CESifo Working Papers

www.cesifo.org/wp

Optimal Income Taxation when Skills and Behavioral Elasticities are Heterogeneous

Laurence Jacquet
Etienne Lehmann

CESIFO WORKING PAPER NO. 5265
CATEGORY 1: PUBLIC FINANCE
MARCH 2015

ISSN 2364-1428

CESifo *Working Paper No. 5265*

Optimal Income Taxation when Skills and Behavioral Elasticities are Heterogeneous

Abstract

We solve a large class of multidimensional adverse selection problems with one observed action, to derive the nonlinear optimal income tax schedule when individuals differ along multiple unobserved characteristics. Based on a perturbation of the optimal allocation, our method allows individuals to have e.g. different skills and different taxable income elasticities. Our optimal tax formula generalizes the one with only one-dimensional source of heterogeneity and is numerically implementable. We find that, compared to the case where individuals differ only in skills, allowing them to also have heterogeous taxable income elasticities leads to substantially different optimal tax schedules and in particular, different asymptotic tax rates.

JEL-Code: H200.

Keywords: optimal taxation, multidimensional screening problems.

Laurence Jacquet
THEMA
University of Cergy-Pontoise
33 boulevand du Port
France – 95011 Cergy-Pontoise Cedex
laurence.jacquet@u-cergy.fr

Etienne Lehmann
CRED (TEPP)
University Panthéon Assas Paris II
12 place du Panthéon 75 231
France – Paris Cedex 05
etienne.lehmann@gmail.com

March 17, 2015
A previous version of this paper has circulated under the title “Optimal Income Taxation when Skills and Labor Supply Elasticities are Heterogeneous”. The authors acknowledge comments made by seminars’ participants at CREST, THEMA, University of Oslo, University of Mannheim, UC Irvine, The Cologne 2014 conference on taxation theory and the 2014 LAGV conference with a particular mention to Pierre Boyer, Pierre Cahuc, Philippe Choné, Helmuth Cremer, Vidar Christiansen, Eric Danan, Olihvier 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, Stefanie Stancheva, Alain Trannoy, Aleh Tsyvinsky, Nicolas Werquin, John Weymark and Danny Yagan. Any remaining errors are ours. Laurence Jacquet’s research is supported by a Marie Curie FP7 Integration Grant within the 7th European Union Framework Program and 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. Laurence Jacquet is also research fellow at CESifo and IRES- Université Catholique de Louvain. Etienne Lehmann is also research fellow at CREST, IRES-Université Catholique de Louvain, IDEP, IZA and CESifo and is junior member of IUF

I Introduction

The nonlinear optimal tax model of Mirrlees (1971), which pioneered the adverse selection literature, assumes individual unobserved heterogeneity to be one-dimensional. This assumption is very restrictive. For instance, it prevents individuals with the same income level to exhibit heterogeneous behavioral responses to tax reforms. One therefore needs to introduce at least a second source of unobserved heterogeneity to derive nonlinear optimal income taxes in more realistic frameworks. Using a first-order approach, this paper solves a class of adverse selection models with one observed action and many unobservable characteristics in order to assess how the nonlinear optimal income tax schedule is modified in such contexts.

The technical difficulty of these adverse selection models lies in characterizing the types of individuals who "pool" by choosing the same action, i.e. by earning the same income. To address this issue, we consider the population as composed of distinct "groups", which are subsets of individuals with the same vector of characteristics except for skill levels. We assume a very general distribution of groups and a continuous skills distribution within each group. We show that, whatever group they belong to, individuals who pool at a given income level must face the same marginal rate of substitution between pretax income and after-tax income, as they face the same marginal tax rate. Because we assume that the single-crossing condition with respect to skill holds within each group, this equality of marginal rates of substitution unambiguously determines the skill levels of individuals in the different groups who pool at any given income level. This characterization of pooling is our first methodological contribution. It allows us to deduce any (smooth) incentive-compatible allocation from its restriction to a single reference group that one arbitrarily chooses.

The pooling of individuals from distinct groups at each level of income induces that we have constraints on state and control variables that hold at endogenous skill levels. This prevents us from using the standard Hamiltonian approach to characterize the optimal allocation. Our second methodological contribution is to overcome this difficulty by adopting an "allocation-perturbation" method. We compute the first-order effects of a small perturbation in the allocation specific to an arbitrarily chosen reference group. Thanks to our characterization of pooling, we can compute how this perturbation in the reference group affects the allocations in all other groups. Our allocation-perturbation methodology uses calculus of variation to deal with endogenous pooling.

We first derive an optimal tax formula expressed in terms of the policy-invariant primitives of the model (Proposition 1), which makes it numerically implementable with real data. This formula also allows us to show that optimal marginal tax rates are positive under Benthamite social preferences and maximin (Proposition 2). We then reformulate our optimality condition to obtain an elasticity-based optimal tax formula in terms of behavioral responses, social welfare weights and income density. We show that the multidimensional context implies that all these terms need to be averaged across individuals who earn the same income (Proposition

3). A difficulty in the averaging procedure lies in the way to integrate the following circular process in the optimal tax formula: In the presence of a nonlinear tax schedule, an individual who faces a tax reform adjusts her income, which in turn induces a further change in the marginal tax rates, which triggers a further behavioral response. This is the reason why we express our optimal tax formula in terms of *total* compensated income elasticity and *total* income responses, which, unlike the usual *direct* compensated income elasticity and *direct* income responses, encapsulate the circular process in their definitions. We show that under multidimensional heterogeneity, the tax formula averages total, and not direct, behavioral responses over individuals who pool at the same income level across the different groups.

We next show that heterogeneous elasticities have dramatic consequences when determining the optimal asymptotic marginal tax rates. As an illustrative example, assume that each group has a distinct taxable income elasticity and a conditional skill distribution that is unbounded Pareto as, is observed empirically. Consider that asymptotic Pareto coefficients differ across groups. When calculating the asymptotic tax rate, only the income elasticity of the group with the fatter-tailed Pareto distribution matters. This elasticity can be drastically different from the average taxable income elasticity among, let us say, the top 1%. However, in the literature, asymptotic tax rates are typically calibrated using this average elasticity among high income earners (Saez, Slemrod, and Giertz, 2012, Piketty and Saez, 2013) which may therefore lead to erroneous recommendations. If one wants to say something about asymptotic tax rates, our theoretical result calls for estimating the income elasticity of the group whose distribution has the fatter Pareto tail.

Moreover, using Current Population Survey data on the US, we illustrate how taking into account multiple dimensions of individual heterogeneity entails important changes in the simulated optimal tax profile. In our numerical illustration, we consider that the elasticity of taxable income is due to both real labor supply and tax avoidance responses. We take the indicator of whether individuals are salary workers or self-employed as a proxy for distinct taxable income elasticities. The self-employed are endowed with higher taxable income elasticity because it is empirically reasonable to consider that they have more possibility to evade income and adjust their labor supply than salary workers (Sillamaa and Veall, 2001, Saez, 2010, Kleven, Knudsen, Kreiner, Pedersen, and Saez, 2011). In this context, we obtain significantly lower (by up to 10 percentage points) optimal marginal tax rates in the upper part of the income distribution where the self-employed are relatively more numerous, compared to the usual scenario where all individuals have the same taxable income elasticity.

It is worth noting 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. Accordingly, when presenting our framework, we show that the latter encompasses many policy-oriented applications as special cases. It can be interpreted to derive the nonlinear optimal income tax schedules, for instance, when individual elasticities of taxable income are due to both real labor supply responses and tax avoidance (as done in our empirical illustration), when individuals earn labor income and non-labor income (capital income, income from renting out property, etc.), when individuals also earn some untaxable non-labor income or when the income of households is taxed jointly. We show that our method also applies beyond optimal taxation, for instance to the nonlinear monopoly pricing problem in a more general framework than Laffont, Maskin, and Rochet (1987).

In our framework with many unobserved characteristics, a single observable action and one instrument, it is obviously crucial to allow for pooling. Conceptually, we can distinguish "pooling" from "bunching", broadly defined in e.g. Rochet and Choné (1998) as a situation where a set of agents of different types choose the same action and are treated identically in the optimal solution. In our context, we define bunching as the specific situation where individuals who belong to the same group but have different skill levels earn the same income. We however assume away bunching which, in our multidimensional case, is not so relevant. Indeed, individuals in the same group have identical characteristics except skills. So assuming that different skill levels always lead to distinct labor incomes is a reasonable assumption which is moreover validated in all our numerical simulations. This assumption allows us to implement a method that relies on the smoothness of the optimal allocation. At first glance, assuming there is no bunching may look at odds with Rochet and Choné (1998) who argue that bunching is generic in adverse selection models with multidimensional heterogeneity. This is because they do not study a model with many unobservable characteristics and one action, but a model where the number of observable actions is the same as the number of unobservable characteristics. In their context, pooling is then irrelevant. Moreover, the bunching Rochet and Choné (1998) have to deal with comes from 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.[1]

Related literature

Our paper relates to the "sufficient statistic" literature as applied to optimal income taxation (e.g.,Piketty (1997), Saez (2001), Diamond and Saez (2011), Piketty and Saez (2013), Hendren (2014), Golosov, Tsyvinski, and Werquin (2014)). 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). In the nonlinear income tax problem, the sufficient statistics are the compensated elasticity and income response, the income density and the social welfare weights. Sufficient statistics are however endogenous and can be different at the optimum and in the actual economy where they are estimated. While this approach is enough to indicate the direction of desirable tax reforms (Golosov, Tsyvinski, and Werquin, 2014), one needs an optimal tax formula expressed

[1] In an appendix available upon request, we introduce a random participation constraint to extend our model for realistic participation responses.

in terms of the policy-invariant primitives of the model to numerically compute the optimal tax schedule. This is exactly what our Proposition 1 does in the context of multidimensional unobserved heterogeneity.

The sufficient statistics approach derives the optimal tax schedule by using a tax perturbation. This method considers the effects of an infinitesimal tax reform on the government's objective. The underlying assumption is that such a reform only triggers first-order changes in individuals' behavior. With a nonlinear income tax schedule, we see however no way to guarantee that individual labor supply decisions will not jump after a small tax reform for individuals who are initially indifferent between two distinct local maxima. Moreover, such tax reforms introduce nonlinearities in the tax schedule, usually through kinks, whose effects are typically ignored by the sufficient statistics approach. This is why we view the tax perturbation method as merely a way to provide an intuition for the economics behind the optimal tax formula, and not as a rigorous proof in itself.[2] In contrast, our allocation-perturbation method avoids these mathematical weaknesses by directly optimizing over smooth incentive compatible allocations. Moreover, the tax perturbation method is unclear about the treatments of pooling and of the circularity process[3] while our method encapsulates it in a clear-cut way. Saez (2001) rigorously shows that a rewriting of Mirrlees (1971)'s tax formula leads to an elasticity-based optimal tax formula, once the latter is expressed in terms of the virtual income density.[4] His proof is however only valid under a one-dimensional unobserved heterogeneity. Our paper provides a formal derivation of the optimal elasticity-based tax formula with multidimensional heterogeneity.

Our paper is part of the growing body of literature dealing with optimal nonlinear income tax under multidimensional unobserved heterogeneity. To the best of our knowledge, this literature assumes that the (intensive) labor supply decision only depends on a one-dimensional aggregation of the multidimensional unobserved heterogeneity, an assumption from which we depart in our approach. Brett and Weymark (2003), Boadway, Marchand, Pestieau, and del Mar Racionero (2002), Choné and Laroque (2010), Lockwood and Weinzierl (2014) make exactly this assumption. The second source of heterogeneity then only matters in the computation of the social welfare weights and enables the government to value differently the welfare of individuals at the same income level. The mean social welfare weights may no longer be decreasing with income which allows optimal marginal tax rates to be negative. In random participation models, individuals differ in skill and in an additional cost of participation (Rochet and Stole, 2002, Kleven, Kreiner, and Saez, 2009, Jacquet, Lehmann, and Van der Linden, 2013) or of migration (Lehmann, Simula, and Trannoy, 2014, Blumkin, Sadka, and Shem-Tov, 2014)

[2] Saez (2001) does not advocate the tax perturbation as a formal proof but calls it a "heuristic proof". Piketty and Saez (2013, Footnote 78, p.436) state that the extension of the optimal tax formula to a multidimensional context "does not seem to have been formally established".

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

[4] defined by Saez (2001, p. 215) as "the density of incomes that would take place [at an income level] z if the tax schedule at $T(\cdot)$ were replaced by the linear tax schedule tangent to $T(\cdot)$ at level z."

that matters only for the extensive margin. In Rothschild and Scheuer (2013, 2014a,b), Scheuer (2013, 2014) and Gomes, Lozachmeur, and Pavan (2014), the total amount of labor earnings of an individual (in distinct sectors of the economy) depends only on a one-dimensional aggregation of her characteristics. This aggregation of characteristics is allowed to depend on the price vector of the different types of labor to encapsulate general equilibrium effects on the wage distribution, hence pooling depends on prices. However in these papers, the one-dimensional aggregation implies that individuals who earn the same pre-tax income are characterized by the same level of aggregated characteristics and are therefore constrained to react identically to any tax reform. Conversely, our method does no rely on an aggregator, so we are able to simultaneously consider heterogeneity in income and heterogeneity in behavioral elasticities.

The rest of the paper is organized as follows. Section II introduces the model and, to emphasize its flexibility, presents how it can easily be adapted to study several policy applications. Section III characterizes optimal incentive-compatible allocations. In particular, it characterizes the set of individuals belonging to different groups who pool at a given income level. Section IV provides optimal marginal tax rates as a function of the primitives of the model and shows that optimal marginal tax rates are positive under Benthamite and maximin social preferences. Section V reinterprets the first-order conditions of our model to obtain an elasticity-based tax formula. This formula is then used to study the optimal taxation of high income earners. Section VI presents our numerical illustration. The last section concludes.

II Model

II.1 The general framework

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 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 density $f(\cdot|\theta)$ which is to be assumed 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 is unable to base the tax on individual types (w, θ). It can only condition taxes and transfers on pre-tax income y through a non-linear income tax function $T(.)$.

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 pre-tax income (for short, income hereafter) $y \geq 0$ with $v_y(\cdot), v_{yy}(\cdot) > 0 > v_w(\cdot)$.[5] Disutility is increasing and convex in income, and decreasing in skill w because earning a given income requires less effort to a more productive agent.[6] Individual preferences are described by the twice differentiable utility function:

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

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:

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

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:

Assumption 1 (Within-group single-crossing condition). *For each $\theta \in \Theta$, and each (c, y), function $w \mapsto \mathscr{M}(c, y; w, \theta)$ maps $\mathbb{R}_+$ onto $\mathbb{R}_+$ with a strictly negative derivative everywhere, so:*[7]

$$\mathscr{M}_w(c, y; w, \theta) < 0 \quad \Leftrightarrow \quad v_{yw}(y; w, \theta) < 0. \tag{3}$$

Assumption 1 also imposes that the marginal rate of substitution decreases from plus infinity to zero. This is a kind of INADA condition that will appear technically convenient. An individual of type (w, θ), facing the nonlinear income tax $y \mapsto T(y)$, solves:

$$\max_y \quad \mathscr{U}\left(y - T(y), y; w, \theta\right) \tag{4}$$

We call $Y(w, \theta)$ the solution to program (4),[8] $C(w, \theta) = Y(w, \theta) - T(Y(w, \theta))$ 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

[5] For any function f of a single variable, we denote f' its first 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.

[6] 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 $v_{yy} > 0 > v_{yw}$.

[7] The assumption $v_{yw}(y; w, \theta) < 0$ is not restrictive. For instance, it holds when preferences are of the form described in Footnote 6.

[8] If the maximization program (4) 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 tax function is differentiable, the first-order condition associated to (4) implies with (2) that:

$$1 - T'\left(Y\left(w,\theta\right)\right) = \mathscr{M}\left(C\left(w,\theta\right), Y\left(w,\theta\right); w,\theta\right) \tag{5}$$

The Government

The government's budget constraint takes the form:

$$\int_{\theta\in\Theta} \left\{ \int_0^{+\infty} \left[Y(w,\theta) - C(w,\theta)\right] f(w|\theta) dw \right\} d\mu(\theta) \geq E \tag{6}$$

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:

$$\int_{\theta\in\Theta} \left\{ \int_0^{+\infty} \Phi\left(U(w,\theta); w,\theta\right)\ f(w|\theta) dw \right\} d\mu(\theta). \tag{7}$$

This welfarist specification allows Φ to vary with individual types (w,θ) which makes it very general. It admits as a particular case *Benthamite* social preferences where $\Phi(U; w, \theta)) \equiv U$. The social objective is then:

$$\int_{\theta\in\Theta} \left\{ \int_0^{+\infty} U(w,\theta)\ f(w|\theta) dw \right\} d\mu(\theta). \tag{8a}$$

Another particular case is weighted utilitarian preferences with type-specific weights $\varphi(w,\theta)$ and $\Phi(U; w,\theta)) \equiv \varphi(w,\theta) \cdot U$. The social objective is then:

$$\int_{\theta\in\Theta} \left\{ \int_0^{+\infty} \varphi\left(w,\theta\right)\ U(w,\theta)\ f(w|\theta) dw \right\} d\mu(\theta). \tag{8b}$$

Our social objective 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:

$$\int_{\theta\in\Theta} \left\{ \int_0^{+\infty} \Phi\left(U(w,\theta)\right)\ f(w|\theta) dw \right\} d\mu(\theta). \tag{8c}$$

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 (2013).

II.2 Policy-relevant applications

This section highlights that the model described above is general enough to encompass many policy-oriented applications that require individuals endowed with multidimensional unobserved characteristics. We explain how redefining and/or reinterpreting variables and individual characteristics make all these applications tractable in our framework. Each example below includes a different source of heterogeneity. These sources are not mutually exclusive and can be simultaneously incorporated in the model. The reader interested in the core model but not in its various applications can skip this section.

II.2.a Optimal Income Taxation with Heterogeneous Skills and Labor Supply Elasticities

Our leading example is nonlinear labor income taxation with heterogeneous skills w and where heterogeneous θ represent distinct *direct* Frisch labor supply elasticities, hereafter, "labor supply elasticities ". In this context, individuals' preferences (1) are assumed to be isoelastic with:

$$\mathscr{U}(c,y;w,\theta) = u(c) - \frac{\theta}{1+\theta}\left(\frac{y}{w}\right)^{1+\frac{1}{\theta}} \qquad \text{with} \qquad \theta > 0 \quad \text{and} \quad u'(\cdot) > 0 \geq u''(\cdot). \tag{9}$$

The optimal tax schedule derived in this paper then directly applies to this context. It is worth noting that Assumption 1 is verified under this specification of preferences.

II.2.b Optimal Joint Taxation of Labor and Non-Labor Income

Our method can also be used to obtain the optimal joint taxation of non-labor income and labor income when the tax function does not distinguish between both types of income. This is the case in countries like France where, for instance, entrepreneurial income and income received from renting property are jointly taxed with labor income. Assume that an argument in the vector of characteristics θ stands for the ability to earn non-labor income z. We note $V(y-z,z;w,\theta)$ the joint disutility of earning labor income $y-z$ and non-labor income z for an individual of skill w who belongs to group θ, with $V_{y-z}, V_z > 0$. The taxable income of this individual is the sum of her labor and non labor income, i.e. y. 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 program (4). This program can be solved sequentially, the last step being the choice of non-labor income z for a given taxable income y. Our model can then be retrieved by defining:

$$v(y;w,\theta) \stackrel{\text{def}}{\equiv} \min_{z} \quad V(y-z,z;w,\theta). \tag{10}$$

Our framework and optimal tax formulas then directly apply whenever Assumption 1 is satisfied, i.e. when the second-order derivatives of $V(\cdot)$ are such that $v_{yw} < 0$.[9] This within-group single-crossing property generally holds in the case where non-labor income z is exogenous. For instance, θ can be rents perceived by landlords who have inherited the property they rent hence, $z = \theta$. When non-labor income z is endogenous, Assumption 1 is still verified when the disutility of income takes the additively separable form $V(y-z,z;w,\theta) = V^{\ell}(y-z;w,\theta) + V^{z}(z;\theta)$, $V^{\ell}(\cdot;w,\theta)$ and $V^{z}(\cdot;\theta)$ are increasing and convex and $V^{\ell}_{yw} < 0$.[10]

---

[9] 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.

[10] We note $z^*(y;w,\theta)$ the solution to (10). 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 the single-crossing assumption $V^{\ell}_{yw} < 0$. 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 $v_{yw} < 0$ so that our method can be used.

II.2.c Optimal Joint Income Taxation of Couples

From our model, we can also obtain the optimal family tax system where the income tax is based on the sum of income from each household members, as in France, Germany or in the US. Everything we just presented in II.2.b then directly applies to the derivation of the optimal family tax system. We simply need to redefine w and θ as the skill levels of each member of the couple and $y - z$ and z as their respective labor incomes. While Kleven, Kreiner, and Saez (2007) and Cremer, Lozachmeur, and Pestieau (2012) stop short of deriving the optimal tax schedule when the couple is the tax unit and each partner decides along the intensive margin, our method allows one to do this.

II.2.d Optimal Income Taxation with Tax Avoidance

Our method is also relevant to solve an optimal income tax problem with tax avoidance for a given tax enforcement. Assume that individuals differ in productivities w and in their abilities to avoid taxation θ. We denote z the sheltered income (i.e. income that is not taxed at all) and y the taxable income, so that labor income is equal to $y + z$ in this context. Consumption is the sum of after-tax income, i.e. $c = y - T(y)$, plus sheltered income z. Assume that preferences are quasi-linear in consumption: $c + z - V(y + z, z; w, \theta)$ where V_{y+z}, $V_{y+z\,y+z}$. Moreover, $V_z, V_{zz} > 0$.[11] To retrieve our model, we simply define:

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

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

II.2.e Optimal Labor Income Taxation with Taxable and Untaxable Non-Labor Incomes

Our framework also allows θ to be some exogenous untaxable non-labor income, as for instance the imputed rent of owner-occupied housing. In this context, consumption is $c + \theta$ since θ is an implicit income which is consumed, and taxable income consists only in labor income y. Assuming that the social objective is $\widetilde{\Phi}(U; w, \theta)$ and that individual preferences over consumption exhibit constant absolute risk aversion (CARA), the individual utility can be stated as $\mathscr{U}(c, y; w, \theta) = -e^{-\gamma(c+\theta)} - \tilde{v}(y; w)$.[12] To solve this model, we simply divide this utility function by $a(\theta) \stackrel{\text{def}}{\equiv} e^{-\gamma\theta}$ which yields individual preferences (1) where:

$$v(y; w, \theta) \stackrel{\text{def}}{\equiv} \frac{\tilde{v}(y; w)}{a(\theta)}$$

[11] 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.).

[12] We here obviously assume that $\tilde{v}_y > 0 > \tilde{v}_w$, that $\tilde{v}_{yw} < 0$ to ensure the within-group single-crossing property and that $\widetilde{\Phi}_u > 0 \geq \widetilde{\Phi}_{uu}$.

and $u(c) = -e^{-\gamma c}$ and, we multiply by $a(\theta)$ the individual utility in the objective function so that the social objective is

$$\Phi(U;w,\theta) \stackrel{\text{def}}{\equiv} \widetilde{\Phi}\left(a\left(\theta\right)\cdot U;w,\theta\right).$$

Our model then applies with these redefined individual and social preferences.

II.2.f Optimal Nonlinear Price Schedule

To illustrate that our approach can be applied beyond optimal tax problems, we now reinterpret our model in the context of the nonlinear pricing framework where a principal observes a one-dimensional action, namely how much of a commodity the agent buys, and where unobserved individual characteristics are multidimensional.[13] Laffont, Maskin, and Rochet (1987)[14] study the problem of a monopolist (the principal) who sells a single product and needs to determine a nonlinear price schedule observing only a one-dimensional action (how much consumers are demanding), while consumers differ both in the slope and in the intercept of their demand functions. The latter correspond to w and θ in our model. The authors derive the optimal quantity assignment function when these two characteristics are independently and uniformly distributed and under restrictive assumptions on preferences (they are assumed linear in income and quadratic in consumption). Our approach can then be used to generalize this nonlinear pricing model under less restrictive assumptions on the distributions of characteristics and on preferences. To do so, reinterpret c as the amount of goods bought by any consumer and y as the amount she pays for the c units. Equation (1) gives the preferences of consumer (w,θ). The firm proposes a nonlinear price schedule that expresses payment y as a nonlinear function of quantity through $y = P(c)$. The firm only values the profit $y - c$ made on each consumer, so the social transformation is $\Phi\left(U;w,\theta\right) \equiv 0$. For a consumer of type (w,θ), the first-order condition of the program $\max_c \mathscr{U}\left(c, P\left(c\right); w,\theta\right)$ is:

$$\frac{1}{P_c'\left(c(w,\theta)\right)} = -\frac{\mathscr{U}_y(c,y;w,\theta)}{\mathscr{U}_c(c,y;w,\theta)}$$

where the retention rate $1 - T'(y(w,\theta))$ we have in (5) is replaced by the inverse of the marginal price $P_c'(c(w,\theta))$. The reinterpretation of our results is then straightforward.

III Incentive-Compatible Allocations

In this section, we characterizes incentive-compatible allocations when unobserved individual characteristics (w,θ) are multidimensional. We start by stating the incentive constraints. Since the individual's objective (4) is maximized for $y = Y(w,\theta)$, we have:

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

[13] Note that we restrict our attention in this paper to deterministic mechanisms. This restriction is natural in the context of optimal income taxation but may be less natural in other contexts.

[14] See Wilson (1993) for a survey of the literature.

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

$$\forall (w, \widetilde{w}, \theta, \widetilde{\theta}) \in \mathbb{R}_+^2 \times \Theta^2 \quad \mathscr{U}\left(C(w,\theta), Y(w,\theta); w, \theta\right) \geq \mathscr{U}\left(C(\widetilde{w}, \widetilde{\theta}), Y(\widetilde{w}, \widetilde{\theta}); w, \theta\right). \tag{11}$$

Equation (11) 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 (4) into account or to directly select an allocation satisfying the incentive-compatible constraints (11). We follow the second approach and characterize the set of incentive-compatible allocations in two steps. We first characterize incentive-compatible allocations $w \mapsto (Y(w,\theta), C(w,\theta))$ within each group θ. In this step, the within-group single-crossing condition (Assumption 1) enables to retrieve the properties that are usual when unobserved heterogeneity is one-dimensional like in Mirrlees (1971). The novelty lies in the second step where we characterize how these within-group allocations need to be set to ensure overall incentive-compatibility, i.e when considering (11) for $\tilde{\theta} \neq \theta$.

Within-Group Incentive Constraints

An incentive-compatible allocation has to satisfy (11). It thus has to verify for each group θ the following set of "within-group incentive constraints":

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

For each θ, characterizing the within-group allocations $w \mapsto (C(w,\theta), Y(w,\theta))$ that verify the within-group incentive constraints (12) is the same problem as characterizing incentive compatible allocations when unobserved heterogeneity is one-dimensional. This is due to the within-group single-crossing assumption 1. Under the Spence-Mirrlees single-crossing condition, the set of incentive constraints can be transformed into two much simpler local conditions, a monotonicity constraint and a differential equation, without any loss of generality (Mirrlees, 1971). When heterogeneity is multidimensional, we retrieve the monotonicity constraint and differential equation 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 which relies on the within-group incentive constraints and within-group single-crossing assumption. 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 the tax literature to consider only smooth allocations where these two "pathologies" do not arise and to follow the so-called "first-order approach". We thus make the following smoothness assumption:

Assumption 2 (Smooth allocations). *For each θ, $w \mapsto Y(w,\theta)$ is differentiable with a strictly positive derivative and maps $\mathbb{R}_+$ onto $\mathbb{R}_+$.*

Assumption 2 rules out bunching, the absence of which is validated by our simulations in Section VI. It also assumes that $Y(w,0) = 0$ and $\lim_{w\mapsto\infty} Y(w,\theta) = \infty$, so that the support of the income distribution for each group is the entire positive real line.[15] Assumption 2 therefore implies that pooling is unavoidable. Moreover, Assumption 2 is very natural under the isoelastic individual preferences (9). The following lemma provides the differential equation. It also shows that the marginal rate of substitution between income and consumption has to be equal to the ratio $\dot{C}(w,\theta)/\dot{Y}(w,\theta)$.[16]

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

$$\dot{U}(w,\theta) = \mathscr{U}_w(C(w,\theta), Y(w,\theta); w,\theta) = -v_w\left(Y(w,\theta); w,\theta\right). \tag{13a}$$

Moreover, Equation (13a) *is equivalent to:*

$$\frac{\dot{C}(w,\theta)}{\dot{Y}(w,\theta)} = \mathscr{M}\left(C(w,\theta), Y(w,\theta); w,\theta\right). \tag{13b}$$

The proof, which is very standard, can be found in Appendix A.2. Integrating Equation (13a) leads to:

$$U(w,\theta) = U(0,\theta) - \int_0^w v_w\left(Y(t,\theta); t,\theta\right)\, dt. \tag{13c}$$

If the government was 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 (12).[17] The observation of θ would then reduce the set of incentive constraints and increase 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, Gahvari, and Lozachmeur (2010), Mankiw and Weinzierl (2010)). In contrast, our paper does not consider tagging so that the government does not condition taxes on the vector of individual characteristics θ. Therefore, we need now to describe how the various within-group allocations $\omega \mapsto (Y(\omega,\theta), C(\omega,\theta))$ coexist to verify the full set of incentive constraints (11).

[15] One may instead assume the existence of a reference group θ_0 such that for all $\theta \neq \theta_0$, $Y(w,\theta_0) \leq Y(w,\theta)$ and $\lim_{w\mapsto\infty} Y(w,\theta) \leq \lim_{w\mapsto\infty} Y(w,\theta_0)$. Assuming the existence of such a reference group for which the support of the income distribution includes the support of the income distribution in all the other groups would not substantially change the results of the paper while it would add notational complexity.

[16] We henceforth use a dot to denote the derivatives with respect to w of functions $Y(\cdot,\theta)$, $C(\cdot,\theta)$ and $U(\cdot,\theta)$.

[17] 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 is in practice ruled out for horizontal equity reasons, preventing the government from using an information that would otherwise improve the equity-efficiency trade-off (Alesina, Ichino, and Karabarbounis, 2011). A similar issue arises when conditioning income taxation on individuals' height (Mankiw and Weinzierl, 2010).

Pooling Types across θ-Groups at each Income Level

In our context of multidimensional heterogeneity, each level of income y is obtained by individuals belonging to different groups θ. The cornerstone of our method is the characterization for each group θ of the skill level w of individuals who earn the same income level. Choose a reference group $\theta_0 \in \Theta$, a skill level w and another group θ. Individuals of type (w, θ_0) earn income $Y(w, \theta_0)$. According to the smoothness Assumption 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 into 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 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. The pooling function $W(.,.)$ characterizes how the distinct within-group allocations $\omega \mapsto (Y(\omega, \theta), C(\omega, \theta))$ need to be set to ensure that the overall allocation satisfies the entire set of incentive compatible conditions (11). For each $\theta \in \Theta$, the pooling function combines two one-to-one differentiable mappings, with a strictly positive derivative everywhere, 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 one-to-one differentiable mapping with a strictly positive derivative everywhere. It obviously verifies $W(w, \theta_0) = w$.

We now characterize the pooling function from the allocation designed for a reference group $\omega \mapsto (Y(\omega, \theta_0), C(\omega, \theta_0))$. We have $Y(W(w, \theta), \theta) \equiv Y(w, \theta_0)$ from the definition of the pooling function. Provided that the allocation is incentive-compatible, it is not possible that individuals with characteristics $(W(w, \theta), \theta)$ 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:

$$Y(W(w,\theta),\theta) \equiv Y(w,\theta_0) \qquad \text{and} \qquad C(W(w,\theta),\theta) \equiv C(w,\theta_0). \tag{14}$$

These simultaneous equalities induce that individuals of different θ-groups who pool at the same income level need to have the same marginal rate of substitution between income and consumption. This property is formally presented in 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 $W(w, \theta)$ is the unique solution in ω to*

$$\mathscr{M}\left(C(w,\theta_0), Y(w,\theta_0); w, \theta_0\right) = \mathscr{M}\left(C(w,\theta_0), Y(w,\theta_0); \omega, \theta\right). \tag{15a}$$

Proof: According to Assumption 1, Equation (15a) admits exactly one solution in ω. Differentiating in w both sides of each two equalities in (14) 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,

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

If the allocation is incentive-compatible, then, according to Lemma 2, Equation (13b) holds, which implies:

$$\mathscr{M}\left(C(w,\theta_0),Y(w,\theta_0);w,\theta_0\right)=\mathscr{M}\left(C(w,\theta_0),Y(w,\theta_0);W(w,\theta),\theta\right). \tag{15b}$$

□

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 (5) they must face the same marginal rate of substitution. A key point here is that, because of the within-group single-crossing condition (Assumption 1), Equation (15a) admits exactly one solution in ω. Hence, Equation (15a) fully characterizes the pooling function $W(\cdot,\theta)$.

The following lemma, which is proved in Appendix A.3, shows that once an incentive-compatible allocation is set for the reference group θ_0, the allocation for another group θ is determined by the equality between their marginal rates of substitution in Equation (15b). This equality is thus critical to guarantee that all incentive-compatible constraints are satisfied.

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* (13b)*. 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* (15a)*. There exists a unique incentive-compatible allocation $(w,\theta)\mapsto(\underline{C}(w,\theta),\underline{Y}(w,\theta))$ whose restriction to group θ_0 is $w\mapsto(C(w,\theta_0),Y(w,\theta_0))$ and that verifies Assumption 2 if and only if, for each θ, $\underline{W}(\cdot,\theta)$ maps $\mathbb{R}_+$ into $\mathbb{R}_+$ and admits a positive derivative $\dot{\underline{W}}(w,\theta)>0$ everywhere.*

The assumption that $\underline{W}(\cdot,\theta)$ admits a positive derivative everywhere plays in our analysis a role similar to the "first-order approach" in the Mirrleesian literature with a one-dimensional 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 (i.e. Equation (15b)) 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:

$$v_y\left(Y(w,\theta_0);w,\theta_0\right)=v_y\left(Y(w,\theta_0);W(w,\theta),\theta\right). \tag{15c}$$

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. 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, which is standard in the literature.

Consider, as an illustration, the case of the isoelastic preferences proposed in (9). The equality in Equation (15c) implies that the pooling function is:[18]

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

The pooling function thus depends on the choice of $Y(\cdot,\theta_0)$ whenever the different groups are endowed with distinct direct labor supply elasticities (i.e. when $\theta \neq \theta_0$). Note that if Θ is compact and if, for the reference group, one takes $\theta_0 = \max\{\theta \in \Theta\}$, it is then sufficient to have that $w \mapsto Y(w,\theta_0)$ verifies Assumption 2 to ensure that the pooling function $w \mapsto W(.,\theta)$ is differentiable, with a positive derivative everywhere and maps $\mathbb{R}_+$ into $\mathbb{R}_+$. Lemma 4 is then valid and Assumption 2 is directly satisfied for all groups.

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. Therefore, the pooling function does not depend on the chosen allocation, unlike in our model. We argue that this restriction 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:

$$\max_{y} \qquad \mathscr{U}(c,y;\Xi(\mathbf{t})). \tag{16}$$

All individuals with the same $w = \Xi(\mathbf{t})$ are therefore making the same intensive decisions and the pooling function is simply obtained by inverting the aggregator $\Xi(\cdot)$. Importantly, it does not depend on the chosen variables $Y(.,.)$ and $C(.,.)$. Since all individuals who pool at the same income level are characterized by the same $\Xi(\mathbf{t})$, they solve the same intensive program (16) and are therefore equally responsive to tax reforms.

Brett and Weymark (2003), Boadway, Marchand, Pestieau, and del Mar Racionero (2002), Choné and Laroque (2010), Lockwood and Weinzierl (2014) 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 matters for the computation of social marginal weights in Boadway, Marchand, Pestieau, and del Mar Racionero (2002), Choné and Laroque (2010), Lockwood and Weinzierl (2014).

Rothschild and Scheuer (2013, 2014a,b), Scheuer (2013, 2014) and Gomes, Lozachmeur, and Pavan (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

[18] Substituting (9) in (15c) yields $Y(w,\theta_0)^{1/\theta_0}/w^{(1+\theta_0)/\theta_0} = Y(w,\theta_0)^{1/\theta}/W(w,\theta)^{(1+\theta)/\theta}$.

individuals in each sector composes the vector of unobserved characteristics $\mathbf{t}$. 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 $\mathbf{t}$ depends only on a one-dimensional aggregation $\Xi(\mathbf{t};\mathbf{p})$ of types $\mathbf{t}$ and of prices $\mathbf{p}$, 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, Kreiner, and Saez, 2009, Jacquet, Lehmann, and Van der Linden, 2013) or in optimal income tax models with migration (Blumkin, Sadka, and Shem-Tov, 2014, Lehmann, Simula, and Trannoy, 2014), individuals differ in skills and in costs of participation (migration). 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.

IV Optimal Structural Tax Formula

In this section, we derive the optimal marginal tax rates as a function of the policy-invariant primitives of the model, which are the individual $\mathscr{U}(\cdot,\cdot;w,\theta)$ and social $\Phi(\cdot;w,\theta)$ preferences and the distributions of characteristics $f(\cdot|\theta)$ and $\mu(\cdot)$. We start by stating the maximization program the government faces. This program cannot be solved with the usual Hamiltonian. Therefore, to obtain the first-order conditions, we propose a new method that relies on a specific perturbation of the optimal allocation.

The government maximizes social preferences (7) under its budget constraint (6) within the subset of incentive-compatible allocations which satisfy (11). For the sake of clarity, we denote $\mathscr{C}(\hat{u},y;w,\theta)$ the consumption level the government needs to provide to a worker of type (w,θ) who earns y to ensure her with the utility level $\hat{u}$. Function $\mathscr{C}(\cdot,y;w,\theta)$ is the reciprocal of $\mathscr{U}(\cdot,y;w,\theta)$ and we have:

$$\mathscr{C}_u(\hat{u},y;w,\theta) = \frac{1}{u'(c)} \qquad \text{and} \qquad \mathscr{C}_y(\hat{u},y;w,\theta) = \frac{v_y(y;w,\theta)}{u'(c)} \tag{17}$$

where the various derivatives are evaluated at $c = \mathscr{C}(\hat{u},y;w,\theta)$. The Lagrangian multiplier associated to the government's budget constraint (6) is denoted λ. The Lagrangian $\mathscr{L}$ of the government's problem is defined as:

$$\mathscr{L} \stackrel{\text{def}}{\equiv} \iint \left[Y(w,\theta) - \mathscr{C}\left(U(w,\theta),Y(w,\theta);w,\theta\right) + \frac{\Phi\left(U(w,\theta);w,\theta\right)}{\lambda} \right] f(w|\theta)dw\, d\mu(\theta). \tag{18}$$

The government's problem consists in maximizing $\mathscr{L}$ within the subset of incentive-compatible allocations that verify (11), and in adjusting the Lagrange multiplier λ to ensure that the budget constraint (6) is satisfied. Following the usual first-order approach, we consider a "relaxed" problem where the government maximizes over the set of allocations that verify for each group the within-group incentive constraints (13a) and the pooling condition (15c). Whenever the solution to this relaxed problem verifies Assumption 2, it also solves the overall program, according to Lemma 4.

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 (15c) induces constraints on state and control variables which hold at endogenous skill levels. This prevents one from using the Hamiltonian approach. We therefore develop a new method that relies on a specific perturbation of the optimal allocation that we now present.

Prototypical Allocation Perturbation

We investigate the effects of a small perturbation of the optimal allocation specific to group θ: In group θ, the income $Y(.,\theta)$ is increased by small amounts of average ΔY for skill levels ω in a small interval $[w-\delta w, w]$, as shown in the upper panel of Figure 1. Because of the within-group incentive constraint (13c), this prototypical perturbation requires, for all skill levels above w, to uniformly increase utility $U(\omega,\theta)$ by:[19]

$$\Delta U = -v_{yw}(Y(w); w, \theta) \cdot \Delta Y \cdot \delta w \qquad \Leftrightarrow \qquad \Delta Y \cdot \delta w = \frac{\Delta U}{-v_{yw}(Y(w); w, \theta)} \tag{19}$$

This is illustrated in the lower panel of Figure 1. Note that the perturbation does not modify the levels of utility of people with skills below $w-\delta w$. The following definition summarizes the prototypical allocation perturbation we just described.

Definition 1. *A prototypical allocation perturbation $\mathscr{P}(\Delta U, \delta w; w, \theta)$ is the following infinitesimal change in the allocation specific to group θ. Income $Y(\cdot,\theta)$ is changed only for skill levels in the small interval $[w-\delta w, w]$, by an average amount ΔY that verifies* (19)*. Such a perturbation does not trigger any change in utility for skill levels below $w-\delta w$. It implies a uniform change in utility equal to ΔU for all skill levels above w to maintain incentive compatibility* (13c) *within-group θ.*

To obtain the optimal structural tax formula, we compute the first-order effects such a prototypical allocation perturbation $\mathscr{P}(\Delta U, \delta w; w, \theta)$ has on the Lagrangian (18). 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 $c = C(w,\theta)$.

Lemma 5. *A perturbation $\mathscr{P}(\Delta U, \delta w; w, \theta)$ induces a first-order change in the Lagrangian* (18) *equal to:*

[19] This equality is only valid as a first-order approximation when $\omega \mapsto Y(\omega,\theta)$ is continuous in the left of $\omega = w$ and δw and ΔY are small enough to consider that for $(y,\omega) \in [Y(w-\delta w) - |\delta_Y|, Y(w) + |\delta_Y|] \times [w-\delta w, w]$, approximating $v_{yw}(y;\omega,\theta)$ by $v_{yw}(Y(w,\theta); w, \theta)$ is a second-order error.

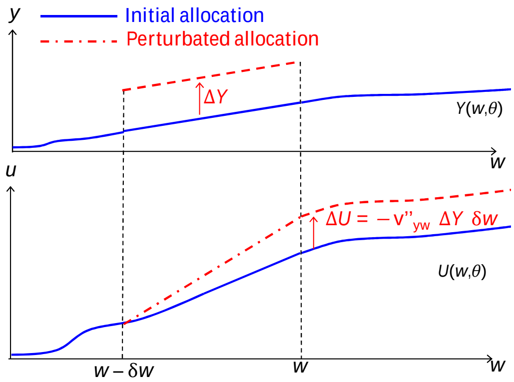

Figure 1: A perturbation of the optimal allocation

$$\left\{\frac{T'\langle w,\theta\rangle}{1-T'\langle w,\theta\rangle}\frac{v_y\langle w,\theta\rangle}{-w\,v_{yw}\langle w,\theta\rangle}\frac{w\,f(w|\theta)}{u'\langle w,\theta\rangle}+\int_w^\infty\left(\frac{\Phi_U\langle x,\theta\rangle}{\lambda}-\frac{1}{u'\langle x,\theta\rangle}\right)f(x|\theta)dx\right\}d\mu(\theta)\;\Delta U. \tag{20}$$

Proof: The perturbation $\mathscr{P}(\Delta U,\delta w;w,\theta)$ implies two first-order changes on the Lagrangian (18). First, it modifies $\omega\mapsto Y(\omega,\theta)$ only within $[w-\delta w,w]$. Using the second equality in (17), the change in $Y(\cdot,\theta)$ implies a first-order change in the Lagrangian (18) equal to:

$$\left(1-\frac{v_y\langle w,\theta\rangle}{u'\langle w,\theta\rangle}\right)\Delta Y\;f(w|\theta)\delta w\;d\mu(\theta) = \frac{1-\dfrac{v_y\langle w,\theta\rangle}{u'\langle w,\theta\rangle}}{-v_{yw}\langle w,\theta\rangle}\,f(w|\theta)d\mu(\theta)\;\Delta U \tag{21a}$$

$$= \frac{T'\langle w,\theta\rangle}{1-T'\langle w,\theta\rangle}\frac{v_y\langle w,\theta\rangle}{-w\,v_{yw}\langle w,\theta\rangle}\frac{w\,f(w|\theta)}{u'\langle w,\theta\rangle}\,d\mu(\theta)\;\Delta U$$

where the right-hand side of the top line is obtained using (19), and the second line is obtained using (5). Second, the perturbation $\mathscr{P}(\Delta U,\delta w;w,\theta)$ does not modify $\omega\mapsto U(\omega,\theta)$ below $w-\delta w$ but it modifies by ΔU the utility levels $U(x,\theta)$ of individuals with skills $x\geq w$. Using the first equality in (17), the change in $U(\cdot,\theta)$ implies a first-order[20] change in the Lagrangian (18) equal to:[21]

$$\int_w^\infty\left(\frac{\Phi_U\langle x,\theta\rangle}{\lambda}-\frac{1}{u'\langle x,\theta\rangle}\right)f(x|\theta)dx\;d\mu(\theta)\;\Delta U. \tag{21b}$$

Adding (21a) and (21b) yields (20). □

Optimal Structural Tax Formula

We now characterize the optimal structural tax formula which expresses the optimal marginal tax rates as a function of the primitives of the model:

[20]The change in utility levels $U(\omega,\theta)$ for skill levels ω within $[w-\delta w,w]$ has an absolute valued bounded by $\delta w\;|\Delta U|$. It is thus of second-order, provided that the size δw of the skill interval is small enough.

[21]To follow the non-welfarist approach of Saez and Stantcheva (2013), one only needs to substitute their generalized social welfare weights for the term $\frac{\Phi_U<x,\theta>}{\lambda}$.

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

$$\frac{T'\langle w,\theta_0\rangle}{1-T'\langle w,\theta_0\rangle}\cdot\int_{\theta\in\Theta}\frac{v_y\langle W(w,\theta),\theta\rangle}{-W(w,\theta)\,v_{yw}\langle W(w,\theta),\theta\rangle}\,W(w,\theta)\,f(W(w,\theta)|\theta)\,d\mu(\theta)$$
$$=\quad u'\langle w,\theta_0\rangle\cdot\iint_{\theta\in\Theta,x\geq W(w,\theta)}\left(\frac{1}{u'\langle x,\theta\rangle}-\frac{\Phi_u\langle x,\theta\rangle}{\lambda}\right)f(x|\theta)dx\,d\mu(\theta)\tag{22a}$$

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

$$\iint_{\theta\in\Theta,x\in\mathbb{R}_+}\left(\frac{\Phi_u\langle x,\theta\rangle}{\lambda}-\frac{1}{u'\langle x,\theta\rangle}\right)f(x|\theta)dx\,d\mu(\theta)=0.\tag{22b}$$

Proof: We consider a prototypical perturbation $\mathscr{P}(\Delta U,\delta w(\theta_0);w,\theta_0)$ in the reference group θ_0. We first determine how this perturbation affects the allocations in the other groups $\theta\neq\theta_0$. The function $\omega\mapsto Y(\omega,\theta_0)$ is modified only within the interval $[w-\delta w(\theta_0),w]$. So, according to Equation (15c), this perturbation does not modify the pooling function $\omega\mapsto W(\omega,\theta)$ outside the interval $[w-\delta w(\theta_0),w]$. Therefore, $\omega\mapsto Y(\omega,\theta)$ is not modified outside the interval $[W(w-\delta w(\theta_0),\theta),W(w,\theta)]$. The perturbation $\mathscr{P}(\Delta U,\delta w(\theta_0);w,\theta_0)$ in the reference group thus triggers a perturbation of the prototypical form $\mathscr{P}(\Delta U(\theta),\delta w(\theta);W(w,\theta),\theta)$ in all other groups θ, where $\delta w(\theta)$ is given by $W(w,\theta)-\delta w(\theta)=W(w-\delta w(\theta_0),\theta)$.

We now need to determine the size $\Delta U(\theta)$ of the perturbations $\mathscr{P}(\Delta U(\theta),\delta w(\theta_0);w,\theta_0)$ in the other groups. To do so, consider two individuals, one in the reference group θ_0 with a skill level x above w and one in another group θ with a skill level $W(x,\theta)$. Both individuals pool at the same income level $Y(x,\theta_0)$ both before and after the perturbations. The perturbations do not modify their income, but modify their utility levels by ΔU for the individual in the reference group and by $\Delta U(\theta)$ for the other individual. These changes in utility occur only through changes in consumption levels, which need to be identical to preserve incentive compatibility. Hence, one must have $\Delta U(\theta)=\Delta U$. Therefore, a perturbation $\mathscr{P}(\Delta U,\delta w(\theta_0);w,\theta_0)$ triggers a perturbation $\mathscr{P}(\Delta U,\delta w(\theta);W(w,\theta),\theta)$ in each of the other groups.

To determine the total effect the prototypical perturbation in group θ_0 has on the Lagrangian, we sum (20) across all groups. If the initial allocation is optimal, this perturbation must imply no first-order effect on the Lagrangian, which leads to (22a) since $T'\langle W(w,\theta)\rangle$ and $u'\langle W(w,\theta)\rangle$ are identical across individuals who pool at the same income $Y(w,\theta_0)$.

To derive (22b), we consider a perturbation of the optimal allocation that uniformly increases $u(C(w,\theta))$ by an amount ΔU for all types $(w,\theta)\in\mathbb{R}_+\times\Theta$ without perturbing income $Y(x,\theta)$. This perturbation maintains incentive-compatibility (11). Using the first equality in (17), it affects the Lagrangian (18) by:

$$\int_{\theta\in\Theta}\left\{\int_{x\in\mathbb{R}_+}\left(\frac{\Phi_U\langle x,\theta\rangle}{\lambda}-\frac{1}{u'\langle x,\theta\rangle}\right)f(x|\theta)dx\right\}d\mu(\theta)\cdot\Delta U.$$

This impact has to be equal to zero at the optimum, which leads to (22b). □

The optimal formulas of Proposition 1 depend only on structural parameters (the utility function and its derivatives, the social welfare function and the distributions of characteristics). 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]

Equation (22a) states that any prototypical tax perturbation induces zero first-order effect on the government's Lagrangian. The term $v_y/(-w\ v_{yw})$ under the integral in the left-hand side of (22a) is equal to $\theta/(1+\theta)$ when the preferences are isoelastic as in (9), in which case it becomes policy-invariant. Equation (22b) measures the welfare impact of uniformly increasing the utility of all individuals. It thus plays the role of the transversality condition at the bottom of the distribution in providing the initial value of the right-hand side of (22a). In combination with the budget constraint (6), it determines the Lagrange multiplier of the budget constraint λ and the initial value of $U(0,\theta_0)$.

The tax formula of Proposition 1 generalizes to multidimensional individual characteristics the structural optimal income tax formula derived by Mirrlees (1971). When the unobserved heterogeneity has only one dimension, Equations (22a) and (22b) simplify to:

$$\frac{T'\langle w\rangle}{1-T'\langle w\rangle}\cdot\frac{v_y\langle w\rangle}{-w\ v_{yw}\langle w\rangle}\ w\ f(w) = u'\langle w\rangle\int_w^\infty\left(\frac{1}{u'\langle x\rangle}-\frac{\Phi_U\langle x\rangle}{\lambda}\right)f(x)\,dx \tag{23a}$$

$$0 = \int_0^\infty\left(\frac{1}{u'\langle x\rangle}-\frac{\Phi_U\langle x\rangle}{\lambda}\right)f(x)\,dx. \tag{23b}$$

The literature obtains these necessary conditions in constructing a Hamiltonian and applying the Pontryagin principle, as we do in Appendix A.4. Comparing these equations with Equations (22a) and (22b) 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.

Signing Optimal Marginal Tax Rates

With multidimensional heterogeneity, the literature has highlighted that negative marginal tax rates can become optimal. In Boadway, Marchand, Pestieau, and del Mar Racionero (2002), Choné and Laroque (2010) and Lockwood and Weinzierl (2014), individuals differ along their skills and preferences for effort, and the social planner has weighted utilitarian preferences (see (8b)). In this context, individuals who pool at the same income level $Y(w,\theta_0)$ are characterized by different social marginal utilities of consumption $\Phi_U(U(W(w,\theta));w,\theta)\cdot u'(C(W(w,\theta)))$. The social marginal utility of consumption is decreasing in skill *within each group* θ due to the concavity of the social welfare function. However, the average social marginal utility of consumption may not be decreasing in income because it requires to aggregate the social marginal

[22]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." (Saez, 2001, p. 223)

utilities of individuals across groups. This *composition effect* in the average social marginal utilities may reduces marginal tax rates (Lockwood and Weinzierl, 2014) and may even induce them to become negative (Boadway, Marchand, Pestieau, and del Mar Racionero, 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 income distribution.

The following proposition points out that in the absence of composition effects in the social welfare weights, the marginal tax rates cannot be negative. We investigate the sign of marginal tax rates with Benthamite preferences (see (8a)) and under maximin. With these preferences, composition effects (in the social welfare weights) cannot exist.

Proposition 2. *Under Benthamite social preferences* (8a) *and maximin, optimal marginal tax rates are positive.*

Proof: Let $u' \langle w, \theta_0 \rangle \, I(w)$ denote the right-hand side of (22a). Under Benthamite preferences, $\Phi_u = 1$ and we get: $I(w) \overset{\text{def}}{\equiv} \int_{x \geq W(w,\theta)} \left(\frac{1}{u' \langle x,\theta \rangle} - \frac{1}{\lambda} \right) \cdot \left(\int_\theta f\left(x | \theta \right) d\mu\left(\theta\right) \right) \; dx$. 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, $\lim\limits_{w \mapsto \infty} I(w) = 0$ and Equation (22b) implies 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 Benthamite preferences.

Under maximin, one has $U(x, \theta) > U(0, \theta)$ for all $x > 0$ from (13a). This implies that, within each group, the most deserving individuals are those whose skill $w = 0$. The maximin objective therefore implies $\Phi_u \langle x, \theta \rangle = 0$ for all $x > 0$. Thereby, $I(w) \overset{\text{def}}{\equiv} \int_{x \geq W(w,\theta)} \frac{1}{u' \langle x,\theta \rangle} \cdot \int_\theta f\left(x | \theta \right) d\mu\left(\theta\right) \; dx$ for all $x > 0$, which again leads to positive marginal tax rates. □

Proposition 2 shows that introducing an endogenous pooling function does not invalidate the result of positive marginal tax rates found in Mirrlees (1971) with a one-dimensional heterogeneity and Benthamite social preferences. In our general framework, the only possibility for optimal marginal tax rates to be negative is if welfare weights increase with income due to a composition effect.

V Elasticity-Based Optimal Tax Formula

This section rearranges the first-order conditions for the government's problem displayed in Proposition 1 to obtain a characterization of the optimal marginal tax rates in terms of sufficient statistics. This is what we call an elasticity-based optimal tax formula.

Individual Behavioral Elasticities

First, we define a set of individual elasticities. To this aim, we consider a specific compensated tax reform around income $Y(w, \theta)$. This reform changes the marginal tax by a constant amount τ around $Y(w,\theta)$, while leaving unchanged the level of tax at income $Y(w,\theta)$. The tax function is then changed to $T(Y) - \tau \cdot (Y - Y(w,\theta))$. The income response effect is defined as the response to a small lump-sum change ρ in tax liability, so the tax function becomes $T(Y,\theta) - \tau(Y - Y(w,\theta)) - \rho$. Individuals of type (w,θ) then solve the following program:

$$\max_y \quad u\left(y - T(y) + \tau(y - Y(w,\theta)) + \rho\right) - v(y;w,\theta). \tag{24}$$

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

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

The second-order condition is $\mathscr{Y}_y(Y(w,\theta),0,0;w,\theta) \leq 0$ with:

$$\mathscr{Y}_y(Y,0,0;w,\theta) = -T''(Y) \cdot u'(Y - T(Y)) + (1-T')^2 \cdot u''(Y - T(Y)) - v_{yy}(Y;w,\theta). \tag{25}$$

Following the literature that gives the optimal tax formula in terms of sufficient statistics, we need to assume further regularity conditions in addition to Assumptions 1 and 2:

Assumption 3. *The tax function $T(\cdot)$ is twice differentiable and, for all $(w,\theta) \in \mathbb{R}_+ \times \Theta$, the second-order condition holds strictly: $\mathscr{Y}_y\left(Y(w,\theta),0,0;w,\theta\right) < 0$.*

Because we assume $\mathscr{Y}_y\left(Y(w,\theta),0,0;w,\theta\right) < 0$, we can apply the implicit function theorem to $\mathscr{Y}(Y(w,\theta),0,0;w,\theta) = 0$. Provided that the sizes of the changes in w, τ and ρ are small enough for the maximum of program (24) to change only marginally, one has for $x = w,\tau,\rho$, that $\partial Y/\partial x = -\mathscr{Y}_x/\mathscr{Y}_Y$ evaluated at $(Y(w,\theta),0,0;w,\theta)$. This leads directly to the following three definitions. The total compensated elasticity of income with respect to the retention rate $1 - T'(.)$ is defined as:[23]

$$\varepsilon(w,\theta) \stackrel{\text{def}}{\equiv} \frac{1 - T'\left(Y(w,\theta)\right)}{Y(w,\theta)} \frac{\partial Y}{\partial \tau} = -\frac{v_y}{Y(w,\theta) \cdot \mathscr{Y}_y} > 0 \tag{26a}$$

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

$$\alpha(w,\theta) \stackrel{\text{def}}{\equiv} \frac{w}{Y(w,\theta)} \dot{Y}(w,\theta) = \frac{w\, v_{yw}}{Y(w,\theta)\, \mathscr{Y}_y} > 0 \tag{26b}$$

which is positive because of the within-group single-crossing condition 1 and because $\mathscr{Y}_y < 0$. In sum, Assumptions 1 and 3 imply that $w \mapsto Y(w,\theta)$ is differentiable with a positive derivative, which is a part of Assumption 2. In this sense, Assumption 3 is more demanding in terms of regularity than Assumption 2. Note that we did not need Assumption 3 to obtain the structural tax formula in Section IV. The total income response effect to a lump-sum change in tax liability is defined as:

$$\eta(w,\theta) \stackrel{\text{def}}{\equiv} \frac{\partial Y}{\partial \rho} = -\frac{u'' \cdot v_y}{u' \cdot \mathscr{Y}_y} \leq 0 \tag{26c}$$

[23]When defining this elasticity, we assume that the tax level is unchanged at earnings level $Y(w,\theta)$ so that we call it a "compensated "elasticity.

which is non-positive due to the additive separability of individual preferences (1). Leisure is therefore a normal good.

Elasticities and income response (26a)-(26c) differ from those in the optimal tax literature by the presence, in their denominators, of a term $T''(Y(w,\theta)) \cdot u'(C(w,\theta))$ which is incorporated in $\mathscr{Y}_Y$ (see Equation (25)). This term accounts for the nonlinearity of the income tax schedule. An exogenous change in either w, τ, or ρ induces a *direct* change in earnings $\Delta_1 Y(w,\theta)$. However, this change in turn modifies the marginal tax rate by $\Delta_1 T' = T''(Y(w,\theta)) \times \Delta_1 Y(w,\theta)$, thereby inducing a further change in earnings $\Delta_2 Y(w,\theta)$. Therefore, a circular process takes place: The income level determines the marginal tax rate through the tax function, and the marginal tax rate affects the income level through the substitution effect. Our definitions of behavioral 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 $T''(Y(w,\theta)) \cdot u'(C(w,\theta)))$ testifies about this. The literature instead considers only the direct effects by assuming that marginal tax rates are exogenous in the computation of behavioral and income responses, thereby taking $T''(Y(w,\theta)) = 0$ in Equations (26a)-(26c). In this case, the tax schedule is locally linear hence total and direct responses coincide.

Conditional and Unconditional Income Densities

We define $h(\cdot|\theta)$ as the conditional density of income y within group θ and we call $H(\cdot|\theta)$ its associated CDF. From Lemma 1, we know that $H(Y(w,\theta)|\theta) = F(w|\theta)$. Differentiating in skill w both sides of the previous equality and using (26b), we get:

$$h(Y(w,\theta)|\theta) = \frac{f(w|\theta)}{\dot{Y}(w,\theta)} \qquad \Leftrightarrow \qquad Y(w,\theta)\, h(Y(w,\theta)|\theta) = \frac{w\, f(w|\theta)}{\alpha(w,\theta)}. \tag{27}$$

Our pooling function $W(w,\theta)$ enable us to express the income density at any income level as a function of individual characteristics w and θ. The unconditional income density at income $y = Y(w,\theta_0)$ is the mass of individuals endowed with distinct w and θ who earn the same income $Y(w,\theta_0)$. It is equal to:

$$\hat{h}(Y(w,\theta_0)) \stackrel{\text{def}}{\equiv} \int_{\theta\in\Theta} h\left(Y(W(w,\theta),\theta)|\theta\right)\, d\mu(\theta). \tag{28}$$

The unconditional income density simply aggregate across all groups the conditional income densities of individuals who earn the same income level $Y(w,\theta_0)$.

Mean behavioral elasticities and mean marginal social weights

We can also define the mean total compensated elasticity at income level $Y(w,\theta_0)$ as:

$$\hat{\varepsilon}(Y(w,\theta_0)) \stackrel{\text{def}}{\equiv} \frac{\int_{\theta\in\Theta} \varepsilon\left(W(w,\theta)|\theta\right)\, h\left(Y(W(w,\theta),\theta)|\theta\right)\, d\mu(\theta)}{\int_{\theta\in\Theta} h\left(Y(W(w,\theta),\theta)|\theta\right)\, d\mu(\theta)}. \tag{29}$$

This is the mean of the total compensated elasticities $\varepsilon(w,\theta)$ across individuals who earn $Y(w,\theta_0)$. We can also define the mean total income effect at income level $Y(w,\theta_0)$ as:

$$\hat{\eta}(Y(w,\theta_0)) \stackrel{\text{def}}{\equiv} \frac{\int_{\theta\in\Theta} \eta\left(W(w,\theta)|\theta\right)\ h\left(Y(W(w,\theta),\theta)|\theta\right)\ d\mu(\theta)}{\int_{\theta\in\Theta} h\left(Y(W(w,\theta),\theta)|\theta\right)\ d\mu(\theta)}. \tag{30}$$

This is the mean of total income effects across all individuals who earn the same income $Y(w,\theta_0)$. In order to define the mean marginal social weight, we first define the (endogenous) marginal social weight associated with workers of type (w,θ), expressed in terms of public funds as:

$$g(w,\theta) \stackrel{\text{def}}{\equiv} \frac{\Phi'_u\left(U(w,\theta);w,\theta\right)\cdot \mathscr{U}'_c\left(C(w,\theta),Y(w,\theta);w,\theta\right)}{\lambda}. \tag{31}$$

Intuitively, the government values giving one extra dollar to a worker (w,θ) as a gain of $g(w,\theta)$ in terms of public funds. Using the latter definition, we can define the mean marginal social weight at income $Y(w,\theta_0)$:

$$\hat{g}(Y(w,\theta_0)) \stackrel{\text{def}}{\equiv} \frac{\int_{\theta\in\Theta} g\left(W(w,\theta)|\theta\right)\ h\left(Y(W(w,\theta),\theta)|\theta\right)\ d\mu(\theta)}{\int_{\theta\in\Theta} h\left(h\left(Y(W(w,\theta),\theta)|\theta\right)|\theta\right)\ d\mu(\theta)}. \tag{32}$$

It gives the average of the marginal social weights of people with distinct characteristics w and θ who earn the same income level $Y(w,\theta_0)$. Intuitively, it tells us how the government values giving one extra dollar to individuals who earn $Y(w,\theta_0)$.

Optimal Elasticity-Based Tax Formula

In Appendix A.5, we rearrange the first-order conditions (22a) and (22b) of our structural optimal tax formula displayed in Proposition 1 to obtain the optimal marginal tax rate in terms of the mean compensated elasticity, mean income effect, mean marginal social weights and the unconditional income density.

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

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

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

Equations (33a)-(33b) generalize the usual elasticity-based optimal tax formula with individuals who differ solely through a skill parameter to the case where individual characteristics are multi-dimensional. Shifting from the model with one-dimensional heterogeneity to the model with multi-dimensional heterogeneity entails replacing, in the tax formula, the marginal social weight, the behavioral and income responses by their means calculated over the individuals (w,θ) who earn the same income level. Importantly, we determine the conditions (Assumptions 1, 2 and 3) under which the elasticity-based tax formula in the one-dimensional context carries over to heterogeneous populations.

Our formulation of the optimal tax formula in (33a)-(33b) differs from the optimal tax formula in (Saez, 2001, Equation (19)) in the way the circularity process is taken into account. Saez (2001) was the first to understand that the circular process has to be taken into account to interpret the optimal conditions of Mirrlees (1971) in terms of empirically meaningful magnitudes. Instead of taking into account the circular process by considering the *total* compensated elasticity and *total* income effects as we do, Saez (2001) uses the standard definitions of these *direct* elasticities along a linear tax schedule, without the term $T'' \ u'$ in $\mathscr{Y}'_y$. His formula then has to rely on a *virtual* income density that corrects the real income density for this circularity process. This difference is a matter of presentation since the product of the real income density times our total compensated elasticity is equal to the product of the direct compensated elasticity times the virtual income density. We however prefer our formulation for the following reasons. First, we think the circularity process is also a behavioral response. Our definitions thus amount to using the total behavioral effects, including those induced by the circularity process and not only the direct effect when marginal tax rates are erroneously considered as independent of the labor supply decisions. Second, with multidimensional heterogeneity, we need to aggregate the different behavioral responses across individuals who pool at the same income level. In Equations (29) and (30), we weight total behavioral responses with real income density, which is natural. What should be the correct averaging procedure with Saez (2001)'s virtual income density remains unintuitive. Third, the optimal formula in Saez (2001) depends both on the virtual income density (when interacted with the compensated elasticity and the income effect) and on the true income density at the optimum (when interacted with social welfare weights), which looks very obscure at first glance.

Formula (33a) expresses the optimal tax rate in terms of three elements: one for behavioral responses to taxes, one for the shape of the income distribution and one for social preferences and income effects. We provide below the intuition behind these elements and emphasize the specificities driven by multidimensional heterogeneity.

In line with Ramsey (1927)'s inverted elasticity rule, optimal marginal tax rates are *ceteris paribus* larger in absolute terms at income levels where the mean compensated elasticity of labor supply $\hat{\varepsilon}(y)$ is lower. In our context, this elasticity is endogenous for different reasons. Firstly, it depends on the curvature of the income tax function, since the circularity process is encapsulated in the expression of the elasticity. This is also the case with one dimension of heterogeneity (Saez, 2001, Jacquet, Lehmann, and Van der Linden, 2013). Secondly, the set of types (w, θ) that pool at any income level is endogenous. Therefore, the mean elasticity $\hat{\varepsilon}(y)$ increases when, among individuals (w, θ) such that $Y(w, \theta) = y$, the proportion of individuals with an elasticity $\varepsilon(w, \theta)$ larger than the average $\hat{\varepsilon}(y)$ increases. This *composition effect* on the compensated elasticity is a novel insight of our setting.

According to the second term in the right hand-side of (33a), the optimal marginal tax rate at a given income y increases in absolute terms with the *distribution term* $(1 - \hat{H}(y))/(y\hat{h}(y))$ at that income level. Intuitively, a nonzero marginal tax rate at income level y distorts labor supply decisions of individuals who earn this income level. The larger the income density $\hat{h}(y)$ or the income level y, and the larger the induced distortions. Moreover, deviating from zero marginal tax rate at income level y modifies the tax liability for all the $1 - \hat{H}(y)$ individuals with income above y, which triggers the magnitude of the equity effects. Empirically, the distribution term varies with income except at the top. This is due to the well-known fact that the top tail is very closely approximated by a Pareto distribution (Diamond, 1998, Saez, 2001). Since the income distribution is typically unimodal, the distribution term is decreasing beyond the mode hence marginal tax rates are decreasing beyond the mode. Finally, it is worth stressing that the distribution term of the income distribution is endogenous, for two main reasons. The first is the endogeneity of the elasticity $\alpha(w,\theta)$ of income with respect to the level of skill due to the curvature of the tax function.[24] The second reason is that the set of individuals (w,θ) who pool at any income level is endogenous, so the distribution term may also be affected by a composition effect, as in Cremer, Gahvari, and Lozachmeur (2010) and Mankiw and Weinzierl (2010).

According to the third term of the right hand-side of (33a), optimal marginal tax rates vary with the mean of social welfare weights $\hat{g}(z)$ and the mean of income effects $\hat{\eta}(z) \cdot T'(z)$ for income levels z above y. The larger $\hat{g}(z)$, the more the government values the well-being of people at this level of income z hence, the lower should be the marginal tax rate they face. The larger the mean income effect $\hat{\eta}(z)(< 0)$ in absolute value for income z above y, the higher should be the marginal tax rate faced by individuals who earn y. Intuitively, an increase in the level of tax paid by workers with income higher than y induces them to work more through income effects. Note that the mean social welfare weight $\hat{g}(z)$ and the mean income effect $\hat{\eta}(z) \cdot T'(z)$ typically vary with earnings. One exception is under a maximin, where $\hat{g}(z) = 0$. Another exception is with quasilinear individual preferences which rule out income effects so that $\hat{\eta}(z) \cdot T'(z) = 0$. Composition effects are also an additional source of endogeneity for $\hat{g}(\cdot)$ and $\hat{\eta}(\cdot)$, as discussed in Boadway, Marchand, Pestieau, and del Mar Racionero (2002) and Choné and Laroque (2010).

Equation (33b) is the elasticity-based version of the transversality condition (22b). If income effects were assumed away, this condition implies that the weighted sum of social welfare weights is equal to 1. If conversely income effects are present, a uniform increase in tax liability triggers a positive income response hence a change in tax revenue proportional to the marginal tax rate which explains the presence of the term $\hat{\eta}(z)T'(z)$.

[24] From (27), we have:

$$\frac{1 - H(Y(w,\theta)|\theta)}{Y(w,\theta) \cdot h(Y(w,\theta)|\theta)} = \alpha(w,\theta) \cdot \frac{1 - F(w|\theta)}{w \cdot f(w|\theta)}$$

which links the distribution term based on the (exogenous) conditional skill density (in the right hand-side) and the distribution term based on the (endogenous) income density (in the left hand-side). Even when θ is homogeneous, the nonlinearity of the income tax schedule implies that the elasticity of income with respect to skill $\alpha(w,\theta)$ depends on the curvature of the tax schedule and is thus endogenous.

Optimal tax rates on top incomes

This section studies the implications of multidimensional heterogeneity for the optimal asymptotic marginal tax rates. To this aim, we follow, in this section, the usual assumptions that lead to the asymptotic tax formula of Piketty and Saez (2013). We consider iso-elastic individual preferences (see Equation (9)) and assume away income effects so that

$$\mathscr{U}(c, y; w, \theta) = c - \frac{\theta}{1+\theta}\left(\frac{y}{w}\right)^{\frac{1+\theta}{\theta}}. \tag{34}$$

We moreover assume that the social marginal weight is asymptotically nil (i.e. $\lim_{y\mapsto\infty}\hat{g}(y) = 0$). By taking the optimal tax formula (33a) to its limit for high income levels, we retrieve the formula of Piketty and Saez (2013) for the optimal asymptotic marginal tax rate:

$$\tau_* = \frac{1}{1+\hat{\varepsilon}_*\, p_*} \qquad \text{where} \qquad p_* = \lim_{y\mapsto\infty} \frac{1-\hat{H}_*(y)}{y\cdot \hat{h}_*(y)} \tag{35}$$

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

We now emphasize that the sufficient statistics $\hat{H}_*(y)$, $\hat{h}_*(y)$ and $\hat{\varepsilon}_*$ which are necessary to implement the optimal asymptotic marginal tax rate (35) are distinct from the sufficient statistics estimated in the actual economy. For this reason, calibrating the optimal asymptotic tax rates using (35) and estimations from the actual economy can be misleading. To illustrate this point, assume that, in the actual economy, the income density within group θ is described by a Pareto density of the form:

$$h_0(y|\theta) = k_\theta \cdot y^{-(1+p_\theta)} \tag{36}$$

where k_θ is the scale parameter and p_θ is the Pareto parameter with $p_\theta > 1$. Both parameters can vary across groups. First-order condition (5) and Equation (34) imply that individuals of type (w, θ) who face the asymptotic marginal tax rate τ earn income

$$y(w, \theta) = (1-\tau)^\theta w^{1+\theta}. \tag{37}$$

Inverting the latter expression, we can write the skill level of individuals belonging to group θ and earning income y in the optimal economy as

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

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:

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

Differentiating both sides of $H_*(y|\theta) = H_0(\tilde{Y}_0(y,\theta)|\theta)$ in y and using (36), we obtain a description of how the asymptotic income density is transformed between the actual economy that is used for calibration and the optimal one:

$$h_*(y|\theta) = k_\theta \cdot \left(\frac{1-\tau_*}{1-\tau_0}\right)^{\theta p_\theta} \cdot y^{-(1+p_\theta)}. \tag{38}$$

Within each group θ, income remains Pareto distributed asymptotically with the same Pareto parameter p_θ. This is because the tax function is asymptotically linear. Therefore, from (26b) and (25), the skill elasticity of income, $\alpha(w,\theta)$, is asymptotically identical in the actual and in the optimal economy (see also Foonote 24). However, the scale parameter of the Pareto density is different because individuals earning the same income in the actual economy where the tax rate is τ_0 and in the optimal economy where the tax rate is τ_* are not endowed with the same skill level. If the tax rate is higher in the optimal economy, a given income level corresponds to a higher skill level, which corresponds to a lower income density at the optimum (see Equation (38)). Because this shift in income density is triggered by labor supply responses, it is larger when the labor supply elasticity θ is larger (see (38)).

This difference in optimal and actual income densities results in distinct optimal and actual asymptotic compensated elasticities. Consider first the case where the Pareto parameter p_θ is identical across groups and equal to p. Then, according to (29) and (38), the optimal asymptotic compensated elasticity is given by:

$$\hat{\varepsilon}(y) = \int_{\theta\in\Theta} \theta \cdot \frac{k_\theta \cdot \left(\frac{1-\tau^*}{1-\tau_0}\right)^{\theta p}}{\int_{\widetilde{\theta}\in\Theta} k_{\widetilde{\theta}} \cdot \left(\frac{1-\tau^*}{1-\tau_0}\right)^{\widetilde{\theta} p} \cdot d\mu(\widetilde{\theta})} \cdot d\mu(\theta) \tag{39}$$

and differs from its value in the current economy. For instance, when the initial asymptotic tax rate τ_0 is lower than the optimal one τ_*, the rise of marginal tax rate from τ_0 to τ_* implies that a given income level is reached by individuals endowed with a higher skill level, as it can be deduced from (37). Along the upper tail of the Pareto income distribution, these individuals, who have a larger income, are less numerous. In other words, the conditional income density at this level of income is lower. This shift is stronger for groups with a higher compensated elasticity θ as it can be seen from (38). The rise in the asymptotic tax rate then reduces the relative weight of high-elasticity groups relative to low-elasticity groups which triggers a *composition effect* (see (39)). This effect reduces the asymptotic compensated elasticity (see (39)) and thereby increases the optimal asymptotic marginal tax rate (see (35)). The reverse occurs when the initial tax rate is above the optimal one.

This composition effect is however quantitatively limited. Consider for instance the case where the economy is composed by two groups of equal size, so that $k_{\theta_1}\mu(\theta_1) = k_{\theta_2}\mu(\theta_2)$. Consider that $\theta_1 = 0.2$ for the low-elasticity group and $\theta_2 = 0.6$ for the high-elasticity group. Moreover, assume the Pareto parameter is $p = 1.5$. Then, neglecting multidimensional heterogeneity leads to (39) being reduced to $\hat{\varepsilon}(y) = \theta$. In the absence of multidimensional heterogeneity, one does not observe two distinct values $\theta_1 = 0.2$ and $\theta_2 = 0.6$ but simply an average value of $\theta = 0.4$. The compensated elasticity is then $\hat{\varepsilon}(y) = 0.4$, and from (35), we obtain an optimal asymptotic marginal tax rate equal to $1/(1 + 1.5 \times 0.4) = 62.5\%$. Now if the asymptotic marginal tax rate in the actual economy is $\tau_0 = 40\%$, then the composition effect implies that the asymptotic compensated elasticity is only $\hat{\varepsilon}_* = 0.369$ and the optimal asymptotic marginal tax rate is equal to $\tau_* = 64.4\%$.[25] The latter is 1.9 percentage points higher than the optimal asymptotic tax rate obtained when neglecting this composition effect. Obviously, this difference shrinks as the asymptotic marginal tax rate in the actual economy gets closer to 62.5%.

Consider now that the Pareto parameter p_θ differs across groups. In this context, we show that taking into account the composition effect drastically modifies the optimal asymptotic tax rate. The group with the fatter upper tail (i.e. with the lowest Pareto parameter p_θ) has a share equal to 1 for asymptotically high income levels. The relevant elasticity for computing the optimal asymptotic marginal tax rate $\hat{\varepsilon}_*$ is therefore the one of the group with the fatter upper tail. The latter can be dramatically different from the one estimated from the average response among, say, the top 1%. Consider as an illustration the case where the economy is composed by two groups of equal size in the top 1% of the population with $\theta_1 = 0.2$ and $\theta_2 = 0.6$. 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, the optimal asymptotic marginal tax rate is $1/(1 + 1.5 \times 0.6) = 52.6\%$ (from (35)) instead of $1/(1 + 1.5 \times 0.4) = 62.5\%$ if one mistakenly calibrates the optimal tax formula from the mean compensated elasticity among the top 1%. If conversely the low-elasticity group has a Pareto parameter slightly below the high-elasticity group, the optimal asymptotic marginal tax rate is $1/(1 + 1.5 \times 0.2) = 76.2\%$. Therefore, heterogeneity in the asymptotic Pareto parameter across groups induce much substantial composition effects that can lead to very big difference in the optimal asymptotic marginal tax rates. 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, Slemrod, and Giertz (2012) and Piketty and Saez (2013).[26] 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.

[25] To obtain the values of τ_* and $\hat{\varepsilon}$, we numerically solve Equations (35) and (39).

[26] 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.

VI Numerical Illustration

This section numerically implements the optimal nonlinear income tax formula of Proposition 1, so as to emphasize the role played by multidimensional heterogeneity. It documents the quantitative impact on the optimal marginal tax rates of erroneously assuming identical behavioral elasticities across individuals who pool at the same income level. The numerical exercises we propose consider the case presented in Subsection II.2.d where individuals choose not only their income but also how much income they evade. The optimal tax profile is then derived for a given tax enforcement.

To calibrate the model, we need to specify individual and social preferences and the distributions of individual characteristics. Regarding individual preferences, we assume away income effects with $u(c) = c$ (as in e.g., Atkinson (1990) and Diamond (1998)) and the disutility of income is given by $v(y, w; \theta) = (y/w)^{1+\frac{1}{\theta}}$. This specification implies that θ is the *direct* taxable income elasticity. For social preferences, we assume Bergson-Samuelson preferences with $\phi(.; w, \theta) = log(.)$. We consider two dimensions of unobservable heterogeneity among individuals: the elasticity of taxable income θ and the skill w.

We calibrate the model from the subsample of singles without kids taken from the CPS data (2013). Beside heterogeneous income, people who earn the same income have distinct elasticities of taxable income θ. We proxy this difference by indexing each individual's elasticity θ on whether she is salary workers or self-employed. The latter 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, Knudsen, Kreiner, Pedersen, and Saez (2011)). We therefore assume two distinct and empirically plausible taxable income elasticities θ that depend on the occupation: $\theta = 0.6$ for the self-employed and $\theta = 0.2$ for the salary workers. We can then recover the skill distribution in each group from individuals' first-order condition (5).[27]

We investigate two scenarii, as shown in Figure 2. In the first scenario, that we call the multidimensional scenario, individuals differ along both their skill levels and their elasticities θ. In the second scenario, that we call the Mirrlees scenario, individuals differ only along their skill levels. Salary workers and self-employed have the same elasticity θ which is assumed to be equal to the sample mean of the direct elasticities θ in the first scenario.

To obtain the optimal tax profiles from our directly implementable tax formula (22a), we use an algorithm which is detailed in Appendix B. The optimal marginal tax rates in the two scenarii are shown on Figure 2 with the percentage of the marginal tax rate on the left-hand side vertical axis. The horizontal axis represents annual pre-tax income y in US dollars. We observe significant differences between the shape of tax profiles obtained in the Mirrlees scenario and in the multidimensional scenario. This is due to variations in the share of self-employed along the income distribution represented on the right-hand side vertical axis. This share matters

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

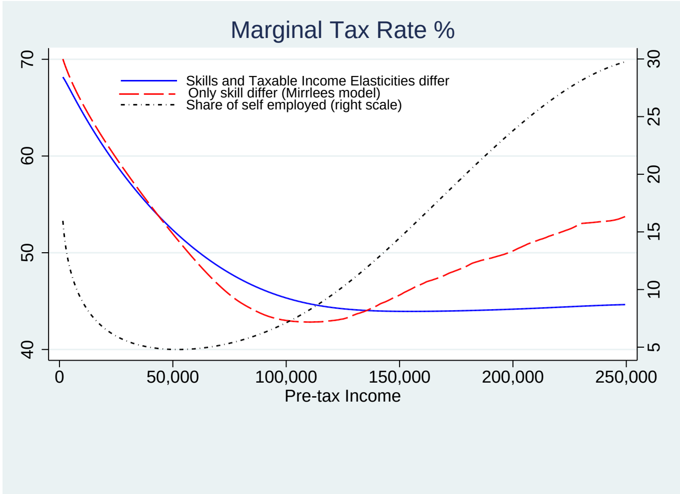

Figure 2: Simulations results

modifies the mean compensated elasticity only 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 10 percentage points when heterogeneous elasticities are taken into account. These numerical results put the stress on the need for including multidimensional heterogeneity when deriving optimal tax policies.

VII Concluding Comments

This paper proposes a method to characterize the nonlinear income tax schedules that one should implement when individuals differ not only in terms of skills, but also in terms of many other characteristics, in particular their taxable income elasticities. We obtain an optimal tax formula in terms of the structural primitives of the model (the individual and social preferences and the distributions of characteristics). It allows us to show that, despite multidimensional heterogeneity, optimal marginal tax rates remain positive under Benthamite and maximin social objectives. We also derive an elasticity-based optimal tax formula which is expressed in terms of empirically meaningful sufficient statistics, i.e. the mean compensated elasticity and income response, the income density and the social welfare weight at any income level. This elasticity-based formula is not only useful to discuss empirically the direction of desirable tax reforms, it also enables us to argue that calibrating the asymptotic tax formula from the estimations of the sufficient statistics among top income earners is misleading when the asymptotic Pareto parameter differ between low-elasticity and high-elasticity groups. The literature typically estimates the mean elasticity of labor supply among people at the very top of the income distribution (see Saez, Slemrod, and Giertz (2012) for a survey). However, to derive the optimal tax rate for very high income earners, one should rather estimate the distinct elasticities of labor supplies at the very top as well as the Pareto parameters by group of people having identical elasticity. We also numerically implement our structural tax formula using the CPS 2013. As an illustration, we consider that individuals differ along their skills and their taxable income elasticities. As a proxy for the latter, we use the indicator of whether individuals are salary workers or self-employed. This numerical exercise exemplifies how neglecting other sources of unobserved heterogeneity beside skills can be misleading.

The paper also shows that many policy relevant adverse selection problems can be solved with our method. It allows one to derive optimal tax policies in presence of real labor supply responses and tax avoidance, in terms of joint taxation of couples, and more generally for any tax problem for which the tax function depends on the sum of different sources of income. Beyond optimal taxation, it also applies to nonlinear pricing problems with consumers who differ along several unobserved dimensions. In our research agenda, we plan to use this framework to study the above set of policy relevant problems.

A Theoretical Proofs

A.1 Proof of Lemma 1

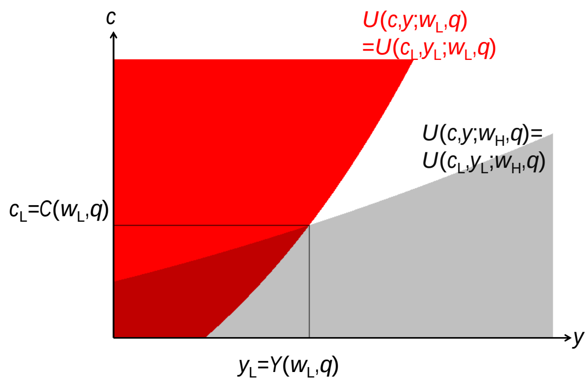

Figure 3: Proof of Lemma 1

Figure 3 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 are labeled $\mathscr{U}(c, y; w_L, \theta)$ and $\mathscr{U}(c, y; w_H, \theta)$ and 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 incentive constraints (12), 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 $\mathscr{U}(c,y;w_H,\theta)$ (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 3, which implies that $Y(w_L,\theta) \leq Y(w_H,\theta)$.

A.2 Proof of Lemma 2

Following, e.g., Salanié (2005), from the taxation principle, individuals choose the type w',θ' that they want to mimic, i.e. they solve:

$$\max_{w',\theta'} \quad \mathscr{U}\left(C(w',\theta'), Y(w',\theta'); w,\theta\right)$$

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 (13b). Differentiating in w both sides of $U(w,\theta) = \mathscr{U}\left(C(w,\theta), Y(w,\theta); w,\theta\right)$ leads to:

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

where the second equality follows (2). Using $\mathscr{U}_w = -v_w$, (13a) holds if and only if (13b) holds.

A.3 Proof of Lemma 4

We first show that there exists at most one allocation $(w,\theta) \mapsto (\underline{C}(w,\theta), \underline{Y}(w,\theta))$ that verifies Assumption 2 and such that $(\underline{C}(w,\theta_0), \underline{Y}(w,\theta_0)) = C(w,\theta_0), Y(w,\theta_0))$. We next show that this allocation verifies the whole set of incentive constraints (11).

To build the entire incentive compatible allocation $(w,\theta) \mapsto (\underline{C}(w,\theta), \underline{Y}(w,\theta))$, we must obviously choose $(\underline{C}(w,\theta_0), \underline{Y}(w,\theta_0)) = C(w,\theta_0), Y(w,\theta_0))$ for any skill level in the reference group θ_0.

For each group θ, $\underline{Y}(\cdot;\theta)$ verifies Assumption 2 if and only if its reciprocal $\underline{Y}^{-1}(\cdot;\theta)$ is differentiable with a strictly positive derivative and maps $\mathbb{R}_+$ into $\mathbb{R}_+$. Let then $y \in \mathbb{R}_+$ be an income level. As $Y(\cdot,\theta_0)$ satisfies 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

$\underline{W}(w,\theta)$ are assigned to the income level $y = Y(w,\theta_0)$.[28] Therefore, $\underline{Y}^{-1}(\cdot;\theta)$ must be defined by:

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

Hence, $\underline{Y}^{-1}(\cdot,\theta)$ is differentiable and is defined over $\mathbb{R}_+$. It admits a positive derivative everywhere and takes value on the whole $\mathbb{R}_+$ if and only if $\underline{W}(\cdot,\theta)$ does. Therefore, $\underline{Y}(\cdot,\theta)$ is a differentiable increasing function with positive derivatives that maps $\mathbb{R}_+$ onto $\mathbb{R}_+$.

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 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 that, given an incentive-compatible allocation $w \mapsto (C(w,\theta_0), Y(w,\theta_0))$ defined with the reference group that verifies Assumption 2, there exists *at most* a unique allocation $(w,\theta) \mapsto (\underline{Y}(w,\theta), \underline{C}(w,\theta))$ that can be incentive-compatible. We now verify that this allocation does verify the entire set of incentive constraints (11).

We first show that this allocation satisfies the within-group incentive constraints. Note that the allocation $(w,\theta) \mapsto (\underline{Y}(w,\theta), \underline{C}(w,\theta))$ is built in such a way that one has:

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

if and only if $\omega = \underline{W}(w,\theta)$ and (15b) holds. Differentiating in w both sides of the above equations, we obtain:

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

Rearranging terms leads to:

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

As $w \mapsto (C(w,\theta_0), Y(w,\theta_0))$ is assumed to verify the within-group incentive-compatible constraints in Equation (13b), 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 (13b). From Lemma 2, it thus verifies the within-group incentive constraints in Equation (12).

We now verify whether the inequality (11) 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

$$\underline{Y}(\omega,\theta) = \underline{Y}(w',\theta') \qquad \text{and} \qquad \underline{C}(\omega,\theta) = \underline{C}(w',\theta')$$

The incentive constraints in (11) are therefore equivalent to:

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

and the latter is verified as $w \mapsto (\underline{C}(w,\theta), \underline{Y}(w,\theta))$ also satisfies Equation (13b). Therefore, from Lemma 2, it satisfies the entire set of incentive constraints (12).

[28]Hence function $\underline{W}(\cdot,\theta)$ coincides with the pooling function $W(\cdot,\theta;\theta_0)$.

A.4 Equations (23a) and (23b)

When the unobserved heterogeneity is one-dimensional, only within-group incentive constraints (12) need to be considered. Under the first-order approach, only the first-order incentive constraint (13a) is considered. Taking $Y(w,\theta)$ as the control variable and $U(w,\theta)$ as the state variable, the Hamiltonian is:

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

Using (17), the necessary conditions are:

$$0 = \left(1 - \frac{v_y\langle w,\theta\rangle}{u'\langle w,\theta\rangle}\right) \cdot f(w|\theta) - q(w|\theta) \cdot v_{yw}\langle w,\theta\rangle \tag{40a}$$

$$-\dot{q}\,(w|\theta) = \left(\frac{\Phi_U\langle w,\theta\rangle}{\lambda} - \frac{1}{u'\langle w,\theta\rangle}\right) \cdot f(w|\theta) \tag{40b}$$

$$0 = q(0|\theta) \tag{40c}$$

$$0 = \lim_{w\mapsto\infty} q(w|\theta) \tag{40d}$$

Combining (40b) with (40d) leads to

$$q(w|\theta) = \int_w^\infty \left(\frac{\Phi_U\langle w,\theta\rangle}{\lambda} - \frac{1}{u'\langle w,\theta\rangle}\right) \cdot f(\omega|\theta) d\omega. \tag{40e}$$

Combining (2), (5), (40a) and (40e) leads to (23a). Combining (40c) with (40e) leads to (23b).

A.5 Proof of Proposition 3

Dividing (26a) by (26b) we get:

$$\frac{\varepsilon(w,\theta)}{\alpha(w,\theta)} = -\frac{v'_y\langle w,\theta\rangle}{w \cdot v''_{yw}\langle w,\theta\rangle}. \tag{41}$$

Plugging (26a) into (26c) 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:

$$\hat{\eta}(Y(w,\theta_0)) = Y(w,\theta_0) \cdot \frac{u''\langle w,\theta_0\rangle}{u'\langle w,\theta_0\rangle} \cdot \hat{\varepsilon}(Y(w,\theta_0)). \tag{42}$$

Let $y \in \mathbb{R}_+$. According to Assumption 2, there exists a single skill level w such that $y = Y(w,\theta_0)$. From (5), we know that:

$$1 - T'\langle w,\theta\rangle = \frac{v'_y\langle w,\theta\rangle}{u'\langle w,\theta\rangle}. \tag{43}$$

The term in the left-hand side integral of (22a) can be rewritten as:

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

The first equality is obtained using Equations (41). The second equality uses (27). It implies with (29) that Equation (22a) can be rewritten as:

$$\frac{T'\langle w,\theta_0\rangle}{1-T'\langle w,\theta_0\rangle}\cdot\hat{\varepsilon}\left(Y(w,\theta_0)\right)\cdot Y(w,\theta_0)\cdot\hat{h}(Y(w,\theta_0))=J(w) \tag{44}$$

where $J(w)$ is defined by the right-hand side of (22a). $J(\cdot)$ 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 (31) has been used. Deriving with respect to the skill w both sides of (14) 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 (27) and (44), $\dot{J}(w)$ can be rewritten as:

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

Using (42) and (32), we get:

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

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

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

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

$$J(w)=\int_{z\geq Y(w,\theta_0)}\left\{1-\hat{g}(z)-\hat{\eta}(z)\cdot T'\left(Y(z)\right)\right\}\cdot\hat{h}\left(Y(x,\theta)\right)\cdot dz. \tag{45}$$

Plugging (45) into (44) gives (33a). Combining (22b) and (45) leads to (33b).

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 (5) 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. As the algorithm actually considers a bounded income distribution, normalize each conditional skill-density $f(\cdot|\theta)$ to ensure that the algorithm consider a total mass of $\mu(\theta)$ over all income levels y_i.

3. Use (22b) to compute the Lagrange multipliers λ.

4. Use (22a) 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 (6).

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

References

AKERLOF, G. A. (1978): "The Economics of "Tagging" as Applied to the Optimal Income Tax, Welfare Programs, and Manpower Planning," *The 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.

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," 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.

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 University Press, Cambridge.

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," (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.

——— (2009): "The Optimal Income Taxation of Couples," *Econometrica*, 77(2), 537–560.

LAFFONT, J.-J., E. MASKIN, AND J.-C. ROCHET (1987): "Optimal Nonlinear Pricing with Two-Dimensional Characteristics," in *Information, Incentives and Economic Mechanism*, ed. by T. Grove, R. Radner, and S. Reiter, 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 (2014): "De Gustibus non est Taxandum: Heterogeneity in Preferences and Optimal Redistribution," NBER Working Paper 18332.

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 *Handbook of public economics, vol. 5*, ed. by M. F. Alan J. Auerbach, Raj Chetty, and E. Saez, vol. 5 of *Handbook of Public Economics*, chap. 7, pp. 391 – 474. Elsevier.

RAMSEY, F. P. (1927): "A Contribution to the Theory of Taxation," *Economic Journal*, 37, 47–61.

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.

——— (2014a): "Optimal Taxation with Rent-Seeking," NBER Working Papers 17035.

——— (2014b): "A Theory of Income Taxation under Multidimensional Skill Heterogeneity," NBER Working Papers 19822.

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

——— (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 (2013): "Generalized Social Marginal Welfare Weights for Optimal Tax Theory," NBER Working Papers 18835.

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

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

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

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 University Press, Oxford.