W}(w,\theta),\theta)\ \dot{\underline{W}}(w,\theta) = \dot{C}(w,\theta_0)$. Rearranging terms leads to:

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

As $w \mapsto (C(w,\theta_0), Y(w,\theta_0))$ is assumed to verify the within-group incentive-compatible constraints in Equation (14b), we know that the left-hand side of the above equation is equal to $\mathscr{M}(C(w,\theta_0), Y(w,\theta_0); w,\theta_0)$. Using the definition of $\underline{W}(\cdot,\theta)$, we have that $w \mapsto (\underline{C}(w,\theta), \underline{Y}(w,\theta))$ also verifies Equation (14b). From Lemma 2, it thus verifies the within-group incentive constraints of (13). We now verify whether the inequality (12) is verified for any $(w,w',\theta,\theta') \in \mathbb{R}_+^2 \times \Theta^2$. We know there exists $\omega \in \mathbb{R}_+$ such that $\underline{Y}(\omega,\theta) = \underline{Y}(w',\theta')$ and $\underline{C}(\omega,\theta) = \underline{C}(w',\theta')$. The incentive constraints in (12) are therefore equivalent to:

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

The latter inequality is verified as $w \mapsto (\underline{C}(w,\theta), \underline{Y}(w,\theta))$ also satisfies Equation (14b). Therefore, from Lemma 2, it satisfies the entire set of incentive constraints (13).

A.4 Derivation of Equation (23)

To derive Equation (23), we must compute the various Gâteaux derivatives at $t = 0$. For $A = C, Y, U$ and a given δ, the Gâteaux derivative of A in the direction $\Delta_Y(\cdot,\cdot;\delta)$ at $t = 0$ is defined by:

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

To facilitate reading, we denote it $\hat{\hat{A}}(x,\theta;\delta)$ in this appendix. We derive $\hat{\hat{Y}}(x,\theta_0;\delta) = \Delta_Y(x;\delta)$, and from (21b) we obtain:

$$\hat{\hat{Y}}(x,\theta;\delta) = 0 \qquad \text{if} \qquad x \in [0, W(w-\delta,\theta)] \cup [W(w,\theta), +\infty). \tag{45a}$$

Equations (21c) imply that the Gâteaux derivatives of utilities are nil for skill below $W(w-\delta,\theta)$. For skills x between $W(w-\delta,\theta)$ and $W(w,\theta)$, Equation (21e) implies:

$$\hat{\hat{U}}(x,\theta;\delta) = -\int_{W(w-\delta,\theta)}^{x} v_{yw}\left(Y(\omega,\theta_0);\omega,\theta_0\right)\ \hat{\hat{Y}}(\omega,\theta_0;\delta)\, d\omega. \tag{45b}$$

For skill above $W(w,\theta)$, according to (21f), we have:

$$\hat{\hat{U}}(x,\theta;\delta) = -\int_{W(w-\delta,\theta)}^{W(w,\theta)} v_{yw}\left(Y(\omega,\theta_0);\omega,\theta_0\right)\ \hat{\hat{Y}}(\omega,\theta_0;\delta)\, d\omega. \tag{45c}$$

Moreover, Equation (21h) implies that the Gâteaux derivatives of income must verify:

$$\int_{w-\delta}^{w} v_{yw}\left(Y(\omega,\theta_0);\omega,\theta\right)\ \hat{\hat{Y}}(\omega,\theta_0;\delta)\, d\omega = \int_{W(w-\delta,\theta)}^{W(w,\theta)} v_{yw}\left(Y(\omega,\theta);\omega,\theta\right)\ \hat{\hat{Y}}(\omega,\theta;\delta)\, d\omega. \tag{45d}$$

Using Equations (18), (45a) and (45c), the Gâteaux derivative of the Lagrangian (22) is:

$$\begin{aligned}
\frac{\partial \mathscr{L}}{\partial t}(0;\delta) = & \int_{\theta\in\Theta}\left\{\int_{W(w-\delta,\theta)}^{W(w,\theta)}\left(1 - \frac{v_y(Y(x,\theta);x,\theta)}{u'(C(x,\theta))}\right)\hat{\hat{Y}}(x,\theta;\delta) f(x|\theta)dx\right. \qquad (46)\\
+ & \int_{W(w-\delta,\theta)}^{W(w,\theta)}\left(\frac{\Phi_U < w,\theta >}{\lambda} - \frac{1}{u' < x,\theta >}\right)\hat{\hat{U}}(x,\theta;\delta) f(x|\theta)dx\\
- & \left(\int_{W(w-\delta,\theta)}^{W(w,\theta)} v_{yw}\left(Y(x,\theta);x,\theta\right)\ \hat{\hat{Y}}(x,\theta;\delta)\, dx\right)\\
\times & \left.\left(\int_{W(w,\theta)}^{\infty}\left(\frac{\Phi_U < w,\theta >}{\lambda} - \frac{1}{u' < x,\theta >}\right) f(x|\theta)dx\right)\right\} d\mu(\theta).
\end{aligned}$$

Dividing the first-order condition $\frac{\partial \mathscr{L}}{\partial t}(0;\delta) = 0$ by $\int_{w-\delta}^{w} v_{yw}\left(Y(x,\theta_0);x,\theta_0\right)\,\hat{\hat{Y}}(x,\theta_0;\delta)\,dx$ implies, using (45b) and (45d), that

$$\int_{\theta\in\Theta} \frac{\int_{W(w-\delta,\theta)}^{W(w,\theta)} \left(1-\frac{v_y(Y(x,\theta);x,\theta)}{u'(C(x,\theta))}\right)\hat{\hat{Y}}(x,\theta;\delta)f(x|\theta)dx}{\int_{W(w-\delta,\theta)}^{W(w,\theta)} v_{yw}\left(Y(x,\theta);x,\theta\right)\,\hat{\hat{Y}}(x,\theta;\delta)\,dx}\,d\mu(\theta) = \tag{47}$$
$$\int_{\theta\in\Theta}\left\{\int_{W(w-\delta,\theta)}^{W(w,\theta)}\left(\frac{\Phi_U < w,\theta >}{\lambda}-\frac{1}{u' < x,\theta >}\right)\frac{\int_{W(w-\delta,\theta)}^{x} v_{yw}\left(Y(x,\theta);x,\theta\right)\,\hat{\hat{Y}}(x,\theta;\delta)\,dx}{\int_{W(w-\delta,\theta)}^{W(w,\theta)} v_{yw}\left(Y(x,\theta);x,\theta\right)\,\hat{\hat{Y}}(x,\theta;\delta)\,dx}f(x|\theta)dx\,+\right.$$
$$\left.\int_{W(w,\theta)}^{\infty}\left(\frac{\Phi_U < w,\theta >}{\lambda}-\frac{1}{u' < x,\theta >}\right)f(x|\theta)dx\right\}d\mu(\theta).$$

We finally take the limit of the latter equality when δ tends to 0. Let us consider the first term in the right-hand side of (47). Since

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

we get that:

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

As the right hand-side of the latter inequality tends to 0 when δ tends to 0, the limit of (47) when t tends to zero leads to:

$$\lim_{\delta\mapsto 0}\int_{\theta\in\Theta}\frac{\int_{W(w-\delta,\theta)}^{W(w,\theta)}\left(1-\frac{v_y(Y(x,\theta);x,\theta)}{u'(C(x,\theta))}\right)\hat{\hat{Y}}(x,\theta;\delta)f(x|\theta)dx}{\int_{W(w-\delta,\theta)}^{W(w,\theta)} v_{yw}\left(Y(x,\theta);x,\theta\right)\,\hat{\hat{Y}}(x,\theta;\delta)\,dx}d\mu(\theta) \tag{48}$$
$$=\iint_{\theta\in\Theta,x\geq W(w,\theta)}\left(\frac{\Phi_U < w,\theta >}{\lambda}-\frac{1}{u' < x,\theta >}\right)f(x|\theta)dx\,d\mu(\theta).$$

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

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

and

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

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

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

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

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

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

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

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

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

Hence, left-hand side of (48) is equal to the left-hand side of (23).

A.5 Equations (25a) and (25b)

With one-dimensional heterogeneity, we only consider within-group incentive constraints. Adopting a first-order approach, only (14a) is considered when building up the Hamiltonian:

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

where $Y(w,\theta)$ and $U(w,\theta)$ are the control and state variables respectively. Using (18), the necessary conditions are:

$$\begin{aligned}
0 &= \left(1-\frac{v_y\left\langle w,\theta\right\rangle}{u'\left\langle w,\theta\right\rangle}\right)\cdot f(w|\theta) - q(w|\theta)\cdot v_{yw}\left\langle w,\theta\right\rangle && \text{(49a)}\\
-\dot{q}\left(w|\theta\right) &= \left(\frac{\Phi_U\left\langle w,\theta\right\rangle}{\lambda} - \frac{1}{u'\left\langle w,\theta\right\rangle}\right)\cdot f(w|\theta) && \text{(49b)}\\
0 &= q(0|\theta) && \text{(49c)}\\
0 &= \lim_{w\mapsto\infty} q(w|\theta). && \text{(49d)}
\end{aligned}$$

Combining (49b) with (49d) leads to

$$q(w|\theta) = \int_w^{\infty} \left(\frac{\Phi_U\left\langle w,\theta\right\rangle}{\lambda} - \frac{1}{u'\left\langle w,\theta\right\rangle}\right)\cdot f(\omega|\theta)d\omega. \tag{49e}$$

Combining (2), (6), (49a) and (49e) leads to (25a). Combining (49c) with (49e) leads to (25b).

A.6 Proof of Proposition 4

Dividing (30a) by (30b) we get:

$$\frac{\varepsilon(w,\theta)}{\alpha(w,\theta)} = -\frac{v'_y \langle w,\theta\rangle}{w \cdot v''_{yw} \langle w,\theta\rangle}. \tag{50}$$

Plugging (30a) into (30c) leads to:

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

It is then straightforward to obtain:

$$\hat{\eta}(Y(w,\theta_0)) = Y(w,\theta_0) \cdot \frac{u'' \langle w,\theta_0\rangle}{u' \langle w,\theta_0\rangle} \cdot \hat{\varepsilon}(Y(w,\theta_0)). \tag{51}$$

Let $y \in \mathbb{R}_+$. From Assumption 2, there exists a single w such that $y = Y(w,\theta_0)$. We know that

$$1 - T' \langle w,\theta\rangle = \frac{v'_y \langle w,\theta\rangle}{u' \langle w,\theta\rangle} \tag{52}$$

from (6). The integrand in the left-hand side of (20a) can be rewritten as:

$$\begin{aligned}\frac{v_y \langle W(w,\theta),\theta\rangle}{-W(w,\theta)\ v_{yw} \langle W(w,\theta),\theta\rangle}\ W(w,\theta)\ f(W(w,\theta)|\theta) &= \frac{\varepsilon\left(W(w,\theta),\theta\right)}{\alpha\left(W(w,\theta),\theta\right)} \cdot W(w,\theta)\ f(W(w,\theta)|\theta) \\ &= \varepsilon\left(W(w,\theta),\theta\right)\ Y(w,\theta_0)\ h(Y(w,\theta_0)|\theta).\end{aligned}$$

from Equations (50) and (32), respectively. Combining with (34a), it leads to rewriting (20a) as:

$$\frac{T' \langle w,\theta_0\rangle}{1 - T' \langle w,\theta_0\rangle} \cdot \hat{\varepsilon}\left(Y(w,\theta_0)\right) \cdot Y(w,\theta_0) \cdot \hat{h}(Y(w,\theta_0)) = J(w) \tag{53}$$

where $J(w)$ is defined by the right-hand side of (20a). $J(\cdot)$ admits for derivative $\dot{J}(w)$ where:

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

where (10) has been used. Deriving with respect to the skill w both sides of (15) and of $C(w,\theta_0) = Y(w,\theta_0) - T\left(Y(w,\theta_0)\right)$, we get that:

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

We thus obtain:

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

Using (32) and (53), $\dot{J}(w)$ can be rewritten as:

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

using (51) and (34c). As $J(w)=\int_{x\geq w}(-\dot{J}(x))dx$, we obtain:

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

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

$$J(w)=\int_{z\geq Y(w,\theta_0)}\{1-\hat{g}(z)-\hat{\eta}(z)\cdot T'(Y(z))\}\cdot\hat{h}(Y(x,\theta))\cdot dz. \tag{54}$$

Plugging (54) into (53) gives (33a). Combining (20b) and (54) leads to (33b).

A.7 Proof of Lemma 5

We first prove that the tax function is twice differentiable (i.e. part *i*) of Assumption 3). The tax function can be retrieved from $T(y)\equiv y-\mathscr{C}\left(U\left(Y^{-1}(y,\theta_0),\theta_0\right),y;w,\theta_0\right)$ where $U(\cdot,\theta_0)$ is differentiable in skill since, according to Assumption 2, the right-hand side of Equation (14a) is continuous. Since $Y^{-1}(\cdot,\theta)$ is continuously differentiable in income from Assumption 2, $T(\cdot)$ is continuously differentiable in income. According to (6), $T'(y)\equiv 1-\mathscr{M}(y-T(y),y;Y^{-1}(y,\theta))$. Using again that $Y^{-1}(\cdot,\theta)$ is continuously differentiable in income (from Assumption 2), $\mathscr{M}(y-T(y),y;Y^{-1}(y,\theta),\theta)$ is also continuously differentiable in income so that the marginal tax rate is continuously differentiable in income. We can conclude that the tax function is twice continuously differentiable.

We now show that $\mathscr{Y}_y(Y(w,\theta),0,0;w,\theta)<0$ (i.e. part *ii*) of Assumption 3).[39] By definition of $Y(w,\theta)$, the first-order condition $\mathscr{Y}(Y(w,\theta),0,0;w,\theta)=0$ must be verified at all skill levels w, in all groups θ. Differentiating with respect to w yields $\dot{Y}(w,\theta)\;\mathscr{Y}_y(Y(w,\theta),0,0;w,\theta) = -\mathscr{Y}_w(Y(w,\theta),0,0;w,\theta)$. From (28), $\mathscr{Y}_w=-v_{yw}$, which is positive from Assumption 1. As $\dot{Y}(w,\theta)>0$ from Assumption 2, we have $\mathscr{Y}_y(Y(w,\theta),0,0;w,\theta)<0$.

Finally, we show that for each $(w,\theta)\in\mathbb{R}_+$, $y\mapsto u(y-T(y))-v(y;w,\theta)$ admits a single global maximum (i.e. part *iii*) of Assumption 3). Assume that there exists $y'\neq Y(w,\theta)$ that also maximizes $y\mapsto u(y-T(y))-v(y;w,\theta)$. According to Assumption 2, there exists $w'\neq w$ such that $y'=Y(w',\theta)$. Moreover, the first-order condition $\mathscr{Y}(Y(w',\theta),0,0;w,\theta)=0$ must be verified for individuals of type (w,θ) at $Y(w',\theta)$. As $Y(w',\theta)$ must also solve the individual program for individuals of type (w',θ), the first-order condition must also be verified at $Y(w',\theta)$ for individuals of type (w',θ), so that $\mathscr{Y}(Y(w',\theta),0,0;w,\theta)=\mathscr{Y}(Y(w',\theta),0,0;w',\theta)$. However, as $\mathscr{Y}_w=-v_w$ from Equation (28), Assumption 1 implies that $\mathscr{Y}_w>0$. Therefore, the equality $\mathscr{Y}(Y(w',\theta),0,0;w,\theta)=\mathscr{Y}(Y(w',\theta),0,0;w',\theta)$ can only happen if $w=w'$, a contradiction that ends the proof.

[39] We are grateful to Kevin Spiritus for encouraging us to emphasize this result.

B Numerical simulations

The calibration is based on the March 2013 supplement CPS distribution of adjusted gross income among singles without dependent. We approximate an unbounded income distribution by considering income until $\$1,000,000$, but showing results only until $\$250,000$. Because of top coding of income in the CPS, we extend it with an exogenous mass at income $\$1,000,000$ to mimic a Pareto density with power $-(1+p) = -2.5$.

We use an algorithm based on a discrete grid of the income distribution, whose $2,001$ elements are denoted y_i and are evenly distributed. The different steps of the k^{th} loop are the following, where integrals with respect to skill are approximated by right Riemann sums.

1. Given a tax function $T_k(\cdot)$, find from the individual's first-order condition (6) for each income level y_i and each group θ the skill level $w_i(\theta)$ such that:

$$1 - T'_k(y_i) = \frac{v'\left(y_i; w_i(\theta), \theta\right)}{u'(y_i - T_k(y_i))}$$

2. For each group, use a kernel density estimation to approximate the conditional skill density $f(\cdot|\theta)$ and extend this density by a mass at the highest income to approximate an unbounded Pareto tail at the top. Normalize each conditional skill-density $f(\cdot|\theta)$ to ensure a total mass of $\mu(\theta)$ over all income levels y_i.

3. Use (20b) to compute the Lagrange multipliers λ.

4. Use (20a) to update marginal tax rate to $T'_{k+1}(y_i)$ through:

$$\frac{T'_{k+1}(y_i)}{1 - T'_{k+1}(y_i)} \cdot \int_{\theta\in\Theta} \left\{ -\frac{v'_y\left(y_i; w_i(\theta), \theta\right)}{w_i(\theta)\, v''_{yw}\left(y_i; w_i(\theta), \theta\right)} w_i(\theta)\, f(w_i(\theta)|\theta) \right\} d\mu(\theta) = u'(y_i - T_k(y_i)) \cdot$$
$$\int_{\theta\in\Theta} \left\{ \int_{\omega \geq w_i(\theta)} \left(\frac{1}{u'\left(y_i - T_k(y_i)\right)} - \frac{\Phi'_u(u(y_i - T_k(y_i)) - v(y_i; w_i(\theta), \theta))}{\lambda} \right) f(\omega|\theta) d\omega \right\} d\mu(\theta)$$

5. Update Tax liability $T_{k+1}(y_i)$ to satisfy the budget constraint (7).

6. Go back to Step 1 until $\max_i \left\{ \left| T'_k(y_i) - T'_{k+1}(y_i) \right| \right\} < 0.1\%$.

B.1 Direct vs total compensated elasticities

Figure 10 displays the mean total and direct compensated elasticities, in the multidimensional scenario. The mean total compensated elasticity is higher than the mean direct one around $\$50,000$. With the former elasticity, we obtain lower marginal tax rates around $\$50,000$ as expected theoretically and as can be seen in Figure 6.

References

Akerlof, G. A. (1978). The economics of "tagging" as applied to the optimal income tax, welfare programs, and manpower planning. *American Economic Review 68*(1), 8–19.

Alesina, A., A. Ichino, and L. Karabarbounis (2011). Gender-based taxation and the division of family chores. *American Economic Journal: Economic Policy 3*(2), 1–40.

Atkinson, A. (1990). Public economics and the economic public. *European Economic Review 34*(2-3), 225–248.

Blumkin, T., E. Sadka, and Y. Shem-Tov (2014). International tax competition: zero tax rate at the top re-established. *International Tax and Public Finance, Forthcoming*.

Boadway, R., M. Marchand, P. Pestieau, and M. del Mar Racionero (2002). Optimal redistribution with heterogeneous preferences for leisure. *Journal of Public Economic Theory 4*(4), 475–498.

Boadway, R. and P. Pestieau (2006). Tagging and redistributive taxation. *Annals of Economics and Statistics* (83-84), 123–147.

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

Chetty, R. (2009). Sufficient Statistics for Welfare Analysis: A Bridge Between Structural and Reduced-Form Methods. *Annual Review of Economics 1*(1), 451–488.

Chetty, R. (2012). Bounds on elasticities with optimization frictions: A synthesis of micro and macro evidence on labor supply. *Econometrica 80*, 969–1018.

Choné, P. and G. Laroque (2010). Negative marginal tax rates and heterogeneity. *American Economic Review 100*(5), 2532–47.

Cremer, H., F. Gahvari, and J.-M. Lozachmeur (2010). Tagging and income taxation: Theory and an application. *American Economic Journal: Economic Policy 2*(1), 31–50.

Cremer, H., J.-M. Lozachmeur, and P. Pestieau (2012). Income taxation of couples and the tax unit choice. *Journal of Population Economics* (25:2), 763–778.

Diamond, P. (1998). Optimal income taxation: An example with u-shaped pattern of optimal marginal tax rates. *American Economic Review 88*(1), 83–95.

Diamond, P. and E. Saez (2011). The Case for a Progressive Tax: From Basic Research to Policy Recommendations. *Journal of Economic Perspectives 25*(4), 165–90.

Gerritsen, A. (2016). Optimal nonlinear taxation: the dual approach. *mimeo*.

Golosov, M., Tsyvinski, and N. Werquin (2014). Dynamic tax reforms. NBER Working Papers 20780.

Gomes, R., J.-M. Lozachmeur, and A. Pavan (2014). Differential taxation and occupational choice. CESifo Working Paper 5054.

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

Hammond, P. J. (1979). Straightforward individual incentive compatibility in large economies. *The Review of Economic Studies 46*(2), 263–282.

Hendren, N. (2014). The Inequality Deflator: Interpersonal Comparisons without a Social Welfare Function. *NBER Working Papers* (20351).

Jacquet, L., E. Lehmann, and B. Van der Linden (2013). Optimal redistributive taxation with both extensive and intensive responses. *Journal of Economic Theory 148*(5), 1770–1805.

Kleven, H. J., M. B. Knudsen, C. T. Kreiner, S. Pedersen, and E. Saez (2011). Unwilling or Unable to Cheat? Evidence From a Tax Audit Experiment in Denmark. *Econometrica 79*(3), 651–692.

Kleven, H. J., C. T. Kreiner, and E. Saez (2007). The optimal income taxation of couples as a multi-dimensional screening problem. CESifo Working Paper 2092.

Kleven, H. J., C. T. Kreiner, and E. Saez (2009). The optimal income taxation of couples. *Econometrica 77*(2), 537–560.

Kleven, H. J. and M. Waseem (2013). Using Notches to Uncover Optimization Frictions and Structural Elasticities: Theory and Evidence from Pakistan. *The Quarterly Journal of Economics 128*(2), 669–723.

Laffont, J.-J., E. Maskin, and J.-C. Rochet (1987). Optimal nonlinear pricing with two-dimensional characteristics. In T. Grove, R. Radner, and S. Reiter (Eds.), *Information, Incentives and Economic Mechanism*, pp. 256–266. Univ. of Minnesota Press.

Lehmann, E., L. Simula, and A. Trannoy (2014). Tax me if you can! optimal nonlinear income tax between competing governments. *Quarterly Journal of Economics 129*(4), 1995–2030.

Lockwood, B. B. and M. Weinzierl (2015). De gustibus non est taxandum: Heterogeneity in preferences and optimal redistribution. *Journal of Public Economics* (124), 74–80.

Mankiw, N. G. and M. Weinzierl (2010). The optimal taxation of height: A case study of utilitarian income redistribution. *American Economic Journal: Economic Policy 2*(1), 155–76.

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

Piketty, T. (1997). La redistribution fiscale face au chômage. *Revue Française d'Economie 12*, 157–201.

Piketty, T. and E. Saez (2013). Optimal labor income taxation. In M. F. Alan J. Auerbach, Raj Chetty and E. Saez (Eds.), *Handbook of public economics, vol. 5*, Volume 5 of *Handbook of Public Economics*, Chapter 7, pp. 391 – 474. Elsevier.

Renes, S. and F. Zoutman (2015). A easy as abc? multi-dimensional screening in public finance. Nhh working papers.

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

Rochet, J.-C. and P. Choné (1998). Ironing, sweeping, and multidimensional screening. *Econometrica 66*(4), 783–826.

Rochet, J.-C. and A. Stole, Lars (2002). Nonlinear pricing with random participation. *The Review of Economic Studies 69*(1), pp. 277–311.

Rothschild, C. and F. Scheuer (2013). Redistributive taxation in the Roy model. *The Quarterly Journal of Economics 128*(2), 623–668.

Rothschild, C. and F. Scheuer (2014). A theory of income taxation under multidimensional skill heterogeneity. NBER Working Papers 19822.

Rothschild, C. and F. Scheuer (2016). Optimal taxation with rent-seeking. *Review of Economic Studies (forthcoming)*.

Saez, E. (2001). Using elasticities to derive optimal income tax rates. *Review of Economic Studies 68*, 205–229.

Saez, E. (2002). Optimal income transfer programs: Intensive versus extensive labor supply responses. *Quarterly Journal of Economics 117*, 1039–1073.

Saez, E. (2010). Do taxpayers bunch at kink points? *American Economic Journal: Economic Policy 2*(1), 180–212.

Saez, E., J. Slemrod, and S. H. Giertz (2012). The elasticity of taxable income with respect to marginal tax rates: A critical review. *Journal of Economic Literature 50*(1), 3–50.

Saez, E. and S. Stantcheva (2016). Generalized social marginal welfare weights for optimal tax theory. *American Economic Review 106*(1), 24–45.

Salanié, B. (2005). *The Economics of Contracts*. Cambridge, MA: MIT Press.

Salanié, B. (2011). *The Economics of Taxation*. Cambridge, MA: MIT Press.

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

Scheuer, F. (2014). Entrepreneurial taxation with endogenous entry. *American Economic Journal: Economic Policy 6*(2), 126–63.

Scheuer, F. and I. Werning (2016). Mirrlees meets diamond-mirrlees. NBER Working Papers 22016.

Sillamaa, M.-A. and M. R. Veall (2001). The effect of marginal tax rates on taxable income: a panel study of the 1988 tax flattening in Canada. *Journal of Public Economics 80:3*, 341–356.

Wilson, Robert, B. (1993). *Nonlinear pricing*. Oxford: Oxford University Press.

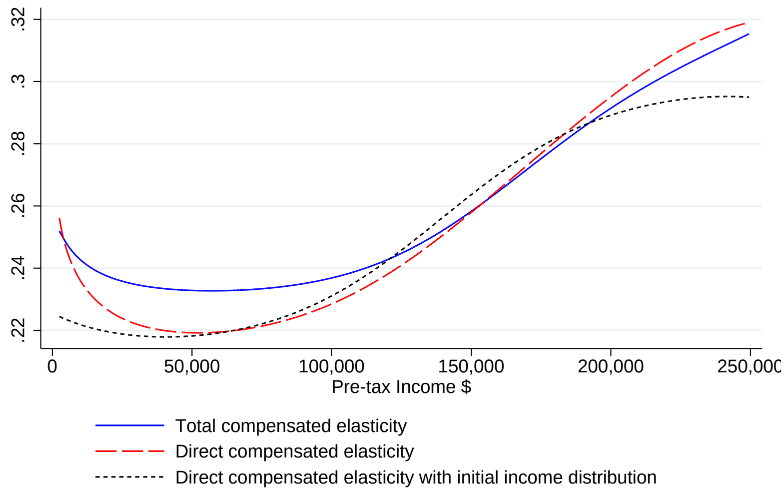

Figure 10: Direct and total elasticities in the multidimensional scenario